Statistics, Technology and Data Science for Economic and Social Development

Book of short papers of the ASA Bologna Conference

Supplement to Volume 35/3

Guest Editors:
Luigi Fabbris, Stefania Mignani, Giorgio Tassinari

Statistica Applicata – ITALIAN JOURNAL OF APPLIED STATISTICS

EDITORIAL TEAM

A.S.A CONTACTS
Principal Contact
Francesco Palumbo (Editor in Chief)
editor@sa-ijas.org

Support Contact
Domenico Vistocco (Editorial Manager)
ijas@sa-ijas.org

JOURNAL WEBPAGE
https://www.sa-ijas.org/ojs/index.php/sa-ijas

Statistica Applicata – Italian Journal of Applied Statistics is associated to the following Italian and international journals:

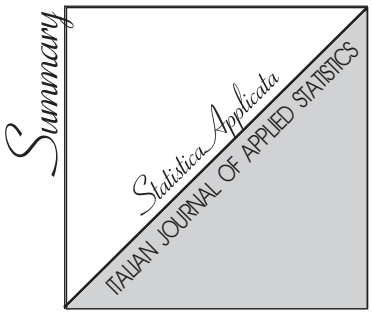
QTQM – Quality Technology & Quantitative Management (http/web.it.nctu.ew/~qtqm/)

SINERGIE – Italian Journal of Management

Pubblished: November 2023

CLEUP sc
'Coop. Libraria Editrice Università di Padova'
via G. Belzoni, 118/3 – Padova Italy
Phone +39 049 8753496 Fax +39 049 9865390
info@cleup.it – www.cleup.it – www.facebook.com/cleup

Statistica Applicata – Italian Journal of Applied Statistics is a four-monthly journal published by the Associazione per la Statistica Applicata (A.S.A.), Largo Gemelli 1 – 20123 Milano, Italy (phone + 39 02 72342904). Advertising:CLEUP sc, via G. Belzoni, 118/3 – 35128 Padova, Italy (phone +39 049 8753496 – Fax +39 049 9865390), email: info@cleup.it.
Rules for manuscript submission: https://www.sa-ijas.org/ojs/index.php/sa-ijas/about/submissions
Subscription: yearly €103.30; single copy €40.00; A.S.A. associates €60.00; supporting institutions: €350.00. Advertisement lower than 70%. Postal subscription Group IV, Milan. Forum licence n. 782/89. CLEUP SC on behalf of ASA., 7 March 2023.



Supplement
Vol. 35, Number 3

Introduction

This volume includes a selection of the papers presented at the ASA (Association for Applied Statistics) scientific conference held in Bologna from 6 to 8 September 2023. The volume keeps the title of the meeting "Statistics, Technology and Data Science for Economic and Social Development". The majority of papers are, in fact, devoted to economic statistics, but there are other papers dealing with other applied statistics fields and official statistics methodology.

All papers have been blindly reviewed by at least two independent experts belonging to the conference scientific committee. The published papers are, then, the outcome of a reviewing process, including suggestions, by academic statisticians and of the consequent \adjustment - both formal and substantive – by authors.

At the conference, the presentations were organized in session homogeneous according to content. The self-selection of some contributors and the severe selection from the reviewers drastically reduced the number of possible articles and so the content-based structure of the works went lost. Finally, we decided to publish the articles in alphabetic order according to the family name of the first author.

This volume was annexed as a supplement of a "normal" volume of the scientific journal Statistica Applicata – Italian Journal of Applied Statistics. Scholars can freely download the articles in an open access format, with the only obligation to cite the paper, possibly as follows:

Author(s) (Year). Paper title. In L. Fabbris, S. Mignani, G. Tassinari (eds) , Technology and Data Science for Economic and Social Development. Book of Short Papers of the ASA Bologna Conference. Padua: Cleup (pages).

The guest-editors:
Luigi Fabbris is president of the Association for Applied Statistics and former professor of Social statistics at the University of Padua, Italy.
Stefania Mignani is professor of Statistics at the Department of Statistical Sciences, the University of Bologna.
Giorgio Tassinari is professor of Economic statistics at the Department of Statistical Sciences, the University of Bologna.

Daiana Colledani (University of Padua)
Davide Colombo (Istat)
Cristina Davino (Federico II University of Naples)
Adriano Decarli (University of Milan)
Loretta Degan (Galgano Group Co., Milan)
Tonio Di Battista (G. D'Annunzio University of Chieti and Pescara)
Enrico di Bella (University of Genoa)
Angela Maria Digrandi (CNR-National Council for Research)
Simone Di Zio (G. D'Annunzio University of Chieti and Pescara)
Matteo Farné (University of Bologna)
Roberto Fasanelli (Federico II University of Naples)
Benito Vittorio Frosini (Sacred Heart Catholic University of Milan)
Antonio Giusti (University of Florence)
M. Gabriella Grassia (Federico II University of Naples)
Andrea Guizzardi (University of Bologna)
Domenica Fiordistella Iezzi (Tor Vergata University of Rome)
Salvo Ingrassia (University of Catania)
Corrado Lagazio (University of Genoa)
Filomena Maggino (La Sapienza University of Rome)
Paolo Mariani (University of Milan-Bicocca)
Marina Marino (Federico II University of Naples)
Roberta Palmieri (Istat)
Francesco Palumbo (Federico II University of Naples)
Chiara Parretti (Marconi University, Rome)
Alfonso Piscitelli (Federico II University of Naples)
Tommaso Pronunzio (Assirm)
Alessandra Righi (Istat, Rome)
Egidio Robusto (University of Padua)
Rosaria Romano (Federico II University of Naples)
Vito Santarcangelo (iInformatica)
Nicola Tedesco (University of Cagliari)
Venera Tomaselli (University of Catania)
Domenico Vistocco (Federico II University of Naples)
Giorgio Vittadini (University of Milan-Bicocca)
Mariangela Zenga (University of Milan-Bicocca)

doi.org/10.26398/asaproc.0001

Predictive modelling of length of domestic stays in Italy: A comparative study of machine learning techniques

Fabrizio Antolini[a], Samuele Cesarini[a]
[a] Department of Business Communication, University of Teramo, Teramo, Italy

1. Introduction

Tourism is a key economic sector, and accurately predicting the duration of domestic tourists' stays is pivotal for tourism planning and development (Li et al., 2018; Gössling et al., 2018). Various authors have underscored the significance of stay length on the tourist incomes generated at destinations (Antolini et al., 2023; Aguiló et al., 2018; Park et al., 2020; Marrocu et al., 2015).

In this field, the contributions of Thrane have guided the studies and methodologies for predicting the duration of tourist stays. Thrane and Farstad (2012) advocated the shift from ordinary least square (OLS) regression to survival models to better investigate tourists' length of stay. This perspective was subsequently adopted by Aguilar and Díaz (2019), Gutiérrez et al. (2020) and Almeida et al. (2021). Furthermore, several researchers have utilised truncated count data regression methods, including Esiyok et al. (2018), Alén et al. (2014) and Alegre et al. (2011). However, Thrane (2015) demonstrated in his study that count data regression models should be considered obsolete as they fail to provide superior estimates compared with traditional OLS models and class survival models.

The approach to these studies has recently shifted, as evidenced by the works of Jackman and Naitram (2023) and Lee and Kim (2021). Employing regression tree models, their research examined the influence of tourists' socio-demographic profiles, trip-related characteristics, distance and economic conditions on the prediction of stay duration. Their analyses revealed substantial heterogeneity, often obscured in simplistic parametric methodologies. Furthermore, an expected benefit of the 'divide-and-conquer' nature of tree methods is that they partition the cases into more homogeneous classes before giving the prediction, further enhancing the precision of these models. The superior 'flexibility' of tree-based methods, eschewing the prerequisite of pre-identifying a specific model, endows these approaches with enhanced predictive capacity for the socio-demographic determinants.

Additionally, to our knowledge, artificial neural network (ANN) models have rarely been utilised for predicting average duration in the tourism context, as their application has been predominantly favoured in the medical field. Moreover, only very few microeconomic studies on tourist demand have focused on analysing this key element.

This paper presents an innovative approach to estimating the duration of domestic tourists' stays in Italy utilising microdata from the 2019 Viaggi e Vacanze survey conducted by the Italian National Institute of Statistics (ISTAT, 2022). Our primary objective is to develop predictive models employing decision trees and deep learning methodologies, incorporating the most pertinent predictive variables identified in scholarly literature. Leveraging the capabilities of machine learning, which is renowned for discerning complex patterns, this study seeks to surpass the constraints of conventional models, providing an accurate and dependable predictive tool.

The structure of the paper is as follows. Section 2 offers a comprehensive overview of the microdata utilised and the variables included in the model and Section 3 briefly outlines the models used. Section 4 details the results, emphasising the root mean square error (RSME) and the predictive efficiency of the model. This leads to the main findings, along with practical and theoretical implications for tourism policymakers.

2. Data

According to Lin et al. (2020), explanatory variables that influence tourist behaviour can be categorised into four main groups: sociodemographic, economic, travel-related, and psychological. The selection of the variables depends on the availability of relevant information, with microdata offering insights into sociodemographic, travel-related and psychological variables but often lacking consistent economic data due to non-response issues. This has led researchers to utilise proxy variables like occupation and education level, as suggested by Marcussen (2011).

The microdata used in this study refers to the 'Viaggi' database included in the statistical survey Viaggi e Vacanze conducted by ISTAT. The reference year is 2019, and by considering only residents who engaged in domestic tourist trips, a total of 3,001 units. Due to the sample nature of the survey, it was possible to estimate the entire Italian domestic tourist population size with the use of expansion coefficients (COEV). Consequently, a dataset containing 48 million and 410 thousand trips was reconstructed.

Based on the data availability and the relevant literature, the following categories and variables were utilised: sociodemographic factors – nationality, sex, age, marital status and region of origin; economic factors – occupation and education; travel-related factors – region of destination, month of travel, accommodation, transport, transport organisation, accommodation organisation, group of travel party and principal activities, and psychological factors – travel motivation.

In machine learning, categorical variables need to be transformed into numerical representations, often using techniques like one-hot encoding. This process converts each category into a binary vector representation, expanding the number of columns in the dataset (from 16 to 128). By encoding the categorical variables as numerical representations, machine learning models can effectively handle complex relationships and patterns. In this study, considering the nature of the explanatory variables (all dummies), partial least squares regression (PLSR) was selected as a reference linear model to reduce dimensionality and address predictor multicollinearity for comparison with the proposed machine learning methods.

3. Method

3.1 Partial least squares

In essence, *PLSR* decomposes the set of independent variables $\boldsymbol{X}$ (the 128 dummy variables) into orthogonal scores $\boldsymbol{T}$ and loadings $\boldsymbol{P}$ according to the relationship $\boldsymbol{X}=\mathbf{TP}$, aiming to explain the maximum covariance between $\boldsymbol{X}$ and $\mathbf{y}$ (Wold et al., 1984). This methodology regresses the dependent variable $\mathbf{y}$ (length of stay) not directly on $\boldsymbol{X}$ but rather on the leading columns of the scores $\boldsymbol{T}$. The optimal number of components a is typically determined through *cross-validation*.

3.2 Random forest

The presence of nonlinearity between the length of stay and independent variables, along with the intricate nature of their relationships, argues against making a priori assumptions through the formulation of parametric models and favours the adoption of tree-based models. The *RF* algorithm improves predictive performance by combining (*ensemble*) multiple decision trees. It starts by randomly selecting a sample from the training dataset to build the first tree, followed by creating additional trees using random subsamples with replacement (*bootstrapping* or *bagging*). Each tree has a unique set of observations and variables for training. *Random feature selection* is used to partition the data within each tree, reducing overfitting and promoting diversity. The model makes predictions by aggregating the individual tree predictions.

In regression, the final prediction is the average of all tree predictions. For further insights into the methodology, it is recommended to refer to specialised texts on the subject (Breiman, 1996, 2001; James et al., 2020).

3.3 Gradient boosting

Boosting is an alternative approach to enhance decision tree predictions, where trees are sequentially grown. Each tree is trained using information from previously grown trees and a modified version of the original dataset. The boosting method takes a different approach compared to fitting a single large decision tree. Instead, it incrementally enhances the model by fitting small trees to the residuals of the current model. This iterative process targets areas where the model performs poorly, and the shrinkage parameter controls the pace of learning.

Unlike *bagging*, which is independent for each tree, boosting relies on the influence of previously grown trees in constructing subsequent trees in the ensemble (James et al., 2020).

3.4 Deep learning

Finally, as highlighted earlier, we also aim to propose an approach based on *ANNs*, a class of machine learning models inspired by the structural and functional properties of the human brain (Kriegeskorte and Golan, 2019; LeCun et al., 2015). A typical artificial neuron with *n* input dendrites can be represented by the following formula:

$$y(x) = f(\sum_{i=1}^{n} w_i\, x_i)$$

In *ANN*s, weights (*w*) assigned to the inputs determine their influence on the overall input signal. The resulting signal is then processed by an activation function *f(x)*, yielding the output axon (Lantz, 2019). Deep learning utilises neural networks with multiple hidden layers to address complex problems (*multilayer perceptron or MLP*). The learning process involves *backpropagation*, where prediction errors are propagated through the network to adjust the weights iteratively until the network's performance reaches a satisfactory level (Rumelhart et al., 1986). Architecture selection and hyperparameter tuning are critical for optimising performance (Lantz, 2019).

To evaluate the predictive performance of the models, a train set and a test set were created, randomly comprising 75% and 25% of the total observations, respectively. The test set is used to make predictions for unseen data. Given the quantitative nature of the length of stay, the correlation between the model's predictions and the actual values in the test set was measured to gauge the strength of the linear relationship and evaluate the accuracy of the model, as suggested in Lantz (2019). Additionally, a comparison between models was conducted by assessing the standard deviation of the residuals (RMSE) they generated.

4. Results and discussion

The estimated results from the *PLS* model trained with 75% of random observations are summarised in Table 1. Specifically, the dimensions (or latent factors) that explain most of the variations in **y** and **X** are identified. The efficiency and reliability of the model are assessed by the percentage of the explained variance (R^2y), goodness of fit (R^2) and RMSE. These latter measures were computed by applying the model to the remaining 25% of observations (test set), as it is done for machine learning models.

Based on *cross-validation*, it was decided to use PLSR with 24 components (a straightforward transformation of the observed data into principal components). This choice was motivated by a satisfactory level (60%) of explained variance of the original predictors (128) and 38% of explained variability of the target variable. Increases in the number of components do not lead to improvements in the latter measure. Using this model, $\boldsymbol{R^2}$ is 0.34 and RMSEP is 4.30 compared with a mean length of stay of 5.32 days. These findings highlight the limitations of the parametric and linear models like PLS in both explaining and predicting this variable. Although successful in

diminishing the number of predictors required for modelling, it fails to provide sufficient predictive power. Moreover, the linear correlation between the predicted and original values is only 0.58. On the other hand, machine learning models, despite their superior predictive performance, do not prioritize dimensionality reduction in their algorithmic structure.

Table 1: PLS model results

Model comp.	R^2X	R^2X *(cum)*	R^2y	R^2y *(cum)*	R^2	*RMSE*
1	4.05	4.05	22.75	22.75	0.20	4.73
2	3.67	7.72	7.07	29.82	0.27	4.52
-	-	-	-	-	-	-
24	0.81	60.36	0	38.04	0.34	4.30

The RMSE and the correlation between the original and predicted values from the three previously described machine learning models are presented in Table 2.

Table 2: Machine learning model results

Model	**RMSE**	**ρ**
Random Forest	2.08	0.93
Gradient Boosting	1.09	0.98
MLP	0.65	0.99

The implemented *RF* algorithm consists of 500 regression trees. To mitigate overfitting, limits were imposed on the maximum number of nodes (300) and the minimum number of observations in each leaf (100) of the regression trees. These limits ensured a moderately complex model, avoiding small subgroups susceptible to noise. Additionally, the model randomly selected 43 out of the 128 explanatory variables for splitting the observations in each tree, enhancing predictive capability and reducing the impact of spurious or uninformative variables (Probst et al., 2019). The *Gradient Boosting* algorithm, on the other hand, trained the regression tree 500 times to progressively reduce the RMSE using a learning rate of 0.05. The tree depth, number of usable variables and number of observations were the same as in the *RF* algorithm. Finally, the *MLP network* employed a backpropagation algorithm to train a neural network with two hidden layers consisting of 16 and 8 neurons, respectively. The output layer of the network utilised a *linear activation function* (Sharma et al., 2020) to predict the domestic Italian touristic length of stay. In this architecture, information flows backwards between the neurons to minimise the error between the prediction and the actual values.

The machine learning models trained on a designated training set and then generalised on a test set of previously unseen observations exhibited strong positive linear relationships. This is evidenced by high correlation coefficients (0.93, 0.98 and 0.99) between the predicted and actual values in the test set. Such results underscore the models' capability to learn from the provided data effectively and suggest their potential for making robust predictions about the duration of stay in the domestic Italian tourism context. Additionally, the relatively low RMSE values (2.08, 1.09 and 0.65) suggest that the model's predictions closely align with the actual length of stay, indicating an accurate representation of underlying data patterns and trends.

The findings of this study hold practical implications for policymakers in the tourism industry. The proposed models employ 128 dummy variables, allowing for the estimation of tourists' stay duration by inputting binary values (1 or 0) based on the target tourist. Unlike traditional linear models that fail to yield satisfactory results, the accurate prediction of tourists' stay duration

obtained through machine learning models can inform and guide decision-making processes aimed at prolonging their visits to different destinations. By leveraging the insights provided by the predictive models, policymakers can develop targeted strategies and interventions to optimise hospitality services, enhance tourism promotion efforts and curate immersive experiences that encourage tourists to prolong their stays. These results offer valuable information for formulating effective policies and initiatives aimed at maximising the economic benefits derived from tourism.

References

Aguilar, M.I., Díaz, B. (2019). Length of stay of international tourists in Spain: A parametric survival analysis. *Annals of Tourism Research*, **79**, 102768.

Aguiló, E., Rosselló, J., Vila, M. (2017). Length of stay and daily tourist expenditure: A joint analysis. *Tourism Management Perspectives*, **21**, pp. 10-17.

Alén, E., Nicolau, J.L., Losada, N., Domínguez, T. (2014). Determinant factors of senior tourists' length of stay. *Annals of Tourism Research*, **49**, pp. 19-32.

Alegre, J., Mateo, S., Pou, L. (2011). A latent class approach to tourists' length of stay. *Tourism Management*, **32**(3), pp. 555-563.

Almeida, A., Machado, L.P., Xu, C. (2021). Factors explaining length of stay: Lessons to be learnt from Madeira Island. *Annals of Tourism Research Empirical Insights*, **2**(1), 100014.

Antolini, F., Cesarini, S., Simonetti, B. (2023). Italian tourist expenses' determining factors: A machine learning approach. *Quality & Quantity* in Methods in the tourism and hospitality industry, *in press*.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, **24**, pp. 123-140.

Breiman, L. (2001). Random forests. *Machine Learning*, **45**, pp. 5-32.

Esiyok, B., Kurtulmuşoğlu, F.B., Özdemir, A. (2018). Heterogeneity in the determinants of length of stay across middle age and senior age groups in thermal tourism. *Journal of Travel Tourism Marketing*, **35**(4), pp. 531-540.

Gössling, S., Scott, D., Hall, C.M. (2018). Global trends in length of stay: Implications for destination management and climate change. *Journal of Sustainable Tourism*, **26**(12), pp. 2087-2101.

Gutiérrez, A., Miravet, D., Saladié, Ò., Clavé, S.A. (2020). High-speed rail, tourists' destination choice and length of stay: A survival model analysis. *Tourism Economics*, **26**(4), pp. 578-597.

ISTAT (2022). Viaggi e Vacanze: File ad uso pubblico. https://www.istat.it/it/archivio/178695 (last access: 06.03.2023).

Jackman, M., Naitram, S. (2023). Segmenting tourists by length of stay using regression tree models. *Journal of Hospitality and Tourism Insights*, **6**(1), pp. 18-35.

James, G., Witten, D., Hastie, T., Tibshirani, R. (2020). *An Introduction to Statistical Learning with Applications in R*. Springer New York, NY.

Lantz, B. (2019). *Machine Learning with R: Expert Techniques for Predictive Modeling*. Packt publishing ltd. ISBN: 9781788295864.

LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *Nature*, **521**(7553), pp. 436-444. 10.1038/nature14539.

Lee, Y., Kim, D.Y. (2021). The decision tree for longer-stay hotel guest: The relationship between hotel booking determinants and geographical distance. *International Journal of Contemporary Hospitality Management*, **33**(6), pp. 2264-2282.

Li, K.X., Jin, M., Shi, W. (2018). Tourism as an important impetus to promoting economic growth: A critical review. *Tourism Management Perspectives*, **26**, pp. 135-142.

Lin, V.S., Qin, Y., Li, G., Wu, J. (2020). Determinants of Chinese households' tourism consumption: Evidence from China family panel studies. *International Journal of Tourism Research*, **23**(4), pp. 542-554.

Marcussen, C.H. (2011). Determinants of tourist spending in cross-sectional studies and at Danish destinations. *Tourism Economics*, **17**(4), pp. 833–855.

Marrocu, E., Paci, R., Zara, A. (2015). Micro-economic determinants of tourist expenditure: A quantile regression approach. *Tourism Management*, **50**, pp. 13-30.

Kriegeskorte, N., Golan, T. (2019). Neural network models and deep learning. *Current Biology*, **29**(7), pp. 231-236.

Park, S., Woo, M., Nicolau, J.L. (2020). Determinant factors of tourist expenses. *Journal of Travel Research*, **59**(2), pp. 267-280.

Rumelhart, D.E., Hinton, G.E., Williams, R.J. (1986). Learning representations by back-propagating errors. *Nature*, **323**(6088), pp. 533-536.

Sharma, S., Sharma, S., Athaiya, A. (2020). Activation functions in neural networks. *International Journal of Engineering Applied Sciences and Technology*, **4**(12), pp. 310-316.

Thrane, C. (2012). Analysing tourists' length of stay at destinations with survival models: A constructive critique based on a case study. *Tourism Management*, **33**, pp. 126–132.

Thrane, C., Farstad, E. (2012). Tourists' length of stay: The case of international summer visitors to Norway. *Tourism Economics*, **18**, pp. 1069–1082.

Thrane, C. (2015). Research note: The determinants of tourists' length of stay: Some further modelling issues. *Tourism Economics*, **21**(5), pp. 1087-1093.

Wold, S., Ruhe, A., Wold, H., Dunn, Iii, W.J. (1984). The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM Journal on Scientific and Statistical Computing*, **5**(3), pp. 735-743.

Probst, P., Wright, M.N., Boulesteix, A.L. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, **9**(3), e1301.

doi.org/10.26398/asaproc.0002

Multivariate permutation-based ranking: An application to college basketball data

Rosa Arboretti [a], Nicolò Biasetton [b], Riccardo Ceccato [b], Livio Corain [b], Giacomo Vezzosi [b]

[a] Department of Civil, Environmental and Architectural Engineering, University of Padua, Padua, Italy

[b] Department of Management and Engineering, University of Padua, Vicenza, Italy

1. Introduction

Players in a basketball team can perform quite differently, even when playing in the same position. It is therefore fundamental for team managers to be supported in their technical and tactical decisions with insights from objective statistics. Data collected during matches and analyzed using appropriate techniques assist managers in choosing the best team for each match. The aim of this paper is to introduce a multivariate nonparametric procedure, based on the work by Arboretti et al. (2014), to enhance decision-making in this context.

The ranking procedure proposed by Arboretti et al. (2014) is indeed a valuable and versatile solution for ranking players according to their performances on the court. This approach exploits multiple pairwise comparisons and uses the achieved p-values to compute a ranking of C multivariate (or univariate) populations $\mathbf{X}_c$. The user is firstly required to identify appropriate statistical tests to assess multiple systems of hypotheses in the form:

$$\begin{cases} H_0: \mathbf{X}_c \overset{d}{=} \mathbf{X}_k \\ H_1: \mathbf{X}_c \overset{d}{>} \mathbf{X}_k, \end{cases} c, k = 1, \dots, C, c \neq k. \tag{1}$$

In the context of basketball, we have multiple performance measures for each player that we want to take into account in our comparisons. Given the multivariate nature of our data and the absence of information on data distribution, we relied on the NonParametric Combination (NPC) technique (Pesarin and Salmaso, 2010) to perform comparisons between pairs of players. NPC is in fact a highly flexible permutation-based methodology which proved to be quite successful in multivariate scenarios.

Let us assume that in a generic comparison (P_c,P_k) we want to compare the performance of two players, namely P_c and P_k. For each player we have information on V variables (i.e., performance indexes). The NPC methodology initially decomposes system of hypotheses (1) into V sub-systems (i.e., one for each performance index). The generic v^{th} sub-system of hypotheses (v = 1, ..., V) is as follows:

$$\begin{cases} H_{0v}: X_{cv} \overset{d}{=} X_{kv} \\ H_{1v}: X_{cv} \overset{d}{>} X_{kv} \end{cases} c, k = 1, \dots, C, c \neq k. \tag{2}$$

The user now needs to choose an appropriate test statistic, such as the difference in mean or the Anderson-Darling test statistic proposed in Pesarin and Salmaso (2010). For the problem at hand we choose the difference in medians to take into account the presence of possible outliers and the numerical nature of the data: $T_{c,k}^{v} = \text{median}_{cv} - \text{median}_{kv}$.

The NPC technique then addresses each sub-system individually and returns V partial p-values achieved through permutation. In doing so, the same permutation mechanism is applied

for the computation of each partial *p*-value (Pesarin and Salmaso, 2010) so that we implicitly take into account the existing dependence between variables. This allows us to compare players in terms of individual performance indexes.

A combination step is then required to achieve an overall evaluation of player performances. The user must firstly choose an appropriate combining function, a choice driven mainly by the number of available sub-problems in which the null hypothesis is expected to be rejected and by the correlation between variables (Langthaler et al., 2022). In this study, we rely on Fisher's combining function because it has been shown to be the best solution in many scenarios. A combination step is therefore undertaken and a combined *p*-value is achieved. This p-value merges the insights provided by the partial *p*-values and can be used to address the multivariate problem of interest (see system 1).

After using NPC to perform all possible directional pairwise comparisons (P_c, P_k), $c,k = 1,...,C$, $c \neq k$ and retrieve the related $C\times(C-1)$ p-values, we move on to the second step of the ranking procedure. Let us suppose that Λ denotes the $C\times C$ matrix containing all the combined p-values $\lambda^{(c,k)}$, $c,k = 1,...,C$. A multiplicity correction is applied to the matrix of p-values Λ and the matrix of adjusted p-values Λ_{adj} is achieved using commonly accepted corrections such as the Bonferroni-Holm-Shaffer (Shaffer, 1986), the Bonferroni-Holm (Holm, 1979), or the Benjamini-Hochberg (Benjamini and Hochberg, 1995). A matrix Z is then created to keep track of the significant comparisons, with $Z_{c,k} = 1$ if $\lambda_{adj}^{(c,k)} \leq \alpha/2$ and $Z_{c,k} = 0$ if $\lambda_{adj}^{(c,k)} > \alpha/2$, where α is the desired significance level. A vector of downward rank estimates is firstly computed as:

$$r_d^k = 1 + \sum_{c=1}^{C} Z_{c,k}\,, k = 1, \dots, C,$$

counting how many populations are stochastically larger than the k^{th} population. A vector of upward rank estimates is calculated as:

$$r_u^c = 1 + \#\left[(C - \sum_{k=1}^{C} Z_{c,k}) > (C - \sum_{k=1}^{C} Z_{c',k})\right], c' = 1, \dots, C, c' \neq c, c = 1, \dots, C,$$

where # means number of times. In this case, we are counting how many populations are stochastically smaller than the c^{th} population. The final vector of ranking estimates is computed using both the downward and upward estimates as:

$$r^c = 1 + \#\left[\frac{(r_u^c + r_d^c)}{2} > \frac{(r_u^k + r_d^k)}{2}\right],$$

$k = 1, ..., C$, $c \neq k$, $c = 1, ..., C$, and can be used to identify the best population according to the considered hypotheses.

2. Case study description

The case study analyzed in this research involves a total of 22 matches (14 from the regular season and 8 from the playoff round) played by the women's basketball team of the Padua University Sports Centre (CUS) during the 2022-2023 season. For each match and each player who took part in the match (both as a starter and a substitute), several pieces of information were collected, and in particular: shooting statistics (points scored; field goals attempted and made, distinguishing between 2 and 3-point shots; free throws attempted and made; throws attempted and made from different areas of the court), assists, offensive and defensive rebounds, blocks, steals, fouls, turnovers and minutes played.

As described in Corain et al. (2019) and Metulini and Gnecco (2023), several indexes were computed to evaluate player performance and allow more systematic comparisons. Based on data collected during the matches, the following indexes were computed:

- Floor impact counter (FIC) (Ferrario, 2021). This index represents a player's evaluation rating with greater importance given to the construction of offensive actions, in particular to offensive rebounds and assists. Other statistics are considered in the computation, such as points, defensive rebounds, steals and blocks (positive), but also missed field goals, missed free throws, turnovers and fouls committed (negative).
- Performance index rating (PIR) (Cene et al., 2018). This metric is useful to assess the efficiency of players in a match and is computed by summing up all the positive actions (i.e., points, assists, total rebounds, steals, blocks, received fouls) and subtracting all the negative actions (i.e., missed field goals, missed free throws, turnovers, fouls committed) performed by the player in the match.
- Adjusted field goal (AFG) (Cene et al., 2018). This index allows us to measure the shooting ability of each player by considering points scored as well as free throws made and attempted field goals.
- Player impact estimate (PIE) (Senatore et al., 2022). This index is useful to compare players and teams and is computed for each player, each team, and each match. It provides a measure of the overall contribution of the player and the team to each match.
- Offensive efficiency (OE) (Lee and Page, 2021). This index measures the quality of a player's offensive contribution and is computed as the ratio between the number of profitable offensive possessions in which the player was involved (sum of field goals made and assists) and the player's total number of potential end-of-possession situations (sum of attempted field goals, assists and turnovers, subtracting offensive rebounds).
- Efficient offensive production (EOP) (Lee and Page, 2021). This index considers the contribution of an assist to the final points scored, estimated using a tuning parameter equal to 0.75, and is computed as: $(0.75 \times \text{assists} + \text{points}) \times \text{OE}$, thus derived from the previous measure.

The conducted analysis, which aims to extend the NPC technique to the analysis of c samples with a dedicated ranking procedure, looked at three players from the same team (identified as Player 1, Player 2 and Player 3) and the previously described performance indexes were computed. The number of matches and minutes played by the three players were comparable therefore this was not taken into consideration. To apply the ranking procedure, only data from matches in which all three players took part at the same time were considered (13 matches out of 14 for the regular season, and all 8 matches for the playoff round).

Figure 1 shows the boxplot of the distributions of the performance indexes computed using data from the regular season and the playoffs. In terms of median value, Player 3 is seen to have performed better overall than her teammates during the regular season. Similarly, during the playoffs, Player 3 seems to outperform the others in terms of median value for almost all computed indexes.

3. Findings and conclusion

The results of the permutation-based ranking procedure for both regular season and playoff matches are reported in Table 1. As anticipated in the descriptive analysis, Player 3 has the best ranking within the group, both in the regular season and in the playoffs, for all indexes (in terms of the median value) as well as globally.

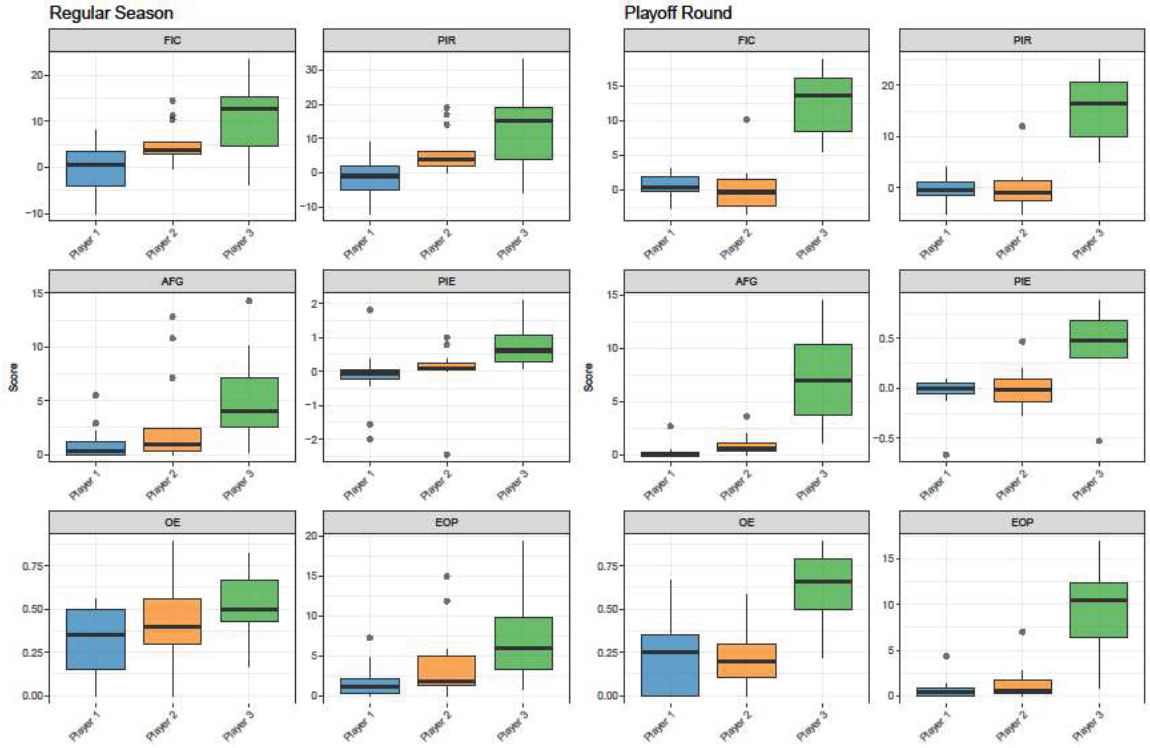


Figure 1: Boxplot of the distributions of the performance indexes per player for both regular season and playoff matches.

Players 1 and 2 have the same global ranking therefore their overall performance can be considered quite similar, even if the results obtained for individual indexes show a better ranking for Player 2 in the regular season. Additionally, in this period, the values obtained for the OE index show substantial performance similarities, therefore all three players have the same ranking if only this index is considered.

Table 1: Ranking of players for each index and globally, computed both on regular season and playoff matches.

	Regular Season			Playoff Round		
Index	Rank Player 1	Rank Player 2	Rank Player 3	Rank Player 1	Rank Player 2	Rank Player 3
FIC	3	2	1	2	2	1
PIR	3	2	1	2	2	1
AFG	3	2	1	2	2	1
PIE	3	2	1	2	2	1
OE	1	1	1	2	2	1
EOP	3	2	1	2	2	1
Global	3	2	1	2	2	1

This study shows the usefulness of the ranking procedure by Arboretti et al. (2014) for the analysis of sports data. In particular, it serves as a valuable tool for ranking players according to their performances, aiding team managers in selecting the best possible team. The proposed procedure can easily be extended to the analysis of larger groups of players and performance indexes.

Acknowledgement

This study was carried out within the framework of the MICS (Made in Italy - Circular and Sustainable) extended partnership and received funding from the European Union NextGeneration EU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) - MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.3 - D.D. 1551.11-10-2022, PE00000004). This article reflects only the authors' views and opinions, and neither the European Union nor the European Commission can be considered responsible for them. This study was also supported by the following projects: BIRD 2022 project entitled "Fuzzy theory in Unsupervised Machine Learning algorithm and Sentiment Analysis"; PNRR - MOST - Centro Nazionale per la MObilità soSTenibile, Spoke8: Maas & Innovative services.

References

Arboretti, R., Bonnini, S., Corain, L., Salmaso, L. (2014). A permutation approach for ranking of multivariate populations. *Journal of Multivariate Analysis*, **132**, pp. 39–57.

Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, **57**(1), pp. 289–300.

Cene, E., Parim, C., Ozkan, B. (2018). Comparing the performance of basketball players¨ with decision trees and TOPSIS. *Data Science and Applications*, **1**(1), pp. 21–28.

Corain, L., Arboretti, R., Ceccato, R., Ronchi, F., Salmaso, L. (2019). Testing and ranking on round-robin design for data sport analytics with application to basketball. *Statistical Modelling*, **19**(1), pp. 5–27.

Ferrario, A. (2021). *Basketball Analytics: The Use of Data Science to Describe and Predict the Performance of an NBA Team*. Master's Thesis, Università degli Studi di Milano Bicocca.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**(2), pp. 65–70.

Langthaler, P.B., Ceccato, R., Salmaso, L., Arboretti, R., Bathke, A.C. (2022). Permutation testing for thick data when the number of variables is much greater than the sample size: Recent developments and some recommendations. *Computational Statistics*, **38**(1), pp. 1–32.

Lee, D.-J., Page, G.L. (2021). Big data in sports: Predictive models for basketball player's performance. *Mathematics in Industry Reports* (https://www.academia.edu/70747795/Big_Data_in_Sports_Predictive_Models_for_Basketball_Players_Performance)

Metulini, R., Gnecco, G. (2023). Measuring players' importance in basketball using the generalized Shapley value. *Annals of Operations Research*, 325(1), pp. 441–465.

Pesarin, F., Salmaso, L. (2010). *Permutation Tests for Complex Data: Theory, Applications and Software*. John Wiley & Sons.

Senatore, J.V., Fellingham, G., Lamas, L. (2022). Efficiency and productivity evaluation of basketball players' performance. *Motriz: Revista de Educação Fısica*, **28**.

Shaffer, J. P. (1986). Modified sequentially rejective multiple test procedures. *Journal of the American Statistical Association*, **81**(395), pp. 826–831.

doi.org/10.26398/asaproc.0003

The *young side of agriculture*: A focus on the Liguria case

Chiara Baggetta[1], Agnese Sechi[1], Yari Vecchio[2]
[1]Department of Economics, University of Genoa, Genoa, Italy
[2]Department of Veterinary Medical Science, University of Bologna, Bologna, Italy

1. Introduction, data and summary statistics

Agriculture is a sector of fundamental importance to the Italian economy, but it has faced significant challenges in recent decades, especially due to an aging farming population and the agricultural production fabric. What has been observed is a paradigm shift: from the cultivation of fields and the raising of livestock to a multiplicity of related activities that over time have shaped the concept of *multifunctionality*. One might hypothesize that young farmers are more adept at actively participating in this transformation compared to others. Not only do they bear fewer established routines, but they also possess the ability to leverage new ICT technologies and fulfil their desire for greater autonomy in the inputs and outputs market (Milone and Ventura, 2019).

However, while it is true that data about youth agricultural entrepreneurship is encouraging, a closer look at the picture shows that not all Italian regions are experiencing the same trend in generational turnover. To understand properly the generational renewal in the agricultural sector, it is insightful to observe data from the Istat 7th agriculture census. In 2020, the proportion of farm leaders aged up to 40 years old decreased to 9.3 per cent from 11.5 per cent with respect to 2010. This decline indicates a minimal influx of young individuals into the agriculture sector. According to Fargione et al. (2022), new agricultural establishments supported by the CAP (Common Agricultural Policy) funds would have been approximately 20 thousand during the last seven-year programming period (2014-2020). This translates to an average of 3 thousand young people per year. In comparison, France, a country with fewer farms than Italy (less than 30 per cent), boasts around 9 thousand young farmers.

The left panel of Figure 1 illustrates the percentage of agricultural young entrepreneurs in Italy. It is possible to observe that Valle d'Aosta, Sardinia, and Sicily are regions with a higher percentage of young agricultural entrepreneurs compared to the total number of enterprises, accounting for about 30, 21.5 and 21 per cent, respectively. Lazio and Abruzzo follow closely, both with almost 27 per cent of young agricultural entrepreneurs.

A plausible reason of this trend would lie in the inadequacy of regulatory instruments designed to facilitate generational turnover and the associated funds. Indeed, to obtain public support for starting a farming business, young people mainly draw on resources from the CAP, while a more marginal role is played by instruments fielded by the Institute of Agricultural and Food Market Services (ISMEA). More specifically, at the European level, the two main measures which interest young agricultural entrepreneurship are the measure 112 (for the programming period 2007-2013) and the sub measure 6.1 (for the programming period 2014-2020). The former is also known as the *first establishment* and consists of a one-time premium for farmers up to 40 years of age who are setting up for the first time in an agricultural enterprise as head of the farm, possess adequate professional qualifications and skills and apply for it through participation in the appropriate calls (Piras, 2018). While the latter is a reinforcement of the first establishment measure. The institution responsible for assisting young agricultural enterprises in Italy is ISMEA.

The right panel of Figure 1 shows the percentage of young farmers who access to dedicated funds: it is evident how the percentage dramatically falls with respect to the percentage of young entrepreneurs, confirming the difficulties in accessing these funds. For instance, Valle d'Aosta stands out with 5.02% of young entrepreneurs who access to dedicated funds, followed by Piedmont with 3.8% and Veneto with 2.17%.

This paper seeks to investigate the influence of two factors on business performance: the presence of young managers and the access to funding (from both European and national sources) specifically aimed at promoting young agricultural entrepreneurship. In this analysis, we take into account Liguria, a sea region located in the north-east of Italy. It is an interesting region, both morphologically and in terms of population. If in fact Liguria has the oldest inhabitants in Italy, the Figure 1 shows that the percentage of young farmers is much higher than in other regions.

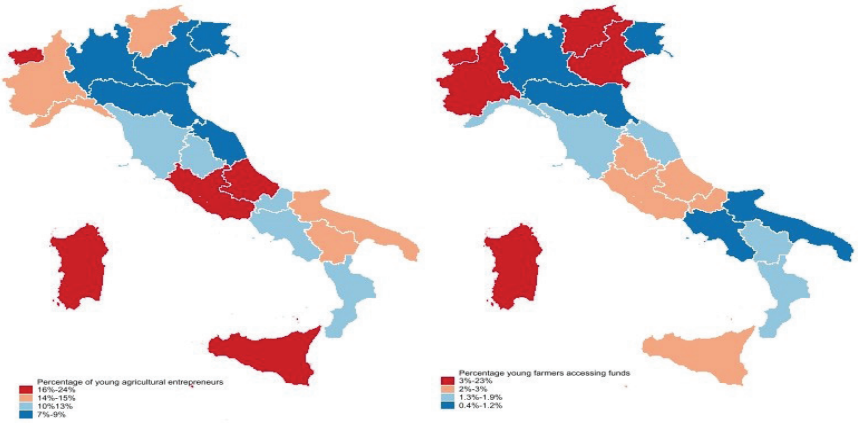


Figure 1: The distribution by regions of Young Entrepreneurs and Young Policies Accesses

We exploit the Farm Accountancy Data Network (FADN)[1]. FADN was introduced in 1965 on an initiative of the European Commission and serves as the authoritative repository of microeconomic data, owing to its adherence to harmonized accounting standards. The responsibility of collecting and organizing the FADN on an annual basis for Italy belongs with the Council for Research in Agriculture and Agricultural Economics (CREA).

Table 1: Summary statistics

	Mean	SD	Min	Max	N
Gender	0.343	0.475	0.000	1.000	5969
Young	0.143	0.350	0.000	1.000	5969
YoungPol	0.008	0.0884	0.000	1.000	5969
ROE	0.005	0.003	-0.002	0.031	5969
Size	1.317	1.835	-0.3489	18.129	5969

The sample of the Italian FADN, comprises 11,000 annual farms and is designed to accurately reflect the diverse range of production types, sizes, and geographical locations (such as region and altitude) found across the entire national territory. It is stratified based on three main variables: location, economic size, and farm production. Specifically, the sample includes 21 NUTS2 regions, 6 economic size classes, and 19 farm types. The validation of data is conducted by the National Institute of Statistics (ISTAT). The system enables a nationwide coverage of 93 % of the

[1] Data Availability Statement: Restrictions apply to the availability of these data. The data were obtained from the Council for Agricultural Research and Economics (CREA) and are accessible at the URL https://bancadatirica.crea.gov.it/Account/Login.aspx with the permission of CREA. Acknowledgments: We gratefully acknowledge the support of CREA for making the RICA data available to the research team (https://www.crea.gov.it/accordi-nazionali).

utilised agricultural area (UAA) and 98% of the value of standard production. The database for the Liguria region counts 1037 farms and covers the period from 2008 to 2020. Among these farms, 228 companies are led by young entrepreneurs[2]. The total number of observations for Liguria is 5969 during the period under scrutiny. All the variables contained in the analysis are taken from FADN.

2. Empirical model

Two econometric specifications are used: one (Equation 1) to analyse the role of young farmers and the other (Equation 2) to investigate the effectiveness of youth agricultural policies (in this case, a sub-sample of young farmers, who are the only possible beneficiaries of these specific funds, is considered).

The dependent variable is the economic performance measured by ROE (Return on Equity). The main independent variable for Equation 1 is a dummy variable called *Young* which assumes value 1 if the farm is run by an under 40 years old individual, and 0 otherwise. For Equation 2, the main independent variable is *YoungPol* which is a dummy variable that assumes value 1 if the farm has benefited from funds dedicated to young entrepreneurs, and 0 otherwise. Then, several control variables are taken into consideration in the two analyses. Firstly, we introduce some individual variables that describe entrepreneurs' characteristics. *Gender* is a dummy variable that assumes value 1 if the farm's tenant is a female and 0 if he is a male. *Education* is a categorical variable and describes the conductor's educational attainment which ranges from *primary school* license to *university degree*. *Legal form* is a categorical variable which represents the different legal forms of each farm. They can be *sole proprietorship, simple partnership, cooperative, limited partnership (società in accomandita semplice, s.a.s.), general partnership (società in nome collettivo, s.n.c.), or other type*. *Management* represents the type of company management: *direct with only family members, direct with a prevalence of non-family members, direct with a prevalence of family members, or with salaried employees*. Furthermore, the categorical variable *ProfCond* details the conductor's professional condition and it can be in *looking for a job, employed outside the farm, part-time employee in the farm, full-time employee in the farm, retired from work, or others*. Another insightful variable is *Settlement*, a categorical variable that describes the entrepreneur's mode of establishment: *rent, free loan, purchase, donation, or others*.

Then, controls related to the farm enter the model. *Organic* is a dummy variable that assumes value 1 if the farm's production is organic and 0 otherwise. The variable *Size* describes the farm's dimension: it is the ratio between UL (*total labour units* including inter-farm labour exchange) and the total surface. Finally, the variable *Altimetric zone* is a categorical variable which defines where the farm is located. It can assume three different levels: *Plain, Hill and Mountain*.

$$\log Y_{i,t} = \beta_0 + \beta_1 \log Young_{i,t} + \beta_3 \log X_{i,t} + \beta_4 \log Z_{i,t} + d_{i,j} + \epsilon_{i,t} \quad (1)$$

$$\log Y_{i,t} = \beta_0 + \beta_1 \log YoungPol_{i,t} + \beta_3 \log X_{i,t} + \beta_4 \log Z_{i,t} + \epsilon_{i,t} \quad (2)$$

$Y_{i,t}$ is the ROE dependent variable. $Young_{i,t}$ and $YoungPol_{i,t}$ are the main independent variables of interest. $X_{i,t}$ are the socio-demographic controls and $Z_{i,t}$ are the farm-specific controls. The baseline profile is a young female farmer (in Equation 2, she accesses policies for young farmers), who has her business in a mountainous area, producing biologics. Her farm is run directly by only family members and is a sole proprietorship that she obtained by inheritance. As far as qualification is concerned, she has a middle school degree and is employed full-time in the farm. The data set is panel data, and the period from 2008 to 2020 is taken into account. Farm fixed effects are also

[2] According to the literature, a young entrepreneur is typically defined as an individual whose age is below 40 years.

considered for Equation 1.

The results are shown in Table 2. The first and second columns relate to the first specification with the entire data set and consist of a simple OLS and an OLS with farm fixed effect, respectively. Instead, the last column refers to the sub-sample of young farmers. Looking at the first specification, for both models with and without fixed effects, it is clear that the presence of young people leading the enterprise has a significant and positive effect. Farm size likewise has a positive impact on the economic result, as does location in a hilly rather than a mountainous area. Furthermore, compared to a totally family-run business, having salaried employees seems to have a positive effect. Regarding settlement, it is observed that free lending has a better impact on the economic result than succession. Finally, educational qualifications do not seem to be particularly significant (the sign often varies when considering fixed effects) as does occupational status. Focusing on the last column and the second specification of the model (sample reduced to 855 observations), the noteworthy finding is that agricultural policies targeting young individuals do not exhibit a significant effect on economic performance. The size and management with salaried employees remain significant and positively impacting.

The results are similar to those of the OLS without fixed effects for the total sample. In addition, female gender and having a university degree seem to have a positive impact even if not significant. Finally, the Hausman test has also been conducted: the fixed effects model is consistent.

3. Conclusions

The aim of this study is twofold: firstly, it explores how the presence of a managerial system led by young farmers contributes to tangible benefits in terms of economic outcomes. Secondly, it examines the effectiveness and suitability of the agricultural policy structure designed for young individuals. The findings, concerning Italy and particularly young farmers in Liguria, reveal that young entrepreneurs actively and effectively contribute to the agricultural sector. However, the current policies seem inadequate. As a potential area for future investigation, it would be essential to explore whether the lack of effectiveness stems from insufficient funding allocated to young farmers or issues related to their access to such funds.

References

European Parliamentary Research Service (2016). *Precision Agriculture and the Future of Farming in Europe*. Scientific Foresight Unit (STOA), Brussels.

Fargione, R., Modesto, L., Riccio, C. (2023). *Rapporto: I giovani in agricoltura. Resistere, crescere e innovare*. r/02, Divulga (https://www.divulgastudi.it/wp-content/uploads/2023/03/Rapporto-giovani-2.pdf).

Istituto di Servizi per il Mercato Agricolo Alimentare. (2012). *La competitività dell'agroalimentare italiano: Check up 2012*. ISMEA, Roma.

McMichael, P. (2007). Feeding the world: Agriculture, development and ecology. *Socialist Register*, **43**:170–194.

Milone, P., Ventura, F. (2019). New generation farmers: Rediscovering the peasantry. *Journal of Rural studies*, **65**, pp. 43–52.

Zagata, L., Sutherland, L. (2015). Deconstructing the 'young farmer problem in Europe': Towards a research agenda. *Journal of Rural Studies*, **38**, pp. 39–51.

Table 2: Regression Output

	Complete data set		*Young Subsample*
	OLS	*Firm Fixed Effect*	*OLS*
	(1)	(2)	(3)
Young	0.001***	0.001***	
	(0.0001)	(0.0001)	
PolGiov			−0.0005
			(0.0004)
Sex F	−0.0001	0.0003	0.0002
	(0.0001)	(0.0002)	(0.0002)
Altimetric zone: Hilly	0.0001*	0.003***	0.0002
	(0.0001)	(0.001)	(0.0002)
Management: Others	0.0002	0.001	−0.001
	(0.0004)	(0.001)	(0.001)
Management: Direct with prevalence of non-family members	−0.0001	−0.0004	−0.002***
	(0.0001)	(0.0003)	(0.0004)
Management: Direct with a prevalence of family members	−0.0001	−0.001***	0.0001
	(0.0001)	(0.0001)	(0.0002)
Management: With salaried employees	0.002*	0.002*	0.005**
	(0.001)	(0.001)	(0.002)
Settlement: Purchase	−0.0002*	0.0004	−0.00002
	(0.0001)	(0.0004)	(0.0004)
Settlement: Rent	0.002***	−0.001	0.002***
	(0.0001)	(0.0004)	(0.0003)
Settlement: Others	0.001***	0.001***	0.001**
	(0.0001)	(0.0003)	(0.0004)
Settlement: Free loan	0.002***	0.001**	0.003***
	(0.0001)	(0.0005)	(0.0005)
Settlement: Donation	0.0002*	0.001	−0.0001
	(0.0001)	(0.0004)	(0.0003)
Organic	−0.0002	−0.0002	−0.001**
	(0.0002)	(0.0004)	(0.0004)
Size	0.0002***	0.0002***	0.0002***
	(0.00002)	(0.00003)	(0.0001)
Degree: High school	0.0002*	−0.0003	−0.00004
	(0.0001)	(0.0003)	(0.0003)
Degree: Professional qualification	0.0001	−0.0001	−0.001*
	(0.0001)	(0.0004)	(0.0004)

doi.org/10.26398/asaproc.0004

Volatile compounds of traditional Bagnolese Pecorino cheese

Andrea Balivo[a], Luigi Fabbris[b], Alfonso Piscitelli[a] Annunziata Piccolo[a], Francesca De Filippis[a], Alessandro Genovese[a]

[a] Department of Agricultural Sciences, Federico II University of Naples, Naples, Italy.
[b] Tolomeo studi e ricerche, Padua, Italy.

1. Introduction

The term "pecorino" typically refers to a family of cheeses made from ewe's milk. The most famous is Pecorino Romano DOP, produced in the regions of Lazio and Sardinia. The strong and robust flavour, combined with the long aromatic persistence, are some of the most appreciated sensory characteristics of pecorino cheese. The sensory properties of pecorino cheese differ based on the region and specific cheese-making methodologies employed (Braghieri et al., 2014). The positive reputation of pecorino cheese is mainly based on tradition and origin from specific territories, where consumers express a liking especially for small-scale local productions. The willingness of consumers to request local dairy products made with milk from pasture-raised animals lies in the greater perception of animal welfare and environmental sustainability, as well as a better nutritional quality of the resulting cheese (Balivo et al., 2023). In fact, Serrapica et al. (2020) found that Pecorino Bagnolese obtained from sheep raised on mountain pastures had a higher amount of unsaturated fatty acids, in particular of the omega-3 series and trans 11-based fatty acids (i.e., rumenic acid and vaccenic acid).

Pecorino Bagnolese is manufactured with raw milk from the local Bagnolese sheep breed raised in the Picentini Mountains, in the Alburni Mountains, in the Piana del Sele and in the Vallo di Diano of Irpinia in Campania region (Comendador et al., 2012). The flock is reared extensively or semi-extensively, taking advantage of the mountain pasture during the summer season. In addition to fresh forage, the sheep's diet may include hay and concentrate. The milk must be processed within 24 hours of milking and is curdled with paste lamb chymosin. The annual production is small and varies between 150-250 quintals of cheese (Comendador et al., 2012). Pecorino Bagnolese cheese is recognised as a traditional agri-food product (PAT) by the Italian Ministry of Agriculture and as a slow food presidium (ONAF, 2020). Maturation lasts from 20 days to over 6 months, generating a soft-to-hard cheese.

In a previous study, Pecorino Bagnolese was distinguished from Pecorino Sardo and Toscano applying multivariate statistical techniques to sensory data assessed by a trained panel (Comendador et al., 2012). So far, there have been no studies examining the volatile compounds (VOCs) in Pecorino Bagnolese. We chose the analysis of volatile compounds since the VOCs in cheese are affected by factors such as animal breed, diet, production methods, maturation, and microbial interactions, and therefore can represent a chemical fingerprint for distinguishing the various types of Pecorino cheeses.

The aim of this research is to differentiate Pecorino Bagnolese cheese from other kinds such as Pecorino Romano, Toscano, Sardo, del Pastore, Canestrato Pugliese and Fiore Sardo based on volatile organic compounds (intrinsic characteristics) and other productive socio-economic variables as breed, rennet, paste, ripening, feeding system and certification (extrinsic characteristics) using multivariate statistics approaches.

The following part of the paper's structure is as follows: Section 2 introduces data sources, outlines SPME-GC/MS[1] for VOCs data collection, and the statistical analysis model. In Section 3, main statistical findings are presented. Section 4 discusses the results with the

[1] Solid-Phase Micro Extraction (SPME) Gas Chromatography-Mass Spectrometry (GC/MS).

existing literature.

2. Data and methods

Thirty-seven samples of Pecorino cheese were purchased from local supermarkets in the Campania region, Italy. Seven types of pecorino cheese, listed below, were included in our study: Pecorino Bagnolese, Pecorino Romano, Pecorino Toscano, Pecorino Sardo, Canestrato Pugliese, Fiore Sardo and Pecorino del Pastore. The samples were randomly selected to ensure a representative assortment of Pecorino cheeses available in the local market. Upon collection, the Pecorino cheese samples were immediately transported to the laboratory under refrigerated conditions and frozen at -20°C.

2.1. SPME-GC/MS

The VOCs were extracted from cheese samples using the Headspace-SPME technique following the procedure described by Sacchi et al. (2020). Frozen samples (25 g) were finely grated, suspended with distilled water (25 mL), and mixed in a 100 mL glass bottle with 50 µL of 2-methyl-3-heptanone (99% purity, Sigma-Aldrich, St. Louis, MO, USA) as internal standard (510 mg L^{-1}, in water solution) and 6.25 g of sodium phosphate (NaH2PO4) (Sigma-Aldrich). The mixture was heated (50 °C for 10 min, then 40 °C for 20 min) and magnetically stirred (150 rpm). VOCs were adsorbed onto a 50/30 µm thick divinyl-benzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS) 2 cm length fibre, which was exposed to sample headspace for 30 min at 40 °C while stirring.

The VOCs were desorbed directly in the injector port of GC kept at a temperature of 250 °C in split mode with a 4:1 split ratio, for 10 minutes. Volatile compound analysis was performed on an Agilent 7890A GC System gas chromatograph coupled to an Agilent 5975C VL MSD with Triple-Axis-Detector mass spectrometer (Agilent Technologies, Inc., Palo Alto, CA, USA). GC was equipped with a Zebron ZB-WAX capillary column (60 m × 0.25 mm i.d. × 0.25 µm film thickness 100% polyethylene glycol; Phenomenex, USA). The carrier gas was helium with a flow of 1 mL min^{-1}. The temperature program was 40 °C for 10 min, then raised at 5 °C min-1 to 240 °C and held for 11 min (Balivo et al., 2023). Mass spectra were recorded at 70 eV. The source temperature was 230 °C, the quadrupole temperature was 150 °C and the interface temperature was 250 °C. Identification of VOCs was based on retention times, mass spectra, and reference compounds. Quantitative data were normalised to the internal standard and analysed using MSD ChemStation 5975 TAD data analysis software (Agilent Technologies, Palo Alto, CA, USA). The SPME-GC/MS procedure allowed us to extract a total of 64 different VOCs molecules from 37 distinct samples of Pecorino cheese. The VOCs of cheese samples were collected in triplicate for each sample type.

2.2 The statistical model

The analysis of the data collected on Pecorino cheese includes the classification variable as a criterion variable and dichotomised as follow: $Y = 1$ if the pecorino cheese is Bagnolese, and $Y = 0$ otherwise. We performed a linear discriminant analysis in order to find a linear combination of variables that characterises or separates two classes of units. Discriminant analysis is also useful in determining whether a set of variables is effective in predicting the class.

If the observed variables are denoted by $x_1, x_2, \ldots, x_p$, then in discriminant analysis the row vector of coefficients $\boldsymbol{a}'$ is sought, which maximizes $\boldsymbol{a}' \boldsymbol{B}\boldsymbol{a} \,/\boldsymbol{a}' \boldsymbol{W}\boldsymbol{a}$, where $\boldsymbol{B}$ and $\boldsymbol{W}$ are the between and the within sum of squares and cross-products matrices. The linear combination of the observed variables involving the elements of $\boldsymbol{a}'$ as coefficients is the best discriminant function, in that it provides for maximum separation on the groups.

In order to obtaining the discriminant function, the coefficients (the a_j) are scaled so that $\boldsymbol{a}'\boldsymbol{a} = 1$ for each composite (the so-called unit norm condition). The discriminant function ($\boldsymbol{d}$)

then is given by:

$$\boldsymbol{d} = \boldsymbol{a}'\boldsymbol{x} = \sum (a_j\, x_j).$$

The similarities and differences among the groups is achieved through the relative contribution of variables towards the separation of groups. When the classes are well-separated, linear discriminant analysis does not suffer of unstable parameter estimates.

The stepwise discriminant analysis with prior probability for the groups calculated from the size of the groups was performed on the VOCs data to evaluate the differences between Pecorino Bagnolese and other Pecorino cheeses. The first variable to enter is the one that maximizes separation among the groups. The next variable to enter is the one that adds the most to further separating the groups, and so on.

Statistical analysis was performed on the 64 VOCs and 8 socio-productive dummy variables (dichotomised in terms of presence = 1 or absence = 0) using the SPSS package (IBM, 2020).

3. Results

The SPME-GC/MS technique involves extracting and analysing the volatile compounds in cheese, which constitute the mixture of odorous molecules perceived by the sense of smell when eating cheese. In particular, SPME is used to sample the volatile compounds released by cheeses in the headspace of a bottle, while GC/MS is used to separate, identify and quantify the individual volatile compounds in the odorous mixture, providing detailed insight of the aroma composition of cheese samples.

Table 1 shows the results of the stepwise selection of discriminant variables. The discriminant function is statistically significant (canonical correlation = 0.941, Wilks' lambda = 0.115, chi-square = 68.152, $p < 0.001$). The square of the canonical correlation for this composite, when converted to a percent, indicates that about 89% of the score variation for the first function is between groups. The fitted discriminant function with seven variables do provide support for the presence of large differences between the Bagnolese and the other Pecorino cheeses.

Table 1. Forward stepwise discriminant analysis, selection of observed variables with the classification variable in two groups (Bagnolese Pecorino cheese; Other pecorino cheese), n =37.

Step		Variables	Wilks'Lambda	F	Sig.	Canonical discriminant function coefficients
1	Included	2-pentanol	0.546	29.142	< 0.001	1.331
2	Included	2-undecanone	0.390	26.567	< 0.001	1.814
3	Included	2-pentanone	0.321	23.307	< 0.001	-------
4	Included	2-heptenal	0.257	23.101	< 0.001	-1.353
5	Included	ethyl-hexanoate	0.212	23.039	< 0.001	1.203
6	Included	ethanol	0.173	23.820	< 0.001	- 1.191
7	Included	2-octanone	0.147	24.094	< 0.001	- 1.315
8	Removed	2-pentanone	0.155	27.286	< 0.001	-------
9	Included	1-hexanol	0.115	31.908	< 0.001	0.765

At each step, the variable that minimizes the global Wilks' Lambda is included – $P(F_{to\_Enter})=0.05$; $P(F_{to\_Remove})=0.10$.

Pecorino Bagnolese exhibited higher levels of 2-undecanone, 2-pentanol, and ethyl hexanoate. Conversely, it showed lower levels of 2-heptenal, 2-octanone, and ethanol.

4. Discussion and conclusion

The aim of our work was to differentiate Pecorino Bagnolese from other pecorino cheeses. The first result highlighted by the statistical analysis is that the extrinsic characteristics of the cheeses are secondary compared to the intrinsic ones.

The development of volatile compounds in cheese is the result of complex biochemical reactions during fermentation and maturation/ripening process (Andiç et al., 2015). 2-undecanone is a methyl ketone formed by the reaction of β-oxidation of free fatty acids followed

by the subsequent decarboxylation of the resulting β-ketoacid (Bertuzzi et al., 2018). Labows et al. (1980) reported that *Pseudomonas aeruginosa* strains are mainly responsible of the production of a series of odd carbon methyl ketones. Methyl ketones in cheese also derive from the metabolism of mesophilic non-starter lactic acid bacteria, i.e., mesophilic *Lactobacillus*, *Pediococcus*, *Enterococcus* and *Leuconostoc* spp. (De Pasquale et al., 2019).

Pseudomonas spp can reduce methyl ketones to secondary alcohols, like 2-pentanol (Cormier et al., 1991). In Pecorino Bagnolese cheese, 2-pentanol resulted from the reduction of 2-pentanone, which derives from the β-oxidation and decarboxylation of hexanoic acid. Hexanoic acid is the main free volatile fatty acid in cheese (Randazzo et al., 2010). High amounts of secondary alcohols derived from methyl ketones have been detected in cultures of *Leuconostoc* spp. and *L. fermentum* (Pogačić et al, 2016).

Ethyl hexanoate is an ester derived from the esterification of hexanoic acid and ethanol. *Pseudomonas* spp are also known for their significant lipolytic activity, producing large amounts of ethyl esters. In addition, *Lactococcus lactis* subsp. *cremoris* thanks to the presence of alcohol acyltransferase, enzyme that catalyse the esterification reaction between free fatty acids and alcohols, favours the production of ethyl hexanoate in cheese (Liu et al., 2004).

Pecorino Bagnolese had a lower amount of ethanol, consumed mainly to produce ethyl esters. A higher quantity of ethanol was found with the addition of starter cultures of *Lactococcus lactis* subsp. *lactis* and *Lactococcus lactis* subsp. *cremoris* (Kondyli et al., 2002). The other Pecorino cheeses were characterised by a higher abundance of 2-octanone and 2-heptenal derived from oxidation reactions of fatty acids.

These findings suggest that the abundance of these VOCs in Pecorino Bagnolese may be directly linked to its unique microbial composition and enzymatic activity during cheese ripening. These compounds result from lipid catabolism, influenced by the indigenous microflora specific to the production area, particularly shaped by the terroir.

In conclusion, our study identifies seven variables that effectively discriminate Pecorino Bagnolese from other Pecorino cheeses. Plausibly, these intrinsic variables could be attributed to the characteristic microbiota of Pecorino cheese, such as the presence of *Pseudomonas* and other indigenous microorganisms, whose enzymatic activity might explain the production of these VOCs.

Further analyses are required to elucidate the specific enzymatic processes responsible for generating these VOCs in Pecorino Bagnolese cheese. These intrinsic characteristics could serve as crucial factors for authenticating Pecorino Bagnolese cheese and hold promising applications within the dairy industry.

References

Andiç, S., Tunçtürk, Y., Boran, G. (2015). Changes in volatile compounds of cheese, in *Processing and Impact on Active Components in Food*, ed. Victor Preedy, Academic Press, London, (UK), pp. 231-239.

Balivo, A., Cipolletta, S., Tudisco, R., Iommelli, P., Sacchi, R., Genovese, A. (2023). Electronic nose analysis to detect milk obtained from pasture-raised goats. *Applied Sciences*, **13**(2), 861.

Balivo, A., Sacchi, R., Genovese, A. (2023). The Noble method in the dairy sector as a sustainable production system to improve the nutritional composition of dairy products: A review. *International Journal of Dairy Technology*, **76**(2), pp. 313-328.

Bertuzzi, A.S., McSweeney, P.L., Rea, M.C., Kilcawley, K.N. (2018). Detection of volatile compounds of cheese and their contribution to the flavor profile of surface-ripened cheese. *Comprehensive Reviews in Food Science and Food Safety*, **17**(2), pp. 371-390.

Braghieri, A., Girolami, A., Riviezzi, A.M., Piazzolla, N., Napolitano, F. (2014). Liking of traditional cheese and consumer willingness to pay. *Italian Journal of Animal Science*, **13**(1),

3029, pp. 155-162.

Comendador, F.J., Moneta, E., Peparaio, M. (2012). Traditional ewes' milk cheeses from Irpinia: Sensory quality and comparison with some renowned PDO "pecorino" cheeses. *Scienza e Tecnica Lattiero-Casearia*, **63**(1), pp. 41-51.

Cormier, F., Raymond, Y., Champagne, C.P., Morin, A. (1991). Analysis of odor-active volatiles from Pseudomonas fragi grown in milk. *Journal of Agricultural and Food Chemistry*, **39**(1), 159-161.

De Pasquale, I., Di Cagno, R., Buchin, S., De Angelis, M., Gobbetti, M. (2019). Use of autochthonous mesophilic lactic acid bacteria as starter cultures for making Pecorino Crotonese cheese: Effect on compositional, microbiological and biochemical attributes. *Food Research International*, **116**, pp. 1344-1356.

Kondyli, E., Katsiari, M.C., Masouras, T., Voutsinas, L.P. (2002). Free fatty acids and volatile compounds of low-fat Feta-type cheese made with a commercial adjunct culture. *Food Chemistry*, **79**(2), pp. 199-205.

Labows, J.N., McGinley, K.J., Webster, G.F., Leyden, J.J. (1980). Headspace analysis of volatile metabolites of Pseudomonas aeruginosa and related species by gas chromatography-mass spectrometry. *Journal of Clinical Microbiology*, **12**(4), pp. 521-526.

ONAF Organizzazione Nazionale degli Assaggiatori di Formaggio (2020). Pecorino di Bagnolese. Available online at https://www.onaf.it/index.php?c=index&a=schedaformaggio&id=1185. Accessed on August 10, 2023.

Randazzo, C. L., Pitino, I., Ribbera, A., Caggia, C. (2010). Pecorino Crotonese cheese: Study of bacterial population and flavour compounds. *Food Microbiology*, **27**(3), pp. 363-374.

Sacchi, R., Marrazzo, A., Masucci, F., Di Francia, A., Serrapica, F., Genovese, A. (2020). Effects of inclusion of fresh forage in the diet for lactating buffaloes on volatile organic compounds of milk and mozzarella cheese. *Molecules*, **25**(6), 1332.

Serrapica, F., Masucci, F., Di Francia, A., Napolitano, F., Braghieri, A., Esposito, G., Romano, R. (2020). Seasonal variation of chemical composition, fatty acid profile, and sensory properties of a mountain Pecorino cheese. *Foods*, **9**(8), 1091.

doi.org/10.26398/asaproc.0005

Participation rates in direct official business surveys: trends and solutions for sustainability

Giampaola Bellini, Francesca Monetti, Pasquale Papa
ISTAT - Italian National statistical institute, Rome, Italy

1. Background

Data collection processes of official direct business surveys carried out by National Statistical Institutes (NSIs) and other institutions show common trends, that require consideration on their role and possible future approaches. They point out on one side decreasing human resources available and increasing intolerance by businesses towards a growing statistical burden, and on the other development of new technologies in order to face arising difficulties. During last years, in order to optimize and increase the overall efficiency of the data collection processes, NSIs have invested a lot in the organizational and management aspects of the survey processes. On this regard, during the year 2016, Istat realized the centralization of data collection processes, both direct surveys and acquisition of administrative sources. The paper, presenting the time series of response rates (in brief *rr*) in main official direct business surveys, aims at analyzing the permanent effects of CDC (Centralized Data Collection), pointing out critical issues and identifying guidelines for planning future surveys.

2. Analysis of data

Tables 1 and 2 show the *rr* recorded in years from 2016 to 2022, for business official surveys. For structural surveys carried out on ana nnual or multi-annual basis, after a sharp increase observed in coincidence with introduction of CDC, *rr* tended to stabilize. A slight downward trend started after 2019 which, if confirmed, could constitute a warning signal. For shorter-term surveys, the sharp increase recorded with introduction of CDC was consolidated in the following years, with a slight decrease in correspondence with health emergency of years 2020 and 2021.

Table 1. Response rates for business surveys by year (years 2016-2022)

Type of data	Pre CDC 2016	Post CDC					
		2017	2018	2019	2020	2021	2022
Structural surveys (*)							
Total units involved	264.698	256.963	229.885	245.555	206.219	262.968	203.186
Response rate	49	54	60	61	61	60	58
Selected Short-term surveys							
Total units involved	34.244	34.804	33.027	32.127	31.998	34.276	34.331
Response rate	61	67	85	90	86	88	90

* Each unit can be included in one or more surveys. For surveys carried out in more than one year, the year of conclusion is considered. In even years, Inward Fats and CIS surveys are not ran.

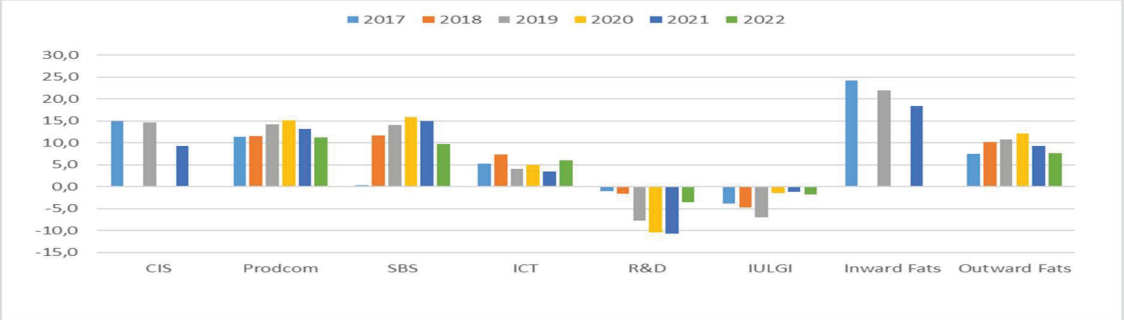
2.1.Response rates in structural surveys

Almost all structural surveys[1], at the beginning, reacted very positively to the new organizational model, except for R&D and IULGI surveys. During the following years, most

[1] CIS - Community innovation survey, Prodcom - Statistics by product, SBS - Survey on enterprise accounting system, ICT-Survey on information and communication technology in enterprises, R&D - Survey on Research and Development in

surveys reached the maximum *rr* in corrispondence with the edition run in the pandemic year 2020. That can be explained by the re-opening of data collection for these surveys during following year 2021. After that, most surveys recorded a slight decline, even if the level reached is still higher than the value they had before CDC (Figure 1). For CIS, Prodcom, and SBS the increase over the values recorded before CDC range between 10 and 15 percentage points (in brief *pp*) while for Outward FATS it ranges from 8 to 12 *pp*. The trend is more stable for ICT survey and such difference stands at around 5 *pp*. For Inward FATS the *rr* recorded a decline since 2017 reducing the increment of 24 *pp* gained that year, even if the difference with the edition run before CDC is still the highest (18 *pp*) (see Figure 1).

Figure 1. Structural surveys response rates variations after CDC per survey and year



The analysis of *rr* for larger enterprises, having at least 250 employees, reported in Table 3, shows marked increases for all surveys right after CDC introduction, that are confirmed in the following editions. The only exception is Outward FATS.

Table 3. Structural surveys response rates for survey units having at least 250 employees, before and after CDC, per year

SURVEYS	Before CDC	After CDC			
	Reference year 2014/2015 (a)	Reference year 2017/2019 (b)	*pp* variation (b-a)	Reference year 2021/2022 (c)	*pp* variation (c - a)
SBS - *all*	75.7	86.2	10.4	87.1	11.3
Prodcom	88.9	90.4	1.5	90.9	2.0
ICT	82.3	92.1	9.8	93.7	11.4
Outward Fats	85.4	91.9	6.5	90.7	5.3

2.2. Response rates in short-term surveys

Figure 2 shows the comparison of the *rr* for the main short-term surveys, among year 2016 (last survey edition run before CDC introduction), year 2017 (first survey edition run after CDC introduction), year 2018 (the first survey edition run after the introduction a the new penalty system[2]), and following years, until 2022. Only from 2021 the new penalty scheme was fully operational. The comparison was carried out at the end of the so called 'useful periods' [Binci, S. et al., 2019]. The Figure show a clear growth in *rr* from 2017 to 2022 in comparison with 2016, the only exception being 2020 due to the health emergency for COVID-19 that occurred

enterprises, IULGI - Updating of the statistical register of economic units ASIA – Local units, Inward FATS - Survey on the activities of foreign controlled enterprises resident in Italy, Outward FATS - Survey on foreign affiliates activities abroad controlled by Italy.

[2] The new penalty system provides for a time articulation on an annual basis, provision of administrative penalties in case of non-compliance for just a single period (month or quarter) and possibility of delayed delivery beyond a certain number of days of tolerance that may vary according to survey (useful period deadline).

during this year [Bellini, G. et al., 2020]. In particular, for DETT and PREIMP surveys the difference between the values recorded in 2017 and in 2022, range from 6 to 45 *pp*, while for IPI and OCC from 3 to 31 *pp*. For PPI and FATT surveys, the trend is more stable recording a *pp* variation from 5 to 16.

Considering only the enterprises having at least 100 employees (Table 4), *rr* for examined surveys increased on average of about 16 *pp* in 2017, starting from 66 percent in 2016. The increase consolidated over the following years, reaching average values of 92 and 95 percent respectively in 2021 and 2022. This upward trend is mainly due to full application, since 2018, of the new penalty management system. It had positive effects also on companies under the penalty threshold.

Figure 2 - Short-term surveys response rates variations after CDC per survey and year

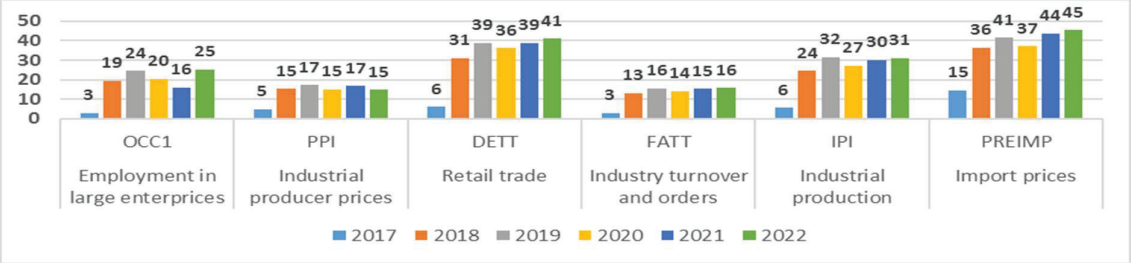


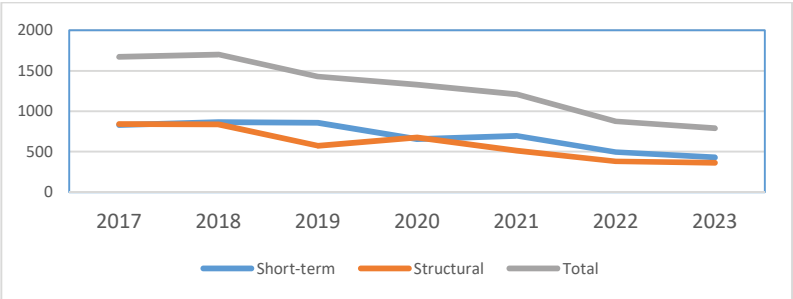
Table 4 - Short-term surveys, average response rates for survey units having at least 100 employees (%), per year

SURVEY		2016 (pre CDC)	2017	2018	2019	2020	2021	2022
Employment in large enterprices	OCC1	68	71	88	93	89	85	93
Industrial producer prices	PPI	77	86	96	97	95	96	96
Retail trade	DETT	52	60	78	82	79	85	91
Industry turnover and orders	FATT	84	88	93	95	93	95	96
Industrial production	IPI	63	72	90	95	90	93	95
Import prices	PREIMP	53	69	91	96	91	96	96

2.3. Trend of human resources employed in the conduction of data collection activities

Figure 3 shows trend of human resources employed in conducting data collection processes since the introduction of CDC in Istat up to today. In particular, the analysis shows a halving of the resources employed. The falling ranges from over 16 FTE (Full Time Equivalent) resources available in 2017 to around 8 FTE resources in 2023. The trend has approximately equally affected the data collection of direct short-term and structural surveys.

Figure 3. Human resources employed in data collection for structural and short-term business surveys. Years 2017 - 2023 (Full Time Equivalent=100).



3. Findings and proposal for new approaches

The generalised increase of *rr* recorded in economic surveys is explained by a set of technological and organizational solutions that were implemented at the same time as the introduction of CDC. The main solutions adopted are schematically reported in the following.

The Business statistical portal, with its main facilities, is a crucial tool supporting users and ensuring adequate participation rates in business surveys. Single point of access, possibility of delegation and updated status of obligations to be performed are examples of services oriented to simplification of statistical obligations required of users.

The design, since 2016, of a centralized contact center service to manage assistance to businesses involved in the surveys is a further solution oriented to efficiency. The service includes different channels for assistance and two levels of management. The first level is completely ran by a specialized external supplier. It concerns repetitive requests and it is provided exclusively on the basis of a detailed set of FAQs designed directly by ISTAT. The second level concerns the management of complex requests both of a thematic and non-thematic nature, which are transmitted by the supplier to specialized Istat personnel through a tool called "shared agenda". The service manages on average around 10,000[3] requests for month specifically for the businesses target.

After CDC introduction, a strict schedule was introduced for management of formal (via certified email) and informal (via ordinary email) communications addressed to units involved in surveys. Deadlines indicated in the informative letter[4] lead the definition of dates for sending communications included in the schedule, determining a different approach for structural and short-term surveys. Texts were recently revised, in order to harmonize them across survey types (short-term, structural) and targets (businesses, institutions, farms) to make them more effective toward involved units. In order to focus the main message of each communication, texts were lightened by removing normative references and inserting specific links to survey pages of the institutional website.

ISTAT pursued, following introduction of CDC, a convergence strategy on two platforms to manage data collection of business surveys. Use of generalized data acquisition software is useful for making data collection processes faster and more efficient both for user's and NSIs.

Application of optimization and standardization techniques in questionnaires design stages, turned out effective for streamlining and rationalizing the filling in task by respondents. This process influenced positively efficiency of the data collection process and output quality, notably in terms of measurement errors control. For this reason, in the framework of the CDC model, Istat created a specific team specialized on this activity.

3.1. New approaches to increase the efficiency of the data collection system

Trends underway in most statistical systems require the introduction of new strategies for data collection of official business surveys.

Convergence towards a multisource approach to data collection, aimed at maximize use of alternative sources is a promising prospect. In this regard, primary sources are those of administrative nature, than those in the form of big data and coming from the web. A greater role will be played also by data science techniques aimed at adequately exploiting the new sources. Concerning the realization of a multisource approach to business surveys, acquisitions of administrative data produced by Revenue agencies, on electronic invoicing and fees (Lima et al., 2023) are solid and promising. They will be able to satisfy, partially, the information needs of surveys on turnover in industry and services and on retail sales. In this regard, Istat is

[3] Thank to Paola Bosso for elaborations carried out on time series of service requests.

[4] The informative letter is the official document that ISTAT uses to communicate to enterprises their involvement in the surveys. It illustrates the objectives of each survey, the regulations in force and the application of possible penalties.

proceeding with defining the agreements for the systematic acquisition and use of this source. Similarly, a strong push is needed to use big data and/or web sources in official statistics. On this front, Istat has recently launched a project aimed at promoting operational use of these sources in the framework of TSS project (Trusted Smart Statistics) (Istat, 2022). A parallel issue to consider in relation to the use of alternative sources consists of the definition of new framework for error control and quality assessment. In this regard, two reference paradigms, although still to be fully developed, are the TSE (Total Survey Error), extended to the use of alternative sources (Biemer et al., 2017) and the KOMUSO project (Eurostat, (2019). A further relevant issue in the perspective of convergence of official statistics towards a multisource approach concerns the confidentiality and protection of personal data, in line with the provisions of Regulation (UE) 2016/679. In fact, the integration of information from different sources prospects new protection issues, that require specific solutions and pose new constraints.

A second relevant topic in designing a future strategy for data collection in official business surveys consists of greater exploitation of opportunities offered by new tecnologies. They will play a crucial role in making data collection processes more efficient and timely. Furthermore, they will contribute to reduce the burden required of the actors involved in the production processes. Main role of new tecnologies is to boost automation of survey processes (Bruni et al., 2022; Snijkers G., 2022). The automation is strictly linked to use of artificial intelligence (AI), web intelligence (WI), as well as the opportunities offered by the development of industry 5.0. An operational condition for the application of new technologies is availability of trained and skilled human resources by NSIs. Further topic of interest in this field concerns the application of interoperability techniques for the real-time acquisition of information available in a structured form by other National Statistical System institutions.

Third strategic issue, dealing with efficiency of data collection processes, concerns application of adaptive survey tecniques. As an example, the reminder activity run by Istat on non-respondent survey units is normally standardized and uniform, as it is based on the assumption that all units have the same behavior in terms of propensity to survey participation. In fact, survey experience shows that there are several factors affecting such propensity, introducing a possible bias on final estimates. That is a reason for considering total unit non-response as a component of the non-survey error. Thus, Istat is introducing a more targeted kind of communication in order to make the reminder activity more effective. According to this approach, the reminder activity is focused on specific target units, whenever specific criteria are met, reducing the bias on the estimates due to the unit non-response.

4. Discussion

Faced with severe decreasing human resources, the producers of official business statistics had necessarily to design solutions aimed at increasing the efficiency of data collection processes, with the constraint of maintaining adequate quality levels. Furthermore, datafication processes of societies and diffusion of increasingly data-driven approaches in the strategic choices of public decision-makers and entrepreneurs involved an increasing demand for statistical data. Efficiency gains achieved during last years concern all organizational, methodological and technological issues. Main factors, in this regard, are centralization of the collection processes, questionnaires optimization techniques, adaptive techniques in processes design, convergence towards generalized acquisition platforms, design of portals for users, design of specialized services for assistance to users involved in the surveys, design of penalty criteria aimed at maximizing participation. All these solutions made it possible to improve the participation rates in the official short-term and structural business surveys and to consolidate them over time. In fact, about seven years after the implementation of the CDC model in Istat, *rr* in the structural business surveys have maintained levels close to 60 percent (about 90 percent for medium-large companies) and about 90 per cent for main short-term surveys (about 95 per

cent in the 100-employee and above companies). Despite signs that participation rates in official business surveys are basically holding up, the trends underway in statistical systems will require, in the next future, a substantial revision of the role of direct surveys. Trend of decreasing human resources, development of new technologies and need to reduce burden on respondents will open new challenges for producers of official statistics. Main answers by NSIs will concern definition of strategies oriented to adoption of a multisource approach and process automation. In this regard, in both fields, it should be noted that the operational results are still limited, despite the various research frameworks set up during last years. It is therefore essential to activate collaboration projects and exchange of experiences among NSIs, on the steps to be taken in order to increase effective use of new technologies and alternative sources.

The outlined context requires a deep review of role and features of direct surveys. They will continue to play a crucial role in business official statistics, but only in situations where recourse to other sources is not possible. They will also tend to take on a more efficient and agile form, with smaller samples, shorter questionnaires, automation of data collection and simpler collection techniques.

Author Contributions

G. Bellini: Sections 2, 2.1, 2.3; F. Monetti: Sections 1, 2.2, 3; P. Papa: Sections 3.1, 4.

References

Bellini, G., De Gaetano, L., Di Paolo, G.G., Monetti, F., Papa P., Pellizzaro, F., Ranaldi, R. (2020). The impact of Covid-19 emergency on ISTAT data collection-transient and permanent effects. *UNECE Expert Meeting on Statistical Data Collection –5-8 October 2020*, Online

Bellini, G., Monetti, F., Papa P. (2020). The impact of a centralized data collection approach on response rates of economic surveys and data quality: the Istat experience. *Statistika, Statistics and Economy Journal,.* **100**(1), ISSN 0322-788X.

Biemer, P.P., de Leeuw, E.D., Eckman, S., Edwards, B., Kreuter, F., Lyberg, L.E., N. Clyde Tucker, Brady T. West. (2017). *Total Survey Error in Practice*. John Wiley & Sons Inc, 2017, Wiley Series in Survey Methodology.

Binci, S., Monetti, F., Papa, P., (2019). Centralised data collection: Effects of a new administrative penalties provision procedure in business short-term surveys. *EESW19, 6th European Establishment Statistics Workshop*, 24-27 September 2019, Bilbao.

Bruni, R., Bianchi, G., Papa, P. (2023). Hyperparameter black-box optimization to improve the automatic classification of support tickets. *Algorithms,* **16**(1), 46. https://doi.org/10.3390/a16010046.

Eurostat, (2019). ESSnet on quality of multisource statistics – KOMUSO WorkPackage 1, *Quality Guidelines for Multisource Statistics (QGMSS)*, Version 1.1, 14 October 2019, https://cros-legacy.ec.europa.eu/system/files/qgmss-v1.1_1.pdf

Istat (2022). Smart statistics from big data, *Available at https://www.istat.it/en/analysis-and-products/smart-statistics-from-big-data* (accessed 10 June 2023)

Lima, B., Poças J., Rodrigues, S., Moreira, A., Saraiva dos Santos, P. (2023). Process automation for treatment of missing values in e-invoice data. *UNECE Expert Meeting on Statistical Data Collection, Rethinking Data Collection*, 12 - 14 June 2023, Online.

Snijkers, G., (2022). System-to-System Data Collection in business surveys applied to an agricultural survey: a Proof of Concept. *UNECE Expert Meeting on Statistical Data Collection, Towards to a New Normal?*, 26 - 28 October 2022, Rome, Italy.

doi.org/10.26398/asaproc.0006

Exploring the gender gap: A preliminary analysis of inclusivity in Italian life sciences industries

Laura Benedan[a], Cinzia Colapinto[b], Paolo Mariani[a], Laura Pagani[a], Mariangela Zenga[c]

[a] Department of Economics, Management and Statistics, University of Milano-Bicocca, Italy.
[b] Venice School of Management, Department of Management, Ca' Foscari University of Venice, Italy.
[c] Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Italy.

1. Introduction

Ensuring women's full and effective participation and equal leadership opportunities at all levels of decision-making in political, economic and public life is one of the key objectives of the United Nations 2030 Agenda for Sustainable Development. Despite the creation of new legislation and the progress of recent years, some disparities between men and women still exist.

Recent economic literature has extensively studied the impact of the presence of women in companies, in particular, the presence of women in top management positions and on boards of directors (BoDs) in several countries. The results of this literature are very heterogeneous and substantially depend on the context in which they are studied (Nguyen, Ntim, and Malagila, 2020; Reddy and Jadhav, 2019; Yu and Madison, 2021).

The introduction of gender quotas on BoDs in several countries, beginning with Norway in 2003, was enforced externally, independent of the companies' preferences. This implementation resulted in an increased presence of women in top management positions. Furthermore, it enabled researchers for the first time to reliably estimate the impact of having women in key business positions.

In Italy, this exogenous force has been represented by the Law 120/2011 (published on July 28, 2011, GU n. 174), also called “Golfo-Mosca” in honour of the proposers, and has forced a change by accelerating and tightening the times. This law concerns listed companies and companies under the public control and aimed to ensure equitable access to the top positions and to instigate a profound cultural shift toward increased female representation in leadership roles and a more robust acknowledgment of women's talents and capabilities in the business realm.

The involvement of women on BoDs has an influence on corporate choices, affecting the value of the company, funding and investment choices (Rossi, Hu, and Foley, 2017). The implementation of gender quotas resulted in a modest rise of women in high-level executive positions and in the percentage of females within individual firms. Moreover, there was an increase in the number of companies led by female CEOs due to the reform (Maida and Weber, 2022). These effects refer to the productive sector as a whole, while limited information is presently available regarding the effects within specific industries, such as that of the life sciences industry.

Besides, special attention to work flexibility and inclusive workplaces balance allows both men and women to balance their professional and personal lives more effectively. This may reduce the gender gap in employment and career progression. Indeed, how employment is organised and compensated to maximise flexibility in scheduling can be a pivotal element in reducing the gender gap (Goldin, 2014).

The life sciences sector represents an interesting sector from the point of view of gender equality, having been a promoter of equal opportunities for years. According to the latest report by Farmaindustria (2023), women represent 44% of employees (53% in R&D), and hold significant roles, reaching 36% of executives and 47% of middle managers in the pharmaceutical industry.

Pharmaceutical companies seem to be very attentive to their workers, promoting training activities and adopting concrete, modern and effective corporate welfare tools, promoting strategies for work-life balance, education, and the well-being of employees and their families, including parenting. In addition, these companies adopt measures aimed at protecting equal opportunities promoting the dissemination of public recognition of the gender equality certification in the sector.

In light of this, our aim was to conduct a more comprehensive exploration of this sector, delving into its peculiarities for a better understanding. Therefore, the present study aimed to explore the world of life sciences companies, medical devices, biotech, and homoeopathy from the point of view of promoting equal opportunities, inclusion and equality. The objective was to collect data about the current scenario to acknowledge the advancements already made and where there is still room to improve.

2. Methods

The survey was promoted by researchers from the University of Milan-Bicocca and the Venice School of Management of Ca' Foscari University of Venice, together with HPS AboutPharma and the Intersectoral Group of Personnel Directors - Human Resources Directors Association (GIDP - HRDA) and aimed to assess the current status of applying and implementing guidelines for gender equality within Italian life science sector companies.

An ad hoc questionnaire was created to investigate the presence of women within corporate management teams and boards. The primary focus encompassed their responsibilities, positions held, and levels of seniority. Besides, the study aimed to estimate the forecasts for the inclusion of female figures within the management team and company initiatives designed to attract and support women. These initiatives ranged from flexible work hours, part-time arrangements, remote work options, supplementary paid leave and provisions for childcare through nursery facilities. This questionnaire has been distributed and addressed to personnel specialised in the human resources of companies in the life sciences sector.

The survey encompassed all pharmaceutical, medical device, biotech, and nutraceutical companies within HPS AboutPharma's databases. Each company was represented by a single recipient of the invitation email, totalling 1,090 companies. The survey was also promoted in the AboutPharma magazine and on select reference websites catering to relevant companies.

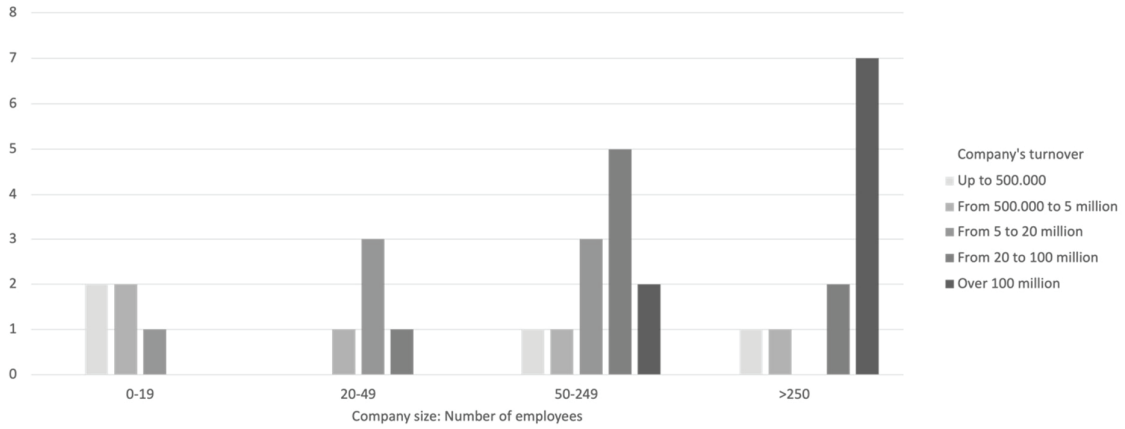
On May 23, 2023, a structured questionnaire was sent via email to the human resources managers of the selected companies. Over the subsequent months until July 18, 2023, a total of 34 comprehensive responses were received and analysed in a descriptive way. At present, the survey remains active and is scheduled to conclude in October 2023. Consequently, this paper will exclusively concentrate on a subset of the acquired results.

The data

In this paper we analyse data about 34 companies of the life sciences sector. Among them, 22 are Italian and 11 are foreign companies; 11 are joint-stock companies (SpA), while 23 are limited reliability companies (Srl).

Figure 1 describes the size of the companies as in terms of the number of employees and the economic turnover in euro.

Figure 1. Company size in terms of employees and turnover (in euro), Italy, 2023



3. Results

The female employment rate ranges across the sampled companies, spanning from 20% to 80%, with an average of 52% *("Referring to the gender of the employees, could you indicate the percentage of female employees?").*

The percentage of women within the company BoD ranged from 0% to 100%, with a mean of 32% (*"Referring to the gender of the Board of Directors, could you please indicate the percentage of female members?").* Illustrated in Figure 2 and Figure 3 are the gender compositions of both the employees and the BoD, categorised according to the total number of employees within the company.

Figure 2. Gender composition of the employees by number of employees – Italy, 2023

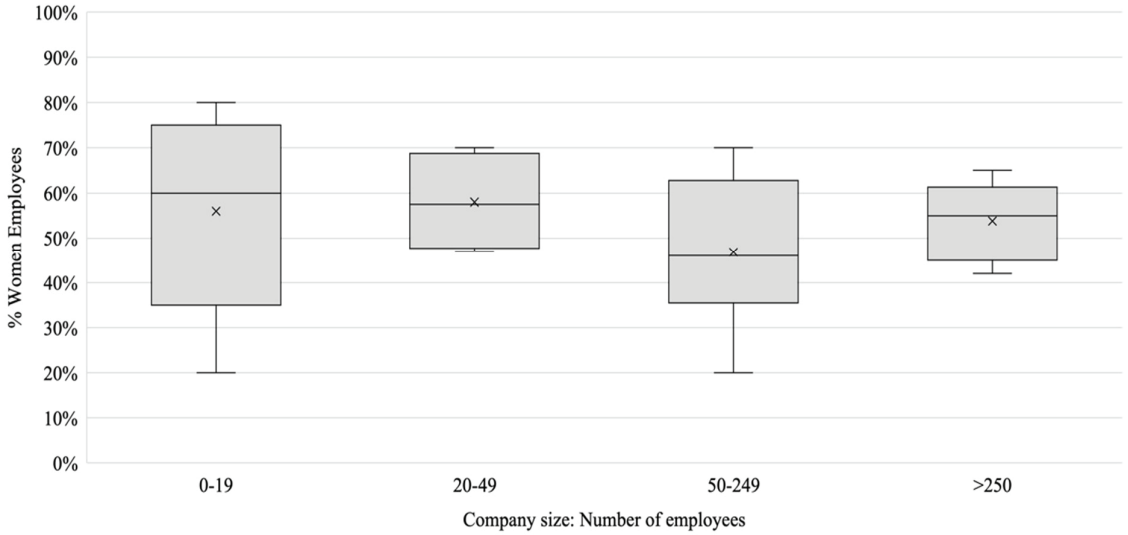


Figure 3. Gender composition of the BoD by number of employees – Italy, 2023

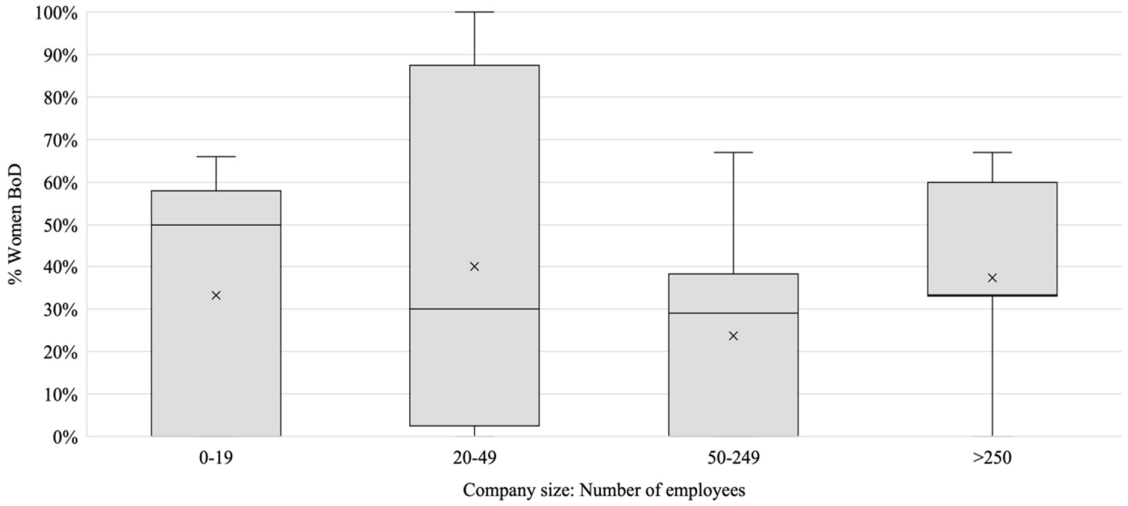


Figure 4. Gender composition of the management team by class size – Italy, 2023

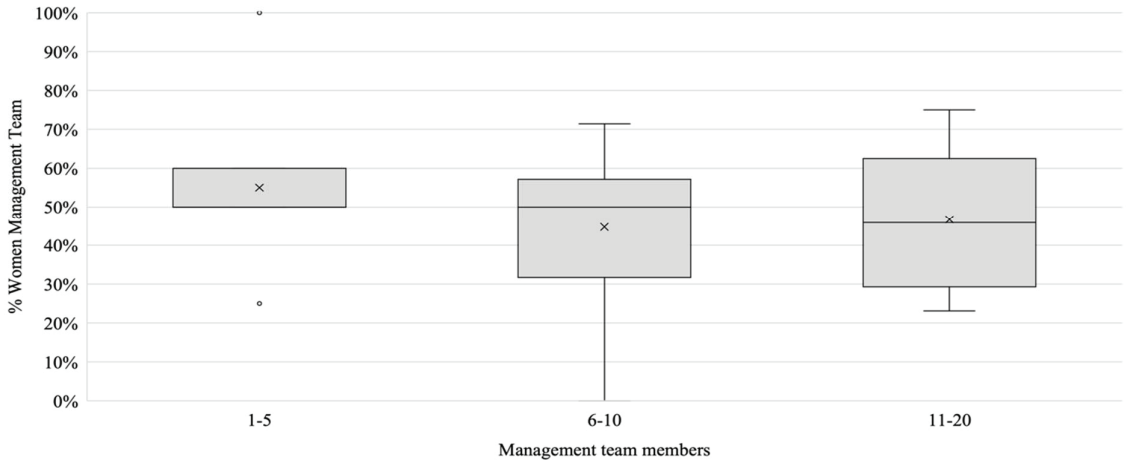
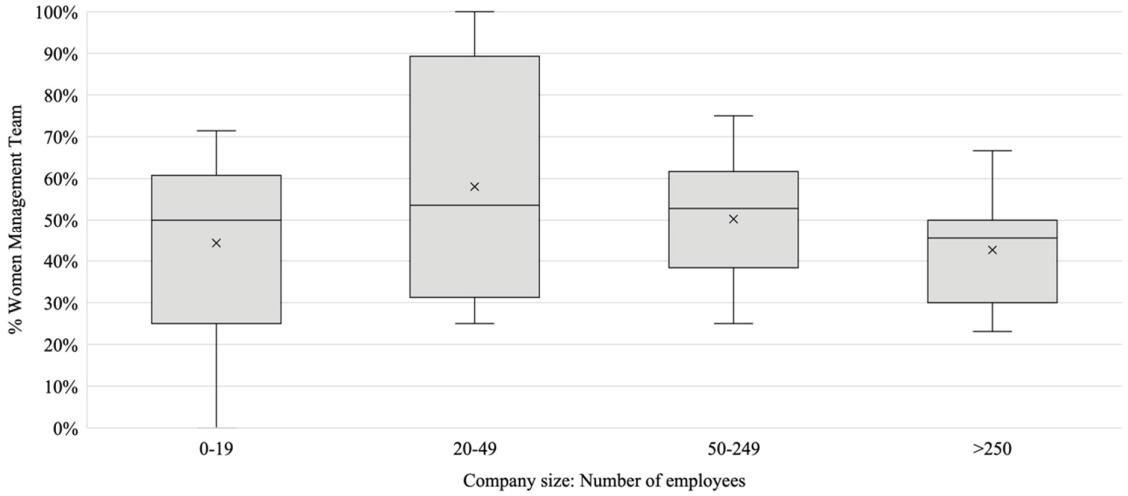


Figure 5. Gender composition of the management team by number of employees – Italy, 2023



The composition of the management team varied, encompassing a range of 1 to 19 members (with an average of 8 individuals). In terms of gender diversity, the representation of women within the management team spanned from 0 to 9 members (averaging 4 women). When assessing the proportion of women relative to the total management team size, this percentage ranged widely from 0 to 100% (with an average of 48%).

Figures 4 and 5 illustrate of the distribution of women within the management team, organised according to the size of the management team class and the total number of employees in the company, respectively (*"With reference to the members of the first managerial line of your company (where first line means only the direct reports of the Chief Executive Officer and the General Manager, level usually called N-1) could you tell us how many total members there are and how many female components are there?").*

4. Discussion and conclusions

The present study aimed to assess gender equality within a sample of Italian life sciences companies, with a specific emphasis on women's presence in upper echelon positions.

The life sciences industry in Italy stands out due to its multitude of organisations that uphold their competitiveness through a profoundly specialised workforce, cutting-edge research, and a burgeoning culture of innovation.

It should be noted that the Golfo-Mosca law is not directly applicable to the companies in our sample, as it is aimed exclusively at listed or publicly owned companies. Nonetheless, its influence might have rippled outward, impacting even those businesses beyond its immediate purview. For the purpose of our study, observing whether this law's influence has extended beyond its intended scope was interesting to understand if its impact had permeated activities beyond its immediate reach, possibly indicating a cultural shift in the longer term.

Despite our analysis is limited to a constrained sample size of companies, we observed a growing presence of women in top roles. Yet, achieving complete gender equality remains a distant goal requiring persistent efforts. The current landscape highlights a significant challenge: the accessibility of women to board positions remains arduous.

One limitation of the study lies in the sampling method, which is not probabilistic but rather convenience-based. It specifically involved companies within the HPS About Pharma network, due to their established professional standing in the industry, whose involvement represented an added value for the research.

A next phase of our study will involve a more comprehensive analysis of all surveyed areas, expanding our sample size once the survey concludes. Additionally, we plan to compare questionnaire responses with broader data from other sources (e.g., public budgets, company websites) to gain a more detailed and expansive understanding, specifically addressing the contemporary and extensive issue of the gender gap in companies.

In conclusion, while our study observed a growing presence of women in top roles within the Italian life sciences companies, achieving complete gender equality remains a distant goal that demands persistent efforts. The current landscape highlights a significant challenge: the accessibility of women to board positions remains arduous despite progress. This persistence of the gender gap underscores the need for a multifaceted approach involving policy changes, cultural shifts, and individual support. Empowering both genders through equal opportunities, challenging stereotypes, and promoting inclusiveness remains crucial. As we continue our research, it's clear that closing this gap requires concerted efforts from various angles to ensure a more equitable and diverse landscape within these organisations and beyond.

References

Farmindustria. (2023). *Indicatori farmaceutici*. Roma, Farmindustria Centro studi.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, **104**(4), pp. 1091-1119.

Maida, A., Weber, A. (2022). Female leadership and gender gap within firms: Evidence from an Italian board reform. *Ilr Review*, **75**(2), pp. 488-515.

Nguyen, T.H.H., Ntim, C.G., Malagila, J.K. (2020). Women on corporate boards and corporate financial and non-financial performance: A systematic literature review and future research agenda. *International Review of Financial Analysis*, **71**, 101554.

Reddy, S., Jadhav, A.M. (2019). Gender diversity in boardrooms–A literature review. *Cogent Economics and Finance*, **7**(1), 1644703.

Rossi, F., Hu, C., Foley, M. (2017). Women in the boardroom and corporate decisions of Italian listed companies: Does the "critical mass" matter?. *Management Decision*, **55**(7), pp. 1578-1595.

Yu, J.J., Madison, G. (2021). Gender quotas and company financial performance: A systematic review. *Economic Affairs*, **41**(3), pp. 377-390.

doi.org/10.26398/asaproc.0007

The quality adjusted labour productivity indicator (QALI): The Italian case

Danilo Birardi, Laura Serbassi, Francesca Tartamella
Directorate for National Accounts - Italian National Institute of Statistics (Istat), Rome, Italy

1. Introduction

Growth and productivity accounts and related indicators play an important role in the political agenda of the European Commission and of other institutions such as the European Central Bank. Moreover, they constitute an important input in the assessment of country-specific developments and in policy evaluation.

"Using total hours actually worked as the input measure for calculating labour productivity changes over time implicitly assumes that each hour worked is of the same quality (that is, there are no differences in the qualifications and skill levels of the labour employed). In other words, each hour worked by a highly skilled person, such as a brain surgeon, is assumed to produce the same quantity and quality of output as each hour worked by an unskilled worker. It is possible to produce a quality-adjusted measure of the labour inputs that takes account of changes in the mix of workers over time by weighting together indicators of quality for different grades of workers" (SNA 2008, § 19.55).

The aim of this work is to present a synthetic index of labour input 'adjusted' to account for quality related variables such as skill level, experience in work and gender, in order to measure changes in the composition of the workforce involved in production. The simple indices, calculated to measure the variation occurred in workforce, are weighted using average hourly wages for a worker of each category (obtained crossing the single categories of each variable previously considered). The premise behind this approach is the neoclassical assumption that workers are paid at their marginal productivity depicted, in this case, by their income.

In this experimental statistical approach, the data processing to obtain quality adjusted labour input (hereafter QALI) considers four different professional statuses: declared and undeclared employees, and declared and undeclared self-employed. For each status we produce a separate matrix of values (hours and incomes) by domain (gender, experience, skill) and economic activity; these matrices are then added together for the final QALI index.

2. Methodology and results

The reason for calculating QALI indices is that the simple change in hours worked by industry is not a sufficient measure to capture all changes in labour force. In order to account for the economic heterogeneity in the measurement of hours worked, these should be disaggregated into groups that reflect these differences, which can be expressed through variables such as skill, experience and gender of the worker (Acheson and Franklin, 2011). On the basis of the information available, the educational qualification (three education levels) was considered as a proxy for skill, while age (three age classes) was used to measure experience. QALI measures only concern the private market economy, and they are computed using a Törnqvist index (Törnqvist, 1936), broken down at industry level. In order to overcome problems of statistical robustness, we adopt a detail of 20 sectors of economic activity (Nace Rev.2 classification by Section), excluding Public administration and defence, compulsory social security and activities of households as employer. The variables considered generate 18 strata domains (3 skill groups * 3 levels of experience * 2 gender groups), for each industry.

The growth in quality-adjusted hours for time *t* with respect to time *t-1* can be represented using a Törnqvist index, defined as a weighted geometric average of the growth rates of hours worked *(h)*, where the weights are labour income *(e)* shares *(w)* across the different groups *(i)*.

In formulas: $$Q_{t-1}^{t} = \prod_i \left(\frac{h_{i,t}}{h_{i,t-1}}\right)^{\left(\frac{w_{i,t}+w_{i,t-1}}{2}\right)} \quad ; \; w_{i,t} = \frac{e_{i,t}}{\sum_i e_{i,t}}$$

A QALI productivity indicator must be consistent with national accounts (NA) productivity measures. To achieve this external consistency, all components of QALI are scaled and constrained to NA figures at industry level (Eurostat, 2016).

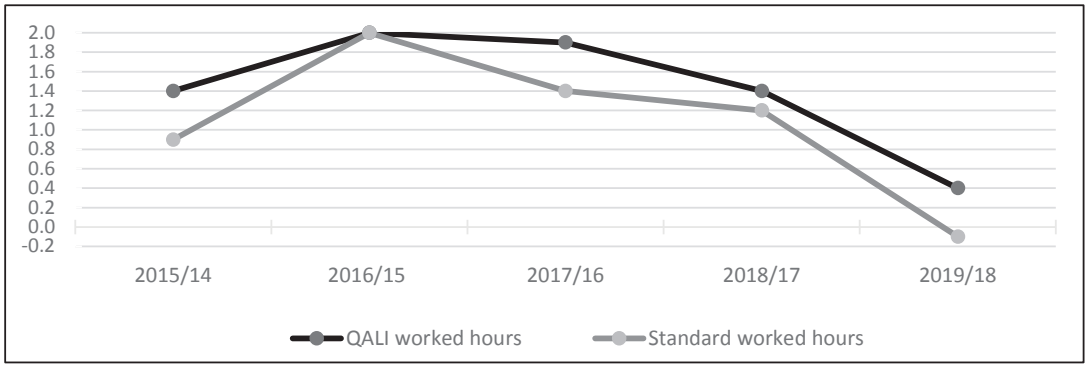
The idea behind the Italian methodology is to consider data on employed and self-employed, and on declared and not declared labour input separately. This allows exploiting the high level of detail of NA estimates (AA.VV., 2015) together with other available annual micro data[1] sources in order to achieve the required stratification by gender, age and skill. This distinction by employment components is a direct consequence of the availability of information for different types of employment, of the existence of separate NA constraints, and, last but definitely not least, of the relevance of not observed economy (NOE) in Italy.

The methodology to obtain QALI index is developed through two different procedures. One for declared employees and one for self-employed and undeclared jobs. The procedure for declared employees is based on register and administrative data, covering all persons employed (census coverage). Instead, the one for self-employed and undeclared jobs is based on sample data derived from the Labour force survey, so quality tests were made to verify the reliability of the results for each strata.

Once the procedure is completed, it is possible to compute, for each sector of economic activity, the simple index of changes in total hours worked. This can be used to compute the geometric average typical of the Törnqvist index, its weights being the 20x18 matrix of compensation (industries-strata).

The results obtained by comparing the changes in hours worked calculated with a simple variation index and those obtained using the QALI index are shown in Figure 1.

Figure 1 - Worked hours: QALI vs simple variation index - Years 2015-2019 (change rates)



Source: elaboration on Istat data

The figure displays that, with the exception of the two-year period 2015/2016, in all the remaining years the QALI emphasizes the dynamic of hours worked, with an average increase of

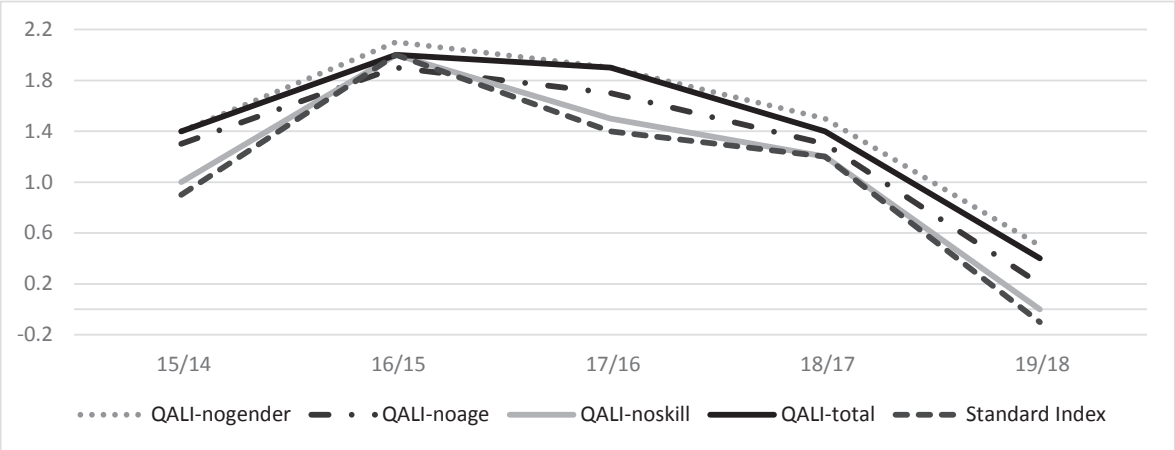
[1]The data sources are: the annual register on employment, wages, hours and labour costs - Racli (https://www.istat.it/it/archivio/234824), the income register (Rtr) used for hourly wages of declared employees in the agriculture sector, the Labour force survey integrated with administrative data (Lfs-Admin database) to obtain the framework of hours worked by sex, age and skill in all industries for self-employed jobs and for undeclared jobs.

just below half a percentage point. Although the analysed period is only six years, it shows that the distance between the two series is fairly stable.

The final additive effect, i.e. the greater dynamism of the QALI hours worked, derives from the greater weight of those strata with higher variability, with respect to total economy, in the geometric average. This is true for each sector of economic activity, with the exception of Agriculture, Accommodation and food service activities and Support services activities, where employment characteristics are more uniform or wages are not differentiated according to them.

The methodology previously described was replicated iteratively, collapsing each time one of the three variables, thus conducting three separate analyses obtained by crossing the categories of the remaining pairs of variables. It allows to investigate the actual significance of the three variables used for the stratification, i.e. how sensitive the QALI index is to each of the socio-demographic characteristics chosen in the selection of the strata.

Figure 2 - Total and partial QALI indices for total economy - Years 2015-2019



Source: elaboration on Istat data

Figure 2 shows the time series of the results obtained by calculating the total QALI index (i.e. including all 3 variables used) and the three partial QALI indices (i.e. obtained by eliminating one variable a time from the calculation), compared with that of the simple variation index of hours.

The differences related with the index no-gender show significantly smaller values than for the other two partial indices, thus highlighting the lower informative contribution of this variable in the calculation of the full QALI. In other words, the exclusion of this variable from the calculation will not produce statistically significant differences in the results[2].

It is interesting to note how Figure 2 shows a prevalent underestimation effect of the full QALI index when compared to the version calculated without the contribution of the gender variable. A negative effect of gender on the results would therefore seem to emerge, since its exclusion from the calculation would produce higher changes in the corresponding QALI index. This effect is in contrast to that of the other two variables which generally tend to make positive contributions instead (especially education).

1. QALI vs standard labour productivity index

Productivity measures allow assessing the contribution of factors that determine output growth. Such measurement is implemented in the framework of the sources of growth (or growth

[2] These results are confirmed by a synthetic indicator computed as the sum of the absolute vales of the deviations between the complete index and the partial ones. The smallest value of this indicator assumed by no-gender confirms that gender has the smallest impact while education the highest (gender=16.8, age=35.4, education=33.8).

accounting) analysis, an analytical approach that allows to decompose the growth of output in the contributions of labour and capital input and of a residual component, the so-called multifactor productivity.

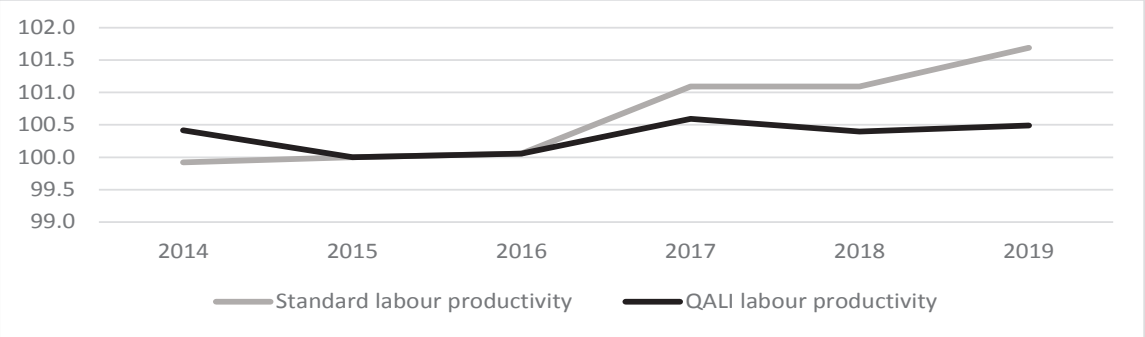
Labour productivity is calculated as the ratio of value added at basic prices in volume to total hours worked by all persons engaged in the production (employees and self-employed), both measures are expressed in chain-linked volumes reference year 2015 (Eurostat, 2021).

Starting from the standard measures of labour productivity (Istat, 2022)[3], we made a comparison with the productivity obtained by using a chained linked QALI index.

Looking at the total economy (net of Real estate activities, Public sector and activities of Households as employers) the analysis of differences (Figure 3) shows that:

- using QALI hours the labour productivity dynamics is lower than the standard one, although by construction, the first year shows an opposite trend with QALI productivity half a point higher than standard one;
- the gap between the two measures widens progressively to a maximum of 1.2 percentage points in 2019.

Figure 3 - Standard and QALI chained linked labour productivity index, 2014-2019 (base year 2015=100)



Source: elaboration on Istat data

This suggests that chain-linked volumes GDP indices are more similar to QALI hours than to standard hours worked indices. In other words, GDP changes appear to be correlated more closely with QALI changes than with simple changes in hours worked. This similarity is more pronounced in recent years (from 2016 onwards).

We can interpret this result by adopting a growth accounting approach (OECD, 2001). The growth accounting approach uses a specific theoretical model to identify empirical measures that can approximate output, input and productivity indices within national accounts frameworks. The theoretical framework of reference derives from the neoclassical theory of production, according to which it is possible to represent technology in terms of a continuous and differentiable production function that relates output, production factors and technical progress. The standard growth accounting model uses a variety of restrictive assumptions, such as perfect competition and the so-called Hicks-neutrality of technical progress, i.e. that technical progress causes marginal factors productivity to change in the same proportion.

The assumptions adopted imply that the elasticity of output with respect to each factor of production is equal to the share of the cost of that factor on the value of output. In this way output value coincides with the total cost incurred in using the inputs, so that the sum of all shares is

[3] Productivity indicators produced by national accounts are released at the end of each year up to t-1, in accordance with NACE Rev.2 classification. Estimates are provided for 20 industries and for the total economy. Both industry level and aggregate estimates are calculated net of Real estate activities, Activities of households, Activities of extraterritorial organizations and of all activities of the General government sector (https://www.istat.it/it/archivio/278143)

equal to unity. According to this approach, the dynamics of output can be expressed as the sum of the growth rates of the production factors (weighted by their respective elasticities) and technical progress. In this way, the rate of change in technical progress (i.e., the shift in the production function over time for given levels of capital and labour) remains the only unobservable quantity, and can be calculated (Solow, 1957) as a residual between the changes in output and those in inputs, thus using only data on prices and quantities of production, capital and labour.

Under these assumptions it can be shown how the evolution of technical progress can be expressed by the change in total factor productivity obtained with the Tornqvist index, which is thus a measure of the change in output not due to changes in the use of factors of production. The greater dynamics of the labour factor (expressed here by hours worked, given the growth of capital), resulting from the use of the QALI index (which reflects qualitative changes in the labor force), can therefore be seen as the "explanation" for a greater share of variability in output, and a consequent downsizing of the residual importance of the evolution of technical progress, unrelated to changes in the factors of production.

Table 1 shows QALI productivity index and the comparison with the standard LPI by economic activity. With the exception of the first year, the productivity obtained with the QALI worked hours is always lower than the standard one in almost all sectors. Two relevant exceptions are Education and Human health and social work activities where productivity is higher with QALI.

Table 1 - QALI chained linked labour productivity and differences with the standard productivity index - Years 2014-2019 (2015=100)

Economic activity Nace Rev.2	QALI labour productivity (2015=100)						Differences with Standard LPI index (*)					
	2014	*2015*	*2016*	*2017*	*2018*	*2019*	*2014*	*2015*	*2016*	*2017*	*2018*	*2019*
Agriculture, forestry and fishing	97.0	100	97.6	95.9	95.5	95.4	0.0	0	0.0	-0.1	0.0	-0.2
Mining and quarrying	117.6	100	138.1	148.0	146.1	136.4	2.2	0	1.4	-0.1	-2.8	-3.7
Manufacturing	98.5	100	101.1	102.6	102.5	102.0	0.9	0	-0.2	-0.9	-1.3	-1.8
Electricity, gas, steam and air cond. supply	105.4	100	99.0	104.2	110.9	113.9	0.5	0	-0.1	-0.6	-0.9	-1.6
Water supply, sewerage, waste manag.	100.9	100	98.5	96.0	95.5	92.4	0.5	0	-1.0	-1.2	-1.2	-1.4
Construction	102.6	100	99.1	99.7	100.2	103.0	0.7	0	-0.3	-0.7	-1.3	-1.4
Wholesale and retail trade services	95.8	100	100.9	102.0	102.1	103.2	0.4	0	-0.4	-0.7	-1.5	-2.4
Transportation and storage	103.5	100	100.4	102.9	101.6	101.6	0.9	0	0.4	-0.1	0.1	-0.3
Accommodation and food service activities	102.6	100	99.3	96.6	96.5	97.8	0.2	0	-0.1	-0.2	-0.4	-0.6
Information and communication	101.0	100	101.9	100.1	97.2	98.5	0.5	0	-0.2	-1.7	-2.5	-2.2
Financial and insurance activities	102.8	100	98.8	98.8	100.6	100.5	1.1	0	-1.0	-2.2	-2.5	-2.7
Real estate activities	-	-	-	-	-	-	-	-	-	-	-	-
Professional, scientific and technical activities	102.2	100	101.7	102.3	104.1	100.4	0.7	0	-0.3	-1.5	-1.7	-1.4
Administrative and support service activities	101.5	100	96.3	94.0	93.1	93.2	0.1	0	-0.1	-0.4	-0.3	-0.9
Public adm. and defence, comp. social securit	-	100	-	-	-	-	-	0	-	-	-	-
Education	102.6	100	98.7	102.8	106.9	103.7	0.2	0	1.0	0.4	1.3	0.6
Human health and social work activities	101.3	100	97.3	98.2	94.9	94.6	-1.1	0	0.6	0.4	0.5	0.5
Arts, entertainment and recreation	103.7	100	106.1	99.0	100.2	108.4	0.4	0	0.5	-2.3	-1.6	-1.1
Other service activities	102.7	100	94.4	97.2	96.7	94.4	0.3	0	-0.6	-1.3	-0.7	-1.3
Activities of households as employers	-	-	-	-	-	-	-	-	-	-	-	-
Total (net Real estate, Act. of households)	**100.4**	**100**	**100.1**	**100.6**	**100.4**	**100.5**	**0.5**	**0**	**0.0**	**-0.5**	**-0.7**	**-1.2**

(*) LPI index disseminated in http://dati.istat.it/?lang=en&SubSessionId=184a9e59-8a3a-4948-969a-781b58352c0d (*Source: our analysis of Istat data*)

In conclusion, development of quality adjusted labour productivity indicators can be seen as an innovative change in the context of labour productivity statistics that affect the actual computation procedures. This is an important step forward to improve understanding of a relevant economic phenomenon as also requested by Eurostat[4]. Unfortunately, as far as Italy is concerned,

[4] https://ec.europa.eu/eurostat/web/experimental-statistics/qali.

this means acknowledging that labour productivity, historically among the lowest in industrialized countries, is even lower.

References

Acheson, J., Franklin, M. (2011). Quality-adjusted labour input: New quarterly estimates for 1993 to 2009 and annual estimates from 1970. *Economic and Labour Market Review,* **5**, pp. 22-46.

AA.VV. (2015). Soluzioni metodologiche per l'utilizzo integrato delle fonti statistiche per le stime dell'occupazione, *Istat Working Papers, N.19.*

Eurostat - DG Joint Research Centre (2016). Methodological document on labour productivity indicators for the EU-28: Quality adjusted labour input: https://ec.europa.eu/eurostat/documents/7894008/8915486/Methodology_QALI.pdf

Eurostat (2021). Eurostat Labour Productivity Indicators (LPIs), Methodological note and quality aspects: https://ec.europa.eu/eurostat/documents/24987/13735645/LPI-methodological note.pdf/22cd66e7-1230-d453-273b-b4b09ab7c7d3?t=1638286567962

Istat (2022) - Misure di produttività – 1995-2021. *Statistiche Report, Novembre 2022:* https://www.istat.it/it/files//2022/11/REPORT-PRODUTTIVITA-2021.pdf

OECD (2001). *Measuring Productivity: OECD Manual,* OECD*,* Paris.

Solow R.M. (1957). Technical change and the aggregate production function. *The Review of Economics and Statistics MIT Press*, **39**(3), pp. 312-320.

Törnqvist, L. (1936). The Bank of Finland's consumption price index. *Bank of Finland Monthly Bulletin,* **10***,* pp. 1-8.

United Nations, European Commission, International Monetary Fund, Organisation for Economic Cooperation and Development and World Bank (2009). *System of National Accounts 2008*. United Nations: New York.

doi.org/10.26398/asaproc.0008

After-covid remote working and towards-the-future disposition in Italy

Luigi Bollani[a], Simone Di Zio[b], Luigi Fabbris[c]
[a] Dept. of Economics, Social Studies, Applied Mathematics & Statistics, Univ. of Turin, Italy
[b] Dept. of Law & Social Science, G. D'Annunzio University of Chieti-Pescara, Italy
[c] Tolomeo Studi e Ricerche, Padua, Italy

1. Introduction

Remote work (RW), also referred to as work from home, underwent a sudden transformation during the COVID-19 pandemic: the outbreak of the pandemic triggered a surge in the demand for RW and the majority of employees worked remotely (OECD, 2020). Prior to the pandemic, the prevalence of RW was growing, although it remained relatively low in most European countries. Eurostat reported that 85% of employees had never worked from home in 2019. Eurofound and the International Labor Office (2017) surveyed telework (another way of naming RW) and information and communication technology-mobile work in 15 countries and concluded that this work arrangement was increasing in all countries, although the large majority of employees had never experienced it. After the pandemic, many workers returned to traditional office work.

The Consumer Expectations Survey (CES) of the European Central Bank (Dias da Silva et al., 2023) explored RW practices in the euro area at different times between 2020 and 2022. While, before the pandemic, over 60% of employees had never experienced working from home, this figure dropped to below 40% in the months following the pandemic's onset. In May 2022, a significant majority of euro area workers reported that they had engaged in remote work for at least one day per week during the previous three months, and among those who worked remotely, the majority did so between 2–4 days per week. This working arrangement will rapidly change the customs of both companies and the workforce. Indeed, changes in working patterns and preferences were accelerated by forced experience, and companies and workers started bargaining on how to hold remote work soon after the pandemic. In many cases, the new way of organising work was a hybrid form: some days in the office and others from home. The CES detected that approximately two-thirds of the surveyed workers desired to continue working remotely for at least one day per week, while about 25% preferred an RW arrangement spanning 2–4 days per week. This new condition allows companies to save space and rent costs, while workers can save commuting time and expenses. In addition, it grants scheduling flexibility and the freedom to work anywhere. Thus, local administrations lowered transportation costs and burdens, and society was rewarded with lower pollution rates. All this represents a win–win situation. On the other hand, companies and workers are both required to organise their activities and lives, respectively, in a very different manner.

RW continues to evolve. Our basic hypothesis is that this business model will not only endure but also flourish in the future. Remote options are poised for expansion, especially as technological advancements, such as artificial intelligence, assume a significant role in managing remote teams. While the prospect of overseeing a remote workforce may appear daunting, with the right technology and dedicated employees, it promises to become a seamless and efficient process. In the medium term, RW can yield significant ramifications for economic and societal development, including an impact on the labour market, transport and housing choices (Williamson, 2022).

To understand workers' preferences regarding RW and the reasons for their choices, we conducted a survey of Italian adults using a computer-assisted web interviewing (CAWI) questionnaire. The survey was part of the RECALL project (*Research on the Effects of Covid-19 on All Lifestyles and social Links)*, aimed at studying the vitality of Italians after the pandemic experience. The project involved research groups at various Italian universities:

Padua, Chieti-Pescara, Turin, Bari and Naples-Federico II.

The research questions under analysis are as follows:

Q1: What is the long-term disposition of Italian workers towards RW?

Q2: Is Italian workers' preference for RW related to their work experience during the Covid-19 pandemic?

Q3: How is Italian workers' disposition to work from remote in a possible future related to their sentiments about this type of work and their work experience during the pandemic?

Q4: Besides opinions on RW and previous work experience, what personal or social resources and problems determine the disposition of workers towards RW?

The rest of the paper is organised as follows: Section 2 introduces the available data and the model for data analysis. Section 3 presents the main results of the statistical analysis of the collected data. Finally, Section 4 discusses the results with reference to the mainstream literature on future RW.

2. Data and methods

An extensive questionnaire was administered from June to November 2021 using a CAWI technique for data collection. We obtained 1,192 responses, including both workers and non-workers. Our data, collected during times of absence of official statistics, are used in what follows for an exploratory analysis. The following analysis concerns only workers, excluding teachers. Teachers were excluded from the analysis because the way they work from remote is peculiar and responses on distance learning have been collected with specific questions. Therefore, the sample reduced to 347 units.

The criterion variable, Y, is the propensity to work remotely in a post-pandemic future. The data are the responses to the following question: "*The health emergency will end. If you continue to work after that, would you rather work from home or at your workplace?*" The four ordinal responses to this question (*1. Only in my office, at my workplace; 2. From home, some days a week, a month, the rest in my office; 3. From home, only at certain times, the rest in my office;* and *4. Only from home, remote working*) were collapsed into two categories: $Y = 1$, indicating a propensity to work remotely (responses 2–4), and $Y = 0$ otherwise (response 1).

The analytical model included two sets of regressors: a first set $A_{01}, A_{02}, Z_{01}, \ldots, Z_{05}$ of control variables forced into the model and a second set $B_{01}, \ldots, B_{06}, C_{01}, \ldots, C_{30}, D_{01}, \ldots, D_{51}$ selected according to its statistical significance through a forward stepwise selection procedure.

The relationship between Y and the regressors may be written as follows:

$$Y=f(B_{01}, \ldots, B_{06}, C_{01}, \ldots, C_{30}, D_{01}, \ldots, D_{51} \mid A_{01}, A_{02}, Z_{01}, \ldots, Z_{05})$$

The between-variable relations can be expressed graphically, as shown in Figure 1, in which homogenous groups of variables are shown.

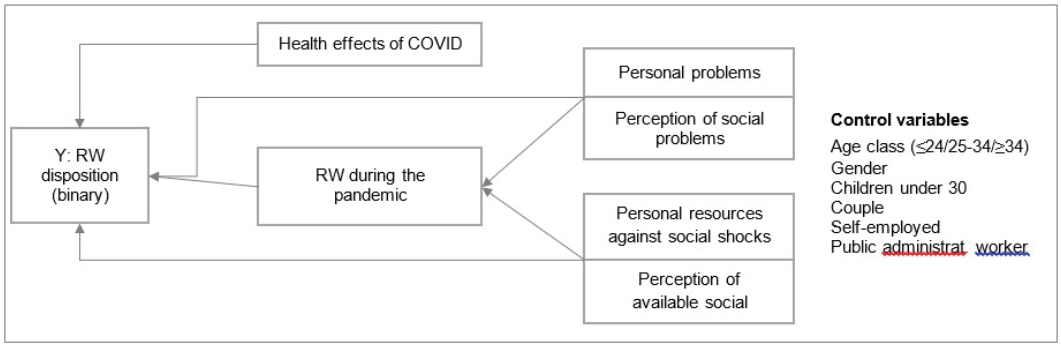


Figure 1. Model of between-variable relations; Y variable is RW disposition.

The logistic regression model for multivariate analysis is (Hosmer and Lemeshow, 2000):

$$logit\ [p(Y=1)] = ln[(p/(1-p)] = \beta_0+\beta_1 B_{01}+\cdots+\beta_j X_j+\beta_{j+1} Z_1+\cdots+\beta_{j+5} Z_5$$

To select the predictors, a significance level < 0.10 was adopted. The R environment (R

Core Team, 2023) was used to estimate the logistic model, adopting the functions *glm* (*MASS* package), *My.stepwise.glm* (*My.stepwise* package) and *PseudoR²* (*DeskTool* package).

3. Results

Table 1 shows how Italian workers' judgements about RW crossed with their work experience during the COVID-19 pandemic. The basic results are that workers' judgement is rather balanced between positive and negative positions, with 45.5% of workers considering RW to be as positive as office work. Indeed, many more workers (67.7%) experienced RW during the pandemic. This means that a relevant quota (43.8%) of workers experiencing RW during the pandemic are not positively oriented towards this type of practice. Nevertheless, those who practised RW during the pandemic tend to have a more positive view of RW, and the more time they spent in RW, the better their judgement: workers practicing RW more than 50% of the time are more positively oriented (67.6%) than those with a lesser practice (35.5%) and much more than those left out from it (23.2%).

Table 1. RW experience during the pandemic and judgment of Italian workers about RW, 2021.

RW experience during pandemic	***Judgment about RW***		
	Negative	***Positive***	***Total***
Null	24.8	7.5	32.3
1 – 50% time	15.6	8.6	24.2
51 – 100% time	14.1	29.4	43.5
Total	54.5	45.5	100.0

If we compute the prevalence of Italian workers available to practice RW in the future with respect to their RW experience during the pandemic and their judgement about RW utility (Table 2), we can see that the disposition to RW ranges from 26.7% in case workers did not experience RW and assign a negative judgement to its social utility to 95.1% in the case of prevailing experience and positive judgement. This highlights that practicing RW is not only relevant to its disposition but suspicion about its utility also lowers the disposition. Of course, suspiciousness can be founded on one's own or other people's previous experiences.

Table 2. Per cent prevalence of Italian workers available to RW in the future, by RW experience during the pandemic and judgement about RW.

RW during the pandemic	***Judgment about RW***		
	Negative	***Positive***	***Total***
Null	26.7	53.8	33.0
1 – 50% time	46.3	86.7	60.7
51 – 100% time	63.3	95.1	84.8
Total	41.8	86.7	62.2

Note: Each cell in the table is obtained by comparing the frequency of available workers with the overall frequency, both referring to the intersection derived from the cell's position in the table.

The effects of specific experience during the pandemic and of value judgement are shown to be compatible with an additive model. Therefore, while fitting a multivariate statistical model to the willingness of Italian workers to practice RW in the future, we analysed the explanatory relevance of these two variables and, on top, of various problems and opportunities related to both personal/familial and social area characteristics.

Table 3 shows the variables selected as significant explanatory variables by stepwise procedure using the logistic model. It shows what follows.

- Demographic variables such as age group, gender, status of living (as a couple) and nature of job (work in public administration) were not significant. This is a very important result because it means that a favourable disposition towards RW is not a

matter of structural variables describing a worker or a family structure or the type of work; rather, propensity is mainly conditioned by situational and personal values descriptors.

Table 3. Beta estimates of the logistic regression model with disposition to RW in the future as criterion variable (forward stepwise selection of regressors, n=347; Nagelkerke R^2=0.619; "RW experience during pandemic", "RW judgment" and control variables were forced into the model; *** < *0.001;* ** < *0.01;* * < *0.05;* ° < 0.10; NS: Not significant; Multi-collinearity checked through VIF (variance inflation factor): VIFi<1.9 for each variable i).

Regressor	$\widehat{\boldsymbol{\beta}}$	$\boldsymbol{se(\beta)}$	***significance***
Intercept	1.537	1.051	NS
Age (classes: <35, 35-64; >64)	-0.056	0.366	NS
Gender: male (binary)	0.033	0.337	NS
Living in couple (binary)	0.313	0.448	NS
Self-employed (binary)	-0.986	0.477	*
Worker in public administration (binary)	0.527	0.510	NS
Having children (binary)	-1.015	0.428	*
RW experienced during the pandemic (6 levels)	0.521	0.109	***
Assessment of RW during the pandemic (binary)	1.728	0.419	***
How RW increases well-being (5 levels)	1.055	0.296	***
How efficient RW is (3 levels; descending order from 1 to 3)	-0.975	0.367	**
I am able to cope with life's challenges (binary)	-0.691	0.339	*
Amount of work to be done after the pandemic: less (binary)	1.196	0.437	**
Personal rejection of home banking (binary)	-1.215	0.665	°
How RW helps to cope with technological challenges (5 levels)	-0.753	0.304	*
Infection: self (binary)	-0.853	0.437	°
How RW influences labour productivity (5 levels)	0.943	0.358	**
Difficulty in handling operational problems (binary)	-0.974	0.579	°
More time available for myself (binary)	1.757	0.943	°
Audits of homework results (binary)	-1.267	0.706	°

- The only family variable related to the propensity towards RW is having children. One could expect it to be positively related to RW disposition; instead, it is significant with a negative sign, which means that workers having children are less available to RW than those without children – all other regressors kept fixed.
- A significant control variable was self-employment ($p < 0.05$), which seems to assume a greater presence in the workplace and is indicated as an unfavourable factor for choosing RW.
- The two main variables assumed in the model (i.e. having experienced RW during the pandemic and having a good opinion of it) are highly significant ($p < 0.001$), and both conditions are favourable to choosing RW, even if other personal and social variables are present in the model.
- Equally significant is the variable indicating that RW increases well-being ($p < 0.001$); this is also reflected in the appreciation of more time for oneself ($p<0.01$).
- Workers inclined to RW also expect to work less than during the pandemic period ($p < 0.01$) but believe that RW can increase labour efficiency ($p < 0.01$) and labour productivity ($p < 0.01$), even if there is a greater difficulty in coping with technological challenges ($p < 0.05$). However, there is no rejection of technological applications, as evidenced by the statement that handling operational problems ($p < 0.1$) and having a proclivity to use home banking ($p < 0.01$) are more frequent among respondents inclined towards RW.
- There is a low degree of control over homeworking activity ($p < 0.1$), which, on the one

hand, suggests a feeling of freedom and trust of the organisation towards the employee and, on the other hand, suggests that organisations can gradually improve RW management.

- Among health descriptors, there is a mild relation between having contracted the virus and the RW propensity ($p < 0.1$), in the sense that the infected are less likely to work remotely. No other variable related to health is present in the model. This is unexpected since RW was a main social policy to circumscribe contagion; instead, it is not a determining factor to compose the decision to be available to RW in the future.
- Psychologically, a low self-efficacy ($p < 0.05$) prevails: people having difficulties in facing life challenges are less disposed to RW. This may be another key factor in the interpretation of the multivariate analysis results.

4. Discussion and conclusion

The pandemic suddenly forced about half the workforce (in our sample, 67.7%) of workers to work from home, and this changed people's mentality towards RW (Fana et al., 2020; Eurofound, 2021, 2022). Before the pandemic, the colony of teleworkers was a lake; then, it became an ocean. We reported that 62.2% of workers were available to experience RW in the future. Comparing these two figures, we can infer that not all workers who experienced RW during the pandemic are available to repeat the experience in the future. In addition, there are others who did not experience it but would be available to do it. We can state that the large majority of the Italian workforce would be available to work remotely if their employers required it.

In the future, RW may be a normal way of working for about half of workers all over the world (Glenn et al., 2019; Willcocks, 2020). Hence, it is important to know in advance the categories that would immediately accept and the ones that would resist. We did not find a category of workers openly favouring or opposing RW; our data show that women are at the same level as men, young workers have propensity levels similar to any other adult and public servants have the same expectations as companies' employees.

Unexpectedly, we have found that people with children and those who are self-employed are less inclined towards RW. The former finding may mean that working at home may affect family organisation and cannot be considered the main avenue for balancing work–family responsibilities. The latter may derive from the concept that a self-employed person is a worker whose only workplace is the field – a laboratory or a customer's house – and in any case, he or she is an entrepreneur in addition to being a worker.

Concerning the antagonism between the presence of children and RW disposition, we may conjecture that the competition during the "lockdown" for personal computers and other communication devices and the attention needed by little boys at home influenced the responses of workers to the questionnaire. In a less disjointed situation, disposition can differ.

Our data showed that sustainability is the underlying factor in people's propensity to work from home (Eurofound, 2022). Once the aim to make their ends meet is given for granted, workers do not seem to be worried by the technology and the protocols needed to make working from home efficient (Sostero et al., 2020), nor are they excessively obliged by possible productivity controls, given that RW implies an advance emotional commitment of employees to their organisation (Felstead and Henseke, 2017; Donnelly and Johns, 2021). Work engagement may be given for granted, and the workplace, as a social site, may be granted by hybrid forms of RW (Delany, 2021; Grzegorczyk et al., 2021).

Nevertheless, there are workers with low levels of self-effectiveness (Parker et al., 2010) who may have difficulty sorting both their family and work lives. Psychological problems related to the sustainability of this double engagement may be a handicap for future RW because our survey data revealed about one-third of respondents with low levels of self-effectiveness.

Finally, health was a dominant topic during the pandemic, and workplaces were considered

major sites for the transmission of the virus. Therefore, the result that having had health problems during the pandemic lowers the propensity for RW was unexpected. It may depend on the perception that a viral contagion, at the end of the day, was not that disruptive, so the risk of a future contagion is a minor factor in the making of people's preferences for working at home or at the workplace.

References

Dias da Silva A., Georgarakos D., Weißler M. (2023). How people want to work – preferences for remote work after the pandemic. *European Central Bank, Economic Bulletin*, **1**/2023, pp. 50-55

Delany, K. (2021). What challenges will organisations face transitioning for the first time to the new normal of remote working? *Human Resource Development International*, DOI: 10.1080/13678868.2021.2017391

Donnelly, R., Johns, J. (2021). Recontextualising remote working and its HRM in the digital economy: An integrated framework for theory and practice. *International Journal of Human Resource Management*, **32**(1): pp. 84–105.

Eurofound (2021). *COVID-19: Implications for Employment and Working Life*. Luxembourg: Publications Office of the European Union.

Eurofound (2022). *Working Conditions in the Time of COVID-19: Implications for the Future*. European Working Conditions Telephone Survey 2021 series, Publications Office of the European Union, Luxembourg.

Eurofound and the International Labour Office (2017). *Working Anytime, Anywhere: The Effects on the World of Work*. Publications Office of the European Union, Luxembourg, and the International Labour Office, Geneva, Switzerland.

Fana, M., Milasi, S., Napierala, J., Fernandez Macias, E., Gonzalez Vazquez, I. (2020). *Telework, Work Organisation and Job Quality during the COVID-19 Crisis: A Qualitative Study*. JRC Paper No. 2020/11, European Commission.

Felstead, A., Henseke, G. (2017). Assessing the growth of remote working and its consequences for effort, well-being and work–life balance. *New Technology, Work and Employment,* **32**(3): pp. 195–212.

Glenn, J.C. and the Millennium Project Team (2019). *Work/Technology 2050. Scenarios and Actions*. Washington, D.C.: The Millennium Project.

Grzegorczyk, M., Mariniello, M., Nurski, L., Schraepen, T. (2021). Blending the physical and virtual: A hybrid model for the future of work. *Policy Contribution*, **14**, Bruegel.

Hosmer, D.W., Lemeshow, S. (2000). *Applied Logistic Regression (2nd ed.).* New York: John Wiley & Sons.

OECD (2020). *Productivity gains from teleworking in the post COVID-19 era: How can public policies make it happen?* (https://www.oecd.org/coronavirus/policy-responses/).

Parker, S.K., Bindl, U.K., Strauss, K. (2010). Making things happen: A model of proactive motivation. *Journal of Management*, **36**(4): 827– 856.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, AT (http:/www.R-project.org, accessed 26 June 2022).

Sostero, M., Milasi, S., Hurley, J., Fernández-Macías, E., Bisello, M. (2020). *Teleworkability and the Covid-19 Crisis: A New Digital Divide?* Seville: European Commission, JRC121193.

Willcocks, L. (2020). Remote working: here to stay? (https://blogs.lse.ac.uk/businessreview/2020/04/02/remote-working-here-to-stay/).

Williamson, H. (2022). *An Evaluation of the Impacts of Remote Working*. Irish Government Economic and Evaluation Service Paper, Department of Enterprise, Trade and Employment (https://enterprise.gov.ie/en/publications/publication-files/an-evaluation-of-the-impacts-of-remote-working.pdf)

doi.org/10.26398/asaproc.0009

Elderly of today, elderly of the past

Emanuela Bologna, Cinzia Castagnaro, Lorenzo Di Biagio, Marina Musci
Italian National Institute of Statistics (Istat), Rome, Italy

1. Introduction

Today, the age threshold defining entry into old age is shifting. The concepts of *active aging* and *silver age* identify a period of life that has gradually expanded, in which people, while aging, continue to fully participate in various spheres of social, economic, political and cultural life.

As it is widely known, in the last century, the combined effect of a high life expectancy and the persistence of a low fertility level have contributed to the progressive increase of elderly population in Italy, both in absolute and relative terms. There is no doubt that population aging is one of humanity's greatest triumphs. But it is also one of our greatest challenges: without adequate policies our system will no longer be sustainable.

The analysis conducted aims to highlight how the lifestyles, behaviours, and perceptions of the elderly population have changed over the last 25 years, with special regard to gender and educational level differences.

2. Data and methods

To gain a deeper understanding of the major changes that have occurred over the last 25 years with respect to the elderly population, a set of indicators calculated from two Istat multipurpose surveys[1] was compared.

These indicators are based on data with a time series that allows a comparison between 1997 (or around that time) and 2022 and covers various aspects of life, including health, lifestyles, cultural and social participation, satisfaction with different areas of life, and concern for the environment.

In order to analyse the differences between the elderly of today and the elderly of the past, for each indicator the ratio between the values for the two groups have been computed (elderly of today/elderly of the past). If the indicator has a negative polarity (i.e., an increase in value implies a worsening in well-being) the inverse ratio has been considered. This default calculation method yields a measure that is not symmetrical around 1 and that has no upper limit, generating possible misunderstanding in graphic representations. This problem can be addressed in two ways: either graphically, using a logarithmic scale (Istat, 2023), or more generally adopting an adjusted parity index (OECD, 2019), in which ratios that exceed 1 are inverted and subtracted from 2. The latter has been chosen because of the advantage to conveniently allow also the representation of ratios much greater than 1. By construction, the parity index and its inverse, after the adjustment, are symmetrical with respect to the parity line (ratio=1) and so different ratios can be correctly compared. The further from 1 the adjusted parity index lies, the greater the disparity between the elderly of today and the elderly of the past. An adjusted parity index value less than 1 indicates disparity in favour of the elderly of the past and a value higher than 1 indicates disparity in favour of the elderly of today. The adjusted parity index lies in the range 0-2 where 0 means that the virtuous behaviour pertains only to the elderly of the past, while 2 means that the virtuous behaviour pertains only to the elderly of today.

For an indicator I with positive polarity the formula for the adjusted parity index (API) reads

[1] The main source of the data presented in this paper is the "Aspects of daily life" Istat survey, with the exception of the data referring to 2000, which comes from "Citizens and leisure time" Istat survey.

as follows:

$$API_{today,past} = \begin{cases} \frac{I_{today}}{I_{past}} \text{ if } I_{today} \leq I_{past} \\ 2 - \frac{I_{past}}{I_{today}} \text{ if } I_{today} > I_{past} \end{cases}$$

Sometimes (OECD 2020) it is considered that the parity index gives no indication of a clear difference between the two groups on which it is calculated if it is within 0.03 points distance to parity. In this paper we have preferred to test the null hypothesis of no clear difference between the two groups with a standard two-sample t-test for the equality of means, with a 5% significance level (Snedecor and Cochran, 1989). Since samples are usually way larger than 30, a normal approximation has been applied. The null hypothesis is rejected (i.e., the values of the indicator for the two groups are considered statistically different) if the test statistic T is larger (in absolute value) than the quantile of order 97.5% (i.e., 1.96) of the standard normal distribution, where

$$T = \frac{I_{today} - I_{past}}{\sqrt{SE_{today}^2 + SE_{past}^2}}$$

and SE_{today}^2 , SE_{past}^2 are the squares of the sample standard errors related to I_{today} and I_{past}.

3. Elderly of today: more participating, more active, but less satisfied

Over the 25-year period, the indicator with the most marked improvement is regular internet use[2], which is about 34 times higher than in the past, for both age groups (Figure 1).

Overall, proactive behaviours related to leisure and participation reveal significant improvements. In fact, there have been positive developments in voluntary activities[3], cultural participation outside the home, and civic and political participation, particularly among the older age group (75+). Use of libraries also shows improvement, with a greater impact seen in the young elderly (65-74) group. Social participation increases only for young elderly (https://www.istat.it/en/well-being-and-sustainability).

Some of these indicators, i.e., use of libraries, voluntary activities, social and cultural participation had reached even higher levels of participation before the Covid-19 pandemic. These advances were interrupted due to the health emergency and the recovery observed in 2022 is not enough to bring the gained advantages back to the 2019 level.

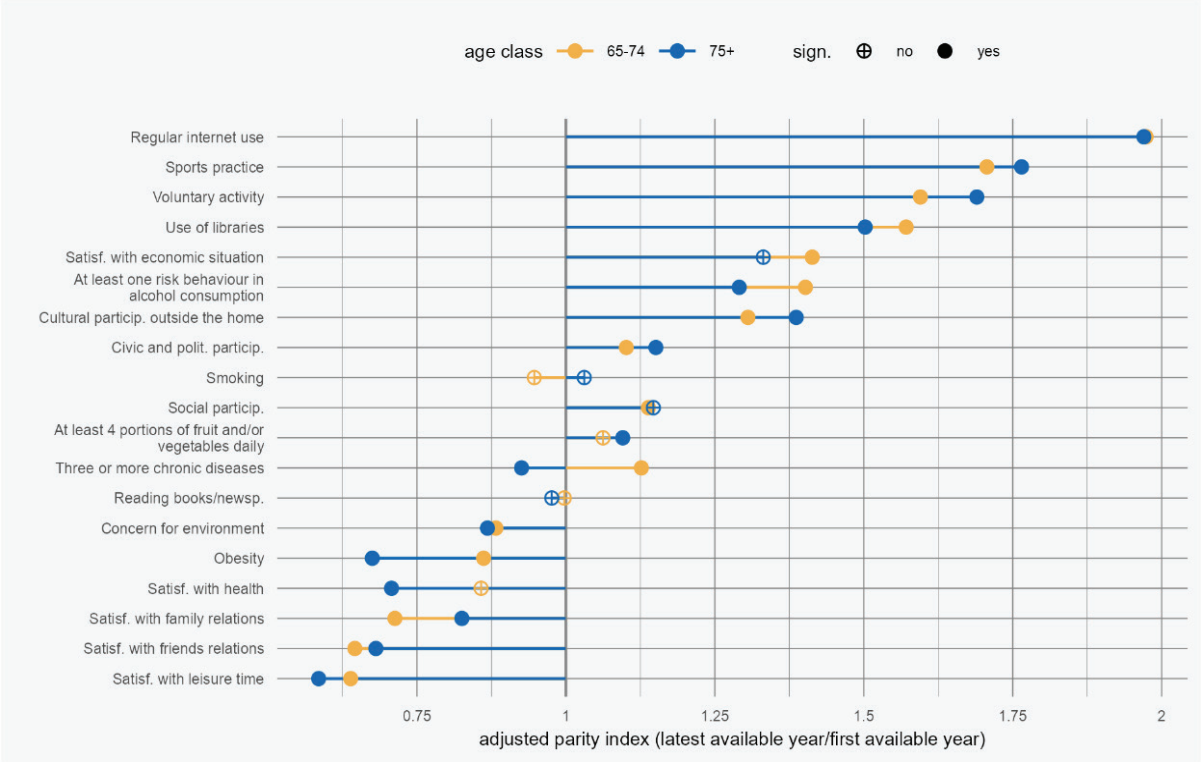
Indicators related to health and lifestyle show important progress, including the widespread of sports practice (continuously or intermittently), a decline in risky alcohol consumption behaviours, and, only for the oldest group, increased consumption of fruits and vegetables (https://www.istat.it/en/well-being-and-sustainability). Looking at the differences between the two age targets, while sports practice increases more among older people, risky alcohol consumption decreases more among people aged 65-74 years. On the other hand, a concerning increase in the share of obesity is observed, particularly among the elderly population.

The only indicator that shows a discordant trend by age is the one referring to multicronicity[4]. Among the younger elderly, there is a notable decrease in the number of individuals with three or more chronic diseases. In contrast, the older elderly group experiences an increase in the prevalence of multiple chronic conditions over the reference period. However, the gains in

[2] People who have used the internet at least once a week in the past three months.
[3] People who have carried out unpaid activities for voluntary organizations or groups.
[4] People who declare to be affected by 3 or more chronic conditions.

survival observed between 1997 and 2022, and the increase in the "oldest old", suggest that this decrease is not necessarily matched by a worsening as such, but it may be an inevitable consequence of the increase in longevity: despite facing more health challenges, the elderly people have higher and higher probability of survival. It is pointed out that, in 25 years, the population over 80 has doubled, the population over 90 has tripled, and the population over 100 and over has even quintupled.

Figure 1 *- Comparison of indicator ratios between 1997(a) and 2022 for age groups 65-74 and 75+. Adjusted parity index (Source: Istat)*



(a) Except for the following indicators: Concern for environment (1998); Civic and political participation (1999); Use of libraries (2000); Regular internet use (2000); Obesity (2001); At least 4 portions of fruit and/or vegetables daily (2003); At least one risk behaviour in alcohol consumption (2003).

Despite an overall improvement on most non-subjective indicators, satisfaction[5] in various aspects of life has declined, except in the economic sphere, which does show improvement for younger elderly. On one hand, satisfaction with leisure time and with friend relations worsen with high and similar intensity for the two age targets. On the other hand, satisfaction with family relationships worsens most strongly for younger older people, while satisfaction with health worsens only for older people aged 75 and older.

Looking at the historical series of indicators related to satisfaction, some considerations can be made. Satisfactions with friendship and family relationships and with leisure time follow a very similar trend in steady decline since the late 1990s for both age groups. In 2021 there is a drop in satisfaction related to the Covid-19 pandemic, which is followed by a recovery in 2022 that in the case of the younger elderly is not enough to recover the 2019 levels, while in the case of the older ones it does.

Satisfaction with health appears to be worse for people aged 75 and over. But looking at the

[5] People very satisfied with various aspects of life.

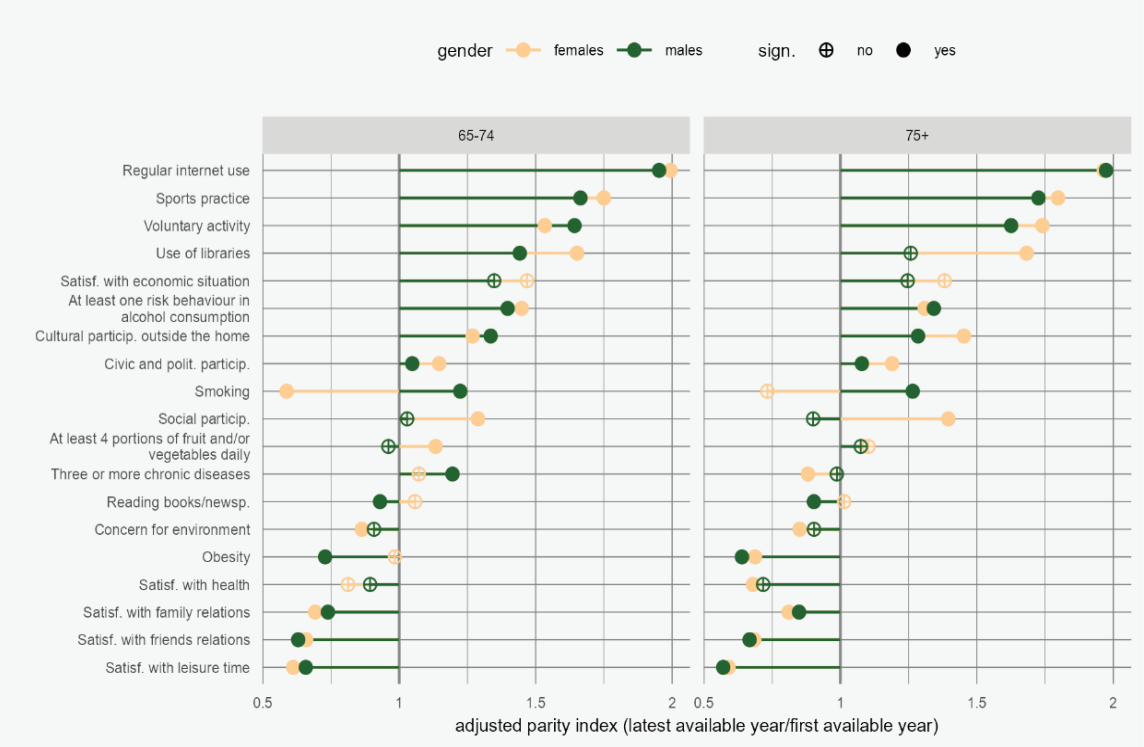
time series, the values of high satisfaction show very fluctuating values until 2005. Since then, high satisfaction levels remain stable. Instead, the share of the elderly who are "very" or "quite" satisfied with their health remained more stable over the period under examination.

Regarding satisfaction referred to the economic situation, the only significant improvement concerns the young elderly, who today are often still active in the labour market or can enjoy their retirement in more active living conditions than in the past.

Finally, concerns about environmental[6] issues seem to be increasing, as reflected in the worsening of the indicator for both age groups.

Performing an analysis by gender, all indicators show concordant trends except for two indicators that show an advantage only for the male component (Figure 2).

Figure 2 *- Comparison of indicator ratios between 1997(a) and 2022 for age groups 65-74 and 75+ by sex. Adjusted parity index (Source: Istat)*



(a) Except for the following indicators: Concern for environment (1998); Civic and political participation (1999); Use of libraries (2000); Regular internet use (2000); Obesity (2001); At least 4 portions of fruit and/or vegetables daily (2003); At least one risk behaviour in alcohol consumption (2003).

Firstly, the smoking habit shows improvement among men in both age groups but worsens for women aged 65 to 74. Secondly, multicronicity displays improvement among men aged 65 to 74 but worsens for women aged 75 and older.

For both age groups, women improve more than men in library attendance and in civic, political and social participation. Only among the elderly aged 75 and older, women improve more than men in volunteering and cultural participation outside the home. However, it is essential to highlight that in 1997, women were significantly less engaged in cultural, social, civic and political activities, as well as volunteering and library use, compared to men. While women have shown greater improvement over time, they have not yet reached the same levels of

[6] People who are concerned about at least 3 environmental issues (out of a maximum of 5 possible answers) from a list of 14 environmental issues.

participation as men in these areas. So, despite progress, the gender gap persists.

An important factors to consider in the changing lifestyles of the elderly is the rise in the educational levels[7]. As a result, the increase in the elderly population is translating into a rapid increase in the population with higher educational levels. According to Istat data[8], from 2004 to 2020, the share of people aged 65 and older with at least a high school diploma more than doubled, rising from 13.2% to 27.8%. The data analyzed for this paper confirm that people with higher educational qualifications participate more in social and cultural activities. However, it is interesting to note that participation levels of the most educated are lower in 2022 than in 1997. This implies that participation is increasing because college and high school graduates have increased, but college and high school graduates participated more 25 years ago than today.

To sum up, the analysis highlights that the elderly of today have a more active lifestyle than in the past, although gender differences persist. One of the main goals of policies should be to reduce these inequalities.

The analysis also shows a decreasing satisfaction with some areas of life. This could incentivize measures to support subjective well-being, a fundamental dimension of quality of life.

Further reflection cannot disregard the demographic scenario and its inevitable evolution.

4. Better conditions of the elderly: is it also a sustainability issue?

The data analyzed highlight the improvements in the living conditions of the elderly population that contributed to the gradual increase in the number of elderly people. Clearly, population aging poses substantial challenges to society, and these challenges are meant to intensify in the coming years. In fact, although today's elderly live in better conditions than in the past, it is inevitable that vulnerable and non-self-sufficient people will increasingly weigh down our social and health care system.

Aging process, in Italy, is fairly recent. Indeed, in the last century, life expectancy increased quickly and fertility had a plummet. These phenomena caused a radical change within the age structure of the Italian population.

In 2023, the population pyramid exhibits distinct changes compared to 1997 (Figure 3). The base of the pyramid has significantly narrowed, while the top has expanded.

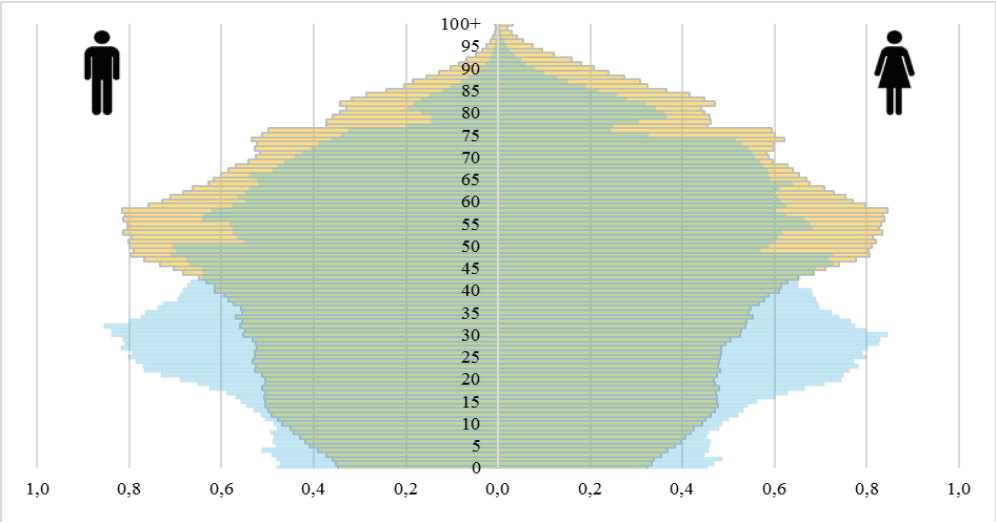


Figure 3 *– Population pyramid. Italy, 1997 and 2023 (a). Per cent data (Istat, provisional data)*

[7] In this regard, it is crucial to recall the turning point represented by the middle school reform (1963) and the introduction of unrestricted university access (1969), the first beneficiaries of which were 65 years in 2020.

[8] Istat, Labour force survey, previous regulation (until 2020). http://dati.istat.it

Notably, the younger population (0-14 years old) has decreased by almost 1 million individuals, comprising 12.5% of the total population as of January 1st 2023. Conversely, the elderly population (aged 65 years and over) has surged, comprising 24% of the total population as of January 1st 2023, representing an increase of almost 4.4 million individuals since 1997.

The *aging ratio*, the best indicators of the demographic balance between the elderly and the young, reveals a significant imbalance destined to increase. In 1997, there were 118.7 elderly individuals every one hundred young people, whereas in 2023, this ratio has climbed to 193.3.

As the Italian Supreme Health Council pointed out (CSS, 2020), one of the most complex challenges of the 21st century is for health systems to optimize available resources in order to respond to the complexity of health demands induced by the continuing growth of chronic-degenerative diseases. As our analysis points out, in 2022 a substantial proportion of the population aged 65 and over (equal to 34.3%) affected by three or more chronic diseases. This share reaches 43.2% among the very old aged 75 and over. This entails an increasing commitment of resources, requiring continuity of care for long periods of time and a strong integration of health and social services. It also requires residential and territorial services that have so far not been sufficiently designed and developed in our country.

Furthermore, the indicator assessing the prevalence of severe limitations reaches 13.1% in 2022 among the population aged 65 and over, with the figure rising to one in every five individuals among those over 75 years old.

5. Conclusion

Over the past decades, Italy's population has undergone a significant aging process, driven by a rapid escalation in life expectancy, coupled with a marked decline in fertility rates.

Comparing the "elderly of today" with respect to the "elderly of the past", the results show a gradual spread of active and proactive behaviours, accompanied by an improvement in lifestyles.

In the area of social, cultural, civic and political participation, women show more marked improvements than men. Nevertheless, they still have not reached men's levels of participation, so the gender gap persists.

On the other hand, today's elderly are less satisfied with friends and family relations, and with leisure time than their peers 25 years ago. This trend has only been accentuated by the Covid-19 pandemic, but the decline was evident even before the health emergency.

Finally, the analysis confirms that over time the use of the internet among people aged 65 and older has strongly increased (while remaining lower than in younger age groups) with a significant impact on everyday life, due to the myriad of opportunities that the internet offers.

As the elderly population continues to grow, accommodating policies aimed at addressing the multiple impacts of population ageing are urgently needed and require a comprehensive and multidimensional approach.

References

CSS (2020). *Invecchiamento della popolazione e sostenibilità del Servizio Sanitario*. Consiglio Superiore di Sanità – Sezione I, Rome.

Istat (2023). *Rapporto Bes 2022: Il Benessere Equo e Sostenibile in Italia*. Istat, Rome.

Istat (2023). *Rapporto annuale 2023. La situazione del Paese*. Istat, Rome.

OECD (2019). *PISA 2018 Results (Volume I): What Students Know and Can Do, PISA*. OECD Publishing, Paris.

OECD (2020). *How's Life? 2020: Measuring Well-being*. OECD Publishing, Paris.

Snedecor, G.W., Cochran W.G. (1989). *Statistical Methods,* Eighth edition. Iowa State University Press, Ames, Iowa 1191.2.

WHO (2000). *A Life Course Approach to Health*. World Health Organization, Geneva.

doi.org/10.26398/asaproc.0010

Testing the effectiveness of public policy interventions on the adoption of 4.0 technologies: The permutation approach

Stefano Bonnini, Michela Borghesi
Department of Economics and Management, University of Ferrara, Ferrara, Italy.

1. Introduction

The main goal of this work is to investigate the effectiveness of public policy interventions to stimulate the adoption of Industry 4.0 technologies by Italian small and medium enterprises (SMEs). The list of 4.0 technologies includes a wide and heterogeneous set of innovative solutions. According to some authors, the advent of Industry 4.0 has led to some harm but also many advantages (Cucculelli et al., 2022). Hence, in recent years, public incentives to support the innovative processes of private companies have been proposed by the governments.

To achieve the Sustainable Development Goals, Industry 4.0 technologies may play a decisive role in assisting organizations' transition. When it comes to sustainability, three aspects should be considered: economy, society, and environment. Even if Industry 4.0 is mainly seen as a means of economic progress, scientific studies proved the environmental and social consequences of it. Industry 4.0 determines efficiency and creates economic and social opportunities (Raj et al., 2020; Bag et al., 2021).

The adoption of Industry 4.0 technologies allows flexible production, improves supply chains, and leads to more effective business management with significant improvements. For the implementation of Industry 4.0, a new combination of traditional and emerging technologies is needed. With regard to operation technologies, the highest growth is linked to additive manufacturing, followed by advanced manufacturing, which has been mainly implemented through the use of collaborative robots (Cugno et al., 2021).

Advanced manufacturing technologies include a wide variety of technical combinations, which support product design and production, higher levels of connectivity, and an optimized resource planning process. Furthermore, the additive manufacturing process enables more product customization options to be activated in new innovation ecosystems (Stornelli et al., 2021; Durão et al., 2017). On the other hand, also the additive manufacturing process is an essential driver of Industry 4.0. By using these techniques, limitations imposed by traditional production methods are eliminated (Ashima et al., 2021).

As said, the aim of this study is to investigate the role of recent public incentives in strengthening the innovative capacity of Italian companies within the Industry 4.0 framework. To monitor the adoption of 4.0 technologies and its consequences, empirical studies based on suitable and advanced statistical methods are necessary. In this work, a nonparametric approach, based on the methodology of combined permutation tests, is applied to assess the impact of public policies in terms of 4.0 technological innovations. For the assumptions on the underlying probability distribution, this approach ensures greater flexibility and robustness. For example, dealing with a multivariate response does not require the multivariate probability distribution to be explicitly assumed or estimated. In this framework, tests of hypotheses may be carried out by using a permutation approach.

The paper is structured as follows. Section 2 focuses on the presentation of the statistical problem and of the nonparametric solution. The results of the application of the proposed method to the case study of manufacturing enterprises in the north of Italy are discussed in Section 3.

Section 4 is dedicated to the conclusions.

2. Statistical problem and nonparametric solution

Let us consider k different public incentives, two main 4.0 technologies, and four different groups of companies. The four groups are defined according to which of the two technologies have been adopted by the companies. Every company in these groups was asked which of the considered public incentives had been exploited.

This work aims to use sample data to test the hypothesis of significant differences between the four groups in terms of the propensity to embrace some public incentives. This propensity is defined in a multivariate sense, jointly considering, for each group, the proportions of companies that have exploited the k policy incentives under consideration.

The problem consists of a multiple test composed of partial tests defined according to:

- the public incentive under consideration (dummy response variable),
- pairwise comparisons,
- directional alternative hypotheses (two-tailed tests decomposed into two one-sided tests).

Many solutions to this multisample problem were proposed in the literature, e.g. Hotelling T^2-test (Agresti and Klingenberg, 2005), Wald-type statistic (Chuang-Stein and Mohberg, 2019), and others. In this work, as said, a solution based on the methodology of Combined Permutation Tests (CPTs) is applied (Bonnini et al., 2014). CPTs consist of a family of permutation tests, appropriate for complex testing problems that can be broken down into partial tests. This approach is distribution-free and robust for the underlying distribution. The iid assumption is not required. The only (milder) necessary condition is exchangeability between samples. This condition, under the null hypothesis, is satisfied (Pesarin, 2001; Pesarin and Salmaso, 2010).

To obtain the global p-value related to the overall testing problem, a suitable combination of the p-values of the partial tests (λ_q) is carried out. The combination can be obtained from the Fisher rule as follows:

$$T_{comb} = - \sum_q \ln(\lambda_q).$$

The possible significance of the global test can be attributed to specific pairwise comparisons and/or response variables, by adjusting the p-values of the partial tests for controlling the family wise error rate.

3. Case study

To investigate the role of recent public policy incentives proposed by the Italian government to enhance the innovative capacity of companies with regard to Industry 4.0 technologies, we will test whether such incentives affected the decision to introduce advanced manufacturing solutions and/or additive manufacturing by Italian companies.

Original data, collected in a sample survey carried out in January 2022, are analysed. The survey was conducted in the northern regions of Italy by the Department of Economics and Management of the University of Ferrara. The questionnaire was administered to a random sample of 3924 manufacturing enterprises in the north of Italy.

The twelve response variables are dichotomous. Each variable takes “1” if the company benefited from the specific incentive and “0” otherwise. The list of variables is the following:

- *hyp-sup-depr*: Hyper and super depreciation,
- *nuova-sab*: Nuova Sabatini,
- *guar-fund*: Guarantee fund,
- *R&D-tax*: R&D tax credit,
- *innov-agr*: Innovation Agreements,

- *develop-contr*: Development contracts,
- *startups-inn-SMEs*: Startups and innovative SMEs,
- *pat-box*: Patent box,
- *spec-comp-centers*: Highly specialized competence centers,
- *tech-transf-centers*: Technology transfer centers,
- *train-tax-credit*: Training tax credit,
- *reg-incent-R&D*: Regional incentive measures for Research and Development and innovation.

The four compared groups are defined according to whether the companies adopted none, one or both of the two considered 4.0 technologies, as follows:

- Group 1: adoption of advanced manufacturing solutions and additive manufacturing,
- Group 2: adoption of only additive manufacturing,
- Group 3: adoption of only advanced manufacturing solutions,
- Group 4: adoption of neither of the two innovations.

We applied the methodology explained in the previous section by carrying out all the possible directional pairwise comparisons between the four groups (for a total of twelve). We point out that such a problem corresponds to a test on multivariate Bernoulli random variables. Thus, we obtained 144 partial tests (twelve response variables for each of the twelve comparisons) and we combined the *p*-values through the application of the CPT. The global (combined) *p*-value of the test is equal to **0.002**. Hence, at the significance level α=0.10, we reject the null hypothesis in favour of the alternative hypothesis of significant difference in the propensity to adopt 4.0 technologies between groups.

In Table 1 we can observe the partial *p*-values adjusted with the MinP closed testing method (Pesarin and Salmaso, 2010). When the partial test is significant, Group 4 (adoption of neither of the two innovations) appears to be increasingly less inclined to use public incentives. Conversely, when significant, the partial tests concerning Group 1 (adoption of advanced manufacturing solutions and additive manufacturing) indicate that this appears to have a greater propensity to use public incentives compared to the other groups. As regards the comparison between Group 2 (adoption of only additive manufacturing) and Group 3 (adoption of only advanced manufacturing solutions), the former appears to have used the *R&D tax credit* incentive the most, while the latter appears to have used *Hyper and super depreciation* and *Nuova Sabatini* incentives the most.

Incentives	1<2	2<3	3<4	2<1	3<2	4<3	1<3	1<4	2<4	4<2	4<1	3<1
hyp-sup-depr	1.000	**0.080**	1.000	**0.080**	1.000	**0.080**	1.000	1.000	1.000	**0.080**	**0.080**	0.587
nuova-sab	1.000	**0.080**	1.000	0.322	1.000	**0.080**	1.000	1.000	1.000	**0.080**	**0.080**	1.000
guar-fund	1.000	1.000	1.000	0.996	1.000	**0.080**	1.000	1.000	1.000	0.234	**0.080**	1.000
R&D-tax	1.000	1.000	1.000	1.000	**0.080**	**0.080**	1.000	1.000	1.000	**0.080**	**0.080**	**0.080**
innov-agr	1.000	0.402	1.000	1.000	1.000	0.459	0.402	1.000	1.000	1.000	0.996	1.000
develop-contr	0.990	1.000	1.000	1.000	0.998	0.996	1.000	1.000	1.000	0.856	1.000	1.000
startups-inn-SMEs	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
pat-box	0.898	1.000	1.000	1.000	0.322	**0.080**	1.000	1.000	1.000	0.134	0.809	1.000
spec-comp-centers	1.000	1.000	1.000	0.801	1.000	0.934	1.000	1.000	1.000	0.991	0.427	0.943
tech-transf-centers	1.000	1.000	1.000	0.820	1.000	0.998	1.000	1.000	1.000	1.000	0.573	0.943
train-tax-credit	1.000	1.000	1.000	0.492	0.998	**0.080**	1.000	1.000	1.000	0.183	**0.080**	**0.080**
reg-inc-R&D	1.000	1.000	1.000	1.000	0.971	**0.080**	1.000	1.000	1.000	0.234	0.514	1.000

Table 1: Adjusted *p*-values of all pairwise comparisons (significance in bold).

4. Conclusions

The presented nonparametric methodology for pairwise comparisons between multivariate Bernoulli distributions, based on the application of combined permutation tests, is a distribution-free, robust and flexible statistical solution. Its application to an original dataset concerning a survey about northern Italian enterprises provides empirical evidence in favour of the hypothesis that the propensity of firms to use public incentives differs depending on whether they have introduced advanced manufacturing solutions or additive manufacturing. In general, the propensity to use incentives is an increasing function of the number of adopted innovations. When the groups which adopted only one innovation are compared, the group with the highest propensity to use an incentive may differ according to the type of incentive.

In conclusion, the incentives represent a useful and effective policy tool to stimulate the adoption of Industry 4.0 technologies by Italian SMEs and positively contribute to Sustainable Development Goals.

Acknowledgements

The authors thank the University of Ferrara, which funded the project entitled "Public policies, 4.0 technologies and enterprise performance. Empirical analyses on a representative sample of manufacturing enterprises of northern Italy (Politiche pubbliche, tecnologie 4.0 e performance d'impresa. Analisi empiriche su un campione rappresentativo di imprese manifatturiere del Nord Italia)" for the period 2022-2024, with the Departmental Research Incentive Fund - FIRD 2022.

References

Agresti, A., Klingenberg, B. (2005). Multivariate tests comparing binomial probabilities, with application to safety studies for drugs. *Journal of the Royal Statistical Society Series C: Applied Statistics*, **54**(4), pp. 691-706.

Ashima, R., Haleem, A., Bahl, S., Javaid, M., Mahla, S. K., Singh, S. (2021). Automation and manufacturing of smart materials in Additive Manufacturing technologies using Internet of Things towards the adoption of Industry 4.0, in *Materials Today: Proceedings*, **45**, pp. 5081-5088.

Bonnini S., Corain L., Marozzi M., Salmaso L. (2014). *Nonparametric Hypothesis Testing, Rank and Permutation Methods with Applications in R*. Wiley series in probability and statistics.

Bag, S., Gupta, S., Kumar, S. (2021). Industry 4.0 adoption and 10R advance manufacturing capabilities for sustainable development. *International Journal of Production Economics*, **231**, 107844.

Chuang-Stein, C., Mohberg, N (2019). A unified approach to the analysis of safety data in clinical trials, in *Drug Safety Assessments in Clinical Trials, G.* Sogliero-Gilbert (ed.). Dekker, New York.

Cucculelli, M., Dileo, I., Pini, M. (2022). Filling the void of family leadership: Institutional support to business model changes in the Italian Industry 4.0 experience. *Journal of Technology Transfer,* **47**, pp. 213-241.

Cugno, M., Castagnoli, R., Büchi, G. (2021). Openness to Industry 4.0 and performance: The impact of barriers and incentives. *Technological Forecasting and Social Change*, **168**, 120756.

Durão, L F.C., Christ, A., Zancul, E., Anderl, R., Schützer, K. (2017). Additive manufacturing scenarios for distributed production of spare parts. *The International Journal of Advanced*

Manufacturing Technology, **93**, pp. 869-880.

Pesarin, F. (2001). *Multivariate Permutation Tests with Applications in Biostatistics*. Wiley: Chichester.

Pesarin F., Salmaso L. (2010). *Permutation Tests for Complex Data: Applications and Software*. Wiley Series in Probability and Statistics.

Raj, P.E., Wahab, S.A., Zawawi, N.F.M., Awang, K.W., Ibrahim, W.S.A.A.W. (2020). The benefits of Industry 4.0 on Sustainable Development and Malaysia's Vision. In *IOP Conference Series: Earth and Environmental Science*, **549**(1), p. 012080). IOP Publishing.

Stornelli, A., Ozcan, S., Simms, C. (2021). Advanced manufacturing technology adoption and innovation: A systematic literature review on barriers, enablers, and innovation types. *Journal of Research Policy*, **50**(6), pp. 104-229.

doi.org/10.26398/asaproc.0011

Italian wines: Ratings & reputations

Eugenio Brentari
Department of Economics and Management, University of Brescia, Brescia, Italy.

1. Introduction

The objective of this paper is to compare two different types of wine evaluation and to measure the concordance between these two different ways of evaluating wine. In particular, reference will be made, firstly, to the Altroconsumo wine guide, where the evaluations of Italian wines are the result of various analyses and assessment made blindly by expert judges; and, secondly, to the Vivino app used by millions of consumers around the world, where the evaluation is made knowing all the characteristics of the wine tasted and also the evaluations of other consumers.

2. Altroconsumo wine guide and Vivino app

From 2006 to 2021, Altroconsumo, the most widespread consumer protection and defence association in Italy, published its own wine guide. This guide considers every year about 300 Italian wines including reds, whites, and rosés, and each year has a dossier dedicated to a particular type of wine (Prosecco, Franciacorta, Lambrusco, etc.).

They are all wines that are generally not very expensive (under 20 €) and chosen to represent all the various types of wines classified according to vineyard, producer, region of origin. Moreover, these wines are frequent in large scale retail trade and wine shops and, in recent years, can be found also online.

The ratings in the Altroconsumo guide are the result of a detailed analysis of the wine: The information contained on the labels (producer, varietal, vintage, origin, colour, alcohol degree), the chemical characteristics (free and total SO2, Ph, …), the sensory characteristics (obtained from a panel of experts) and the market price.

An important note is that the evaluation of the wines in this case is made blindly by the experts, i.e., without any information about the wine and using advanced experimental sensory analysis methods.

At last of this evaluation, there is a rating scale rating from 1 to 100, divided into four bands: a score less than 50 means that it is a poor quality wine; if the score is from 50 to 59, we have medium quality wine; from 60 to 69, a good quality wine; from 70 on, an excellent quality product (Figure 1).

The other wine evaluation tool considered in what follows is the Vivino app, the world's largest wine app for platform sales. Founded as a startup in 2010 and launched as an app in 2011 by Heini Zachariassen and Theis Søndergaard, two Danish entrepreneurs, this app is available in six languages (English, Italian, Deutsch, French, Spanish, and Portuguese) since 2014.

With over 50 million downloads, and more than 13 million wines from over 231,000 wineries, the app has millions of ratings and reviews. Vivino makes hundreds of thousands of wines available for purchase in 17 markets worldwide.

The Vivino app uses an ordinal categorical rating scale similar to Likert's, from one to five stars, and also allows half a point. In addition to the average price and average rating, Vivino reports consumer reviews, taste characteristics, scents of wine, even food pairings, and information on the winery that produced it.

Figure 1: How to read a data sheet in the Altroconsumo guide (Wine Guide 2020)

Over the years, the distribution of Vivino scores was found to have a close-to-normal distribution, with mean and median scores both being 3.6. Surprisingly, 15% of the wines had a score of 4 (85th percentile).

Thanks to the large number of users of Vivino, the ratings of each wine on the app can be seen as a kind of reputation that the wine enjoys.

One question arises at this point: how do these two instruments behave with respect to the same wine, the two evaluation methodologies being quite different?

Or rather, is a wine reputation (resulting from consumers' judgments) a synonym for its quality (according to consumers)?

To answer these questions a simple statistical tool was considered in this paper: The analysis of correlation between Altroconsumo ratings and Vivino ratings.

3. The dataset and results

In the present study, a sample of 300 Italian wines (134 white wines, 160 red wines, and 6 rosés) was taken into consideration. The scores assigned to them by the Altroconsumo guide and those on the Vivino app were collected and compared.

As is well known, the Bravais-Pearson linear correlation coefficient r takes values from -1 to 1; positive values indicate agreement between the two assessments, while negative values indicate disagreement between the two assessments. Values close to zero indicate no relationship (Piccolo, 2010).

As can be seen in Figure 2, the assessments of Altroconsumo and Vivino are far apart: The value of r is close to zero (0.068).

An emblematic case is the first value on the left side of Figure 1, which appears as an outlier. It is a *Lambrusco Modena DOC Vino Frizzante Secco* (2015 vintage) from a well-known company that scored extremely low for Altroconsumo (21/100) while Vivino evaluations assign to it an average score of 3.3/5.

Inversely, the lowest rating given by Vivino was 2.4/5 for a *Sangiovese Merlot Rubicone IGP Assieme* red wine (2014 vintage), whereas the Altroconsumo's score for the same wine was 57/100.

The first discouraging conclusion is trivial: The Vivino evaluations are very distant from those

expressed in the Altroconsumo wine guide.

Even a separate analysis for white and red wines did not yield particularly interesting results: For white wines the value of the linear correlation coefficient *r* is 0.105, while for red wines it is 0.084. Only in the case of rosés the value of *r* turns out to be somewhat higher (0.33) but only six rosé wines were considered.

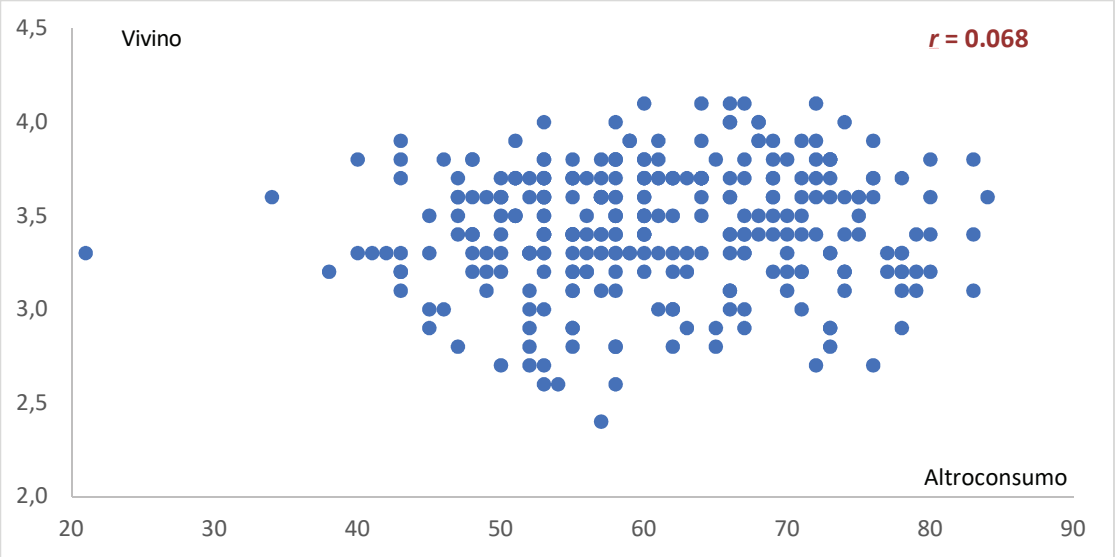


Figure 2 *Scatter plot of rankings of Altroconsumo Guide and Vivino App*

To understand the reason for this difference in ratings, the wines considered by Vivino as of medium quality (scores between 3 and 4) were studied. All the *r*-values are still extremely low: when all wines are considered, the *r*-value is 0.033, while if only white wines are considered, *r* = 0.021, and for red wines, the *r*-value increases slightly but only stands at 0.073.

The evaluation of medium-quality wines is particularly difficult if we consider the mid-range wines, i.e., those judged as such by both Altroconsumo and Vivino (for Altroconsumo scores from 50 to 70; for Vivino from 3 to 4), the value of the *r* coefficient further reduces from 0.033 to 0.030.

Hence, the panorama is certainly not exciting.

However, it should be noted that if one considers only the high-end wines and the low-end wines, in this case, the evaluations of Altroconsumo and Vivino are more in agreement with each other. In fact, if we consider the scores assigned by Vivino higher than or equal to 4 and lower than 3, the linear correlation coefficient *r* increases significantly, especially for white wines where it reaches a value of 0.501, while for red wines it stands at 0.348 and for all types of wine it is just under 0.3 (Figure 3).

Some improvement is obtained if the average wines of Altroconsumo are kept aside, but in this case, the residual number of wines is very small. For example, a very high *r*-value is obtained for red wines, but only three wines are considered.

Also, an attempt was also made to calculate the *r*-index jointly for all wines considered to be either very good or much less good for both Altroconsumo and Vivino. For example, by excluding all those wines within a circle with a center in the medians of the two distributions, the results change little and even in this case, are calculated on a small number of wines and therefore not very reliable for the present study.

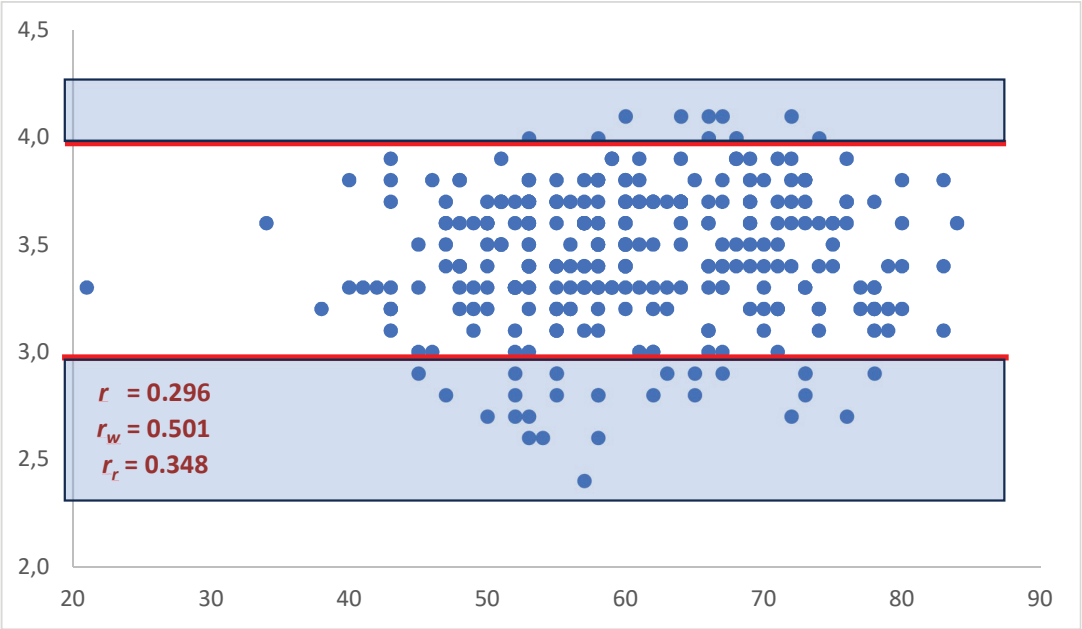


Figure 3 *Calculation of* r *for very good and bad wines of Vivino App*

An alternative tool for the analysis is the between-rank Spearman index ρ, which measures the degree of agreement or disagreement between two rankings.

In the presence of rankings with peer units, as in the present case, the Spearman index can be easily computed. The interested reader can refer, among the others, to Leti (1983, pp. 808-812). By means of this index, the analyses done previously were traced back with quite similar results.

If we analyse all the 300 wines, the Spearman index takes on a value $\rho = 0.121$, which is certainly higher than the value of the linear correlation coefficient ($r = 0.068$), but not enough to state that the evaluations of Altroconsumo and Vivino agree.

For white wines alone, the index value slightly increases ($\rho = 0.118$, very similar to $r = 0.105$), while for red wines alone, it is $\rho = 0.082$, practically the same value obtained with the linear coefficient ($r = 0.084$).

Things somewhat change when only the best and worst wines in the two rankings are considered together. All the values of ρ calculated are significantly higher than those of r, a sign that there is some agreement between the two rankings on good and bad wines.

Considering white and red wines together, the Spearman index takes the value $\rho = 0.359$. The value of the index increases when only red wines are considered; in fact, for red wines, the index is 0.410. Among the 20 red wines considered, there is little agreement on the best wines, but the rankings agree on the worst ones.

A surprising agreement qualifies white wines: the index is close to one ($\rho = 0.977$). Both evaluation guides agree, for instance, in considering a *Lugana DOC* from 2018 as the best among the white wines considered; just as they agree in considering a *Falanghina Beneventano IGP* (2014 vintage) as the worst.

Even with the Spearman's index, if one considers only the high-end wines and the low-end wines, the evaluations of Altroconsumo and Vivino are more in agreement with each other.

It should be emphasized that knowledge of wine market characteristics may heavily condition wine evaluations. In particular, knowledge of the price is positively correlated with judges' scores and this may be the case with Vivino's ratings (Table 1). In fact, while Altroconsumo's evaluations, obtained in cases the judges do not know the price of the wine, have a very low (indeed negative) correlation with the price. The same can be noticed regarding the alcoholic degree: the knowledge

of the alcoholic degree while evaluating a wine leads to a linear correlation coefficient $r = 0.429$; instead, a much lower correlation ($r = 0.241$) is shown between alcoholic degree and the Altroconsumo's evaluations.

	Alcoholic grade	Price	Altroconsumo quality	Vivino quality
Alcoholic grade	1			
Price	0,194	1		
Altroconsumo quality	**0,241**	**-0,056**	1	
Vivino quality	**0,429**	**0,313**	0,068	1

Table 1: Estimates of r between wine characteristics

4. Conclusion

In conclusion, it seems very difficult to give an objective assessment of wine without established scientific criteria, such as those used for example in sensory analysis.

A striking example comes from the evaluation of the wine considered as a "best buy" and "best test" in the Altroconsumo wine guide of 2021. In that year, the best wine was a 2018 Cirò DOC white wine from the *Calabrian company Caparra & Siciliani*, which scored 76/100 (the highest score in that guide) and the price was only € 3.64 (Figure 1). The same wine for Vivino consumers was rated 2.5/5 stars, highlighting once again the difference in judgement between the Altroconsumo and Vivino approaches.

I like to recall the advice given by Giovanni (Gianfranco) Manzoni (1923-2010), director for many years of the San Michele all'Adige Agricultural Institute (Trento, Italy), who answered the question "*How do you judge if a wine is good?*" simply by saying "*If you like it, it is good!*".

Acknowledgments

The author wishes to thank the anonymous referees for their comments concerning the analytical techniques adopted in this paper. Moreover, the author expresses his gratitude to colleagues and friends Luigi Fabbris, Maria Gabriella Grassia and Alfonso Piscitelli for other suggestions.

References

Altroconsumo (2006 - 2021). *Wines Guide*. Altroconsumo Edizioni S.r.l., Milano
Leti, G. (1983). *Statistica descrittiva*, Il Mulino, Bologna
Piccolo, D. (2010). *Statistica*, Il Mulino, Bologna
Vivino (2023). https://www.vivino.com/app

doi.org/10.26398/asaproc.0012

The new preventive alert mechanism in the Italian Crisis Code: A statistical approach for predictive insolvency analysis

Eugenio Brentari, Gloria Ermici, Andrea Zanetti
Department of Economics and Management, University of Brescia, Brescia, Italy

1. Introduction

The Italian bankruptcy law has always provided tools to act on companies that were already in a state of crisis, instead of creating a preventive alert mechanism. As a result, the bankruptcy law reform and the Crisis Code were introduced (Legislative Decree 14/2019).

This problem has been studied in detail by various authors in order to create ratings or predictive tools to anticipate the business crisis and the analysis of business continuity and especially the going concern (Chye Koh and Kee Low, 2004).

The "Going Concern Principle" is "the accountant's firm model" (Sterling, 1968) and is the basis for the financial statements and represents, according to the international standard on auditing ISA 570[1], that the financial statements are prepared on the assumption that the entity is a going concern and will continue its operations for a foreseeable future. It is a mandatory requirement for the Italian Accounting Organism (2018) and the Italian Civil Code[2] for the Italian financial statement.

One of the consequences of the lack of going concern is the company's economic and financial disequilibrium that leads the company to bankruptcy. For example, according to ISTAT's data (Istat, 2023), bankruptcy in Italy in the first quarter of 2023 has had a 19.1% increase compared to the previous quarter.

Altman's Z score model is one of the best known international predictive models (see Altman, 1968; 1970; 1993; 2002; Altman and Hotchkiss, 1995) and many authors proposed to adapt it to the Italian prospective (Alberici, 1975; Altman, Danovi and Falini, 2013). However, the Altman's Z score model is not well applicable to the Italian company structure and size, characterized by micro and small enterprises (Bottani, Cipriani and Serao, 2004).

The new code has delegated to the *Consiglio Nazionale dei Dottori Commercialisti ed Esperti Contabili* (National Council of Chartered Accountants and Accounting Experts – CNDCEC, 2019) the task of developing warning indicators able to promptly detect and prevent the state of insolvency of companies in compliance with business continuity. The companies with the major prejudicial out of range indicators have to communicate the crisis state to the Chamber of Commerce.

The CNDCEC also determines five more indicators. If all these five indicators are out of range, the company has to communicate the crisis state to the Chamber of Commerce.

The application of the New Crisis Code, introduced in 2019 and subsequently amended, has seen the extension of the administrator's responsibility, implying the need to adopt adequate organizational, administrative and accounting structures that are able to detect insolvency situations and to adopt corrective elements.

In July 2022 the New Crisis Code was modified, introducing different indicators, with a focus

[1] International Standard on Auditing 570 (2016), *Going Concern*. https://www.ifac.org/_flysystem/azure-private/publications/files/ISA-570-(Revised).pdf (Last access: 16/05/2023)

[2] Art. 2423 bis Italian Civil Code, "the valuation of the items has to be done according to prudence and in view of the continuation of activity".

on cash and payables. The previous indicators remain a tool for companies to check their organizational, administrative and accounting structures.

This paper aims to verify: *Are the indicators determined by the Italian law in 2019 the right tools to predict the companies' failure?*

2. Indicators

The selection of the indexes was carried out by the CNDCEC (2019) with the collaboration of the Central Financial Statements of Cerved, the database of the Italian chambers of commerce.

For this selection, it was necessary to consider a group of indicators that, combined with each other, were able to select a cluster of companies close to insolvency.

The objective was to identify a combination of indicators which were representative of income, equity or financial imbalance.

The selection steps are as follows:

- Univariate analysis of the indices: analysis of the forecasting capacity of the default of each index. With this approach, the most representative indicators have been identified for each area.
- Multivariate analysis of the indices of the previous point: identification of the signals which, when suitably combined, were able to achieve the preventive warning objective.

An objective quantitative approach has therefore been used. This approach is based on the computation of the median of the index concerning the subjects who presented a default situation in the following 36 months.

There were different critical thresholds, depending on the companies belonging to different sectors. For this reason, specific alert indices have been introduced.

3. The dataset

The dataset includes data about 20,000 small, micro and medium-sized enterprises (SME)[3] with headquarters in the province of Brescia, Italy.

The dataset was downloaded by the AIDA dataset, following the definition by the European Union Commission Recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises. The sample was determined with some mandatory requirements, as the headquarters in the province of Brescia and available financial statements concerning 2020.

Additional sample requirements were used to select only the micro enterprises as defined by the European Union Commission, that represents in 2020 the 94.8% of Italian active enterprises (Istat, 2022).

These additional requirements consider the enterprises with two of these three thresholds in 2020 financial statements:

- Total production value[4] (A) < 2,000,000 EUR
- Employees < 10
- Total balance sheet assets[5] < 2,000,000 EUR

[3] Small, micro and medium-sized enterprises, as defined by the European Union Commission Recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises.

[4] *Total Production Value* is the sum of all the components of the Group A of the Economic Statement (Revenues from sales and services, change in work in process and finished goods, change in work in process on long-term contracts, internally generated fixed assets and other revenues and income except of financial incomes).

[5] *Total balance sheet* is the sum of Receivables from Shareholders for Share capital, Fixed assets, Current Assets and Accruals and prepayments.

Each record considers economical e financial data from the 2020 financial statements published by Brescia's Chamber of Commerce. Firstly, the dataset calculates net equity > 0, prejudicial indicator about the going concern of each SME. The dataset includes the early-warning indicators chosen by the CNDCEC with the thresholds calculated for the manufactured and industry business:

1. Financial expenses / Revenues from sales and services < 3.0%
2. Net equity / Total balance sheet assets > 7.6%
3. (Current assets + Cash and cash equivalents) / Short-term liability > 93.70%
4. Cash flow / Total balance sheet assets > 0.5%
5. Tax and social security debts / Total balance sheet assets > 4.9%

Some of the SMEs' indicators are not available because of the simplified Italian balance sheets required for the SMEs. Especially the fifth indicator is available for only 161 SMEs. Therefore, for the data analysis and for the last results, it turned out to be necessary to consider the only first four indicators.

We proceed by comparing the results determined by the indicators of the 2020 financial statements with:

- Failed companies list from the *Court Law of Brescia*[6];
- Inactive or ceased companies list from the *Brescia Chamber of Commerce*.

The reason of these comparisons is connected with the research question: Are t*he indicators determined by the Italian Law the right tools to predict the companies' failure*?

4. Results

Table 1 shows the results of the prejudicial indicator (net equity > 0) and the following four indicators.

Indicator	*Total Companies in the sample* (with data available) (A)	*Number of out of range Companies* (B)	*% of the sample* (B/A)
Net equity > 0	19,999	2,389	11.95
Financial expenses/Revenues from sales and services < 3.0%	16,732	2,872	17.16
Net equity/Total balance sheet assets > 7.6%	19,855	4,933	24.85
(Current assets + Cash and cash equivalents)/ Short term liabilities > 93.70%	19,787	4,837	24.45
Cash flow/Total balance sheet assets > 0.5%	19,961	6,752	33.83

Table 1 *Result of the analyzed indicators*

Data show that the prejudicial indicator is out of range for 2,389 companies. With the Italian law, these companies (about 12% of the total sample) have to begin the early warning procedure with the local chamber of commerce.

The other companies (17,610) must verify how many indicators are out of range. This test is

[6] The data includes the failures from 01/01/2020 to 31/10/2022 declared by the Court of Law of Brescia

shown in the Table 2. The 65 SMEs with four out of range indicators must begin the early warning procedure with the chamber of commerce. With the application of the new Italian Law, the 12,27% of the total sample must begin the early warning procedure.

Number of out of range indicators	*Companies* (A)	*% of the sample* (A/17.610)
0	8,472	48.11
1	4,000	22.71
2	1,948	11.06
3	644	3.66
4	65	0.37
Lack of 1 or more indicators[7]	2,481	14.09

Table 2: ***Out of range indicators***

Number of out of range indicators	*Companies* (A)	*% of the sample* (A/1,600)
0	186	11.63
1	158	9.88
2	160	10.00
3	171	10.69
4	65	4.06
Lack of 1 or more indicators	860	53.75

Table 3: ***Inactive or ceased companies***

After these analyses, how many companies have really ceased or failed? We compare this data with:

- the failed companies list of the *Brescia Court of Law;*
- inactive or ceased companies list from the *Brescia Chamber of Commerce*.

The companies which are declared failed by the Brescia Court of Law in the period 01/01/2020 – 31/10/2022 are about 399. None of these are included in our sample. This point could be explained by the Italian Failure Law, which declares that some dimensional requirements are necessary for the declaration of failure. These requirements exclude most of the micro and small enterprises. With the sample of the SMEs in Brescia we can analyze the inactive or ceased companies in 2020, 2021 e 2022 and their out of range indicators with Table 3.

Number of out of range indicators	*% of the total sample* (17,610)	*% of the inactive or ceased Companies (1,600)*	*Difference*
0	48.11	11.63	-36,48
1	22.71	9.88	-12,83
2	11.06	10.00	-1,06
3	3.66	10.69	+7,03
4	0.37	4.06	+3,69
Lack of 1 or more indicators	14.09	53.75	+39,66

Table 4: ***Comparison of inactive or ceased companies (total sample)***

[7] The simplified balance sheet for the Italian SMEs doesn't allow to calculate some of the indicators.

Data show a higher presence of out of range indicators in the inactive or ceased companies list if compared to the total sample. Data are also shown in Table 4.

5. Conclusions

This paper analyses the 2020 financial statements of the SMEs in Brescia with the failure companies list of the Court of Law of Brescia and the inactive or ceased companies by the Brescia's Chamber of Commerce. None of the companies in the sample failed in 2020, 2021 and 2022: this point is not enough to declare that indicators are not the right tools to predict the crisis state for the Italian SMEs.

One of the possible causes could be the technical and legal times for the bankruptcy declaration by the Court of Law, especially in 2020 with the suspension introduced by the Law decree 23/2020 in the Covid-19 period (09.03.2020 – 30.06.2020)[8].

A further possible cause could be that the lack of going concern emerges after the three years analyzed, therefore, for example, in 2023 or 2024. For this reason, the research could be extended with the 2023 and 2024 bankrupt data from the Brescia Court of Law.

A final possible cause could be the non-compliance by companies with the reporting obligation and absence of calculation of the indicators.

Data also show a correlation between the presence of out of range indicators and ceased or inactive companies. These indicators seem not to be the best tools to predict the companies' failure, especially for the SMEs with the Italian simplified balance sheets, but in our opinion and by analyzing data, these are a step forward for the introduction of prediction analysis and for business management.

The research could be extended with the 2021 or 2022 financial statements or with the analysis of others Italian provinces. Furthermore, the research could be extended with the new indicators introduced in the 2022 modifications.

References

Alberici, A. (1975). *Analisi dei bilanci e previsione delle insolvenze: affidabilità bancaria e informativa del mercato mobiliare*. Isedi, Milano

Altman, E.I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *The Journal of Finance*, **23**(4). pp. 589-609.

Altman, E.I. (1970). Ratio analysis and the prediction of firm failure: A reply. *The Journal of Finance*, **25**(5), pp. 1169-1172.

Altman, E.I (1993). *Corporate Financial Distress and Bankruptcy: A Complete Guide to Predicting & Avoiding Distress and Profiting from Bankruptcy*, Wiley, New York.

Altman, E.I. (2002). Revisiting credit scoring models in a Basel 2 environment (May 2002). *NYU Working Paper No. S-FI-02-11*, SSRN: https://ssrn.com/abstract=1298829.

Altman, E.I., Falini, A. Danovi, A. (2013). Z-score models' application to Italian companies subject to extraordinary administration (May 1, 2013). *Bancaria*, **04-2013**, SSRN: https://ssrn.com/abstract=2275390.

Altman, E.I., Hotchkiss, E (1995). *Corporate Financial Distress and Bankruptcy: A Complete Guide to Predicting and Avoiding Distress and Profiting from Bankruptcy,* Wiley, New York.

[8] D.L. 23/2020, art. 10, "*All appeals pursuant to articles 15 and 195 of the R.D. 16 March 1942, n. 267 and 3 of the legislative decree of 8 July 1999, no. 270 deposited in the period between 9 March 2020 and 30 June 2020 are not proceedable".*

Bottani, P., Cipriani, L., Serao, F. (2004). Il modello di analisi Z-score applicate alle PMI, *Amministrazione & Finanza,* **1**/2004, pp. 50-53.

Chye Koch, H., Kee Low, C. (2004). Going concern prediction using data mining techniques. *Managerial Auditing Journal*, **19**(3), pp. 462-476.

Consiglio Nazionale dei Dottori Commercialisti ed Esperti Contabili (2019). *Crisi d'impresa – Gli indici dell'allerta, 20/10/2019*. http://press-magazine.it/wp-content/uploads/2019/10/codice-della-crisi_definizioni-indici_ott-2019.pdf.

International Standard on Auditing 570 (Revised) (2016). *Going Concern*. https://www.ifac.org/_flysystem/azure-private/publications/files/ISA-570-(Revised).pdf.

ISTAT (2022). *Annuario statistico italiano 2022*. Chapter 14, Istat, Roma https://www.istat.it/storage/ASI/2022/capitoli/C14.pdf,

ISTAT (2023). *Registrazione e fallimenti di imprese, I trimestre 2023* Istat, Roma https://www.istat.it/2023/05/fallimenti_imprese_I-TRIM-2023.pdf.

Italian Accounting Organism (2018). *OIC 11, Finalità e postulati del bilancio d'esercizio*, Marzo 2018. https://www.fondazioneoic.eu/wp-content/uploads/2011/02/2018-03-OIC-11.pdf

Sterling R. R. (1968). The going concern: An examination. *The Accounting Review*, **43**(3), pp. 481-502.

doi.org/10.26398/asaproc.0013

Road accidents in Italy: New indicators, at province level, based on geographic information system open data

Marco Broccoli[a], Silvia Bruzzone[b]
[a] Directorate for Methodology and Statistical Process Design, Istat, Rome, Italy
[b] Directorate for Social Statistics and Welfare, Istat, Rome, Italy

1. Introduction

Road safety performance indicators (RSPIs) give a multidimensional approach for accidents investigation concerning roads, vehicles and persons involved. Combining the use of statistical surveys, administrative geographical information systems (GIS) and big data (BD) sources, the result gives new elements on planning infrastructure solutions and applying policies to reduce deaths and serious injuries.

Nowadays there is a clear information bias as regards the appropriate reference denominators placed as basis in construction of statistical indicators linked to road accidents. Resident population, used as a common proxy for exposed at risk in a specific geographical area, usually is not an appropriate solution due to the seasonal nature of road accidents. Vehicle fleet might provide a comprehensive overview of the vehicles present within a certain region and should give a more accuracy on the statistical indicators. The "deductible distortion" is primarily attributed to the mobility of road users. The length of the road network gives a consistent first set of information concerning the different territories. This information is not available from official statistics at national level. Although there are archives and detailed road graphs for each municipality, province and region, a harmonised and systematic national road registry has not been established yet.

The present research focuses on the use of existing administrative sources, the scouting of new sources and the analysis of integrated and auxiliary data to calculate road accidents rates, mortality and harmfulness indexes, comparing these measures to the correspondent length in meters of carriageway by road direction from the Open Street Map (OSM). This approach has been used for the first time, by the authors, to build road accidents indicators and adopted by the "experimental statistics" at Istat (Broccoli and Bruzzone 2019; 2021). Although OSM is mentioned as a data source in many scientific articles, the main purpose was for GIS-based traffic simulation (Zilske et al., 2011). The final purpose of the project is to expand statistical information with the supply of traffic flows (vehicles / km) on the national road network. This would make it possible to calculate the probability of being involved in the accident, considering the different exposure to risk of accidents.

2. Methodology

GIS is a geographic system designed to capture, store, manipulate, analyse, manage and present spatial or geographic data. To relate information from different sources, GIS uses a spatial location as the key index variable (key reference by position). Just as a relational database containing text or numbers can relate many different tables using common key index variables, GIS can relate otherwise unrelated information by using location as the key index variable. This key characteristic of GIS is nowadays a valid alternative on statistical dissemination. Any variable that can be located spatially using an x, y, and z coordinates, representing longitude, latitude, and elevation, respectively. These GIS coordinates may represent other quantified systems of territories (polygons), road networks (lines) and point of traffic (points).

"Join attributes by location"[1] is the algorithm that takes an input vector layer and creates a new vector layer, adding on the extended version the additional characteristics of the second layer. Specifically, shapes of the Italian localities, wide areas on which one or more houses are scattered, were used (2011 census latest available for the whole country in Italy) and consequently the Italian municipalities[2] were aggregated according to the locality layer. The aim of the project is to harmonise the location statistical variable of the Istat survey with the class of road graph. Therefore, the street layer is harmonized with the classification of the accidents location and permits to create new safety indicators by road type. The localities are classified as: urban areas, small-inhabited areas, productive areas and widespread houses[3] [4].

OSM[5] is a collaborative project aimed on creating free content maps of the world. The project aims at a collection world of geographical data, with the main purpose of creating maps and cartography. The key feature of the geographic data present in OSM is having a free license, the Open Database License. It is therefore possible use them freely for any purpose with the only constraint of mentioning the source. Everyone can contribute by populating or correcting data. The maps are created using the data recorded by portable GPS devices, aerial photographs, and other free sources.

Although OSM is an **o**pen-**s**ource tool based on information from a community, the product provides data to be considered reliable and consistent, so much that a major part of GPS Android and iOS navigation software on portable devices are powered by OSM, for example WisePilot, Maps.me, NavFree, Scout, ShareNav, MoNav, Navitel, etc.

The OSM vector layers, used in the present research, daily updated and free downloadable data consist in: road graph; point of traffic (POT); buildings; use of the land; natural places; POWF (point of worship); POIS (Point of interest); railways; transport; water areas; and water ways.

Table 1 contains the list of the main different type of road arch by OSM used to provide the length in meters of carriageways by road type and Italian provinces. The arch road types selected to calculate the indicators analysed in the present research are referred to the motorized vehicles flow: motorway, trunk, primary, secondary, tertiary, unclassified, residential, living street, motorway link, trunk link, primary link, secondary link, tertiary link, service, unknown. Pedestrian, track, track grade, bridleway, cycle way, footway, path, steps, also defined in OSM road types, are not considered part of the observed domain.

Besides, to calculate road accidents indicators, with denominator represented by the arches length in OSM, we built a "bridge matrix" between road categories, classified by functional road type and the categories linked to the roads holder, used by Istat road accidents survey. The systematic classification of road arches, classified in the categories used by Istat, has been modified, in respect of a first release (Broccoli and Bruzzone 2019; 2021).

A new analytical classification has been adopted, using a more refined technique of attributing single road arches, about three and a half million in total, to the Istat classification groups (Table 2).

---

[1] Join attributes by location is a functionality of QGIS software. The authors adapted the implementation of the algorithm for the specific data used in the present study - 27.1.17. Vector general — QGIS Documentation documentation

[2] Italian municipalities codes list at 1/1/2023 https://www.istat.it/it/archivio/6789

[3] 1.Urban area is an aggregate of contiguous or near houses with roads, squares and similar, or however areas characterized by services or public activities. 2.Small-inhabited area is an area without a place of collection, and it is based on a group of at least fifteen contiguous and near buildings, if the distance between the buildings does not exceed thirty meters and it is lower than the distance between the centre and the nearest of the houses clearly scattered. 3.Productive area is an extra-urban area not included in the centres or residential areas with more than 10 local units. 4.Widespread houses represent an area with scattered house in the municipal territory at a distance not enough to constitute a built-up area.

[4] Istat - Basi territoriali e variabili censuarie https://www.istat.it/it/archivio/104317 ; Istat - Descrizione dei dati geografici e delle variabili censuarie delle Basi territoriali per i censimenti: anni 1991, 2001, 2011 Descrizione Dati Geografici (istat.it)

[5] OpenStreetMap provides geographic data on thousands of websites, mobile and hardware devices. OpenStreetMap is built by a community of mappers, who contribute, update and monitor data on roads, cafes, railway stations and much more, all over the world - OSM: OpenStreetMap

Table 1: ***Open Street Map main road arches classification***

OSM Road type	Road type Description
Motorway	Free or by toll payment highways, including motorway connections.
Motorway Link	The link roads (slip roads/ramps) leading to/from a motorway from/to a motorway or lower-class highway. Normally with the same motorway restrictions
Trunk	Roads type between Motorway, Motorway connections and Primary Road. The junction section of a motorway-ring road that leads to the city centre can also be classified as trunk.
Trunk Link	The link roads (slip roads/ramps) leading to/from a trunk road from/to a trunk road or lower-class.
Primary	Roads of national and regional importance not classified as motorways, trunks, or their link. They connect the main cities to each other. Mostly, they are classified as SS (Main Roads) or SR (Regional).
Primary link	The link roads (slip roads/ramps) leading to/from a primary road from to a primary road or lower-class highway.
Secondary	Secondary Roads of regional and provincial importance. They link together the main municipalities of a region. They are, usually classified as SP (provincial roads) but there are some exceptions. In urban areas, they are normally classified as important streets with two lanes in each direction
Secondary link	The link roads (slip roads/ramps) leading to/from or from/to a secondary road or lower-class highway.
Tertiary	Roads of local rank. They connect smaller municipalities together. In urban areas, they are side roads to primary and secondary roads with a medium flow of traffic.
Tertiary link	The link roads (slip roads/ramps) leading to/from or from/to a tertiary road or lower-class highway.
Residential	Roads in a residential area, which serve as an access to housing, without function of connecting settlements.
Living Street	Residential road where pedestrians have legal priority over cars, speeds are kept very low.
Service	Access roads or internal service areas, beaches, camping, industrial areas, shopping centres, residences, parking places, landfills, installations, etc.
Unknown/Unclassifie	Not classified

Table 2*: Bridge coding table between roads arches classification by OSM, localities and road type (a)*

Road Arches classification by Open Street Map	Localities at Census 2011			
	Urban areas + Small		Productive areas + Wide	
	Road Localisation by Istat Road accidents survey			
	Motorways	Urban Roads	Motorway	Rural Roads
Motorway and Trunk	x		x	
Primary, Secondary and Tertiary		x		x
Residential and Living Street		x		x
Motorway Link and Trunk link	x		x	
Primary, Secondary and Tertiary Link, Service		x		x
Unknown		x		x

3. Data and results

Using data from the OSM source, has been done new experimental statistics on road mortality and injuries. It has been compared the most common rates at province level out of the resident population or vehicles fleet amount. The rates consist in a ratio between a numerator based on data from Istat road accidents survey and a denominator given by the length of the road network, by type or the average resident population in Italy or the vehicle fleet counted by ACI.

All road accidents are resulting in deaths or injuries during the years 2018, 2020 and 2022, involving at least a vehicle circulating on the national road network and documented by police.[6] Moreover, to enrich data, an additional information has been obtained from OSM, in particular, data on the traffic points detected on the road arches (PoT). New road accident indicators, "weighted" with the information on traffic intensity, were built too, considering the kilometres of carriageway with the presence of a traffic point on the arch.

Finally, a group of composed and synthetic road accidents indicators completes the set of indicators. The method used is the arithmetic mean (z-scores) MZ method (Software Ranker) [7],

[6] Istat - Road accidents in Italy Istat.it Road accidents

[7] Istat Ranker tool desktop software: http://www.istat.it/en/tools/methods-and-it-tools/analysis-tools/ ranker; I.Ranker

(Mazziotta and Pareto, 2013). The values of indices have been calculated according to the MZ (arithmetic average Z scores)[8] method for road arches, vehicle fleet and population. This method allows to build indicators excluding the traffic volumes on the road and to consider the cause of the accident in a different light, linked to the driving behaviour and the characteristics of the infrastructures and vehicles, with or without the traffic flow influence. The results show that the ranking of the Italian provinces, in respect of the new synthetic indicator, is modified if compared with the ranking of the single indexes and underlines a new profile of accidents risk at territorial level.

3.1. Comparison of indicators and risk of road accident ranking in Italian provinces

The proposed indicators refer to the ratio between road accidents, deaths, injuries, and vehicles to the road length, considering the additional information on points of traffic too, the resident population and the vehicle fleet. In addition to what already disseminated by *Istat Experimental Statistics* (Broccoli and Bruzzone 2019 and 2021), updated information on road accidents for the years 2018, 2020 and for the last available 2022 were used in the present study. A comparison between the three key-years representative for the previous, during and post pandemic period, was useful to explain the change in mobility habits and the difference through the country counties. Some evidence and results are presented in the following Section 3.2.

Concerning the different indicators built, is interesting to highlight the difference between the measures of the same events, road accidents ratio or road mortality ratio, for example, but considering different denominators. In support of this thesis, an interesting result is given by the covariance matrix results, obtained by means the synthesis of all indicators by year, province and road type and processed with the tool *Ranker* (Ref. Footnote n.5). The application of different weighting criteria leads, in fact, to very divergent results. Table 3 shows the covariance values, referred to the variation of each variable contained in the matrix in respect of all others. The values show, in fact, that the risk to be involved in a road accident, within the province of residence (0.2929-0.3712) or within the vehicle registration province (0.4138-0.4861) is significantly lower if compared to the indicator out of the length road arch (including PoT), where the accident occurred. The road accidents indicators, referred to the road length by province, therefore, allows to obtain a result closer to the road accident risk measures, using traffic flows data. The last data would represent, in fact, the best and correct denominator for road accidents rates.

Focusing on the year 2022 on the Road accidents indicators by road length "weighted" by the PoT information (number of accidents, vehicles involved, deaths and injuries per 100 kilometres of carriageway in the province) "adjusted" by the effect on accidents of the presence of traffic points on the roads. Some evidence shows that a maximum exposure to the risk of accidents and number of vehicles involved, for motorways and urban roads in main cities. Milan records a maximum for accidents rates per 100 km of carriageway, on motorways and urban roads (respectively 929.37 and 629.48 per 100 km), followed by Rome (851.44 and 447.97) and Genoa (703.51 and 492.74). On rural roads, in contrast, the medium-sized provinces result most affected, Monza and Brianza has the maximum value (518.55 road accidents per 100 km). Similar results are detected for vehicles and injuries, slightly different the results for road mortality rates that highlight Rome disadvantage on urban roads. Meanwhile Naples on rural roads and Rimini on the motorways. Considering the rates out of the resident population, the ranking is completely different, for road accidents rates (per 1 ml of inhabitants), in fact, on motorways the highest value is for the province of Savona, on rural roads, for Latina and Genoa on urban roads.

Milan and Rome, in the first positions for road accidents risk on motorways, in respect of road length, went down in 8th and 10th positions out of the resident population (11th and 12th

web application https://i.ranker.istat.it

[8] The three methods evaluated are: MZ - arithmetic mean (z-scores); MR - relative indices method (IR); and Mazziotta-Pareto index (Mazziotta and Pareto, 2013). The criterion applied after the robustness analysis is MZ.

position for urban roads) (Figures 1 and 2).

Table 3*: Covariance matrix between computing indicators (a)*

Years	2018			2020			2022		
Denominators Type	Population	Vehicle fleet	Road graph	Population	Vehicle fleet	Road graph	Population	Vehicle fleet	Road graph
Population	1.0000	0.9066	0.3712	1.0000	0.9148	0.2929	1.0000	0.8914	0.3688
Vehicle fleet		1.0000	0.4861		1.0000	0.4138		1.0000	0.4802
Road graph			1.0000			1.0000			1.0000

a) Results obtained by Istat Ranker application.

Figure 1: ***Road accidents rate by province. Year 2022 (per 100 km road length - Motorways)***

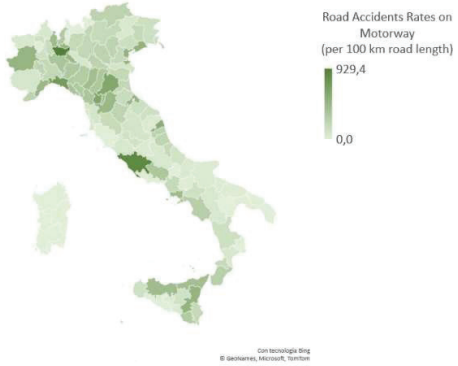
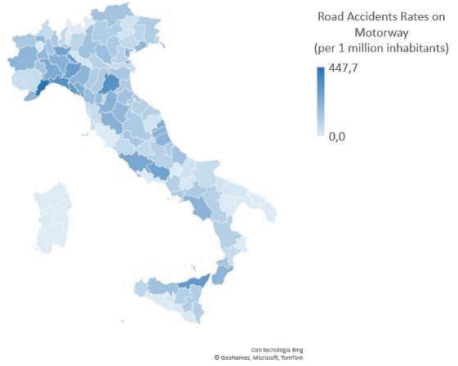


Figure 2: ***Road accidents rate by province. Year 2022 (per 1 million inhabitants - Motorways)***



The biggest provinces as Rome, Milan, Naples, Palermo and Turin with a high level of population lead to moderate values of indicators in comparison to reality. The high density of roads on these territories is correctly highlighted by road indicators ranking them as the worst provinces on exposure risk to be involved on an accident. At the same time, small provinces as Monza and Brianza and Varese, as the biggest one, get influenced on the difference between population and road extension on the two methods ranking. Traffic flows on this small area explode risk indicators to high-level values.

Some of medium size province as Grosseto, Savona, Rieti, Piacenza and Macerata using population reference give them a dangerous exposure risk harmonizing with their road infrastructure, the ranking gives a more accurate value with level of security better than the national average.

3.2. The effect of pandemic season on road accidents

During the year 2022, after the years in which the pandemic has seen the most acute phase, a clear recovery in mobility and in road accidents occurred. Road accidents, deaths and injuries compared on 2020, showed an overall growth above the months in which the traffic and mobility limitation measures enforced on containing of the virus. With respect to 2018, the road accidents and casualties in 2022, anyway, showed a decrease.

Concerning the comparison between the three years 2018, 2020 and 2022, the results in Table 4 show the ranking of the five best and worst performance by province in Italy of the synthesis indicators (method MZ processed by the Istat software Ranker, see footnote 5). Looking at the classification, Milan and Monza and Brianza appear constantly the provinces with the highest risk for the road safety, followed by Genoa, in 2018, the year of the collapse of Morandi Bridge, and by Rome, Naples, and Trieste in 2020 and 2022. The best performance, considering the compendium of the indicators, instead, is in towns of the southern regions: Basilicata, Sicily and Sardinia.

Table 4*: Best and worst performance in the road accidents risk. Ranking first and last five provinces by year (ranking by arithmetic average Z scores). Years 2018, 2020 and 2022*

Best Z Road graph performance (a)							
2018		**2020**		**2022**		**2022 No Traffic**	
Isernia	0.9593	**Potenza**	0.9613	**Isernia**	0.9953	**Isernia**	0.9356
Agrigento	0.9376	**Isernia**	0.9447	**Benevento**	0.9488	**Benevento**	0.8894
Benevento	0.9224	**Benevento**	0.9320	**Sud Sardegna**	0.9057	**Campobasso**	0.8519
Oristano	0.9133	**Agrigento**	0.9094	**Campobasso**	0.8877	**Sud Sardegna**	0.8514
Nuoro	0.8514	**Nuoro**	0.8566	**Potenza**	0.8765	**Potenza**	0.8460
Worst Z Road graph performance (a)							
2018		**2020**		**2022**		**2022 No Traffic**	
Naples	-1.6180	**Genoa**	-1.3306	**Trieste**	-1.8683	**Naples**	-1.8181
Rome	-1.8673	**Naples**	-2.1097	**Naples**	-1.9743	**Trieste**	-1.9996
Genoa	-2.1538	**Rome**	-2.2092	**Rome**	-2.2733	**Rome**	-2.2437
Monza- Brianza	-2.9951	**Monza -Brianza**	-2.5266	**Monza-Brianza**	-2.2891	**Monza - Brianza**	-2.8428
Milan	-3.3958	**Milan**	-2.9771	**Milan**	-3.3239	**Milan**	-3.8396

(a) The values of indices have been calculated according to the MZ (arithmetic average Z scores)

4. Conclusion

Big data are not generated directly for statistical purposes, for a statistical use, a suitable methodology must be able to link events to refer mainly to units of the population of interest for official statistics. Other elements required are to process data with the aim to collect them coherently with the statistical framework, to weight data to guarantee representativeness with respect to the target population and to estimate aggregates of interest including quality measures by spatial and time dimensions. The project of the use of OSM for the investigation of road accidents pattern is adherent to the features cited and the activities in progress may provide new developments in the future.

References

Broccoli, M., Bruzzone, S. (2019). Use of the open street map to calculate indicators for road accidents on the Italian roads. Updating with 2017 data. *Istat Experimental Statistics*, Rome https://www.istat.it/en/archivio/257384.

Broccoli, M., Bruzzone S. (2021). Use of open street map for accident investigation on the road and motorways networks. *Rivista Italiana di Economia Demografia e Statistica*, **LXXV**(4), pp. 157-168.

Farchi, S., Molino, N., Giorgi Rossi, P. et al. (2006). Defining a common set of indicators to monitor road accidents in the European Union. *BMC Public Health,* **6**, pp. 183.

Hakkert, A.S., Braimaister, L. (2002). The uses of exposure and risk in road safety studies, *Report documentation SWOV Institute for Road Safety Research,* Leidschendam (NL).

Harbison, F.H., Maruhnic, J., Resnick, J. R. (1970). Quantitative analyses of modernization and development, *Princeton University Press,* New Jersey, pp. 224.

Mazziotta, M., Pareto, A. (2013). Methods for constructing composite indices: One for all or all for one?. *Rivista Italiana di Economia Demografia e Statistica,* **LXVII**(2), pp. 67-80.

Retallack, A.E., Ostendorf, B. (2020). Relationship between traffic volume and accident frequency at intersections. *International Journal of Environmental Research and Public Health,* **17**(4), pp. 139.

Zilske, M., Neumann, A., Nage, K. (2011). OpenStreetMap for traffic simulation. In: *Proceedings of the 1st European State of the Map: OpenStreetMap Conference*, Wien: OpenStreetMap Austria, pp. 126–134.

doi.org/10.26398/asaproc.0014

Testing the implications of a different approach for estimating capital input on the computation of productivity measures

Francesca Brunaccini, Carmine Fimiani, Antonio Regano
Directorate for National Accounts, Istat, Rome, Italy.

1. Introduction

The Growth and Productivity Account (GPA) project was launched by Eurostat at the end of 2018 under the priorities of the European Statistical Programme with the aim to develop high-quality indicators for labour and capital productivity for all EU Member States. ISTAT joined the GPA project and applied for financial assistance in this area at the end of 2019 with the aim to improve the quality and the coverage of the productivity measures already provided and to develop new indicators.

The empirical final results concerning capital productivity indicators (CAPI) presented in this short paper were part of a larger Eurostat-funded research activity[1] carried out in ISTAT National Accounts Directorate in the area of productivity indicators. One of the several goals of this activity was the extension of the CAPI computation, currently available at A*38 industry breakdown, to A*64 industry breakdown in order to provide a database able to be used for analysing capital productivity by industry at a more detailed level. The coverage at A*64 industry level required additional methodological efforts because historical time series on gross fixed capital formation (GFCF) are not available. In fact, the method most commonly used in national statistical offices is the perpetual inventory method (PIM) with linear depreciation function which, however, requires the availability of particularly long time series of investments. When this is cannot be the case, the literature suggests adopting a geometric model for estimating the capital stock.

Therefore, the investigation on broadening CAPI data disaggregation by economic activity gave us the opportunity to test how much capital input, CAPI and total factor productivity (TFP) can be affected by the choice of a different method of capital stock estimation. The paper is structured as follows. In Section 2 a short overview of the geometric model *vs.* the standard linear depreciation function is provided. Section 3 and 4 highlight the empirical final results mainly in terms of capital input, capital productivity and TFP growth rate patterns between the two methodological approaches along the all period 1995-2020. Section 5 concludes.

It is worth noting that the capital stock and productivity indicators, in the framework of our study and, more generally, according to Istat release on productivity statistics, are computed with the exclusion of the industries belonging to NACE Rev2 sections L, O, P (only for non-market services), Q (only for non-market services), T and U.

2. The computation of capital stock in the capital productivity formula: two alternative methodological approaches

Capital productivity can be defined as the ratio of the volume index of an output measure

[1] Eurostat Grant agreement n. 887026 – AMD-887026-2 - 2019/IT/GPA, "Development and improvement of productivity indicators: coverage and practice in the national statistical office of Italy", Coordinator: Carmine Fimiani, Beneficiary: Istituto Nazionale di Statistica (ISTAT), Italy.

(value added in our case) to the volume index of capital input, i.e. as value added per unit of capital input. Several approaches can be followed to estimate the denominator of the capital productivity formula. In the Italian national accounts capital input is measured by the flow of productive services provided by the existing stock of capital goods (referred to as capital services) and summarised in a volume index of service flows generated by each type of capital goods. For each industry, the capital volume index is obtained by aggregating the rates of change of the stock of productive capital of 14 types of assets through the Tornqvist formula, with weights given by the share of each type of capital good in the total cost for capital services of the industry.

Various steps, with necessary assumptions, must be followed to calculate the flows of capital services to be used in the capital productivity formula. The starting point is represented by the calculation of gross capital stock (the cumulative flow of volume investments, corrected for retirement) and net capital stock (the stock of assets still in use and adjusted for depreciation). The approach currently used for the capital stock measurement by asset is the PIM. In the PIM[2], gross capital stock is defined as the cumulative flow of past investments, corrected for retirement. In the gross stock, assets are treated as new until they are retired and it is assumed that they retain their full productive capacity until removed from the stock. The PIM adopted by Italian National accounts assumes constant average service life, a normal bell-shaped retirement pattern function[3] and a straight line model of depreciation[4] for almost all capital goods except R&D and Weapons systems. The GFCF series by industry are available since 1870 for Construction, (Dwellings and Buildings other than dwellings and Other structures) and for the other asset categories since 1951, except for R&D and Weapons systems (available since 1995).

The next stage is represented by computing, for each type of asset, the productive capital stock defined as the sum of investments made in the current and previous years, weighted with weights reflecting both the progressive loss of efficiency[5] that the asset suffers over time due to wear and tear, and the process of withdrawal of capital goods from the production process. It therefore measures the residual capacity to provide a flow of productive services from capital goods purchased in the past and not yet withdrawn from the production process.

Finally, capital services are computed. They refer to the user cost weighted aggregate of the asset types in the productive stocks. The user cost ($\mathbf{u_{it}}$) measures the price of productive services generated by a capital good. It is assumed to be the sum of the nominal rate of return ($\mathbf{r_t}$), the rate of economic depreciation the capital good suffers over its average service life ($\mathbf{d_{it}}$) and the holding gains or losses due to increases/decreases in the purchase price of a capital good ($\mathbf{p_{it}}$). The nominal rate of return is calculated on the assumption that capital income (gross operating surplus plus the mixed income component that is considered remuneration of capital) should equal the total cost of capital services[6].

As mentioned before, historical time series of GFCF are not currently available at A*64

[2] Usually, the PIM is used to estimate gross capital stocks; a retirement function and a depreciation function are used to calculate Consumption of Fixed Capital (CFC). Net capital stocks is then obtained by subtracting the accumulated CFC from gross measures.

[3] The retirements are distributed around the average service life (constant over time) according to a truncated normal distribution (with truncation limits set at +/- 35% of average service life and the variance set so that 90% of retirements occur between +/- 25% of the average service life). A truncated normal distribution is chosen because it is excluded that an asset, once entered, is never retired.

[4] The straight line (linear) model of depreciation means that the market value of an asset is assumed to decline by the same amount each period. In other words a homogeneous asset with a service life of T years loses a constant proportion (1/T) of the initial asset value each period, until the asset's value becomes zero at the end of year T.

[5] For all types of capital goods, an efficiency loss profile concave to the origin is selected, i.e. production capacity is assumed to decline at an increasing rate with the age of the good.

[6] The nominal rate of return (known in the literature as the endogenous rate) is given by the value of $\mathbf{r_t}$ for which the user cost relationship ($\mathbf{u_{1t}}$) is satisfied. The other components of $\mathbf{u_{1t}}$ are specific to each type of asset. The purchase price of the new capital good is measured by the corresponding implicit national accounts investment deflator. $\mathbf{d_{1t}}$ is obtained as the ratio of consumption of fixed capital to the net capital stock (both measured at chain-linked values). The term expressing holding gains/losses ($\mathbf{p_{1t}}$) is thought of as a measure of expected inflation. This has been approximated as a moving average of the rate of change of the implicit deflator over the previous three years.

industry breakdown. Therefore, the approach followed to overcome this lack of information starts from the geometric perpetual inventory equation, that can be used to calculate the capital stock (OECD, 2009; Griliches, 1980). According to the geometric model, net capital stock and consumption of fixed capital are obtained by means of the formula: $\mathbf{K_t = K_{t-1}\ (1-\ \delta) + I_t}$ where K_t and K_{t-1} are, respectively, the net stock at the end of period t and t-1, I_t is GFCF (at chain-linked values) at year t and δ represents the depreciation rate. Consumption of fixed capital in period t is simply obtained multiplying the depreciation rate δ for the capital stock at the end of period t-1.

The main computational problem with the equation above is determining the initial capital stock, considering that the series of GFCF are available only since 1995. The initial capital stock have to be calculated for the year 1994, by means of the following formula (OECD, 2009): $\mathbf{K_{t-1} = I_t/(\ \delta + \theta)}$, where I_t is GFCF at period t, δ represents the depreciation rate and θ is the long term-growth rate of GFCF. The parameter θ is estimated as the transform of the slope coefficient (m) of a linear regression of ln (I_t) on calendar time t, following the initial proposal made by Griliches (1980) who showed that under reasonable conditions the unobserved growth rate of the net stock is well approximated by the growth rate of investment. In our analysis, the parameter θ has been necessarily calculated on the shorter GFCF series available at A*64 industry level over a number of years starting from 1995 onwards for all assets. The depreciation rate δ, is estimated by using information about average service life of assets through the "declining balance method" formula in which the declining-balance rate (DBR), in the absence of any empirical evidence on the measure of this parameter, is set equal to 2 for all assets (OECD, 2009).When assuming geometric profiles for computing capital stock, the productive and the net stock are identical and the entire process of computation appear to be easier and faster compared to the PIM with linear depreciation function. The steps to compute the flow of capital services are identical to the ones described before.

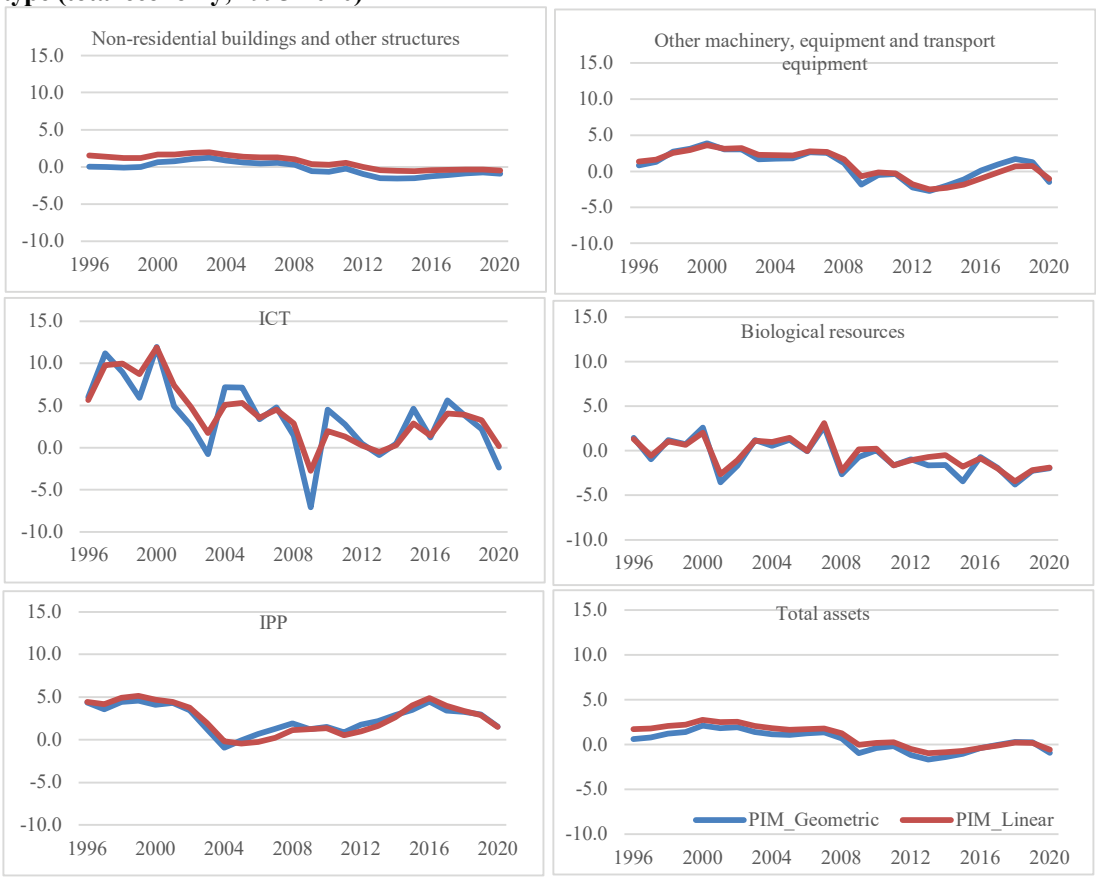
3. Impact on capital input (productive capital stock) estimates

It is interesting to look at the implications of applying the PIM with geometrical depreciation function and the PIM with linear depreciation function and truncated normal distribution of retirements, first of all, on the estimation of productive capital stock the latter being the method currently adopted in Istat productivity statistics computation.

The results are presented in terms of growth rates patterns of total productive capital stocks time series computed with the two methods for total economy throughout the period 1995-2020 as depicted in Figure 1. Growth rates refer to stocks in chain-linked volumes (with 2015 as reference year) where chain-linked values come from stocks in current and previous-year replacement costs. Figure 1 shows trends in the rates of growth of the productive capital stock by asset type referred to total economy and for total assets in the last graph of Figure 1. In general, at total capital input level (total assets), it can be seen how growth paths tend to converge as we approach 2020. This is also partly true for the assets of construction, IPP and biological resources. These findings are more evident by looking at the growth rate differences, that is to say in terms of asset time series revisions. The new method particularly affects the asset of construction with longer average service life compared to the other capital goods, but the ICT component is not exempt from this as well. This is especially evident for those years in which the estimated initial capital stock (1994) are nearest to the current time series available (from 1995 up to 2020).

Specifically, the differences in the growth rates for the asset 'Non-residential buildings and other structures' tend to diverge more from 1996 to 2000 compared to the remaining part of the period (the differences between the linear and geometric methods vary between 1.48 and 1 percentage point, while those for the asset 'ICT' tend to diverge on average up to 1.50 percentage point over the entire period analysed). Less marked appear the differences for the remaining assets although they do exist.

Figure 1: Pattern of productive capital stock growth rates based on the two methodologies by asset type (total economy; 1995-2020)



4. Impact on capital productivity measures

In the Italian National accounts input data for the production of CAPI at A*64 industry breakdown, specifically gross value added and capital input, are available beyond A*38 only for value added. As a consequence, it is possible to extend the calculation of CAPI at this more detailed level of disaggregation by means of the methodology for calculating capital stock described in Section 2.

It may be interesting to look at the impact of the two methods of estimating capital input (geometric depreciation function versus linear depreciation function) on productivity measures, in particular capital productivity and total factor productivity. Table 1 shows average annual growth rates for the entire period 1995-2020, for some sub-periods identified as boom and bust cycles (2003-2009, 2009-2014, 2014-2020) and annual growth rate for 2019 and 2020 at the total economy level for capital input (in terms of capital services), capital productivity and total factor productivity.

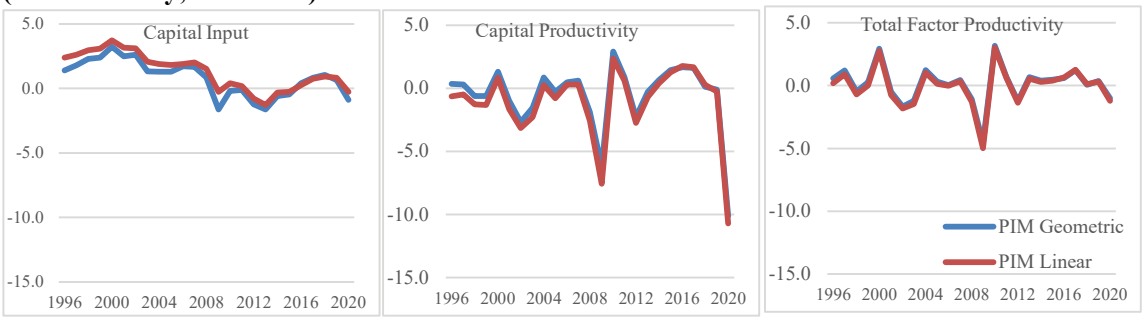
Table 1: Growth rates of capital input, CAPI and TFP with PIM (geometric depreciation function) and PIM (linear depreciation function) – Total economy; 1995-2020.

	Capital Input		Capital Productivity		Total Factor Productivity	
	PIM Geometric	PIM Linear	PIM Geometric	PIM Linear	PIM Geometric	PIM Linear
1995-2020	0.8	1.3	-0.6	-1.0	0.1	0.0
2003-2009	0.9	1.5	-1.1	-1.7	-0.6	-0.8
2009-2014	-0.8	-0.4	0.4	0.0	0.7	0.6
2014-2020	0.3	0.4	-0.9	-1.0	0.3	0.3
2019	0.7	0.8	-0.1	-0.3	0.4	0.3
2020	-0.9	-0.3	-10.1	-10.7	-1.0	-1.2

Source: elaboration on Istat data

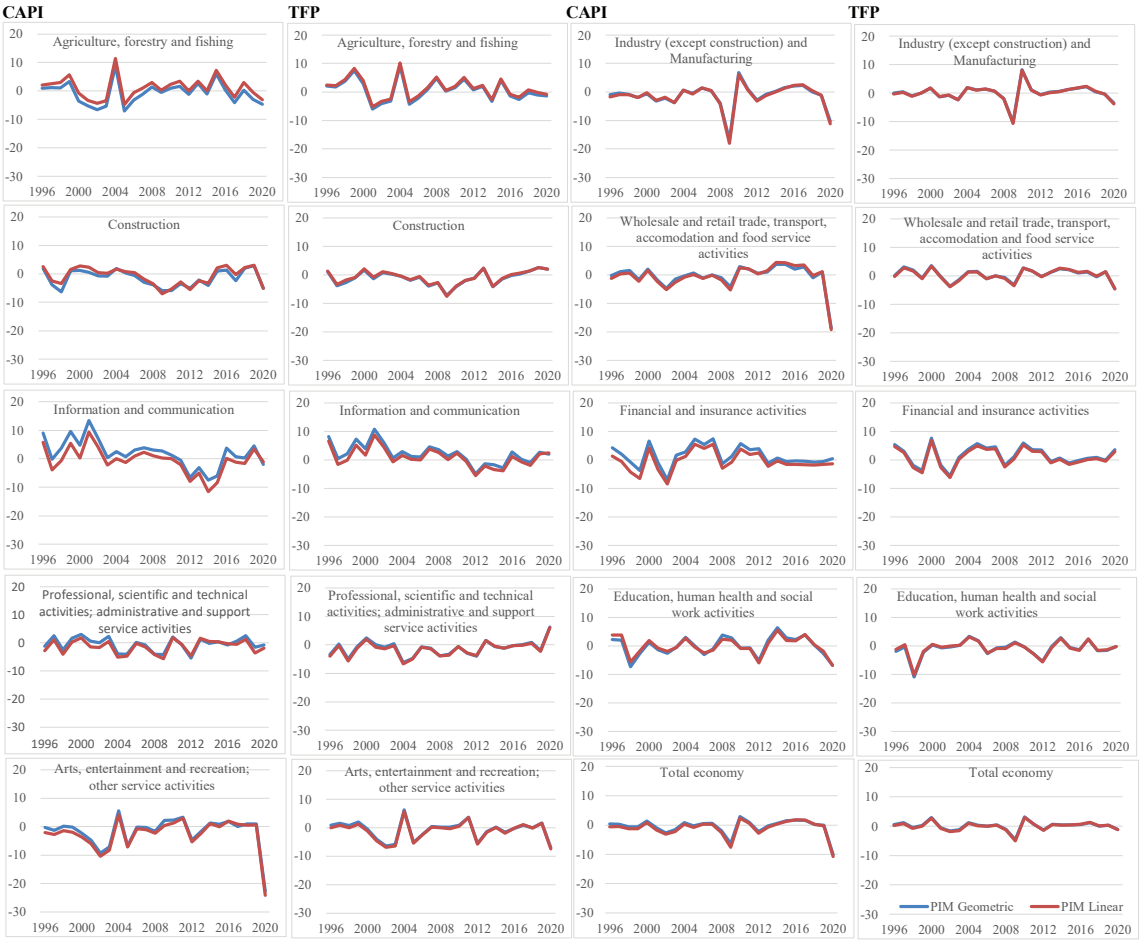
In general, it can be noted that capital input growth rates deriving from computation of the capital stock with the geometric model are lower if positive, higher if negative compared to those estimated with the current method (linear model of depreciation). The opposite occurs in the case of capital productivity growth rates. Certainly not negligible appear to be the differences in growth rates relative to the capital services and capital productivity estimated by the two methods. In contrast less remarkable is the impact on total factor productivity growth rates compared to capital services and CAPI ones as shown by TFP revisions.

Figure 2: Pattern of capital input, CAPI and TFP growth rates based on the two methodologies (Total economy; 1995-2020)



Source: elaboration on Istat data

Figure 3: Pattern of CAPI and TFP growth rates based on the two methodologies by A*10 industry 1995-2020



Source: elaboration on Istat data

Figure 2 shows trends in the rates of growth of capital input and productivity measures (CAPI and TFP) referred to total economy. From the graphical representation, it is possible to note how growth paths of capital input and capital productivity tend to diverge from 1995 to 2009 and to converge as we approach 2020. In order to make the comparison at industry level possible, the productivity measures (CAPI and TFP) computed at A*64 industry breakdown were aggregated at A*10 industries.

The results of our simulation are presented at a level of detail of 10 industries in Figure 3, that shows the outcomes in terms of CAPI and TFP growth rate patterns. The impact on TFP growth rates is lighter than for capital productivity as shown by graphical representation. With reference to the impact on CAPI, Figure 3 show a wide variability across industries, in particular for 'Information and communication', 'Financial and insurance activities' and 'Arts, entertainment and recreation; other service activities', where capital productivity growth rates deriving from computation of the capital stock with the geometric model are higher compared to those estimated with the linear model of depreciation. While for 'Agriculture, forestry and fishing' and 'Construction' capital productivity growth rates estimated with the geometric model are lower compared to the growth rates calculated with the linear model.

5. Conclusions

The results from the simulation carried out at A*64 industry breakdown showed that, in the absence of particularly long time series of investment, when the initial stock is estimated for a year closer (1994) to the current date (1995 up to 2020), the estimates of capital input obtained with the two depreciation methods (geometric *vs*. linear) are quite different both in level and in terms of growth rates as shown in our analysis.

These differences in capital input estimates inevitably have an impact on the calculation of productivity indicators mainly if we look at the results at industry level. It is widely recognized that estimates obtained by the PIM are robust if sufficiently long time series of GFCF data are available for each asset. This is true both for linear depreciation functions combined to a bell-shaped retirement functions and for geometric depreciation patterns. As a consequence, the lack of historical GFCF time series at a greater industry detail could imply lower robustness of the estimates unless efforts are made to refine the process of estimating the parameters (mainly declining balance rate and long-term investment growth rate) peculiar to the geometric model.

References

Eurostat (2013). *European System of Accounts (ESA2010)*. Publication Office of the European Union, Luxembourg, (LU).

Griliches, Z. (1980). Returns to research and development expenditures in the private sector. in *New Developments in Productivity Measurement*, eds. J.W. Kendrick, B.N. Vaccara, *Studies in Income and Wealth*, Vol. 44, The National Bureau of Economic Research, University of Chicago Press, Chicago, Ill., pp. 419-462.

Italian National Institute of Statistics, Directorate for National Accounts (2021). *Italy Gross National Income Inventory 2019 (ESA 2010)*. Istat, Rome, (IT).

OECD (2001). *Measuring Productivity: OECD Manual*. OECD, Paris, (FR).

OECD (2009). *Measuring Capital Productivity*. OECD, Paris, (FR).

Stehrer, R. (2020). Choice of capital productivity indicators (CAPI) and multi-factor (MFP) productivity indicators according to data availability and policy needs, in *Deliverable D1.1 – Results of Sub-Task 1.1*, The Vienna Institute for International Economic Studies (WIIW), Wien (AT).

United Nations, European Commission, International Monetary Fund, OECD, World Bank (2009). *System of National Accounts 2008*. United Nations, New York, (USA).

doi.org/10.26398/asaproc.0015

Compiling sub-national spatial price indices using scanner data and traditional data sources

Alessandro Brunetti[a], Anna Di Franco[a], Barbara Dramis[a], Tiziana Laureti[b], Federico Polidoro[c], Agostina Zanoli[a]

[a] National Institute of Statistics (ISTAT), Rome, Italy

[b] Department of Economics, Engineering, Society and Business Organization, University of Tuscia, Viterbo, Italy

[c] World Bank, Rome, Italy

1. Introduction

Sub-national spatial price indices (SN-SPIs) or regional purchasing power parities (RPPPs) measure the differences in price levels across regions within a country at a given point of time. They express how many currency units a given quantity of goods or services costs in different areas. SN-SPIs are essential for comparing real income, standards of living and consumer expenditure patterns. In countries characterized by large territorial differences in consumer preferences as well as in household characteristics, the calculation of SN-SPIs acquires considerable importance.

The Italian National Institute of Statistics (Istat) is one of the few national statistical offices that carried out official experimental sub-national spatial price indices computations. The analyses, based on price data coming from consumer price indices (CPIs) and *ad hoc* surveys, focused on the comparison of consumer prices across the 20 Italian regions (Istat, 2010). The availability of scanner data, used since 2018 in the CPIs production process, provided new impulse to the research (Laureti, Ferrante, Dramis, 2017; Laureti, Polidoro, 2022).

Moreover, the increasing request for sub-national indicators both at the national and European levels has pushed Istat to develop a project to estimate SPI at the regional level on a regular basis. Due to the wide information amount necessary, a multi-source approach has been used to select for each product the best available information. Data sources used in the analyses are scanner data, traditional CPI data and *ad hoc* surveys.

The methods adopted to properly consider the information provided by each source at the basic heading (BH) level are weighted and unweighted regional product dummy (RPD) methods. BH-SPIs are obtained and aggregated at higher levels of classification by Gini-Elteto-Koves-Szulc (GEKS) technique based on Fisher indexes (in accordance with international practices).

Experimental results (Istat, 2023), referred to the first three expenditure divisions of the European classification of individual consumption according to purpose (ECOICOP) (food, including tobacco, beverages, clothing and footwear), are presented and discussed in the rest of this paper: data sources are illustrated in Section 2; the methodological aspects are addressed in Sections 3; Section 4 provides a picture of the results. Finally, in Section 5 some concluding remarks are sketched.

2. Data sources

SN-SPIs computation needs many detailed data at a local level that satisfies two main requirements: comparability and representativeness. The first one entails that items being compared should be identical or at least equivalent, to ensure that differences in prices are not influenced by differences in quality. As for the second requirement, products that enter into the calculation of the SN-SPIs should be those that are frequently purchased and widely available at the regional level.

To this aim, a multi-source approach has been used, coherently with the approach adopted for the compilation of the Italian CPIs.

The first source consists of scanner data coming from a sample of outlets of twenty-one retail trade chains [1] covering 60% of the total turnover for the hard-discount channel and over 90% for each other distribution shop types at national level. Scanner data are referred to processed food, including non-alcoholic and alcoholic beverages.

The outlets sample[2], widespread in all the municipalities within the 107 Italian provinces, is the same sample extracted for the inflation estimates. The sampling plan is probabilistic, stratified by province and store type, with the probability of selection of the outlets proportional to their turnover.

The detailed information contained in these data (turnover and quantities sold for each global trade item number-GTIN in each outlet) can be used to compile weighted SPIs at a highly detailed territorial level and allows to fulfil both the requirements for spatial comparison of prices: comparability and representativeness.

Data for turnover and quantities sold in the first three weeks of each month (weekly data) are used to calculate the annual unit value price for each GTIN, at the outlet level, as the ratio between annual turnover and annual quantities sold (in the outlet). These annual unit values are then aggregated across outlets of the same province to obtain the average price for GTIN at the provincial level. For the aggregation, a weighted average is used, with outlet and outlet-type weights. Only GTINs sold in at least two regions are considered: 140,197 for Food and non-alcoholic beverages and 25,121 for Alcoholic beverages.

In addition, CPI data coming from the traditional price collection has been used for fresh products such as fish, fruits and vegetables for which CPI definition is enough accurate to guarantee comparability. Moreover, products included in CPI basket are widespread in almost all 80 provinces. The data have been cleaned for outliers and only the prices of the months in which these products are "in season" have been included. The annual provincial average prices are the unweighted mean by product at the provincial level of all the prices since the weights relating to the quantities sold are not available. Therefore, the annual price quotations included in the analysis were 87,784. The regional average price per product was then calculated as a weighted mean of average provincial prices, using population as weights, in line with what occurs in inflation calculating.

Thirdly, *ad hoc* surveys have been carried out by product category for which the use of other sources is prevented by comparability issues. A specific basket of comparable products is defined as subset of the one used for International Comparison Program (ICP) at European level[3].

The twenty-one municipalities involved in the survey are the 19 regional and the two autonomous provincial capitals (with the exception of L'Aquila, replaced by Pescara). The outlet sampling design in each municipality is the one defined for the CPI survey. The *ad hoc* price surveys are based on a cyclical system of surveys coinciding with those of international comparison program for purchasing power parity (PPP) computations. Overall, there are six surveys, each on a specific basket of goods and services, which take place over a three-year cycle.

The basket of food products included in the index calculation is made up of 39 items, while 79 items are considered for clothing and footwear. The number of quotations entered in the indices calculation for food products was 8,945, while for clothing and footwear 7,885. The trimming of invalid quotations had an impact above all on clothing and footwear (16% vs. 5.5% in food division).

In general, *ad hoc* surveys involve a considerable effort by municipalities in terms of organization, time and human resources. The number of collected and validated quotations for the municipalities of Napoli, Pescara and Potenza, in clothing and footwear, is small for several products. Therefore, the results for the Campania, Abruzzo and Basilicata regions have to be read taking into consideration this limit and require significant further study. Despite these difficulties, the BHs with the greatest weight are overall well represented.

Given that the prices collected through *ad hoc* surveys refer to the two months in which each

[1] Scanner data are provided by Nielsen to Istat, thanks to an agreement with the retail trade chains.

[2] The sample of outlets includes about 4,000 shops of different type (supermarkets, hypermarkets, minimarkets and discounts).

[3] Eurostat is a partner in the ICP. EUROSTAT-OECD Methodological manual on purchasing power parities (PPPs)

survey was conducted, the monthly data were made representative of the year, by considering the inflation recorded in the Municipality in the other months of the year. Temporal adjustment factors (TAFs) were calculated based on harmonized index of consumer price (HICP) data.

Finally, for tobacco products, the administrative source has been used (Customs and Monopolies Agency).

The first three ECOICOP expenditure divisions represent about 32% of household expenditure based on the 2021 HICP basket. The first division (20.5%) is affected by the all three data sources, even if with different importance: 53% scanner data, 27% *ad hoc* survey and 20% CPI traditional price collection. The second division represents about 3.8% of Italian consumption and the drinks are entirely covered by scanner data. The third division data (approximately 7.5% of the basket) comes entirely from *ad hoc* surveys. The tobacco products whose prices have no variability across the Italian territory (but have different weights in the regions) have been included in the calculation with parity equal to 100.

3. Methodological approach

PPP indicators are calculated using recognized methodological tools and used in numerous international studies and empirical experiences conducted in various countries. The ICP published a guideline in 2021 (ICP, 2021), to which Italy contributed, for the calculation of sub-national PPP at different territorial or geographical area (frequently regions within a country). In accordance with these guidelines, the methods used are:

- RPD model at BH level
- GEKS (Gini - Èltetö-Köves-Szulc) method for aggregations above BH level.

3.1 Regional Product Dummy

The RPD method is the regional version of the country-product-dummy (CPD) method used in international comparisons. The idea behind this method (Rao and Hajargasht, 2016; World Bank, 2013) is that the price p_{nr} of an item n (n=1…N) belonging to the BH in a region r (r=1...R) is function of a specific regional factor $RPPP_r$ (parity or general price level of the area considered with respect to other areas), of the average price of the *n*-th commodity/item P_n and of a random error u_{nr}:

$$p_{nr} = P_n * RPPP_r * u_{nr} \quad (1)$$

Considering the logarithms, equation (1) can be written as:

$$ln\, p_{nr} = \sum_{r=1}^{R} a_r D_r + \sum_{n=1}^{N} b_n D_n^* + v_{nr} \quad (2)$$

where: D_r is the dummy variable that takes value equal to 1 if the price quotation is from region *r* and *0* otherwise, D_n^* is the dummy variable for commodity n which takes value equal to 1 when item considered is *n* and *0* otherwise, a_re b_n are the differences in the effects associated with the regions and the commodity type, respectively, v_{nr} are random error normally distributed with a zero mean and variance σ^2.

Parameters of this model can be estimated using ordinary least squares, imposing a restriction that a coefficient corresponding to a specific area or region is set equal to zero (a_i=0) or equivalently PPP_i =1 thus considering it as reference area to which the coefficient estimates are referred.

The SN-PPP between an area r and the reference area is $\widehat{RPPP}_r$=exp($\widehat{a_r}$). The parities thus estimated satisfy the property of transitivity and invariance of the basis.

Having weights in terms of value or quantity for each commodity, the weighted RPD model can be written as:

$$\sqrt{w_{nr}}\, ln\, p_{nr} = \sum_{r=1}^{R} a_r \sqrt{w_{nr}} D_r + \sum_{n=1}^{N} b_n \sqrt{w_{nr}} D_n^* + \sqrt{w_{nr}} v_{nr} \quad (3)$$

where w_{nr} are the weights in terms of value or quantity share that reflect the economic importance of the different commodities consumed in the area.

The model was estimated for each BH, making the most of the information available depending on the data source used. Within each BH, only one of the three considered sources was used.

In the case of scanner data, they provided information on turnover and quantities sold for each GTIN, in all 107 Italian provinces. The revenue-weighted model was then used to obtain the estimates. Furthermore, a two-step procedure was adopted (Laureti, Polidoro, 2022) for each BH:

- Step 1: In each region, a model (equation 3) was estimated to obtain provincial PPP. These indicators have been used to 'deflate' the initial prices and the turnover of each GTIN within each region.
- Step 2: Model (equation 3) was estimated to obtain RPPPs, using the 'deflated' prices and turnover obtained in step1.

In the case of the data from the local survey of consumer prices and the *ad hoc* surveys, weighted model cannot be used (the quantities sold are not available) and the unweighted RPD (equation 2) has been used to estimate RPPPs at the BH level.

3.2 GEKS (Gini - Èltetö-Köves-Szulc)

Parities at the upper level of the BH are calculated using the GEKS method aggregating RPPP estimated at the BH level. The weights, based on household expenditure, are the same used in HICP. The index compiled by the GEKS method satisfies the properties of transitivity and invariance of the basis. Furthermore, it is as close as possible to the corresponding binary indices (Diewert, 2013). Aggregate RPPPs are obtained as follows:

$$RPPP_{jk}^{GEKS} = \prod_{l=1}^{R} (F_{jl} \cdot F_{lk})^{1/R}$$

The parity for region *k* with reference to region *j* chosen as the base is given by the geometric mean of the Fisher indices of all direct comparisons between region *j* and region *k*, and indirect across all possible links between the *R* regions ($l, k, j \in R$). Fisher type (F_{jl}) indices are obtained as the geometric mean of the corresponding Laspeyres type and Paasche type indices, calculated based on the parities for BH weighted respectively by the expenses of the base region (Laspeyres) and of the partner region (Paasche).

These first results lack data on some BHs, because of difficulties in data collection. The weights of these BHs were distributed among the BHs of the same consumption segment or at a higher level if it was the only BH of the segment.

To express the parities referring to the national average, each RPPP is divided by the geometric mean of the price level indices of the participating regions, and conventionally, multiplied by 100.

4. Results

The analyses show significant differences in consumer price levels between the Italian regions, in 2021. In general, considering the aggregation of the first three ECOICOP expenditure divisions, the prices recorded in the northern regions are higher than those of the central ones, except for Tuscany, and of the south and islands, except for Sardinia (Table 1).

The most expensive regions are Alto Adige with prices 5.3 % higher than the national average (Italy=100), Lombardy (+ 5%) and Liguria (+ 4.7%). The least expensive regions, compared to the national average, are Campania (prices 9.5% lower than the average), Abruzzo (-6.2 %) and Basilicata (- 5.2%). Thus, the difference in price levels between Campania (least expensive) and Alto Adige (most expensive) is almost 15 percentage points. The estimated RPPPs for Food Products show that the regions with price levels above the national average are both in Northern Italy and in the Centre, while those with price levels systematically lower than the national average are in the South (except for, also in this case, Sardinia). Lombardy and Alto Adige have price levels

above the Italian average by more than 6 %, while Campania and Basilicata have price levels below the average by 11% and 9 % respectively. In this case, the difference in price levels between Campania (least expensive) and Lombardy (most expensive) is almost 17 percentage points.

Soft drinks and Alcoholic beverages show the least heterogeneity in consumer price levels between regions and low differences in price levels compared to the Italian average. There is not territorial trend that leads to characterize groups of regions as in the other product categories. It should be noted, in fact, that in both cases a northern region is the least expensive (Veneto for soft drinks, Lombardy for alcoholic ones). This seems to highlight how for products with a longer supply chain, greater efficiency in logistics and infrastructure tends to change the traditional comparative geography of price levels between Italian regions.

Table 1. Consumer spatial price indices by ECOICOP divisions, groups of product and regions, year 2021- Italy=100

Region	Food and non-alcoholic beverages			Alcoholic beverages and tobacco			Clothing and footwear			First three divisions
	Food	Non-alcoholic beverages	All items	Alcoholic beverages	Tobacco	All items	Clothing	Footware	All items	
Piedmont	103.27	100.06	103.00	98.04	100.00	99.29	99.10	95.23	98.32	**101.49**
Valle d'Aosta	102.66	103.23	102.70	101.90	100.00	100.71	103.00	97.10	101.82	**102.29**
Lombardy	106.59	98.50	105.90	96.71	100.00	98.79	106.08	104.92	105.82	**105.01**
Trentino	100.96	99.63	100.85	99.18	100.00	99.70	101.31	117.89	104.40	**101.57**
Alto Adige	106.08	102.14	105.77	102.27	100.00	100.87	107.05	103.81	106.37	**105.33**
Veneto	103.86	95.20	103.16	98.03	100.00	99.26	101.42	117.93	104.50	**103.00**
Friuli-Venezia Giulia	103.91	97.64	103.40	100.06	100.00	100.04	99.24	113.76	101.94	**102.67**
Liguria	104.47	102.21	104.28	101.25	100.00	100.47	111.55	98.59	108.92	**104.77**
Emilia-Romagna	104.78	98.39	104.27	98.35	100.00	99.38	104.48	106.56	104.87	**103.80**
Tuscany	103.94	97.45	103.40	97.91	100.00	99.26	102.97	95.24	101.38	**102.46**
Umbria	100.65	98.28	100.45	99.13	100.00	99.71	96.37	84.14	93.81	**98.85**
Marche	103.07	99.96	102.81	100.70	100.00	100.27	88.64	88.35	88.55	**98.95**
Lazio	101.12	102.72	101.25	100.88	100.00	100.33	93.84	97.48	94.53	**99.59**
Abruzzo (*)	94.54	101.63	95.08	101.62	100.00	100.53	87.25	85.19	86.77	**93.75**
Molise	94.94	99.42	95.28	100.34	100.00	100.12	93.40	101.86	95.21	**95.80**
Campania (*)	88.99	98.22	89.69	99.79	100.00	99.95	84.28	100.41	87.61	**90.47**
Puglia	93.31	98.66	93.72	99.21	100.00	99.76	109.88	103.51	108.41	**97.52**
Basilicata (*)	91.03	99.99	91.71	100.49	100.00	100.18	101.44	99.28	100.93	**94.72**
Calabria	94.42	100.66	94.90	100.38	100.00	100.14	107.19	108.26	107.42	**98.25**
Sicily	97.54	101.14	97.85	101.55	100.00	100.50	103.54	98.55	102.47	**99.16**
Sardinia	102.67	105.41	102.91	102.45	100.00	100.77	103.52	90.62	100.66	**102.15**
Min	*88.99*	*95.20*	*89.69*	*96.71*	*100.00*	*98.79*	*84.28*	*84.14*	*86.77*	*90.47*
Max	*106.59*	*105.41*	*105.90*	*102.45*	*100.00*	*100.87*	*111.55*	*117.93*	*108.92*	*105.33*
Coefficient of Variation	*5.09*	*2.28*	*4.70*	*1.56*	*0.00*	*0.56*	*7.12*	*9.10*	*6.58*	*3.87*

(*) Due to the low number of observations for clothing and footwear division data of Basilicata, Campania and Abruzzo regions are not reliable.
Source: Istat

For Clothing and footwear, five southern regions are more expensive than the national average, almost in line with the northern ones. Instead, the central regions show price levels on average lower than the national average (except for Tuscany, which ranks slightly above). There are more than 20 percentage points of difference between the least expensive region (Abruzzo) and the most expensive one (Liguria). The most expensive regions are Puglia and Liguria, with values more than 8 points above the average, while Campania and Abruzzo are the least expensive. Likely, this more accentuated dispersion of the results, related to clothing and footwear, is due to the greater volatility present in the elementary data, attributable to the limited number of price observations recorded

(especially in Campania, Basilicata and Abruzzo) and deserving of further insights.

4. Concluding remarks

Our multi-source approach provides the opportunity to achieve the important objective of compiling sub-national spatial price indices for the first three expenditure divisions of the ECOICOP classification.

The preliminary results provide very interesting insights in the comparison of prices between the Italian regions (it is not obvious that are the southern ones those with the lower prices).

The project uses mainly data sources already acquired by Istat for the inflation estimates (scanner data and traditional price collection), thus optimizing the available resources, in terms of data, software used, personnel skills. Specifically, scanner data bring such an important value added, given the availability of information on quantities sold that allow a more precise consideration of the requirement of representativeness.

The data source with major issues is the *ad hoc* survey, which has a low territorial coverage and the least number of quotations used in the process. In several cases, it was not possible to collect a number of price quotations sufficient to guarantee robust results. Possible solution to this problem should be to extend the data collection to other municipalities and/or alternative sources of data.

The results will be updated and progressively extended to other divisions. The next release will cover ECOICOP division 04 (Housing, water, electricity, gas and other fuels) and 11 (Restaurant and accommodation services). SN-SPIs or RPPPs produced on a regular basis will improve our knowledge of the real economic territorial differentiation in Italy and in particular, the measurement of relative poverty by taking into account the price dimension, which is now not considered, given that a uniform threshold at the national level has been adopted.

References

Diewert, W.E. (2013). Methods of aggregation above the basic heading level within regions, in *Measuring the Real Size of the World Economy,* World Bank, Washington D.C., pp. 121-167

ICP (2021). *A Guide to the Compilation of Subnational Purchasing Power Parities (PPPs).* World Bank.

Istat (2010). La differenza nel livello dei prezzi al consumo tra i capoluoghi delle regioni italiane. *Comunicato Stampa.* Roma, Italy.

Istat (2023). Indici spaziali dei prezzi al consumo – anno 2021, *Statistica Sperimentale*, 3 agosto 2023.

Laureti, T., Ferrante, C., Dramis, B. (2017). Using scanner and CPI data to estimate Italian sub-national PPPs, in *Proceedings of 49th Scientific Meeting of the Italian Statistical Society – SIS*, Palermo, Italy, pp. 581–588.

Laureti, T., Polidoro, F. (2022). Using scanner data for computing consumer spatial price indexes at regional level: An empirical application for grocery products in Italy. *Journal of Official Statistics,* **38**(1), pp. 23-56.

Rao, D.S.P, Hajargasht, G. (2016). Stochastic approach to computation of purchasing power parities in the International Comparison Program (ICP). *Journal of Econometrics,* **191**(2), pp. 414-425.

World Bank (2013). *Measuring the Real Size of the World Economy: The Framework, Methodology, and Results of the International Comparison Program (ICP).* World Bank.

doi.org/10.26398/asaproc.0016

The role of STEAM disciplines and metaverse for the targets of the UN 2030 Agenda

Marco Vito Calciano[a], Alessandro Verderame[b], Diego Carmine Sinitò[b], Antonio Ruoto[b], Federico Usai, Fabio Stasi[b], Davide Scintu[b], Vito Santarcangelo[b], Marialuisa Sabino[c]

[a] Zio Startup Srl, Matera, Italy.

[b] iInformatica Srl, Matera, Italy.

[c] Liceo Scientifico Statale "Dante Alighieri", Matera, Italy.

1. Introduction

STEAM (Science, Technology, Engineering, Art and Mathematics) disciplines play a pivotal role in schools in a world increasingly driven by innovation and technology as, through their interdisciplinary approach to learning, they encourage students to develop a wide range of competences and skills that are essential to meet the challenges of the 21st century by preparing them to become knowledgeable and critical citizens, equipped with the practical skills needed to meet the demands of Industry 4.0. Introducing students to these disciplines from an early age not only prepares them for their future employment, but also equips them with useful tools to tackle global challenges in a creative and informed manner.

In this research paper we will present the role and potential of STEAM disciplines and the metaverse within schools to contribute to the goals of the UN 2030 Agenda and their importance in the world of industry by finally presenting the Robocom kit, a tailor-made high-tech solution for schools developed by the innovative SME iInformatica Srl that represents a new way of enhancing the educational experience through the combination of innovation, empathy and inclusion in line with the goals of the UN 2030 Agenda and the Faro Convention.

2. STEAM disciplines and their importance in the world of industry

The STEAM disciplines represent an interdisciplinary educational approach that integrates Science, Technology, Engineering, Art and Mathematics. The acronym STEAM is a variation of the acronym STEM (Science, Technology, Engineering and Mathematics), to which art is added to emphasise the importance of creativity and artistic expression in a learning context based on science and technology. The STEAM approach seeks to overcome traditional disciplinary barriers and promote collaboration between different areas of knowledge, encouraging students to develop a more comprehensive understanding of the world around them, to creatively apply their knowledge and to address contemporary challenges in innovative ways. STEAM knowledge and skills are particularly relevant in the era of technological innovation and Industry 4.0, as they prepare students for careers that require multidisciplinary skills and the ability to tackle complex and ever-changing problems, and are critical to solving many global challenges such as solving environmental problems, developing advanced technologies and creating new jobs.

Industry 4.0 is a concept referring to the fourth industrial revolution that describes the advanced integration of digital technologies, automation and the Internet of Things (IoT) within industrial processes, and in effect represents a radical transformation of the way companies produce and deliver goods and services. This revolution is driven by the use of advanced technologies such as the IoT, artificial intelligence (AI) and augmented reality (AR). In total, we have 9 enabling technologies, also identified within the "Industria 4.0" National Plan, that enable companies to embark on a true journey towards the factory of the future:

1. Advanced manufacturing solutions: advanced interconnected and modular productive systems, therefore high-flexible and performing. Among the main systems we find robotics with collaborative robots;
2. Additive manufacturing: systems of production that improve the resource efficiency of materials, as 3D printer connected to software of digital development;
3. Augmented reality: use of augmented reality systems to support production processes;
4. Simulation: simulation between intelligent and inter-connected machines to increase the productivity and to optimize operations;
5. Horizontal/vertical integration: integration of information and data among all the areas of the supply chain, from vendors to end users;
6. Industrial internet: multi-directional communication among elements of production, both internal and external, through the introduction of the internet;
7. Cloud: implementation of cloud computing solutions and data management techniques in open systems;
8. Cyber-security: new security rules to protect data, more and more exposed to the dangers of impairment for the several internal and external connections;
9. Big data analytics: analysis of large amounts of data to optimize production processes.

The STEAM world forms the foundation on which Industry 4.0 is built, representing the intellectual and practical heart of the fourth industrial revolution. This connection stems from the intrinsic importance of STEAM skills in the evolution of production processes, technologies and operational strategies in industry. Indeed, scientific knowledge provides us with an in-depth understanding of the laws of nature, materials and fundamental processes while technology, on the other hand, transforms this scientific knowledge into concrete solutions. Industry 4.0 is based on advanced technologies such as artificial intelligence, machine learning, robotics and IoT, all based on scientific knowledge and applied to automate and improve industrial processes. Art also plays a key role in Industry 4.0 through the concept of aesthetic, empathic and functional design, where aesthetics acquires a fundamental role in the ergonomics of products, in the design of intuitive user interfaces and in the visual integration of technologies in industrial operations, thus fostering their acceptance by users. Last but not least, mathematics is in effect the language of Industry 4.0; data analysis, mathematical modelling and algorithms are crucial for making informed decisions and developing intelligent solutions, as they help to understand process models, optimise operations and drive machine learning. Industry 4.0 is not just about implementing advanced technologies, but is based on a holistic approach involving scientific knowledge, engineering design, artistic creativity and mathematical logic.

As a result, STEAM training not only provides the skills needed to work in Industry 4.0, but also offers a flexible and multidisciplinary mindset that is crucial for tackling the complex challenges and innovative opportunities that this industrial revolution brings.

The metaverse can also be a useful learning tool within schools in the form of the 'metaverse for learning, inclusion and sustainability', as it can complement traditional classroom learning by offering immersive and interactive experiences that enhance engagement and knowledge retention, thanks to the interaction of avatars through voice and reactions beyond any barrier, the presence of narrating characters who can provide information on the study subjects with related quizzes, the presence of inclusive sports, health and wellness, information panels and displays of school excellence or digital twin educational experiments with STEAM subjects.

3. Results of the survey conducted on STEAM disciplines, metaverse and UN Agenda 2023

STEAM disciplines, as well as the metaverse for learning, positively influence the pursuit of the UN 2030 Agenda goals. Indeed, the link between STEAM and sustainability is very close; indeed, technologies developed under STEAM can help solve environmental problems and

develop sustainable solutions for future challenges such as climate change, resource scarcity and population growth (e.g. solar technology and renewable energy sources were developed through the application of science and engineering). Energy efficiency technologies, such as sensors and smart grids, can help reduce energy waste. Furthermore, artificial intelligence and data-driven solutions can help predict and manage environmental risks and identify opportunities for more sustainable production.

A survey was carried out among various participant types, including university students, teachers, and professional workers. The survey aimed to analyse their comprehension of STEAM disciplines and the metaverse, and their contributions to the accomplishment of UN's 2030 Agenda objectives.

The data in Table 1 show that 84% of the respondents know about STEAM disciplines, while all respondents (100%) know about the metaverse; of these, only 24% do not know about the UN 2030 Agenda goals. Finally, 96% of the respondents believe that both STEAM disciplines and the metaverse can contribute to achieving some of the UN 2030 Agenda goals.

	Yes
Do you know the meaning of STEAM?	84%
Are you familiar with the metaverse?	100%
Do you know the goals of the UN Agenda 2030?	76%
Do you believe that STEAM disciplines can contribute by training new recruits to pursue the objectives of the UN Agenda 2030?	96%
Do you believe that the metaverse can contribute with its educational potential to the pursuit of the objectives of the UN 2030 Agenda?	96%

Table 1. Percentage of positive opinions, in relation to the questions present, out of a total of about 30 respondents

The data in Table 2, on the other hand, show the percentages of the UN 2030 Agenda goals that respondents believe can be achieved through STEAM disciplines and the metaverse. As it turns out, 87.5% of respondents link STEAM disciplines and the metaverse to Goal 4 - Quality education, as they can complement traditional classroom learning by offering immersive and interactive experiences that improve engagement and knowledge retention. The link with Objective 9 - Industry, innovation and infrastructure, on the other hand, was found by 70.8% of the respondents. Goal 11 - Sustainable cities and communities, was voted by 50% of respondents, as by integrating sustainability principles into STEAM education, schools can foster future innovators for the development of sustainable cities, where the metaverse can provide virtual simulations to explore sustainable urban planning and design concepts. Finally, 41.7% of respondents voted for Goal 5 - Gender equality, as STEAM education plays a key role in promoting gender equality by encouraging girls' participation in traditionally male-dominated fields such as science and engineering.

4. Robocom

The innovative SME iInformatica Srl, a private research centre enabled on the MUR (Ministry of University and Research) list, has designed and developed Robocom, a tailor-made solution for schools (primary and secondary) that is highly technological and empathetic and patented on MEPA (Public Administration Electronic Marketplace) and includes a series of paper and digital teaching tools (easily accessible via 'quick approach' directly from the dashboard) to enhance the educational offer through the use of innovative technologies such as virtual and

augmented reality, metaverse, IoT, artificial intelligence and haptic technology.

The 17 Goals	%
Goal 1: No poverty	4%
Goal 2: Zero hunger	0%
Goal 3: Good health and well-being	40%
Goal 4: Quality education	87.5%
Goal 5: Gender equality	41.7%
Goal 6: Clean water and sanitation	8%
Goal 7: Affordable and clean energy	28%
Goal 8: Decent work and economic growth	32%
Goal 9: Industry, innovation and infrastructure	70.8%
Goal 10: Reduced inequalities	44%
Goal 11: Sustainable cities and communities	50%
Goal 12: Responsible consumption and production	28%
Goal 13: Climate action	40%
Goal 14: Life below water	8%
Goal 15: Life on land	16%
Goal 16: Peace, justice and strong institutions	28%
Goal 17: Partnerships for the goals	36%

Table 2. Percentage of positive opinions, in relation to the goals of the UN 2030 Agenda that can be achieved through STEAM disciplines and the metaverse, out of a total of about 30 respondents

a) Interactive 360-VR environment: 360-VR reconstruction of the school grounds (exterior and some building interiors) for opendays and promotion via multimedia content/recognition/trails;
b) Metaverse environment for learning and sustainability: metaverse environment with restricted access with avatars (synchronous interaction) with voice interaction and reactions beyond all barriers, informative storytelling characters/quizzes, inclusive sports, health and wellbeing, information panels and exhibition of school excellence, digital twin of STEAM educational experiment in English;
c) Chemistry in Augmented Reality: app to reproduce in augmented reality, by scanning ARtags, molecules and atoms;

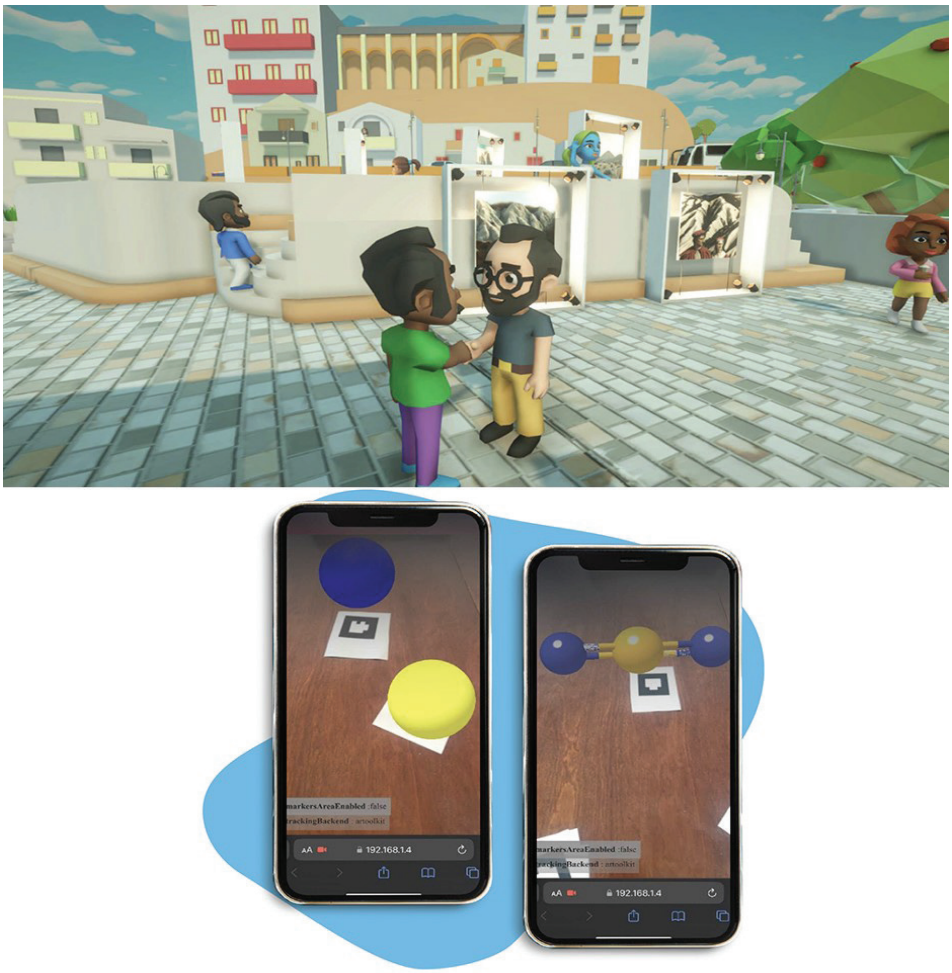


Figure 1. Metaverse environment for learning and sustainability (up), Chemistry in Augmented Reality (down)

d) Kit with NFC (Near Field Communication)/QRcode boards for the study and learning of school subjects (Mathematics, Science, English), card and board games on geography for the enhancement of geographical knowledge (Italy and Europe) and learning through gamification;
e) Inclusion kit consisting of humanoid chatbot for inclusive screening and skills enhancement, Haptic App with haptic and acoustic interaction for presentation of graphical content for visually impaired and learning enhancement through sounds and vibrations, and VR scenario for sustainability journey between civic education and ecology.

The Robocom empathic solution represents a new way of enhancing the educational experience through the combination of virtual reality, artificial intelligence, human-machine interaction and an inclusive haptic approach, in line with the goals of the UN 2030 Agenda and the Faro Convention.

5. Conclusions

The paper described what STEAM disciplines are, how their interdisciplinary approach to learning encourages students to develop a wide range of competences and skills essential to meet

the challenges of the future creatively and consciously, and how they are relevant in schools in relation to the emerging enabling technologies of Industry 4.0. STEAM disciplines enable to increase the digital competences of citizens and increase access to digital tools and services, especially for vulnerable social groups. In this sense, STEAM disciplines have a close link to the NRRP (National Recovery and Resilience Plan) as it has Mission 1 (digitisation, innovation, competitiveness, culture and tourism) and Mission 4 education and research. In this regard, it also introduced Robocom, a tailor-made kit for primary and secondary schools to enhance the educational offer through the use of innovative technologies such as virtual and augmented reality, metaverse, IoT, artificial intelligence and haptic technology.

The STEAM disciplines and Robocom fully espouse the two NRRP missions and are among the most innovative responses to the many projects being carried out by public administrations between now and 2026. Resilient tools that enable public bodies to have innovative solutions for communities and schools to have state-of-the-art equipment and technology. Improved early childhood education and care systems, as well as the skills of the entire population, including digital skills that enable the development of a social policy for new generations, children and young people. These solutions include the transversal priorities of the NRRP, i.e. the protection and enhancement of young people, overcoming territorial divides (south) and increasing awareness of accessibility for all possible users.

A tool, Robocom, that promotes digital transformation and the production and training processes, supporting investments for the innovation of Italy, today and tomorrow. Finally, the survey indicates a high awareness of STEAM disciplines and the metaverse among respondents, as well as a general knowledge of the UN Agenda 2030 goals. The 96% believe that STEAM and metaverse can contribute to the achievement of these goals. Respondents associate STEAM and metaverse with Goal 4 (87.5%), and with Goal 9 (70.8%). Overall, the results suggest a positive perception of the potential impact of STEAM disciplines and the metaverse on the achievement of several 2030 Agenda goals.

References

Beck, D., Morgado, L., O'Shea, P. (2023). Educational practices and strategies with immersive learning environments: Mapping of reviews for using the metaverse. *IEEE Transactions on Learning Technologies.*

Bertrand, M.G., Namukasa, I.K. (2020). STEAM education: student learning and transferable skills. *Journal of Research in Innovative Teaching & Learning*, **13**(1), pp. 43-56.

Idin, S. (2018). An overview of STEM education and industry 4.0. *Research highlights in STEM Education*, pp. 194.

Martinelli, A., Mina, A., Moggi, M. (2021). The enabling technologies of industry 4.0: examining the seeds of the fourth industrial revolution. *Industrial and Corporate Change*, **30**(1), pp. 161-188.

Moraes, E. B., Kipper, L. M., Hackenhaar Kellermann, A. C., Austria, L., Leivas, P., Moraes, J. A. R., Witczak, M. (2023). Integration of Industry 4.0 technologies with Education 4.0: Advantages for improvements in learning. *Interactive Technology and Smart Education*, **20**(2), pp. 271-287.

Sá, M. J., Serpa, S. (2023). Metaverse as a learning environment: Some considerations. *Sustainability*, **15**(3), 2186.

Shatunova, O., Anisimova, T., Sabirova, F., Kalimullina, O. (2019). STEAM as an innovative educational technology. *Journal of Social Studies Education Research*, **10**(2), pp. 131-144.

Skowronek, M., Gilberti, R. M., Petro, M., Sancomb, C., Maddern, S., Jankovic, J. (2022). Inclusive STEAM education in diverse disciplines of sustainable energy and AI. *Energy and AI,* **7***,* 100124.

doi.org/10.26398/asaproc.0017

Combining real-time spatial Delphi judgments and artificial intelligence for the development of future scenarios

Yuri Calleo[a], Simone Di Zio[b], Francesco Pilla[a]
[a] College of Engineering and Architecture, Dublin, Ireland.
[b] Department of Legal and Social Sciences, G. d'Annunzio University of Chieti-Pescara, Pescara, Italy.

1. Introduction and theoretical framework

The exponential increase in complex and dystopian phenomena of environmental nature is rapidly spreading within territories, endangering the overall global population and ecosystem. Governments, local authorities, and scientists believe that if policies are not implemented immediately, dangerous effects could arise, jeopardizing human safety (Xu et al. 2020). In this context, proper planning becomes fundamental to counteract or prevent the effects of the ongoing climate impacts and those that could further increase in the future. Nevertheless, it cannot be adopted without a careful analysis of past and present information, aiming to avoid future threats. The complexity of the territorial structure is a multifaceted and intricate challenge that confronts planners and policymakers. It encompasses several interrelated factors, including geographical features, land-use patterns, infrastructure development, environmental considerations, socio-economic dynamics, and the aspirations of the community (Faludi, 2000). It follows, therefore, that specific approaches must be developed to achieve future objectives.

The study of futures remains an important aspect within the research line of *futures studies* (FS) with the aim of identifying hypothetical scenarios to act in the present to counteract or facilitate, respectively, future threats or opportunities (Kosow and Gaßner, 2008). In particular, scenarios can be considered as "*[…] an internally consistent view of what the future might turn out to be – not a forecast, but one possible future outcome*" (Porter, 1985). In other terms, the ultimate aim is not to predict a future event but to envision different futures to better manage present policies. In the spatial context, scenarios are predominantly crafted using spatial statistical models, which prove highly valuable in examining spatial data, detecting patterns, and making informed decisions regarding the studied phenomena. However, in FS, such models are often challenging to adopt for the following reasons: 1) *Data availability*: they require a substantial amount of data that is often either unavailable or only partially present for small areas. 2) *Prediction*: models produce *forecasts*, which have long been eschewed in the realm of *foresight* due to the inherent impossibility of attaining a singular view of the future. Over time, a transition occurred from the conventional approach of forecasting (which leaves no room for manoeuvring in changing the future) to embracing the exploration of multiple potential futures through the lens of foresight (Martin, 1995).

Among the many methods used to develop scenarios, *mixed-methods* remain a valid solution in order to have both a quantitative and qualitative perspective. Specifically, in this paper, we refer to the method proposed by Di Zio et al. (2017), namely the Real-Time Spatial Delphi (RTSD). RTSD is a customized approach combining the Real-Time Delphi (Gordon and Pease, 2006) and Spatial Delphi (Di Zio and Pacinelli, 2011), specifically designed to aid in foresight and decision-making. It leverages Geographic Information Systems (GIS) and multiple spatial technologies to facilitate expert communication and collaboration within a virtual environment with the final aim to obtain a convergence of opinions on the territory. In this process, experts' judgments become fundamental, both in combination with statistical models and for facilitating final decisions. RTSD in fact, solves one of the main problems of the traditional version of the Delphi method (Linstone and Turoff,

1975), namely the lack of spatial references in the process. In the scientific literature, RTSD found several applications in various research areas, including urban security and decorum (Di Zio et al. 2017), health, air quality, and energy (Castillo et al. 2017). Recently, Calleo et al. (2023), introduced for the first time the concept of "Delphi-based spatial scenarios" through an application in the climate change context, combining RTSD and the scenario method proposed by Bishop et al. (2007).

In this paper, the overall objective is to continue this line of research, providing innovation in the method by adopting a hybrid approach, combining RTSD and Artificial Intelligence (AI). In particular, since RTSD facilitates expert consensus-building on geographical locations, the final outputs are judgments expressed in the form of geographic coordinates (x, y), with a circle representing the consensus achieved—analogous to the interquartile range (IQR) of the classical Delphi procedure—and possible textual comments. However, the outputs produced are not in the form of scenarios and do not provide a narrative or a picture of the plausible future reality. With the rise of AI models, different methods can be used to visualize specific outputs starting from general inputs, including Generative Adversarial Networks (GANs) and Text-to-Image (T2I) models. Overall, this paper proposes to:

1. $\boldsymbol{O_1}$**:** combining Real-Time Spatial Delphi and Text-to-Image models to have a real vision, in the form of images, of the experts' consensus on the territory.
2. $\boldsymbol{O_2}$**:** exploring the capacity of images to foster consciousness and facilitate informed policy-making choices.
3. $\boldsymbol{O_3}$**:** developing a new hybrid method useful in the visioning phase of scenario planning (Bishop et al. 2007).

To showcase our novel method, we involve a panel of 26 experts in the process, asking to evaluate plausible impacts of climate change in 2050 for the city of Dublin, adopting a "Real-Time Geo-Spatial Consensus System (RT-GSCS – www.rtgscs.com, see Calleo et al. 2023). The judgments expressed by the experts in the form of geographic coordinates and textual comments within the consensus circle are implemented in a T2I tool (Adobe Firefly, www.firefly.adobe.com) in order to visualize the future impacts. By embracing this innovative approach, policymakers and experts can enhance the visualization of proposed policies, leading to a more effective assessment of their potential impacts with heightened accuracy.

2. Materials and methods

The method proposed in this paper combines RTSD and Text-to-Image models with the aim of developing future scenarios of possible impacts in 2050 for the city of Dublin and visualising renderings of possible threats in reality. To meet the research objective, we adopt the method proposed by Calleo et al. (2023) implementing it with an additional phase where AI is adopted (*visioning phase*).

The method is composed of the following phases:

1) *Framing*: where desk research is performed. Specifically, we develop the methodology of the study, acquiring spatial data available, and the area of interest in Dublin city. The city of Dublin is part of the SCORE H2020 EU project and is facing multiple challenges posed by coastal flooding in the upcoming years. For these reasons, we want to explore the climate impacts in a reasonable time horizon, identified as 2050.

2) *Scanning*: in this phase, a list of key drivers is extracted. Usually, in the traditional version of the Delphi method, this involves workshops and focus groups with experts, however, in our case, to speed up the procedure, we extract the main drivers from the project proposal since the drivers have been already refined by a group of researchers. In our study we identify six main hazards possibly affecting the future of Dublin in 2050: coastal flooding, land flooding, landslides, heatwave, storm surge, and coastal erosion. From these drivers, we can formulate the questions to be posed to our panel: $RQ1$: "Thinking about 2050, what area will be most at risk of flooding?" $RQ2$: "Thinking about 2050, what area will be most at risk of erosion?" $RQ3$: "Thinking about 2050, what area will

be most affected by extreme events?" Once we have a list of questions validated by the research team, in terms of transparency and clarity, we can proceed with the upload to the platform (RT-GSCS). The chosen panel adheres to the fundamental principles of the traditional Delphi method (Calleo and Pilla, 2023), considering the diverse range of expertise among the participating experts. In fact, we select a cohort of experts as part of two main categories: i) *Internal experts*: members of the project (SCORE H2020), including academics, stakeholders, and local authorities. ii) *External experts*: with a strong level of expertise and strong professional experience, including representatives from companies, local and governmental authorities, and NGO members. We contacted 12 internal experts and 50 external experts, and out of these $E = 26$ experts agreed to participate, including 6 internal and 20 externals.

3) *Forecasting*: in this phase, the Real-Time Spatial Delphi survey is performed. We sent a registration form to each panellist by email, including technical guidelines to access the platform. To pursue the objectives of this paper, we adopt RT-GSCS, a web-based open platform developed in 2023 (Calleo et al. 2023), to achieve a spatial convergence of opinions among panellists, with multiple tools including spatial analysis and real-time algorithms. Once the experts successfully register to the platform, the exercise can start. In this case, the experts can select the questions from a sidebar and answer by placing one or more points on the map. From this point, an automatic circle appears, moving, shrinking, and expanding in real time based on the anonymous responses from other experts. The experts have the option to justify their judgments at any time by providing comments. The statistical algorithm implemented in the platform is suggested by Di Zio and Pacinelli (2011) and aims to obtain convergence of opinions on the territory. Following this logic, spatial convergence is achieved by considering a geometric element identified by a circle C, the smallest among all the potential circles. In this case, C includes 50% of the N judgments – with $N \geq E$, since each expert can give more than one point for each question – (analogue of the IQR of the traditional Delphi). Once the experts place one or more points on the map, we have a vector of judgments $(n_1, n_2, \ldots, n_N)$ for each question, where each n_i is in the form of geographical coordinates (x, y). The main aim of the algorithm is to find a minimum area A_i of a circle C_i covering half of those points $A_i \supseteq T_{(N/2)}$, where $T_{(N/2)}$ denotes a set containing 50% of the N points. Nevertheless, since there are an infinite number of circles (C_i) that satisfy these conditions, we have the constraint that C_i must have its centre in one of the N points. Hence, for each question the algorithm determines a vector $A = A_1, A_2, A_3, \ldots, A_N$ where A_i represents the area of a circle containing 50% of the N points and centred at point n_{i}. Then, $min(A)$ – the smallest among all those circles – corresponds to the geo-consensus. With this approach, we have two types of final outcomes: i) *Geographical results*: the judgments represented in an interactive map. ii) *Non-geographical results*: the spatial and textual results. Geographical results offer an instant visualization, however, they do not depict the specific process of convergence. For this reason, the spatial Delphi (Di Zio and Pacinelli, 2011) involves the calculation of three main indicators to evaluate spatial data. $M_1 = FC(km^2)$ corresponds to the final circle (FC) area in km^2 useful for the identification of the portion of the territory identified. Nonetheless, this measure is absolute and does not consider the study area boundaries and the size of the initial circle. To address this challenge, we also consider as second indicator: $M_2 = 1 - \frac{FC}{S}$, calculated as the ratio between the final circle's area (FC) and the surface (S) of Dublin $(S = 117.8\ km^2)$. This indicator illustrates the level of geo-consensus, and the closer the measure is to 1, the smaller the consensus circle is relative to the surface. The third indicator, measures a dynamic process of the spatial convergence: $M_3 = \frac{FC}{IC} \cdot 100$, where IC is the initial circle area (set a priori as 50 km^2), and the higher the value (closer to 100%), the poorer the convergence of opinions; conversely, the closer it is to zero, the stronger the convergence. Since our process is in real-time, to end the exercise we must take into consideration a stopping criterion, identified in the literature as stability over time (von der Gracht, 2012). For this reason, in our case, we perform time series analysis, and we end the exercise when there is not a significant variation in the total distribution of the N points (usually under 5% of the

points).

4) *Visioning*: the novelty of the method is implemented in this phase. When we obtain the final results, we identify – for this preliminary study – the centre of the final circle area (A_i) of each scenario, but any other location inside the circle of consensus could be used. Currently, the point coordinates of A_i are imported into Google Street View (Google, n.d.) and the related real image of the examined area, is considered as input for our text-to-image model. In this instance, Adobe Firefly is employed as an artificial intelligence tool capable of generating images from textual inputs. This model adopts Generative Adversarial Networks (GANs) consisting of two neural networks: a Generator (G), responsible for generating images, and a Discriminator (D), which differentiates between real and counterfeit images through an adversarial process. At this stage, we bring the corresponding real image for each scenario into the model using the "Generative fill" tool. We then proceed to select the Regions of Interest (ROI), which represent specific areas of the image where the AI performs the generation process from the specific text input. Once the ROI has been identified we used a textual prompt to generate the modified image: for Sc.1 "*Generate a flooded road*", for Sc.2 "*Generate a coastal erosion*", for Sc.3 "*Generate extreme weather condition*" (for Sc. 3 given the general nature of the prompt, the ROI in the image is more expansive and encompassing). Ultimately, as a result, the T2I model generates a modified image based on the provided prompt. This capability can be exceptionally valuable for envisioning future events/scenarios and raising awareness among experts and the general public in order to develop efficient policies in the present.

3. Results and discussion

The study successfully addressed the research objectives, yielding 3 main spatial scenarios. It officially ran from November 1, 2022, to December 5, 2022, following a double stability check. For $RQ1$, 58 expert judgments with 13 comments were collected. $RQ2$ resulted in 54 judgments and 16 comments, while $RQ3$ recorded 40 points and 11 comments. The experts achieved a substantial reduction in the initial circles, exceeding 99% in all final results (referenced as M_2 in Tab. 1), indicating a remarkable level of convergence. In both Sc.1 and Sc.2, the initial circle underwent significant reduction, with M_2 values of 0.993 and 0.999, respectively. The initial circle of Sc.1 measured 8.24 km^2 and decreased to 0.77 km^2, while Sc.2 started at 3.25 km^2 and reduced to 0.15 km^2. In the case of Sc.3, the initial circle had a smaller size of 2.92 km^2, and it was ultimately reduced to 0.54 km^2, representing a reduction of 0.995, as indicated by M_2. In a conventional Delphi study, the Interquartile Range (IQR) is often used as a measure of consensus achieved when the IQR is less than 20% of the measurement scale employed. Likewise, in the Spatial Delphi method, consensus can be considered achieved when M_3 is less than or equal to 20%. In our study, the M_3 values of 9.34%, 4.61%, and 18.49% for the three research questions indicate that the experts achieved a high level of consensus for Sc.1 and Sc.2. However, for Sc.3 there was a slightly lower level of consensus due to the presence of multiple clusters on the territory.

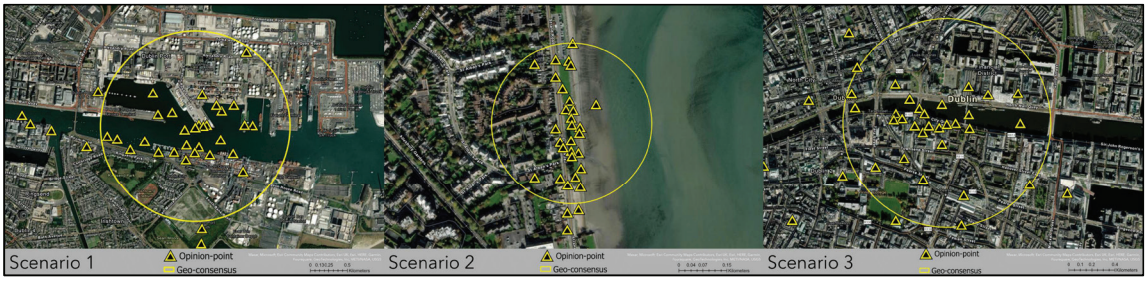
Table 1. Measures of spatial consensus

Scenario	S (km^2)	IC (km^2)	FC (km^2) M_1	M_2	M_3	N
Sc.1	117.8	8.24	0.77	0.993	9.34%	58
Sc.2	117.8	3.25	0.15	0.999	4.61%	54
Sc.3	117.8	2.92	0.54	0.995	18.49%	50

As stated in Section 2, while consensus is an important aspect when determining the survey's stopping criterion, it is not the sole factor, and stability also holds significance. Overall, Sc.1 experienced multiple changes before stabilizing from the 18th to the 20th day. Nevertheless, after validation, three additional changes occurred, highlighting the contentious

flooding issue in the Dublin area and uncertainty regarding appropriate solutions. Sc.2 demonstrated the strongest consensus, remaining stable from the 10th day and only undergoing two changes during validation, signifying agreement on future spatial erosion dynamics. Sc.3 exhibited significant changes in the circle's radius within the first 15 days, reflecting debates about potential extreme event scenarios but stabilizing after the 18th day. Once we illustrated the dynamic process of convergence, spatial analysis is performed adopting ArcGIS PRO (Fig. 1).

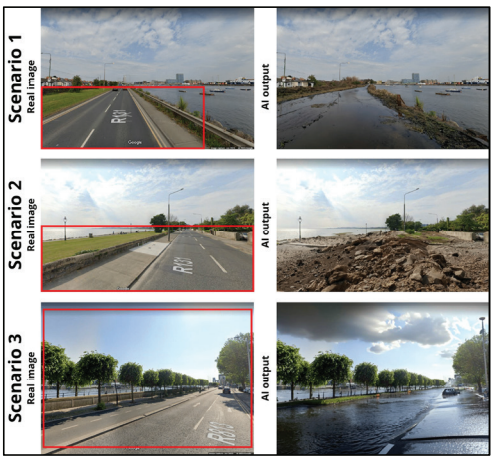
Figure 1. Delphi-based spatial scenarios



Sc.1 depicts that by 2050, the central part of Dublin city, between the banks of the River Liffey, faces the highest threat of flooding. Experts are concerned about potential harm to buildings, infrastructure, essential services, the environment, and even loss of life. Moving to Sc.2, eastern coastal regions of Dublin are identified as most susceptible to erosion by 2050. Coastal erosion here could lead to the loss of valuable real estate, infrastructure, risks to public safety due to unstable cliffs, and negative impacts on tourism and fishing. Regarding Sc.3, the central area of Dublin is seen as most likely to be impacted by various extreme events like storms, floods, and heat waves. Following the experts' comments, consequences include damage to buildings, infrastructure, disruptions to daily life, threats to public safety, strain on emergency services, healthcare, and environmental implications affecting local ecosystems and habitats.

The results provide immediate insights from the experts' judgments; however, they are spatial representations and may not fully convey the reality and magnitude of the threat. Policy makers and citizens might not be aware of the potential implications of future threats. To address this, we generated plausible visual scenarios adopting T2I models, with the aim of providing a clearer understanding of the possible outcomes.

Figure 2. Results from the Text-to-Image model



The generated images are highly significant and offer a clear and well-defined representation of potential future scenarios. In Sc.1, the ROI is replaced by a visually suggestive depiction of a flooded road. Likewise, for Sc.2, we illustrate the erosion phenomena affecting the road. Lastly, Sc.3 demonstrates the impact of extreme events, resulting in the flooding of the road and disrupting transportation.

4. Conclusions and future works

This paper proposed a novel hybrid method combining Real-Time Spatial Delphi and Artificial Intelligence to represent experts' judgments. We employed T2I models to generate plausible visual scenarios, providing clearer insights into potential future threats. These visually suggestive representations offer valuable information to policy makers and citizens, helping them understand the magnitude and implications of the identified threats. However, it is essential to emphasize that these images are hypothetical visions designed to raise awareness among citizens and policymakers, encouraging them to take appropriate actions in the present. In future works, this method can be used to generate realistic images of concrete policies to be adopted in the present in order to facilitate the work of policymakers.

References

Calleo, Y., Pilla, F. (2023). Delphi-based future scenarios: a bibliometric analysis of climate change case studies. *Futures*, **149**, p.103143.

Calleo, Y., Di Zio, S., Pilla, F. (2023). Facilitating spatial consensus in complex future scenarios through Real-Time Spatial Delphi: A novel web-based open platform. *Futures & Foresight Science*, e155.

Castillo-Rosas, J.D., Diez-Rodríguez, J.J., Jiménez-Vélez, A.F., Núñez-Andrés, M.A., Monguet-Fierro, J.M. (2017). Collection and integration of local knowledge and experience through a collective spatial analysis. *ISPRS International Journal of Geo-Information*, **6**(2), p. 33.

Di Zio, S., Pacinelli, A. (2011). Opinion convergence in location: A spatial version of the Delphi method. *Technological Forecasting and Social Change*, **78**(9), pp. 1565-1578.

Di Zio, S., Rosas, J. D. C., Lamelza, L. (2017). Real Time Spatial Delphi: Fast convergence of experts' opinions on the territory. *Technological Forecasting and Social Change*, **115**, pp. 143-154.

Faludi, A. (2000). The performance of spatial planning. *Planning practice and Research*, **15**(4), 299-318.

Gordon, T., Pease, A. (2006). RT Delphi: An efficient,"round-less" almost real time Delphi method. *Technological Forecasting and Social Change*, **73**(4), pp. 321-333.

Kosow, H., Gaßner, R. (2008). Methods of future and scenario analysis: overview, assessment, and selection criteria, **39**, p. 133. DEU.

Linstone, H.A., Turoff, M. (Eds.). (1975). *The Delphi Method*. Reading, MA: Addison-Wesley.

Martin, B.R. (1995). Foresight in science and technology. *Technology Analysis & Strategic Management*, **7**(2), pp. 139-168.

Porter, M.E. (2011). *Competitive Advantage of Nations: Creating and Sustaining Superior Performance*. Simon and schuster.

von Der Gracht, H.A. (2012). Consensus measurement in Delphi studies: review and implications for future quality assurance. *Technological Forecasting and Social Change*, **79**(8), pp. 1525-1536.

Xu, R., Yu, P., Abramson, M. J., Johnston, F.H., Samet, J.M., Bell, M.L., Guo, Y. (2020). Wildfires, global climate change, and human health. *New England Journal of Medicine*, **383**(22), pp. 2173-2181.

doi.org/10.26398/asaproc.0018

On the measurement of cognitive interdisciplinarity with OpenAlex's concepts

Giulio Giacomo Cantone[a], Venera Tomaselli[b], Paul Nightingale[a]
[a] Science Policy Research Unit, Business School, University of Sussex, UK
[b] Department of Economics and Business, University of Catania, IT.

1. Introduction

Scientific interdisciplinarity is intuitively understood as a property of scientific production. The more research is interdisciplinary, the more is able to bring together assumptions, perspectives, and results that belong to separate disciplinary traditions (Wagner et al., 2011). Under these premises, the potential benefits of fostering interdisciplinary research (IDR) have been widely discussed. Proponents argue that it stimulates innovation by encouraging the exploration of novel perspectives and facilitating breakthrough discoveries. Reservations against IDR revolve around the difficulty of conducting fair evaluations of IDR activities Seeber et al., 2022).

Empirical evidence on the outcomes of interdisciplinary research presents a mixed picture, with a (possibly misleading) prevalence of positive outcomes reported (Yegros-Yegros et al., 2015). Rafols et al. (2012), demonstrating that IDR is systemically undervalued and possibly underfunded. The intuitive argument is that scientific careers are defined by the choice to maximise the chances to get research funds. If research products too divergent from the core topics of famous scientific journals are poorly evaluated, the intellectual stimulation coming from IDR is actively impeded by the evaluation systems. Indeed, proponents of IDR also advocate for alternative models of research evaluation that do not penalise interdisciplinary forms of innovation (Wagner et al., 2011).

This study highlights novel possibilities to adopt bibliometric data provided by the open database OpenAlex (OpAl). Section 2 presents the state-of-the-art in indicators of interdisciplinarity. These indicators are canonically defined on data of citations, instead in Section 3 we will adapt them for the measurement of IDR through the 'scientific concepts' introduced by OpAl. Along the canonical measure for interdisciplinary *Diversity*, we introduce a novel indicator of *Difformity* as a divergence between the observed disciplinary profile of a published paper and the typical expectation from the journal where it is published.

2. How interdisciplinary is measured

Across the rich literature on the measurement of interdisciplinarity, it is possible to identify two main dichotomies for the operative definition of IDR activities. These are:

- Cognition *vs.* Organisation. It refers to the unit of analysis of IDR. The 'organisation' approach is interested in atypical combinations of authorships. The cognitive approach is much less interested in combinations of authorships, and much more in the transmission of concepts across papers and journals. This can happen directly, with the application of text mining techniques, or, as more frequently seen, comparing lists of references and citations. On a more abstract level, 'cognition' implies that scientific authors, who operate within disciplines, still 'recognise' the relevance or the usefulness of concepts lying outside the typical bounds of their own disciplines (Abramo et al., 2018).

• Integration *vs*. Diffusion. Integration can be defined as the capacity to combine a list of different inputs and arrange them in a coherent way. Commonly, these inputs are usually observed through analytical operations on the features of the reference lists of the papers. Diffusion refers to the capacity of a scientific product to be cited, mentioned, and replicated across different other papers, and it can be also recognised as a dimension of the scientific impact.

Although there have been attempts to provide a more complete operative definition of Integration (Rafols, 2014), this concept is often conflated with the measurement of diversity in the disciplinary profile of the paper. *Diversity* is paradigmatically defined as the interaction of three factors:

$$Variety \times Balance \times Disparity \quad (1)$$

Variety (a.k.a. ecological richness) implies that the authors relate their production to many disciplinary categories, independently of the difference of relevance given to these and to how these categories are similar to each other. Balance (a.k.a. ecological evenness) implies that the variance in the relevance of the categories is minimal. These two factors are usually measurable jointly (or not) with conventional indicators, e.g. Simpson index of repeat Rousseau (2018). The third factor of Disparity measures the similarity across the categories and it is commonly associated with the introduction of a matrix of similarity-dissimilarity in the equation, as in the Rao index of quadratic entropy (Stirling, 2007).

3. Measurement of integration in cognitive interdisciplinarity

We propose that in the bibliometric methodology exists a paradigmatic model of Normal Interdisciplinarity, which, referencing the aforementioned dichotomies is the degree of cognitive integration of a disciplinary profile. This model of measurement is not only the most frequent in literature, but it is also the most justifiable in practical terms. Typically, papers have only a few authors but many references, so reference lists have a higher size. The variation in the size of references across papers also has a finite variance, propriety not holding for lists of citing papers (Diffusion), because the number of citations grows over time.

According to Mugabushaka et al. (2016), indicators of interdisciplinarity are historically understood as a procession of 'generations'. Assuming a $X{:}\{i,j,\dots\}$ system of countable elements or numeric traits, the 1^{st} generation of indicators of diversity consists of measures of entropy of first order (including Gini-type entropies). An accepted indicator of *Diversity* of '2^{nd} generation' is the aforementioned Rao-Stirling index of quadratic entropy:

$$D_{RS}(X) = \sum_{(i,j)} [p(X,i) \cdot p(X,j)]^{\beta} \cdot d_{(i,j)}^{\alpha} \quad (2)$$

where p_i and p_j are couples of the relative frequencies of the elements or the normalised score (i.e. $\frac{\text{score}}{\sum\text{score}}$) of the trait, and $0 < d_{(i,j)} < 1$ is a value for the dissimilarity between i and j. Indeed, $d_{(i,j)}$ acts as prior about the expected value of $p(i) \cdot p(j)$, unconditional to X; low dissimilarity penalises the apportion of the couplet to $D_{R\square}(X)$. α and β are modelling parameters which are canonically set equal to 1 in parametrically naive measurement models[1].

[1] Notice that for $p(i) = p(j)$ when $d_{(i,j)} = 1$ (or, alternatively, $\alpha = 0$), Eq. 2 collapses into the Simpson index of repeat Rousseau (2018), also known as Hirschman index. Hence, the naive parameterisation of Rao-Stirling is considered the canonical 2^{nd} generation index.

Eq. 2 has been criticised for two reasons:

1. It lacks the propriety referred in many ways: "trueness", "replication principle" or "composition principle". Trueness means that given n sets $X_1 \dots X_n$, such that their $D_{RS}(X)$ is equal for all of them and all of them have no elements in common, then the diversity of their union is n times $D_{RS}(X)$, i.e. $D_{RS}\left(\cup\,(X)\right) = n \cdot D_{RS}(X)$. As a corollary of the lack of "trueness", D_{RS} has a low discriminant power as a test statistic (e.g. to answer questions as "is X_1 significantly more diverse than X_2?") (Zhang et al., 2016).
2. The Rao index inherits propriety of Simpson's: non-monotonicity to balanced addition. This propriety implies that by adding any new non-empty category ($p_i \neq 0$), the index will differ, and this is even in the case of $p_i = \bar{p}_i$. This propriety contrasts with the monotonic behaviour of the Gini index based on the Lorenz curve [2]. According to authors who debated on this feature, the adoption of Rao-Stirling implies that Variety and Balance are not dual but a unique feature (Leydesdorff et al., 2019; Rousseau, 2019).

Authors such as Leydesdorff et al. (2019) and Rousseau (2019) proposed to adopt a different index based on Gini's index to solve the second issue. On the contrary, Zhang et al. (2016), inspired by the systematisation of the theory of Hill-type measures in Leinster and Cobbold (2012) noticed that adopting the transformation

$$D_{RS}^{(T)} = \frac{1}{1-D_{RS}} \tag{3}$$

the first issue is solved (Mugabushaka et al., 2016). These advancements constitute the 3^{rd} generation of indexes of *Diversity*. In the following application, we will propose how to expand this framework to welcome new variables of IDR present in OpAl's database.

3. Application

3.1 Sampling Frame

Let a paper be symbolised by X, a journal as K, and an author as A. Papers have many authors, so the fractional contribution to the authorship of an author is $a_A(X) = \frac{1}{N(A_X)}$.

We queried the OpAl catalog (Aria and Le, 2023) for papers published in the years 2018, 2019, 2021, and 2022 in 939 Class A journals in the official list for Disciplinary Area 13 of the Italian National Agency of Evaluation of the University and Research System (ANVUR). These are considered by ANVUR the most relevant international journals for Economics, Business, Management, Finance Statistics, and Demography.

We fetched 31,632 papers with at least one author classified as affiliated with an Italian university. We sampled 64 research units (U) as departments of Economics, Management, Statistics, Business or Finance[3]. Only 7,280 papers have been authored by at least one author affiliated with these 64 research units.

---

[2] In statistical software R, `abdiv::simpson(c(1,1)) != abdiv::simpson(c(1,1,1))` holds, while `DescTools::Gini(c(1,1)) != DescTools::Gini(c(1,1,1))` it does not hold.

[3] Two departments from the same university would be counted as different research units. This is a representative and sature sample of departments of Area 13 - ANVUR in Italy. Only small departments have been ecluded from the sample. In some cases (e.g. Bocconi Schools, or University of Calabria) it was convenient to consider an aggregation of departments as a research unit, instead.

In this context, a research unit (U) consists of a list of authors. The sum of their authorship can roughly estimate the nominal total research output of the research unit, i.e. $y_U = \sum a_A(X) \mid A \in U$.

3.2 Methods

In OpAl each unitary record (an X journal, a K journal, etc.) is related to a 'concept' through a c score. Concepts are labels associated with scientific activities and are classified through different levels. At level 1, they work as disciplinary labels. In the context of OpAl, it is convenient to treat disciplinary labels as observable traits (i) of the record.

Even concepts have a score that relates to each other, so for the determination of the aforementioned matrix of similarities we adopted the normalised scores of similarity provided by OpAl:

$$d_{(i,j)} = \begin{cases} 0; & i = j \\ \frac{\delta_{ij}}{\max(\delta_{ij})}, \delta_{ij} = \frac{c(i,j)}{\sum c(i,j)}; & i \neq j \end{cases} \quad (4)$$

where $c(i,j)$ is the similarity score that OpAl assigns to disciplinary labels i and j.

Each u research unit is evaluated through its average score of *Diversity* and *Difformity*. The second is a dimension still unexplored in literature, which is a measure of how unexpected is the disciplinary profile of a paper within the context of the journal where it is published.

3.2.1 Measurement of Diversity through OpenAlex's concepts

The estimator $p_i = \frac{c_{X,i}}{\sum c_{X,i}}$ is considered for the determination of the *Diversity* (Eq. 3) of a paper. The proposed method has the advantage of requiring much less information and computation over the canonical alternative of formulating p_i as the disciplinary proportion of the reference list (Zhang et al., 2018).

Since $D_{RS}^{(T)}$ has the propriety of trueness (or, 'composition'), its linear aggregation is not biased. In this case, in order to balance the difference in research outputs across research units (y_U), we are interested in the average

$$D_{RS}^{(T)}(\bar{U}) = \frac{\sum_{A \in U} D_{RS}^{(T)}(X_A) \cdot a_A(X)}{\sum_{A \in U} a_A(X)} \quad (5)$$

3.2.2 Measurement of Difformity through OpenAlex's concepts

Difformity is the dissimilarity of the X paper from the expectation of the archetypical $\langle X_K \rangle$ paper published in its K_X-journal.

Let the normalised absolute divergence between the portions of disciplinary scores observed by OpAl between X and K_X be

$$\phi(X,i) = \frac{|p(X,i) - p(K_X,i)|}{\sum |p(X,i) - p(K_X,i)|} \quad (6)$$

then the 'true' estimator for *Difformity* ($\Phi(X)$) can be derivated by Eq. 2 and Eq. 3

$$\Phi^{(T)}(X) = \frac{1}{1 - \sum_{(i,j)} \phi(X,i) \cdot \phi(X,j) \cdot d_{(i,j)}} \quad (7)$$

and the estimator for the average *Difformity* of the research unit can be derived by Eq. 5.

4. Results

Figure 1 shows the distribution of the 64 research units across the two dimensions and two-time windows. Papers published in the years 2018 and 2019 are considered a sample of scientific production just *before* COVID-19, while those in 2021 and 2022, *after* it.

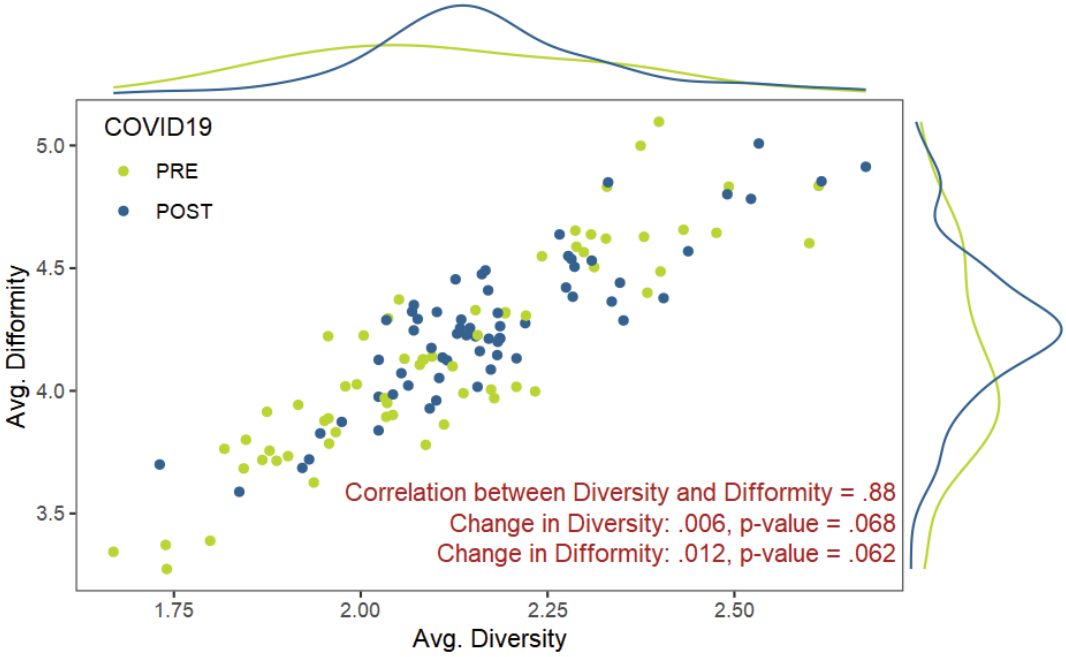


Figure 1. Average *Diversity* and *Difformity* before and after 2020 in 64 research units.

Maybe after COVID-19 research units focused more scientific production on a diversified research agenda, possibly as feedback from high-quality journals accepting different papers from their own disciplinary canon. However, evidence is still not sufficient to claim that Italian research in high-quality journals became on average more diversified after COVID-19.

There is a significant reduction in the dispersion of *Diversity* (F-Levene: 5.69, p-value = $.019$) and *Difformity* (F-Levene: 8.25, p-value = $.005$). Another observed change is the following: before COVID-19 no significant correlation is observed between research output (y_U) and *Diversity*. However, in years after COVID-19 there is an ambiguous positive correlation between the two (Pearson corr. coefficient: $.23$, p-value = $.06$). Considering that y_U depends on the size of the departments, even the increase in average research output before and after 2020 (6.95 more authorships, p-value = $.051$) is ambiguous, too.

These results should be calibrated with the credibility of OpAl's assessment of discipline. In Figure 1, *Difformity* is usually around double of *Diversity*. This is by-effect of the addition of *Richness* of the journal's discipline in Eq. 7. In other words: in many cases, disciplinary labels diverge between journal and paper. This could be a signal that OpAl's level 1 concepts are not an accurate representation of a disciplinary reality in scientific papers. A suggested development is to consider the taxonomy at level 0 to improve robustness of results.

Acknowledgements and Funding

This research was supported by a grant from the Schmidt Family Foundation to the Science Policy Research Unit at the University of Sussex.

References

Abramo, G., D'Angelo, C.A., Zhang, L. (2018). A comparison of two approaches for measuring interdisciplinary research output: The disciplinary of authors vs the disciplinary diversity of the reference list, *Journal of Informetrics*, **12**(4), pp. 1182–1193.

Aria, M., Le, T. (2023). *OpenAlexR: Getting Bibliographic Records from 'OpenAlex' Database Using 'DSL' API*.

Leinster, T., Cobbold, C. A. (2012). Measuring diversity: the importance of species similarity, *Ecology*, **93**(3), pp. 477–489.

Leydesdorff, L., Wagner, C. S., Bornmann, L. (2019). Interdisciplinarity as diversity in citation patterns among journals: Rao-Stirling diversity, relative variety, and the Gini coefficient, *Journal of Informetrics*, **13**(1), pp. 255–269.

Mugabushaka, A. M., Kyriakou, A., Papazoglou, T. (2016). Bibliometric indicators of interdisciplinarity: the potential of the Leinster–Cobbold diversity indices to study disciplinary diversity, *Scientometrics*, **107**(2), pp. 593–607.

Rafols, I. (2014). Knowledge integration and diffusion: Measures and mapping of diversity and coherence. In Y. Ding, R. Rousseau, D. Wolfram (eds.). *Measuring Scholarly Impact: Methods and Practice*, pp. 169–190. Springer International Publishing, Cham.

Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., Stirling, A. (2012). How journal rankings can suppress interdisciplinary research: A comparison between Innovation Studies and Business & Management, *Research Policy, Elsevier*, **41**(7), pp. 1262–1282.

Rousseau, R. (2018). The repeat rate: from Hirschman to Stirling, *Scientometrics*, **116**(1), pp. 645–653.

Rousseau, R. (2019). On the Leydesdorff-Wagner-Bornmann proposal for diversity measurement, *Journal of Informetrics*, **13**(3), pp. 906–907.

Seeber, M., Vlegels, J., Cattaneo, M. (2022). Conditions that do or do not disadvantage interdisciplinary research proposals in project evaluation, *Journal of the Association for Information Science and Technology*, **73**(8), pp. 1106–1126.

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, **4**, pp. 707–719.

Wagner, C.S., Roessner, J.D., Bobb, K., Klein, J.T., Boyack, K.W., Keyton, J., Rafols, I., Borner, K. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature, *Journal of Informetrics*, **5**(1), pp. 14–26.

Yegros-Yegros, A., Rafols, I., D'Este, P. (2015). Does interdisciplinary research lead to higher citation impact? The different effect of proximal and distal interdisciplinarity, *Plos One*, **10**(8): e0135095. Publisher: Public Library of Science.

Zhang, L., Rousseau, R., Glanzel, W. (2016). Diversity of references as an indicator of the interdisciplinarity of journals: Taking similarity between subject fields into account, *Journal of the Association for Information Science and Technology*, **67**(5), pp. 1257–1265.

Zhang, L., Sun, B., Chinchilla-Rodriguez, Z., Chen, L., Huang, Y. (2018). Interdisciplinarity and collaboration: on the relationship between disciplinary diversity in departmental affiliations and reference lists, *Scientometrics*, **117**(1), pp. 271–291.

doi.org/10.26398/asaproc.0019

Beyond the composite indices of development: The development profile approach

Sara Casacci[a], Adriano Pareto[a]
[a] Department for development of methods and technologies for production and dissemination of statistical information, Italian National Institute of Statistics (Istat), Rome, Italy

1. Introduction

In recent decades, the need to go beyond GDP (gross domestic product), as a measure of 'welfare' (Stiglitz et al., 2009), and find complementary measures that can provide a holistic and multidimensional view of development has led to a variety of composite indices. The most important are the Human Development Index (HDI) and the Gender Development Index (GDI), published by the United Nations Development Programme (UNDP), and the Better Life Index (BLI), created by the Organisation for Economic Cooperation and Development (OECD) (Boarini and Mira D'Ercole, 2013). However, several studies point out that composite indices need to be accompanied by 'dashboards' that give a very detailed description of the observed phenomenon. In both cases, a set of individual indicators with different units of measurement must be normalized to make them comparable. Generally, the individual indicators are normalized independently of each other using univariate methods. For example, in the HDI, they are linearly transformed into indices between 0 and 1, and this normalization does not take into account the curvilinear relationships typical of development data (McGranahan et al., 1985). However, normalization has a strong impact on the results as it creates a 'correspondence system' among different variables. Therefore, the normalization method should take into account the relationships that exist in the data.

In this paper, a nonlinear multivariate method is used to normalize a set of development indicators, aggregating them into a composite index and creating profiles of country performance. The method allows to overcome the limitations of classical methods based on linear models, such as regression, principal component analysis and path analysis.

The basic idea is that given a set of development indicators for a number of countries, at a fixed time, the point cloud is distributed, in multidimensional space, along a curvilinear path extending from the lowest to the highest level of development. Therefore, the aim is to estimate the single curve in multidimensional space that best fits the scatter of points. Once the multidimensional curve is identified, it is used to linearize the relationships between the variables, normalize them, and create a 'correspondence grid' showing the original values of the development indicators for each successive level of the common scale. For each country, the indicators values can be plotted on this grid to draw a development profile. The correspondence grid allows comparisons of levels among indicators (for each country), among countries (for each indicator) and among values at different times.

2. The alternative approach

The alternative methodology is based on an idea by McGranahan et al. (1985) and allows a set of individual indicators to be normalized together to construct development profiles for a number of countries, without making assumptions about the normality of the data, linearity and causality or dependence between variables.

It rests on the use of nonlinear principal component analysis (NLPCA), which transforms a set of variables simultaneously to maximize the variance accounted for by the principal

components in the transformed variables. The first principal component in the transformed indicators is the BFL and its projection onto the space of the original indicators yields a set of corresponding points, each of which represents a set of values covering all the indicators involved. The table describing the correspondence system between the original values of each indicator and the values of the common scale (the scores of the first principal component) is called the *correspondence grid*.

The main steps of the procedure are as follows.

1. Given the matrix $\mathbf{X}=\{x_{ij}\}$, where x_{ij} is the value of development indicator j for country i, compute the transformed matrix $\mathbf{Y}=\{y_{ij}\}$ with $y_{ij} = f_j(x_{ij})$, where f_j is a monotonically increasing spline transformation such that the largest eigenvalue of the PCA of the matrix $\mathbf{Y}$ is maximized. This step is obtained by NLPCA of the matrix $\mathbf{X}$.
2. Given the factor loading a_{j1} for development indicator j on the first factor (the BFL), calculate the predicted transformed value of indicator j by the formula $y_{zj} = z\, a_{j1}$, where z ranges from approximately -3 to +3 with step 0.1 (or as desired).
3. Construct the correspondence grid $\mathbf{G}=\{g_{zj}\}$ with $g_{zj} = f_j^{-1}(y_{zj})$, where g_{zj} is the value of development indicator j corresponding to score z of the common scale[1].
4. Compute the normalized matrix $\mathbf{Z}=\{z_{ij}\}$ with $z_{ij} = y_{ij} / a_{j1}$, where z_{ij} is the value of normalized indicator j for country i.
5. Construct a development profile and/or a composite index, for each country, based on normalized values.

3. An application

In this section, an application to a set of wealth and consumption indicators for the 107 Italian provinces, in 2022, is presented (source: https://github.com/IlSole24ORE).

The individual indicators used are: "Value added per capita (thousands of euros)" (X_1), "Consumer household bank deposits (thousands of euros per capita)" (X_2), "Invoice payments after 30 days (% of commercial invoices to suppliers)" (X_3), "Household expenditure on durable goods (euros per year)" (X_4), "Investments (euros per inhabitant)" (X_5), "Citizen income recipients (households per thousand inhabitants)" (X_6). Note that X_3 and X_6 have negative polarity[2].

The summary statistics – mean, median, standard deviation, skewness and kurtosis – for each indicator are presented in Table 1.

Table 1. Descriptive statistics for X_1-X_6

Statistic	X1	X2	X3	X4	X5	X6
Mean	26.14	19.17	10.11	2,687.18	55.34	15.96
Median	26.00	20.16	8.80	2,746.00	51.60	10.89
Std. dev.	7.20	4.34	4.11	473.62	29.46	12.92
Skewness	0.77	-0.62	-0.83	-1.30	-0.97	0.85
Kurtosis	0.57	-0.31	0.45	-0.28	0.36	1.31

[1] Note that if the data were normalized by z-scores, at level 0 of the correspondence grid we would have the set of the mean values of the individual indicators. This correspondence is correct only if the individual indicators are linearly interrelated.

[2] The polarity of an individual indicator is the sign of the relation between the indicator and the phenomenon to be measured (Mazziotta and Pareto, 2020).

None of the individual indicators are distributed according to the normal bell-shaped curve, or an approximation to it. X_2, X_3, X_4 and X_5 have left-skewed distributions (negative skewness). In addition, X_3, X_4 and X_5 have a bimodal distribution with two distinct peaks. In contrast, X_1 and X_6 have right-skewed distributions (positive skewness), with the highest frequency concentrated around the lowest values. In this case, the median of the two indicators is smaller than the mean.

Table 2 contains the correlation matrix for the original and normalized values. The correlations between the individual indicators when using the original data vary, in absolute value, from 0.743 (X_1 versus X_6) to 0.886 (X_3 versus X_6). When the individual indicators are normalized the correlations increase and range from 0.822 (X_1 versus X_3) to 0.918 (X_3 versus X_6). The normalized data correlations are much higher than the correlations of the original data, mainly because they are not affected by curvilinearity. The most pronounced curvilinearities – and the greatest consequent increase in correlation with data normalization – can be found in the relationships between X_6 and the other indicators.

Individual indicator	X1	X2	X3	X4	X5	X6
			Original values			
X1	1.000	0.844	-0.765	0.803	0.764	-0.743
X2	0.844	1.000	-0.824	0.797	0.794	-0.824
X3	-0.765	-0.824	1.000	-0.840	-0.846	0.886
X4	0.803	0.797	-0.840	1.000	0.796	-0.844
X5	0.764	0.794	-0.846	0.796	1.000	-0.744
X6	-0.743	-0.824	0.886	-0.844	-0.744	1.000
			Normalized values			
X1	1.000	0.855	0.822	0.870	0.846	0.864
X2	0.855	1.000	0.837	0.829	0.823	0.868
X3	0.822	0.837	1.000	0.838	0.881	0.918
X4	0.870	0.829	0.838	1.000	0.847	0.907
X5	0.846	0.823	0.881	0.847	1.000	0.857
X6	0.864	0.868	0.918	0.907	0.857	1.000

Table 2. Correlation matrix for X_1-X_6 (original and normalized values)

Figure 1 shows the bivariate relationships between each pair of original indicators and the BFL. It provides a set of correspondence points that allow us to construct the correspondence grid.

The correspondence grid is presented in Figure 2, where two extreme cases with very different profiles – Roma and Bari – are shown. Roma has a fairly high level of wealth and consumption (mean of normalized values of 0.160), but a set of very 'unbalanced' individual indicators. In particular, it has a deficit in X_3, X_5 and X_6 and a surplus in X_1 and X_2. In contrast, Bari has a lower level of wealth and consumption (mean of normalized values of -0.610) and all individual indicators are more or less at the same level.

The development profile is a descriptive analytical tool that provides a general picture of each country that is superior to any attempt at a crude listing of its values on various indicators (the 'dashboard') or a single composite index. It can lead to the identification of areas where the government needs to take action and support decision-making in allocating the resources needed to implement a strategy. It can also help to monitor actions, evaluate the results achieved and make decisions on how to change policy in the future.

Figure 1. Matrix plot of X_1-X_6. Original values and best-fitting line (BFL)

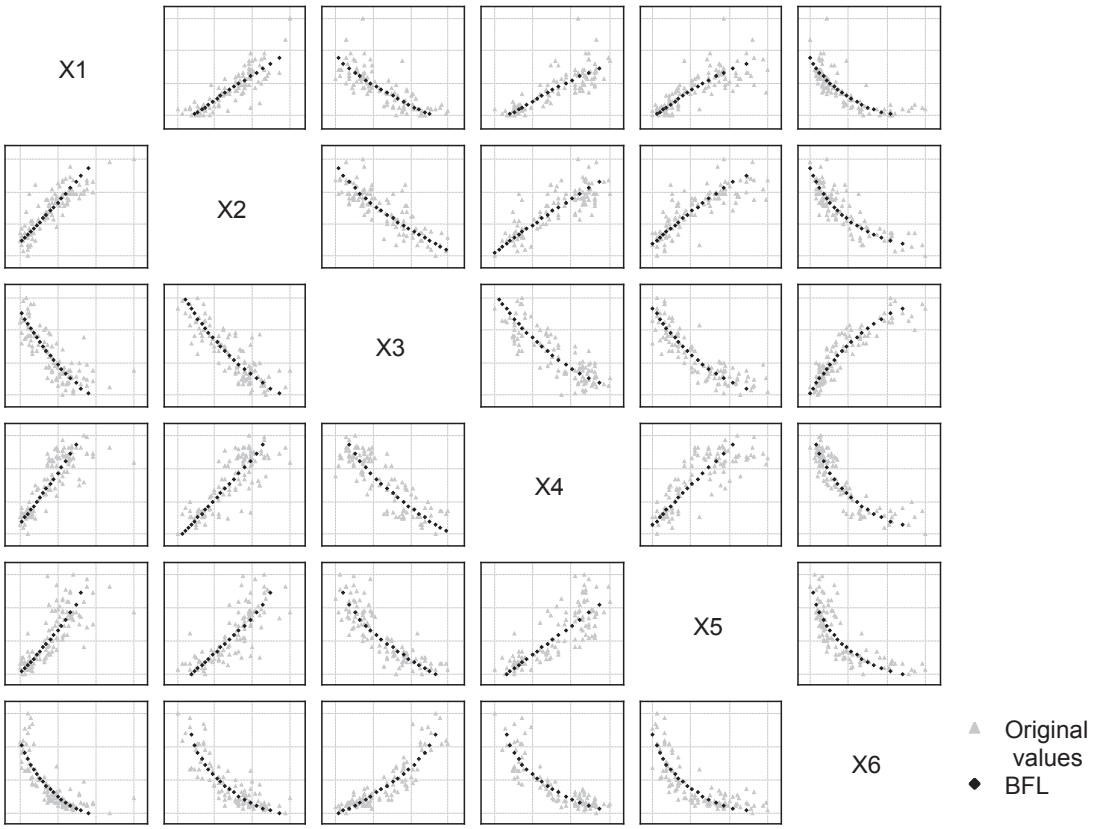


Figure 2. Development profile. Roma vs. Bari

Index level		X1	X2	X3	X4	X5	X6
2.00		43.6	-	-	-	-	-
1.80		40.4	-	-	-	-	-
1.60		37.9	27.5	3.9	-	-	0.7
1.40		35.8	25.9	4.7	-	101.1	2.1
1.20		33.9	24.5	5.4	3,315	88.4	3.7
1.00		32.2	23.3	6.2	3,159	79.3	5.2
0.80		30.6	22.2	7.0	3,027	71.8	6.8
0.60		29.2	21.2	7.8	2,910	65.3	8.5
0.40	**Roma**	27.8	20.3	8.6	2,802	59.4	10.2
0.20	**0.160**	26.5	19.4	9.3	2,703	54.1	12.0
0.00		25.2	18.6	10.1	2,610	49.1	13.9
-0.20		24.0	17.8	10.9	2,523	44.5	15.9
-0.40	**Bari**	22.8	17.0	11.7	2,441	40.1	17.9
-0.60	**-0.610**	21.7	16.3	12.4	2,362	36.0	20.1
-0.80		20.7	15.6	13.2	2,287	32.0	22.5
-1.00		19.6	14.9	14.0	2,215	28.2	25.0
-1.20		18.6	14.3	14.8	2,145	24.6	27.7
-1.40		17.6	13.6	15.5	2,078	21.0	30.8
-1.60		16.7	13.0	16.3	2,013	17.6	34.3
-1.80		15.7	12.4	17.1	1,950	14.4	38.4
-2.00		-	11.8	17.8	1,888	11.2	44.2
-2.20		-	11.3	18.6	1,829	-	-
-2.40		-	10.7	19.4	1,771	-	-
-2.60		-	10.2	-	1,714	-	-

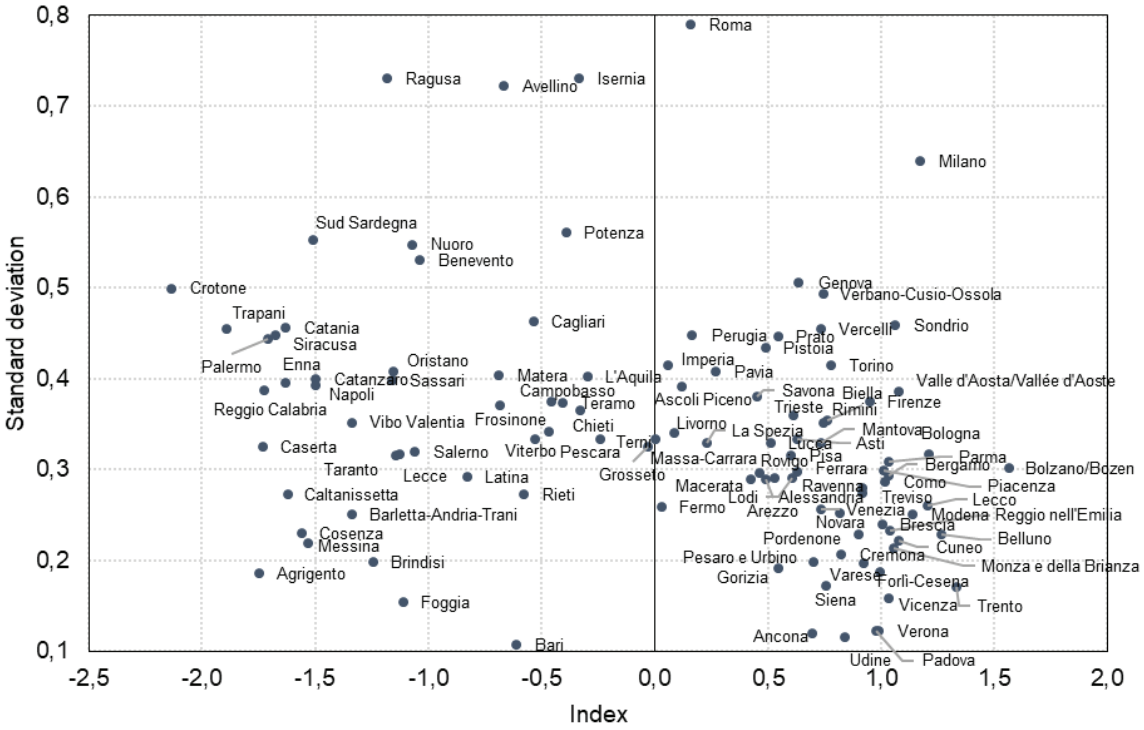
For example, the Roma profile shows both a high level of household bank deposits and a high level of citizen income recipients. This could suggest to policy makers that reducing inequalities among households should be a priority on the policy agenda.

However, it should be noted that the development profile helps in diagnosis and points out possible imbalances in the phenomenon under study, but it is not a direct assessment tool. It tells how the pattern of a province is in relation to the others, but not how it should be. Therefore, it can also be used to show that provinces with the same general level of wealth and consumption and different natural conditions tend to develop differently.

Of course, the composite index resulting from the mean of normalized indicators can be represented graphically in various ways. Figure 3 displays the relationship between the composite index and the standard deviation of normalized indicators, for the 107 provinces.

The lower right part of the diagram shows the most balanced provinces (low values of the standard deviation) with the highest level of development (high values of the composite index). These are mainly the northern provinces (e.g., Trento, Vicenza, Verona, Monza). The upper left part there shows the most unbalanced provinces (high values of the standard deviation) with the lowest level of development (low values of the composite index). Roma, Ragusa, Avellino, Isernia and Milano show the greatest imbalance between the normalized indicators. However, Milano has a very high level of development, while Ragusa, Avellino and Isernia have a rather low level of development. The provinces of Toscana, Umbria and Marche have a medium-high level of development (index values greater than 0) and a medium-low standard deviation of normalized indicators. Instead, the provinces of Lazio (with the exception of Roma) and Abruzzo have a lower average level of development. The lowest values of the composite index are recorded in some provinces of Calabria, Sicily and Campania. In particular, Crotone, Trapani, Agrigento, Caserta, Reggio Calabria and Palermo are the most critical areas. However, the profiles of these provinces are not similar: Crotone, Trapani and Palermo also show a strong imbalance between the normalized indicators, while in Agrigento all individual indicators are more or less at the same level.

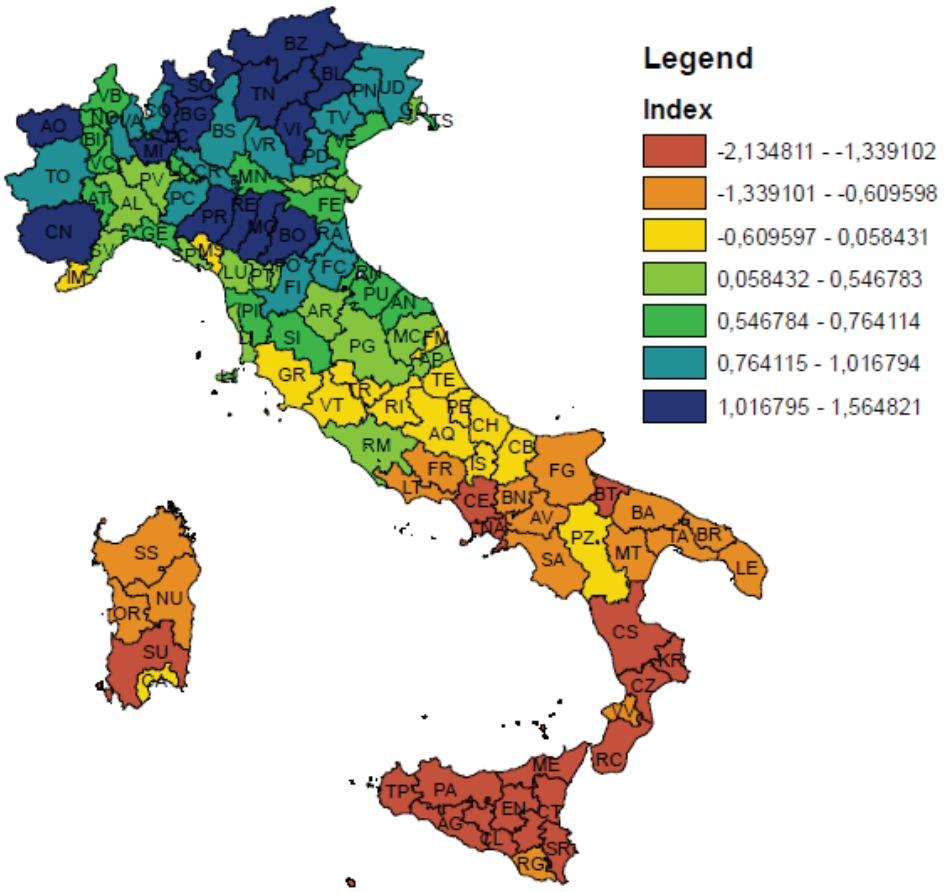
Figure 3. Composite index vs. standard deviation of normalized indicators



Finally, Figure 4 shows a cartogram of the composite index using a quantile classification in which provinces are divided into 7 classes so that the total number of provinces in each class is approximately equal. As can be seen, wealth and consumption levels are highest in the north of the country, particularly in Trento, Bolzano, Aosta and some provinces in Lombardy, Piedmont and Emilia-Romagna. On the other hand, the lowest levels are recorded in the south and in the islands.

These results are consistent with those of the Istat survey on household consumption expenditure, and the Bank of Italy survey on household income and wealth, for 2022.

Figure 4. Cartogram of the composite index by province



References

Boarini, R., Mira D'Ercole, M. (2013). Going beyond GDP: An OECD Perspective. *Fiscal Studies*, **34**(3), pp. 289-314.

Casacci, S., Pareto, A. (2021). A nonlinear multivariate method for constructing profiles of country performance. *International Journal of Multicriteria Decision Making,* **8**(4), pp. 295–312.

Mazziotta, M., Pareto, A. (a cura di) (2020). *Gli indici sintetici*. Giappichelli, Torino, (IT).

McGranahan, D., Pizarro, E., Richard, C. (1985). *Measurement and Analysis of Socioeconomic Development*. United Nations Research Institute for Social Development, Geneva, (CH).

Stiglitz, J., Sen, A., Fitoussi, J. (2009). *Report by the Commission on the Measurement of Economic Performance and Social Progress*. http://www.stiglitz-sen-fitoussi.fr.

doi.org/10.26398/asaproc.0020

How is the use of AI perceived in a classroom environment?

Rosanna Cataldo, Maria Gabriella Grassia, Marina Marino, Violetta Simonacci
Department of Social Sciences, University of Naples Federico II, Naples, Italy

1. Introduction

We are living in an era defined by remarkable technological advances. Digital transformation and artificial intelligence (AI) have become the driver of the transformation of the economy and society, transforming our lives and our ways of communicating. Recent developments in technology in general and in AI, in particular, have also impacted education (Sekeroglu et al., 2019; Kumar, 2019). Applications of AI in the domain of education for predicting student performance, detecting undesirable student behaviour, or providing feedback for supporting instructors and students, are becoming more common (Baker et al., 2009).

This work is part of the evaluation proposal for the experimentation of the ClassMate Robot (CMR) project, promoted by the Protom Group (with Protom Robotics and Scuolab srl), in four Italian schools (junior high and high school level). The project includes the collaboration with the Projects of Intelligent Robotics and Advanced Cognitive System (PRISCA) Lab of the University of Naples Federico II for the development of the software infrastructure and the scientific support of the Department of Social Sciences (DiSS) of the same university. The idea behind CMR is to use AI, by introducing a social robot archetype, to bring upon the Italian school framework innovative teaching and learning processes. The experimentation consists in testing how a newly developed AI device for social education is received in a classroom environment. A post-experimental investigation was carried out to evaluate the performance of the CMR.

The work aims to develop an easy evaluation tool for the CMR that decision-makers can adopt. The work first implements an initial exploratory study of survey data and then investigates different dimensions that affect the students' evaluation analyzing how these dimensions impact this evaluation. The dimensions concern aspects relating to students' general perception of the CMR, their comfort level using the CMR, their perception of the CMR's impact on school results, and perception of platform likability. Structural equation modeling, and in particular Partial least squares - path modeling (PLS-PM), is used to examine the relationships between these dimensions. According to PLS-PM, student satisfaction may be defined as a multidimensional latent variable (LV) related to its manifest variables (MVs) and linked to other LVs, that represent the variables dimensions. The goal is to determine which aspects of this product need to be altered to boost student satisfaction.

2. Theoretical framework

PLS-PM is a framework for analyzing multiple relationships between blocks of observed and latent variables (Wold, 1974). This approach consists of two elements: the measurement model (also known as the outer model), which describes the relationships between each construct (LV) and its associated observed variables, also called indicators or items (MVs); and the structural model (also known as the inner model), which describes the causal-predictive relationships between the constructs.

According to Lauro et al. (2018) *the PLS-PM approach to SEM consists of an iterative algorithm that compute the estimation of the LVs and the relationships between them by means of an interdependent system of equations based on multiple and simple regression*. The aim is to determine the scores of the LVs through a process, that, iteratively, computes first an outer and

then an inner estimation (Lauro et al., 2018), for making decisions and predictions.

The properties of this algorithm favor its application even when the number of observations is smaller than the number of the MVs. In recent years, the number of publications on PLS-PM has increased significantly. For a review of the PLS-PM approach and its advantages see Hair et al. (2021); Hair et al. (2019); Lauro et al. (2018); Latan and Noonan (2017); Esposito Vinzi et al. (2010).

One of the most important outputs of PLS-PM is the IPMA (Importance-Performance Matrix). It provides information on the relative importance of the dimensions in explaining other dimensions for conclusions. In this context, the IPMA can be considered as a valuable decision making tool (Ringle and Sarstedt, 2016) to identify the drivers to increase the student's satisfaction. It is based on a very intuitive scatter plot where each dimension is positioned according to its importance and performance with concerning the target construct: the vertical axis represents the performance of the attributes from poor performance to good performance, while the horizontal axis represents the perceived importance of the attributes from not very important to very important. In this kind of analysis, we have the possibility to analyze the strengths, weaknesses, opportunities, and threats of constructs, that are considered in the model in order to estimate a latent concept.

3. The model

The blocks relating to students' general perception of the CMR, their comfort level using the CMR, their perception of the CMR's impact on school results, and perception of platform likability are considered dimensions of the PLS-PM model. In this model, student satisfaction is conceived as a latent dimension linked to three dimensions of questionnaire (Figure 1).

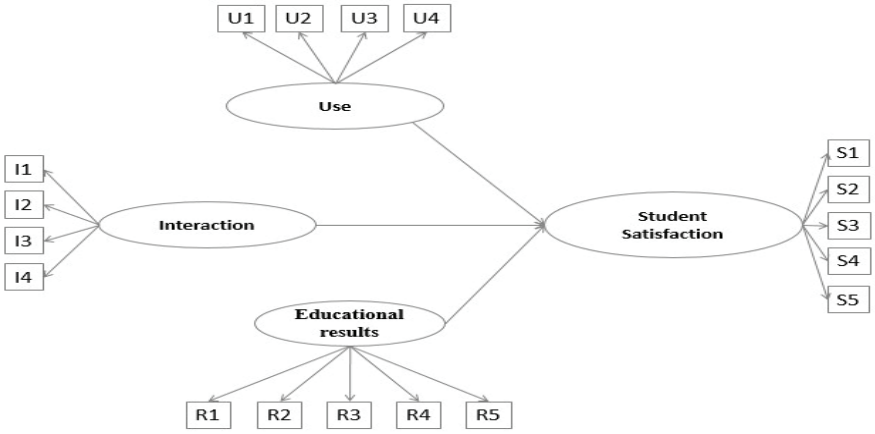


Figure 1: Student satisfaction model

The dimensions of interest were defined first from a qualitative perspective and then translated into survey items by also taking into consideration literature on satisfaction, perceived usefulness and perceived ease of use.

The study focuses on a reflective measurement model, assuming that each dimension is a common factor that reflects itself in its respective MVs, and formative structural model, in which the latent dimension is obtained as a linear combination of the three corresponding under dimensions. The *cSEM* package in the R programming language (Rademaker and Schuberth, 2021) was used to estimate the model.

4. The project and the data

The project experimental phase took place between October 2022 and May 2023 in four

Italian schools with a compatible profile (focus on technology and robotics) selected by the project coordinators. In detail, 21 students from a class in Carrù (2nd year Junior High), 29 from Dalmine (4th year High School), 19 from Rome (5th year High School) and 27 from Verona (5th year High School) were involved in the project, for a total of 96 individuals. The designated classes, which already actively participated in defining the educational framework and co-projecting software requirements, were provided with one of the few available CMR prototypes to test during school hours. The DiSS was responsible for carrying out a full assessment of this pilot study. Such assessment comprised both qualitative and quantitative tools, including a student questionnaire administered in classrooms by teachers at the end of the year.

Table 1: Survey questions, median and per cent distribution of answers

Dimension		**Definition**	**Median**	**1**	**2**	**3**	**4**
Use	U1	The CMR made lessons more fun	3	2	21	59	18
	U2	The CMR facilitates the study of difficult subjects	2	18	46	31	5
	U3	The CMR makes me feel more involved	2	14	42	35	13
	U4	Thanks to the CMR I can contribute to the definition of educational contents	3	8	22	57	13
Interaction	I1	I feel comfortable using the CMR	3	1	12	68	19
	I2	I am calm before a test or a question where I can use the CMR	3	23	17	44	16
	I3	I think the robot in the classroom can be a fun tool to use	3	2	14	51	32
	I4	I think interacting with CMR is simple	2	24	49	22	5
Educational results	R1	The ClassMate Robot motivates me to study	2	31	37	31	1
	R2	The CMR has increased my interest in the subjects studied	2	24	32	43	1
	R3	My academic results in educational activities where we used the CMR have improved	2	17	45	33	5
	R4	Teachers are happy with the results I have obtained in the teaching activities in which we have used the CMR	3	1	9	64	26
	R5	Parents are happy with the results I have achieved in the educational activities in which we have used the CMR	3	10	23	52	15
Satisfaction	S1	The experience with the CMR helped me feel more confident	2	32	45	20	3
	S2	The experience with the CMR has strengthened my skills in the use of digital technologies	2	14	43	33	10
	S3	The experience with the CMR has strengthened my interest in digital technologies	2	13	44	39	4
	S4	The experience with the CMR has improved my relationship with teachers	2	27	27	44	2
	S5	The experience with the CMR has improved my relationship with classmates	2	12	44	40	4

The questionnaire was developed by a multidisciplinary team of experts with the purpose of measuring the end-user satisfaction and provide insights on how to improve performance. In detail, questions were specifically developed to measure: students' general perception of the use of the CMR (4 items on usability and likability), students' comfort level in interacting with the CMR (4 items), students' perception of CMR impact on school results (5 items), students' overall satisfaction (5 items). All Likert scale items are on a 4-point system where the options were "1 = Not at all", "2 = A little", "3 = Enough", "4 = A lot". A 4-point Likert scale was preferred to discern between positive/negative aspects without a neutral choice. Being an experimental project, the participation of the schools was on a voluntary basis. Only 78 viable questionnaires were

returned to the DiSS by the schools.

Table 1 summarizes the questions, median, and distribution of answers for each item. In general, students are quite satisfied their experiences of CMR because the median is 2 and 3 out of 4 for most answers. If we look at the distribution of answers, we notice that the highest frequency of responses is concentrated around 2 and 3.

5. Evaluation of the model

The evaluation of the PLS-PM results begins with an assessment of the reflective measurement models. Table 2 shows the results and evaluation criteria outcomes.

Table 2: Assessment results of reflective measurement model

		Convergent Validity	Internal consistency reliability	
LVs	**AVE**	**Cronbach Alpha**	**Reliability ρ_A**	**Composite reliability ρ_C**
Threshold	*>0.50*	*0.70-0.90*	*>0.70*	*>0.70*
Use	0.54	0.720	0.824	0.824
Interaction	0.51	0.709	0.762	0.762
Educational results	0.49	0.736	0.824	0.824
Satisfaction	0.53	0.786	0.852	0.852

All four reflective measurement blocks meet the relevant assessment criteria. More specifically, all AVE values are above 0.50, providing support for the measures' convergent validity. The composite reliability ρ_C is clearly above the expected minimum level of 0.70. Moreover, the Cronbach's alpha values range between 0.709 and 0.786, which is acceptable. Finally, all composite reliability ρ_A values meet the 0.70 threshold. These results show that the outer model is well specified and that the LVs are well measured by the MVs, their synthesis being good.

We can proceed with the assessment of the structural model, starting to check the VIF values of all predictor constructs in the model. All VIF values are below the conservative threshold of 3, concluding that collinearity among the predictor constructs is not a critical issue in the structural model.

Analyzing the path coefficient estimates of the structural model (Table 3), it is possible to note that not all dimensions are important in order to estimate student satisfaction.

It mainly depends on educational results (path coefficient of 0.366) and, subsequently, on interaction (path coefficient of 0.253); Use is not relevant with a path coefficient of 0.189 and with not significant effect at the 5% probability of error level.

Table 3: The inner estimation

LVs	Estimate	Std.Error	t-value	Pr(>|t|)	CI 95\% percentile
Use	0.189	0.1184	1.6016	0.1092	[-0.0303; 0.4196]
Interaction	0.253	0.1085	2.3350	0.0195	[0.0558; 0.4764]
Educational results	0.366	0.1127	3.2514	0.0011	[0.1464; 0.5707]

6. Analysis – IPMA

The IPMA was considered to identify which aspects can be considered as drivers to increase the student's satisfaction. It is important to focus attention in the lower right quadrant because there are all important aspects that need to be improved. According to the IPMA matrix (Figure

2), the block of educational results is in the critical area because it represents an essential aspect of student satisfaction but has a low performance. This aspect is a critical driver to increasing the CMR evaluation. If we look the IPMA of educational results block we note that three out five aspects are in the critical area, aspects related to motivation to study, interest in study subjects and better educational results. So the analysis found that the educational results dimension is a key factor in increasing the student's evaluation of CMR but has low performance, suggesting that developers need to focus more on this component to improve CMR evaluation.

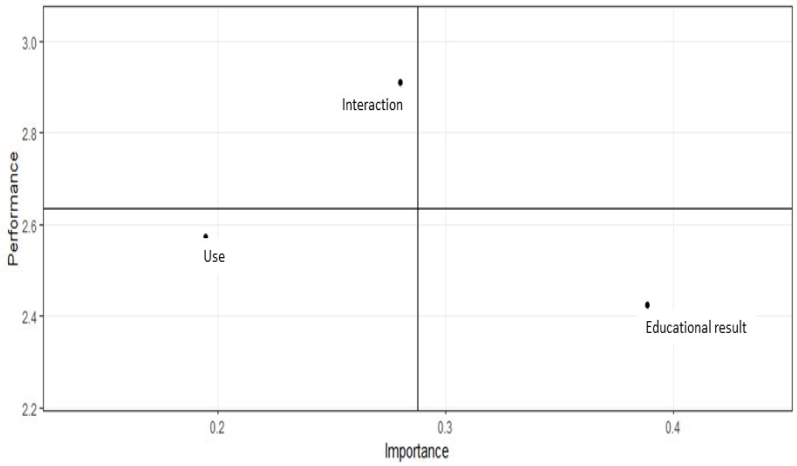


Figure 2: Importance performance analysis on student satisfaction

Figure 3 reports the aspects for each dimension. If we look the IPMA of educational results block we note that three out five aspects are in the critical area, aspects related to motivation to study, interest in study subjects and better educational results. So the analysis found that the educational results dimension is a key factor in increasing the student's evaluation of CMR but has low performance, suggesting that developers need to focus more on this component to improve CMR evaluation.

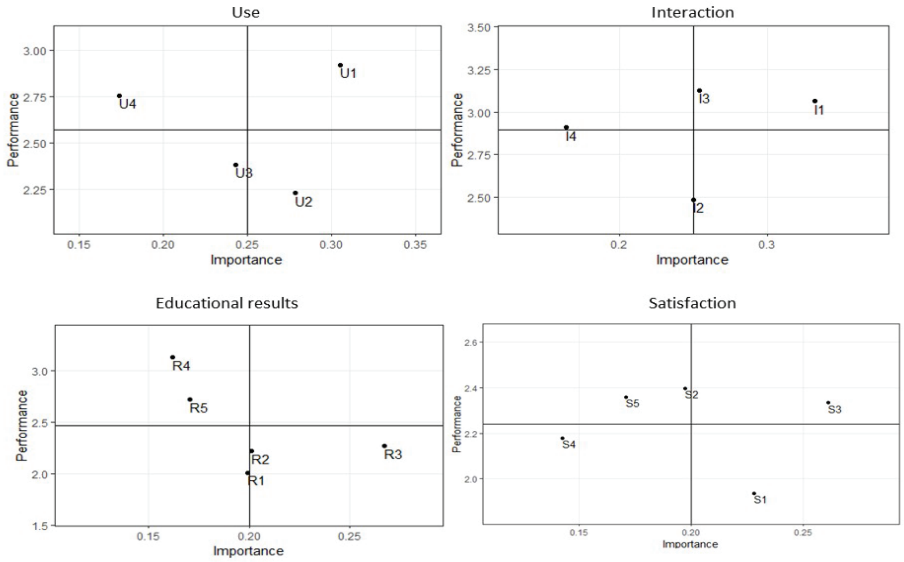


Figure 3: Importance performance analysis of dimensions

7. Remarks

This study, applying PLS-PM and IPMA, investigates the dimensions that affect the students'

evaluation of CMR. The analysis shows that the relevant dimensions are related to between-student interaction and to educational results. The study also found that the educational results dimension is a key factor in increasing the student's evaluation of CMR but has low performance, suggesting that developers need to focus more on this component to improve CMR evaluation.

It may be underlined that the results of this paper should be interpreted in light of the small non-representative sample, a consequence of the cost and complexity of the prototype, which did not allow for mass production. This work should thus be considered as a pilot study which can be improved only after the adoption of CMRs by other schools.

References

Baker, R., Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, **1**, pp. 3–17.

Chin, W. W. (1998). The partial least squares approach to structural equation modeling. *Modern Methods for Business Research*, **295**(2), pp. 295-336.

Esposito Vinzi, V., Chin, W.W., Henseler, J., Wang, H. (2010). *Handbook of Partial Least Squares: Concepts, Methods and Applications*. Heidelberg, Dordrecht, London, New York: Springer.

Hair Jr, J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M. (2021). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM).* Sage Publications

Hair Jr, J.F., Risher, J.J., Sarstedt, M., Ringle, C.M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, Emerald Publishing Limited

Kumar, N. (2019). Implementation of artificial intelligence in imparting education and evaluating student performance. *Journal of Artificial Intelligence*. **1**, pp. 1-9.

Latan, H., & Noonan, R. (Eds.) (2017*). Partial Least Squares Structural Equation Modeling: Basic Concepts, Methodological Issues and Applications*. Berlin/Heidelberg: Springer

Lauro, N.C., Grassia, M.G., Cataldo, R. (2018). Model based composite indicators: New developments in partial least squares-path modeling for the building of different types of composite indicators. *Social Indicators Research*, Springer, **135**(2), pp. 421-55.

Rademaker, M., Schuberth, F. (2021). cSEM: Composite-based SEM.

Ringle, C., Sarstedt, M. (2016). Gain more insight from your PLS-SEM results: The importance-performance map analysis. *Industrial Management & Data Systems*, **116**, pp.1865-1886.

Sánchez, G., Aluja, T. (2006). Pathmox: a PLS-PM segmentation algorithm, *Proceedings of KNEMO*, Citeseer, pp. 69.

Sekeroglu, B., Dimililer, K., Tuncal, K. (2019). Artificial intelligence in education: Application in student performance evaluation. *Dilemas Contemporaneos: Educacion, Politica Y Valores*. **7**.

Simonacci, V., Gallo, M. (2017). Statistical tools for student evaluation of academic educational quality. *Quality & Quantity.* **51** pp. 565-579.

Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y-M., Lauro, C. (2005). PLS path modeling. *Computational Statistics & Data Analysis*, Elsevier, **48**(1), pp. 159-205.

Wold, H. (1974) Causal flows with latent variables: partings of the ways in the light of NIPALS modelling. *European Economic Review*, **5**, pp. 67-86.

doi.org/10.26398/asaproc.0021

The analysis of the informal economy at the province level through cash usage indicators from the Aggregate Anti-Money Laundering Reports

Luciano Cavalli[a], Stefano Iezzi[b], Nazzareno Renzi[b]
[a] Italian National Institute of Statistics – ISTAT, Rome, Italy
[b] Financial Intelligence Unit - UIF, Banca d'Italia, Rome, Italy

1. Introduction

The relation between the use of cash and the underground economy is based on the evidence that irregular transactions or transactions aimed at laundering money deriving from illegal activities are mainly settled in cash to avoid tracing and subjection to tax obligations, or protect them from investigative activities for criminal aspects; this is mainly due to the guarantee of anonymity that characterizes this payment instrument. Despite the consensus on this relationship, empirical evidence is limited, especially for lack of comprehensive data at the sub-national level (Immordino e Russo, 2018; Marmora and Mason, 2021).

In Italy, recent studies have explored this issue in light of changes to cash usage limits. Russo (2022) found that reducing the threshold to €1,000 in 2011 led to a more significant decrease in cash usage among potential tax evaders compared to the general population. Giammatteo et al. (2022) analysed the causal relationship between cash use and undeclared work at the provincial level, showing that a higher ratio of cash transactions was associated with an increase in the share of undeclared value added. Additionally, an increase in the cash usage threshold to €3,000 in 2016 resulted in a slight rise in undeclared value added. Rainone (2023), exploiting high-frequency data on cash transactions by denomination of banknote, showed that the 2016 threshold modification and the access to bank data by the Revenue Agency led to a contraction in cash circulation in Italy, with the latter policy having a more significant impact on regions prone to tax evasion.

The aim of this paper is to contribute to the literature along two lines. On one hand, it has been built a unique dataset, which allows to associate comprehensive measures of underground economy with indicators of the propensity to use cash for the years 2016-2020, both at the province and economic sector level, for the first time in Italy. On the other hand, two sets of indicators for the shadow economy and the cash usage are analysed via canonical correlation analysis for each economic sector separately in order to extract information regarding the correlation structure between the two phenomena. This exploratory analysis does not aim to establish causal relationships between the phenomena, but to reveal the potential interconnection between cash and shadow economy across the sectors.

The statistical relationship between cash usage and the shadow economy holds significance as it could underscore the value of using cash transaction data to improve estimations of the informal economy. Furthermore, this potential connection could strengthen the capabilities of anti-money laundering (AML) authorities in identifying anomalies in cash transactions.

2. The data

Non-Observed Economy (NOE) includes economic activities that escape to direct statistical observation. In Italy, it represents a relevant part of the economy; NOE amounts to about 174 billion euro, accounting for 9.5% of the GDP and 10.5% of the value added generated by the business system in 2020 (ISTAT, 2022). Underground production represents the largest component of NOE – in Italy it accounts for 90% in 2020 (157 billion euro). It is mainly composed

of under-reporting of value added – which is connected with the false declarations aimed at under-reporting production and/or over-reporting costs in order to reduce tax payments on profits (slightly less than 80 billion euro in Italy) – and the value added generated by the employment of un-registered workforce (about 62 billion euro in Italy).

This study focuses on the three main determinants of the underground economy: under-reporting of value added by business units, the value added produced by employing unregistered employees and self-employed individuals. Under-reporting of value added occurs when business units intentionally conceal a portion of their income by providing incorrect declarations of turnover and/or costs. Specifically, under-reporting takes place when one or more components of production are inaccurately reported, or conversely, when certain components are over-reported. In either case, the end result is an underestimation of profits, which corresponds to a underestimation of value added. The estimation of under-reporting of value added also encompasses the amount of VAT fraud without complicity, wherein the buyer pays VAT to the seller, while the latter does not declare it. [1]

The un-registered workforce component of NOE refers to the value added generated through the employment of un-registered labour input. In particular, the component relating to self-employment measures the value added generated by un-registered self-employed, both in their operating surplus and in the mark-up over the remuneration of un-registered employees. The un-registered employee component, on the other hand, measures the value added generated by the employment of un-registered employees measured by their remuneration.

The estimation of these components, for the use proposed here, was carried out with a mixed approach in successive steps. A bottom-up approach was used for the under-reporting component, in which the territorial and sectoral intersections were estimated by aggregation of micro-data referring to business units. A top-down approach was used, instead, for the un-registered workforce component, in which, starting from the domain, the estimates at territorial and sectoral level were obtained by successive breakdowns, consistently with the proposed analysis. This allowed the construction of a dataset by province and economic sector.

For the purposes of this analysis, the three dimensions of the shadow economy are used to construct seven indicators, which differ according to the variable used for normalization, which can be the provincial total value added[2], the total value of underground economy (as sum of the three components), or the gross value added at the province and sector level, which includes only the under-declaration component (Table 1).

Indicators S1, S2 and S5 measure the extent of under-declared value added, respectively, in comparison to the total provincial economic output, in comparison to the gross provincial economic output, including under-declaration, and as a ratio of the total value of the shadow economy. Indicators S3 and S6 measure the contribution of irregular independent work, respectively, to the province's overall economic activity, and to the total shadow economy. Finally, indicators S4 and S7 are similar to S3 and S6 but related to the irregular dependent work.

The source of data we use for cash is the Aggregate Anti-Money Laundering Reports (S.AR.A. is the Italian acronym) which the Financial Intelligence Unit for Italy (UIF) collects from the obliged intermediaries. The Italian anti-money laundering law (Legislative Decree 231/2007) mandates banks and other financial intermediaries to report on a monthly basis to the UIF all

[1] For further information on the methods for estimating the sub-declaration of value added and the calculation of VAT with and without complicity see Sallusti and Cavalli (2019).
[2] The estimate of total value added both at the province and sector level is not available.

transactions amounting to €15,000 or more,[3] after aggregation according to given criteria[4], regardless of any money laundering suspect.

The reports refer to the operations carried out by the customers of the obliged intermediaries and, as the data are aggregated, they are reported in an anonymous format. We have constructed seven indicators of cash usage propensity based on S.AR.A. data (Table 1). The first four indicators, C1 to C4, serve as general metrics for cash usage, as they quantify the total value or number of cash transactions, encompassing both incoming and outgoing flows. The remaining three indicators, on the other hand, are more specific in nature, as they focus on the interplay of cash with other transaction types. Specifically, C5 assesses the portion of total cash value used in conjunction with wire transfers in relation to the total value of wire transfers. Indicator C6 pertains to the role of cash in loan repayment and provision operations, while indicator C7 concerns transaction cash involving the sale or purchase of securities.

Table 1: Description of the indicators

Indicator	Description
	Shadow economy
S1	Ratio between the under-declaration of value added and the total provincial value added
S2	Ratio between the under-declaration of value added and the gross provincial value added which includes only the under-declaration component
S3	Ratio between the value of irregular independent work and the total provincial value added
S4	Ratio between the value of irregular dependent work and the total provincial value added
S5	Share of the under-declaration of value added on the total value of the shadow economy
S6	Share of the value of irregular independent work on the total value of the shadow economy
S7	Share of the value of irregular dependent work on the total value of the shadow economy
	Use of cash
C1	Share of value of incoming cash transactions on total incoming transactions
C2	Share of value of outgoing cash transactions on total outgoing transactions
C3	Share of the number of incoming cash transactions on the total number of incoming transactions
C4	Share of the number of outgoing cash transactions on the total number of outgoing transactions
C5	Share of value of cash (incoming and outgoing) used in combination with wire transfers on the total value of wire transfers
C6	Share of value of cash (incoming and outgoing) used in combination with operations of re-payment or provision of loans on the total value of operations of re-payment or provision of loans
C7	Share of value of cash (incoming and outgoing) used in combination with operations for selling or purchasing securities on the total value of operations for selling or purchasing securities

For our analysis, shadow economy and cash indicators are computed only for the private non-financial sectors. The customer's economic sector, together with the province[5] in which the branch of the intermediary where the operation took place is located, are the two criteria used to associate ISTAT indicators on the underground economy with S.AR.A. indicators on cash.

3. The canonical correlation analysis

This study employs canonical correlation analysis to assess the concurrent relationship between shadow economy and cash usage across the provinces. The analysis is conducted

[3] The presence of a reporting threshold for cash payments does not automatically exclude transactions below that threshold. For instance, a retailer may receive multiple smaller cash payments throughout the day and eventually deposit the total amount at the end of the day or week. If the total deposited amount exceeds the €15,000 threshold, the retailer's bank will record these transactions for anti-money laundering reporting purposes. Financial intermediaries are also required to report data on fractional operations, which are unitary operations with a total amount equal to or higher than €15,000 but conducted through several smaller operations within a seven-day period. Starting from 2021, the reporting threshold has been lowered to 5,000 euros and fractional operations are no longer reported.

[4] Aggregation criteria include the type of transaction, the intermediary's branch where the transaction took place, the client's residence (at municipality level) and economic sector. Each aggregate record includes information on the total amount transacted, the corresponding cash component and the number of individual transactions being aggregated.

[5] Italy comprises a total of 107 provinces; however, the provinces within the region of Sardinia have undergone numerous changes over the past decade. As a result, conducting a consistent analysis across all regions has become challenging. To overcome this issue, we opted to consider all provinces in Sardinia as a single entity, reducing our cross-sectional dimension to 103 provinces.

separately for each economic sector on an annual basis. For brevity, we present the results of the canonical correlation analysis for the year 2019. Notably, the analysis for 2020 is heavily influenced by the pandemic crisis.

Table 2 showcases the 2019 results, featuring eigenvalues, canonical correlations, and the cumulative proportion of explained variability for the first two canonical variables in each sector. Canonical variables beyond the second are disregarded due to their eigenvalues not significantly differing from zero across the analysed economic sectors. The first canonical correlation varies between 0.387 and 0.887, which would appear to be substantially larger than any of the between-set correlations. Strong and significant canonical correlation can be found in most of the economic sectors, especially in producer households and retail trade, capturing a cumulative proportion of explained variability between 0.525 and 0.883. The second and remaining canonical correlations are significant at 1 per cent level only for food industry.

Table 2: First and second canonical variables by economic sector (year 2019)

	First canonical variable			Second canonical variable		
	Eigenvalues	Canonical correlation	Cumulative proportion of explained variability	Eigenvalues	Canonical correlation	Cumulative proportion of explained variability
Mining, energy, petrochemicals, iron and	0.683	0.637 ***	0.566	0.257	0.452 **	0.779
Construction	0.739	0.652 ***	0.566	0.263	0.456 **	0.768
Manufacture of machinery and equipment	0.455	0.559 **	0.593	0.192	0.401	0.842
Food industry	1.310	0.753 ***	0.680	0.391	0.530 ***	0.883
Textile industry	0.371	0.520 **	0.525	0.209	0.416	0.821
Other manifactured products	0.257	0.452	0.592	0.126	0.335	0.882
Wholesale trade services	0.560	0.599 ***	0.636	0.159	0.370	0.816
Retail trade services	1.255	0.746 ***	0.721	0.318	0.491 *	0.904
Accommodation and food services	0.697	0.641 ***	0.568	0.459	0.561 *	0.941
Transportation and storage	0.658	0.630 ***	0.606	0.190	0.400	0.781
Real estate and financial auxiliaries	0.526	0.587 **	0.762	0.104	0.307	0.912
Waste treatment	0.176	0.387	0.534	0.087	0.283	0.798
Health services	0.176	0.387	0.570	0.083	0.277	0.838
Other services	0.904	0.689 ***	0.813	0.098	0.299	0.901
Producer households	3.690	0.887 ***	0.883	0.197	0.406 *	0.931
Others	0.273	0.463	0.736	0.070	0.255	0.923

Asterisks refer to the likelihood ratio chi-squared statistics test that the first/second canonical correlation and all the other are zero; *: significant at 10 per cent; **: significant at 5 per cent; ***: significant at 1 per cent.

The standardized canonical score of the underground economy indicators have been summed up along the three dimensions of the shadow economy: in particular, we compute a cumulative standardized canonical score for all indicators regarding the under-declaration of value added (indicators S1, S2 and S5); one for the irregular independent work (indicators S3 and S6); and one for the irregular dependent work (indicators S4 and S7). The results indicate that all three dimensions play a crucial role in explaining the canonical correlation with the cash use indicators. However, noticeable variations in the impact of underground economy indicators can be observed across different sectors. Specifically, under-declaration of value added holds greater significance for the transportation sector, health services, food industry, and construction. Irregular independent work demonstrates higher relevance for the food industry, textile industry, and construction. Meanwhile, irregular dependent work undeniably exhibits greater relevance for producer households (Figure 1 – panel a).

When considering the impact of cash use indicators, it is evident that deposits hold the highest level of importance across most sectors, except for the manufacture of machinery and equipment, and waste treatment sectors. In particular, for the former, withdrawal indicators appear to be very relevant in explaining the overall correlation between underground economy and cash use. The three specific indicators of anomalous cash use generally exert a lesser influence on the overall correlation. However, there is an exception in the case of cash used in combination with wire

transfers, which appears to be strongly associated with the informal economy in the food and textile industries. Additionally, it is worth noting that the indicator of cash use in combination with operations for selling/purchasing securities has a relatively significant impact on correlation for the construction and transportation sectors (Figure 1 – panel b).

Figure 1: Cumulative standardized canonical scores by economic sector (year 2019)

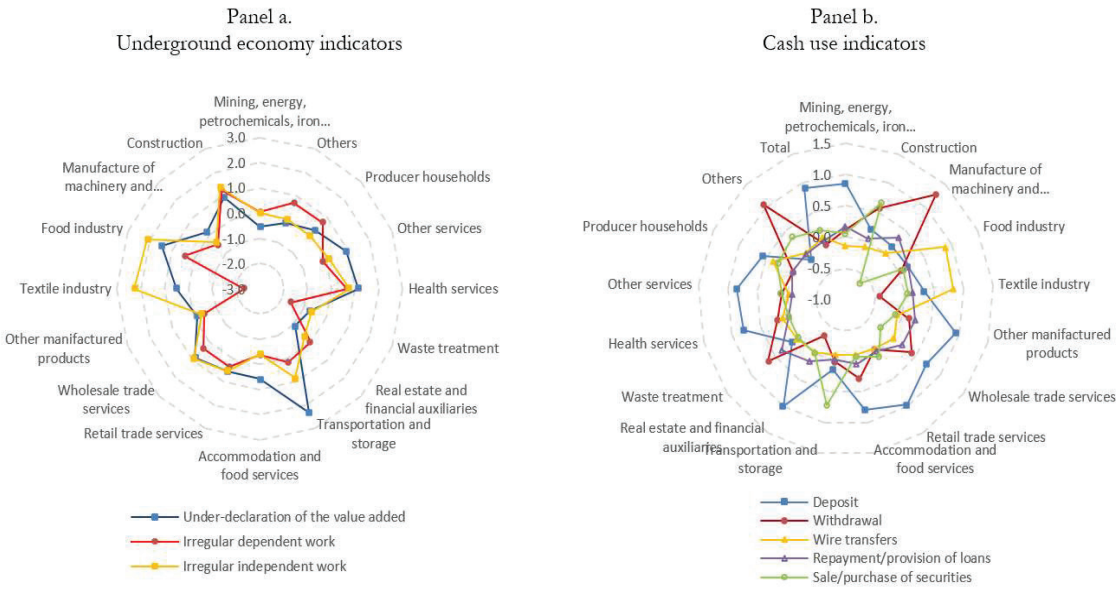


Figure 2: Per cent change of first canonical correlation by economic sector

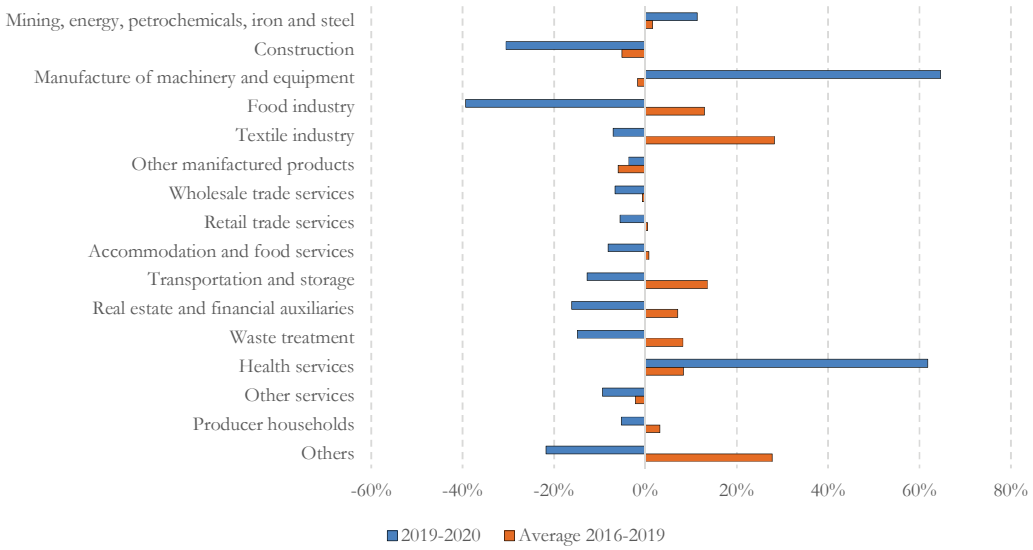


Figure 2 compares the average percentage change in the first canonical correlation between 2016 and 2019 with the percentage change between 2019 and 2020. Canonical correlation shows a general stability in the period 2016-2019, with the exception of the increase recorded in the textile industry. On the contrary, between 2019 and 2020, correlation between underground economy and cash use shows marked negative variations in many sector, presumably attributable

to the detrimental impact of the COVID pandemic on the Italian economy.[6] Note, in particular, the decrease of correlation for food industry and construction by more than 30 per cent, and in transportation, real estate and waste treatment by more than 10 per cent. Note, however, that health sector and the manufacture of machinery and equipment recorded a sharp increase in canonical correlation by more than 60 per cent during the first year of pandemic. The strong increase of correlation in health sector, in particular, seems to be a confirmation of the marked proliferation of cases of financial crimes linked to the health emergency (UIF, 2021).

4. Conclusions

The aim of this work was to build a dataset allowing to associate comprehensive indicators of underground economy with indicators of cash usage for the years 2016-2020, both at the province and economic sector level, for the first time in Italy. The canonical correlation analysis of the two sets of indicators has highlighted the strong statistical association between the two phenomena across various sectors. Notably, this relationship is particularly pronounced in producer households and the retail trade sector.

The purpose of this analysis is twofold. On one hand, since cash data are available only with a three-month delay, the signalling power of indicators of cash might provide valuable information on the underground economy for specific economic sectors ahead of its estimate release. On the other hand, such potential association could enhance the tools employed to highlight anomalies in cash usage for AML purposes.

References

Giammatteo, M., Iezzi, S., Zizza, R. (2022). Pecunia olet. Cash usage and the underground economy. *Journal of Economic Behavior & Organization*, **204**, pp. 107-127.

Immordino, G., Russo, F.F. (2018). Cashless payments and tax evasion. *European Journal of Political Economy*, **55**, pp. 36–43.

ISTAT (2022), L'economia non osservata nei conti nazionali. *Statistiche report*, Rome.

Marmora, P., Mason, B.J. (2021). Does the shadow economy mitigate the effect of cashless payment technology on currency demand? Dynamic panel evidence. *Applied Economics*, **53**(6), pp. 703–718.

Rainone, E. (2023). Tax evasion policies and the demand for cash. *Journal of Macroeconomics*, **76** (https://doi.org/10.1016/j.jmacro.2023.103520).

Russo, F.F. (2022). Cash thresholds, cash expenditure and tax evasion. *Fiscal Studies*, **4**(4), pp. 387-403.

Sallusti, F., Cavalli, L. (2019). Detecting under-reporting of value added and VAT fraud in national accounts, in *7th IMF Statistical Forum 2019*.

UIF (2021). *Rapporto Annuale per il 2020*, Rome.

[6] The economic crisis fueled by the COVID pandemic has determined a decrease in both the total amount of cash operations recorded in S.AR.A. database and the level of unobserved economy (UIF, 2021; ISTAT, 2022).

doi.org/10.26398/asaproc.0022

Dealing with index numbers of ratios. A case study with short-term unit labour cost indicators

Anna Ciammola[a], Eleonora Cimino[a], Donatella Tuzi[a]
[a] Italian National Institute of Statistics, Rome, Italy.

1. Introduction

Since 2003 Istat has been releasing quarterly indices of total labour cost and its components, wages and other costs in terms of full time equivalent (FTE) units. The reference population are all Italian enterprises in the private sectors (excluded agriculture and activities of households and of extra-territorial organisations) with at least one employee. Indices, that are released at section level of Nace Rev.2 economic classification, are produced through the multi-source statistical process *Oros* based on census administrative and survey data (Aquilini et al., 2019).

These indicators entail a system of statistics whose *components* are wages and other costs by Nace section and *aggregates* are: i) total labour cost (C) by Nace section and upper Nace levels and ii) wages (W), other costs (O) and total cost by upper Nace levels. In this system, items i) and ii) represent *horizontal* and *vertical constraints*, respectively.

When variables are reported in terms of FTE, i.e. are expressed as *ratios*, severe consistency problems between components and aggregates may emerge. In particular, while in the horizontal constrain consistency between components and aggregates is guaranteed by construction (Aquilini *et al*., 2019), for the vertical dimension it may happen that ratio values for the aggregates are outside the range delimited by the minimum and maximum of ratio values for components. This consistency issue of ratio variables may get worse when they are reported as index numbers and strictly depends on the formula chosen among those available in literature (Diewert, 2007). Furtherly, inconsistency may be amplified when indices are treated for seasonality or calendar effects, implying time-consuming solutions to restore consistency that are difficult to reconcile with the timeliness required for short-term indicators. Moreover, different formulations may stress different effects and results when dramatic changes affect data as during the Covid-19 outbreak.

Stimulated by the compilation of Oros indices during the pandemic, and focusing on the critical issue of consistency, in this paper we investigate complex number indices of ratios and analyse the relationships between different formulations, scrutinizing the effects of both fixed and time varying weighting systems and assessing their feasibility in the context of indirect seasonal adjustment. Results are reported for both Oros data (highlighting the pandemic span) and simulated time series.

2. The labour cost "non-standard" index numbers: An analytical comparison with more conventional price indices

The Oros unit labour cost indices, according to the practice followed in Istat for analogous indicators, are elementary indices for both components and aggregates (Aquilini *et al*., 2019). Focusing on the total cost, a stylized formulation of this index for a generic sector k at quarter t is:

$$I^k(t) = \frac{\frac{C^k(t)}{U^k(t)}}{\frac{\bar{C}^k(b)}{\bar{U}^k(b)}} = \frac{c_U^k(t)}{\bar{c}_U^k(b)} \quad (1)$$

where b is the reference base year, C^k (U^k) the total labour cost (FTE units), $\bar{C}^k(b) = \frac{1}{4}\sum_{t_b=1}^{4} C^k(t_b)$ ($\bar{U}^k(b) = \frac{1}{4}\sum_{t_b=1}^{4} U^k(t_b)$) the average quarterly cost (FTE units) at b.

According to this approach, the index for a generic aggregate A can be expressed as:

$$I^{A,N}(t) = \frac{\frac{\sum_{k\in A} c^k(t)}{\sum_{k\in A} U^k(t)}}{\frac{\sum_{k\in A} \bar{c}^k(b)}{\sum_{k\in A} \bar{U}^k(b)}} = \frac{\frac{\sum_{k\in A} c^k(t)}{\sum_{k\in A} \bar{c}^k(b)}}{\frac{\sum_{k\in A} U^k(t)}{\sum_{k\in A} \bar{U}^k(b)}} = \frac{\sum_{k\in A} I_C^k(t)\times\frac{\bar{c}^k(b)}{\sum_{k\in A}\bar{c}^k(b)}}{\sum_{k\in A} I_U^k(t)\times\frac{\bar{U}^k(b)}{\sum_{k\in A}\bar{U}^k(b)}} = \frac{\sum_{k\in A} I_C^k(t)\times\bar{\omega}_C^k(b)}{\sum_{k\in A} I_U^k(t)\times\bar{\pi}_U^k(b)} = \frac{I_C^A(t)}{I_U^A(t)} \quad (2)$$

where: $\sum_{k\in A}\bar{\omega}_C^k(b) = \sum_{k\in A}\bar{\pi}_U^k(b) = 1$.

In equation (2) both indices $I_c^A(t)$ and $I_U^A(t)$ can be derived as weighted averages (i.e., as convex linear combinations), the former representing the index of the value $C(t)$ and the latter the index of the "quantity" $U(t)$. In this relation, the weighting systems $\bar{\omega}_C^k(b)$ e $\bar{\pi}_U^k(b)$ do not guarantee the consistency among the index of the aggregate $I^{A,N}(t)$ and the indices of its components $I^k(t)$. It turns out that the current index expressed by (2) is a *non-standard* aggregated index that cannot be derived through a weighted average of indices.

In order to show the differences with more conventional price indices like Laspeyres and Paasche, we consider alternative formalizations of the index $I^{A,N}(t)$. Firstly, let $I^{A,L}(t)$ be the index of the aggregate *A* calculated through the *Laspeyres* formula:

$$I^{A,L}(t) = \frac{\sum_{k\in A} I^k(t)\times\bar{c}^k(b)}{\sum_{k\in A}\bar{c}^k(b)} = \sum_{k\in A} I^k(t)\times\bar{\omega}_C^k(b) \quad (3)$$

where $\bar{\omega}_C^k(b)$ represents the sector k share in terms of the total labour costs of the aggregate *A* in the base year. We are interested in finding out the relation between the non-standard index (2) and the Laspeyres index (3):

$$I^{A,N}(t) = \frac{\frac{\sum_{k\in A} c^k(t)}{\sum_{k\in A} U^k(t)}}{\frac{\sum_{k\in A} \bar{c}^k(b)}{\sum_{k\in A} \bar{U}^k(b)}} = \frac{\sum_{k\in A}\frac{\frac{c^k(t)}{U^k(t)}\bar{c}^k(b)\quad U^k(t)}{\bar{c}^k(b)\ \bar{U}^k(b)\sum_{k\in A}U^k(t)}}{\frac{\sum_{k\in A}\bar{c}^k(b)}{\sum_{k\in A}\bar{U}^k(b)}} = \sum_{k\in A}\frac{c_U^k(t)}{\bar{c}_U^k(b)}\frac{\frac{\bar{c}^k(b)}{\sum_{k\in A}\bar{c}^k(b)}}{\frac{\bar{U}^k(b)}{\sum_{k\in A}\bar{U}^k(b)}}\frac{U^k(t)}{\sum_{k\in A}U^k(t)}$$

that can be simplified as: $I^{A,N}(t) = \sum_{k\in A} I^k(t)\times\bar{\omega}_C^k(b)\times\frac{\pi_U^k(t)}{\bar{\pi}_U^k(b)}$ (4)

Relations (3) and (4) show that $I^{A,N}(t)$ e $I^{A,L}(t)$ coincide when $\frac{\pi_U^k(t)}{\bar{\pi}_U^k(b)} = 1$, that is when the sectorial quarterly share of FTE units does not change with respect to the reference base.

Secondly, using a *Paasche* index for the calculation of aggregate A, we have:

$$I^{A,P}(t) = \frac{\sum_{k\in A} c_U^k(t)\times U^k(t)}{\sum_{k\in A}\bar{c}_U^k(b)\times U^k(t)} = \frac{\sum_{k\in A}\frac{c_U^k(t)}{\bar{c}_U^k(b)}\times\bar{c}_U^k(b)\times U^k(t)}{\sum_{k\in A}\bar{c}_U^k(b)\times U^k(t)} = \sum_{k\in A} I^k(t)\times\tilde{\omega}_C^k(t) \quad (5)$$

where $\tilde{\omega}_C^k(t)$ represents the sector k share in terms of a *pseudo total labour cost* of the aggregate calculated as the product between the cost per FTE in the base year and the FTE in the *current quarter*. Similarly, we are interested in the relation between the non-standard index (4) and the Paasche index (5):

$$I^{A,N}(t) = \frac{\frac{\sum_{k\in A} c^k(t)}{\sum_{k\in A} U^k(t)}}{\frac{\sum_{k\in A} \bar{c}^k(b)}{\sum_{k\in A} \bar{U}^k(b)}} = \frac{\sum_{k\in A}\left(\frac{c_U^k(t)}{\bar{c}_U^k(b)}\bar{c}_U^k(b)\frac{U^k(t)}{\sum_{k\in A}U^k(t)}\right)}{\frac{\sum_{k\in A}\bar{c}^k(b)}{\sum_{k\in A}\bar{U}^k(b)}} = \sum_{k\in A}\frac{c_U^k(t)}{\bar{c}_U^k(b)}\frac{\bar{c}_U^k(b)U^k(t)}{\sum_{k\in A}\bar{c}_U^k(b)U^k(t)}\frac{\sum_{k\in A}\bar{c}_U^k(b)U^k(t)}{\frac{\sum_{k\in A}\bar{c}^k(b)}{\sum_{k\in A}\bar{U}^k(b)}\sum_{k\in A}U^k(t)}$$

that can be simplified as: $I^{A,N}(t) = \sum_{k\in A} I^k(t)\times\tilde{\omega}_C^k(t)\times\frac{\sum_{k\in A}\bar{c}_U^k(b)U^k(t)}{\frac{\sum_{k\in A}\bar{c}^k(b)}{\sum_{k\in A}\bar{U}^k(b)}\sum_{k\in A}U^k(t)}$ (6)

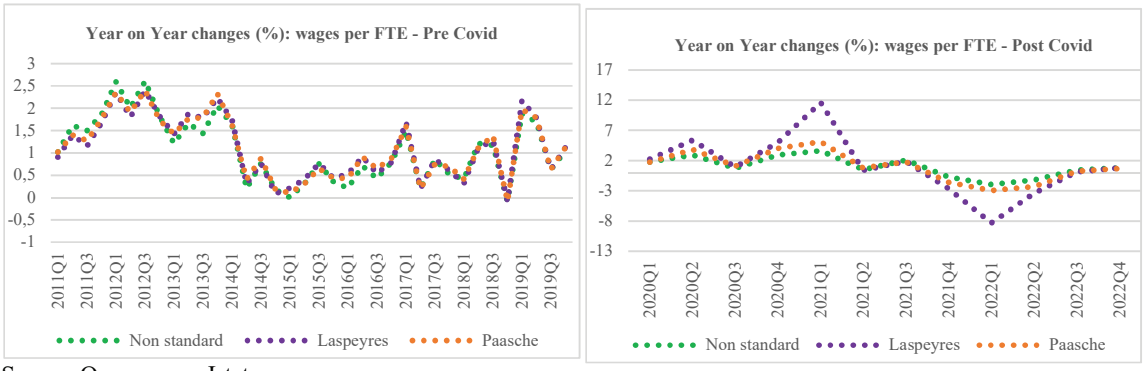
Relations (5) and (6) show that $I^{A,N}(t)$ e $I^{A,P}(t)$ coincide when $\frac{\sum_{k\in A}\bar{c}_U^k(b)U^k(t)}{\frac{\sum_{k\in A}\bar{c}^k(b)}{\sum_{k\in A}\bar{U}^k(b)}\sum_{k\in A}U^k(t)} = 1$, i.e., when $\sum_{k\in A}\bar{c}_U^k(b)U^k(t) = \frac{\sum_{k\in A}\bar{c}^k(b)}{\sum_{k\in A}\bar{U}^k(b)}\sum_{k\in A}U^k(t)$, implying $\bar{U}^k(b) = U^k(t)$, that is when FTE in the

current quarter does not change with respect to the base year in all sectors of A.

The performances of the three indices are exemplified with FTE wages in the aggregate P to S (social and personnel services) as example. Despite the generalized increasing employment in the analysed period (Q1:2010-Q4:2022) this aggregate suffered, more than others, the pandemic outbreak that caused a sharply decreasing importance of some of its components. Furthermore, in this aggregate, the sectorial average levels of wages per FTE (and other costs) differ largely from one another. As a result, the three indices produce significant differences mainly in the quarters following the pandemic outbreak. Both alternatives to the non-standard imply generally higher index values (67% of cases) and the non-standard formula is generally much more adherent to the Paasche. Year-on-Year (Y-on-Y) growth rates emphasize such differences. The left panel of Figure 1 refers to the sub span Q1:2011-Q4:2019 when Y-on-Y growth rates on Laspeyres and Paasche indices are smoother than those calculated on non-standard index.

The right panel refers to the sub span Q1:2020-Q4:2022, marked by the pandemic outbreak: Laspeyres greatly amplifies Y-on-Y growth rates in comparison to Paasche and non-standard indices in at least three quarters (Q2:2020, Q1:2021 and, as a bounce effect, Q1:2022). This is due to how changes in the composition of labour input between the components affect wages per FTE variations of the total aggregate. In Q1:2021 the Y-on-Y growth rate (11.6%) recorded by the Laspeyres, that fixes this composition at the base year, is more than twice as large as the other two indices whose weights depend on FTE at t (5.9% Paasche and 3.6% non-standard).

Figure 1: Non-standard, Laspeyres and Paasche indices for wages per FTE - Aggregate P to S. Y-on-Y growth rates (%). Q1:2010-Q4:2022.



Source: Oros survey, Istat.

Even if unit wages are more responsive that the other costs to an employment dynamic that implies "quality effects" (such as composition by qualification, working time, etc.), the main evidence just described for the aggregate P to S, is confirmed by almost all the other aggregates and variables, particularly in the services sector, and may be synthetized as follows: i) in quite all the aggregates, for more than 50% of the observations, the rates of growth of wages are downsized by the non-standard formula due to a downturn in FTE and a change in its sectoral composition from the base year; ii) the more stationary the employment, the smaller differences among the three indices as shown by the industrial sector excluding constructions (where there is an extensive presence of larger enterprises and government measures are usually put in place to contrast the risk of economic crises) in contrast to the business services sector (where more volatility in employment is recorded).

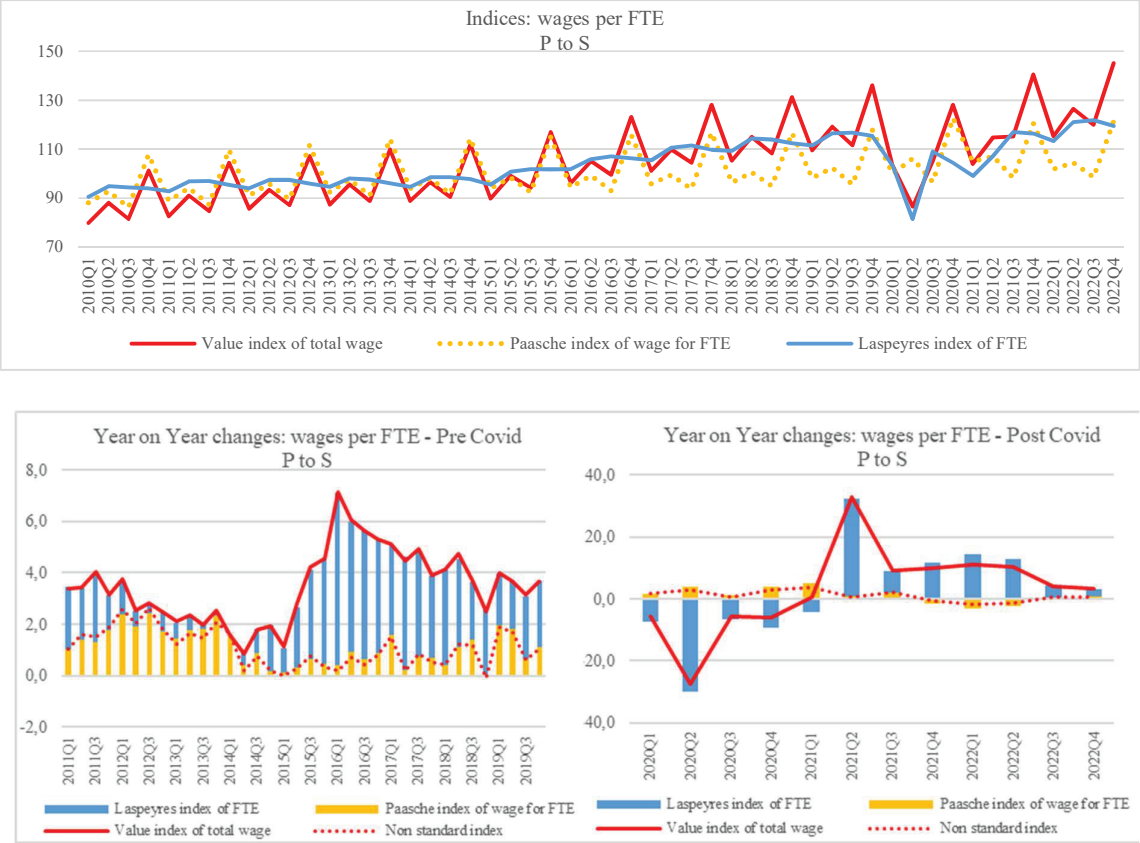
3. From the current system of indices to a more complete framework

A more structured alternative to the single solutions analysed above could be to implement a *system of indices*. It is based on three kinds of indices: i) "value" indices for each variable $v \in \{W, O, C\}$, easily obtained as no weighting structure is needed; ii) "price" Paasche indices for each unit labour cost variable $v_U \in \{w_U, o_U, c_U\}$; and iii) "quantity" Laspeyres indices for FTE, derived

as ratio between i) and ii).[1]

The framework proposed is once again applied on wages per FTE of aggregate P to S (Figure 2). From Q4:2011 to the beginning of 2014 the contribution of wages per FTE (i.e., the price component of the value index) is the dominating effect, being the increase of labour input in a period of strong deceleration[2]. In the period post crisis, starting from Q4:2014, when the increase of wages per FTE becomes lower, total wages increases largely supported by a significant rise of FTE, especially during the years 2016-2018 affected by the active policies of those years, aimed at increasing and stabilizing employment through social contributions' relief for employers.

Figure 2: Value index, Paasche prices index and Laspeyres quantities index for wages per FTE in the aggregate P to S. Indices and year on year growth rates. Q1:2010-Q4:2022.



Source: Oros survey, Istat.

One of the most interesting advantages coming from such structured system of "value" and "price" indices (Paasche or Laspeyres) implies a relevant simplification and a quality improvement on the seasonal and calendar adjustment process, giving the opportunity to adopt the indirect approach. Nevertheless, a drawback is that the weighting structure underlying the Paasche indices (for prices or for quantities) shows a seasonal pattern. In the seasonal adjustment process (SA), this

[1] An alternative could also consist in the implementation of "value" indices for each variable and "price" Laspeyres indices for each unit labour cost variable. As in the previous version, an implication is that the ratio between a value index and a Laspeyres prices index is a Paasche quantities index (FTE), whose weighing structure is given by a pseudo total value of variable *v* where prices are referred to *t* and quantities to *b*. Furthermore, it is worth noting that, in the case of Oros indicators, the use of the Fisher index formula (as done for instance by the Istat survey on Foreign Trade), verifying the property of the reversibility of factors, is not feasible due to the complexity of the weighting system required for implementing the indirect approach in the seasonal adjustment.

[2] This evidence is clearer in the market sectors of both industry and services, whose results are not reported in this paper, where behind high Y-on-Y changes of wages per FTE there are negative Y-on-Y changes of FTE in both the two period of crisis of 2009 and 2010.

seasonality may have implications in terms of residual seasonality on the seasonally adjusted (SA) aggregated series obtained indirectly, i.e. aggregating seasonally adjusted components. This issue is dealt with in the next section.

4. Evidence from a simulation exercise and conclusions

Our proposal, based on an integration of value and Paasche price indices, represents an interesting novelty in the spectrum of the official systems of indices released by Istat. In fact, the survey on external trade apart, short-term indicators computed as number indices are all achieved applying the Laspeyres formula involving weights referred either to the previous year (for chained indices) or to the last base year (for fixed-base indices). This is not motivated by theoretical considerations but by practical reasons since weights are usually available with a certain delay. Being built on a timely census data source, Oros framework easily overcomes the above issue and, therefore, the choice between Laspeyres indices and Paasche indices as feasible alternatives to the current non-standard indices depends on other factors, in particular the ability to represent the underlying phenomenon and the size of revisions generated by new formulas, both extensively investigated in the previous sections.

Although theoretical considerations and empirical evidence point to the Paasche index as a workable alternative for unitary labour cost, Istat has no experience in the management of weights that change monthly/quarterly or in the indirect approach to seasonally adjust Paasche aggregated indices. To cope with this lack of experience, an experiment was carried out adapting the semiparametric bootstrap data generation process described in Davidson and MacKinnon (2006) to Oros time series by NACE sectors (namely FTE units, wages, and other costs) and to ARIMA models (Box, Jenkins and Reinsel, 2008). The preference given to such models among the large class of dynamic linear models relies on the fact that they underlie the seasonal adjustment method implemented in Tramo-Seats (Gomez and Maravall, 2000), the software utilized to produce official seasonally adjusted data by Istat. This section is aimed at detailing the steps of our experiment, reporting the main results achieved for Nace sections B, C, D and E and their aggregation B to E (Industry excluding construction).

The main feature characterizing the bootstrap for time series is the time dependence: data are not IID and the resample procedure based on drawing each observation of a bootstrap sample randomly, with replacement, from the original sample cannot be applied. In the experiment here proposed, the residual bootstrap is used to overcome the problem and residuals are derived estimating ARIMA models on the available time series (Efron and Tibshirani,1994).

The bootstrap experiment is carried out through the following steps: i) automatic pre-treatment or linearization (to remove outliers and calendar effects), identification and estimation of ARIMA models on Oros time series (FTE units, wages and other costs) with Tramo-Seats implemented in JDemetra+ 2.2.3 (time series span the quarters from Q1:2000 to Q4:2019 in order to avoid the observations of the pandemic and consequently possible problems in the automatic model identification); ii) storage of ARIMA models, linearized time series, logarithmic transformation (if applied), model ARIMA orders and parameter estimates, and residuals; iii) resampling residuals with replacement, generating 200 replications per series (although usual for time series, resampling without replacement is not considered here because not suitable to deal with missing residuals at the beginning of time series); iv) recursive generation of bootstrap time series utilizing the previous ARIMA models, initialized with fixed starting values given by the linearized data (given the non-stationarity of ARIMA models, this is the only appropriate procedure as stated in Davidson and MacKinnon, 2006).

It should be noted that the above bootstrap time series are not used for inference, particularly confidence intervals and hypothesis testing, but to generate a huge number of elementary and aggregated time series of index numbers, to find out a confirmation of results derived from real data, and to explore the performance of the indirect SA when Laspeyres and Paasche weighting systems are applied to aggregate SA component indices.

The next necessary steps are: i) generation of sectoral total labour cost time series adding up bootstrap times series referred to wages and other costs; ii) generation of aggregated wages, other costs, total labour cost and FTE units referred to aggregate B to E adding up the respective sectoral time series; iii) calculation of elementary indices of wages per FTE unit, other costs and total labour cost referred to NACE sections B to E; iv) application of the three index formulas (Paasche, Laspeyres and non-standard) to derive indices referred to the aggregate B to E.

Summing up: i) 200 bootstrap time series are generated for each sector (B to E) and for each "independent" variable (wages, other costs and FTE units) with a total of 2400 time series, allowing to obtain 800 time series referred to the sectoral total labour cost; ii) through such datasets, 200 time series of indices referred to the aggregate B2E are derived for each of the three unit variables and for each of the three formulas with a total of 1800 time series of indices.

The simulated time series are then utilized to assess consistency among components and aggregates, in both unadjusted (levels and Y-on-Y growth rates) and SA form (levels and quarter-on-quarter (Q-on-Q) growth rates). Moreover, SA Laspeyres and Paasche aggregates are derived through direct and indirect approaches. Table 1 reports the shares of quarters with inconsistencies.

Table 1: Per cent shares of quarters with inconsistencies between SA components and aggregates

Indices	Unadjusted		SA direct approach		SA indirect approach	
	Levels	Y-o-Y	Levels	Q-on-Q	Levels	Q-on-Q
Non-Standard	0.40	0.46	0.56	4.14	-	-
Laspeyres	0	0	0.21	3.08	0	0
Paasche	0	0.36	0.25	3.63	0	0.88

Finally, the aggregates seasonally adjusted with the indirect approach are treated with Tramo-Seats to verify if further seasonality is extracted from them. As far as the Paasche indices are concerned, results are extremely encouraging since seasonality is extracted only from 11 series out of 600 (5 series for Laspeyres indices). Moreover, a careful analysis of such series reveals a weak and negligible seasonal component. Although further studies are needed, utilizing seasonal quarterly weighting systems does not introduce seasonality during the aggregation of SA components and, to conclude, nothing hinder the implementation of the proposal described above.

References

Aquilini E., Ceccato F., Cimino E., Lattanzio M., Pogelli F.R., Tuzi D. (2019). La rilevazione trimestrale Oros su occupazione e costo del lavoro: indicatori e metodologie. *Letture statistiche - Metodi*. Istat, Rome.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C. (2008). *Time Series Analysis*. John Wiley & Sons, Inc.

Davidson, R., MacKinnon, J.G. (2006). Bootstrap methods in econometrics, in K. Patterson and T.C. Mills (eds). *Handbooks of Econometrics: Vol. 1 Econometric Theory*, Palgrave, pp. 812-838.

Diewert W. E. (2007). *Index Numbers*. New Palgrave Dictionary of Economics, Second edition. Palgrave Macmillan.

Efron, B., Tibshirani, R.J. (1994). *An Introduction to the Bootstrap*. Chapman and Hall/CRC.

Gómez, V., Maravall, A. (2000). Programs Tramo and Seats. Instructions for the user, *Banco de España - Servicio de Estudios. Documento de Trabajo,* 9628.

doi.org/10.26398/asaproc.0023

Informal old adult care provision and caregivers' health status in Italy

Elisa Cisotto[a], Giulia Cavrini[a], Alessandra De Rose[b],
[a] Faculty of Education, Free University of Bozen-Bolzano, Bolzano, Italy.
[b] Department of Methods and Models for Economics, Territory and Finance, Sapienza University of Rome, Rome, Italy.

1. Introduction and background

With improvements against mortality, a growing number of individuals in Western societies are surviving to older ages, leading to a higher likelihood of adult children facing the care need of ageing parents or parents-in-law. Additionally, since the beginning of the 2000s, many European countries have emphasized in-home care over institutionalized care (Colombo et al. 2011). This shift results in a growing number of adults facing the responsibilities of providing informal care to elderly and ill family members. Moreover, due to the demographic ageing of the population structure, the working-age people responsible for both economic and care support for those in need will also decrease over time.

To explore the dynamics between adult children and their parents and the potential role of the former as caregivers for the latter, we can draw upon the well-known 'intergenerational solidarity model', initially developed by Bengtson (Bengtson 1975; Bengtson and Roberts 1991). The model offers valuable insights into the mechanisms of support and care exchange within families across different generations. It identifies six dimensions of intergenerational family solidarity, which determine the likelihood of a child supporting a parent (and vice versa): structural, associational, affectual, consensual, functional, and normative. Notably, normative solidarity, characterized by the sense of mutual obligation as a family norm, plays a crucial role in influencing the support provided to older parents. The model's usefulness extends to investigating the relationship between adult children and their parents-in-law as well. Besides, the solidarity model has undergone further developments and has been applied in subsequent research. Notably, both the theory of social breakdown and the ambivalence theory (Luescher and Pillemer 1998) explore the potential negative consequences of extensive involvement of adult children in their parents' lives. These theories provide evidence that, in certain cases, excessive care from children can adversely impact the well-being of their parents. Advancements in the solidarity paradigm now also recognize that conflict or negative emotions can coexist with emotional closeness in relationships (Bengtson et al. 2002; Suitor, Gilligan, and Pillemer 2011).

Based on the current body of literature, providing care is linked to negative health and well-being consequences (Bom et al. 2019; Bremer et al. 2015; Houtven et al. 2019; Pinquart and Sörensen 2003, 2007). Evidence consistently reveal that caregivers experience diminished levels of subjective well-being, psychological well-being, and self-rated overall health (Van Den Berg, Fiebig, and Hall 2014; Bremer et al. 2015; Labbas and Stanfors 2023). As caregiving intensity increases, these adverse effects tend to worsen, encompassing the amount of time devoted to the role (Houtven et al. 2019) and co-residence with the individual receiving care (Kaschowitz and Brandt 2017; Litwin, Stoeckel, and Roll 2014).

Moreover, the perception of stress experienced by an individual in a caregiving role can change depending on the circumstances and the extent of the available social support (Pearlin et al. 1990). In this regard, Italy provides a remarkable context for investigating the evolution of informal elderly care provision. Indeed, the country boasts exceptional demographic characteristics, including remarkable life expectancy, substantially aged population structure, and resilient family ties (Billari and Tomassini 2021). Furthermore, Italy lags other EU countries,

exhibiting a shortage of public care facilities for older adults. This scarcity places strain on family relationships when elderly parents require care, underscoring the enduring significance of families as primary sources of support for adult individuals (Kalmijn and Saraceno 2008; Tomassini et al., 2020). Moreover, in Italy, the provision of care for older parents/parents-in-law remains a common and normative experience for adult children and traditional gender norms still structure the household division of labour, with women doing more unpaid work than men (ISTAT 2018; Saraceno and Keck 2011; Saraceno and Keck 2010; Zannella and De Rose 2019).

In this context, the primary objective of this study is to assess the relationship between informal caregiving and adverse health outcomes among caregivers in Italy, utilizing data from the 2016 national representative Survey on families, social subjects, and the life cycle (FSS). Our specific research objectives are as follows: (1) to determine the prevalence of elderly care responsibilities within Italy; (2) to examine the correlation between providing informal care and caregivers' self-reported health (SRH) status; (3) to ascertain whether this correlation is influenced by factors such as gender and the extent of caregiving.

2. Data and method

We use data from the FSS carried out by the Italian National Institute of Statistics in 2016 to analyse the characteristics of Italian individuals aged 35-74, who are potentially providers of care to their older parents/parents-in-law. The sample counts 11,211 respondents (51% men) having at least one parents/parent-in-law alive and non-missing information on self-reported health status. After conducting descriptive statistics, multiple logistic regression models are applied to study the association between intergenerational care and caregivers' health status. The outcome indicator selected is the good perceived health as declared by the respondent answering 'good' or 'very good' to the question 'How is your health in general?' (With the following possible answers: very good, good, fair, poor and very poor).

We define informal caregivers as unpaid individuals who provide care for family members lacking self-sufficiency due to factors such as old age, chronic illness, or disability. To identify these caregivers, we employ a comprehensive approach that revolves around analysing responses to a series of questions from the survey. Firstly, participants are asked about their involvement in providing care or assistance to individuals living with them who face challenges due to aging, chronic illnesses, or infirmities on a weekly basis. Subsequently, data from household members is utilized to precisely identify respondents who not only engage in such caregiving but also share a residence with a parent or parent-in-law who has reported limitations in their daily activities, be they severe or non-severe. In cases where a respondent shares a residence with multiple household members facing activity limitations, a binary variable is created to consider the potential dual caregiving responsibility.

Secondly, the FSS asks participants if they have provided no-cost aids to individuals outside their household in the past 4 weeks. Based on this query, respondents who predominantly assist a parent or parent-in-law with activities such as healthcare provision, adult support, domestic tasks, companionship, and administrative tasks are classified as caregivers. Individuals who are caring for both a co-residing parent/parent-in-law and a non-co-residing one are also categorized as engaging in 'double caring'.

Finally, to provide a proxy measure of caregiving involvement, we consider time spent caregiving using a categorical variable: under 10 hours, 10-20 hours, and over 20 hours.

Gender acts as a foundational factor in caregiving dynamics. To address this, we incorporate it as a binary control variable within the model. Additionally, we utilize gender as a stratifying variable for separate analyses. As for respondents' characteristics we control for a set of variables measuring age, partnership (married or cohabiting), having children and at least one child under 15 in the household, area (North, Centre, South and Islands) and degree of urbanisation or the municipality of residence (high, middle and low). As indicator of socio-economic condition, we account for education level (low, middle and high), employment status (employed, retired and

other) and home ownership (owner, renter, other). Further, we control for respondents' functional limitations (severe, light or no limitations) and chronic conditions.

Finally, due to the descriptive nature of this article, normalized weights based on the population's marginal distribution coefficients provided by ISTAT are utilized for all analyses and descriptive tables. All the analyses are performed using Stata 16.

3. Preliminary results and discussion

Table 1 provides an overview of the main variables of interest for both the total sample and the exclusive subset of caregivers, as defined through the analytical framework of the FSS 2016. Considering our focus on respondents with at least one living parent or parent-in-law (potentially engaged in intergenerational caregiving), and ensuring non-missing self-reported health status information, the caregiving proportion among respondents stands at approximately 13%. Among caregivers, over a quarter reside with the individuals they are caring for, approximately 3% are engaged in double care, tending to more than one person, and women outnumber men. Additionally, it's worth noting that among caregivers, around one-third provide high-intensity care (more than 20 hours a week), while the remaining two-thirds offer less intensive care.

Generally, caregivers report being older than non-caregivers, less likely to be partnered, and having a slightly higher proportion without children. Both groups show similar levels of education and employment status, while caregivers display a higher rate of homeownership compared to non-caregivers. Additionally, a greater proportion of caregivers report poorer self-reported health conditions, chronic diseases, and a slightly higher percentage reports limitations in activities of daily living (ADLs) compared to non-caregivers.

Results from the logistic regression confirm that unpaid caregiving for parents/parents-in-law is significantly associated with lower self-reported health status (Table 2, model 1). Specifically, the odds ratio (OR) of 0.795, linked to the risk of caregivers reporting good health, indicates a statistically significant 21% reduction in the likelihood of caregivers reporting good health compared to non-caregivers (the reference group). As expected, when analysing care intensity, the risk of reporting good health undergoes changes: for low-intensity care, the odds ratio is 0.821, representing a 17.9% decrease in the odds, while for high-intensity care, the odds ratio is 0.742, signifying a 25.8% reduction in the likelihood of reporting good health compared to non-caregivers.

To delve deeper into our analysis, we conducted a gender-stratified examination (Table 2, models 2 and 3). The results reveal a notable gender disparity regarding the association between caregiving and good self-rated health. Among women, a discernible negative link is observed, emphasizing the potential impact of caregiving on their health outcomes. Specifically, women demonstrate an odds ratio of 0.748, a statistically significant finding that underscores a substantial 25.2% reduction in the likelihood of reporting good health, compared to women who are not engaged in caregiving responsibilities. Although the outcomes lack statistical significance for men, the consistency in the direction of the association is noteworthy.

Further examination of the models focusing on caregiving intensity unveils a nuanced picture. Among women, it is the realm of high-intensity caregiving that emerges as more detrimental, displaying a statistically significant OR of 0.590. This finding translates to a substantial 41% reduction in the likelihood of reporting good health. In contrast, among men, a significant OR of 0.730 is observed for low-intensity care, signifying a 27% reduction in the likelihood of reporting good health.

These distinct patterns of caregiving impact underscore the imperative for in-depth exploration of differences in care intensity within the realm of caregiving. Moreover, these results underscore the necessity for heightened attention to the unique challenges faced by women, who predominantly assume primary caregiving roles. Addressing these gender-specific dynamics is crucial for ensuring effective support for caregivers and promoting better health outcomes.

Table 1. Sample characteristics. Weighted proportions (%) of caregiving and non-caregiving to parent(s)/parent(s)-in-law in Italy, 2016.

	Total sample (n=11,211)	Non-caregivers (n= 9,661)	Caregivers (n=1,550)
Share providing any care	13.4		
Of which			
Low intensity care			66.2
High intensity care			33.8
Coreside with the carer			25.0
Double care			3.1
Good or very good SRH	76.6	77.3	71.7
Women	48.8	47.9	54.9
Age group:			
35-44	37.4	39.1	26.2
45-54	37.6	36.7	43.6
55-64	19.6	18.7	24.9
65-74	5.5	5.5	5.3
Family status:			
Partnered	76.0	76.4	73.5
No child	22.9	22.5	25.5
No child under 15	40.5	39.3	48.1
Child under 15	36.6	38.2	26.4
Employment status:			
Employed	63.9	64.1	62.4
Retired	26.4	26.5	25.7
Not in employment	9.7	9.4	11.9
Education level:			
High	18.4	18.2	19.8
Medium	41.0	40.9	41.4
Low	40.6	40.9	38.8
Dwelling tenure status:			
Rent or sublet	18.7	19.9	10.9
Owner	70.4	69.0	79.5
Other	10.9	11.1	9.7
Degree of urbanisation:			
High	27.0	27.0	26.9
Middle	43.4	43.8	40.7
Low	29.6	29.2	32.5
Geographical area:			
North	46.8	46.8	46.8
Centre	20.9	20.8	21.6
South & Islands	32.4	32.5	31.6
Functional limitations in daily activities:	8.4	8.3	9.3
Reporting chronic illness:	16.9	16.1	21.6

Source: Families, Social Subjects and life cycle (FSS) ISTAT, 2016, weighted data

Table 2. Logistic regression estimates (ORs) for providing unpaid care to parent(s)/parent(s)-in-law and good SRH in Italy, 2016

	Total sample (mod.1)	Women (mod.2)	Men (mod.3)
No care (ref)			
Any care	0.795**	0.748**	0.840
	(0.072)	(0.087)	(0.120)
By caregiving intensity			
No care (ref.)			
Low intensity	0.821*	0.880	0.730*
	(0.092)	(0.135)	(0.118)
High Intensity	0.742**	0.590***	1.151
	(0.104)	(0.097)	(0.312)

Source: Families, Social Subjects and life cycle (FSS) ISTAT, 2016, weighted data.
Note: All models control for age, partnership status, children and children under 15, education level, employment status, home ownership, area of residence and degree of urbanization, chronic health conditions and ADL limitations. Robust standard errors.
***p<0.01, **p<0.05, *p<0.10

References

Bengtson, V., Giarrusso, R., Mabry, J.B., Silverstein, M. (2002). Solidarity, conflict, and ambivalence: Complementary or competing perspectives on intergenerational relationships?. *Journal of Marriage and Family*, **64**(3), pp. 568-576.

Bengtson, V.L. (1975). Generation and family effects in value socialization. *American Sociological Review*, **40**(3), pp. 358-371.

Bengtson, V.L., Roberts, R.E.L. (1991). Intergenerational solidarity in aging families: An example of formal theory construction. *Source Journal of Marriage and Family*, ***53***(4), pp. 856-870.

Bom, J., Bakx, P., Schut, F., Van Doorslaer, E. (2019). The impact of informal caregiving for older adults on the health of various types of caregivers: A systematic review. *Gerontologist*, **59**(5), pp. e629-e642.

Bremer, P., Cabrera, E., Leino-Kilpi, H., Lethin, C., Saks, K., Sutcliffe, C., Soto, M., Zwakhalen, S.M., Wübker, A., RightTimePlaceCare Consortium (2015). Informal dementia care: Consequences for caregivers' health and health care use in 8 European countries. *Health Policy*, **119**, pp. 1459-1471.

Colombo, F., Llena-Nozal, A., Mercier, J., Tjadens, F. (2011). *Help Wanted?: Providing and Paying for Long-Term Care.* OECD Health Policy Studies, OECD Publishing, Paris (FR).

Houtven, C. Van, Carmichael, F., Jacobs, J., Coyte, P.C. (2019). The economics of informal care. *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press, Oxford, UK.

ISTAT (2018). *Condizioni di vita, reddito e carico fiscale delle famiglie*, www.istat.it.

Kalmijn, M., Saraceno, C. (2008). A comparative perspective on intergenerational support: Responsiveness to parental needs in individualistic and familialistic countries. *European Societies*, **10**(3), pp. 479-508.

Kaschowitz, J., Brandt, M. (2017). Health effects of informal caregiving across Europe: A longitudinal approach. *Social Science and Medicine*, **173**, pp. 72-80.

Labbas, E., Stanfors, M. (2023). Does caring for parents take its toll? Gender differences in caregiving intensity, coresidence, and psychological well-being across Europe. *European Journal of Population*, **39**(18).

Litwin, H., Stoeckel, K. J., Roll, A. (2014). Relationship status and depressive symptoms among older co-resident caregivers. *Aging and Mental Health*, **18**(2), pp. 225-231.

Luescher, K., Pillemer, K. (1998). Intergenerational ambivalence: A new approach to the study of parent-child relations in later life. *Journal of Marriage and Family*, **60**(2), pp. 413-425.

Pearlin, L.I., Mullan, J.T., Semple, S.J., Skaff, M.M. (1990). Caregiving and the stress process: An overview of concepts and their measures. *The Gerontologist*, **30**(5), pp. 583-594.

Pinquart, M., Sörensen, S. (2003). Differences between caregivers and noncaregivers in psychological health and physical health: a meta-analysis. *Psychology and Aging*, **18**(2), pp. 250-267.

Pinquart, M., Sörensen, S. (2007). Correlates of physical health of informal caregivers: A meta-analysis. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, **62**(2), pp. 126-37.

Saraceno, C., Keck, W. (2011). Towards an integrated approach for the analysis of gender equity in policies supporting paid work and care responsibilities. *Demographic Research*, **25**(11), pp. 371-406.

Saraceno, C., Keck, W. (2010). Can we identify intergenerational policy regimes in Europe? *European Societies*, **12**(5), pp. 675-696.

Suitor, J.J., Gilligan, M., Pillemer, K. (2011). Conceptualizing and measuring intergenerational ambivalence in later life. *The Journals of Gerontology Series B, Psychological Sciences and Social Sciences*, **66**(6), pp.769-781.

Tomassini, C., Cisotto, E., Cavrini, G. (2020). You are my favourite! Parent–child relationships and satisfaction in later life in Italy. *Ageing and Society*, **41**(11), pp. 2467-2483

Van Den Berg, B., Fiebig, D. G., Hall, J. (2014). Well-being losses due to care-giving. *Journal of Health Economics*, **35**(1), pp. 123-131.

Zannella, M., De Rose, A. (2019). Stability and change in family time transfers and workload inequality in Italian couples. *Demographic Research*, **40**(3), pp. 49-60.

doi.org/10.26398/asaproc.0024

New perspectives in the integration of data sources in official statistics

Giuliana Coccia[a], Fabio Crescenzi[b]
[a] ASviS, The Italian Alliance for Sustainable Development, Rome, Italy.
[b] Department for Statistical Production, National Institute of Statistics, Rome, Italy.

1. Introduction

The large amount of information contained in administrative and other sources (big data, remote sensing, etc.) offers opportunities to respond to the growing need of data. The integration of multiple sources produces significant gains in terms of efficiency and effectiveness and reduce the burden of surveys. For these reasons in recent years much research has been devoted to data integration.

The Italian National Institute of Statistics (Istat) has set up the Integrated System of Registers (SIR) and adopted a model to produce statistics based on the integration of multiple sources. The SIR is powered by a multiplicity of data, obtained primarily from administrative sources, then from surveys and alternative sources.

When multiple sources are used, the coherence of the produced data acquires great importance in the framework of the quality of statistics. As known the conceptualizations of coherence often remain vague, with proposals for measures not fully specified. In this paper we assume as coherence the attitude of surveys to produce consistent results when different mix of sources are used. When, for various reasons, the results differ, we assume them as incoherence events.

Incoherencies among administrative and other sources may arise (Eurostat, 2003; Krapavickaitė, 2018; Pannekoek, 2017; Thagard et al., 1998; Thagard et al., 2002; Zhang, 2017). Measuring the incoherencies between the administrative sources, big data, and surveys, in some cases leads to differences that are acceptable or physiological; in others require specific interventions.

New frontiers in machine learning and artificial intelligence may provide some help in this context. In this paper, we discuss in general terms the need of a treatment and measurement system to reduce risks of incoherencies for data producers, in view of the expansion of the outputs.

Istat has launched the project "Coherence of survey data and data from administrative sources (CoeReS)" to develop a system for evaluating and treating incoherencies. Otherwise, the increase of the data disseminated exposes official statistics to future risks. National and international scientific institutions have already started this methodological action (Agafiţei et al., 2015; Zhang, 2017).

This paper presents some general considerations and some preliminary results of the CoeReS project.

2. Coherence among sources

Coherence is a fundamental dimension of quality (ABS, 2009; Agafiţei, 2015; Ehling et al., 2007; Eurostat, 2015; HCSO, 2014; Statistics Canada, 2009; UNECE, 2017) and often statistical quality guidelines deal with this dimension and the importance of its measurement. Although it is such a recognized aspect, the conceptualizations of coherence in the framework of the quality of statistics often remain vague, with proposals for measures not fully specified.

Regulation (EC) no. 223/2009 of the European Parliament and of the Council establishes the criteria (dimensions) of statistical quality, and presupposes the need to measure these criteria in

the quality assessments of statistical processes, products, etc. The regulation contains brief definitions of these criteria, although we cannot expect comprehensive conceptual and/or methodological explanations from the regulation. Important guidelines were written, such as the very important Code of Practice (CoP) (Eurostat-European Statistical System, 2011). It consists of fifteen principles, the fourteenth of which concerns coherence. The quality assurance framework of the European Statistical System (QAF) offers further details to each principle and is a "fundamental supporting document aimed at assisting the implementation" of the CoP (European Statistical System, 2015). Overall, there are some still open related conceptual problems that should be more clearly delineated.

The most important characteristics of the 14th principle concern internal coherence (considering the single source) and the coherence over time and over space. Principle 14.4 is dedicated to multi-source consistency. It establishes that "statistics from different sources and of different periodicity are compared and reconciled". However, there is no specification of measures and actions to be carried out for this purpose.

In the start-up phase of Istat CoeReS project, while referring to the coherence between different sources, to that which occurs between data acquired from administrative sources and data acquired through surveys, we will first propose a combined evaluation of indicators of internal coherence of the sources (Daas et al., 2011). The proposed goal is to collect the results of Istat current practices to support the formulation of principles and indications of a more general nature.

3. SIR contents and benefits of integrated information

The SIR integrates more than 26 registers on three main dimensions: economic units, individuals, and places. Additional registers play a bridging function between the three main dimensions as in the case of employment and income.

Economic units are entities of production in the Italian territory. The base registers on economic units have a longer and more consolidated tradition in Istat: There are base registers on agriculture, enterprises, public and private institutions. A series of thematic and extended registers are connected to the economic units, in particular the registers concerning economic values. The other fundamental unit of analysis is that of individuals: The base register on individuals (RBI) uses a greater number of sources and provides information on all individuals who have crossed the Italian territory and provides information on family compositions. Within it, it is possible to identify the subset of resident units, which form the basis of the census.

The objective of the base statistical register of places is to strengthen the ability to geo-reference statistical units (families, individuals, businesses, farms, etc.) and enhance the release of data on the Institute's GIS (Gistat). The geo-referencing of data allows Istat to produce richer statistical and geographical information and to learn more about the structure and composition of Italian territory. The employment register acts as a bridge between the individual dimension and the economic unit dimension. The unit of analysis is the job position, referable to the individual-company relationship, therefore connections remain possible both on the business side (or economic unit), and on the individual side with RBI and the registers connected to it.

Strengths of the new model to enhance statistics are:

- Valorisation of outputs. The large amount of information contained in the SIR offers opportunities to better respond to the growing need for data on complex phenomena (for the purpose of knowledge, research, definition, and evaluation of policies, etc.).
- Spatial and longitudinal data. New opportunities in the production of spatial and longitudinal data and of data from different fields of analysis linked to each other.
- Efficiency and effectiveness. The SIR also provide significant gains in terms of efficiency and effectiveness, with reference to the reduction of the burden due to the surveys.

4. CoeReS project

The objective of the project CoeReS is to build an integration system to monitor the coherence among statistical registers and surveys. It has these specific objectives:

- Increasing the common heritage of knowledge and evaluation of the incoherencies between data acquired from different sources, to create an integration system for the coherence between statistical registers and surveys, outlining intervention strategies. Determining the cases of incoherence and the main factors that determine them.
- Identifying potential actions to reduce burden by replacing variables detected in the field with data from other sources. Propose measures of the quality that would ensure and of the effects of containment of the operational burden.
- Classifying the Istat surveys in a taxonomy defined based on how the sources are involved in the surveys (Conceptual framework and taxonomy of the surveys). We will work from existing classification tools, for example, classifications of the registers, big data (BD) and / or surveys available in the Institute's programs.
- Quality measurement indicators: Looking for measure of incoherence and order them by importance. Identify proxy indicators of incoherencies.
- Construction of an assessment and rating system for the impact of incoherencies: The system is organised based on the taxonomy cases defined in the conceptual framework, defining acceptability thresholds and consistency ratings based on quality indicators.

As first step we focused on 6 case studies concerning the following surveys: 1. Labour market; 2. Income and EUSILC (European Union statistics on income and living conditions); 3. Usually resident population; 4. Agriculture; 5. Disability and morbidity; 6. ICT Survey.

For these case studies we designed a questionnaire to compare quality of data from administrative sources and sample survey sources and this questionnaire was filled in by the thematic experts. We named source *A* the registers or BD, and source *B* the sample or censual surveys and adopted the following classification of variables:

1.1 Variables obtained from sources A and for which there is no auxiliary data from source B.
1.2 Variables obtained from sources A and for which auxiliary data from sources B are available, but which have not been subject to integration.
2.1 Variables obtained from sources B, and for which no has auxiliary data from sources A.
2.2 Variables obtained from sources B and for which auxiliary data is available from sources A, but which have not been subject to integration.
3 Registry variables estimated by integrating source data A and B.

To investigate on coherence, we need variables for which are available data from both the kind of sources A and B. (item 1.2, 2.1, 3). We found that in the case of disability and morbidity and ICT Survey there are no such variables.

5. Main results and next steps

Here the main results of the case studies:

1. Overall assessment of coherence between sources A and B. Only in two cases a high degree of coherence is indicated between sources of type A and B, and these are the cases referring to labour market and usually resident population.
2. Use of indicators or methods for the assessment of coherence. The assessment of the degree of coherence is not currently based on the use of harmonized indicators or common methods. Only ad hoc methods are employed.
3. Reasons of incoherencies. The most frequent reasons of incoherence among data are differences due to the time shifts, measurement errors, differences in definitions and classifications adopted.

Having in mind the specific objectives of the CoeReS project listed in the previous section, we

are going to study the remaining four case studies:

- Labour market,
- Income and EUSILC,
- Usually resident population,
- Agriculture.

6. Challenges for official statistics

More in general, official statistics are facing big challenges (Wallgren, 2007). Here some of the most important challenges:

- Change of survey methods. Survey methods have to be reviewed rapidly and dramatically, from traditional surveys that produce the best possible estimates from an observed sample to surveys that produce the best possible estimates using the integration of all available data obtained from multiple sources. Official and scientific institutions already started the review of methods; however, a more incisive and extensive cooperation is required.
- Classifying and measuring discrepancies. There may be discrepancies between estimates obtained from data from different sources. In some cases, the differences can be considered physiological and acceptable; in others, they may require the adoption of specific intervention tools. Istat is working to classify and measure discrepancies, to assess the risk associated with every estimation strategy, and define possible intervention strategies.
- BD methodological framework. Although some BD sources have already entered statistical production (scanner data, web data) the BD methodological framework has yet to be designed. Different BD sources pose a wide range of integration issues to address, which are different by type of source type of use. An occasion in Istat is the use of web data on agritourisms in the census of agriculture. Web data will be used both to update the list of activities and to collect new variables or confirm information already acquired.

CoeReS project is ongoing and the results that we will obtain within this project will be useful to face these challenges.

References

ABS (2009). *ABS Data Quality Framework*. Australian Bureau of Statistics, http://www.abs.gov.au/ausstats/abs@.nsf/mf/1520.0.

Agafiţei, M., Gras, F., Kloek, W., Reis, F., Vâju, S. (2015). Measuring output quality for multisource statistics in official statistics: Some directions. *Statistical Journal of the IAOS,* **31**, pp. 203-211.

Daas, P., Ossen, S. (2011). *BLUE-ETS: Report on Methods Preferred for the Quality Indicators of Administrative Data Sources (Deliverable 4.2)*. European Commission http://www.pietdaas.nl/beta/pubs/pubs/BLUE-ETS_WP4_Del2.pdf.

Ehling, M., Körner, T. (Eds.) (2007). *Handbook on Data Quality Assessment Methods and Tools*. Wiesbaden: Eurostat.

Eurostat (2003). Assessment of quality in statistics, *in Sixth meeting: Item 4.2: Methodological Documents* ESSnet KOMUSO Work Package 3.

Eurostat (2015). *ESS Handbook for Quality Reports*. 2014 edition. Luxembourg: Publications Office of the European Union.

Eurostat-European Statistical System (2011). *European Statistics Code of Practice.* Eurostat. https://unstats.un.org/unsd/dnss/docs-nqaf/.

European Statistical System (2015). *Quality Assurance Framework of the European Statistical System. European Statistical System*.

HCSO (2014). *Quality Policy of HCSO*, Hungarian Central Statistical Office.

Krapavickaitė, D. (2018). *A Glance to Coherence*. Manuscript, internal document within the project group Komuso: https://cros-legacy.ec.europa.eu/system/files/essnet_sga2_literature_review.pdf.

Pannekoek, J. (2017). *An Approach to Coherence Measures? A Few Additional Remarks*. Manuscript,

internal document within the project group Komuso: https://cros-legacy.ec.europa.eu/system/files/essnet_sga2_literature_review.pdf.

Regulation (EC) No 223/2009 of the European Parliament and of the Council. *Official Journal of the European Union*, **52**(L87), pp. 164-173.

Statistics Canada (2009). *Statistics Canada Quality Guidelines*. Fifth Edition. Ottawa: Statistics Canada.

Thagard, P., Verbeurgt, K. (1998). Coherence as constraint satisfaction. *Cognitive Science*, **22**(1), pp. 1-24.

Thagard, P., Eliasmith, C., Rusnock, P., Shelley, C. (2002). Knowledge and coherence, in *Common Sense, Reasoning, and Rationality*. R. Elio (ed.), Oxford University Press. New York. pp. 104-131.

UNECE (2017). *Quality Indicators for the Generic Statistical Business Process Model (GSBPM) for Statistics Derived from Surveys*. United Nations Economic Commission for Europe.

Wallgren, A., Wallgren, B. (2007). *Register-Based Statistics: Administrative Data for Statistical Purposes*. Chichester: John Wiley & Sons.

Zhang, L.-C. (2017). *An Approach to Coherence Measures?* Manuscript, internal document within the project group Komuso: https://cros-legacy.ec.europa.eu/system/files/essnet_sga2_literature_review.pdf.

doi.org/10.26398/asaproc.0025

Population census: Methodology for the list survey

Fabio Crescenzi[a], Emanuela Scavalli [b]
[a] Department for Statistical Production, Istat, Rome, Italy
[b] Directorate for Social Statistics and Population Census, Istat, Rome, Italy

1. Introduction

Starting from 2018, the Italian National Institute of Statistics (Istat) launched the permanent population and housing census (PPHC), based on the data integration of administrative sources with sample surveys. The first round of the PPHC (2018-2021) was designed starting from the register of the Italian population (RBI), a register obtained integrating multiple administrative sources, among which the local population registers of Italian municipalities (LAC).

In addition to the RBI data, address data from the statistical register of places (RSBL) and data from the education and employment registers were used and two sample surveys were designed and carried out annually (both area and list surveys) to assess coverage errors in the RBI and collect data for variables that are not fully available in the registers (Agafiţei et al., 2015).

In 2020, given that it was not possible to carry out the surveys due to the COVID-19 pandemic, the estimation method was changed using RBI data and "signals of life" (SoL) (Gallo and Zindato, 2021). [1] Taking advantage of this experience, the second round of the PPHC (2022-2026) integrates RBI data and SoL forming an extended population register (Solari et al., 2023).

SoLs are obtained from AIDA (Integrated Database of Usual Residents), which contains integrated data over time on individuals from more than forty administrative sources. SoL profiles are defined aggregating individuals in sub-populations having similar over-coverage attitude. An indicator function based on SoL profiles is defined (Bernardini et al. 2022), according to which any individual in a profile is classified as either included or excluded from the population count.

New techniques have been explored that use survey data to improve SoL profiles (latent class models and other data science methods) in relation to population groups that are not easily identifiable by deterministic criteria alone (Gallo and Zindato, 2021). At the same time, survey data can be used to estimate the bias in population counts and better define SoL profiles.

Two surveys are conducted to support the registers:

- An area survey, which consists of a sample of addresses that includes all households in an address: this survey aims to estimate the possible under-coverage of the population register.
- A list survey carried out on a sample of households from the LAC registers with the aims of estimating the distribution of the census variables of interest (occupation, education, dwelling, etc.) and defining SoL profiles.

The data from administrative sources was integrated with survey data (collected through list or areal sampling) through a statistical model. This paper describes the methodology employed in developing the list survey.

2. The list survey design

The list survey design is a two-stage sampling with stratification of the primary units (municipalities); the secondary units are households. The characteristics of all individuals

[1] Administrative 'signs of life' refer to activities of individuals that can be inferred from administrative records. Being self-employed or working for a company, being a civil servant, having a regular annual rent for a dwelling, attending school or university courses are examples of direct administrative life signs. An indirect life sign is instead an identifier of a status or condition, such as being an income or pension recipient.

belonging to the same family are recorded. In each municipality, the sample households are selected from the RBI list with equal probabilities.

The municipalities (updated until 2021/06/30) are divided into two groups:

i) self-representative (SR) municipalities annually participating in the survey (1,188 units);
ii) non self-representative (NSR) municipalities (6.716 units), participating in the survey only once in five years (approximately 1,343 per year).

The SR municipalities have been defined according four criteria:

1. municipalities with more than 17.800 inhabitants
2. municipalities provincial capital
3. municipalities that are SR in the labour force survey (not rotating)
4. critical municipalities according to parameters derived from the analysis of the surveys carried out in the previous census cycle 2018-2021.

Each SR municipality is considered as a separate stratum and is included in the sample each year.

Table 1. Number of sampled municipalities - period 2022-2026

	SR municipalities	NSR municipalities	TOTAL municipalities
Population ≥17.800	600		600
Provincial capital municipalities	109		109
SR municipalities in the LFS	916		916
Critical municipalities	149		149
TOTAL	**1,188**	**6,716**	**7,904**

Within each province, the NSR municipalities were grouped into strata of five units according to population size. So, we obtained 1,381 strata, 85 of which contained less than five municipalities because they were residual strata in the province. 61 strata have been constructed with border municipalities[2] (including a total of 305 out of 313 border municipalities –see Table 2). The other 8 border municipalities were included in the remaining 1,320 strata. To define the year in which the NSR municipalities had to be surveyed, we excluded the municipalities that participated in the 2021 survey.

Table 2. Number of border municipalities in the Italian provinces

Provinces	Border municipalities
12 - Varese	103
13 - Como	139
97 - Lecco	21
103 - Verbano-Cusio-Ossola	50
Total	313

3. Allocation of the households

The households to be interviewed were assigned to municipalities adopting a mixed strategy based on cost and organisational criteria as well as the sampling errors of the main estimates. The

[2] Municipalities with workers who are employed in a border zone of another State but who return daily, or at least once a week, to the neighbouring State in which they reside and of which they are nationals.

allocation strategy considered the following parameters:

- 100 is the minimum number of households to interview in a municipality;
- municipalities with less than 300 inhabitants are exhaustively surveyed (446 units);
- municipalities with a sampling fraction >90% and with difference between population and sample <40 households are exhaustively surveyed (8 units);
- households of border municipalities are oversampled in proportion to the border workers.

Table 3. Allocation of household sample by year - period 2022-2026

Type of municipalities	Municipalities	Households sample
SR every year	1,188	604,062
NSR year 1	1,343	394,683
NSR year 2	1,343	394,522
NSR year 3	1,343	394,898
NSR year 4	1,345	395,272
NSR year 5	1,342	395,129

Approximately 2,531 municipalities (NSR+SR) and about 1 million households (see Table 4) were surveyed each year.

Interviews were conducted using a mixed mode technique (CAWI, CAPI, CATI), with a first phase of so-called "spontaneous response" and a second phase of field follow-up of non-respondents.

Table 4. Annual sample of Municipalities and Households. period 2022-2026

Year	Municipalities	Households sample
2022	2,531	998,745
2023	2,531	998,584
2024	2,531	998,960
2025	2,533	999,334
2026	2,530	999,191

4. Weighting system

The weights assigned to the sample units are obtained through a complex procedure enabling to correct for possible non-response bias from households that were unavailable or refused to be interviewed and ensure that the sample estimates are consistent with known totals of auxiliary variables.

The distributions of the population by gender, age class and citizenship (Italian *vs.* non-Italian) for each province and 12 main municipalities are known from the RBI. [3]

Indicating with ${}_kX$ (k=1, …, c) the known total of the k-th auxiliary variable for the generic province and with ${}_kX_{hij}$ the value assumed by the k-th auxiliary variable for the respondent unit hij, the above described condition is expressed by the following equation:

$$ {}_k X = {}_k \hat{X} = \sum_{h=1}^{H} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} W_{hij\,k} X_{hij} \quad (k=1, \ldots., c) $$

where H indicates the number of strata. In case a stratum is empty, it is aggregated to the neighbouring one.

[3] Torino, Genova, Milano, Verona, Venezia, Bologna, Firenze, Roma, Napoli, Bari, Palermo, Catania

The final weights for the respondent sampling units are determined as follows:

1. initial weights are obtained as the inverse of the inclusion probability of the units in the stratum;
2. correction factors for total non-responses are worked out as the reciprocal of the response ratio in the municipality to which each unit belongs;
3. basic weights corrected for non-response are then computed multiplying initial weights by these correction factors for total non-response;
4. provincial correction factors are obtained solving a minimisation problem under constraints, which make it possible to satisfy the condition of equality between the known totals of the auxiliary variables and the corresponding sample estimates;
5. final weights are computed multiplying the basis weights by the correction factors obtained at the previous step.

The correction factors of point 4 are obtained by solving a constrained minimum problem (Deville and Särndal, 1992), where the function to be minimised is a distance function (suitably chosen) between the basic weights and the final weights, the constraints regard the estimates of some auxiliary variables that have to be equal to the totals in the reference population derived from the RBI. The chosen distance function is the truncated logarithmic function; the adoption of this function guarantees that the final weights are positive and within a predetermined range of possible values, thus eliminating the extreme positive weights (too large or too small).

All estimation methods resulting from solving a constrained minimum problem of the type described above belong to a general class of calibration estimators.

5. Concluding remarks

The Italian permanent census involves only representative samples each year, unlike the traditional census, which involves all citizens and all households at the same time. In any case, the former returns estimates of the population at municipal level (Carbonetti et al. 2023).

The design of the list survey, as defined in the first round of surveys (2018-2021), was changed in the second round (2022-2026). The annual estimates concerning the municipalities are incomplete since the NSR units are only surveyed in one of the five-year cycle.

For this reason, the annual surveys are used to collect information that, when integrated with data from administrative registers, may help to define profiles that support the enumeration of the population and the estimation of census variables using alternative methods to the traditional ones.

The list survey was used to determine the RBI over-coverage error measure for the 2018 and 2019 population censuses. After 2022, these surveys will continue to play a crucial role in collecting data for variables that cannot (or can partially) be replaced by administrative data and in providing quality measures for the fully register-based population size estimation.

The implementation of more efficient registers is also fundamental because they are a starting point for the other sample surveys carried out by Istat, in order to integrate the information collected in a surveys with that stemming from administrative registers.

References

Agafiţei, M., Gras, F., Kloek, W., Reis, F. & Vâju, S. (2015). Measuring output quality for multisource statistics in official statistics: Some directions. *Statistical Journal of the IAOS,* **31**, pp. 203-211.

Bernardini, A., Chieppa, A., Cibella, N., Gallo, G., Solari, F., Zindato, D. (2022). Evolution of the Italian permanent population census. Lessons learnt from the first cycle and the design of the permanent census beyond 2021. *Conference of European Statisticians, Group of Experts on Population and Housing Censuses*, Geneva, *Switzerland, 21-23 September 2022*. Economic Commission for Europe.

Carbonetti, G., De Matteis, G., Di Zio, M., Fardelli, D., Ferrara, R., Lipizzi F. (2023).

Enumeration area imputation methods for producing sub-municipal data in the Italian permanent population and housing census. *Statistical Journal of the IAOS,* **39**(1), pp. 123-136.

Deville, J.C., Särndal, C.E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, **87**(418), pp. 376-382.

Gallo, G., Zindato, D. (2021). Italy: The combined use of survey and register data for the Italian permanent population census count in UNECE, *Guidelines for Assessing the Quality of Administrative Sources for Use in Censuses.* https://unece.org/statistics/publications/CensusAdminQuality.

Solari, F., Bernardini, A., Cibella, N. (2023). Statistical framework for fully register based population counts, *Metron*, 10.1007/s40300-023-00244-5.

Zhang, L.C. (2022). Complementarities of survey and population registers, in N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri, J.L. Teugels (eds.) *Wiley StatsRef: Statistics Reference Online*. https://doi.org/10.1002/9781118445112.stat0835

doi.org/10.26398/asaproc.0026

Artificial intelligence and sustainability

Alessandro D'Alcantara [a], Fabrizio Drago[a], Angelo Lamacchia[a], Giulio Setzu[a], Gioele Gargano[a], Nicolò Montesano[a], Massimiliano Giacalone[b], Angelo Romano[a]

[a] iInformatica Srl

[b] Università degli studi della Campania Luigi Vanvitelli Srl

Introduction

Artificial intelligence (AI) is increasingly becoming a disruptive technology across various human activities (Păvăloaia, 2023), including medical diagnosis (Cabitza, 2023; Biswas, 2023; Lu, 2023), autopilot systems (Maleki, 2016), education (Lo, 2023; Grassini, 2023; Pavlik, 2023), and agriculture (Partha, 2023). Its application scope is vast, transcending traditional scientific domains (Johnson,2023). Notably, AI-driven chatbots like ChatGPT have demonstrated their capabilities beyond conventional technology, excelling in tasks like poem and movie script writing (Dwivedi et. al, 2023). Despite potential risks in everyday applications, AI holds promise as a powerful tool for enhancing human work and life (Lund, 2023; Wang, 2023). It's crucial, however, to ensure AI does not supplant human autonomy in decision-making and does not compromise human ability to think critically (Parry, 2016; Singh, 2023). Instead, it should serve as a valuable resource for solving complex issues and aiding humans in informed decision-making (Van den Berg, 2023). Our previous research highlighted ChatGPT's proficiency in generating texts akin to natural human language (Santarcangelo, 2023).

In this study, we explore ChatGPT's effectiveness and accuracy in supporting entities and governments with the UN 2030 agenda implementation. We specifically investigate ChatGPT's utility in identifying and assessing problems related to each UN 2030 agenda goal, both qualitatively and quantitatively. The results were scrutinized against traditional literature research to gauge their reliability. Employing a similar methodology, ChatGPT was also evaluated for its ability to propose viable solutions for achieving the UN 2030 agenda objectives. Finally, we conducted a survey among employees of public entities, university faculty, and secondary school teachers to evaluate their awareness and understanding of AI chatbots, like ChatGPT.

2. Structure of the test

The study was structured taking into consideration the UN 2030 agenda goals, that encompass 17 distinct targets. These goals address various themes, including poverty eradication, efficient and affordable healthcare, environmental sustainability, and more. The primary objective was to assess the capability of the AI-based ChatGPT chatbot in understanding and addressing the UN 2030 agenda themes, both qualitatively and quantitatively. Additionally, the study aimed to evaluate ChatGPT's effectiveness in proposing strategies and solutions for specific problems, thereby testing its utility as a planning tool for entities and governments implementing the UN 2030 agenda. Focusing on the first goal of the UN 2030 agenda, which aims to eliminate extreme poverty by 2030, we designed a test involving 19 diverse questions fed into the chatbot. These questions began with general inquiries about the UN 2030 agenda and global poverty issues, evaluating ChatGPT's ability to focus on these topics and to give some suggestion of how to solve it. In particular, the test was designed to assess the ChatGPT chatbot's comprehension of the UN 2030 agenda by posing six diverse, general questions:

1) *What is UN 2030 Agenda* ?;
2) *Could you give me some numbers about the poverty in the world ?*;
3) *Thanks, but give me some more detailed numbers*;
4) *Ok, until 2021 it is fine;*
5) *Which are the most affected countries ?;*
6) *Which are the causes of the poverty in these nations ?*

The subsequent questions, numbers 7, 8, and 9, sought general advice on global poverty alleviation strategies. Specifically:

7) *Suggest to us how we can solve the problem of poverty in the worlds according to UN 2030 agenda*;
8) *Suggest how to eliminate the extreme poverty (people that live with less than 1,9 dollar /day);*
9) *Suggest how, by 2030, we could reduce by at least half the amount of people of all ages living in extreme poverty in all its dimensions according to UN definitions.*

The test then progressed to simulate ChatGPT's application in a real-world scenario, examining its capacity to provide information about specific regions. For this, we selected two contrasting areas: The Zinder region in Niger, Africa, and Basilicata in Southern Italy. Although Basilicata does not experience extreme poverty like the Zinder region, it was included to test the chatbot's ability to discern and adapt strategies to different regional contexts.

The same set of questions was presented to ChatGPT for both regions, in the same order, to assess the system's reliability in differentiating and tailoring its strategy suggestions for each area. A key part of the test involved simulating the allocation of 1 billion dollars towards specific economic development plans in these regions. ChatGPT's responses were compared with classical literature research to evaluate the system's accuracy in processing relevant data and offering practical suggestions. In particular, the questions for the first test, pertaining to the Zinder region in Niger, were as follows:

10) *Suggest me a detailed program of economic development of Zinder region;*
11) *For each point could you give me a more defined idea of how act with specific actions ?;*
12) *Then simulate me a plan of 1 billion dollars of investment in Zinder region and how I could use these money for promoting socio-economic development;*
13) *Could you give me an idea of return of investment for each points ?;*
14) *Simulate me a realistic ROI in number;*
15) *Suggest me how I could actuate each points and how I should start;*
16) *Do you have some example of african region that actuate similar plan ? which ROI did they have ?;*
17) *According to this data: infrastructure Development (25% - $250 million), how should I invest it ? small infrastructure or big infrastructure? could you give me an idea ?;*
18) *Simulate me a detailed investment program according to the actual infrastructure of Zinder region;*
19) *According to Zinder region conditions could you suggest me a touristic development plan of 50 million dollars ?.*

The questions of the second test, related to Basilicata region were in order:

10) *Suggest me a detailed program of economic development of Basilicata region;*
11) *For each point could you give me a more defined idea of how act with specific actions ?;*
12) *Then simulate me a plan of 1 billion dollars of investment in Basilicata region and how I could use these money for promoting socio-economic development;*
13) *Could you give me an idea of return of investment for each points ?;*
14) *Simulate me a realistic ROI in number;*

15) *Suggest me how I could actuate each points and how I should start;*

16) *Do you have some example of italian region that actuate similar plan ? which ROI did they have ?;*

17) *According to this data: infrastructure Development (15% - $150 million), how should I invest it ? small infrastructure or big infrastructure? could you give me an idea ?;*

18) *Simulate a detailed investment program according with the actual infrastructure of Basilicata region;*

19) *According to Basilicata region conditions could you suggest me a development plan of 200 million dollars in Agri-Tourism and Agri-Food ?.*

These 19 questions were input into ChatGPT in two distinct stages, one for the Basilicata region and the other for the Zinder region.

3. Results of the test

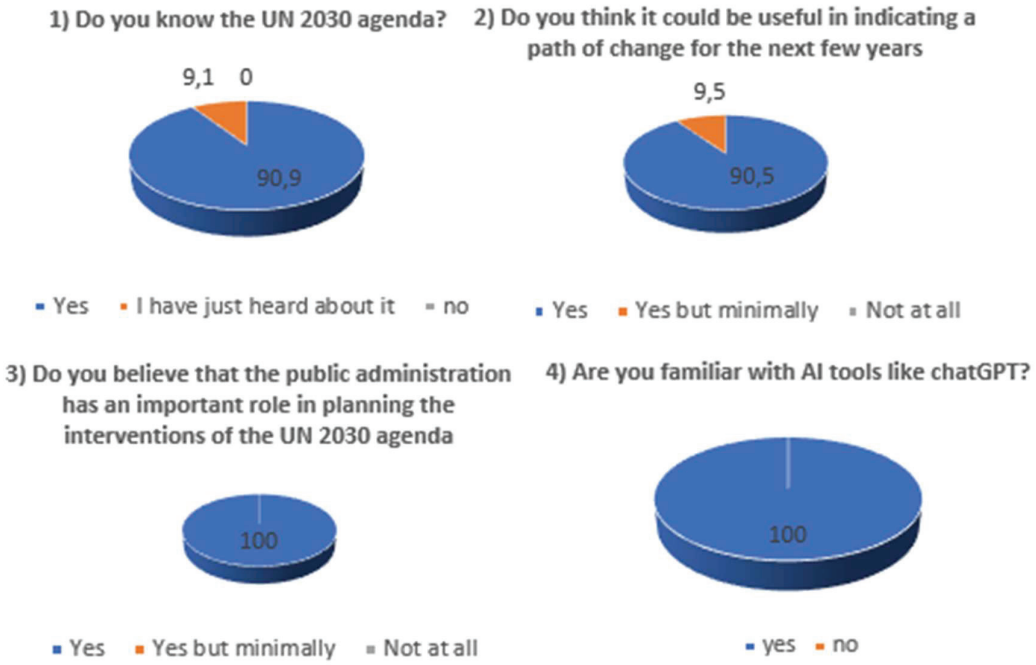
We observed ChatGPT's proficiency in grasping the topic and providing succinct yet detailed answers. In response to the first question, the chatbot accurately identified all 17 targets of the agenda, offering remarkably consistent responses across both tests. This demonstrated ChatGPT's ability to generate similar and coherent answers to identical questions, albeit using varied vocabulary and different lines of reasoning. Regarding questions 2 and 3, the chatbot precisely characterized the issue of poverty, providing accurate numbers and tailoring its responses for each test. It is noteworthy that ChatGPT's algorithm consistently presents data up until 2021. In response to question 5, ChatGPT accurately listed the poorest countries, in line with available data, primarily located in Africa and Asia. This confirmed the chatbot's reliability in delivering general information. For question 6, ChatGPT varied its responses between the tests, offering a more detailed explanation in the first instance and a more concise one in the second. Nevertheless, in both cases, the chatbot effectively conveyed the primary causes of global poverty, demonstrating its clarity and efficiency in addressing complex issues.

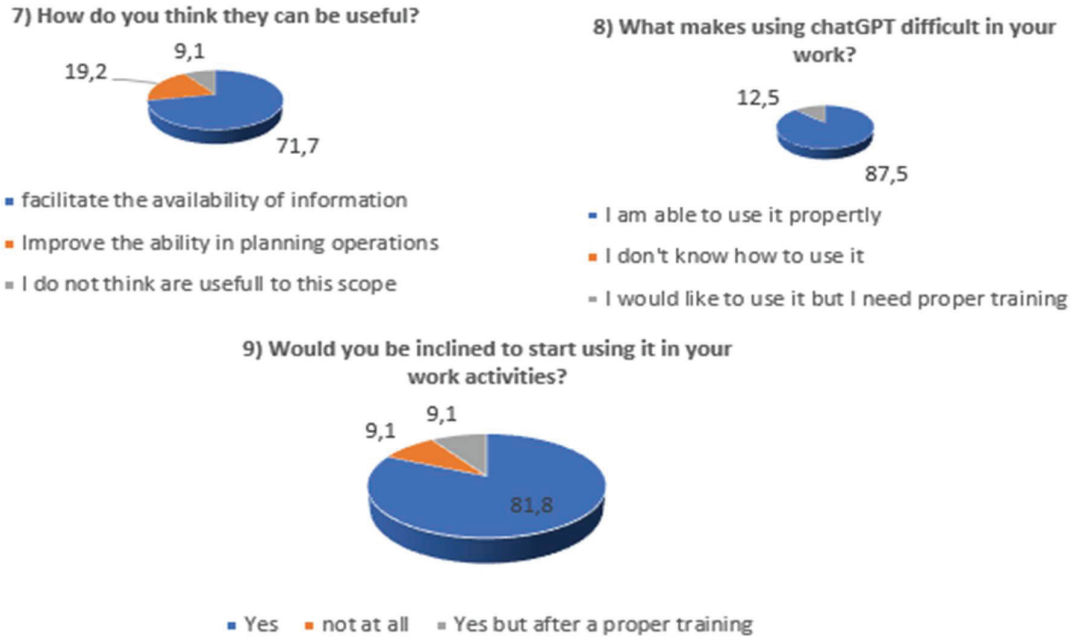
Regarding question 7, 8, and 9, in both tests, ChatGPT's responses to these questions were notably similar, offering general concepts and ideas rather than specific solutions. This was likely due to the broad nature of questions 7, 8, and 9. To garner more concrete strategies for poverty reduction, we posed more detailed inquiries (questions 10 to 19) to the chatbot. These questions were designed to focus on specific regions, allowing us to evaluate ChatGPT's ability to propose different strategies tailored to the unique characteristics of each area. We concentrated on the Zinder region in Niger and the Basilicata region in Southern Italy. Given that both regions are classified as undeveloped, but with Zinder located in one of the poorest countries in the world and Basilicata in a wealthier nation, we expected distinct strategic recommendations for each area. The responses provided by ChatGPT's algorithm to these questions were notably intriguing. Specifically, for questions 10 and 11, the answers, while somewhat general, varied between the two regions. ChatGPT recommended investments in education, infrastructure, and sustainability for both areas, which aligns with typical needs of undeveloped regions. However, it uniquely suggested investments in gender equality and microfinance for the Zinder region, highlighting these as more critical issues there than in Basilicata. In response to question 12, where we asked the chatbot to simulate a $1 billion investment strategy, the results demonstrated ChatGPT's impressive potential. The algorithm proposed two distinct strategies, allocating specific investment amounts across various sectors (like infrastructure, education, and agriculture) for each region. Notably, for the Basilicata region, it recommended a greater focus on renewable energy, infrastructure, and agriculture, compared to education, and did not suggest investments in gender equality and microfinance. In contrast, the strategy for the Zinder region allocated a significant portion of the investment to gender equality, microfinance, and education, reflecting the stark differences between the two regions. Question 13 yielded no significant new information in either case. Question 14, however, provided similar yet slightly varied ROI percentages for similar investments in both

regions, perhaps indicating the different potential impacts of the invested funds. Question 15 elicited general suggestions in both instances. For question 16, ChatGPT presented diverse case studies of investment plans implemented in African and Italian regions. Delving deeper with question 17, we asked the chatbot to simulate an investment plan for a specific aspect of the previously mentioned strategy. While ChatGPT did not provide a detailed investment plan, it offered valuable insights into how the funds could be effectively utilized in both regions. With question 18, we sought a more comprehensive investment plan. Here, the algorithm provided a step-by-step investment breakdown for both Basilicata and Zinder, simulating a practical investment strategy. Finally, question 19 involved simulating a detailed investment plan in tourism and agri-tourism for both regions. In this case, ChatGPT's responses varied according to each region, detailing specific investment steps. The suggestion to invest in the Grand Mosque of Zinder and the Sultan Palace was particularly noteworthy, affirming tailored investment strategies for the Zinder region. These results underscored ChatGPT's potential as a valuable tool in addressing the challenges outlined in the UN 2030 agenda. By posing precise and targeted questions, ChatGPT was able to provide more detailed and specific investment strategies.

4. AI Chatbots and the UN 2030 Agenda: A survey of public sector awareness and use

Additionally, we conducted a survey among workers in public institutions, including ministries, universities, and secondary schools, to assess their awareness and expertise regarding AI chatbots like ChatGPT. The survey initially gauged the participants' familiarity with the UN 2030 agenda and the role of public administration in executing its goals. Subsequently, it probed their knowledge and proficiency in using tools like ChatGPT. Lastly, participants were queried about their views on the potential of ChatGPT to assist public administrations in developing solutions for the UN 2030 agenda goals and their willingness to use such tools in their professional activities. The results of the survey, summarized in what follows, provided insightful data on these aspects





5. Conclusions

In this study, we evaluated ChatGPT's efficacy in supporting organizations and governments with the implementation of the UN 2030 agenda. Specifically, we explored the use of ChatGPT for analyzing and addressing the challenges associated with each goal of the UN 2030 agenda, both qualitatively and quantitatively. The focus was on the agenda's first goal: the eradication of extreme poverty by 2030. To this end, we programmed the chatbot with 19 distinct questions. These questions were administered twice, initially concentrating on the Zinder region in Niger, followed by the Basilicata region in Southern Italy. The outcomes of this test highlighted ChatGPT's potential as a valuable tool in identifying and brainstorming solutions for the objectives of the UN 2030 agenda. The chatbot provided broad ideas for resolving each issue and, when prompted with specific inquiries, offered more detailed action plans, both qualitatively and quantitatively.

Additionally, a survey conducted among public administration workers shed light on the growing relevance of the UN 2030 agenda and the emerging role of artificial intelligence chatbots in everyday professional tasks. The survey findings indicated that a majority of the respondents recognize the potential of ChatGPT in aiding institutions to achieve the UN 2030 agenda goals. Crucially, following adequate training, most interviewees were open to incorporating ChatGPT into their work processes.

References

Biswas S. S. (2023). Role of Chat GPT in public health. *Annals of Biomedical Engineering*, **51**(5), pp. 868–869.

Cabitza, F., Campagner, A., Ronzio, L., Cameli, M., Mandoli, G.E., Pastore, M.C., Sconfienza, L. M., Folgado, D., Barandas, M., Gamboa, H. (2023). Rams, hounds and white boxes: Investigating human-AI collaboration protocols in medical diagnosis. *Artificial Intelligence in*

Medicine, **138**, 102506.

Dwivedi, Y.-K., Kshetri, N., Hughes, L., et al. (2023). Opinion paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy, *International Journal of Information Management,* **71**, 102642.

Grassini S. (2023). Shaping the future of education: Exploring the potential and consequences of AI and ChatGPT in educational settings. *Education Sciences*. **13**(7): 692.

Johnson, M., Albizri, A., Harfouche, A., Tutun, S. (2023). Digital transformation to mitigate emergency situations: Increasing opioid overdose survival rates through explainable artificial intelligence, *Industrial Management & Data Systems*, **123**(1), pp. 324-344.

Lo, C.K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature, *Education Sciences*, **13**(4): 410.

Lu, Y., Wu, H., Qi, S., Cheng, K. (2023). Artificial intelligence in intensive care medicine: Toward a ChatGPT/GPT-4 way?. *Annals of Biomedical Engineering*, **51**(9), pp. 1898–1903.

Lund, B.D., Wang, T. (2023), Chatting about ChatGPT: How may AI and GPT impact academia and libraries?. *Library Hi Tech News*, **40**(3), pp. 26-29.

Maleki, K.N., Ashenayi, K., Hook, L.R., Fuller, J.G., Hutchins, N. (2016). A reliable system design for nondeterministic adaptive controllers in small UAV autopilots. *IEEE/AIAA 35th Digital Avionics Systems Conference (DASC), Sacramento, CA, USA*, pp. 1-5.

Parry, K., Cohen, M., Bhattacharya, S. (2016). Rise of the machines: A critical consideration of automated leadership decision making in organizations. *Group & Organization Management*, **41**(5), pp. 571-594.

Partha, P.-R. (2023). AI-assisted sustainable farming: Harnessing the power of ChatGPT in modern agricultural sciences and technology. *ACS Agricultural Science & Technology*, **3**(6), pp.460-462.

Păvăloaia, V.-D., Necula, S.-C, (2023). Artificial intelligence as a disruptive technology—A systematic literature review, *Electronics*, **12**(5), 1102.

Pavlik, J.V. (2023). Collaborating with ChatGPT: Considering the implications of generative artificial intelligence for journalism and media education. *Journalism & Mass Communication Educator*, **78**(1), pp. 84-93.

Santarcangelo, V., Lamacchia, A., Massa, E., Crisafulli, S., Giacalone, M., Basile, V. (2023). Sustainability explained by ChatGPT artificial intelligence in a HITL perspective: innovative approaches. *Conference: Italian Statistical Society (SIS) 2023, Ancona - June 21-23, 2023 Volume: Statistical Learning, Sustainability and Impact Evaluation*

Singh, G. (2023). Leveraging ChatGPT for real-time decision-making in autonomous systems. Eduzone: *International Peer Reviewed/Refereed Multidisciplinary Journal*, **12**(2), pp. 101–106.

van den Berg, G., Du Plessis, E. (2023). ChatGPT and generative AI: Possibilities for its contribution to lesson planning, critical thinking and openness in teacher education. *Education Sciences*, **13**(10): 998.

Wang, C., Liu, S., Yang, H., Guo, J., Wu, Y., Liu, J. (2023). Ethical considerations of using ChatGPT in health care. *Journal of Medical Internet Research*, **25,** e48009.

doi.org/10.26398/asaproc.0027

A composite index to assess the complexity of environmental consciousness: Empirical evidence from European consumers

Ida D'Attoma[a], Marco Ieva[b],
[a] Department of Statistical Sciences, University of Bologna, Bologna, Italy.
[b] Department of Economics and Management, University of Parma, Parma, Italy.

1. Introduction and theoretical framework

Population growth, industrial production processes and new consumption models have generated a huge number of environmental protection challenges (Fontecha et al., 2022). Worldwide policy makers undertake efforts to promote sustainable behaviours among companies and consumers such as reducing energy consumption and greenhouse gas emissions, purchasing products or services that have a lower negative impact towards the environment and implementing garbage recycling.

Environmental consciousness has been found to lead consumers to undertake behaviours with a significant and positive impact on environment (Abd'Razack et al., 2017). Consumers represent a very important part of the production-consumption chain, as their choices influence sustainability practices among companies, thus becoming a driver for corporate and public actions in implementing sustainability practices and green innovations (Caeiro et al., 2012). Moreover, measuring environmental consciousness over time could be useful to policy makers to understand the current levels of awareness of the public and to measure the effectiveness of their environmental promotion campaigns over time.

Despite the relevance of environmental consciousness, several unresolved issues exist as far as its measurement is concerned. The available measures are country-specific, and they have been developed with a predominant U.S. focus (Diamantopoulos et al., 2003). Current measures of environmental consciousness are mostly based on convenience samples that are unrepresentative of the total population: this entails limitations in terms of interpretability and external validity (Diamantopoulos et al., 2003). Moreover, the environmental consciousness cannot be measured by a single descriptive indicator, but instead it should be represented with multiple dimensions. To the authors' knowledge no unique environmental consciousness index was developed. Therefore, given the relevance of the topic and in light of the above gaps, the present study proposes a composite indicator of environmental consciousness.

The present paper adopts the theoretical framework developed by Sanchez and Lafuente (2010), that offers an operationalisation of environmental consciousness based on the attitude structure theory (Dunlap et al., 2000). The framework assumes that environmental consciousness can be explained by four dimensions: affective, cognitive, dispositional and active. The affective dimension of environmental consciousness captures the concern for the environment and the endorsement of a pro-environmental global perspective of environmental issues and reflects to what extent the individual perceives the importance of environmental issues. The cognitive dimension is related to the amount of information and knowledge about environmental issues and to what extent the individual is aware of environmental issues and of its causes and consequences. The dispositional dimension addresses the personal involvement and intention to take responsibility and to support personal costs when undertaking pro-environmental behaviours and following environmental policies. The active dimension refers to different pro-environmental behaviours.

2. Data presentation and data analysis step

The development of the present composite indicator was based on secondary data Eurobarometer 92.4 (European Commission, 2019) on the 'Attitudes of European citizens towards the environment'. The advantage of using this survey is that it includes indicators for the four dimensions of environmental consciousness considered in the theoretical model. Moreover, we had the opportunity to compare indicators across countries. The topic of EB 92.4 is related to the European citizens' attitudes toward perceptions of climate change, attitudes toward environmental issues, sources of information on environmental issues, environmental activities individual sustainability, concerns with respect to environmental issues, assessment of the environmental behaviour, environmental issues of institutional actors, attitudes toward plastic recycling and air quality.

The anonymized EB 92.4 data include information on 27,498 citizen-consumers from the 28 EU member states, roughly 1000 individuals per country. Following the before mentioned theoretical model of Sanchez and Lafuente (2010), we have considered 19 primary indicators[1]: 11 for the affective dimension (all 11 indicators collected the perceived importance of various environmental issues such as, to cite, few, protecting the environment, climate change, impact of environmental issues on life, health and consumption habits, impact of product made of plastic, microplastics and chemicals in everyday products), 1 for the cognitive dimension (level of available information), 2 for the dispositional dimension (individual sense of responsibility, personal costs) and 5 for the active dimension (engagement on behaviour, like for instance 'use of more environmentally-friendly way of travelling', or on buying behaviour aimed at reducing the impact on the environment). All of the variables selected refer to questions with a different response scale. In some cases, it was necessary to rotate the response scale so that an increase in the normalized indicators corresponds to an increase in the composite indicator (Salzman, 2003). All missing values were imputed or removed from the dataset. Consequently, the sample size used in the analysis was reduced to 27378 respondents.

The analysis proceeded in several stages. First, indicators were normalized. Second, indicators weights were obtained using the PCA-based strategy illustrated in the methodology section. Third, indicators were aggregated by considering different levels of compensability. After that a further level of aggregation was employed to obtain an index by country. Finally, the index by country was compared to an index based on Google Trends query searches popularity.

3. Methodology

3.1. The construction of the environmental consciousness index

The main factors to take into account in the construction of the composite indicator are: i) indicator selection and normalization, ii) weighting of indicators and iii) aggregation of indicators. In the present work the approach based on the distance to the ideal point measured by different degrees of compensability developed by Diaz-Balteiro and Romero (2004) along with a principal component analysis (PCA) based strategy to compute weights has been chosen for implementation. Let us consider the sample of n elementary units (the flash EB respondents). Each respondent i (i=1,2,…, n) is evaluated according to m primary indicators of environmental consciousness j, with j=1,2,…, m. Let I_{ij} be the value of the generic jth primary indicator for the generic ith respondent.

The first step toward the construction of the composite indicator will involve normalizing of the m indicators. Following Diaz-Balteiro and Romero (2014) we use the normalization procedure as in (1), that is appropriate when the indicator will assume the meaning of '*more is better'*:

$$\bar{I}_{ij} = 1 - \frac{I_j^* - I_{ij}}{I_{*j} - I_j^*} = \frac{I_{*j} - I_{ij}}{I_{*j} - I_j^*} \quad (1)$$

[1] The full list of the 19 indicators along with a detailed description of them is available upon request.

with I_j^* as the optimum value of the *j*th indicator of environmental consciousness (ideal value)[2], I_{*j} as the worst score assigned (anti-ideal), $\bar{I}_{ij}$ as the normalized score assigned by the *i*th respondent with respect to the *j*th indicator of environmental consciousness.

Using the above normalization system, the indicators of environmental consciousness have no dimension and are also bounded between 0 and 1. An 'objective' weighting is then used by constructing weights using PCA. Once the principal components have been retained, variables' weights were attributed by multiplying the contribution of each *j*th primary indicator (I_{ij}) to the K most important components retained k – say L_{jk} −with their proportion of explained variance (λ_k) as in (2):

$$W_j = \sum_{j=1}^{m} \sum_{k=1}^{K} \left| L_{jk} \right| \cdot \lambda_k \quad (2)$$

with W_j as the weight of the *j*th primary indicator, L_{jk} as the loading value of the *j*th primary indicator on the principal component k and λ_k as the proportion of the explained variance of the *k*th PC. Final weights were rescaled to sum up to one. After variables' normalization and weighting, the aggregation of the different indicators into a composite one, which measures the environmental consciousness as a whole, represents a crucial and complex problem. In this work, following Diaz-Balteiro and Romero (2004), the composite indicator of environmental consciousness – ECI_i- is established for each respondent by calculating the complement of the distance between normalized scores and the ideal vector as in (3)

$$ECI_i = \sum_{j=1}^{m} W_j^p \, \bar{I}_{ij}^{\,p}, \forall\, i \quad (3)$$

Therefore, the respondent that maximizes the expression (3) is the 'most environmental conscious' respondent.

For p=1, we obtain a linear additive aggregation method. Because the weighted additive system ignores interactions between variables (Diaz-Balteiro and Romero, 2014) and implicitly assumes total compensability between indicators (Gómez-Limón et al. 2020), following the more general framework developed by Biaz-Balteiro and Romero (2014) for the sustainability case, we construct the environmental consciousness composite indicator as in (4):

$$ECI_i = (1-\lambda)[\min_j (W_j \bar{I}_{ij})] + \lambda \sum_{j=1}^{m} W_j \bar{I}_{ij}, \forall\, i \quad (4)$$

By considering various values of λ in (0,1] we are able to account for different degrees of compensability. In particular, we take into consideration two different situations: a) total compensability (λ=1) and b) various degree of partial compensability (λ=0.25, λ=0.5, λ=0.75). Zero compensability (λ=0) was not taken into account since we have assumed that primary indicators were substitutable.

3.2 Robustness analysis

To test the robustness of our composite indicator we analyse its relationship with Google Trends (GT) data related to the number of searches in multiple countries on environment-related topics. Searching for information is a key driver of pro-environmental behaviour (Ritter et al., 2015). Consumers displaying higher environmental consciousness are more likely to search for additional information on the environmental features of products (Testa et al., 2020). Searching information can be thought to be positively related to each of the dimensions of 'environmental consciousness': it means acquiring information (cognitive), it might stem from a concern for environment (affective), it signals a personal involvement in environmental matters (dispositional) and reflects an actual behaviour oriented towards the environment (active). Therefore, we expect a positive relationship between our composite index at the country level and the number of searches on Google Trends per country.

[2] This value represents a maximum value if (as in our case) the indicator is of the type 'more is better'.

Thus, we use online search behaviour recorded on the Google Trends website (trends.google.com) to measure public awareness environment related issues. In response to queries, GT reports how often the worldwide population of internet users does Google searches for a given term in proportion to the total of all Google searches during the given range of dates. Google excludes repeated queries from a single user over a short time-period to avoid pseudo-replication of web searches. The reported output provides normalized search data presented on a scale from 0 to 100. GT measures search popularity in relative terms based on a randomly drawn sample of all search terms used in queries within the chosen period (e.g., Rousseau and Deschacht, 2000). Therefore, the search popularity indicator (SPI) might be expressed as in (5)

$$SPI_{i,t,c}^{z} = \frac{\vartheta_{t,z,c}^{i}}{\max_{t\in T, i\in I}\{\vartheta_{t,z,c}^{i}\}} \times 100 \quad (5)$$

where I is the set of search terms (or topics) in which a researcher is interested (e.g. 'climate change'), t is a time unit within the time horizon T (e.g. a day in 2019), z the chosen region (e.g. France) and c the chosen category (e.g. animals), and the search popularity indicator of search term i element of set I is defined as in (6):

$$\vartheta_{t,z,c}^{i} = \frac{\#\ searches\ for\ term\ i\ in\ category\ c, during\ time\ unit\ t\ in\ region\ z}{\#\ searches\ in\ category\ c, during\ time\ unit\ t\ in\ region\ z} \quad (6)$$

Moreover, if the number of searches in the numerator of Eq. (6) is smaller than an (unspecified) threshold, it is set to zero. Google first calculates the relative search volume for each search term i of the set I, in each unit period during a predetermined time horizon T, for region z. This relative search volume is then rescaled for the maximum relative search volume for all search terms in the set I. This maximum is set equal to 100, and all other relative search volumes are expressed as a proportion of this maximum and thus range from 0 till 100.

4. Results

To assign weights to each primary indicator PCA was conducted using the 19 normalized primary indicators as input. Seven principal components, explaining the 72.65% of the total variance, have been retained. Once constructed weights following Diaz-Balteiro and Romero (2004) we aggregated primary indicators using different level of compensability as in (4), thus obtaining the composite one. After calculating the composite indicator, we converted it on a scale from 0 to 100.

Table 1: Descriptive statistics of the composite indicator by compensation parameter (λ)

Degree of compensability	Mean	Median	Min	Max	St.Dev	CV
λ=0.25	57.37	57.72	0	100	13.87	24.17
λ=0.50	57.38	57.73	0	100	13.87	24.17
λ=0.75	57.39	57.74	0	100	13.87	24.17
λ=1	57.39	57.74	0	100	13.87	24.17

After simply looking at the distribution of our indicator (Table 1), we compared the averages and medians of European countries so as to see which countries were the ones whose citizens show higher environmental consciousness. From Figure 1 we can see that the distribution of our indicator `is quite heterogeneous' among the different geographical areas considered. The map shows areas with more intense environmental consciousness. As we would have expected, areas in the Scandinavian countries (e.g. Sweden) find themselves at the top of our ranking and such a result is in line with other studies (e.g., Harju-Autti and Kokkinen, 2014). By varying lambda (λ=0.25,

λ=0.50, λ=0.75) the distribution of countries in terms of their environmental consciousness appears almost the same[3].

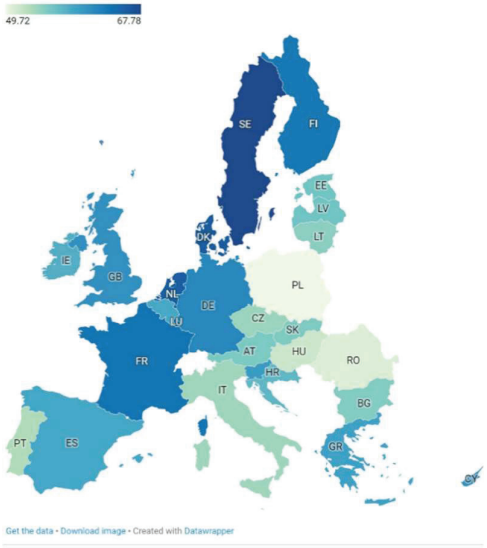


Figure 1: Median environmental consciousness in EU countries (2019) (λ=1).

The country ranking derived by aggregating ECI by country (Figure 1), was then compared with the country ranking based on the search popularity of three environmental-related topics. For the purpose of this study, using Google Trends, 3 environmental-related search topics were queried: "climate change", "environmental issues", "sustainability".

To ensure the temporal consistency of the comparison, we chose to take as the reference period of Google Trends searches the period from 06-12-2019 to 19-12-2019, which basically follows the reference period of the Eurobarometer survey data collection.

In response to investigator queries, GT reports how often the worldwide population of internet users searches Google for a given term as a proportion of the total of all Google searches during the given range of dates. Google excludes repeated queries from a single user over a short time-period to avoid pseudo-replication of web searches.

The search volume was recorded for each of the 14 days on a sample randomly drown from a population of billions of searches per day. Google Trends provides a search volume value (SPI) that is relative and normalized. The SPI is computed for each search term of the topic, in each unit period of the time horizon. The resulting values are scaled based on the topic's proportion to all searches of a topic in a range of 0 to 100, starting from the maximum relative search volume that is set to 100. Once the daily data was downloaded from GT, the average number of searches was calculated in the time span considered. We used the average values to make a ranking of countries with higher interest in the topics 'climate change', 'environmental issues' and 'sustainability'. Therefore, we compared the rankings for 28 countries obtained as a result of constructing the indicator by varying lambda and according to the search popularity of the three different topics. The Kendall's W statistic indicated a significant level of concordance among the set of ranks obtained using GT daily data and the set of ranks obtained using different levels of compensability for the composite indicator.

5. Conclusions

[3] To evaluate the degree of agreement among the set of m ranks, with m>2, (for the n countries produced by the composite indicator, ECI) we use the Kendall's coefficient of concordance W (Kendall and Babington-Smith, 1939; Teles, 2012)

The developed indicator aimed to be a comprehensive measure of consumer environmental consciousness that allows to capture heterogeneity at the country level in multiple countries. As of geographical heterogeneity, we found a great degree of cross-country variability in the composite index with Sweden, Denmark and Netherlands outperforming the rest of the countries and northern and central-western European countries (citizens) showing -on average- higher environmental consciousness than southern and central-eastern ones. The robustness of the developed indicator is also assessed by employing an additional data source, namely user search data on Google Trends at the country level.

Results show a strong level of concordance among the set of ranks obtained considering the actual consumer search patterns on Google and the set of ranks obtained using different level of compensability for the composite indicator developed using survey data. Given the relevance of the phenomenon for both policy makers and citizens, a composite index certainly provides relevant benefits. First, on the empirical side we contribute to the stream of literature studying the awareness on environment related issues by enlarging the geographical scope. Second, a composite index can be easily disseminated to the general public and immediately understood by the users.

References

Abd'Razack, N.T., Medayese, S.O., Shaibu, S.I., Adeleye, B.M. (2017). Habits and benefits of recycling solid waste among households in Kaduna, North West Nigeria. *Sustainable Cities and Society*, **28**, pp. 297-306.

Caeiro, S., Ramos, T.B., Huisingh, D. (2012). Procedures and criteria to develop and evaluate household sustainable consumption indicators. *Journal of Cleaner Production*, **27**, pp. 72-91.

Diamantopoulos, A., Schlegelmilch, B.B., Sinkovics, R.R., Bohlen, G.M. (2003). Can socio-demographics still play a role in profiling green consumers? A review of the evidence and an empirical investigation. *Journal of Business Research,* **56(6)**, pp. 465-480.

Diaz-Balteiro, R., Romero, C. (2004). In search of a natural system sustainability index, *Ecological Economics*, **49**, pp. 401-405.

Dunlap, R.E., Van Liere, K.D., Mertig, A.G., Jones, R.E. (2000). New trends in measuring environmental attitudes: Measuring endorsement of the new ecological paradigm – A revised NEP scale. *Journal of Social Issues*, **56(3)**, pp. 425-442.

Fontecha, J.E., Nikolaev, A., Walteros, J.L., Zhu, Z. (2022). Scientists wanted? A literature review on incentive programs that promote pro-environmental consumer behavior: Energy, waste, and water. *Socio-Economic Planning Sciences*, **82A**, 101251.

Ritter, Á.M., Borchardt, M., Vaccaro, G.L., Pereira, G.M., Almeida, F. (2015). Motivations for promoting the consumption of green products in an emerging country: exploring attitudes of Brazilian consumers. *Journal of Cleaner Production*, **106**, pp. 507-520.

Rousseau, S., Deschacht, N. (2020). Public awareness of nature and the environment during the COVID-19 crisis, *Environmental and Resource Economics*, **76**, pp. 1149-1159.

Salzman, J. (2003). *Methodological Choices Encountered in the Construction of Composite Indices of Economic and Social Well-being.* Technical Report, Center for The study of Living Standards, Ottawa.

Sanchez, M, Lafuente, R. (2010). Defining and measuring environmental consciousness, *Revista International de Sociologia*, **68**, pp. 731-755.

Testa, F., Iovino, R., Iraldo, F. (2020). The circular economy and consumer behaviour: The mediating role of information seeking in buying circular packaging. *Business Strategy and the Environment*, ***29***(8), pp. 3435-3448.

doi.org/10.26398/asaproc.0028

The Italian NEETs from the economic crisis to the aftermath of the COVID-19 pandemic and the role of the channels used to find a job

Giovanni De Luca[a], Paolo Mazzocchi[a], Federica Pizzolante[a], Antonella Rocca[a], Claudio Quintano[a]

[a] Department of Management and Quantitative Studies, Parthenope University of Naples, Naples, Italy

1. Introduction

In the European Union, the financial and economic crisis of the past decade had a particularly severe impact on Mediterranean countries. Among the EU countries, even within a common decreasing trend for the NEET rates, since 2014, Italy consistently had the highest proportion on the total young people of those who are Not in Employment, Education, or Training (NEET). This indicator is considered more comprehensive than the youth unemployment rate to describe the young people's vulnerability in the labour market as it considers not only unemployment but also the exclusion from the labour market due to inactivity. Therefore, it provides a complete picture of the unproductive potential of the youth, with significant consequences at the macroeconomic, social, and personal levels.

Reducing the share of NEETs is one of the sub-goals explicitly included in the UN Agenda, aligned with the Sustainable Development Goals (SDGs). Specifically, it falls under indicator 8.6, within the pillar of Decent work and economic growth. The reasons for these high NEET rates in Italy are in part connected to the high levels of unemployment and an education system that is not effectively aligned with the needs of the labour market. Indeed, the Italian education system fails to equip young people with the skills required by employers and struggles to keep them engaged in education for an extended period.

In comparison to the other EU countries, Italy shows low percentages of tertiary graduates and a high rate of early school leavers, who drop out school after completing compulsory education (De Luca et al., 2020). Moreover, Italy's investment in education and active labour market policies are relatively low compared to other EU countries. Additionally, the institutions designed to support young people in transitioning from school to work are underdeveloped and lack the necessary resources to effectively reach all those in need and therefore are unable to provide tailored services. An example refers to the Public Employment Service (PES) which in Italy have one of the lowest proportions of individuals registered among the EU-27 countries (Pacifico et al., 2018). However, in response to the pandemic, PES made strong efforts to improve productivity, favouring the match between the supply and the demand of work, and received more financial resources particularly to support the digitalization of the services provided (OECD, 2020).

This paper aims to analyse the NEET rates from 2012 to 2021 and examine the impact of different job search channels on these rates. Specifically, we focus on NEETs, with the scope to identify which is the role of the different job search channels used by unemployed NEETs. By understanding the job search behaviour and preferences of NEET individuals, targeted interventions can be designed to enhance their access to suitable job opportunities and improve their employment outcomes. Indeed, even if the economic literature dedicated to youth unemployment great attention, only few studies (among others, Ghignoni, 2021) focused on the effectiveness of search channels for young people. This topic is even more important in view of the strong occupational mismatches highlighted by recent data showing that many enterprises find very difficult to hire the professional profiles they need.

The paper is organized as follows: in Section 2, an overview of the economic literature on NEETs and the job search channels is presented; Section 3 shows data and the methodology, while in Section 4 we report the main findings. Finally, Section 5 concludes.

2. The framework of analysis

Even if the general level of unemployment plays an important role in determining the share of NEETs, other factors are relevant to explain this outcome. Indeed, comparing the youth and the adult unemployment rate, in 2021 in Italy the ratio between these two indicators was 3.54, higher than in the majority of the other EU countries, denoting a stronger young people's vulnerability in the labour market. Even worse the situation in 2021, when the ratio was 3.92.

Young people are in the NEET condition when they, after completing education, enter the labour market and search for a job, or when they look for a job after a previous job experience. Finally, they can be in the NEET condition when they are not working and not searching for a job. Indeed, NEET refers to young people who are unemployed or inactive. Previous studies have demonstrated that among young people, the condition of inactivity is very often connected with discouragement, especially in Italy (Pastore et al., 2021; Quintano et al., 2018). Discouragement implies that a person renounces to search for a job in the conviction that no job is available. High rates of discouraged NEETs are certainly linked to the high levels of unemployment, but even to the underdeveloped institutions that should help young people to find a job and more in general assist them during their school-to-work transition. The most important institution within this framework is the PES, whose main goals consist in connecting the supply of work with its demand, that is employers and aspiring employees. Other important functions of these offices consist in providing assistance to those who are searching for a job, giving information about possible training opportunities, and advice on how to prepare the curriculum or on how to deal with a job interview.

However, the Italian PES, such as all the public administration offices, shows many key issues. The challenges that they are facing in the last years concern mainly the shortage of skilled staff, that act as an obstacle in providing adequate personalized services, and a strong decentralization, that can lead to still deeper territorial differences (Sultana and Watts, 2005). As Eurostat (2010) highlights, there is a need for a more efficient, restructured, and digitalized public administration, to enhance quality and citizen satisfaction, and raise productivity. However, in the aftermath of the COVID-19 pandemic, many resources have been addressed to PES, accelerating, in particular, the conversion of many services to the digital mode. PES are recently even investing a lot of their efforts in improving their activities, developing specific services tailored to the different people needs. The main innovations consist in creating advanced statistical profiling techniques (Linfante et al., 2023) and realize the conversion of many services in the digital mode. During the last decade, indeed, the digital revolution led to deep changes in job search channels. This implied the conversion of many of them to the digital mode, but even the creation of new channels, based on digital platforms. Unfortunately, the official statistical sources still do not provide specific information on the new digital channels. However, they allow us to analyse the changes in the use of the traditional ones. Due to these great issues still in part not being solved, data reveals that Italian young people exhibit a higher propensity for using informal channels and direct application to employers compared to their counterparts in other EU countries (Eurostat online database). Previous studies have shown that a greater reliance on informal channels tends to connect with a lower proportion of tertiary-educated individuals, consequently leading to negative effects on youth employment rates (Ghignoni, 2021).

This paper analyses the NEET rate in Italy in the period from 2012 to 2021, with the aim to verify the effect of the different job search channels. The period under analysis is of great interest, as it includes the recovery from the 2007-2010 financial crisis, as well as the subsequent crisis triggered by the COVID-19 pandemic. These events likely brought about permanent shocks. Additionally, the digital revolution, accelerated by the pandemic, further transformed in these

years the dynamics of interactions among the various actors in the labour market. Finally, the significant labour market reforms implemented during this period in Italy undeniably influenced the NEET rates and their propensity to engage with PES (De Luca et al., 2019). Notable reforms include the introduction of the Youth Guarantee Fund in 2014, subsequently revised in 2021 in response to the pandemic crisis, as well as the implementation of the citizenship income in 2019.

The Youth Guarantee is a collection of active labour market policies designed to promote the integration of young people into the workforce. It ensures that all the registered beneficiaries gain access to job offers or quality training courses within four months from registration (Council Recommendation, 2020/C372/01). In 2021, following the COVID-19 crisis, a second phase of the Youth Guarantee was funded, but limited to the poorest regions of Basilicata, Calabria, Campania, Puglia, Sicily, as well as the transitional regions of Abruzzo, Molise, and Sardinia. The target beneficiaries remained young people in the NEET condition aged 15-29, with an extension of up to 34 years.

The other important reform of the labour market involving the unemployed was the "citizenship income". It was introduced as a reform in 2019, providing a subsistence income to combat poverty. It was addressed to unemployed people of any age and served as a guaranteed minimum income, conditional on receiving assistance from job centres for job searching and accepting subsequent job offers. Both these reforms aim to engage NEETs by providing them with job experiences or training opportunities. The anticipated effect of these reforms was a reduction in NEET rates, while also potentially increasing the number of young people registered with PES, since registration is required to access these benefits. However, another possible effect of the introduction of the citizenship income could have acted in the opposite direction leading to an apparent increase in NEET rates and overall unemployment rates. Indeed, like other unemployment benefit programs, it may have increased the incentive for individuals to register and receive economic incentives without actively seeking employment. This was one of the declared reasons for its abolition in July 2023.

3. Data and methods

Data are from the Istat labour force survey (LFS), which is specifically designed to generate and monitor labour market indicators for short-term analysis, providing insights into employment figures, job characteristics, working conditions, and job search activities. The age class considered in this study for the NEET indicator is 15-34 years, in view of the longer Italian transitions to work (Pastore et al., 2022). We constructed time series of the NEET rates and information pertaining to the job search channels for those NEETs who declare to be searching for a job during the period from 2012 to 2021, encompassing the latest available data. Indeed, by utilizing the specific week of the survey interview, we were able to calculate indicators based on a four-week period, resulting in a total of 13 observations per year. Consequently, our time series dataset consists of 130 observations in total. Given the significant differences by gender observed in NEET rates, we decided to analyse the data separately for men and women. Historically, Italian women have displayed a higher inclination towards inactivity and have even faced higher unemployment rates. Examining the phenomenon by gender allows us to verify whether significant gender differences also extend to the job search channels employed, significantly contributing to the existing literature on the topic.

The reference to the NEET rates rather than to the youth unemployment rates for the analysis of the job search channels was driven by the consideration that in most of cases even inactivity, especially for young people, arises from motivations related to the absence of an adequate labour market functioning. Looking at the causes of inactivity (Mascherini, 2016), the most frequent ones for Italian young people are discouragement or caregiver needs. The former relates to people who want to work but do not search for a job because they are convinced that no work is available for them. The latter is related to persons who need to take care of people who are not self-sufficient and, in most of cases, would work if adequate welfare services were available.

The methodology employed to study the NEET rates is an auto-regressive integrated moving average (ARIMA) model. A delay of one or two periods (referred therefore to 4 or 8 weeks) has been considered for the job search channels, to account for the time needed for them to be efficacious.

With reference to job search channels, the labour force survey investigates the use of different methods, including those of both formal and informal types. Among the first type, we find the interaction with public and private agencies finalized to promote the match between the demand and the offer of work (methods A and B, respectively), the direct application to employers (Method C), or the execution of test, interview or examination (Method G). Channels linked to newspapers and journals include the insertion or response to an advertisement (Method E) and the study of advertisements (Method F). Finally, methods H and I concern the request for land, permits, or equipment and the request for permits, licenses, and financial resources, respectively. Among the more passive methods, we find waiting for a call from PES (method K), for the results of an application (method J), or for a competition for recruitment in the public sector (method L). Finally, for the informal channel, it was considered the network of friends, relatives, trade unions, and others (method D).

4. Results

During the period under analysis, from 2012 to 2021, the NEET rate in Italy remained substantially stable, passing from 25% at the beginning of 2012 to 24% at the end of 2021. The decreasing trend in the remaining EU-27 countries was more pronounced (from 15.8% to 12.6% for the same period). Even the use of the job channels analysed suffered a substantial decreasing trend. In particular, the share of young people who used PES reduced in this period by 16%. The decrease is confirmed, even if less pronounced even in the case we limit the analysis to the pre-COVID-19 situation (6% until 2019). A similar reduction has been registered in the share of people who used private employment agencies (12%) or who made tests, interviews or examinations (14%).

Table 1: ARIMA model for the NEETs job search channels, by gender

NEETs	Women		Men	
	Coefficients	st.dev.	Coefficients	st.dev.
NEET (1)	0.5677***	0.0757	0.7036***	0.0671
A PES (1)	0.0830***	0.0248		
B Private agency (1)				
B Private agency (2)	0.0657**	0.0377	0.0751***	0.0342
C Applied employers (2)	-0.0519*	0.0380		
D asked friends (1)			0.0573*	0.0381
D asked friends (2)	-0.0686***	0.0285		
E inserted or answered adv (2)	0.0528***	0.0256		
F studied adv (1)			-0.0487*	0.0306
G took a test (2)	-0.0513**	0.0310		
H looked for land (2)			0.1499*	0.1071
I permits (2)	-0.4158***	0.1461		
J awaiting results of an application (1)				
J awaiting results of an application (2)	-0.0702***	0.0276		
L awaiting results competition (1)	0.2375***	0.0878		
Intercept	0.3806***	0.0378	0.1895***	0.0369
Log-Likelihood	392.59		391.29	
AIC	-761.17		-768.58	
N	128		128	

* p<0.1; ** p<0.05; *** p<0.01

The ARIMA model findings are shown in Table 1 and indicate that public and private agencies, as well as passive behaviours such as waiting for a call from PES or competition results,

are positively and significantly associated with NEET rates. On the other hand, certain job search channels have a significant impact on reducing NEET rates, but the effects differ by gender. More in particular, waiting for application results is linked to a decrease in NEET rates for both men and women. Conversely, applying directly to employers, seeking help from friends, taking tests, and requesting permits are all associated with a significant reduction in NEET rates, but only for women. In contrast, studying advertisements is found to be the most effective channel for reducing NEET rates, specifically for men.

These results are in part surprising, as they seem to indicate that recurring to formal channels to search for a job increases, rather than reducing, the share of NEETs. However, this positive association is due mainly to the common decreasing trend that characterized both the NEET rates and the use of such types of job search channels. On the other side, the use of PES may not necessarily immediately translate into a transition to employment but could even be connected to the involvement of people in training opportunities or other initiatives finalized to increase employment opportunities.

5. Conclusions

In this paper, we tried to identify the more effective job search channels for reducing the share of NEETs. Since 2014, Italy showed the highest proportion of young people who are NEETs among Mediterranean countries (only in the last year surpassed by Romania). This is attributed to high levels of unemployment, an education system that is not effectively aligned with the needs of the labour market, low investment in education and active labour market policies, and underdeveloped institutions designed to support young people in transitioning from school to work. Italian young people exhibit a higher propensity to using informal channels and direct application to employers compared to their counterparts in other EU countries. This reliance on informal channels is a cultural aspect, that has its roots in a past of inefficiencies and inefficacious of PES, considered underdeveloped and unable to help young people during their school-to-work transition. Another important aspect is the lack of trust in institutions, which tends to connect especially with the low-educated individuals (Ghignoni, 2021). The period under consideration includes the recovery from the financial and economic crisis of 2007-2010 and the subsequent crisis caused by the COVID-19 pandemic. These events have substantially changed the equilibria in the labour market. Significant labour market reforms, such as the Youth Guarantee and the Citizenship Income, have been implemented in this period to address the NEET issue in Italy. However, despite them, the use of formal channels, such as PES, consistently decreased.

From the ARIMA models, it is evident that for women the most efficacious channels for job search are to take tests, to ask for permits, to directly apply to employers, and even to ask friends and relatives. Conversely, for men, study advertisement results in the most efficacious action. No specific positive effect in the NEET reduction is associated with asking PES or private agencies. This outcome requires a deep reflection on the role that these agencies could have and only in part are able to provide. These results are certainly due even to the digital revolution which changed the habits, and the way young people interact even with the labour market. This is one of the reasons for which it is relevant to adequately drive the conversion to digitalization of any channel, to improve and facilitate the interactions between young people and the institutions aimed to help them during their transition to work.

Reducing the share of NEETs requires different types of interventions and policies, involving the education system, and the services that in the labour market should help young people in their school-to-work transition, including the PES, which represents the most important institution in this sense. Policymakers, workforce development agencies, and social support organizations can utilize these insights to design effective strategies to reintegrate NEET individuals into the labour market, improve their access to job opportunities, and enhance their employment outcomes.

Future research should build upon the findings of this study and explore additional avenues to identify the most effective job search channels for reducing the share of NEETs. It would be

valuable to investigate the impact of specific interventions, such as the Youth Guarantee Fund and the Citizenship Income, on NEET rates and employment outcomes in Italy. By examining the long-term effects of these reforms, researchers can assess their effectiveness and identify areas for improvement. And again, decomposing the NEET rates by the main cause of their condition could help to distinguish the different types of actions they need to leave this condition and access to employment. Overall, further research in this field can contribute to the ongoing efforts to achieve the Sustainable Development Goals and reduce NEET rates in Italy. By evaluating the effectiveness of existing policies, exploring new interventions, and understanding the complexities of the labor market, researchers can assist policymakers in designing evidence-based solutions to enhance youth employment and promote a smooth transition from education to work.

References

De Luca, G., Mazzocchi, P., Quintano, C., Rocca, A. (2019). Italian NEETs in 2005–2016: Have the recent labour market reforms produced any effect? *CESifo Economic Studies*, **65**(2), pp. 154–176.

De Luca, G., Mazzocchi, P., Quintano, C., Rocca, A. (2020). Going behind the high rates of NEETs in Italy and Spain: The role of early school leavers, *Social Indicator Research*, **151**, pp. 345–363.

Eurostat (2010). *Public Employment in European Union Member States*, State Secretariat for the Public Service Directorate-General of the Public Service, Madrid.

Ghignoni, E. (2021). Informal recruitment channels, family background and university enrolments in Italy, *Higher Education*, **81**, pp. 815–841.

Linfante, G., Radicchia, D., Toti, E. (2023). Strumenti per l'attuazione dell'assessment. La profilazione quantitativa degli utenti dei Servizi per l'impiego, *Metodologie e approfondimenti*, ANPAL, Rome.

Mascherini, M. (2016). *Exploring the diversity of NEETs*, Eurofound, Dublin.

OECD (2020). *Public Employment Services in the Frontline for Jobseekers*, Workers and Employers, April.

Pacifico, D., Browne, J., Fernandez, R., Immervoll, H. (2018). *Faces of Joblessness in Italy: A People-Centred Perspective on Employment Barriers and Policies*, OECD Social, Employment and Migration Working Papers, 208.

Pastore, F., Quintano, C., Rocca, A. (2021). Some young people have all the luck! The duration dependence of the school-to-work transition in Europe, *Labour Economics*, **70**, 101982.

Pastore, F., Quintano, C., Rocca, A. (2022). The duration of the school-to-work transition in Italy and in other European countries: A flexible baseline hazard interpretation, *International Journal of Manpower*, **43**(7), pp. 1579-1600.

Quintano, C., Mazzocchi, P., Rocca, A. (2018). The determinants of Italian NEETs and the effects of the economic crisis, *Genus*, **74**(5). https://doi.org/10.1186/s41118-018-0031-0.

Sultana, R.G., Watts, A.G. (2005). *Career Guidance in Europe's Public Employment Services*, European Commission, Directorate-General for Employment, Social Affairs and Equal Opportunities Unit A3, October.

doi.org/10.26398/asaproc.0029

Public opinion and textual data analysis: Towards an automatic framing approach

Alberto Maria De Mascellis[a], Michelangelo Misuraca[b]
[a] DiSS, University of Naples Federico II, Naples, Italy
[b] DiScAG, University of Calabria, Rende Italy

1. Introduction

Facing the ever-increasing need to process large amounts of information in an automated way (Matthes and Kohring, 2008), the attention of social and political researchers has turned to a large branch of communication which concerns the automatic identification and the analysis in a collection of texts of *frames*. In public opinion studies, one of the most incisive definitions of *framing* is the one provided by Riker (1986): "framing is the central process by which the government or the journalists exert political influence among each other or on the public". However, a more operational definition comes from Entman (1993), who speculated that newspapers operate through *selection* and *salience*. In other words, anyone who provides information could "select the aspects of a perceived reality and make them more salient within a text, thus promoting a particular definition of a problem, a causal interpretation, a moral evaluation and/or a recommendation for the object described". Entman refined his view in a study dated 2003, streamlining and expanding the definition of framing: "framing is selecting and highlighting facets of events or problems and creating connections between them to promote a particular interpretation, evaluation and/or solution".

The reference to the connections/relationships of elements is particularly interesting because it implies that the frames encompassed in a text can be seen as complex structures involving multiple dimensions. Furthermore, this interpretative paradigm seems to be coherent with the definitions of frames typically used in the qualitative literature of political studies, in which frames are seen as *closed field of meaning* (Schütz, 1972), *interpretative package* (Gamson and Modigliani, 1989), or *boundary of expectations* (Goffman, 1993). An analysis of the reference literature about framing theory suggests that texts contain *framing judgments* in the form of specific keywords, stock phrases, stereotyped images, and favourite sources. Fairhurst (2005) talks about frames in terms of "a choice of language aimed at framing people's actions as if they were framed by a telescope". In contrast, Iyengar (1994) defines them as a "subtle alteration in the statement or presentation of both a judgment and problematic choices". From a quantitative viewpoint, one of the research paths that seem to be more interesting relies on the possibility of treating the frames as communities of terms that occur in a text, obtained by text mining techniques based on network analysis (Misuraca and Spano, 2020). We can claim that most human activities can be modelled as networks (or graphs), represented as nodes connected by some criteria (Patgiri et al., 2023).

This work intends to pave the way to an automatic framing approach based on community detection. Several aspects have to be considered. First, there is a lack of a universal definition of community since network structures cannot be mapped *a priori* without some form of arbitrariness. It is then necessary to consider techniques capable of inferring distinct partitions without referring to preconceived structures or ground truths. Moreover, as the complexity of the task increases with the amount of data, it is important to consider computational costs (Nicholls and Culpepper, 2001). Finally, once the frames have been obtained, it is important to establish a procedure that allows for the unique identification and correct interpretation of the partitions.

To test the effectiveness of our proposal, we carried out a preliminary case study by investigating a topic broadly discussed in Italy in the last years, the so-called "reddito di cittadinanza" (RC, citizenship income). After collecting the articles published from 2018 to

2023 by five prominent Italian newspapers, we mapped the frames used to present this topic to public opinion and their temporal evolution.

2. Materials and methods

To understand in which way RC has been released to public opinion and explore which frames have been eventually used by different counterparts representing conflicting interest groups, we collected the articles published by five Italian newspapers (Il Corriere della Sera, Il Giorno, Il Resto del Carlino, La Nazione, and La Stampa) in the period that goes from May 2018 to June 2023. The articles were selected from the *Nexis Uni* repository[1] with a relevance criterion and an imposed limit of one thousand documents per search. The newspaper's name, publication date, the section containing the article, the title and the full-text body have been retained from each original document. After filtering out the articles with a length lower than 200 characters (associated with short comments on video articles or advertisements) and articles with missing meta-data, the total number of documents in the collection amounted to 3891.

Since data embodied in texts are unstructured, it was necessary to pre-process the collection with several procedures:

- removal of special characters, punctuation, numbers, extra whitespaces, and URLs;
- recoding of specific terms in unambiguous forms (e.g., names of politicians or parties' names);
- lemmatization of the text to bring each term back to its basic form;
- removal of stop words with null analytical value (e.g., conjunctions).

It was also decided to reduce the heterogeneity of the texts by keeping only the lexical POS (parts of speech), specifically nouns, adjectives, and proper nouns. After structuring the articles as vectors and juxtaposing them in a lexical table, we built a *term*×*term* co-occurrence matrix and depicted the knowledge structure of the collection as a network.

One of the most used approaches to study networks is their deconstruction into sub-units, the so-called *communities* (Fortunato and Castellano, 2007), viewed as groups of nodes that have a strong internal structure but weak external connections. Several computational methods have been developed over time to identify communities. Considering the nature of the data to be explored, extracted from texts written in natural language, and to implement an automatic approach to frame analysis, we decided to use an algorithm that relies on the concept of modularity (Girvan and Newman, 2002). Aiming at assigning each object/keyword to a single community, these algorithms operate very differently with respect to the divisive algorithms that remove inter-community links and the agglomerative ones that instead merge the most similar nodes (Radicchi et al., 2004). These methods try to maximize an objective function and return the modularity of the formed partitions, i.e. a scalar value between -1 and 1 that measures the density of the edges within communities compared to those formed between them (Newman, 2006). In particular, we referred to an approach known as *Leiden* algorithm (Traag et al., 2019), which starts from an already consolidated approach known as *Louvain* (Blondel et al., 2019) overcoming some of its well-known problems, such as the presence of bad internal connections between the elements of a community. To guarantee well-connected communities, the Leiden algorithm essentially operates by considering the local movement of nodes and refining the initial partition. An aggregated network is obtained starting from $P_{refined}$ instead of P (as happens in the Louvain approach) so that the algorithm can identify high-quality partitions more effectively. In more specific terms, it starts with a singleton partition, where each node is in its own community and then joins the nodes locally.

We applied the Leiden algorithm to our collection and obtained the communities of meaning to highlight the different frames used by the newspapers. The analysis has been carried out in the `R` environment with the `igraph` library after pre-processing the data with `quanteda`. In

[1] www.lexisnexis.com

the following, some early findings of the strategy are presented and discussed.

3. Preliminary results

In Tab. 1, for each newspaper, we reported the total number of articles included in the analysed collection and the average length (in characters) of each article.

Table 1: Main figures on the analysed collection.

Newspaper	N. of articles	Avg. length
Corriere della Sera	764	894
Il Giorno	887	2645
Il Resto del Carlino	779	427
La Nazione	762	2294
La Stampa	851	3536

We noted a heterogeneous average length, ranging from 894 characters (Corriere della Sera) to 3536 (La Stampa). It should be noted that for La Stampa all the 851 articles available in the repository are concentrated in 2018–2019.

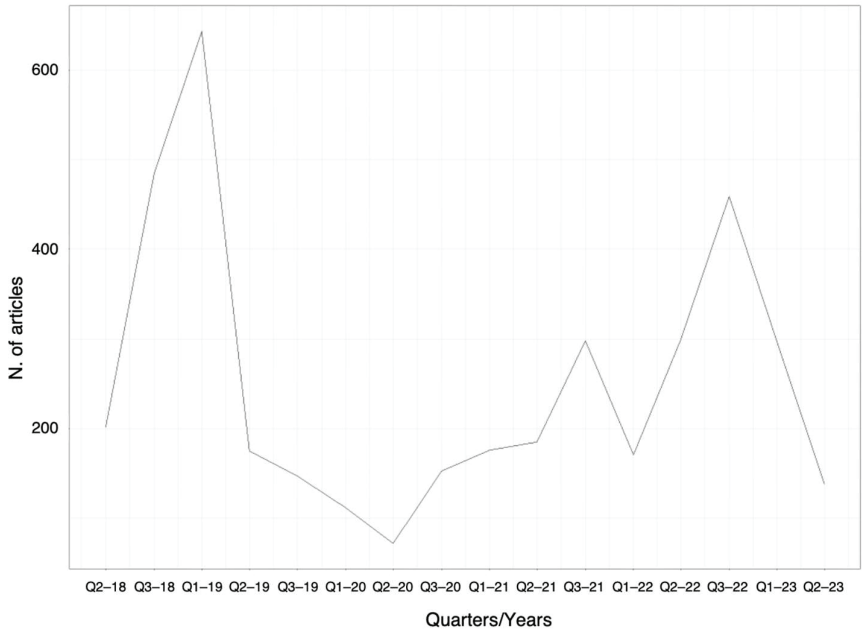


Figure 1: Year-wise distribution of articles about RC (2018–2023)

The chart reported in Fig. 1 is influenced by the latter anomaly, although it is still possible to note a peak in January 2019, the date of the Decree Law n. 4 that established the RC, and the declining curve of September-December 2022, after it has been revoked with the Budget Law (art. 1, subsec. 318). In Fig. 2, we can see the community detection results represented with a GEM force-directed graphic layout. We specify that the weights have been re-attributed based on membership for the sole purpose of an easier reading. The size of the nodes is scaled according to the modularity.

Figure 2: Graph of the main communities/frames about RC

We found multiple communities of meaning that can be structured as frames to guide the readers' debate. It is possible to group the communities in 4 different categories of frames:

- frames that could be referred to as the civil-social aspects of the RC, denoted by communities of terms such as *projects*, *inclusion* and *Puc* (projects useful to the community)
- which the beneficiaries of RC should be required to carry out), but also terms like *safeguard*, *cultural* and *artistic*;
- frames linked to the everyday-economic aspects, represented by the terms *family*, *child*, *age* and *pension*, but also *small-medium enterprises*;
- frames of political-economic nature, represented by all the names of politicians (*Di Maio*, *Salvini*, *Tria*) and by communities of a more complex economic nature, such as *Pil*, *deficit* and *manoeuver*, or *absolute*, *poverty* and *contrast*.
- frames of institutional-judicial type that refer to communities of terms such as *Inps*, *application* and *submission*, *controls*, *activities* and (job) *placement*, but also *substitutive* and *false* near *declaration*.

4. Final remarks and future developments

Bearing in mind that this is an embryonic proposal, the automatic approach to frame detection here introduced offers itself as a significant solution not only in terms of interpretation but also of the readability and immediacy of the outputs, even if they were to be provided to journalists and policy-makers, to broaden their communication and workflows. What we reported aligns with the results of other qualitative and quantitative studies on the surveyed subject. We highlighted, for example, how the contribution of the RC to *absolute family poverty* has been strongly debated, but also to RC as "a political problem, even before a policy one" (Vittoria, 2020). Other scholars described the problem as "individuals penalized in the access phase", in line with our outputs on the institutional dimension (Sgritta, 2020). In the interpretation phase, these data could also suggest communication gaps. For example, no relevant communities are referring to terms like *north* and *south*, while there are terms connected to the political-economic macro-category like *Italian* and *foreign* near *citizen*.

This work intends to be the first step towards an automatic approach to frame analysis, investigating in-depth the available alternatives and structuring a path that can account for a phenomenon as complex as framing. Further studies can certainly implement new data sources, new strategies in identifying communities and eventually a more effective way to connect the resulting frames more organically with the framing entities, such as newspapers or even the debate information in its entirety, taking into consideration other determining factors such as the temporal or spatial axis.

References

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, **2008**(10), P10008.

Entman, R.M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, **43**(4), pp. 51-58.

Entman, R.M. (2003). Cascading activation: Contesting the White House's frame after 9/11. *Political Communication*, **20**(4), pp. 415-432.

Fairhurst, G.T. (2005). Reframing the art of framing: Problems and prospects for leadership. *Leadership*, **1**(2), pp. 165-185.

Fortunato, S., Castellano, C. (2007). Scaling and universality in proportional elections. *Physical*

Review Letters, **99**(13), 138701.
Gamson, W.A., Modigliani, A. (1989). Media discourse and public opinion on nuclear power: A constructionist approach. *American Journal of Sociology*, **95**(1), pp. 1-37.
Girvan M., Newman M.E.J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, **99**(12), pp. 7821-7826.
Goffman, E. (1993). *Rahmen-Analyse: Ein Versuch über die Organisation von Alltagserfahrungen*. Suhrkamp Verlag.
Iyengar, S. (1994). *Is Anyone Responsible? How Television Frames Political Issues*. University of Chicago Press. Chicago (IL).
Matthes, J., Kohring, M. (2008). The content analysis of media frames: Toward improving reliability and validity. *Journal of Communication*, **58**(2), pp. 258-279
Misuraca, M., Spano, M. (2020). Unsupervised analytic strategies to explore large document collections, in *Text Analytics: Advances and Challenges*, eds. D.F. Iezzi, D. Mayaffre and M. Misuraca, Springer Nature, pp. 17-28
Newman, M.E.J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, **103**(23), pp. 8577-8582.
Nicholls, T., Culpepper, P.D. (2021). Computational identification of media frames: Strengths, weaknesses, and opportunities. *Political Communication*, **38**(1-2), pp. 159-181.
Patgiri, R., Deka, G.C., Biswas, A. (2023). *Principles of Big Graph: In-depth insight*. Academic Press.
Radicchi, F., Castellano, C., Cecconi, F., Loreto, V., Parisi, D. (2004). Defining and identifying communities in networks. *Proceedings of the National Academy of Sciences*, **101**(9), pp. 2658-2663.
Riker, W. (1986). *The Art of Political Manipulation*. Yale University Press.
Schütz, A. (1972). *Gesammelte Aufsätze: I Das Problem der sozialen Wirklichkeit*. Springer Science
Sgritta, G.B. (2020). Politiche e misure della povertà: il reddito di cittadinanza. *Social Policies*, **7**(1), pp. 39-56.
Traag, V.A., Waltman, L., Van Eck, N.J. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific Reports*, **9**(1), 5233.
Vittoria, A. (2020). La "scomparsa dei poveri". Una prima valutazione di policy sul Reddito di Cittadinanza. *Social Policies*, **7**(3), pp. 525-544.

doi.org/10.26398/asaproc.0030

Statistical analysis needs for an evolving wine sector

Valentina Di Chiara, Elena Maggio, Eugenio Pomarici
Department of Land, Environment, Agriculture and Forestry, University of Padua, Italy.

1. Introduction

Data and again data, nowadays in crucial sectors such as finance, transport, environment, medicine and healthcare statistics play a role of fundamental importance. In recent years, scholars have continuously applied statistical methods in various areas of the wine sector. Statistical tools are essential to analyse the wine market, to understand vine behaviour and improve the techniques for vineyard management and protection, to optimise winemaking processes. As wine is a fascinating product, several statisticians and quantitative economist not specialised in wine have dedicated part of their time to studies about wine.

In 1991 a group of "general" economists and statisticians founded the Vineyard Data Quantification Society, and created the neologism "enometrics", as the scope of the association was to "promote exchanges of research in theoretical and applied enometrics". As additional example, Orley Ashenfelter, professor of economics at Princeton University, received the title of Doctor Honoris Causa at the Bordeaux University on the 6th of April 2023, also as consequence of his long-lasting dedication to wine data analysis, started with the "ranking of the chateaux of Bordeaux" based on wine auction prices and not on "expert opinion" (Ashenfelter, 1988). However, the central theme of Ashenfelter's research has been the "Bordeaux equation", a model for predicting vintage quality of Bordeaux red wines based on wine's age and weather conditions during the growing season of the grapes (Ashenfelter et al., 1995). The Ashenfelter equation has paved the way for a new strain of research in wine and finance (Storchmann, 2012). Other scholars have addressed issues such as the impact of wine critics on wine prices (Combris et al., 1997; Hay, 2010; Masset et al., 2015; Kaimann et al., 2023), still others on the quality of wine, ideally related to intrinsic characteristics (Dodd et al., 2005; Carbonell et al., 2008; Rahman et al., 2014) such as visual, olfactory, and gustatory ones (Fabbris and Piscitelli, 2021).

Even today the wine sector can benefit from quantitative analysis, applying statistical methods in all areas of sector and within it. In order to outline the current needs and opportunities of statistical analysis for the wine sector, it is useful to sketch out the main features and issues that characterise this sector in his global dimension. The wine production is almost stable in the new century, fluctuating according to weather conditions in the different producing areas, between 295 and 250 million hl, with an average over the last 5 years of 270 million hl. Main producer countries are still France and Italy, usually with a production higher than 45 million hL, followed by Spain. These three countries together account for 51% of the world wine production in 2022. The USA are the fourth producer (on average 25 million hl), while Argentina, Chile, Australia, South Africa and Germany contribute to global wine supply each one with on average 10 million hl and other producers with smaller quantities.

Wine consumption worldwide now looks stabilised around 230 million hl; the forecast by Statista is of a moderate growth over the next three years but such positive evolution is threatened by the economic situation worldwide. Since the beginning of the century consumption grew from 225 up to 250 million hl immediately before the 2008 financial crisis; then it decreased and remained stable around 240 million hl between 2011 and 2017 when above all the sharp reduction of Chinese consumption brought the consumption to the current level. The consumption stagnation affects the international trade which since 2017 has stabilised around 110 million hl in volume, starting from 60 million hl in 2000. On the other hand, the value of export is still increasing and since the beginning of the century has tripled; the still increasing value is driven by the phenomenon called "premiumisation", the tendency

of consumers to move toward wines with higher price, and since 2021 by inflation.

The reduction of consumption while the production remains on average stable is fostering competition among players, forcing all producers to constantly improve their value proposition in order to defend their competitive advantage. The tougher competition between countries and single companies is occurring in a playground characterised by two main issues: social and institutional demand for environmental and social sustainability (Pullman et al., 2010; Santini et al., 2013; Pomarici and Vecchio, 2019; Forbes et al., 2020;) and climate change (Keller, 2010; Fraga et al., 2012; Cardell et al., 2019). As consequence, wine producers have to constantly improve their supply trying to satisfy increasingly sophisticated, and differentiated, consumers preferences, producing with processes able to reduce the environmental impact and managing the effect of a climate change; climate change that, on one hand, is modifying the characteristics of grape at the harvest and consequently the sensory profile of wines and, on the other hand, for the more frequent extreme meteorological phenomena make more unpredictable the quality and quantity of grape at the harvest, determining a considerable increase of the risk affecting the wine business.

Considered such scenario, this paper intends to highlight the different area where statistics may support the wine sector, proposing some examples of recent applications and focusing mainly on the digitalisation linked with two main topics: wine market and production.

The following part of the paper is structured as follows: Section 2 introduces an overview of the digitalization of data collection and new ways of data analysis in the wine sector. The main remarks in light of the existing literature are discussed in Section 3.

2. Digitalization and statistical applications in the wine sector

The wine sector can be supported in meeting new and old challenges by the continuous collection of vast amounts of data and information allowed by digitalisation thanks to proximal and remote sensing systems in the vineyard, monitoring devices in the winery and new market data collection systems. There is a broad range of enabling technologies currently available to the agriculture sector (Hansen et al., 2022) which have huge potential to increase food production and improve health and nutrition as well as maintain environmental integrity (Olum et al., 2020). Especially, these technologies offer the opportunity to collect data remotely with a greater resolution in space and time, then integrate, analyse and generate actions from these data in new and improved ways. To derive insights for making farming decisions these data require more sophisticated treatment and analyses.

Data collection and analysis can be implemented using different approaches, including wireless sensor networks internet of things (IoT), artificial intelligence (AI), remote sensing and machine learning (ML) or deep learning (DL). The combination of technologies and approaches and the subsequent sharing of data and information can contribute to efficient and sustainable farm management and create value along the entire supply chain. For example, data analytics, including AI and ML enables a more accurate information processing by converting raw data into a form suitable for decision making. Computational decision-support tools, instead, use data to develop recommendations for management and optimise farm tasks like more precise application of pesticides, efficient water and soil management, closer monitoring of the fermentation process, and more sustainable approach. Besides allowing frequent, real-time communication between all the actors of the supply chain, resources and computation resources the digital communication tools can be helpful digital tools.

According to the International Organisation of Vine and Wine (OIV, 2021), digitalisation in the vine and wine sector will bring great benefits in all different phases of the value chain: vineyard, winery and distribution. Benefits that can be the improvement of data collection at source (acidity, quality, weight, etc.), vineyard productivity and yield, product traceability, a more efficient use of data to increase productivity, and improvements leading to the introduction of new distribution and marketing channels.

The increase of available data can give a new impulse to the development of decision support system (DSS) opening up room for new applications of statistics in the wine sector. In fact, the use of statistical methods for decision support in the wine sector is very widespread (Cozzolino et al., 2009; Perrot et al., 2015; Ghiglieno et al., 2019; Vieira et al., 2023). Furthermore, statistics plays a fundamental role in constructing mathematical models, calibrating the parameters of the mathematical model, evaluating the goodness of fit to the data, estimating the margins of error associated with the predictions (Cozzolino et al., 2008). In this regard, the adoption of structure and standards based on statistical principles including a systematic assessment of model uncertainties and parametric sensitivities, can certainly help make them more reliable (Saltelli, 2019).

2.1. Wine market

Also, statistics can play a key role in helping industry members to understand consumer needs and preferences. Data mining approaches can be used to predict human wine taste preferences, as in the study by Cortez et al. (2009), in which the authors, developing a method that performs simultaneous variable and model selection using neural networks (NNs) and Support Vector Machines (SVMs) techniques, were able to predict taste preferences starting from data based on physicochemical analysis.

Big data can be used to understand consumer preferences and optimize marketing strategies, such as exploring the potential role of online channels in the valorisation of the country of origin of wines in international markets (Pegan et al., 2020). Through the statistical analysis of Google AdWords and Google Trends output some useful information about consumer preferences can be identified. In this regard, defining for example buyer personas by cross-referencing demographic and preference data can be useful in understanding consumer needs and optimizing social media campaigns (de Luca et al., 2019; Sottini et al., 2019).

In addition, in recent years, several platforms that can analyse the wine market were developed (e.g., Enolytics, BigDataGrapes project). Based on data collected from wine partner websites or from social networks of wine passionate users, these tools are able to identify the competitive potential of a wine, considering some variables such as price and sales area. This is possible thanks to a series of algorithms that are able to predict the penetration rate in a given market of a wine with specific characteristics by analysing current consumer behaviour (Neveu and Damasio, 2020). Statistics can also be used to support the operation of IoT devices. Indeed, the data collected by these tools need to be analysed and processed. One example is MEMORvINO, a tool developed by Wenda startup, that integrating near field communication (NFC) technology on a smart glass, is able to collect information about consumers and what they taste. The information collected is statistically analysed to be used to conduct targeted marketing communications and operations (Spadoni et al., 2019).

2.2 Wine production

To improve the efficiency and sustainability of the viticulture activities it is necessary to make a reliable prediction and undertaken strategic, tactical and operational decision-making (Ho, 2019). The digitization of information allows us to collect data faster. For instance, the vite.net system, which was developed by Horta through an integrated system of proximal or remote sensors, monitoring tools, and IoT technologies, is able to collect data related to the vineyard and the cultivation environment.

Statistical analysis and expert knowledge allow us to analyse and interpret the information and data collected, making reliable predictions, so bridging the gap between vineyard data and decision-making (Guimapi et al., 2020; Knowling et al., 2021). Consequently, winegrowers are able to implement more rational and informed decisions. In this context, Perrot et al. (2015) proposed a methodology to implement robust predictive tools enable to handle the several

factors involved in complex and coupled physiochemical and biochemical reactions, such as grape maturity indexes (i.e., sugar and total acidity), which are climate dependent. Laroche-Pinel et al. (2021) instead used the remote sensing in combination with Sentinel-2 sensors to perform supervised multiple regressions to achieve better predictions in monitoring water status for vineyards.

Considering the winemaking challenges imposed by climate change, statistics can assist researchers in defining decision-support tools that can help producers to reduce inputs and predict the evolution of wines over time. In fact, system dynamics models, based on sets of interrelated variables, are useful for wine producers to select appropriate production strategies. For example, the wine production is influenced by area under vines and yield per hectare, and, in turn, the amount of wine produced influences wine sales (Sá et al., 2022). While statistical approaches, such as random forest and support vector regression, or mathematical modelling are able to support winemakers in monitoring the fermentation process and predicting the evolution of fermentation kinetics or tannin and anthocyanin extraction kinetics have also been developed for the winery, based on historical data and taking into account the chemical and physical characteristics of the wine (Tindal et al., 2021). Instead, statistical tools, such as ridge regression or artificial neural network, can be applied in cellar to predict potential wine quality and on the base of its physicochemical characteristics identifying possible action to achieve quality goal, although their effectiveness in predicting the quality of a wine varies based on the data available (Dahal et al., 2021).

All these statistical tools can be useful in supporting the winemaker in optimizing the wine production process. However, when wine leaves the cellar other variables can play a role in influencing its quality. In this regard, statistical tools capable of evaluating the possible influence of exogenous variables on wine, such as packaging, storage and transport, could be useful as they can influence wine shelf-life.

3. Final remarks

In recent years, wine research has increasingly focused on the digitalization of information and data collection, through the implementation of new technologies and the use of the www for their storage, but some problems may arise when they have to be processed. The use of statistical methods allows us to bring together data from different sources, that have heterogeneous vocabularies, different measurement scales and modes of data production. At this point, statistics needs new approaches to be applied to structure, connect and exploit all available data simultaneously.

On the other hand, data are difficult to find. In particular, data collected through 4.0 technologies that are owned by wine companies are actually only privately exploited by technology companies and are not used instead by farmers to improve their performance. In this regard, to generate the data using a common language that allows them to be linked, they should be collected in accessible repositories and analysed clearly and automatically to facilitate knowledge generation. Finally, another key point in data processing is the speed with which the data must be processed. There is a clear need for real-time data processing to enable producers, in particular winemakers and oenologists, to be able to promptly act (Neveu and Damasio, 2020). The adoption of digitalisation in the wine sector is still at a low level of maturity but with a high growth rate and potential, and the availability of continuous data free from measurement errors opens new possibilities for the application of statistical methods.

References

Ashenfelter, O. (1988). A new objective ranking of the chateaux of Bordeaux (Based on auction prices, of course!). *Liquid Assets*, **5**, pp. 1–6.

Ashenfelter, O., Ashmore, D., Lalonde, R. (1995). Bordeaux wine vintage quality and the weather. *Chance*, **8**(4), pp. 7–14.

Carbonell, L., Izquierdo, L., Carbonell, I., Costell, E. (2008). Segmentation of food consumers according to their correlations with sensory attributes projected on preference spaces. *Food Quality and Preference*, **19**(1), pp. 71–78.

Cardell, M.F., Amengual, A., Romero, R. (2019). Future effects of climate change on the suitability of wine grape production across Europe. *Regional Environmental Change*, **19**, pp. 2299–2310.

Combris, P., Lecocq, S., Visser, M. (1997). Estimation of a hedonic price equation for Bordeaux wine: does quality matter?. *The Economic Journal*, **107**(441), pp. 390–402.

Cortez, P., Cerdeira, A., Almeida, F., Matos, T. Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, **47**(4), pp. 547–553.

Cozzolino, D., Kwiatkowski, M.J., Dambergs, R.G., Cynkar, W.U., Janik, L.J., Skouroumounis, G., Gishen, M. (2008). Analysis of elements in wine using near infrared spectroscopy and partial least squares regression. *Talanta*, **74**(4), pp. 711–716.

Cozzolino, D., Cynkar, W.U., Shah, N., Dambergs, R.G., Smith, P.A. (2009). A brief introduction to multivariate methods in grape and wine analysis. *International Journal of Wine Research*, **1**, pp. 123–130.

Dahal, K.R., Dahal, J.N., Banjade, H. Gaire, S. (2021). Prediction of wine quality using machine learning algorithms, *Open Journal of Statistics*, **11**(2), pp. 278–289.

de Luca, P., Pegan, G., Gonzalo-Penela, C. (2019). Insights from a Google keywords analysis about Italian wine in the US market. *Micro & Macro Marketing*, **28**(1), pp. 93–116.

Dodd, T.H., Laverie, D.A., Wilcox, J.F., Duhan, D.F. (2005). Differential effects of experience, subjective knowledge, and objective knowledge on sources of information used in consumer wine purchasing. *Journal of Hospitality & Tourism Research*, **29**(1), pp. 3–19.

Fabbris, L., Piscitelli, A. (2021). Wine preferences based on intrinsic attributes: A tasting experiment in Alto Adige/Südtirol province. In B. Bertaccini, L. Fabbris, A. Petrucci, (Eds.). *ASA 2021 Statistics and Information Systems for Policy Evaluation. Book of short papers of the opening conference*, Firenze University Press, Firenze, pp. 129–134.

Forbes, S.L., De Silva, T.A., Gilinsky, A. (2020). *Social Sustainability in the Global Wine Industry*. Palgrave Macmillan, Cham, CH.

Fraga, H., Malheiro, A.C., Moutinho-Pereira, J., Santos, J.A. (2012). An overview of climate change impacts on European viticulture. *Food and Energy Security*, **1**(2), pp. 94–110.

Ghiglieno, I., Simonetto, A., Donna, P., Tonni, M., Valenti, L., Bedussi, F., Gilioli, G. (2019). Soil biological quality assessment to improve decision support in the wine sector. *Agronomy*, **9**(10), p. 593.

Guimapi, R.A., Mohamed, S.A., Biber-Freudenberger, L., Mwangi, W., Ekesi, S., Borgemeister, C., Tonnang, H.E.Z. (2020). Decision support system for fitting and mapping nonlinear functions with application to insect pest management in the biological control context. *Algorithms*, **13**(4), p. 104.

Hansen, B.D., Leonard, E., Mitchell, M.C., Easton, J., Shariati, N., Mortlock, M.Y., Schaefer, M., Lamb, D.W. (2022). Current status of and future opportunities for digital agriculture in Australia, *Crop & Pasture Science*, **74**(6), pp. 524–537.

Hay, C. (2010). The political economy of price and status formation in the Bordeaux en primeur market: the role of wine critics as rating agencies. *Socio-Economic Review*, **8**(4), pp. 685–707.

Ho, N. (2019). Developing decision support systems for multiple knowledge domains: A proposed solution, *Journal of Computers*, **14**, pp. 339–348.

Kaimann, D., Bru, C. L.M.S., Frick, B. (2023). Ratings meet prices: The dynamic relationship of quality signals. *Journal of Wine Economics*, **18**(3), pp. 1–19.

Keller, M. (2010). Managing grapevines to optimise fruit development in a challenging environment: a climate change primer for viticulturists. *Australian Journal of Grape and Wine Research*, **16**, pp. 56–69.

Knowling, M.J., Bennett, B., Ostendorf, B., Westra, S., Walker, R.R., Pellegrino, A., Edwards, E.J., Collins, C., Pagay V., Grigg, D. (2021). Bridging the gap between data and decisions: A review of process-based models for viticulture, *Agricultural Systems*, **193**, p. 103209.

Laroche-Pinel, E., Duthoit, S., Albughdadi, M., Costard, A.D., Rousseau, J., Chéret, V. Clenet, H. (2021). Towards vine water status monitoring on a large-scale using sentinel-2 images, *Remote Sensing*, **13**(9), p. 1837.

Masset, P., Weisskopf, J.P., Cossutta, M. (2015). Wine tasters, ratings, and en primeur prices. *Journal of Wine Economics*, **10**(1), pp. 75–107.

Neveu, P. Damasio, C. (2020). Big data and artificial intelligence: Examples of applications for the vine and wine sector, *Internet Journal of Viticulture and* Enology, **6**(1), pp. 1–4.

OIV. (2021). *Digital Trends Applied to the Vine and Wine Sector: A Comprehensive Study on the Digitalisation of the Sector*, OIV publications, Paris. Available online at https://www.oiv.int/sites/default/files/2023-10/digital-trends-applied-to-the-vine-and-wine-sector.pdf. Accessed on August 18, 2023.

Olum, S., Gellynck, X., Juvinal, J., Ongeng, D., De Steur, H. (2020). Farmers' adoption of agricultural innovations: A systematic review on willingness to pay studies, *Outlook on Agriculture*, **49**(3), pp. 187–203.

Pegan, G., Vianelli, D., de Luca, P. (2020). Online channels and the country of origin. In G. Pegan, D. Vianelli, P. de Luca (Eds.). *International Marketing Strategy*. International Series in Advanced Management Studies. Springer, Cham, CH. pp.149-180.

Perrot, N., Baudrit, C., Brousset, J.M., Abbal, P., Guillemin, H., Perret, B., Goulet, E., Guerin, L., Barbeau, G., Picque, D. (2015). A decision support system coupling fuzzy logic and probabilistic graphical approaches for the agri-food industry: Prediction of grape berry maturity. *PloS one*, **10**(7), e0134373.

Pomarici, E., Vecchio, R. (2019). Will sustainability shape the future wine market? *Wine Economics and Policy*, **8**(1), 1–4.

Pullman, M.E., Maloni, M.J., Dillard, J. (2010). Sustainability practices in food supply chains: how is wine different? *Journal of Wine Research*, **21**(1), pp. 35–56.

Rahman, I., Stumpf, T., Reynolds, D. (2014). A comparison of the influence of purchaser attitudes and product attributes on organic wine preferences. *Cornell Hospitality Quarterly*, **55**(1), pp. 127–134.

Sá, J., Pinto Ferreira, L., Dieguez, T., Sá, J., Silva, F. (2022). Industry 4.0 in the wine sector – Development of a decision support system based on simulation models. In Machado, J., Soares, F., Trojanowska, J., Ivanov, V. (eds) *Innovations in Industrial Engineering.* Springer, Cham, pp. 371–384.

Saltelli, A. (2019). A short comment on statistical versus mathematical modelling. *Nature Communications*, **10**(1), p. 3870.

Santini, C., Cavicchi, A., Casini, L. (2013). Sustainability in the wine industry: key questions and research trends. *Agricultural and Food Economics*, **1**(1), pp. 1–14.

Sottini, V. A., Barbierato, E., Bernetti, I., Capecchi, I., Fabbrizzi, S., Menghini, S. (2019). Winescape perception and big data analysis: An assessment through social media photographs in the Chianti Classico region. *Wine Economics and Policy*, **8**(2), pp. 127-140.

Spadoni, R., Nanetti, M., Bondanese, A., Rivaroli, S. (2019). Innovative solutions for the wine sector: The role of startups, *Wine Economics and Policy*, **8**(2), pp. 165–170.

Storchmann, K. (2012). Wine economics. *Journal of Wine Economics*, **7**(1), 1–33

Tindal, R.A., Jeffery, D.W., Muhlack, R.A. (2021). Mathematical modelling to enhance winemaking efficiency: a review of red wine colour and polyphenol extraction and evolution, *Australian Journal of Grape and Wine Research*, **27**(2), pp. 219–233.

Vieira, A.A., Figueira, J.R., Fragoso, R. (2023). A multi-objective simulation-based decision support tool for wine supply chain design and risk management under sustainability goals. *Expert Systems with Applications*, **232**, 120757.

doi.org/10.26398/asaproc.0031

Geostatistics for the analysis of complex ecosystems

Angela Maria Digrandi [a], Pasquale Cimmino [b], Germano Monteleone [c]
[a] CNR IRISS, Naples
[b] Istat, Naples
[c] Tales-Alenia Space, Rome

1. Introduction

The paper aims to describe the interaction between statistics and geo-data in a context of continuous, capillary, sometimes dissonant, information, where many phenomena risk losing concreteness concerning the time and territory in which they happen. To process information that comes from big data, it is necessary to build methodologies of coherence between traditional sources and sources from the digitalized world in which there are no shared data archiving rules, standardized languages, and metadata so AI systems can adapt information processing. But can predictive algorithms become the new way of knowing and deciding? Are they transparent, objective, and scientifically robust? Do they meet the requirements of official statistics in terms of ontologies and criteria for disseminating open data?

In this paper we present the preliminary results of a study that starts from geostatistics of ecoregions and uses data captured by satellites on phenomena that in the past were obtained only through official statistics (mobility, pollution, land use). At state-of-the-art, adequate statistical tools for analyzing intangible infrastructures are not available as the scalar interpretation provided by official statistics is based on the structure of the subdivisions of administrative borders. At the same time, the industrial districts had taught us (and the Covid-19 epidemic has confirmed) that economic, social, and cultural human relations follow network designs that do not depend on administrative boundaries.

Usually, the territory is analyzed according to altimetry, or by differentiating between coastal municipalities and inland ones, or by demographic size. Our analysis allows us to take a step forward in the construction of complex ecosystem analyses that offers statisticians and policymakers who are not experts in geology and climatology the possibility of interpreting the impact of nature on the anthropic environment in a more conscious way.

Moreover, it helps nature experts, say, botanists, geologists, and climatologists, to deeply understand the social impact of their studies. Indeed, in a climate that changes the nature of places, following the impact of extreme events on the geo-physical structure of the territories requires new accounting tools for intangible investments and more efficient satellite monitoring systems.

2. Measurement of socio-economic and environmental phenomena

New economic analysis models are needed to guide policies, incorporate relational intangible investments into corporate accounting, and measure investments and use of intangible capital visible in the income statement management. The difficulty that emerges in the statistical measurement of intangible investments in the field of knowledge and innovation is even higher in the field of relational development of territorial intangible investments.

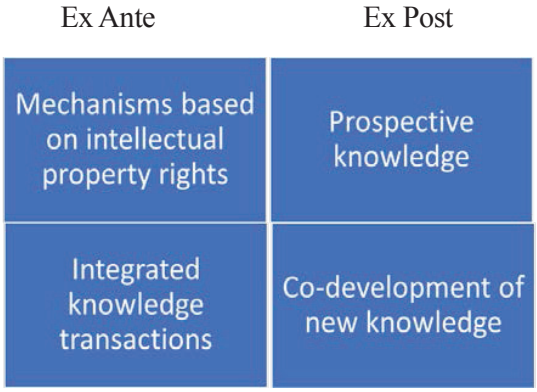
The four quadrants that are represented in Figure 1 make explicit the flows to which it is necessary to assign human and economic resources. It is configured as an explanatory framework of the management and of the results achieved with the specific management and with the available resources.

As a starting point for an interdisciplinary reflection on intangible assets, we propose a

subdivision scheme of economic activities and services (products or purchases) into two subsets: existing knowledge and prospective knowledge (Figure 2).

Figure 1: Intangible services: ex-ante and ex-post

Divided into mechanisms based on intellectual property rights and mechanisms of incorporation into the transaction relating to other goods and services not explicitly connected to innovation and knowledge.



Divided into agreements where companies/institutions hire an external company to provide personalized knowledge and agreements where both parties contribute to the development of a knowledge product.

Figure 2: Economic items to be included in company financial statements to measure investments in intangible assets.

Existing knowledge	**Prospective knowledge**
Mechanisms based on intellectual property rights	**Search for knowledge solutions**
• Confidentiality agreements • Intellectual property license • Pooling agreements for intellectual property • Sale or assignment of intellectual property rights • Inclusion of intellectual property in franchising agreements • Know-how contracts (transfer in tangible form through technical data)	• Consultancy services • Search services • Crowdsourcing rewards for search results.
Integrated knowledge transaction	**Co-development of new knowledge**
• Transfer of intellectual property rights and other knowledge-based capital through mergers and acquisitions of equipment • Turnkey project agreements (delivery of a plant with ready-to-use embedded technology) • Agreements for transfer/use of materials and data	• Co-development programs • Research joint venture • Research/commercialisation alliances • Temporary secondments to share or exchange personnel • Network Membership Agreements

In recent years, the awareness for environmental matters has activated financial instruments aimed at the conservation and enhancement of ecosystems: the payment for ecosystem services (PESs). The research project proposes the economic evaluation of one of the main ecosystem services: the sequestration of CO_2 by the forest heritage. Covering about a third of the earth's surface, today the forest heritage represents one of the largest and most biologically diversified ecosystems on the planet, enclosing, in fact, 80% of the world's biodiversity. This biodiversity is a

source of essential services and benefits that the forest heritage provides to the environment and man, thanks to the ability to absorb and store the CO_2 present in the atmosphere.

Public accounting must therefore initiate serious reasoning regarding payments for ecosystem services. Although the main objective is the protection of the territory and its resources, PES schemes can help reduce inequalities and poverty and improve the production system, in close connection with the objectives of the 2030 Agenda. The application of this tool is the economic evaluation of ecosystems and the services they perform. Of course, the entire economic evaluation process starts with the collection and processing of data regarding the forest categories of the analyzed area, the estimate of annual removal and storage of organic carbon, and its conversion into CO_2. The data is currently made available by the forest register, which will be updated in 2025, and the economic values of the CO_2 capture by forest category and hectare of soil are available from studies disseminated in the literature. It is possible to derive an estimate of the monetary value of the stock of CO_2 absorbed annually and the assessment process based on the cost-benefit scheme can be set up on this. This is a first step towards PES.

3. A linkage between satellite and statistical data

Earth ecosystems are constantly changing due to natural and atmospheric conditions and under the pressure of human activities and need to be updated over time.

The use of advanced statistical systems in earth monitoring satellites is a crucial element in guaranteeing the acquisition and optimal processing of collected data. These statistical systems make it possible to face the challenges related to the vastness and complexity of terrestrial information, providing a solid basis for the analysis and interpretation of the spatial and temporal variations of the monitored phenomena. The statistical systems used in these contexts include a wide range of techniques, methods, and models aimed at processing, analyzing and interpreting the data collected by satellites to obtain significant information on the terrestrial environment. Some key aspects of the statistical systems used include:

1. Multiple data fusion from different sources; statistical systems combine and integrate heterogeneous data to create a complete and accurate picture of environmental conditions;
2. Spatial and temporal analysis is used to examine how environmental phenomena vary over space and time, allowing you to identify trends, cycles and seasonal patterns;
3. Regression models are used to identify correlations between deforestation and climate change;
4. Time series analysis reveals seasonal patterns, long-term trends or outliers, helping to understand environmental changes;
5. Classification and anomaly detection using machine learning techniques;
6. Uncertainty estimation provides a more accurate evaluation of the information.

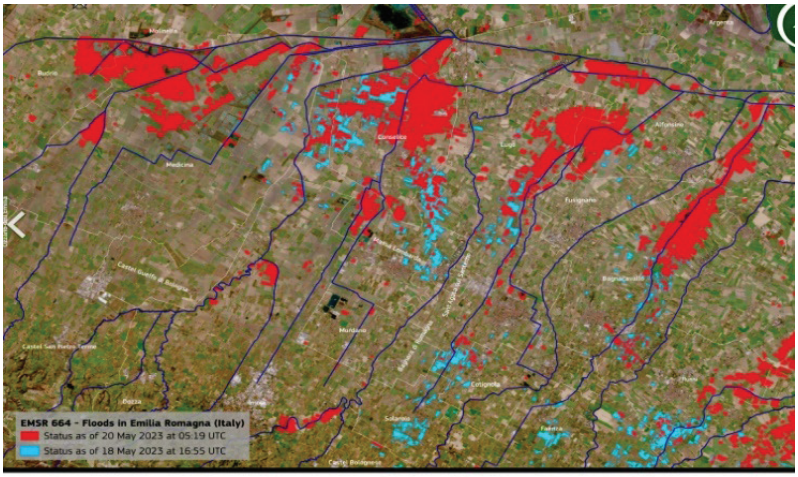
The use of statistical systems in earth monitoring satellites plays a key role in facilitating informed decision-making, contributing to a better understanding of our planet and our environment. The Copernicus satellite program, an initiative of the European Union, represents an important milestone in terrestrial monitoring (https://dataspace.copernicus.eu/). It offers an integrated earth observation system through a constellation of dedicated satellites, known as "sentinels", in synergy with an extensive network of in situ sensors.

The Sentinel-1 satellite is one of a pair of satellites consisting of Sentinel-1A and Sentinel-1B. These are aperture synthesis radar satellites (SAR) which allow monitoring of river waters both day and night and in all weather conditions. Thanks to the ability to penetrate through clouds and rain, Sentinel-1 provides high-resolution images of flooded areas or streams with detailed information on hydrological aspects, such as water distribution, river morphology and level of floods. This information is essential for water resources management, flood monitoring and evaluation of river dynamics. Sentinel-2: The Sentinel-2 satellite consists of two satellites,

Sentinel-2A and Sentinel-2B, which offer high-resolution multispectral optical imaging. These satellites are used for monitoring river waters by observing changes in land cover and hydrological characteristics. Sentinel-2's multispectral imagery helps identify and analyze flood-affected areas, surface water distribution, vegetation, and land use surrounding waterways.

These data are essential for monitoring erosion, river course variations and changes in the river ecosystem, providing crucial information for sustainable water resource management, flood prevention (Figure 2), water resources planning, and the understanding of environmental dynamics at the river level.

Figure 2: Map of floods in Emilia Romagna. May 1-3 and 16-17, 2023



Source: dataspace Copernicus

The use of advanced detection and analysis technologies, together with statistical models and data fusion systems, allows us to obtain a clear and complete vision of the environmental dynamics of our planet.

The combined use of data from the Sentinel-1 and Sentinel-2 satellites allows for a more complete and detailed view of river waters, providing crucial information for sustainable water resource management, flood prevention, water resources planning, and the understanding of environmental dynamics at the river level.

Thanks to the multidisciplinary approach, Copernicus manages to involve different communities, including scientists, public authorities, businesses, and citizens, guaranteeing free and open access to information. The use of advanced detection and analysis technologies, together with statistical models and data fusion systems, allows us to obtain a clear and complete vision of the environmental dynamics of our planet.

4. Synthetic land consumption indicators for hydro-climatic risk measurement

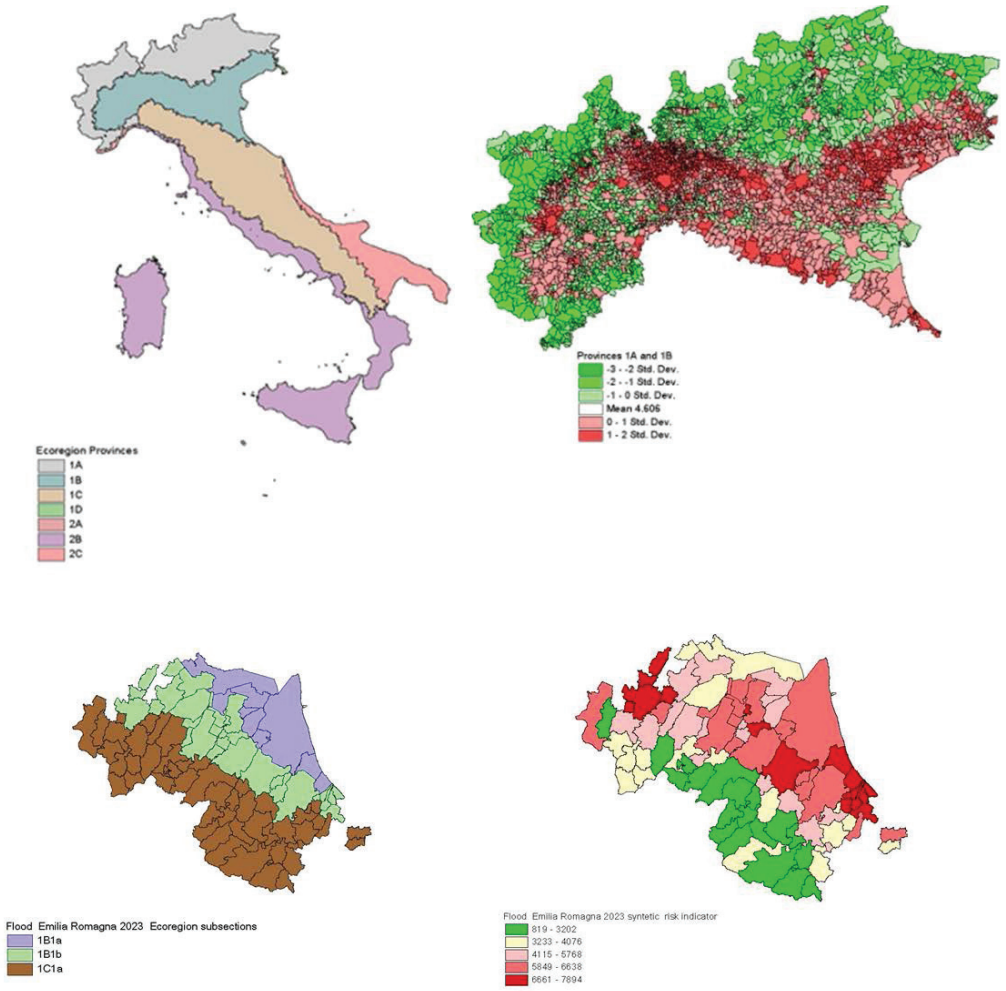
A synthetic index is a mathematical combination (or aggregation) of a set of elementary indicators (variables) that feature the different components of a multidimensional concept to be measured. In our case, we intend to measure the water-climate risk which cannot be captured by a single indicator. As with other complex indicators, its construction has proved to be complex. In particular, the validation process takes the form of an iterative path both as regard the data to be analyzed and the methodology. There are numerous techniques whose use is facilitated in terms of subsequent tests by the use of the Istat I.ranker product that allows the construction of co-graduation and correlation matrices between the indices (https://i.ranker.istat.it/)

The Subsections 1B1a and 1B1b belong to the same section of the ecoregions. The sub-section 1C1a belongs to the provinces 1C, the Toscana and Emilia-Romagna Apennine.

Subsection 1C1a is almost totally characterized by terrigenous sediments. Lagoon Subsection 1B1a is very flat, reaching a maximum altitude of 15 m, with an almost clastic and organic sedimentary lithology. The Central Plain Subsection 1B1b is characterized by a varied litho-morphology, with terraced clastic plains and smooth hills in the high Po River floodplain, rising to 600 m.

The various lithological structure of the soil in the three subsections affected by the flood of May 2023 was subjected to high anthropic pressure concentrated on a highly fragile territory, as demonstrated by the results of the hydro-climatic risk indicator developed by the authors. The three subsections reacted differently (landslides or flooding) about to the specific lithological structure.

Figure 3: Ecoregions, provinces, subsections, synthetic hydro-climatic risk indicator flooded areas in May 2023 in Emilia Romagna, Italy (analysis of Ispra-http://www.consumosuolo.isprambiente.it and Istat data)



The capillarity of the georeferenced data (municipal limits are updated to the year 2020) and the use of indices as basic data (density of local units and employees, of resident population, of employees in the tourist services sector, percentage consumption of municipal land) suggest, after various simulations, to choose to the synthetic indicator of the rankings which proved to be consistent with the theoretical framework.

The synthetic hydro-climatic risk indicator highlights the territories in which the impact of man is concentrated (Figure 3), determining areas of very high risk that must be connected,

through the reading of the geography of the ecoregions, to the geophysical and climatic characteristics which are profoundly different and therefore require safeguard approaches differentiate. The need in the policy for new governance to reduce the climate crisis must be calibrated considering these characteristics.

5. Conclusion

The new paradigm is reading the territory not only by altimetry levels, or demographic size but interpreting the impact of nature on the anthropic environment more consciously and adopting a methodology for sharing knowledge. We put forward that new accounting tools for intangible investments and more efficient satellite monitoring systems can help us. It is an approach of integration between information on the natural and geoclimatic structure of the territory and the structural and functional characteristics of human and productive settlements, reconnecting the material infrastructures with social and immaterial ones.

The tools of geo-statistics support the scientific skills of experts in the natural structure of the earth, vegetation, and ongoing climate change, by integrating them with the results of remote monitoring of natural and social phenomena taking place on the planet. The linkage between satellite data and official statistical data produces enormous quantities of information that data scientists also integrate through AI tools.

The ability to know and apply statistics to concrete reality must be considered as one of the fundamental functions of knowledge at all levels of education, from elementary to the various methods of continuing adult education, and geostatistics can assume the role of facilitator of statistical learning for the definition of scenarios within the territories on which public policies and collective and individual choices must act to affect the future of planet.

References

Blasi, C., Capotorti, G., Copiz, R., Guida, D., Mollo, B., Smiraglia, D., Zavattero, L. (2018*). Terrestrial Ecoregions of Italy. Map and Explanatory notes.* Global Map srl., Firenze.

Chakrabarty, D. (2021). C*lima, Storia e Capitale*. Nottetempo Edizioni.

Digrandi A.M., Persico P., Quagliuolo M. (2022). *Nuove tassonomie sui beni culturali come infrastruttura complessa di area vasta.* CNR Collana: Studi e Ricerche per l'innovazione.

Jaakkola, T. (2019). *Artificial Intelligence Course*, M2U1V2, February 27, 2019.

Istat (2019). *Classificazione dei comuni secondo le Ecoregioni*, Istat, Roma.

Mazziotta, M., Pareto, A. (2020). *Gli indici sintetici*. Torino: Giappichelli.

OECD (2018). *Oslo Manual*. https://www.oecd.org/science/oslo-manual-2018-9789264304604-en.htm.

doi.org/10.26398/asaproc.0032

Bioimpedance data analysis to identify electrical frequencies relevant for monitoring fruit ripeness

F. Marta L. Di Lascio[a], Pietro Ibba[b], Selene Perazzini[c], Luisa Petti[b]
[a] Faculty of Economics and Management, Free University of Bozen-Bolzano, Italy.
[b] Faculty of Engineering, Free University of Bozen-Bolzano, Italy.
[c] DMS Statlab, Department of Economics and Management, University of Brescia, Italy.

1. Introduction

Bioimpedance is the ability of a biological tissue to impede a flow of electric current. Bioimpedance analysis is gaining prominence in the field of food engineering due to its ability to detect changes in the electrical properties of biological tissues, as they allow us to gain insights into their underlying physio-chemical modifications without causing damage to the food. Consequently, bioimpedance analysis has emerged as a powerful tool for effectively managing products from the point of harvest to transportation, thereby potentially mitigating food waste.

In the context of fruit ripening, bioimpedance measurements are typically conducted across a range of frequencies, known as electrical impedance spectroscopy (EIS). Diverse electrical frequencies penetrate cell membranes differently, and monitoring a range of frequencies over time allows us to capture multifaceted changes in the physiochemical properties. Consequently, bioimpedance data constitute large three-way panel datasets, with a total of N fruits observed over T time intervals at J electrical frequencies. For each frequency, it is possible to identify a change point indicating the moment at which the fruit ripening undergoes a transition. However, bioimpedance time series for nearby electrical frequencies often exhibit considerable similarity. An intriguing aspect pertains to the identification of frequency groups that can effectively capture the multifaceted aspects of the ripening process while avoiding redundancy of information.

In this work, we address this concern by applying a clustering method to an innovative data set comprising bioimpedance measurements of 150 bananas collected over an 11-day period. The data set includes measurements obtained from two multi-frequency EIS devices: a bench-top "Impedance Analyzer" (IA) and a portable device known as "FruitMeter" (FM). The IA and FM devices have been observed to yield slightly distinct bioimpedance measurements, despite aiming to capture the same phenomena. Notably, the bench-top EIS device is bulky and more powerful than the FM, resulting in higher costs, increased maintenance requirements, and a greater need for technical expertise. Conversely, the FM is compact, lightweight, portable, and adaptable to various situations and applications. Ongoing discussions revolve around the potential substitution of bench-top instruments with portable ones. Therefore, this study also aims to contribute to the comparison between the performance of the two instruments. To achieve this objective, we separately analyze the data obtained from each device. As a preliminary step, we employ the three-way panel threshold regression model recently proposed by Di Lascio and Perazzini (202x) to eliminate serial dependence within the bioimpedance time series of the bananas. Subsequently, we perform a cluster analysis on the residual bioimpedance time series. This allows us to i) identify clusters of frequencies that provide similar information regarding the ripening process and ii) compare the two EIS devices, helping to define the optimal way to detect fruit bioimpedance.

The remainder of the paper is organized as follows: Section 2 describes the methodology, Section 3 presents the empirical analysis, and Section 4 discusses the results.

2. Methodology

The purpose of this work is twofold: we want to assess if there are frequencies leading to similar information on fruit bioimpedance and compare two different EIS devices. To this aim we perform a cluster analysis on bioimpedance panel data and propose a two-step analysis: the first step involves eliminating the serial dependence within the fruit bioimpedance time series and the second step concerns a cluster analysis of the bioimpedance residual time series.

Three-way panel threshold regression. Since bioimpedance time series exhibit a point of change that indicates a switch in the ripening process, we employ the three-way panel threshold regression model (Di Lascio and Perazzini, 202x) to remove serial correlation. Following the empirical analysis conducted by Di Lascio and Perazzini (202x), we model the bioimpedance y_{ijt} observed for the i-th banana over the t time and the j electrical frequencies as follows:

$$y_{ijt} = \phi_{1j}\mathbb{I}\{y_{ij(t-1)} \leq \gamma_j\} + \phi_{2j}\mathbb{I}\{y_{ij(t-1)} > \gamma_j\} + \varepsilon_{ijt} \quad (1)$$

where ϕ_{1j} and ϕ_{2j} represent the intercepts of the lower and upper regime, respectively, for the j-th frequency, the indicator function $\mathbb{I}$ captures the regime change, determined by the threshold parameter γ_j, which can vary across frequencies, the lagged dependent variable $y_{ij(t-1)}$ serves as the endogenous transition variable, and $\varepsilon_{ijt} = \mu_i + \lambda_j + \upsilon_{ijt}$ represents the error term. The error term is comprised of the unobserved individual fixed effect μ_i, the unobserved third-way fixed effect λ_j, and the zero-mean random error υ_{ijt} satisfying the condition $E(\upsilon_{ijt-1}) = 0$, where t represents a natural filtration at time t. The model is estimated using the two-step first-differenced generalized method of moments (GMM) estimator developed in Di Lascio and Perazzini (202x). The estimation algorithm can be summarized in the following methodological steps:

1) the first difference transformation is applied to the model in Eq. *(1)*;
2) a set of instrumental variables is selected; in our application, we choose the lagged values of y_{ijt};
3) the first-step GMM estimation is performed:
   a) the weighting matrix for GMM estimation is defined based on the chosen instrumental variables;
   b) a set of possible values of the threshold parameter γ_j is identified for each j based on the empirical distribution of the threshold variable;
   c) using a grid search algorithm over the set of possible values of γ_j, the GMM objective function is minimized, and the first-step regime-specific and threshold parameters are computed for each frequency j separately;
   d) the first-step first-differenced residuals are computed;
4) The second-step GMM estimation is performed:
   a) the second-step GMM weighting matrix is constructed using the first-step first-differenced residuals;
   b) the grid-search algorithm outlined in step 3c) is reiterated to estimate the second-step GMM threshold parameters $\widehat{\gamma_j}$ and the first-differenced parameters $\widehat{\beta_j}$ with $\beta_j = (\phi_{1j} - \phi_{2j})$;
   c) the second-step first-differenced residuals $\Delta\varepsilon_{ijt}$ are computed.

Cluster analysis. We consider the first-differenced residuals $\Delta\varepsilon_{ijt}$ and summarize the information of the batch of bananas by summing them over i and obtaining 9 observations for each frequency j. To identify possibly homogeneous frequency clusters from bioimpedance measurements, a hierarchical clustering algorithm (see, e.g., Kaufman and Rousseeuw, 1990) has been used.

Hierarchical clustering requires a pairwise dissimilarity measure to compare singletons and a linkage rule to compare clusters. As for the latter, we employ the complete linkage rule since its superiority over other linkages, such as the single and the average ones, has been found repeatedly in the literature. As for the dissimilarity measure between two frequencies j and j', we use the following measure based on Kendall's τ correlation coefficient: $d_{jj'} = \sqrt{1-\tau}$, which captures the comonotonic dependence relationship between the two compared frequencies. As it is well-known, hierarchical clustering results in a nested structure that does not require choosing a priori the number of clusters, say K. To this aim, we employed the well-known Dunn and Silhouette index (see, e.g., Kaufman and Rousseeuw, 1990).

3. Empirical analysis

The dataset used in this study comprises innovative bioimpedance measurements obtained from a production batch of 150 bananas over an 11-day period. The experiment was conducted by the Sensing Technologies Laboratory at the Free University of Bozen-Bolzano as part of an interdisciplinary research project involving the Faculty of Economics and Management and the Faculty of Engineering.

Bioimpedance was measured using specialized instruments known as bioimpedance analyzers. In this particular experiment, two analyzers were employed: a bench-top EIS device referred to as IA, which is a commercially available equipment, and a portable custom-made device named FM (Ibba et al., 2021). Both analyzers utilized two electrodes placed on the surface of each fruit to measure bioimpedance across a range of electrical frequencies (Ibba et al., 2020). An alternating current was applied through one electrode, and the resulting voltage measured at the other electrode was used to calculate the bioimpedance at different frequencies. For more technical details regarding the experiment, please refer to Ibba et al. (2020) and Ibba (2021).

The bioimpedance measurements provided in the dataset consist of 189 frequencies ranging from 20 to 13,668,764.956 Hz for the IA and 181 frequencies from 10 to 99,000 Hz for the FM. It should be noted that although the frequency values for the IA and FM are similar, they are not identical. All the fruits in the batch were harvested simultaneously and marketed together, ensuring the homogeneity of the initial stage of ripening. The measurements were conducted daily for a total of 11 days, until the fruits visibly deteriorated in order to monitor the progression of ripening under controlled room temperature and humidity.

The hierarchical clustering procedure described in Section 2 was applied to the residuals of the three-way panel regression model separately for the IA and FM data. To ease interpretability, the maximum number of clusters was set to 8, and the optimal number of clusters K was determined based on the Dunn and Silhouette indexes. The top plots of Figure 1 show that for the IA, the Dunn index suggests $K = 2$, while the Silhouette index indicates $K = 7$. Since $K = 2$ is a local maximum of the Silhouette index and the latter does not display a considerable difference between the two considered solutions (namely 0.50 for $K = 7$ and 0.49 for $K = 2$), we select the 2-cluster solution that also leads to more easily interpretable clusters. Regarding the FM, both the Dunn and Silhouette indexes agree on $K = 3$ (Figure1). The selection of the optimal number of clusters is further supported by the dendrograms depicted in Figure 2.

Figure 3 presents the partitions of frequencies obtained for the two EIS instruments. Note that the plot reports the logarithm of the frequencies for readability. The obtained clusters can be summarized as follows: IA frequencies within the range 19,070 Hz to 268,390 Hz (approximately) belong to cluster 2, and the other frequencies to cluster 1. For the FM, three subsequent clusters emerge: the first cluster comprises frequencies approximately below 6,000 Hz, the second cluster includes frequencies from 6,000 Hz to less than 24,000 Hz, and the third cluster encompasses frequencies of roughly 24,000 Hz and above.

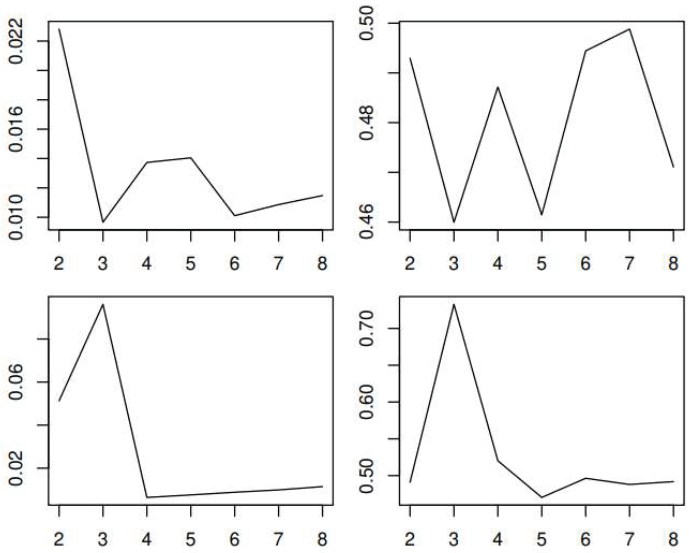


Figure 1: The x-axis indicates the number of clusters, the y-axis reports the index value. Left: Dunn index; Right: Silhouette index. Top: IA. Bottom: FM.

Comparing the partitions obtained from the two instruments reveals a certain correspondence between the frequency clusters. Particularly, IA cluster 2 exhibits similarities with FM cluster 3. Additionally, it is important to note that the second cluster includes electrical frequencies that may have been slightly affected by a technical issue during the FM experiment. Minor changes in electrode placement could have impacted the data collection. Despite this, the technical issue did not compromise instrument reliability, and the corresponding frequencies were not excluded from the analysis. However, considering this, the similarity between clusters 1 of the two instruments becomes more evident.

These findings suggest a strong correspondence between the measurements obtained from the IA and FM instruments. Notably, the most significant result pertains to the highest frequencies, exclusively measured by the IA. Bioimpedance measurements collected at frequencies greater than 268,390 Hz yield similar information to those below 19,070 Hz. This indicates that using higher frequencies is unnecessary, providing further evidence supporting the FM instrument.

4. Conclusions

Through the hierarchical cluster analysis of bioimpedance data collected from two different EIS instruments, the IA and FM, we identified meaningful clusters of electrical frequencies that capture the multifaceted aspects of the ripening process. This cluster analysis allowed us to group similar frequencies and select informative frequency clusters that can be used for fruit quality control.

Furthermore, the correspondence between the frequency clusters obtained from the IA and FM instruments suggests a good agreement between the EIS devices in measuring fruit bioimpedance. These findings support the potential of the portable FM instrument as a reliable alternative to the bench-top IA.

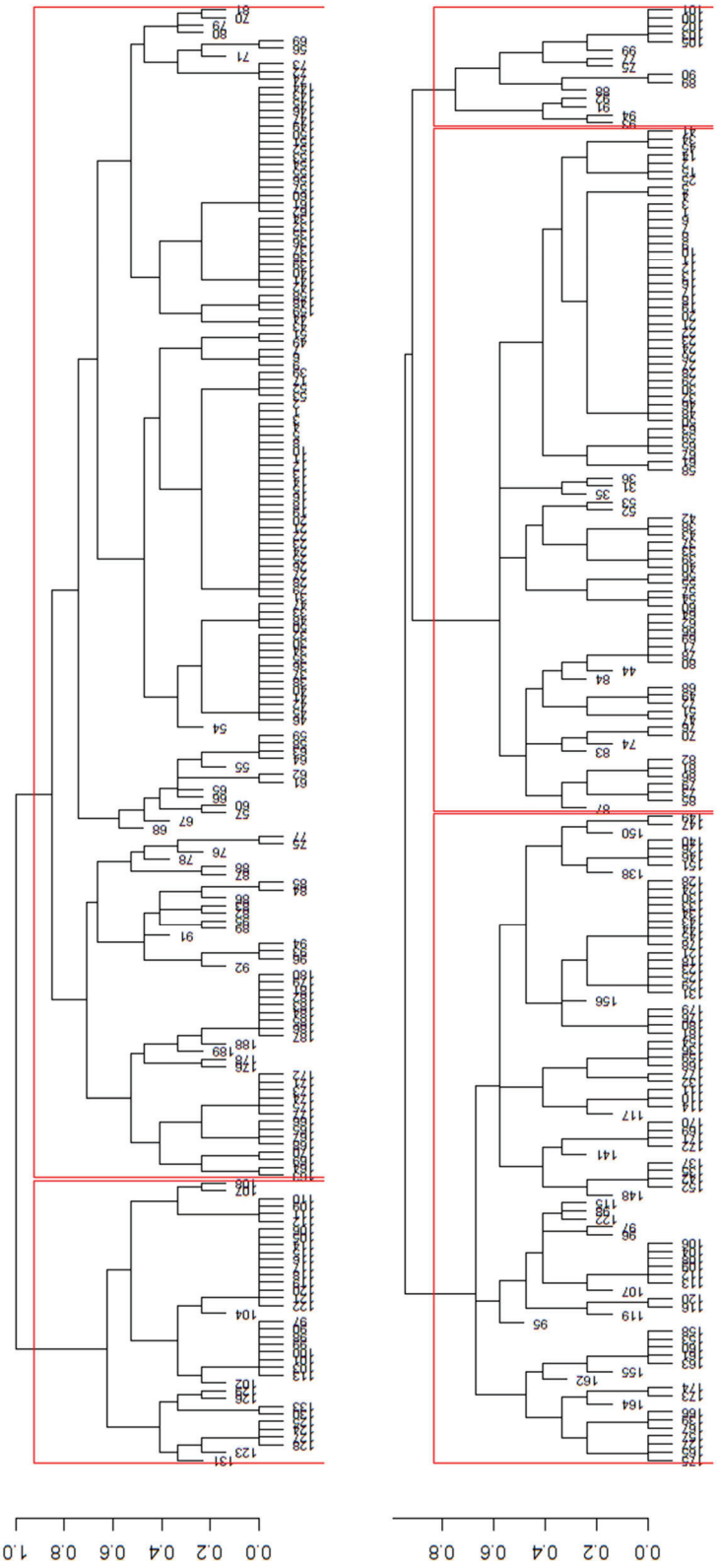


Figure 2: Dendrograms of the hierarchical clustering applied to the electrical frequencies used for measuring bioimpedance with two EIS devices: IA (top) and FM (bottom). The red lines outline the selected partitions of the data

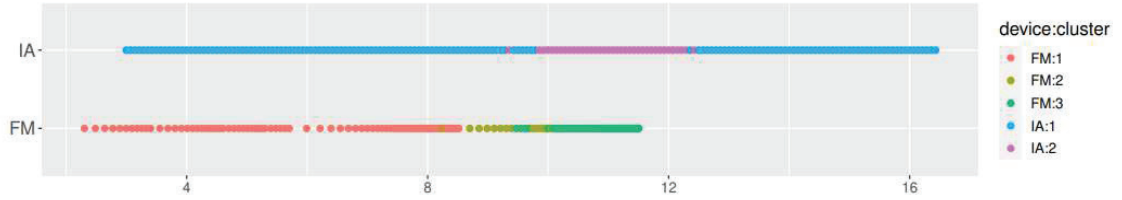


Figure 3: Partitions resulting from the applied hierarchical clustering algorithm. *The x-axis represents the logarithm of electrical frequencies, with colors indicating the grouping of frequencies within each device's cluster. Top: IA. Bottom: FM.*

Acknowledgments *The authors acknowledge the support of the Free University of Bozen-Bolzano via the interdisciplinary research project "New directions in statistical methods for bio-impedance analysis of fruit ripeness (BIOFRUIT)". The first author also acknowledges the financial support from the Italian Ministry of University and Research (MIUR) under the Research Project of National Interest (PRIN) (grant 2017TA7TYC).*

References

Di Lascio, F.M.L., Perazzini, S. (2022). Change point detection in fruit bioimpedance using a three-way panel model, in *Book of Short Papers - SIS2022*, p. 1184-1189. A. Balzanella, M. Bini, C. Cavicchia, R. Verde, (Eds.), Pearson Publ.

Di Lascio, F.M.L., Perazzini, S. (202x). Fitting three-way dynamic panel threshold regression model of bioimpedance data for change point detection. Submitted.

Ibba, P. (2021). *Fruit Quality Evaluation Using Electrical Impedance Spectroscopy*. Free University of Bozen-Bolzano. (https://hdl.handle.net/10863/18446).

Ibba, P., Crepaldi, M., Cantarella, G., Zini, G., Barcellona, A., Rivola, M., Petrelli, M., Petti, L., Lugli, P. (2021). Design and validation of a portable AD5933-based impedance analyzer for smart agriculture. *IEEE Access*, **9**, pp. 63656-63675.

Ibba, P., Falco, A., Abera, B.D., Cantarella, G., Petti, L., Lugli, P. (2020). Bio-impedance and circuit parameters: An analysis for tracking fruit ripening. *Postharvest Biology and Technology* **159**, p. 110978.

Kaufman, L., Rousseeuw, P.J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, NewYork.

doi.org/10.26398/asaproc.0033

Coping strategies and perception of the after-pandemic future of Italians

Francesco D. d'Ovidio[a], Angela Maria D'Uggento[a], Simone Di Zio[b], Luigi Fabbris[c]
[a] Dept. of Economics and Finance, University of Bari Aldo Moro, Italy
[b] Dept. of Legal & Social Sciences, G. D'Annunzio University of Chieti-Pescara, Italy
[c] Tolomeo Studi e Ricerche, Padua, Italy

1. Introduction

COVID-19 pandemic affected millions of people worldwide. While the majority of the diseased recovered, a significant proportion continued to suffer from persistent symptoms and functional limitations referred to as long COVID (LC) or post-acute sequelae of SARS-CoV-2 infection (PASC). Among these symptoms, depression and anxiety have emerged as important health problems. Several studies have reported a higher prevalence of depression in individuals with LC compared with those who fully recovered. The chronic nature of LC, sometimes exacerbated by physical limitations and social isolation, contributes to an increased risk for depression. In addition, dysregulation of immune and inflammatory responses during acute infection may have long-lasting effects on the brain that may lead to the development of depression symptoms.

Efforts are underway to establish comprehensive rehabilitation programs tailored to the needs of people with LC, including multidisciplinary care involving mental health professionals and other specialists. Recent studies (Marchetti and Pössel, 2023) have explored the relationship between personal cognitive traits (such as beliefs about the self, the world and the future) and depressive symptoms, while others have focused on the association between an individual's orientation towards the future and his/her depression symptoms (Zheng et al., 2019). Other scholars (Servidio et al., 2022) have focused on the relationships between future orientation, resilience and depression during the pandemic. However, further research is needed to highlight the mechanisms linking depression and perception of the future to develop effective interventions.

Our research aims to understand how Italians perceive their future in the aftermath of the Covid 19 pandemic, and whether and how disposition toward their future can be strengthened to make them better prepared to cope with their future. We hypothesize that the situation in recent years, characterized by two global shocks (pandemic and Ukraine war) may have increased the difficulty for young people to look into the future. Table 1 describes the aftermath of the pandemic. The data indicate the proportion of respondents who have difficulty imagining their own future, broken down by personal characteristics and the nature of pandemic concerns.

The rest of the paper is organized as follows: Section 2 describes the sample and presents the methodological approach used in the analysis; Section 3 reports the results of the analysis conducted on young and adult people to shed light on the different coping strategies correlated with their future perception. Section 4 discusses the main outcomes in line with the mainstream literature and concludes.

2. Data and methods

2.1. The data

The survey was conducted from May 2021 to May 2022, near the end of the pandemic, using computer-assisted web-based interviews (CAWI) and social media participation. Data collection

was conducted by administering the electronic questionnaire via email to a large convenience sample. The questionnaire was accessible via a link provided in a personal email. As some questions related to health could have been perceived as sensitive, the questionnaire was kept strictly anonymous. It was therefore not possible to ask participants from the first wave who had not responded to cooperate. The final coverage rate was around 6%. A total of 1,164 Italian respondents completed the electronic questionnaire. Research groups from several Italian universities were involved in the study: Bari, Padua, Turin, Naples-Federico II and Chieti-Pescara.

To better understand which variables (sociographic characteristics, personality traits, health experiences, hang-over of the pandemic) affected individuals' vision of the future, an overall analysis and a focus on young people (18-39 years old n=549; 47.16%) were performed. The set of descriptors is shown in Table 1. Our target variable was based on respondents having a clear or, alternatively, blurred vision of what they expected to do after the pandemic. The question was: *Do you have sufficiently clear ideas about what you will do after the pandemic, or do you find it difficult to imagine if and what your life will be like afterwards? R1. I have sufficiently clear ideas about what I expect to do in the future. R2. I imagine my future with difficulty.* Below, we analyze the responses to this question.

Some of the multi-item variables were reduced to a single index, namely: Attitude, Lifestyle, Depression.

Attitude (toward the future) was based on a set of eight items selected from Zimbardo's Time Perspective Inventory (Zimbardo and Boyd, 1999) measured on a dichotomous scale. The factor scores of the responses were used to create the Attitude index allowing us to obtain three groups of respondents scoring about 30% very low and 30% very high. Henceforth, we labeled very low values as "passive" attitude toward the future, medium values as "reactive", and very high values as "proactive".

The second index, **Lifestyle**, was calculated by algebraically summing the reliable raw scores of the 10 items that survived the step-by-step removal process to reach a Cronbach's alpha ≥ 0.7. Subsequently, a one-dimensional factor analysis was performed to determine the quantile thresholds corresponding to three groups: never practicing, indifferent, still practicing social habits.

The third index, **Depression**, measured full-blown depression with the nine-item PHQ9–Patient Health Questionnaire by Spitzer et al. (1999), translated into Italian by Mazzotti et al. (2003). Based on the response scores, three degrees of depression were identified: full-blown depression, at risk of depression, normal. In what follows, only full-blown depression was used.

2.2. Statistical methods

A classification tree, through CART (Breiman et al., 1984), fitting the target variable Y related to the presence of a blurred vision of the after-pandemic future, allowed us to identify the interactions between the main variables in the classification procedure. The predictors selected by the classification tree are listed in Table 1. The predictors were used as independent variables in the subsequent analysis based on a logistic regression model for a binary response variable.

The multiple logistic regression model is expressed as follows (Hosmer & Lemeshow, 2000):

$$logit[p(Y=1)] = \beta_0 + \beta_1 X_1 + \cdots \beta_n X_n$$

where $logit(p) = ln[(p/(1-p)]$ and β_i measures the relation between Y and X_i when all other variables in the model remain fixed. The predictors were divided into two blocks: the control variables that were forced into the model and the set of all other variables that were to be selected according to the backward stepwise criterion. The Nagelkerke pseudo-R^2 was used for the variable selection. Classification tree and logistic model were performed using the software IBM SPSS rel.25.

3. Results

Overall, 26.9% of respondents have a blurred idea of their post-pandemic role, with a significant difference between respondents aged 18-34 (35.2%) and 35 and over (18.3%). Although we hypothesize that young people are less likely to imagine their future than adults (Bollani, Di Zio and Fabbris, 2023), it is interesting to understand to what extent this might depend on individual pandemic experience, mental health status, and character traits.

Table 1. Percentage of Italians having difficulty imagining their own future, by age group, gender, higher education degree, family type, health problems and type of after pandemic worry.

	Particularly worried for the future of:				Sample size	
	Italian economy	Same or no worry	Health	Total	**%**	**N**
18-34 years old	31.8	36.6	46.7	35.2	47.2	549
35 and more	17.6	20.4	29.4	19.3	52.8	615
Male	18.7	22.1	25.7	20.3	40.7	474
Female	29.0	30.6	49.2	31.3	52.3	690
Couple	18.7	23.0	27.3	20.9	58.2	678
Higher education	20.5	23.5	38.8	23.1	57.4	668
Covid infected	18.6	23.4	13.2	20.1	17.0	198
Had physical damages	30.4	46.7	41.7	38.1	9.9	115
Full-blown depression	48.5	55.8	68.4	54.3	28.0	326
No chronical disease	25.4	27.8	30.8	27.0	50.3	585
One chronical disease	18.6	18.2	52.6	22.0	15.4	179
2 or more diseases	24.5	32.9	52.2	29.6	34.3	400
Total	24.2	27.8	40.4	26.8	100,0	1,164

The question was: *When you think about Italy, are you more worried about the future of the economy or the risk to your health? R1. More about the future of the economy R2. Both equally or no worry; R3. More about the risk of contagion.*

When the group of young respondents is further subdivided, there is an astonishing 38.1% among 18-24 year olds who have no clear idea of their future. The complementary group of 25-34 year olds, of course, accounts for a smaller but still relevant 28.6%.

Regarding gender, and in line with other studies (Greene and DeBacker, 2004), men have a significantly better idea of the future than women: the difference in our data is 11 percentage points. In our sample, the proportion of Italians unable to imagining their future, being depressed, is 54.3%. At the same time, people with a blurred vision of the future are more likely to be severely depressed. This highlights a very strong relationship between the difficulty of imagining the future and the dark side of people's minds. Among depressed respondents, most are young people, (41.7%), and the correlation with lack of future imagination is striking ($r=0.385$; $p<0.001$). This means that young people have been exposed to a psychological stress more than adults and the aftermath should be given special attention in the LC of young people.

Couples generally worry less than average (20.9% *versus* 26.8%). This may be because people feel strong in a partnership when it comes to coping with economic and social difficulties, while health, remains a higher-order concern. College graduates also feel stronger than the average (23.1%), but they are also more aware of their powerlessness in the face of the agents of disease. So the virus is a scary factor in that it makes it difficult for individuals to organize their lives in the face of something completely unexpected. Indeed, in the first phase of the pandemic, the apocalyptic news and the astonishing mortality rate of the virus triggered a sense of helplessness that 21st century man would never have expected. Suddenly, daily life was turned upside down, with no idea if or when that worst nightmare would end.

Over time, better knowledge of the epidemic, more efficient organization of the hospital system and vaccines probably enabled most people to revise their fears. Thus, not entirely unexpectedly, Covid 19 infection seems to have little effect on perceptions of health future

(13.2%), but if the respondent is chronically ill, the greater the perceived frailty, the worse the future is viewed.

Table 2 shows the strong relationship between individual optimistic views and perceptions of the future: those who are unable to imagine their future account for 51.3% of those with passive attitudes while the corresponding percentage for those with proactive attitudes is 5.0. When focusing on youth, the two opposite percentages are 59.9 and 5.6, respectively, confirming that the threefold relationship between depression, passive attitude and difficulty in imagining the future is even stronger among youth.

Another important concern related to the future is the belief that a healthy lifestyle (with physical activity) is more conservative than vaccines and that a poorer society is unavoidable after a social shock such as a pandemic. In addition, younger people believe that intensive livestock farming should be closed because it can easily transmit diseases to humans.

Table 2. Percentage of Italians having difficulty imagining their future, by psychological disposition and age.

	Age			Sample size	
Psychological disposition	18-34	35 and more	Total	(%)	N
Passive attitude	59.9	40.5	51.3	30.1	350
Reactive attitude	31.8	18.7	25.2	39.8	463
Proactive attitude	5.6	4.6	5.0	29.5	343
Poorer in the future (a)	42.3	30.8	35.1	17.5	204
Stop to progress (b)	29.5	23.3	25.9	10.1	117
Epidemics inevitable (c)	41.8	22.0	30.1	23.5	274
Closure of livestock (d)	43.1	22.6	31.9	27.6	321
Lazy public servants (e)	35.7	21.0	28.2	22.9	267
Annoyed by migrant landings (f)	41.0	26.4	32.9	12.7	148
No travelling abroad (g)	41.1	23.7	30.9	27.4	319
Mask in case of cold (h)	38.6	19.3	27.9	50.3	585
Only electric or hybrid cars (i)	30.4	20.7	25.3	24.5	285
Physical activity immunizes (j)	47.4	29.6	37.0	7.9	92
Total	35.3	19.3	26.9	26.9	1,164

(a) We will become and stay poorer in the long run; (b) It would not be bad if progress stopped, even if we take a step backward; (c) Epidemics are inevitable, the next one is coming; (d) Intensive livestock farming must be closed, no matter what the cost, because it transmits diseases; (e) Public servants took advantage of working remotely to work even less than usual; (f) Annoyed by migrant landings during the pandemic; (g) I will give up travelling abroad for long time; (h) In case of cold, I will put on the mask; (i) I will only buy electric or hybrid cars; (j) Physical activity makes a person more immune than vaccines.

Table 3. Classification tree in tabular form (Y=unable to imagine future).

Node	Blurred vision (%)	N	Variable/Category	Node	Blurred vision (%)	N	Variable/Category
1	16.2	838	Not Depr	14	22.8	136	Couple=Else
2	54.3	326	Depr=Yes	15	20.7	92	Couple=Yes
4	35.3	187	PassiveAttitude=1	17	64.0	114	Closing livestock=NotAgree
5	12.2	41	ProactiveAttitude=1	18	80.0	50	Closing livestock=Agree
8	17.0	342	ReactiveAttitude=1	21	14.9	181	Stop to progress= NotAgree
9	27.7	148	Psy_Damages=No	22	0.0	25	Stop to progress=Agree
10	64.1	39	Psy_Damages=Yes	23	29.0	69	Higher education=Yes
11	68.9	164	ReactiveAttitude=0	24	16.4	67	Higher education=Lower
12	48.8	121	ReactiveAttitude=1	25	8.3	24	Age=<= 18-34 y.o.
13	13.1	206	Couple=Yes	26	25.0	68	Age=> 18-34 y.o.

A CART analysis with CRT growth method was performed, the difficulty of imagining the future being the response variable. The results of the analysis are summarized in Table 3: the interaction between depression and attitudes is highlighted by the fact that 80.0% of the depressed and passive respondents being unable to see their future believe that intensive livestock farming should be closed because it is one of the greatest threats to human health. On the other hand, all

non-depressed respondents who are a couple can see their future perfectly and believe that it would not be a bad thing even if progress stopped and the society took a step back.

The logistic model (Table 4) proved that people tend to perceive their future better as they shift from a passive to a reactive attitude toward life. Depression also confirms its strong role (Servidio et al., 2022). An important finding is that confidence in the future is enhanced when people can share their future plans with a partner and when they are employed. The role of gender as a predictor is meaningful in itself and is made even more interesting by its relationship to depression, which suggests a gender difference (Greene and DeBacker, 2004). Age and college degree have no effect on respondents' views, as their effects are substituted by the other variables included in the model.

Table 4. Beta estimates of the regression model with blurred vision of future as the criterion variable (stepwise backward selection of regressors, n=549; Nagelkerke pseudo-R^2=35.8%).

Regressor	β	S.E.	p	Exp(β)	***Regressor***	β	S.E.	P	Exp(β)
Sex(baseline=Male)	0.352	0.164	0.032	1.421	Couple	-0.346	0.170	0.042	0.708
Age	-0.013	0.213	0.951	0.987	Reactive Attitude	-1.645	0.272	0.000	0.193
Employed	-0.439	0.205	0.033	0.645	Passive Attitude	2.634	0.275	0.000	13.933
Higher Education	0.168	0.174	0.336	1.183	Depressed	1.265	0.164	0.000	3.542
Constant	-1.430	0.205	0.000	0.239					

4. Discussion and conclusion

The results of our analysis show that the ability to see into the future belongs to states of mind, but also to social conditions. Variables that may explain the ability to see the future include gender (Greene and DeBacker, 2004), being in a partnership (Padawer et al., 2007), being employed, and, most importantly, depression and a passive or reactive attitude toward life. Indeed, the relationship between a person's future orientation and depression symptoms is well established in the mainstreaming literature (Servidio et al., 2022; Schubert et al., 2020; Zheng et al., 2019): being oriented toward the future protects against both the life stress and the risk of depression especially to adolescents and younger adults (Zheng et al., 2019).

Instead, a possible Covid-19 infection did not affect future prospects. Much of the research highlights that the pandemic has significantly worsened future prospects, particularly for young people (Hytman et al., 2023; Singh et al., 2022; Padawer et al., 2007). The prevalence of depression and anxiety symptoms increased among youth during the pandemic (Kleine et al., 2023) and had a strong impact on people's difficulty of projecting themselves into the possible futures. Given the strong association between mental disorders and negative views of the future, it is reasonable to conclude that the pandemic *as a social shock* has worsened attitudes toward a positive future, especially among young people. Depression makes people more vulnerable and therefore exposes them to the most unconscious and ancient fears, sometimes fueled by false information from dubious sources. This could explain our findings that intensive livestock farming may cause health problems to humans and contribute to amplify the greenhouse effect by producing CO_2 in the atmosphere.

However, whether a positive attitude toward the future would moderate the relationship between COVID-19 impact and psychological well-being (Hytman et al., 2023) is a hypothesis that should be further explored in future studies, especially as a moderator to LC.

We can conclude that the examined data revealed clear signals for effective interventions, suggesting the implementation of policies to reinforce positive images of the future (Singh et al., 2022) so to improve the mental health of young people in the post-pandemic society and prepare them for life.

References

Bollani, L., Di Zio, S., Fabbris, L. (2023). Are Italian youngsters adequately equipped for an after-pandemic upswing? in *ASA 2022: Data-Driven Decision Making*. Enrico di Bella, Luigi Fabbris, Corrado Lagazio (Eds.), Firenze University Press-Genoa university Press, pp. 35-40.

Breiman, L., Jerome, H. Friedman, R., Olshen, A., Stone, C.J. (1984). *Classification and Regression Trees*. Wadsworth, Belmont, CA.

Greene, B.A., DeBacker, T.K. (2004). Gender and orientations toward the future: Links to motivation. *Educational Psychology Review*, **16**(2), 91-120.

IBM *SPSS Statistical Package for Social Sciences for Windows, Version 25.* IBM Corp.: Armonk, NY, USA.

Hosmer, D.W., Lemeshow, S. (2000). *Applied Logistic Regression (2nd ed.).* New York: John Wiley & Sons.

Hytman L., Hemming M., Newman T., Newton N.J. (2023). Future time perspective and psychological well-being for older Canadian women during the COVID-19 pandemic. *Journal of Adult Development*, https://doi.org/10.1007/s10804-023-09445-8

Kleine, R., Galimov, A., Hanewinkel, R., Unger J., Sussman S., Hansen J. (2023). Impact of the COVID-19 pandemic on young people with and without pre-existing mental health problems, *Scientific Reports*, **13**, 6111. https://doi.org/10.1038/s41598-023-32918-5.

Marchetti I., Pössel P. (2023). Cognitive triad and depressive symptoms in adolescence: Specificity and overlap, *Child Psychiatry & Human Development*, **54**, pp. 1209–1217, https://doi.org/10.1007/s10578-022-01323-w.

Mazzotti, E., Fassone, G., Picardi, A., Sagoni, E., Ramieri, L., Lega, I., Camaioni, D., Abeni, D., Pasquini, P. (2003). II Patient Health Questionnaire (PHQ) per lo screening dei disturbi psichiatrici: Uno studio di validazione nei confronti della Intervista Clinica Strutturata per il DSM-IV asse I (SCID-I). *Italian Journal of Psychopathology*, **9**(3), pp. 235-242.

Padawer E.A., Jacobs-Lawson J.M., Hershey D.A., Thomas D.G. (2007). Demographic indicators as predictors of future time perspective. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues*, **26**(2), pp. 102–108. https://doi.org/10.1007/s12144-007-9008-4.

Servidio, R., Scaffidi Abbate, C., Costabile, A., Boca, S., (2022). Future orientation and symptoms of anxiety and depression in Italian university students during the COVID-19 pandemic: The role of resilience and the perceived threat of COVID-19. *Healthcare*, **10**, 974. https://doi.org/10.3390/healthcare10060974.

Singh, L., Rathbone, C.J., Moulds, M.L., Holmes, E.A. (2022). Future self-imagery of young people in Sweden during the COVID-19 pandemic: An exploratory mixed methods analysis. *Current Psychology*, https://doi.org/10.1007/s12144-022-04100-z.

Schubert T., Eloo R., Scharfen J., Morina N. (2020). How imagining personal future scenarios influences affect: Systematic review and meta-analysis, *Clinical Psychology Review*, **75**, https://doi.org/10.1016/j.cpr.2019.101811.

Spitzer, R.L., Kroenke, K., Williams, J.B. (1999). Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. *Journal of the American Medical Association*, **282**(18), pp. 1737-1744.

Zheng L., Lippke S., Chen Y., Li D., Gan Y. (2019). Future orientation buffers depression in daily and specific stress. *PsyCh Journal*, **8**(3), pp. 342-352. doi: 10.1002/pchj.283.

Zimbardo, P. G., Boyd, J.N. (1999). Putting time in perspective: A valid, reliable individual differences metric, *Journal of Personality and Social Psychology*, **77**, pp. 1271–1288.

doi.org/10.26398/asaproc.0034

Assessing the sub-urban frailties: The case of Florence

Gianni Dugheri [a], Giulio Grossi [b], Emilia Rocco [b]
[a] Municipality of Florence
[b] Department of Statistics, Computer Science and Applications, University of Florence

1. Introduction

In an era marked by rapid urbanization and societal transformations, the intricate interaction between socio-economic and demographic factors has become a central concern for researchers, policymakers, and urban planners alike. In such a context, understanding the socio-economic dynamics within an area becomes essential for its sustainable development and effective governance.

The city of Florence is not exempt from the forces of change sweeping across the global landscape and this evolution has brought forth complex challenges that warrant meticulous analysis and strategic intervention. Even within a small territory, such as the municipality of Florence, we can study the heterogeneity in the living conditions of inhabitants. A pivotal and currently widely adopted approach, to understanding and addressing the nuanced intricacies of suburban areas' fragility is through the construction and assessment of a socio-economic and demographic fragility index at the suburban level.

In this paper, with the aim of quantifying and qualifying the multifaceted vulnerabilities embedded within the suburban areas of Florence, an index of such kind is suggested. By amalgamating an array of socio-economic and demographic indicators, this index unveils underlying patterns, identifies potential risk factors, and sheds light on the fragility thresholds that may undermine the sustainable development and well-being of these suburban communities. Our aim is to provide a comprehensive understanding of the complex web of factors contributing to fragility, offering valuable insights that can inform targeted policies and interventions. Through this exploration, we seek to contribute not only to the scholarly discourse on urban studies and statistics but also to the practical efforts aimed at enhancing the resilience and prosperity of Florence's suburban areas.

2. Suburban Florentine fragilites

One of the main targets of public policy is to contrast social frailties. Providing tools useful for this purpose is a developing area in social statistics and public policy areas, see Saisana and Philippas (2012) and Khan (1991). The investigation and evaluation of social vulnerabilities necessitate a rigorous examination and quantification of them. In recent years, a substantial body of research has emerged, focusing on the measurement of these intricate concepts and leading to the development of a variety of composite indicators.

Composite indicators serve as tools that amalgamate a collection of elementary indicators into a cohesive measure of complex phenomena, exemplified by metrics such as the Human Development Index (HDI) or the Environmental Performance Index (EPI). For a comprehensive understanding of the construction of synthetic indicators, a valuable resource is OECD et al. (2008). Additionally, an insightful review of existing literature, emphasizing the primary objectives of indicator construction and the attendant challenges, can be found in Greco et al. (2019). The principal aim of a synthetic indicator is to faithfully encapsulate the information inherent in each elementary indicator while minimizing the loss of data. Furthermore, these indicators facilitate transparent ranking of areas (units) for which are calculated, enabling comparisons across different

spatial and temporal contexts. Thus, they offer an effective means of monitoring progress in intricate scenarios.

Effectively handling this wealth of information necessitates the implementation of feature extraction techniques that distill the most pertinent insights. Common methodologies include the arithmetic and the geometric means weighted or unweighted, which condense diverse data points into a singular score. These methodologies also include the Adjusted Mazziotta-Pareto Index (AMPI) Mazziotta and Pareto (2018) a sort of "reasoned" average of several elementary indicators, originally designed to measure well-being and the one used in this work to measure the socio-economic and demographic frailties within the suburban areas of Florence. Other examples of the use of AMPI or its variations can be found in Massoli et al. (2014) and Mazziotta and Pareto (2013). MPI and AMPI are still the golden standard for measuring sustainable and fair wellbeing (BES) in Italy. A fine-grain level of analysis, as the one performed in this work, could benefit surely both the researchers and policymakers interested in disentangling the drivers for fragility. Moreover, to the best of our knowledge, there were no previous published works on this topic, at this fine level of aggregation for the municipality of Florence.

Specifically, the construction of the index is in several steps. Let denote with $Y_{i,j}$ the socioeconomic or demographic outcome regarding the unit i and the indicator j. Then, the scaled matrix of indicators $R_{i,j}$ is defined as

$$R_{i,j} = \left[\frac{Y_{i,j} - min(Y_j)}{max(Y_j) - min(Y_j)}\right] 60 + 70$$

Subsequently, the AMPI index is obtained as

$$AMPI_i^{\pm} = M(R_i) \pm (S(R_i) \times cv(R_i))$$

where M, S, and cv are the mean, the standard deviation, and the coefficient of variation respectively, over the indicators for unit i. The sign of the elementary indicators depends on the nature of the phenomena we are willing to represent, for negative outcomes (such as fragility) we will add, while subtraction is used when a positive phenomena is evaluated, as in the case of Human development Index (HDI). See for further reference on the index construction Mazziotta and Pareto (2018).

In order to calculate the AMPI, first of all we have to identify the elementary index able to represent some main drivers for the identification of socio-economic and demographic fragility within the Florentine area. In particular, regarding the demographic aspect we consider the following elementary indicators: the share of the population aged 80 or more, the natural balance between births and deaths and the % population difference in the last 5 years. In this way, we should be able to catch up on some relevant demographic frailties: aging, low fertility and depopulation. Focusing on the economic aspects, we use as fragility indicators: the share of people under the poverty line (60% of the median income), the share of families under the poverty line, and the share of people paying rent, as a proxy of an insufficient capital accumulation. Lastly, we use several indexes for representing social fragility: the share of the 65+ population living alone, the share of minors in a single-parent family, the share of foreign-born minors over the minors, the share of unoccupied flats, and the share of the graduated population.

All the mentioned elementary indicators are based on data collected during 2021, with data sources shown in Table 1. Demographic and the social indexes are collected by the civil registry of Florence, while economic indexes are collected by the Italian Revenue Agency (Agenzia delle Entrate AdE) and further elaborated by the Municipality of Florence. Lastly, the % of graduates and the % of unused dwellings is derived from the 2011 census. Table 1 reports some relevant descriptive statistics for these elementary indexes.

As units of observation, we assume $N = 74$ suburban units in which the area of Florence is partitioned. These units represent a middle-level aggregation between the census areas and the administrative partitions of Florence, which could be too large for this study. Even if these units

stem from administrative sources, they represent homogeneous partitions of the city, particularly relevant for our purposes. Two of them were excluded from the analysis, as their population is too scarce to have reasonable estimates (under 100 inhabitants).

Table 1: Descriptive statistics of the elementary indexes - Mean, standard deviation, minimum, maximum median, first and third quantile.

Elementary Index	Mean	st.dev.	Min	25%	50%	75%	Max	Source
% Over 80	9.552	2.076	1.260	8.185	9.663	10.938	15.028	Fl. civil registry
Δ population	-2.826	3.584	-11.337	-4.512	-2.942	-1.554	16.098	Fl. civil registry
Natural Balance	-28.542	21.286	-94.200	-40.550	-25.700	-12.900	4.000	Fl. civil registry
% Over65 living alone	9.101	1.840	0.840	8.107	9.445	10.075	12.973	Fl. civil registry
% Under18 foreigners	15.948	8.407	2.308	10.322	14.375	19.197	39.159	Fl. civil registry
% Under 18 - Single parent	42.279	5.985	18.518	38.646	42.153	44.867	64.785	Fl. civil registry
% Unused dwellings	3.936	4.442	0.000	1.651	2.976	4.668	33.663	2011 Census
% Graduated	37.812	10.524	14.796	30.544	38.352	46.441	57.083	2011 Census
% New inhabitants	3.537	1.577	1.230	2.895	3.270	3.895	15.210	Fl. civil registry
% people under poverty line	33.110	3.925	21.171	30.780	32.696	34.801	44.231	AdE
% families under poverty line	20.401	5.037	7.353	17.696	19.343	21.381	34.499	AdE
% rents	20.366	6.823	8.092	15.506	19.109	25.119	37.841	2011 census

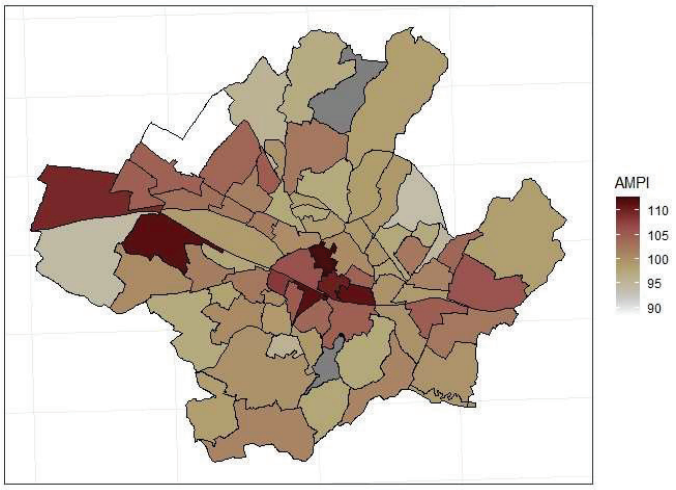


Figure 1: Map of Florentine socio-economic and demographic frailties, AMPI index

The comprehensive fragility map (Figure 1) highlights regions with sub-optimal performance, notably the city center and the northwestern quadrant, both displaying pronounced fragility. Interestingly, this finding might appear counterintuitive, considering the presence of esteemed artistic landmarks juxtaposed with challenging socio-economic conditions. Yet, a closer examination rationalizes this outcome: the historic city center features some of the oldest, relatively affordable residences, which could attract newcomers like migrants or students.

This conclusion gains further support when analyzing areas with significant mean-median income disparities. As shown in Figure 2, the urban core emerges as an income inequality hub, while the northwestern outskirts showcase near-parity between mean and median income.

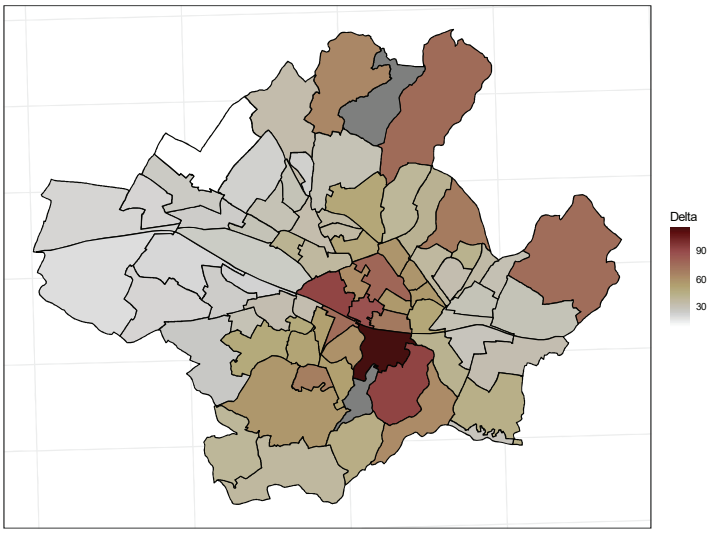


Figure 2: Income inequality measured as the difference between mean and median income, over the median income

3. Cluster analysis

Socio-economic fragility could be a phenomenon that could arise across different units, spatially contiguous or having similar characteristics. For this purpose, we perform a cluster analysis on our units, to investigate whether are present clusters of suburban areas with similar characteristics and enriching so the information set for policymakers. To classify units in the clusters, we adopt a clustering approach based on Gaussian mixture model, and implemented in R using the package mclust (Fraley and Raftery (2002), McLachlan et al. (2019)).

Cluster analysis points out the presence of five major clusters in Florence, as shown in Figure 3. The fifth cluster is composed only by one area, the airport, as it have completely different characteristics with respect the rest of the city. It is worth noting that the cluster analysis draws pretty well some homogenous areas, as follows:

The city center cluster: in pink, represent those areas within the old town of Florence

1. The hills cluster: in green, represents the bordering areas of Florence, encompassed into the hilly areas around the city. We can spot a higher mean age, with an associated higher demographic fragility.
2. The suburbs: in orange, represents the cluster of peripheral areas, located mainly in the northwestern part of the city, and the southeastern part. We notice these areas for a higher level of social fragility and economic fragility, but lower inequality.
3. The semicentral areas: in blue, represent the majority of Florence, with not many frailties to be spotted.

Figure 4 shows the weighted fragilities into the three domains of analysis and at the aggregate level for the aforementioned clusters. We can note that the city center is the most fragile and unequal area of Florence, with considerable issues both on the economic environment and on the social environment. In particular, we can see that the percentage of people and families living under the poverty line is abnormal with respect to the rest of the city. These results can be explained by the particular social composition of the city center, with many students and migrants that move inside the city because of more affordable rents. Similarly, we can notice social frailties in the city center and in the suburbs clusters, mainly due to the high percentage of foreigners and minors in single-parent families.

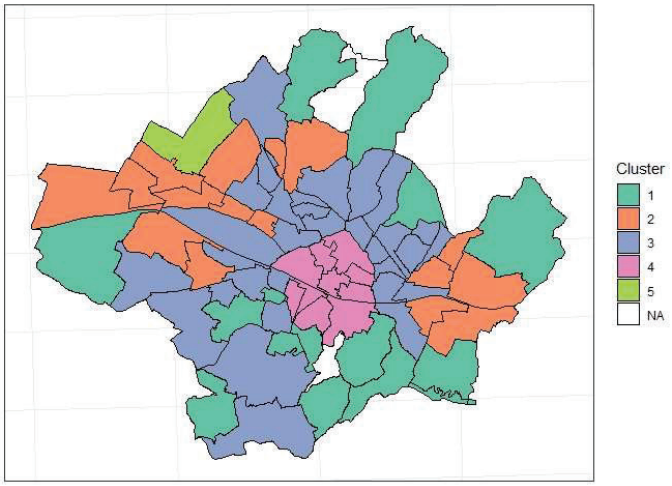


Figure 3: Cluster classification of suburban areas of Florence

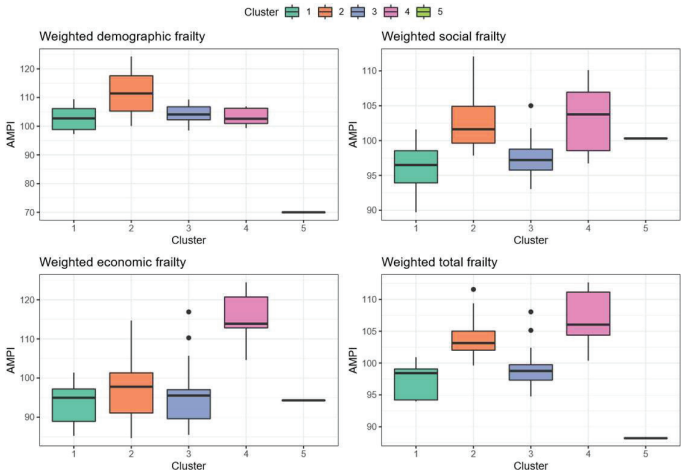


Figure 4: Boxplots of AMPI index for cluster frailties

4. Conclusion

In this work we aimed to assess suburban fralties of Florence, following three pillars: demographic, economic and social frailty. We showed that the use of synthetic indexes aggregated at suburban level could improve the local policymaker action, allowing for a detailed and focused action on the urban frailties. Our findings highlight substantial differences across suburban areas, thereby suggesting that fragility evaluations should be performed at the lowest possible level to capture potential heterogeneity, even at the municipal level. Intriguingly, areas with the highest fragility coincide with those exhibiting elevated economic inequality.

Cluster analysis has proven to be a valuable tool for validating, further explicating, and enhancing the interpretation of the succinct outcomes derived solely from the calculation of the synthetic index.

References

Fraley, C., Raftery, A.E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, **97**(458), pp.611–631.

Greco, S., Ishizaka, A., Tasiou, M., Torrisi, G. (2019). On the methodological framework of

composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research*, **141**, pp.61–94.

Khan, H. (1991). Measurement and determinants of socioeconomic development: A critical conspectus. *Social Indicators Research*, **24**, pp.153–175.

Massoli, P., Mazziotta, M., Pareto, A., Rinaldelli, C. (2014). Indici compositi per il BES. *Giornate della Ricerca in Istat, 11-12 Novembre 2014*, pp. 10–11 (https://www.istat.it/it/files/2014/10/Paper_Sessione-IV_Massoli_Mazziotta_Pareto_Rinaldelli.pdf)

Mazziotta, M., Pareto, A. (2013). Methods for constructing composite indices: One for all or all for one. *Rivista Italiana di Economia Demografia e Statistica*, **67**(2), pp.67–80.

Mazziotta, M., Pareto, A. (2018). Measuring well-being over time: The adjusted Mazziotta–Pareto index versus other non-compensatory indices. *Social Indicators Research*, **136**, pp. 967– 976.

McLachlan, G. J., Lee, S. X., Rathnayake, S. I. (2019). Finite mixture models. *Annual Review of Statistics and Its Application*, **6**, pp. 355–378.

OECD, European Union, JRC- European Commission (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide*. https://doi.org/10.1787/9789264043466-en.

Saisana, M., Philippas, D. (2012). Sustainable society index (ssi): Taking societiesa pulseˆ along social, environmental and economic issues. *Environmental Impact Assessment Review*, **32**, pp. 94–106.

doi.org/10.26398/asaproc.0035

Impact of global turmoil days on the G7 stock markets during the Covid-19 pandemic and Russian-Ukrainian war

Pierdomenico Duttilo[a], Tonio Di Battista[b], Stefano Antonio Gattone[b]
[a] "G. d'Annunzio" University of Chieti-Pescara, Italy
[b] DISFIPEQ, "G. d'Annunzio" University of Chieti-Pescara, Italy

1. Introduction

Considered as economic barometers, stock markets rapidly reflect the impact of important events on traded stocks (Kartal et al., 2022). These events can lead to financial turmoil days, often associated with tail risk events, i.e. extreme and unexpected market movements that fall in the tails of a probability distribution (Bollerslev and Todorov, 2011). Thus, analysing the features of turmoil days can contribute to a better understanding of tail risk, allowing for improved risk management strategies to mitigate the impact of rare market fluctuations. The 2015-2016 stock market sell-off, the US-China trade war, the Covid-19 pandemic, and the Russian-Ukrainian conflict, are cited as examples (Awan et al., 2021; So et al., 2022; Izzeldin et al., 2023). Several contributions (Yousef, 2020; Shehzad et al., 2020) point out that the stock market volatility increased over crisis periods leading to a sudden and unpredictable stock market behaviour.

Common approaches to assess the impact of specific events on stock markets include the use of subjectively identified dummy variables (Yousef, 2020; Chaudhary et al., 2020; Shehzad et al., 2020; Duttilo et al., 2021), sub-periods (Han et al., 2019), and Markov-switching GARCH models (Ardia et al., 2019; Ouchen, 2022).

This work implements a two-step approach (Duttilo et al., 2023) to study the impact of global turmoil days (or global tail risk events) on the G7 stock markets returns and volatility. Firstly, turmoil days were objectively detected with a two-component mixture of generalised normal distributions (MGND) on the Dow Jones Global Index (W1DOW) returns. The MGND model can fit the returns distribution and its non-normal features like skewness and excess kurtosis (Wen et al., 2022; Duttilo et al., 2023). Returns are generated by turmoil days if they belong to the mixture component with the lowest shape parameter which accounts an extreme tail behaviour. Secondly, returns and volatility were estimated with an AR(1)-NAGARCH(1,1)-M model for each G7 stock market including the three exogenous dummy variables defined in the first step.

The following questions are addressed: Do turmoil days have a significant impact on returns and volatility of G7 indices? Which market has been most affected by turmoil days? and in what period? The rest of the paper is organized as follows. Data and methods are described in Section 2. Findings are showed in Section 3, while some conclusions are given in Section 4.

2. Data and methods

The study considers the returns of the daily closing prices of the W1DOW and stock market indices of the G7 countries, namely GSPTSE (Canada), FCHI (France), DAX (Germany), FTSE.MIB (Italy), N225 (Japan), FTSE (UK) and S&P500 (US). The period taken for the study is from January 4, 2016 to June 30, 2023. Daily returns of all indices under study were calculated as follows

$$r_{i,t} = \ln\left(\frac{P_{i,t}}{P_{i,t-1}}\right)100 \tag{1}$$

where: $r_{i,t}$ is the daily return on index *i* at time *t*, $P_{i,t}$ is the daily closing price of index *i* at time *t* and $P_{i,t-1}$ is the daily closing price of index *i* at time $t-1$.

Usually, mixtures of normal distributions are applied to fit the empirical distribution of financial returns (Kon, 1984; Behr and Pötter, 2009; Bellalah and Lavielle, 2002). However, they impose a priori specific constraints on the shape (kurtosis) of returns distribution. The MGND model can overcome this critical issue thanks to the flexibility provided by the additional shape parameter which determines the tail weights. Wen et al. (2022) found that the MGND model better describes the skewness and excess kurtosis of daily returns compared to mixtures of normal distributions. The two-component MGND model is defined as follows

$$f(r_{i,t}|\theta) = \sum_{k=1}^{2} \pi_k \frac{\nu_k}{2\sigma_k \Gamma(1/\nu_k)} \exp\left\{-\left|\frac{r_{i,t}-\mu_k}{\sigma_k}\right|^{\nu_k}\right\} \text{with } \Gamma(1/\nu_k) = \int_0^{\infty} t^{1/\nu_k - 1} \exp^{-t} dt, \tag{2}$$

where: $\theta = \{\pi_k, \mu_k, \sigma_k, \nu_k\}$ for k = 1,2; π_k is the *k*-th mixture weight satisfying $\sum_{k=1}^{2} \pi_k = 1$ and $\pi_k > 0$, μ_k is the *k*-th location parameter ($\mu_k \in \mathbf{R}$), σ_k is the *k*-th scale parameter ($\sigma_k > 0$), and ν_k is the *k*-th shape parameter ($\nu_k > 0$). By varying the shape parameter ν_k, it is possible to capture a wide range of statistical distributions. For example, the normal distribution and Laplace distribution are obtained with $\nu_k = 2$ and $\nu_k = 1$, respectively (Wen et al., 2022). A smaller shape parameter identifies a distribution with thicker tail and then a higher standard deviation. It is assumed that there are two distinct trading days: stable and tumultuous. Turmoil days are more volatile and less frequent than stable days (Kon, 1984; Kim and White, 2004). A daily return is classified as a turmoil day if it belongs to the mixture component with the lowest shape parameter. Using the naïve Bayes classification, each return is assigned to the class with the highest posterior probability (Kon, 1984). The resulting classification was encoded in three different dummy variables: pre-Covid-19 turmoil days (2016-2019), Covid-19 turmoil days (2020-2021) and Russian-Ukrainian conflict turmoil days (2022-2023).

GARCH models were introduced by Engle (1982) to capture the volatility clustering phenomenon. Hentschel (1995) developed the family GARCH model which nests the most popular symmetric and asymmetric GARCH models. Notable among these is the nonlinear-asymmetric GARCH (NAGARCH) model proposed by Engle and Ng (1993). This model allows to capture the leverage effect by a right-shift of the news impact curve (NIC)[1]. In this work, the AR(1)-NAGARCH(1,1)-M model with exogenous dummy variables is specified as:

Conditional mean equation

$$r_{i,t} = \mu + \eta_1 \text{pre-COVID-19} + \eta_2 \text{COVID-19} + \eta_3 \text{WAR} + \phi_1 r_{i,t-1} + \lambda_1 h_{i,t}^2 + \epsilon_{i,t}, \tag{3}$$

Conditional variance equation

$$h_{i,t}^2 = \omega + \delta_1 \text{pre-COVID-19} + \delta_2 \text{COVID-19} + \delta_3 \text{WAR} + \alpha_1 h_{i,t-1}^2 (z_{i,t} - \gamma 1)^2 + \beta_1 h_{i,t-1}^2$$

$$\text{with } z_{i,t} = \frac{\epsilon_{i,t}}{\sqrt{h_{i,t}^2}} \sim \text{SGND}(0, 1, \nu, s) \tag{4}$$

In Eq. (3) μ is the constant term. The coefficients η_1, η_2, η_3 determine the impact of turmoil days on the conditional mean during the pre-Covid-19 λ_1. $\epsilon_{i,t}$ period, Covid-19 pandemic

[1] The NIC associates past returns shocks to current volatility. The leverage effect occurs when negative past returns shocks increase volatility by a larger amount than positive returns shocks (Engle and Ng, 1993).

and Russian-Ukrainian conflict, respectively. Consequently, pre-Covid-19, Covid-19 and WAR are three dummy variables which assume the value of 1 during turmoil days, otherwise they are equal to 0 (stability days). ϕ_1 is the first-order autoregressive coefficient. Based on the ideas of Engle et al. (1987), Eq. (3) includes the risk premium coefficient indicates the error terms of stock index *i* at time *t*.

In Eq. (4), $h_{i,t}^2$ denotes the conditional variance of stock index *i* at time *t*. ω is the constant term. The coefficients δ_1, δ_2, δ_3 determine the impact of turmoil days on the conditional variance during the pre-Covid-19 period, Covid-19 pandemic and Russian-Ukrainian conflict, respectively. The dummy variables are defined as in Eq. (3). The parameters α_1 and β_1 capture the ARCH and GARCH effects. Since $\alpha_1 > 0$ and $h_{i,t-1}^2 > 0$ the leverage effect is observed if $\gamma_1 > 0$. $z_{i,t}$ is the standardized residual of stock index *i* at time *t*. Standardized residuals are modelled with a skewed generalised normal distribution (SGND) to capture non-normal features of returns. The AR(1)-NAGARCH(1,1)-M model with exogenous dummy variables was estimated by the R's *rugarch* package for all G7 indices (Ghalanos, 2022).

3. Results

Table 1 shows descriptive statistics and some preliminary statistical hypothesis tests. The negative skewness, excess kurtosis and JB test stated that G7 indices are not normally distributed. Moreover, the ADF tests highlight that all indices are stationary in their first difference, while the ARCH-LM tests confirm the presence of ARCH effects and heteroscedasticity.

Table 1: Descriptive statistics and statistical tests results

G7 Index	Mean	SD	Skewness	Kurtosis	JB	ADF	ARCH-LM
GSPTSE	0.0223	1.0358	-1.7955	43.1601	132726**	-12.12**	746.26**
FCHI	0.0259	1.2213	-1.1107	14.2981	14819**	-12.66**	336.76**
DAX	0.0203	1.2511	-0.7521	13.2396	12568**	-12.18**	363.94**
FTSE.MIB	0.0208	1.4705	-1.9337	22.8831	38116**	-12.30**	187.36**
N225	0.0212	1.2520	-0.2189	4.6937	1574**	-12.66**	220.26**
FTSE	0.0044	1.0448	-1.0696	15.2159	16711**	-12.62**	451.59**
S&P500	0.0430	1.2124	-0.8726	16.6982	19951**	-12.86**	736.31**

Note. ** denotes p-value significance at the 1% level of statistical hypothesis tests: Jarque-Bera test (JB), Augmented Dickey-Fuller test (ADF) and ARCH-LM test

Table 2 shows the estimated two-component MGND on W1DOW returns. The stable component is predominant compared to the turmoil component which has a negative location parameter and a higher scale parameter. In addition, the shape parameter of the turmoil component is lower than the shape parameter of the stable component. As a result, the standard deviation of the turmoil component (2.0339) is higher than the standard deviation of the stable component (0.7119). These findings are also confirmed in panel a. of Figure 1 where it is evident as the turmoil component (dashed line) captures the extreme tail behaviour of W1DOW returns. Panel b. of Figure 1 shows the W1DOW daily returns and the detected turmoil days (grey vertical lines).

Table 3 shows the estimated coefficients of the conditional mean and variance equations. The coefficients α_1 and β_1 are positive and statistically significant for all G7 indices, underlining the presence of ARCH and GARCH effects. In addition, the leverage effect (γ_1) and risk premium (λ_1) are statistically significant for all indices, except for the N255 (Japan) where the

risk premium term is low and not significant. The SGND suitably captures the non-normal features of daily returns as the coefficients v and s are statistically significant for all indices. Turmoil days decrease the conditional mean ($\eta_i < 0$ for $i = 1,2,3$) and increase the conditional variance ($\delta_i > 0$ for $i = 1,2,3$).

Table 2: Estimated parameters of the two-component MGND on W1DOW

Stable component				Turmoil component			
π_1	μ_1	σ_1	ν_1	π_2	μ_2	σ_2	ν_2
0.8900	0.0781	0.6737	1.2262	0.1100	-0.3097	1.7347	1.1313

Figure 1: Left tail of estimated MGND model (panel a.) and turmoil days (panel b.)

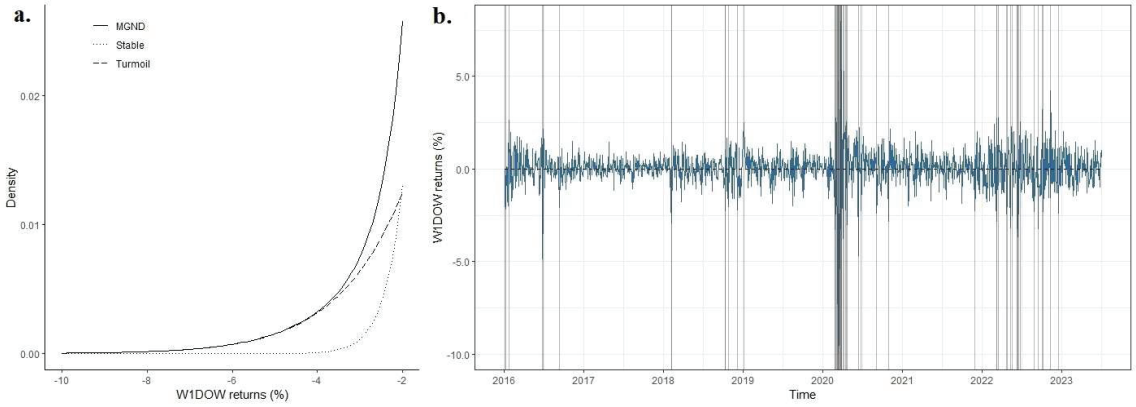


Figure 2 provides an easy view of the turmoil impact on the conditional mean (panel a.) and variance (panel b.). The magnitude of impact varies across periods (pre-Covid-19, Covid-19 and Russia-Ukrainian conflict) and among indices. With the only exception of N225 (Japan), pre-Covid-19 turmoil days have a low impact on the conditional mean and variance. On the other hand, Covid-19 turmoil days greatly affect the conditional mean and variance compared to the other two periods. Specifically, whereas S&P500 (US) and DAX (Germany) were the most affected indices in term of conditional mean, the impact on the conditional variance is almost homogeneous among G7 indices. The impact of the Russia-Ukrainian conflict turmoil days on the conditional mean is lower than the impact of pre-Covid-19 turmoil days, except for S&P500 (US) and DAX (Germany). By contrast, the impact on the conditional variance is approximately the same, except for N225 (Japan).

4. Conclusions

This study adds to the literature as it investigates the reactions of G7 indices to turmoil days in three different periods: pre-Covid-19 (2016-2019), Covid-19 (2020-2021) and Russia-Ukrainian conflict (2022-2023). Findings highlight that the impact of turmoil days varies across markets and periods. Specifically, Covid-19 turmoil days greatly affect the conditional mean and variance compared to the other two periods. In contrast, the impact of Russia-Ukrainian conflict turmoil days on the conditional variance is almost equal to the impact of pre-Covid-19 turmoil days, except for N225 (Japan).

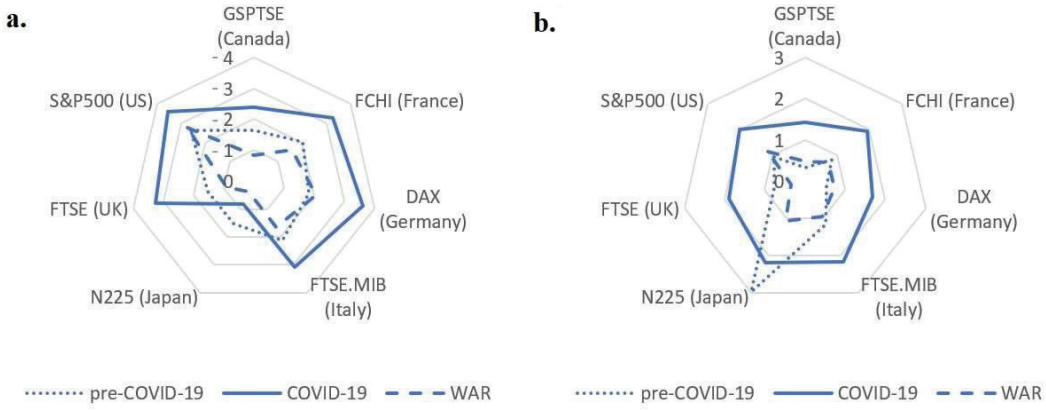
A possible extension of this work could be a comparison between the two-step approach and Markov-Switching GARCH models which identify stable and turmoil regimes directly within the GARCH equations (Ardia et al., 2019).

Table 3: Estimated coefficients of the AR(1)-NAGARCH(1,1)-M model

	GSPTSE	FCHI	DAX	FTSE.MIB	N225	FTSE	S&P500
Conditional mean equation							
μ	0.0198	-0.0065	-0.0607**	-0.0064	0.0430	-0.0608**	0.0218
η_1	-1.6350**	-2.0333**	-1.8392**	-2.1130**	-1.5196*	-1.5262**	-2.6136**
η_2	-2.3978**	-3.2706**	-3.6107**	-3.0733**	-0.8107	-3.2512**	-3.5712**
η_3	-0.8332**	-1.5956**	-2.0588**	-1.6348**	-0.3717	-0.8998**	-2.7547**
Φ_1	0.0181	-0.0077	-0.0281*	-0.0341	-0.0163	-0.0421*	-0.0383
λ_1	0.0483*	0.0677**	0.1149**	0.0466*	0.0025	0.1264**	0.0769**
Conditional variance equation							
Ω	0.0169**	0.0458**	0.0570**	0.0765**	0.1121**	0.0528**	0.0250**
δ_1	0.3262**	0.8267**	0.5294**	1.1566*	2.9796**	0.7684**	0.9547**
δ_2	1.4338**	1.9458**	1.6791**	2.1641**	2.1865**	1.8920**	2.0159**
δ_3	0.4726**	0.7307**	0.7363**	0.9548**	1.0635*	0.3341*	1.1547**
α_1	0.0866**	0.0561**	0.0419**	0.0671**	0.0827**	0.0809**	0.1053**
β_1	0.7954**	0.6316**	0.6395**	0.7131**	0.6877**	0.7287**	0.7387**
γ_1	0.8968**	2.1652**	2.4852**	1.5542**	1.2111**	1.1511**	1.0493**
N	1.7202**	1.3568**	1.2578**	1.3669**	1.4119**	1.3200**	1.4822**
S	0.8391**	0.8974**	0.8982**	0.8863**	0.9544**	0.8838**	0.9663**

Note. ** and * denote p-value significance at the 1% and 5%, respectively

Figure 2: Turmoil impact on the conditional mean (panel a.) and variance (panel b.)



References

Ardia, D., Bluteau, K., Boudt, K., Catania, L., Trottier, D.A. (2019). Markov-switching Garch models in R: The msgarch package. *Journal of Statistical Software*, **91**(4), pp. 1–38.

Awan, T., Khan, M., Haq, I., Kazmi, S. (2021). Oil and stock markets volatility during pandemic times: A review of G7 countries. *Green Finance*, **3**, pp. 15–27.

Behr, A., Pötter, U. (2009). Alternatives to the normal model of stock returns: Gaussian mixture, generalised logf and generalised hyperbolic models. *Annals of Finance*, **5**, pp. 49–68.

Bellalah, M., Lavielle, M. (2002). A decomposition of empirical distributions with applications to the valuation of derivative assets. *Multinational Finance Journal*, **6**(2), pp. 99–130.

Bollerslev, T., Todorov, V. (2011). Tails, fears, and risk premia. *The Journal of Finance*, **66**(6), pp. 2165–2211.

Chaudhary, R., Bakhshi, P., Gupta, H. (2020). Volatility in international stock markets: An empirical study during covid-19. *Journal of Risk and Financial Management*, **13**(9). https://doi.org/10.3390/jrfm13090208.

Duttilo, P., Gattone, S.A., Di Battista, T. (2021). Volatility modeling: An overview of equity markets in the euro area during covid-19 pandemic. *Mathematics*, **9**(11), 1212.

Duttilo, P., Gattone, S.A., Iannone, B. (2023). Mixtures of generalized normal distributions and egarch models to analyse returns and volatility of esg and traditional investments. *AStA Advances in Statistical Analysis*, pp. 1–33.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**(4), pp. 987–1007.

Engle, R. F., Lilien, D.M., Robins, R.P. (1987). Estimating time varying risk premia in the term structure: The Arch-m model. *Econometrica*, **55**(2), pp. 391–407.

Engle, R.F., Ng, V.K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance*, **48**(5), pp. 1749–1778.

Ghalanos, A. (2022). *rugarch: Univariate GARCH models*. R package version 1.4-9.

Han, L., Yan, H., Zheng, C. (2019). Normal mixture method for stock daily returns over different sub-periods. *Communication in Statistics – Simulation and Computation*, **48**(2), pp. 447–457.

Hentschel, L. (1995). All in the family nesting symmetric and asymmetric Garch models. *Journal of Financial Economics*, **39**(1), pp. 71–104.

Izzeldin, M., Muradoglu, Y.G., Pappas, V., Petropoulou, A., Sivaprasad, S. (2023). The impact of the Russian-Ukrainian war on global financial markets. *International Review of Financial Analysis*, **87**, pp. 102598.

Kartal, M.T., Ertugrul, H.M., Ulussever, T. (2022). The impacts of foreign portfolio flows and monetary policy responses on stock markets by considering covid-19 pandemic: Evidence from Turkey. *Borsa Istanbul Review*, **22**(1), pp. 12–19.

Kim, T.-H., White, H. (2004). On more robust estimation of skewness and kurtosis. *Finance Research Letters*, **1**(1), pp. 56–73.

Kon, S.J. (1984). Models of stock returns: a comparison. *The Journal of Finance*, **39**(1), pp. 147– 165.

Ouchen, A. (2022). Is the esg portfolio less turbulent than a market benchmark portfolio? *Risk Management*, **24**(1), pp. 1–33.

Shehzad, K., Xiaoxing, L., Kazouz, H. (2020). Covid-19's disasters are perilous than global financial crisis: A rumor or fact? *Finance Research Letters*, **36**, 101669.

So, M., Mak, A., Chu, A. (2022). Assessing systemic risk in financial markets using dynamic topic networks. *Scientific Reports*, **12**, 2668.

Wen, L., Qiu, Y., Wang, M., Yin, J., Chen, P. (2022). Numerical characteristics and parameter estimation of finite mixed generalized normal distribution. *Communications in Statistics – Simulations and Computation*, **51**(7), pp. 3596–3620.

Yousef, I. (2020). Spillover of covid-19: Impact on stock market volatility. *International Journal of Psychosocial Rehabilitation*, **24**(6).

doi.org/10.26398/asaproc.0036

Modeling gaps in FinTech knowledge

Rosa Fabbricatore[a], Maria Iannario[b], Lucio Palazzo[b], Francesco Palumbo[b]
[a] Department of Economics and Statistics, Federico II University of Naples, Naples, Italy.
[b] Department of Political Sciences, Federico II University of Naples, Naples, Italy.

1. Introduction

Technological innovations in worldwide finance impose new challenges and lead to the development of a new knowledge domain called *FinTech*. National governments aim to exploit these innovations by boosting new technologies, such as blockchain and cryptocurrencies, or using artificial intelligence to ensure automatically managed financial services for more sustainable approaches and tailored solutions to individual needs. Therefore, fostering the broadest access to these new high-technological financial markets has become a crucial policy governments' goal to ensure inclusive economic growth. Indeed, people actively participating in the financial system are more and better able to start and expand businesses, invest in their children's education, and absorb economic shocks (Robb and Woodyard, 2011).

Italy is significantly affected by the spread of digital technologies in financial intermediation. Even before the pandemic crisis, there had been a progressive growth in the use of alternative payment instruments to cash - especially cards - due to both rapid technological development and regulatory interventions to support payment efficiency and security. Nonetheless, the Italian government is showing the willingness to move forward with the intentions stated in the 2018 FinTech Action Plan[1], aiming to expand the functionality and security of traditional electronic payments and to address and cope with new instruments (such as crypto-assets), and are therefore closely related. This work intends to contribute to this topic by investigating individual heterogeneity concerning the knowledge of FinTech tools, with a specific focus on the Italian context.

Knowledge level - and financial knowledge is no exception - can be modeled as a not directly observable construct, namely a latent variable, measured through a set of observed items (manifest variables). Therefore, latent variable models represent a relevant reference framework in this context. In particular, a multilevel IRT model (Fox and Glas, 2001) was exploited to evaluate the level of FinTech individual knowledge and the differences according to Italian regions. At the country level, Italian regions may differ regarding their updating and advancement in FinTech, making it reasonable to suppose that individuals' financial knowledge is more similar for those in the same region.

The paper is organized as follows: Section 2 introduces the data adopted for the study, while Section 3 describes the methodological framework. In conclusion, Section 4 discusses the main results.

2. Data

Data at hand come from the survey on "knowledge and use of FinTech products" administered as part of the European project CA19130 FinTech and artificial intelligence in finance[2]. The involved sample consists of 598 Italian individuals from 10 different Italian regions, predominantly female (56%) and with a mean age of $m = 24.44$ ($sd = 7.36$).

[1] "FinTech action plan: For a more competitive and innovative European financial sector", COM/2018/0109.
[2] https://www.cost.eu/actions/CA19130

The considered observed variables contribute to a broad understanding of the individual financial knowledge referred to in the modern FinTech framework, including the following digital financial tools (FT): Crowdfunding (FT1), Cryptocurrencies (FT2), Instant insurance (FT3), Instant payments (FT4), Roboadvisor (FT5), Peer to peer services (FT6), Cloud computing (FT7), DLT/Blockchain (FT8), Artificial intelligence (FT9), Internet of things (FT10), Machine learning (FT11), Biometric identification (FT12). For each financial tool, participants were asked to indicate their knowledge level on a 6-point Likert response scale, ranging from 1 = "no knowledge" to 6 = "expert".

3. Methodology: Multilevel IRT model

A multilevel approach (Stapleton, McNeish, and Yang, 2016) to data analysis was adopted due to the hierarchical structure of the data, with individuals at the first level (individual level) of the hierarchy and regions at the second level (cluster level). Moreover, as stated before, financial knowledge can be conceived as a latent variable measured by several manifest indicators, herein represented by a set of ordinal-type items. Accordingly, the *item response theory* (IRT; Hambleton and Swaminathan, 2013) framework is considered for the measurement part of the proposed multilevel model.

The adopted methodology allows measurement of individuals' knowledge of FinTech tools and, simultaneously, a rank of Italian regions according to the knowledge levels of their citizens.

More formally, let Y_{ijk} be the response on item k $(k = 1, \dots, 12)$ of individual i $(i = 1, \dots, n)$ belonging to the Italian region j $(j = 1, \dots, 10)$. Each item has $m = 6$ response categories indicated with $c = 0, \dots, m-1$. Moreover, let θ_{ij} denote the financial knowledge of individual i belonging to region j, where θ_{ij} is a normally distributed random variable.

Two components can be distinguished in the multilevel IRT model (Fox and Glas, 2001): a measurement part following an IRT parameterization and a structural component. More in-depth, the measurement model specifies the conditional probability of the observed item responses given the latent financial knowledge $P(\boldsymbol{Y}|\boldsymbol{\theta})$, where the probability of individual i in the region j answering c to item k is given by

$$P\left(Y_{ijk} = c \middle| \theta_{ij}, Y_{ijk} \in c-1, c\right) = F\left[\gamma_k\left(\theta_{ij} - \delta_{ck}\right)\right], \quad c = 1, \dots, m-1,$$

where γ_k and δ_{ck} are the discrimination and the step-difficulty parameters, respectively. $F[\cdot]$ is a cumulative probability distribution function; when the logistics distribution is considered, the generalized partial credit model results (Muraki, 1992). The latter reduces to the partial credit model (Masters, 1982) when discrimination parameters γ_k are constrained to 1 for all items.

The structural component of the model accounts for the hierarchical structure of data by defining the following relations:

$$\theta_{ij} = \beta_{0j} + e_{ij},$$

$$\beta_{0j} = \beta_0 + u_j,$$

where β_0 is the financial knowledge overall average, β_{0j} is the financial knowledge average for the region j, and u_j and e_{ij} are normally-distributed random effects indicating the variation at regional- and individual-level, respectively. Model parameters estimation was carried out through the GLLAMM procedure (Rabe-Hesketh, Skrondal, and Pickles, 2004) of STATA software.

4. Results

Looking at the measurement part of the model, item difficulty parameters provide some

insights into the most widely known (less popular) digital financial services and tools. In particular, average item difficulty (also called item location) reported in Figure 1 shows that item FT7 is the easiest item, indicating that cloud computing is the most popular modern financial tool followed by instant payments (FT4) and artificial intelligence (FT9). On the other hand, the most difficult items (less popular services) are FT3 and FT5 referring to instant insurance and roboadvisor, respectively.

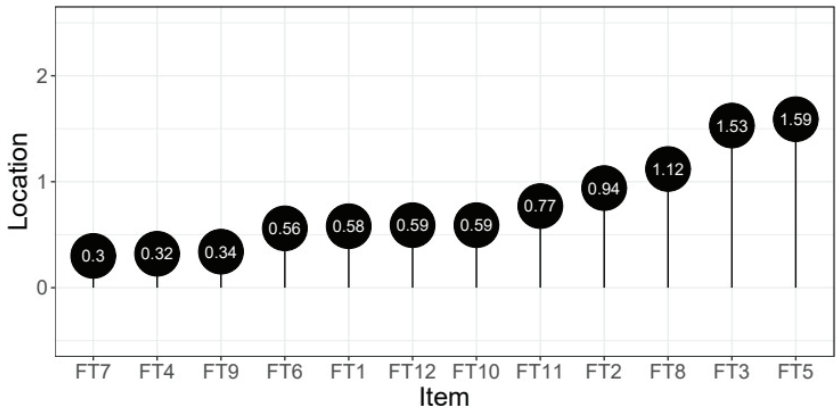


Figure 1: Average item difficulty

Regarding the structural part of the model, the estimates of second-level random effects u_j $(j = 1, \dots, 10)$ allow accounting for the regional impact of belonging on the individual financial knowledge. Accordingly, Figure 2 displays the caterpillar plot with the ranking of the 10 considered Italian regions according to point estimates of u_j (and the related confidence interval at 95%), where increasing values of u_j denote increasing levels of financial knowledge.

Results show that almost all the confidence intervals overlap zero, highlighting a non-significant effect of regions, with the exception of Campania, Lombardy, and Veneto. Specifically, confidence intervals indicate that financial knowledge in Campania and Lombardy is significantly worse than in Veneto.

5. Conclusions

Nowadays, evaluating individuals' knowledge level of modern FinTech tools (e.g., instant payments, artificial intelligence, crowdfunding) is a crucial concern, as they represent the new frontier of digital finance.

The literature has highlighted that financial knowledge can be related to more general geographical macro-area differences (see, among others, D'Alessio et al., 2021).

The present work deepened this issue by exploiting a multilevel approach to further explore Italian regional differences in the FinTech framework, concurrently accounting for the latent nature of the FinTech knowledge construct through the IRT parameterization. As such, the proposed approach allows investigating if individual differences also reflect differences at the regional level regarding, for example, technical advancement and governmental initiatives.

The obtained results, and in particular the uncommon position of the Lombardy region, suggest to investigate in future research the individual (e.g., sex, age, education level, and numeracy) and regional (e.g., GDP and unemployment rate) characteristics to provide a deeper understanding of the regional differences and individual-level heterogeneity. Indeed, as pinpointed by Grohmann and Menkhoff (2020), individual differences in financial inclusion and literacy raise questions regarding their impact on the new FinTech contests: For example, several studies focused on the "FinTech gender gap" (Khera et al., 2022).

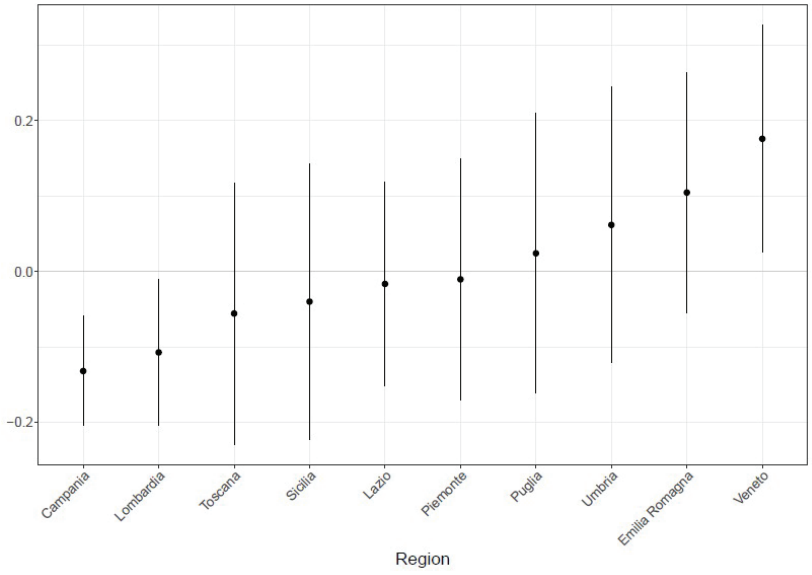


Figure 2: Caterpillar plot: estimated second-level random effects with 95% confidence interval

References

D'Alessio, G., De Bonis, R., Neri, A., Rampazzi, C. (2021). Financial literacy in Italy: The results of the Bank of Italy's 2020 survey. *Politica economica*, **37**(2), pp. 215–252.

Fox, J.P., Glas, C.A. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, **66**, pp. 271–288.

Grohmann, A., Menkhoff, L. (2021). The relationship between financial literacy and financial inclusion, in *The Routledge Handbook of Financial Literacy*, eds. G. Nicolini, B. J. Cude, Routledge, New York, (NY), pp. 517–530.

Hambleton, R. K., Swaminathan, H. (2013). *Item Response Theory: Principles and Applications*. Springer Science & Business Media, Berlin.

Khera, P., Ogawa, S., Sahay, R., Vasishth, M. (2022). *Women in Fintech: As Leaders and Users*. International Monetary Fund, Washington, (US).

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, **47**(2), pp. 149–174.

Muraki E. (1992). A generalized partial credit model: application of an EM algorithm. *Applied Psychology Measurement*, **16**, pp. 159–76.

Rabe-Hesketh S, Skrondal A, Pickles A. (2004). *GLLAMM Manual*. UC Berkeley division of biostatistics working paper series. University of California, Berkeley, CA.

Robb, C.A., Woodyard, A. (2011). Financial knowledge and best practice behavior. *Journal of Financial Counseling and Planning*, **22**(1), pp. 60–70.

Stapleton, L.M., McNeish, D.M., Yang, J.S. (2016). Multilevel and single-level models for measured and latent variables when data are clustered. *Educational Psychologist*, **51**(3-4), pp. 317–330.

doi.org/10.26398/asaproc.0037

Urbanisation and counter-urbanisation in Italy, 2001-2022[a]

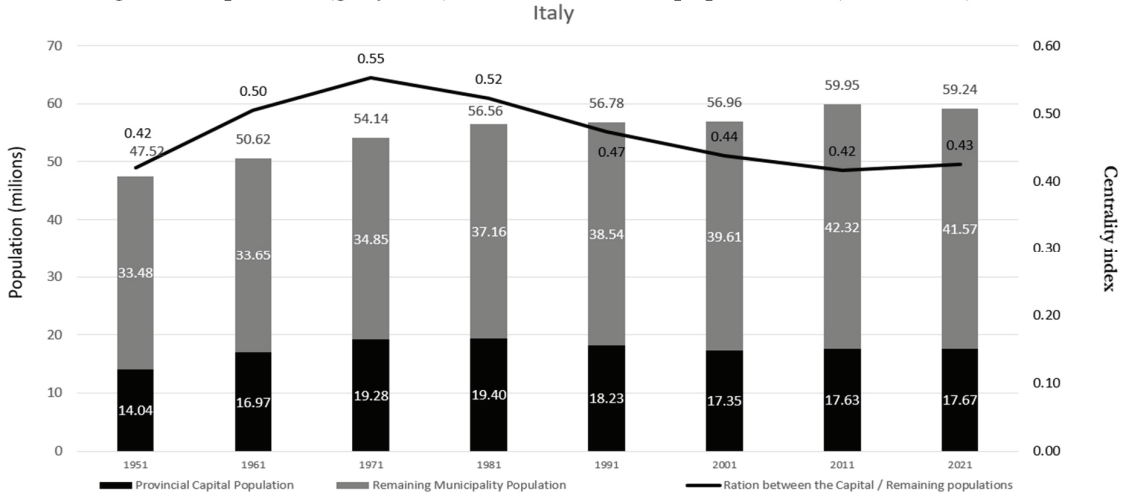
Luigi Fabbris[*], Paolo Feltrin[*], Sergio Maset[**],
[*] Tolomeo Studi e Ricerche, Padua and Treviso, Italy
[**] Idea, Treviso, Italy

1. Introduction

In the year 1981, Italy's population reached almost 57 million inhabitants compared to 47 million after World War II. In the following twenty years, the population basically stagnated, after which it only experienced a new (brief) expansionary phase in the new millennium. Three phases can be identified, based on the rise or decline of the population in the provincial capital cities (from now on also PCCs), relative to the rest of the provincial territories. From 1951 to 1981, an intense process of urbanisation is observed; this declines in the second phase (1981-2001), marked by significant dynamics of de-urbanisation; while in the third phase (2001-2021), significant, albeit limited, signs of re-urbanisation were observed.

Figure 1 shows the time series of Italian population from 1951 to 2021. In year 1 of each decade, the population in the provincial capital city or in the province excluding the capital has been calculated. When the ratio between the two, represented by the piecewise linear curve, increases, the population in the capital rather than in the provincial areas increases.

Figure 1: Total Italian population, divided between provincial capitals (black bar) and remaining municipalities (grey bar), ratio of the two populations (black line).



In the period between 1981 and 2001, the population remained substantially stable, but with a shift towards a higher share of inhabitants residing outside the PCCs. At least three dynamics come together in this movement towards the periphery: the coming of age of the baby-boom generation, which moves from the capitals towards the urban belts and into the province, attracted by better housing conditions; the demand for industrial labour, which makes it convenient to build new low-cost factories in the industrial areas built in rural municipalities; the growth of public and private services (schools, clinics, supermarkets, etc.) also in the peripheral areas. The population of the PCCs contracted from 19,400,000 in 1981 to 17,300,000

[a] The authors wish to thank Marco Bressan, Luca Garavaglia and Luca Pavan (Idea) for their collaboration.

in 2001, against the background of a very slight growth in the total population. The period coincides with the phase of so-called post-Fordist economic development that began in the second half of the 1970s (Feltrin, Pero 2021; Rullani 2021). In the first phase, population growth mainly concerned the capital cities: in 1971, in fact, the census recorded the highest concentration in the PCCs, with a resident population equal to 36% of the total,[b] which remained roughly similar in 1981 (Figure 1).

In the first decade of the 2000s, a new phase of overall population growth was recorded, mainly fuelled by immigration from abroad, resulting in population growth both in the PCCs and in the rest of the provinces. A first trend reversal is thus observed, i.e. a positive balance in the PCCs' resident population, something that had not been recorded since the 1970s. In the second decade of the 21st century, the change of phase is confirmed, with an increase in the relative weight of the provincial PCCs regarding the resident population. What has been observed at the national level results from the combination of phenomena of the same type, albeit with different intensities, in the macro-areas of the country. Moreover, there are numerous phenomena that have occurred over time that can be analysed in correlation with the urbanisation, de-urbanisation and new urbanisation phases. In addition to the already mentioned periodisation of the economic cycle (Rullani, 2023), other significant factors can be found in lower birth rates, immigration and the combination of both with the transformation of labour demand.

Only some of these elements are developed in this paper. Specifically, we will evaluate the weight of the following socio-demographic components: social burden, labour market attractiveness, renewal rate. Have the cities that had an older population between the late 1990s and the beginning of the new millennium seen their age composition remain unchanged, or is the resurgence of centrality accompanied by a relative slowdown in the ageing process? Is the PCCs attractiveness for the workforce also increasing over time? Furthermore, an attempt will be made to understand the relationship with the overall growth processes of the territory: is there a relationship, and if so of what type, between variation in the provincial population and the weight of the provincial capital city? Finally, the existence of an ecological correlation between the variation in the degree of relative specialisation in the main economic macro-sectors and the variation in the centrality index of the PCCs will be investigated.

The rest of this paper is organised as follows. Section 2 introduces the data and the model used for data analysis. Section 3 presents the main results of the statistical analysis. Section 4 discusses the results with reference to the mainstream literature on urbanisation.

2. Data and analytical model

Given:
P, a population (the number of people living in a certain province capital or remaining region);
W, a workforce (the number of people working in a certain region and / or economic sector)
$c = 1, \ldots, 107$ a PCC;
$r = 1, \ldots, 107$ a remaining area (province without the PCC);
p-q = a range of ages (e.g., 0-18 means between 0 and 18; 65+ means 65 or greater);
s = *{Warehousing, Manufacture, Land services, Expansion support services, Tourism, Construction, Personal services, On-demand business services}* is an economic sector.

Note that c and r in the same formula refer to the same province. The capitals have been reclassified into three size classes and are distributed in the macro-areas, as established by the Italian National Institute of Statistics (ISTAT).

The following indicators were analysed:

[b] The 1971 peak is also recorded when considering the median of the index value, although obviously lower than the index referring to the cumulative data (0.338 vs. 0.553).

Centrality index : $$I1 = \frac{P_c}{P_r} \; ; (1)$$

Δ Productive attractiveness : $$\Delta I3 = I3_c - I3_r \; ; (2) \quad \text{where } I3_c = \frac{W_c}{P_{c_{35-64}}} \text{ and } I3_r = \frac{W_r}{P_{r_{35-64}}} \; ;$$

Δ Social load : $$\Delta I4 = I4_c - I4_r \; ; \; (3) \quad \text{where } I4_c = \frac{P_{c_{0-18}} + P_{c_{64+}}}{P_{c_{19-64}}} \text{ and } I4_r = \frac{P_{r_{0-18}} + P_{r_{64+}}}{P_{r_{19-64}}} \; ;$$

Δ Demographic renewal : $$\Delta I5 = I5_c - I5_r \; ; \; (4) \quad \text{where } I5_c = \frac{P_{c_{0-18}}}{P_{c_{65+}}} \text{ and } I5_r = \frac{P_{r_{0-18}}}{P_{r_{65+}}} \; ;$$

Relative population change between 2019 and 2012: $$\Delta P_{c_{2019-2012}} = \frac{P_{c_{2019}} - P_{c_{2012}}}{P_{c_{2012}}} * 100 \; ; \; (5)$$

Relative Specialisation Index 2019: $$SRI_{cs} = \frac{\frac{W_{cs}}{W_c}}{\frac{W_s}{W_{area}}} \; ; \; (6)$$

3. Results

Table 1 shows the statistical population frequencies of PCCs grouped by geographic area. Considering the demographic size, it can be observed that the dynamics of growth in the concentration index tend to characterise the larger municipalities and, to a lesser extent, those of medium size, while the smaller PCCs are affected by relative weight stability. If we cross-reference size and area, the groups of statistical units are too small to provide reliable statistics, so, for further analyses, we group them into two macro-areas (North+centre, South+islands).

Table 1: Number of Italian provincial capital cities in 2023 by demographic size and geographical macro-area.

Inhabitants	Geographical area					
	Nord west	North east	Centre	Islands	South	Italy
Less than 50,000	9	3	5	4	4	25
50,000 - 150,000	12	10	12	7	15	56
More than 150,000	4	9	5	3	5	26
Total	25	22	22	14	24	107

In the north-west of Italy, the maximum level of the centrality index of the largest PCCs was reached in 1961 and a steady decline was recorded until 2010, when a shift took place. The numerous small PCCs showed no appreciable change. In the North-east, the maximum centrality in the largest municipalities was reached in 1981 and here too there was a steady decline until 2010, the year after which an even more noticeable change of phase can be noted. The medium-sized PCCs show a similar trend to the larger ones, with the peak in 1971 and a stabilisation since 2000. In the larger PCCs of the Centre, the maximum weight is recorded in 1971, and here too the rebound occurs around 2010. The medium-sized municipalities show a similar trend, while the variation in smaller municipalities is not significant. Similar trends are also recorded in the larger southern cities, with the index stabilising from the beginning of the new millennium and, unlike elsewhere, the weight of the PCC in the smaller towns tends to increase steadily. However, this increase of centrality is shown in a context of demographic decline.

The correlation coefficients between the change in population and the change in the centrality index of the PCCs were also analysed (Table 2). Where a statistically significant relationship emerges, the sign is always negative, a fact that depends on two different situations: at times when the provincial population increases, the provincial capital loses relative importance; when the population decreases, the capital increases in importance or experiences a smaller decline in population than the rest of the province. The periods in which the relationship is significant correspond to the phase of first urbanisation and then, in the first decade of the 2000s, to the re-urbanisation phase. Looking at the data by individual macro-areas, it can be observed that the dynamic is driven by the North-west.

Table 2: Correlation coefficients between change in the relative population and change in the centrality index, Italy.

Years	*Correlation coefficient*	*p-value*	
51-61	-0.389	<.001	***
61-71	-0.489	<.001	***
71-81	-0.188	0.053	
81-91	0.012	0.904	
91-01	-0.216	0.025	*
01-11	-0.404	<.001	***
11-21	-0.152	0.117	

Secondly, it should be noted that over the last twenty years the PCCs have increased their *productive attractiveness*, measured by the ratio of those employed in the capital city to the population aged 35 to 64 in the province. The same index was calculated for the rest of the province and the difference (Δ) was calculated. The medium-sized and large capitals of the northern and central regions significantly increased their attractiveness over the past two decades. Although with some variations depending on the size class, the productive attractiveness of the capital cities also tends to increase in the South and the Islands. The difference (Δ) of the *social load index* has been decreasing over the last decade in the central and northern provinces and increasing in the southern ones. Finally, the difference (Δ) of the *demographic renewal index* has been increasing in the PCCs of the northern and central regions over the last twenty years, while in the southern ones it increased only in the larger cities and decreased in the Islands.

Analysing the two components of the Δ separately, we find that the renewal index is decreasing both in the PCCs and in the rest of the province with, however, a relatively more favourable ratio in the capitals. Thus, a certain similarity emerges between the macro regions of the Centre and the North with respect to the South and the Islands.

Also, we analysed the relationship between the 2012-2019 change in the capital centrality index and the change over the same period in the relative specialisation index in the various sectors (Table 3), separately for the north and centre, on the one hand, and the south and the main islands of Italy, on the other.[c] Regarding the North+centre area, the multivariate analysis showed a significant positive relationship with an increasing variation in the relative specialisation index in the sectors of construction, advanced services and tourism. However, a negative relationship emerges with an increase in the relative specialisation index in personal services and business services. So, the hypothesis that cities where the concentration index increases showed a relative specialisation in advanced services increase, while their relative specialisation in conventional services decreased, appears to be confirmed. Regarding the other southern regions, the PCCs that recorded an increase in relative specialisation in the hotel and restaurant sector in the 2012-2019 period also see their concentration index increase.

Finally, using the relative change of the population ($\Delta P_{c2019\text{-}2012}$) as a dependent variable, 16 regression analyses were carried out, one for each of the specialisation indices for year 2019 (separately for two groups, below and above the median of the indices) crossed with the two Italian partitions. A significant relationship emerged with the relative specialisation index in

[c] Sectors are identified by the Ateco code (https://www.istat.it/it/archivio/17888): Land services: 35-39; Expansion support services: 62, 70, 72, 73; Tourism: 55, 56; Construction: 41-43; Personal services: 45, 47, 75, 85-88, 90-93, 95, 96; On-demand business services: 46, 49-51, 53, 58-61, 63-66, 68,69,71, 74, 77, 79-82; warehousing: 52; Manufacture: 10-33.

advanced services (Expansion support services) only for the North+centre PCCs in the "below the median" group. No significant relation was instead observed in the southern regions.

Table 3: Estimates and significance of beta regression coefficients of the relative change in the concentration index, by Italian region

	North+centre			South and islands		
Predictors	b	Se	T	b	Se	T
Construction	0.303	0.124	2.449*	-0.319	0.181	0.089
Warehousing	0.164	0.119	1.372	-0.236	0.186	0.217
Manufacture	-0.153	0.110	-1.383	-0.227	0.164	0.177
Expansion support services	0.411	0.117	3.509***	0.248	0.168	0.153
Personal services	-0.206	0.110	-1.866	0.027	0.208	0.896
Land services	0.118	0.114	1.040	0.199	0.166	0.242
On-demand business services	-0.391	0.128	-3.058**	0.061	0.209	0.771
Tourism	0.211	0.113	1.867	0.427	0.193	0.037*
Overall provincial population	-0.038	0.127	-0.300	-0.039	0.172	0.823

Table 4: Regression coefficients for the two groups in the northern+central regions, above and below the median (39.05) of the specialisation index for advanced tertiary services.

	b	se	T	p-value	
Below median provinces	0.161	0.062	2.607	0.012	*
Above median provinces	0.015	0.013	1.242	0.219	
Contrast	0.145	0.063	2.319	0.024	*

An analysis of the 2019-21 pandemic period, shows no difference for all indices with respect to the previous years. The municipalities with a larger specialization index are in almost all cases recognisable as university poles. The only exceptions are the cities of Turin and Vicenza, which, despite a high level of the index, recorded a decrease in population.

4. Discussion and conclusion

Compared to the United States, England, and France (Marksuen, Schrock, 2006; Power et al., 2010; Moretti 2012; Dijkstraet al., 2013; Martin et al., 2016; Piketty, 2023), in Italy the phenomenon of re-urbanisation was delayed - undeniable only after 2010 – and less evident (namely, the demographic decline in the PCCs is slower than in the rest of the province). Especially in the northern and central regions, there are cities that perform well in advanced services and see their weight grow more. In the southern regions and the main islands, a positive relationship emerges only with tourism. As has been observed in the literature (Moretti, 2012; Lamorgese, Petrella, 2018; Accetturo et al., 2019; Istat, 2023), tertiarisation has been the engine of economic growth over the last thirty years in many developed countries and this was responsible for the new leadership role assumed by large cities.

In Italy, at certain moments in its economic history, backwardness was turned into a competitive advantage. One might then ask whether we are really in the midst of a new phase (tertiarisation plus re-urbanisation) or we are observing the tail end of a process of establishing the network economy. In the first case, our country's backwardness should be quickly bridged with policies similar to those adopted in other countries to trigger processes of metropolitan tertiarisation (Martin et al., 2014). If, on the other hand, we are at the end of a path and are witnessing a progressive reduction in the relevance of the (metropolitan) factor of scale, also as a consequence of the post-covid readjustment, then a new window of opportunity would open

up – especially for latecomers like Italy – as a consequence of the diffusion of remote working and flexible work organisation.

On the other hand, Koolhaas (2020) assumes that urban life necessitates a return to the countryside, in which a new organisation and automation is on the way. China is working to connect the countryside to 4g and China's rural areas are now more connected than American ones (Aresu, 2022). The question of which is the real driver of contemporary development – advanced services or the network economy – and what are its medium-term effects remains open, inevitably destined, as Kleinknecht (2020) suggests, to be the subject of further investigation.

Hence, these analyses should be further articulated considering the development not only of the provincial capital city but also of the metropolitan cities or of the urban system identified by the capital and its belt, in order to verify whether a lack of growth in the PCC was passed on to its urban belt or to a larger set of municipalities. The relationship with value-added indicators is also to be verified.

References

Accetturo, A., Lamorgese, A., Mocetti, S., Sestito, P. (2019). Sviluppo locale, economie urbane e crescita aggregata, *Questioni di Economia e Finanza*, **490**, Roma, Banca d'Italia.

AMO, Koolhaas, R. (2020). *Countryside. A Report*, Cologne, Taschen.

Aresu, A. (2022). *Il dominio del XXI secolo. Cina, Stati Uniti e la guerra invisibile della tecnologia*, Milan, Feltrinelli.

Aa. Vv. (2021). *Dalla prima alla quarta rivoluzione industriale. Storia delle relazioni industriali dei metalmeccanici,* Milan, Solferino-Rcs.

Dijkstra, L., Garcilazo, E., McCann, P. (2013). The economic performance of European cities and city-regions: Myths and realities*, European Planning Studies*, **21**(3), pp. 334-354.

Feltrin, P., Pero, L. (2021). *Breve lettura del lungo addio al fordismo (1970-2020)*, in Aa.Vv. (2021), pp. 424-532.

Istat (2023). P*rofili delle città metropolitane. Molte fragilità ma anche potenzialità dei contesti urbani*, Focus Statistiche, 2 febbraio, Roma, Istat.

Kleinknecht, A. (2020). The (negative) impact of supply-side labour market reforms on productivity: An overview of the evidence, *Cambridge Journal of Economics*, **44**(2), pp. 445–464.

Lamorgese, A., Petrella, A. (2018). Le città italiane: definizioni, caratteristiche e crescita, *Questioni di Economia e Finanza* (Occasional Papers), **454**, Roma, Banca d'Italia.

Markusen, A., Schrock, G. (2006). The distinctive city: divergent patterns in growth, hierarchy and specialisation, *Urban Studies*, **43**, pp. 1301-1323.

Martin, R.L., Gardiner, B., Tyler, P. (2014), *The Evolving Economic Performance of UK Cities: City Growth Patterns*, 1981-2011, Working Paper, Foresight Programme on The Future of Cities, London, Government Office for Science.

Martin, R.L., Sunley P., Tyler P., Gardiner B. (2016). *Divergent Cities in Post-Industrial Britain*, Working Paper, Foresight Programme on The Future of Cities, London, Government Office for Science.

Moretti, E. (2012). *The New Geography of Jobs*, Boston, Houghton Mifflin Harcourt.

Piketty, T. (2023). France and its territorial divides, in Le blog de Thomas Piketty, *Le Monde.fr*, 11 Juillet (https://www.lemonde.fr/blog/piketty/2023/07/11/france-and-its-territorial-divides/).

Power, A., Plöger, J., Winkler, A. (2010). *The Fall and Rise of Great Industrial Cities*, Bristol, The Policy Press.

Rullani E. (2021). *Lavoro in transizione: trasformazioni delle imprese e nuove forme di lavoro*, in Aa. Vv. (2021).

Rullani E. (2023). Fordismo, postfordismo, digitalizzazione: l'orizzonte strategico del capitalismo. Intervista a Enzo Rullani di A. Alia, *Machina*, 24/5/2023 (https://www.machina-deriveapprodi.com/post/fordismo-postfordismo-digitalizzazione-l-orizzonte-strategico-del-capitalismo).

doi.org/10.26398/asaproc.0038

Comparing parametric and nonparametric models: An application to university students' rental prices

Matteo Farnè[a], Elena Grimaccia[b], Annalisa Sonnati[a]
[a] Department of Statistical Sciences, University of Bologna, Bologna, Italy.
[b] Istat – Istituto Nazionale di Statistica, Roma, Italy.

1. Introduction

This study analyses two statistical methodologies based on different approaches: multiple linear regression, which is a parametric approach relying on assumptions about the shape of the target variable distribution (i.e., assume a normal distribution) in the underlying population and about the form or parameters (i.e., means and standard deviations) of the assumed distribution, and regression tree, a non-parametric approach which presents the disadvantage to provide no statistical inference, but the advantage to be much more flexible than the linear regression function and to provide a hierarchy of explanatory variables best explaining the dependent variable. These methodologies are applied on a topic of overwhelming relevance in present-day Italy: rental prices for students' accommodation.

Previous literature (Capozza et al., 2002; Reed, 2016; Grum and Govekar, 2016) shows that house rental and sales prices are sensitive to the demographic and socioeconomic conditions of the areas in which houses are located. Likewise, it is well known that significant urban changes occur due to the increasing proportion of students within a city: the concept of studentification (Smith, 2002; Smith and Holt, 2007) has been introduced to illustrate how increased participation of young people in higher education has caused the explosion of student housing demand (Smith, 2002; Munro, Turok and Livingston, 2009). The increased population in large city centres then inevitably leads to a significant increase of rental prices (Égert and Mihaljek, 2007).

In Italy, 660,000 Italian students, representing 39.9% of university enrolments, were living away from their parents' homes in 2022 (Il Sole 24 Ore, May 2023). As a reaction to the recent increases of rental prices, the "tent protest" emerged in May 2023, which brought thousands of students to demonstrate in front of university campuses throughout Italy. Triggered by that event, we study the macroeconomic and sociodemographic characteristics of Italian university cities that determine the rental prices of single rooms for students, with reference to the single room monthly price in 2022.

We employ a sample of 32 Italian chief towns, drawn by the *Immobiliare.it Insights* report, reporting monthly single room average prices in 2022. Following the relevant literature, we have then considered 50 variables, collected by ISTAT from various official sources referring to years 2021 and 2022, that describe the macroeconomic, environmental, and sociodemographic conditions of these 32 cities. These variables pertain to immigration, income, (un)employment, poverty, crime, pollution, and living costs. In what follows, we aim to determine which of these indicators, measuring population, air pollution, labour market, families' conditions, and several other socio-economic features, do affect 2022 rental prices of single rooms in Italian university cities. After a thorough sample description, we proceed by a regression analysis and a regression tree approach, to infer which of these indicators influence single room prices the most.

2. Sample description

Figure 1 reports the distribution of single rooms monthly prices in 2022 across the 32 Italian cities of the sample. From the graph, one can visualize that major university cities record the highest rents for single rooms. According to the *Immobiliare.it Insights* survey, a student in Milan pays an average of €620 per month. Milan is followed by four cities, Rome, Padua, Florence, and

Bologna, which all present values between €450 and €500. The most expensive city in southern Italy is Naples, with an average of €337 per month. In other southern Italian cities, the price of a single room is less than €280/month. The least expensive city to move to is Pescara.

The explanatory variables included in this study cover demographic aspects such as residents, foreigners, and university population for each city included in the sample. Variables related to the economic condition of citizens, such as wages and unemployment rate, are included. Some indicators of citizens' quality of life, such as the crime rate and available living space in individual cities, are considered. Environmental factors, such as pollution levels studied through concentrations of particulate matter (with a diameter of 10 and 2.5 microns) and of nitrogen dioxide, are also examined. Additionally, household spending is incorporated at a macro-regional level.

Figure 1. Distribution of monthly single room prices (source: *Immobiliare.it Insights*, 2022)

3. Statistical methodology

To reach our goal, i.e., to model single rooms monthly prices in 2022 across 32 Italian cities, we consider two different statistical methods: multiple linear regression (parametric) and regression (nonparametric) tree.

The former method requires to specify a linear regression model, and to estimate it by ordinary least squares, which imposes that the residuals are homoscedastic, uncorrelated, and normally distributed, and the explanatory variables are not strongly correlated. For this reason, it is first necessary to test for the normality of our dependent variable. From the results of Shapiro-Wilk test (Shapiro and Wilk, 1965), it can be observed that the test for the price of single rooms returns a significant p-value (0.0015): therefore, it can be concluded that our dependent variable is not normally distributed.

This fact can also be deduced from the histogram of the dependent variable (see Figure 2), which shows positive skewness. From Figure 2, it can be inferred that there is one observation whose value significantly deviates from the others, namely, the city of Milan, which is the most expensive city to move to for studying with its average price of 620€ per month. This fact suggests including in the model a specific dummy for Milan, to catch its stand-alone position in the Italian distribution of monthly single room prices. Similarly, the peak between €450 and €500 represents Rome, Padua, Florence, and Bologna, which also present a specific position to be caught in the model.

From these preliminary findings, we consider the use of Box-Cox transformations (Box and Cox, 1964) for the dependent variable and all the other variables presenting a non-normal distribution shape (that is, a significant p-value for Shapiro-Wilk test). For each variable, we select the parameter of Box-Cox transform which maximizes the normal log-likelihood of the transformed sample. In the end, each transformed distribution shape will be as close as possible to the normal one.

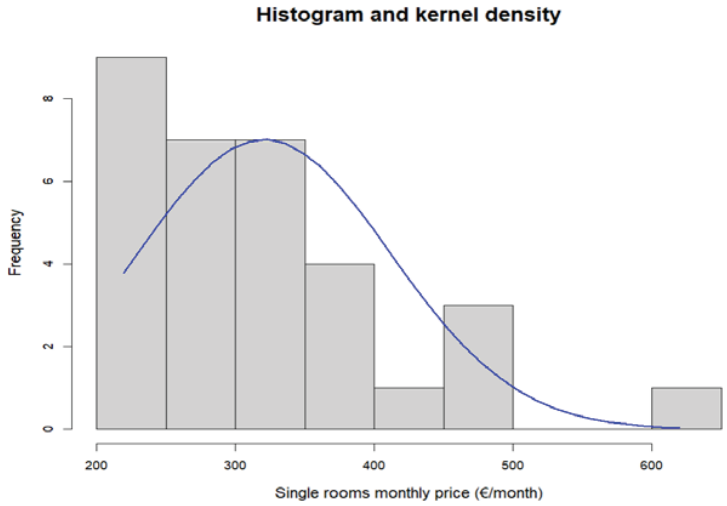


Figure 2. Histogram and kernel density of monthly single room prices in 32 Italian cities in 2022

Whenever there is a large set of independent variables, the need arises to identify a subset of factors that actually influence the dependent variable. To achieve this, there are specific methodologies aiming to construct regression models including only the most influential independent variables in the model, while still ensuring a good predictive accuracy. Following this path, to reduce multicollinearity, a subset of significant explanatory variables has been retrieved by a backward selection approach (Miller, 1990).

On the other hand, we build a regression tree by CART approach (Breiman, 1984). CART procedure iteratively subdivides the domain of explanatory variables to obtain the minimum within deviance of the resulting partition. The sub-regions of the domain in the partition are the leaves of the tree. This recursive process has the relevant advantage to be parameter-free and to identify the most important covariates in explaining the dependent variable, which are responsible for the very first splits in the tree growth. At the end of this process, homogeneous subsets of units are obtained within the leaf nodes.

To prevent the construction of an overly complex tree, the final tree is then "pruned" (Breiman, 1984) to obtain a parsimonious and interpretable tree with a limited number of leaves. This final step is performed by minimizing a proper cost-complexity function. In the end, this method allows to obtain a sequence of decision rules to estimate the expected value of the target variable according to the predictor values, and to establish a ranking of predictors according to their predictive power.

4. Statistical results

Three linear regression models are implemented, differing in the scale of dependent and independent variables. All models include two dummy variables: POLO and MILANO, which characterize the most important university cities in Italy. POLO assigns value 1 to the cities of Rome, Bologna, Florence, and Padua, and 0 to the others. MILANO identifies the Lombardy capital, which registers anomalous values on both sides of the regression models. This territorial

component is included to understand if the dynamics of the rental market differs significantly among cities, i.e., if there is a territorial polarization.

The first estimated model maintains the variables in their original scale, the second model includes Box-Cox transformations where necessary to achieve normally distributed variables, and the third model uses the natural logarithm to stabilize residual variance (Table 1). Backward elimination is employed to select the most significant variables. After this process, the first model yielded the best result due to the best fitting properties, and more importantly, because the residual diagnostics of the first model is more closely aligned with the theoretical assumptions behind OLS estimation.

	Dependent	Independent	AIC	BIC	R^2	Adj-R^2
1	Original Scale	Original Scale	220.9	325.7	0.9276	0.9065
2	Inverse (1/Y)	(some) Box-Cox	-504.1	-395.1	0.9104	0.8258
3	Natural log (log(Y))	Original Scale + log(wages) + log(residents)	-142.4	-30.5	0.9272	0.8677

Table 1. The three regression models implemented. The reported indicators are not directly comparable, because of the different scale of variables, apart from the R^2, which is obtained by a retro-transformation of predicted values to the original scale.

The final model output is reported in Table 2. It includes four regressors: residents (Residents_22), 15-24 years old employment rate (Employ_rate_15-24_22), and the two dummy variables identifying the type of city (POLO and MILANO). Estimated coefficients show that resident population, employment conditions, and being a university city are strongly significant predictors of the monthly single room prices in 2022 in Italy. As expected, their impact on the single room prices is strongly positive.

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	1.281e+02	2.332e+01	5.492	8.14e-06	***
Residents_22	3.128e-05	7.404e-06	4.224	0.000244	***
Employ_rate_15-24_22	7.006e+00	1.032e+00	6.790	2.72e-07	***
MILANO	2.383e+02	3.567e+01	6.680	3.60e-07	***
POLO	1.221e+02	1,794e+01	6.750	3.01e-07	***

Table 2. Final linear regression model output. The symbols '.', '*', '**', '***' denote statistical significance at 10%, 5%, 1%, 0.1%, respectively.

Regarding CART approach, two trees were constructed: the first includes only the quantitative variables used in the regression, while the second also considers a UNIVERSITY_CENTER categorical factor with three values: POLO, MILANO and OTHER. In the first tree, not reported, the most explanatory features are crime rate, youth employment rate, and average annual wages. The second tree, reported in Figure 3, attributes instead a considerable importance to the qualitative factor UNIVERSITY_CENTER. The two trees are indeed strongly connected, in that the cities identified by POLO and MILANO are also characterized by higher crime rates, youth employment rates, and higher wages, because these features constitute a very good proxy to identify the largest Italian university cities. Figure 3 highlights that being a university city is by far the most relevant predictor of high rent prices, as clearly shown by the first tree branch length.

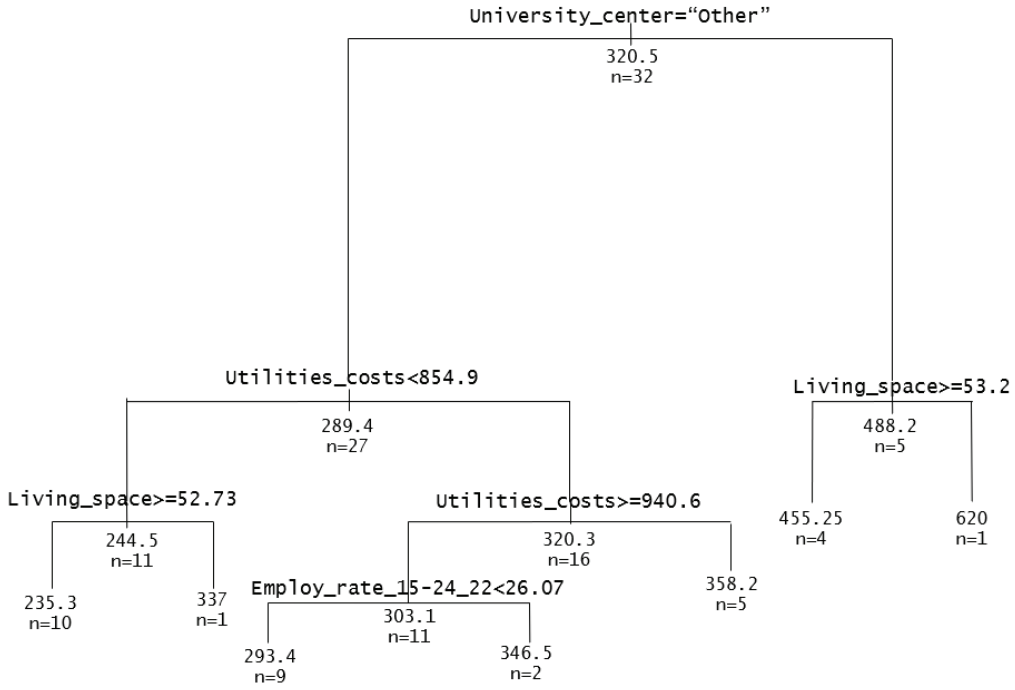


Figure 3. Pruned regression tree output with the factor UNIVERSITY_CENTER.

5. Conclusions

This study has shown, by a regression analysis and a regression tree approach, that the Italian rental market is strongly polarized between major Italian university cities and the other cities. Regression analysis identifies as significant predictors population density (measured by the number of residents), and employment (captured by the youth employment rate), beyond a strong major university city effect, which uncovers that five Italian cities present a significantly higher level of monthly room rental prices compared to others: Rome, Bologna, Florence, Padua, and above all, Milan. Regression tree shows that the same major university city effect is the most relevant predictor of a high rent price, followed by far from utilities costs, young employment rate, and living space.

This study could be further expanded by investigating rental costs in other Italian cities as soon as that data becomes available. New variables related to cities could also be included, such as inflation rate or university characteristics, since the dynamics of the Italian real estate market may depend on a multitude of factors. However, the insights provided by this research could be useful to design public policies aiming to guarantee an equitable access to university level studies for any Italian student.

References

Box, G.E., Cox, D.R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, **26**(2), pp. 211-243.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. (1984). *Classification and Regression Trees.* Wadsworth, Belmont, CA.

Capozza, D.R., Hendershott, P.H., Mack, C., Mayer, C.J. (2002). *Determinants of Real House Price Dynamics*. National Bureau of Economic Research, Cambridge, MA, USA.

Égert, B., Mihaljek D. (2007). *Determinants of House Prices in Central and Eastern Europe*.

Switzerland: Bank for International Settlements Press and Communications CH-4002 Basel.

Grum, B., Govekar, D.K. (2016). Influence of macroeconomic factors on prices of real estate in various cultural environments: Case of Slovenia, Greece, France, Poland and Norway. *Procedia. Economics and Finance,* **39**, pp. 597-604.

Miller A.J. (1990). *Subset Selection in Regression*. Chapman and Hall.

Munro, M., Turok, I., Livingston, M. (2009). Students in cities: A preliminary analysis of their patterns and effects. *Environment and Planning A*, **41**(8), pp. 1805-1825.

Reed, R. (2016). The relationship between house prices and demographic variables: An Australian case study. *International Journal of Housing Markets and Analysis,* **9**(4), pp. 520-537.

Shapiro, S.S., Wilk, M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, **52**(3/4), pp. 591-611.

Smith, D. (2002). Patterns and processes of 'studentification' in Leeds. *The Regional Review*, **12**, pp. 14-16.

Smith, D.P., Holt, L. (2007). Studentification and 'apprentice' gentrifiers within Britain's provincial towns and cities: Extending the meaning of gentrification. *Environment and Planning A*, **39**(1), pp. 142-161.

doi.org/10.26398/asaproc.0039

E-invoicing as essential administrative source for the future of official business statistics: The case of short term statistics

Salvatore Filiberti, Giuseppe Amato, Danilo Orsini,
Giancarlo Bruno, Fabio Massimo Rapiti
Istat, Rome, Italy

1. Introduction

In a world that nowadays is almost completely and - in perspective - will totally be digitised, the use, for statistical purposes, of data acquired by central administrations will represent the standard model for the production, by the National Statistical Institutes (NSI), of official quantitative information on the main economic variables. In particular, the integration of survey and administrative data makes it possible to build a robust economic statistics system capable of returning very detailed and timely information to users, while keeping the statistical burden on respondents at an acceptable level.

The recent introduction of the rules on electronic invoicing has opened up new scenarios in Italy, also with regard to the compilation of official economic statistics. Data acquired by the Revenue Agency (RA)[1] and used for statistical purposes by the NSI could represent one of the main sources of administrative quantitative information on economic variables of enterprises. The practice of using administrative data is also envisaged and strongly suggested by the European legislation on official statistics with the aim to reduce the statistical burden on businesses, especially for small economic units.

There are many concrete perspectives of using administrative information about e-invoicing for statistical purposes, for example as up-to-date information to establish the activity status of an enterprise and thus to update business registers, or as information on the turnover levels of enterprises, which is necessary to produce short-term turnover indices on a monthly basis.

E-invoicing may also be used in surveys, as an auxiliary variable, both when defining the sample design, but also in the estimation procedures, according to the calibration estimators approach, as an additional reference variable in order to obtain more accurate and detailed final estimates of the variables of interest.

Moreover, turnover from e-invoices may represent an interest variable in the frame of structural business statistics and then used to produce the extended business statistics registers. It is also possible to explore flows of goods/services between the transferor/provider and the transferee/customer in order to identify the production chains and quantify/qualify trade from and to abroad, etc.

The focus of this paper is to show first experimental results concerning the use of electronic invoicing data as a source for the compilation of monthly short-term economic statistics on companies' turnover. In Section 2 some basic information on main short-term business statistics sources on turnover and the tax administrative source of e-invoicing is provided. The preliminary analyses conducted on survey and administrative data and main results obtained are shortly shown in Section 3. Section 4 summarises some considerations regarding the prospect of using e-invoicing data for short-term business statistics.

[1] Special thanks go to the Revenue Agency, Resources Division, Central Directorate for Technology and Innovation, Database Sector, for the cooperation and willingness to provide the necessary data for the experimentation.

2. Short term business statistics and e-invoicing

Istat's system of short-term economic statistics is based on well-established system of direct statistical surveys that provide quality results with reference to the main indicators, also in compliance with Regulation (EU) 2019/2152 and within the strict deadlines provided for. The production processes of the surveys, however, are not without operational costs, especially for the contacted enterprises, which have to spend time and resources to provide the required information.

The recent introduction of the law on the obligation of companies to notify the RA of the details of e-invoicing transactions has made it possible to set up a continuous data flow and a database with a huge amount of information. Such information, always respecting the confidentiality of the data, can be available for the institutional purposes of other administrations. Istat, in particular, could use this fiscal administrative source for various purposes including the preparation of short-term turnover statistics.

The experimentation conducted at Istat within the Directorate of business statistics aims at assessing use of administrative e-invoicing data for their progressive use in production.

Short-term business statistics surveys on industrial turnover *Fat* (Istat, *Fatturato dell'industria*, Statistiche f!ash) and turnover in services *Fas* (Istat, *Fatturato dei servizi*, Statistiche f!ash) are involved in the experimentation. These direct surveys produce, on a regular monthly basis, turnover indices (unadjusted, seasonal adjusted and calendar adjusted indices) within 60 days from the reference period at a Nace rev.2 economic activity sector detail (sections B, C) for Fat survey and at a Nace rev.2 division/group economic activity sector detail (sections G, H, I, J, M, N) for Fas survey. In total, about 28,000 enterprises are involved in the surveys and then requested to provide information via an e-questionnaire about their total monthly turnover.

Electronic invoice data from businesses are acquired via the Interchange System (SdI) according to a well-defined record layout including a large amount of information with respect to transferor and transferee. E-invoice data contain also information about the date of issue of the invoice, the country of the assignee, the type of counterparty (B2B, B2C, etc.), the type of document (invoice, simplified invoice, advance on invoice, credit/debit note, etc.).

Since the entry into force of e-invoicing (January 2019), there have been and are still ongoing evolutions of the regulations. In addition, the fiscal regimes and consequently the number of companies concerned by the e-invoicing regulations is growing. All information on the evolution of the regulations is available online.

3. Early experiments and main results

The main objective of the experiment was to compare the turnover levels of enterprises obtained via the RA source (e-invoices) with those obtained by the statistical data source (Fat and Fas surveys) and to assess their coherence and quality. Thus, turnover indices can be constructed according to the currently adopted methodologies[2].

An assessment was made of the correspondence, in both sources in question, between the survey/analysis units and between the variable recorded (turnover of the enterprise) in the time unit. Both the unit of survey/analysis (legal-economic unit) and the measured variable (turnover) definitions are compatible as far as definitions are concerned. In particular, turnover comprises the total amounts invoiced by the observation unit during the reference period and corresponds to the value of market sales of goods or services provided to third parties.

For experimental purposes, a subset of companies was selected in order to meet both Istat's requirements of representativeness in terms of sectors of economic activity and size (in term of number of persons employed), and the constraints indicated by the RA in order to be able to

[2] The interesting results obtained from an initial work with this objective (Amato et al., 2021) demonstrated the coherence between the two sources at the aggregate Nace rev.2 division level.

manage data in an efficient way. In fact, each economic unit can correspond to a very large number of invoices.

It was therefore necessary to build a database that was as consistent as possible with the information of interest in terms of economic units involved, variables, reference period, and completeness of data during the year, etc.

For this purpose, some preliminary steps were carried out: acquiring the data and metadata files from the Agency; reading of e-invoices data files and standardisation of formats of variables; study of the invoice items included in the transmission path. Then the turnover variable was obtained for each company, according to the indications provided in the technical notes of the RA, and it was linked to those of the Istat short-term turnover surveys.

It was essential to link also the records so obtained to the business register (Istat, *Registro statistico delle imprese attive*) in order to retrieve structural information for each economic unit (Nace rev.2 economic activity classification code, number of persons employed, etc.). The key for the links was the fiscal code of the economic unit.

In order to make an initial assessment of the suitability of the tax source as an auxiliary for business statistics, some comparisons at micro level were made.

As a first analysis, turnover levels of e-invoicing administrative data were compared with the turnover data coming from the surveys.

A wide time span was chosen (from 2019 to 2022) so that the correspondence between the turnover values recorded by the survey and those acquired through the administrative source can be better assessed over time.

Table 1 shows the percentage of enterprises in the e-invoicing out of the total enterprises in the samples of the industry and services turnover surveys by Nace rev.2 Sections. The overall coverage of 84.3% seems to guarantee excellent results in terms of coverage of the units included in the samples. In addition, it should be considered that regulatory developments have provided for and continue to provide for a progressive increase in the number of companies and operations subject to the e-invoicing obligation, so that the coverage of administrative data is already even more extensive today.

Table 1

Per cent of enterprises with e-invoicing data in the Fat and Fas survey samples by sector of economic activity (year 2021)

Nace rev.2 section	Enterprises (%)
Total	84.3
B - MINING AND QUARRYING	98.3
C - MANUFACTURING	92.3
G* - WHOLESALE AND RETAIL TRADE; REPAIR OF MOTOR VEHICLES AND M	87.1
H - TRANSPORTATION AND STORAGE	91.1
I - ACCOMMODATION AND FOOD SERVICE ACTIVITIES	73.1
J - INFORMATION AND COMMUNICATION	92.6
L - REAL ESTATE ACTIVITIES	37.3
M - PROFESSIONAL, SCIENTIFIC AND TECHNICAL ACTIVITIES	72.8
N - ADMINISTRATIVE AND SUPPORT SERVICE ACTIVITIES	88.0

(*) excluding G47 - Retail trade, except of motor vehicles and motorcycles

The units considered for the experiment were chosen among those included in the Istat short-term statistics surveys on industry and service turnover, in total about 17.7K units (14.5K units for services sector and 3.2K units for industry) without missing data during the year of reference.

Initial experiments are aimed at understanding whether and to what extent tax data can be used as an auxiliary source and/or as in substitution of survey data. There may be cases in which, however, the tax data must be evaluated in the most appropriate manner (large enterprises, units belonging to specific estimation domains with a low number of enterprises, etc.) so that it may be

preferable to use the survey data.

Preliminary analyses show, as expected, a very good correlation between turnover from e-invoice and turnover from surveys. These analyses were conducted both with respect to different estimation domains and with reference to different periods.

For greater clarity of presentation, it was chosen to show, here, only the graphs relating to some of the numerous subpopulations of units belonging to the domains of interest. In particular, in Figure 1 and Figure 2 the micro comparisons are made between the turnover of service and industry sectors enterprises for 2020, as recorded by the Fas and Fat surveys and as acquired from the tax source. A high correlation seems to be evident; the graphs also shows some cases of enterprises for which the surveyed data and the administrative data are not well aligned. These are precisely the cases to be investigated in order to identify the reasons, which may stem from various causes, e.g. accounting, etc.

Figure 1:
E-invoicing vs Fas, total (year 2020, EUR)

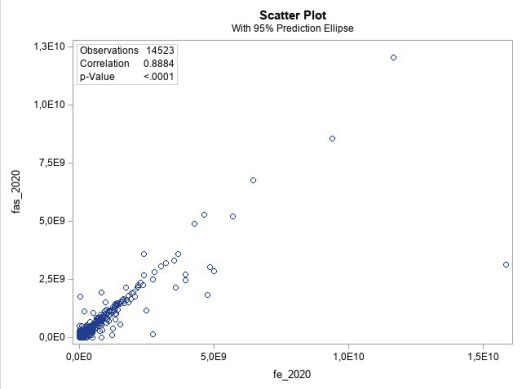
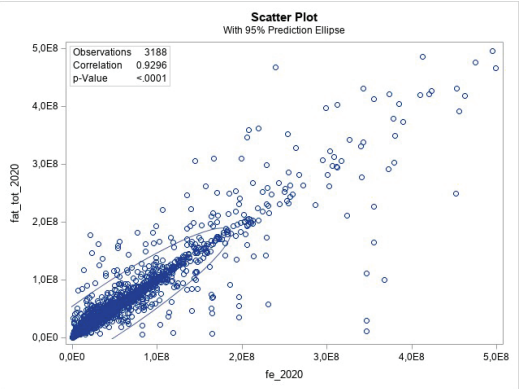


Figure 2:
E-invoicing vs Fat, enterprises < 500mln EUR turnover (year 2020, EUR)



Comparisons and analyses were also made at macro level, with reference to the following domains of interest: sector of economic activity, (Nace rev.2 divisions and groups) per quarter and total, as well as size class in terms of persons employed (1-9; 10-19; 20-49; 50-249; 250+), or turnover classes (< EUR 3 mln by size class; < EUR 1 mln; 1-5 mln; 5-10 mln; ≥ EUR 10 mln). Comparisons at the micro level, also per quarter were carried out[3].

Below are the graphs of some representative cases showing the good degree of alignment between the data collected by surveys and the administrative data.

Figure 3:
E-invoicing vs Fas by Nace rev.2 group (year 2020, EUR)

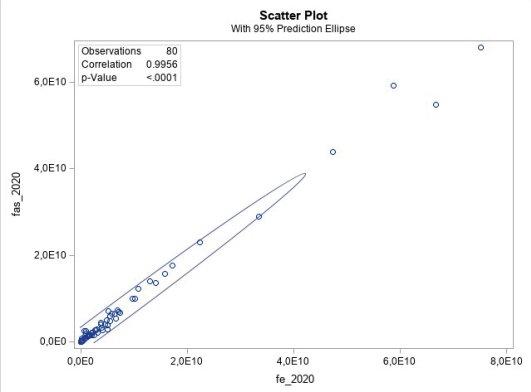
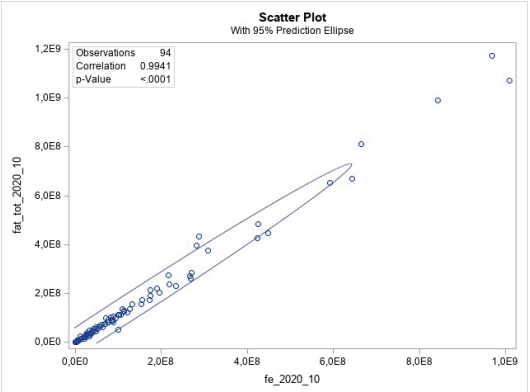


Figure 4:
E-invoicing vs Fat by Nace rev.2 group (Oct. 2020, EUR)



[3] The results and graphs relating to the economic activity sectors not included in this work are possibly available upon request.

Figure 3 and Figure 4 show the high correlation between statistical and administrative data respectively at Nace rev.2 group-level for the service sector for 2020 and for the industry sector for October 2020.

Analyses of changes over time between surveyed turnover levels and e-invoice data were made for quarters 2021_1 to 2022_3 for each Nace rev.2 division.

Below are the graphs with the trend changes for G45 (Wholesale and retail trade and repair of motor vehicles and motorcycles) and G46 (Wholesale trade, except of motor vehicles and motorcycles) Nace rev.2 divisions, which together represent 56.0% of the entire services sector[4], therefore with a greater relative weight, in terms of turnover, compared to the other Nace rev. 2 sections.

With reference to each quarter, shown in the abscissa, the changes were calculated at the stratum level (constructed by crossing economic activity and class of employees), considering the set of microdata available for e-invoicing and Fas survey respectively (uncommon observations are also included). Observations with a trend change outside the 2nd-98th percentile range of the distribution of changes in the reference quarter are not included. The stratum indices were subsequently aggregated using the weight structure of the services turnover survey.

Figure 5:
G45 Wholesale and retail trade and repair of motor vehicles and motorcycles (trend change %)

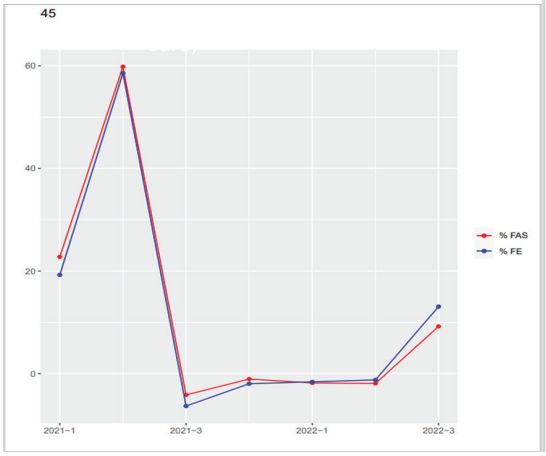
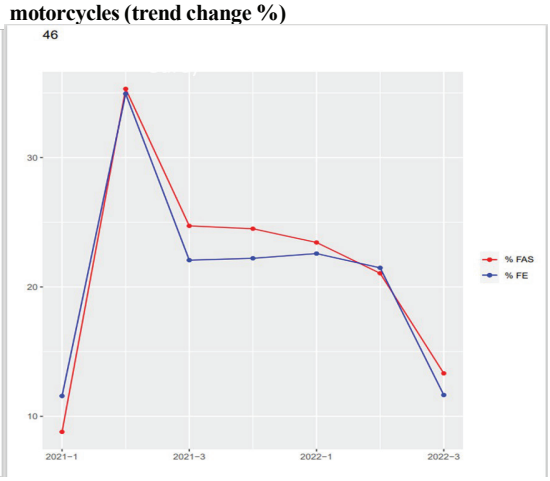


Figure 6:
G46 Wholesale trade, except of motor vehicles and motorcycles (trend change %)



Also this analysis, which is of an exploratory nature and should not be considered exhaustive, confirms on the whole the very good approximation of the trend changes based on administrative data compared to those obtained through survey data.

4. Some remarks

First results from the comparison of company turnover data with e-invoicing data seem to be extremely positive and provide with some relevant indications for the use, when operational, of the tax source for official statistics.

The evaluation of data quality, made possible through the analysis of the correlation coefficients between the levels of turnover from administrative data and survey data, allows to hypothesize that the use of tax data on company turnover would represent an unprecedented opportunity that should be exploited in the best possible way. It would also represent an investment of strategic importance for Istat, which offers prospects of absolute interest in the field of production of official economic statistics. The scenarios are in fact considerably interesting

[4] Excluding Nace rev.2 division 47 - Retail trade, except of motor vehicles and motorcycles, that is covered by Retail Trade survey.

both with regard to the improvement in the quality of the statistics produced (timeliness, coverage, etc.) and with regard to the considerable reduction or even elimination of the costs incurred by companies to participate in direct surveys.

The quality of short-term statistics on the turnover of companies is supposed to improve with the massive use of fiscal data. Indeed, a reduction of both sampling error and bias and thus a reduction of the Mean Squared Error (MSE) is expected. Accuracy, which is relative to the coverage of enterprises belonging to the economic sectors of interest, appears to be very good looking at the results of initial analyses and in perspective, as data availability is extended to all units in the field of observation. Timeliness and punctuality of acquisition are also very good for the purposes of short-term statistics. When operational, e-invoice data will be available 20-25 days after the end of the reference period, therefore with a very good timing considering the production process deadlines of short-term statistics and so improving their overall efficiency.

Obviously, the positive perspectives outlined above will then have to be verified once the project is actually implemented. In fact, some potential critical issues related to the acquisition, management, maintenance and utilisation of administrative data flows and files may occur, as they must be compatible and integrated into Istat production processes.

References

Amato, G., Ciammola, A., Filiberti, S., Margani, P., Orsini, D., (2021). Indagini congiunturali sul fatturato e fatturazione elettronica a confronto: prime evidenze e prospettive. *Quattordicesima conferenza nazionale di statistica*. https://www.istat.it/storage/14-Conferenza-nazionale-statistica/poster/008.jpg Accessed 30 November 2021.

Istat, *Fatturato dell'industria*, Statistiche f!ash, https://www.istat.it/it/archivio/fatturato+industria. Accessed 15 November 2023.

Istat, *Fatturato dei servizi*, Statistiche f!ash, https://www.istat.it/it/archivio/fatturato+servizi. Accessed 15 November 2023.

Istat, *Registro statistico delle imprese attive*, https://www.istat.it/it/archivio/263692. Accessed 23 November 2021.

Regulation (EU) 2019/2152 of the European Parliament and of the Council - https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R2152&qid=1700157163629. Accessed 01 January 2022.

doi.org/10.26398/asaproc.0040

The estimate of consumption of fixed capital and stock of vehicles in Italian national accounts

Paolo Forestieri, Paola Santoro
ISTAT - Italian National Institute of Statistics, National Accounts Department.

1. Introduction[1]

In this paper we describe the recent work carried out by ISTAT with the aim of implementing improvements in the Italian official estimates of consumption of fixed capital (CFC) and capital stocks (net and gross) of vehicles.

In Italian national accounts, such estimates are obtained through the Perpetual Inventory Method (PIM). The PIM generates an estimate of gross capital stock by cumulating past flows of gross fixed capital formation (GFCF) of assets over their estimated service life, i.e. the length of time that assets are retained in the capital stock before to be retired. Retirement refers to the removal of an asset from the capital stock because the asset is exported, sold for scrap, dismantled or abandoned.

As not all assets in a cohort (i.e. all the assets of a particular type that are acquired in specific year) are removed at the same time, retirement functions are required in the estimation method. The most common ones are bell-shaped: The probability of retirement is low in the early years of an asset's life, gradually increases to a peak at the average service life (ASL) of the cohort and gradually falls in the years after the average service life.

Moreover, aggregate stock figures (namely, net capital stock) should reflect the fact that similar assets of different ages have different values because of depreciation (or CFC), that is the loss in value of an asset or a class of assets as they age. The depreciation function shows how the asset value declines due to physical deterioration (wear and tear) and normal obsolescence. With a 'straight line' depreciation, an asset with a service life of T years loses a constant proportion ($1/T$) of the initial asset value each period, until the asset's value becomes zero at the end of year T.

The PIM is the most common method used worldwide. In practice, national statistical institutes (NSIs) may apply it in very different ways, depending on their choices of functions and key parameters. The best option for determining them is to obtain empirical evidence at national level but, generally, this information is sparse and often dated. Consequently, national accountants are often forced to make assumptions.

With the aim of supporting NSIs in the compilation of stocks of capital assets and CFC, Eurostat established the *DMES Task Force on fixed assets and estimation of consumption of fixed capital under ESA 2010 (TF FIXCAP)*. A set of recommendations was developed, focusing on the main parameters and functions of the PIM method for countries to use in their estimates in order to improve harmonization of practices and across countries comparability. The TF emphasized the importance of obtaining empirical evidence for setting assumptions underlying the PIM.

ISTAT participated in this international debate as a member of the TF FIXCAP, supporting its work also through a project granted by Eurostat titled "*Improvement in the Quality of Consumption of Fixed Capital and Capital Stock Estimates in Italy*". As different assumptions underlying the PIM could have significant impact on the estimates of stocks and CFC and on international comparisons, ISTAT investigated the available data sources and methods useful to choice functions and parameters. We developed case studies and sensitivity analyses, sharing the main results with the other members of the TF FIXCAP.

1 The views expressed in this paper are those of the authors and do not necessarily reflect the views of the ISTAT. The sections were authored as follows: 1, 3 and 5 were written by P. Santoro and 2 and 4 by P. Forestieri.

In this context, ISTAT explored a new data source (the administrative dataset of scrapping) and a new method to define service lives for motor vehicles empirically. The strengths and weaknesses of the new data source were checked. We tested the feasibility of updating the functions and parameters underlying the PIM through this new approach, as described in Section 2.

The results obtained were compared with the current assumptions used so far by ISTAT for its official statistics, in order to verify whether they are supported by empirical evidence or whether they need to be updated/revised. To this end, the impact on the official Italian estimates of capital stock of vehicles was calculated through sensitivity analyses, plugging alternative scenarios in the PIM, as presented in Section 3. Finally, conclusions and future plans are presented.

2. Source, method and main outcomes

The source used to estimate motor vehicles services lives in this exercise is the new administrative dataset of scrapping. UNRAE (Italian National Union of Foreign Vehicles Representatives), which is a member of the Italian Statistics System (SISTAN), provides the administrative data used in the exercise. The dataset reports the total number of vehicles scrapped every year cross-classified by type of vehicle, technical features of vehicle and year of first registration. The data refer to the following years of scrapping: 2013-2020. We have used these data to analyse the distribution of scrapping and to estimate the average age at the time of scrapping. The dataset distinguishes between vehicles scrapped by companies and vehicles scrapped by households, allowing us to analyse exclusively vehicles belonging to capital stock before scrapping.

The available data on scrapping are a very important source of information to estimate average service lives of vehicles, but they are not free of limitations. The first one regards major maintenance expenditures. The service life that is used in the PIM model should be the service life in absence of major maintenance expenditures (GFCF data used in the PIM model include also major maintenance expenditures, separately from vehicles). Major maintenance, such as the substitution of engines, extends the use of vehicles. The actual length of time before a vehicle is scrapped (and the average age at the time of scrapping) depends on the maintenance policy followed by its owner. In this context the extended warranty formulas, usually linked to maintenance programs, can effect on time of scrapping (major maintenance) and on the efficiency of vehicles (ordinary maintenance). For this reason, it is likely that the average age at the time of scrapping is an upward biased estimate of the actual service life (especially for what concerns trucks and buses). Another important issue regards the correspondence between the scrapping date and the actual exit from the productive process: it could happen that vehicles not working are left months or years stopped before official scrapping takes place. In this case we have an upward biased estimate of services life as above.

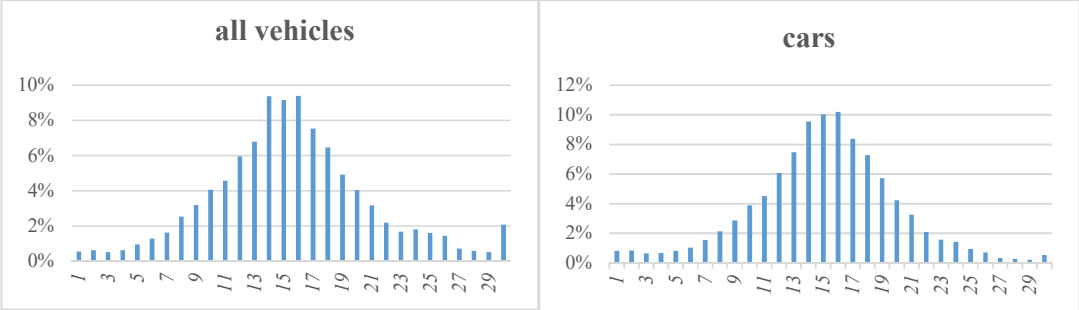
The proper data that should be used to calculate service lives of vehicles are the total number of new vehicles registered in a given year cross-classified by type of vehicle and year of scrapping. Unfortunately, such data were not available. If registration and scrapping of vehicles follow a balanced process (i.e. the distribution of scrapping is stable over time), the calculation of average service life from data on vehicles scrapped classified by year of registration and the calculation from data on vehicles registered in a given year classified by year of scrapping give the same result, otherwise the former calculation doesn't give a correct estimate.

We have calculated the distribution of vehicles scrapped according to their age at the time of scrapping (e.g. according to the year of first registration). We have made this calculation for each of the four types of vehicles. Figure 1 shows the distribution for cars and for all vehicles respectively, year 2016. The main characteristics of the distribution is that there is a non-negligible number of vehicles that are either early scrapped (about 3% are scrapped when they are 4 years old or less) or that are scrapped when they are very old (about 20% are scrapped when they are 20 years old or more). Late scrapping regards mainly trucks and buses, while this event is limited for cars.

We have used the distribution of scrapping as weights to estimate the average age at the time of scrapping for each of the four type of vehicles. Average ages of all vehicles have been calculated as weighted averages using the relative shares in terms of GFCF as weights (actually we have used

the average share for the whole period).

Figure 1 - Distribution of cars and all vehicles scrapped according to their age at the time of scrapping, 2016



Furthermore, to calculate the average life of the vehicles, in our exercise we truncated the age to the maximum level of 30 years, assuming that the vehicles are withdrawn from production at the maximum age of 30 even if scrapped later.

For every type of vehicle, the average age at the time of scrapping is almost constant in every year of scrapping (equal to 16-17 years for the whole aggregate). We consider this as a sign that the distribution of scrapping is quite stable and then our results are a good approximation of the results we should have obtained using data on vehicles classified by year of registration. Cars and trucks <=3.5 ton are retired when they are 14-16 years old, while trucks >3.5 ton 19 years after their first registration, and bus around 20 years. Trucks that weight less than 3.5 tons have an average age of scrapping only slightly higher than cars, while trucks that weight more than 3.5 tons and buses are retired after more years of use (see Table 1).

Table 1 - Average age at the time of scrapping

	2013	2014	2015	2016	2017	2018	2019	2020
Cars	14	14	15	15	16	16	16	16
trucks<=3.5t	14	14	15	15	15	16	16	16
trucks>3.5t	19	19	19	19	19	19	19	18
buses	20	18	20	20	21	21	21	21
all vehicles	**16**	**16**	**16**	**16**	**17**	**17**	**17**	**17**

It is interesting to underline that the average service life, according to the age of scrapping, tends to increase slightly in the period considered. This trend is confirmed by data of ANFIA (Italian National Association of Car Producers) about the average age of the UE car fleet (see Table 2): the average age of fleet increases of 2.6 years from 2007 to 2016[2]. It reflects a delay in terms of vehicles retiring probably due to the effect of 2007-2008 economic crisis: uncertainty holds back GFCF.

The results obtained confirm a higher age of vehicles at the time of scrapping then previous researches; the lack of replacement after 2007 crisis can justify it only in part. In fact, it can also depend on innovations in gasoline-diesel vehicles that make them efficient for a longer period. On the other hand, the effects of new technologies, such as electric or plug-in hybrid engines, on the durability of vehicles will probably be clear in the future, when we will have a substantial withdrawal from the production processes of these types of vehicles.

Data on scrapping, used here for the first time, not only test the services lives used now in the official statistics, but also provide a source for checking them periodically. This is a very important factor because it allows us to capture changes due to both technological progress and the introduction of new regulations in the vehicles market.

[2] https://www.anfia.it/it/automobile-in-cifre/statistiche-internazionali/parco-circolante

Table 2 – Average age of the EU car fleet

2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
8,4	8,5	8,6	8,9	9,1	9,3	10,4	10,6	10,8	11

Finally, we highlight that the time of scrapping is a proxy of service life for vehicles, and the biases of this variable we describe above tend to overestimate services life. Therefore, the level higher than expected could depend in part on the change in the economic frame, in part on the characteristics of the variable used.

3. Sensitivity analyses: testing the empirical evidences on the stock estimates

A straight line depreciation function and a bell-shaped retirement function are currently used for calculating the official estimates of capital stock and CFC for almost all fixed assets published by ISTAT, vehicles included.

Except for the service lives of some types of assets (hardware, telecommunications equipment, furniture, other machinery and equipment), generally no empirical evidences drove our choices: other country practices, expert advice and international manuals (OECD, 2019, 2020) are the main sources. A peculiarity of the Italian practice is that the retirements are distributed around the average service life according to a truncated normal distribution: we exclude that an asset, once entered, is never retired, with truncation limits set at +/- 35% of the average service life and the variance set so that 90% of retirements occur between +/- 25% of the average service life. ISTAT defined these parameters, used in official statistics, on the bases of several tests and preliminary works carried out using alternative assumptions (Lupi and Mantegazza, 1994).

In Italian accounts, the service life for motor vehicles is assumed equal to 10 years and is constant over time. The average age resulting from the administrative data of scrapping is 16-17 years, as described in Section 2. Even if it is likely that the average age at the time of scrapping is an upward biased estimate of the actual ASL, the results are consistent with the assumption of an ASL higher than 10 years for Italy (for instance, 12 or 14 years).

Furthermore, the distributions of vehicles scrapped according to their ages at the time of scrapping are not inconsistent with a bell shaped mortality function. On the other hand, the main characteristics of the distribution is that there is a non-negligible number of vehicles that are either early scrapped or that are scrapped when they are very old, mainly trucks and buses. Instead, in our current hypothesis, retirements are much more concentrated around the ASL.

We have tested the sensitivity of the capital stock estimates to different assumptions regarding the ASL and the mortality pattern, under different hypotheses, for the period 1995-2020 (see Chart 1). This analysis is conducted to check how far the estimates obtained with the different assumptions deviate from each other and from the official data published so far by ISTAT. By comparing the results obtained, it is possible to define the assumptions that need to be updated to produce accurate estimates and those that have no impact (or at least no significant impact) on the estimates published by ISTAT.

Chart 1 - Alternative hypotheses underlying the PIM method for vehicles estimation

Hypothesis	ASL	Retirement function
Hp1	10	The distribution variance is such that 90% of the group of assets installed in a given year are retired in the interval [-25%, +25%] of the average service life. Truncation limits are –35% and +35% of the average service life.
Hp2	12	
Hp3	14	
Hp4	10	The distribution variance is such that 65% of the group of assets installed in a given year are retired in the interval [-25%, +25%] of the average service life. Truncation limits are –80% and +80% of the average service life.
Hp5	12	
Hp6	14	

The first scenario (Hp1) represents the current practice used in Italy for calculating the capital

stock of vehicles (estimates officially published by ISTAT). The hypotheses Hp2 and Hp3 use the same mortality function of Hp1 but different service lives; therefore, it is possible to measure the impact on estimates of the change in ASL only.

In hypotheses Hp4, Hp5 and Hp6, given the services lives indicated above (respectively 10, 12, 14 years), a new retirement function is used, that is more consistent with the empirical results on the distributions of vehicles according to their age at the time of scrapping.

As shown in the following table (table 3, columns a and b), increasing average service life from 10 to 12 years (an increase equal to 20%) and from 10 to 14 years (an increase equal to 40%), given the same mortality function, causes an increase in the level of net capital stock at chain linked values respectively of about 17% and 34% (over the time series 1995-2020). In both cases, it reduces the variability of the series both with respect to negative and positive rates of growth (comparing Hp1, Hp2 and Hp3, Figure 2).

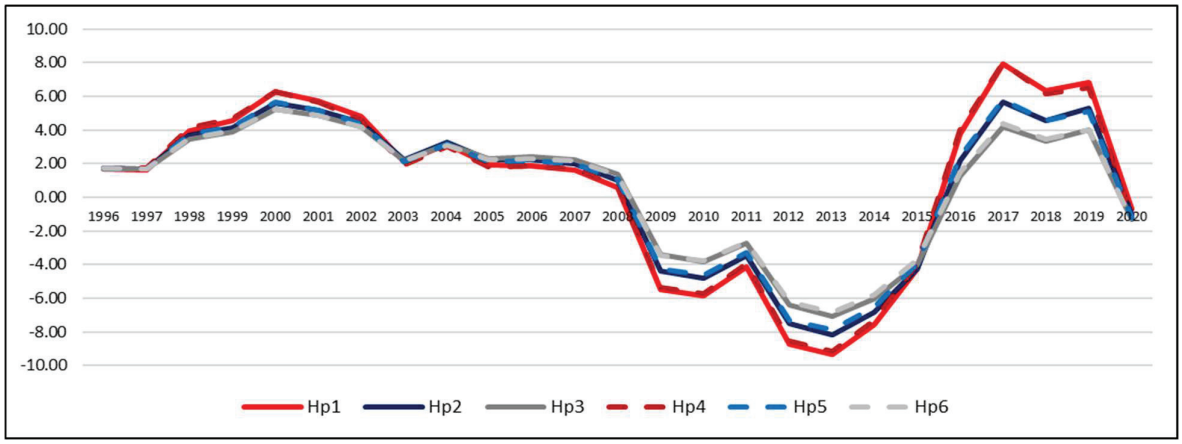
Changing the mortality function, given the same ASL, has not a significant impact (Hp4/Hp1, Hp5/Hp2, Hp6/Hp3), causing slight revisions on the estimates of the net capital stock of vehicles (less than 1%, columns from c to e of the table). Due to the low revisions on levels, growth rates differ very marginally (comparing Hp4 to Hp1, Hp5 to Hp2, Hp6 to Hp3 in figure 2).

The combined effect of both changes, in ASL and in mortality function (Hp5/Hp1 and Hp6/Hp1), is driven by the new ASL: revisions are the same ones obtained just modifying the ASL (columns f and g compared to the columns a and b, respectively). The ASL seems to be the most significant parameter affecting estimates.

Table 3 – Alternative estimates of net capital stock at chained values. Averages ratios over the time series 1995-2020

Period	a	b	c	d	E	f	g
	Hp2/Hp1	Hp3/Hp1	Hp4/Hp1	Hp5/Hp2	Hp6/Hp3	Hp5/Hp1	Hp6/Hp1
1995-2020	1.17	1.34	0.99	1.00	1.00	1.17	1.34

Figure 2 - Alternative estimates of net capital stock. Annual rates of growth. 1995-2020



4. Second-hand market and future development

The data provided by UNRAE allow us to describe the second-hand market, as UNRAE makes available administrative records of all transactions among companies, households and rest of the world. Every year companies sell to households a relevant part of their stock of cars: second-hand cars represent 22% of the total value of cars bought by households and sold by companies in year 2016. It is interesting to notice that the stock of cars sold by companies to households has an average life around 6 years in 2016-2020.

According to the PIM method, these cars should be retired from the stock when they are sold

to households, but we do not have enough information to consider the second-hand market in the path of retirement precisely. More in details, with a linear depreciation combined with a bell-shaped retirement function (assumptions underlying the Italian approach), it is not so straightforward to trace the second-hand transactions in the PIM model in detail. Nevertheless, the second-hand market is caught through the estimate of GFCF year by year: the value of cars sold to households represents a negative component of GFCF, affecting stock estimation indirectly.

In the future it would be really interesting to investigate systematically the relationship between GFCF, with the second-hand component, and PIM method for the stock of cars, especially if we refer to productive stock.

5. Conclusions and future plans

Enhancing the harmonisation of compilation practices and improving the quality of stocks of fixed assets and consumption of fixed capital data has gained importance and attention over recent years. In this context, ISTAT has conducted several analyses and activities in order to review and update its compilation practice for national estimates of capital stocks and CFC.

The exercise presented in this paper tests the feasibility of estimating service life and functions to be used in the perpetual inventory method for estimating vehicles using the administrative dataset of scrapping.

We can conclude that the new data source and the new procedure allow us to adopt choices in the PIM for estimating stock and depreciation for this asset that are empirically based and, therefore, in line with the real Italian production system. The obtained results are quite accurate, despite the weaknesses of the source already described in the previous sections.

Bases on the outcome of this work, ISTAT is planning to use a higher service life for vehicles than the current one, as empirically indicated by the data on scrapping; the sensitivity analysis shows that the estimates are significantly sensitive to the revision of this parameter. Less essential is the need to revise the parameters of the retirement function, since the impact on the estimates is not material in light of the tests performed. We will probably keep the same current assumptions, common to all the other fixed assets.

Instead, an attempt will be made to abandon the restrictive assumption of a constant service life and introduce a variable one in the time series. In fact, as demonstrated empirically, it is not constant over time. There are events and factors influencing the length of time that vehicles are retained in the capital stock, such as technological progress, the intensity of their use over the business cycle, and the product cycle (the latter dependent, for example, on user behavior and expectations and regulations). The new administrative data source of scrapping allows to collect empirical evidence consistent over time to measure the dynamic of service life for vehicles.

Revisions will be introduced in our estimates in 2024. Indeed, as also expressed by the TF FIXCAP, NSIs are expected to implement improvements, if necessary, to their data on capital stock and CFC during the next EU harmonized benchmark revisions of national accounts, scheduled for 2024.

References

Eurostat (2010). *European System of Accounts ESA 2010*, Luxemburg.

Eurostat (2014). *Manual on Measuring Research and Development on ESA 2010*, Luxemburg.

Eurostat, OECD (2015). *Survey of National Practices in Estimating Net Stock of Structures*, Luxemburg.

Lupi C., Mantegazza S. (1994). Ricostruzione delle serie degli investimenti per branca utilizzatrice per branca proprietaria e calcolo dello stock di capitale. *Quaderni di ricerca* n.14. Roma, ISTAT.

OECD (2019). *Measuring Capital.* Second edition, Paris.

OECD (2010). *Handbook on Deriving Capital Measures of Intellectual Property Products*, Paris.

UN Statistical Commission (2008). *System of National Accounts*, New York.

doi.org/10.26398/asaproc.0041

Measuring tourism sustainability for Italian regions: A proposal

Francesca Giambona, Alessandro Magrini
[a] Department of Statistics, Computer Science and Applications, University of Florence, Florence, Italy

1. Introduction

Tourism is certainly one of the main drivers for the economic development of territories, as clearly shown by economic and employment indicators in territorial areas with strong tourist vocation. However, the development of tourism needs to be strengthened and improved to ensure long-term competitiveness for the tourism and travel sector. In recent years, tourism sustainability has significantly gained attention also due to Sustainable Development Goals (SDGs), among which target 8.9 aims to "*by 2030, devise and implement policies to promote sustainable tourism that creates jobs and promotes local culture and products*".

Sustainability, both in general and in tourism, is a very complex topic. In the current literature there are no shared agreements regarding its definition and measurement. Among others, the World Tourism Organization (UNWTO) defines sustainable tourism as "*tourism that takes full account of its current and future economic, social and environmental impacts, addressing the needs of visitors, the industry, the environment and host communities*", with a clear reference to the multidimensionality of the concept, which requires the simultaneous consideration of economic, social and environmental indicators for its evaluation (UNWTO, 2022).

Several international initiatives have placed emphasis on the creation of systems of indicators with the purpose of assessing and monitoring tourism sustainability over time. These indicators are often aggregated into composites to obtain a synthetic measure of sustainability. Composite indicators synthesize the evaluation of a complex and multidimensional phenomenon, facilitating its interpretation and the definition of benchmarking practices.

However, the construction of composite indicators is a critical task because several arbitrary choices must be made that may influence the final results, especially the aggregation method, which can be compensatory or non-compensatory (Munda and Nardo, 2009), and the weighting scheme, which can be subjective or endogenous (Nardo *et al.*, 2005; Greco *et al.*, 2019). Although several international guidelines on the construction of composite indicators have been advanced (Saisana and Tarantola, 2002; OECD, 2008), so far there is no international consensus on which methodology is the most appropriate. Besides these challenges, most existing studies have the limitation to consider a single or few temporal points, which prevents to assess the evolution of sustainability over time.

In this article, we propose a composite indicator to measure tourism sustainability for Italian regions in the period 2010-2021 (12 years). Our composite has a two-level structure accounting for the contributions of the economic, social and environmental dimensions of sustainability, and is based on a non-compensatory (geometric) aggregation of 14 basic indicators through weights endogenously determined according to the Benefit of Doubt (BoD) approach (Cherchye et al., 2007; Zhou et al., 2010). In this way, we explore not only the level of tourism sustainability and its dimensions, but also their evolution over time. Furthermore, geometric aggregation allows a small degree of compensation to reflect the fact that tourism sustainability is achieved only when all or most individual sustainable goals are pursued. Finally, the BoD weighting scheme permits to account for policy priorities without relying on subjective opinions, and has the advantage of scale invariance, which makes normalization of basic indicators unnecessary.

This article is structured as follows: data and methods are presented in Section 2, the results are

reported and discussed in Section 3, concluding remarks are provided in Section 4.

2. Data and methods

In the construction of our composite indicator, we followed the set of 58 indicators proposed by Blancas and Lozano-Oyola (2022). By collecting data from the Istat database, we obtained a complete dataset on 14 of them (4 for the economic, 3 for the social and 7 for the environmental dimension) with yearly measurement in the period 2010-2021 (12 years). For the economic dimension, we considered: "Average length of stay in accommodation facilities" (ECO1), "Gross domestic product at market prices per inhabitant" (ECO2), "Added value of accommodation and catering services" (ECO3), "Area of Sites of Community Importance (SIC) on regional area (%)" (ECO4). For the social dimension, we considered: "Beds in residential social-welfare and social-health facilities per 10,000 inhabitants" (SOC1), "Workers, schoolchildren and students aged 3 and over who habitually use the train to go to work, kindergarten or school (%)" (SOC2), "Overnight stays in accommodation facilities per inhabitant" (SOC3). For the environmental dimension, we considered: 'Robbery per 1000 inhabitants" (ENV1), "Presence of elements of degradation in the area where you live (%)" (ENV2), "Perception of crime risk (%)" (ENV3), "Air quality - PM2.5 (%)" (ENV4), "Municipal waste produced (kg per inhabitant)" (ENV5), "Electricity from renewable sources (%)" (ENV6), "Area of Special Protection Areas (SPAs) on the regional area (%)" (ENV7). In order to account for the negative polarity of some indicators, we take the complement to 100 (ENV2, ENV3, ENV4) or the reciprocal (SOC3, ENV1, ENV5).

Let $i = 1, \dots, n$ denote the regions, $j = 1, \dots, p$ the basic indicators, and $t = 1, \dots, T$ the time points. Also, let x_{ijt} and w_{ijt} be, respectively, the measurement and the weight of the basic indicator X_j for region i at time t. The score in sustainability for region i at time t is defined as the geometric aggregation (weighted product) of basic indicators:

$$\mathrm{SUS}_{it} = \prod_{j=1}^{p} x_{ijt}^{w_{ijt}}$$

Following Magrini and Giambona (2022), this score is decomposed into the product of the scores in the economic (ECO), social (SOC) and environmental (ENV) dimensions:

$$\mathrm{SUS}_{it} = \underbrace{\prod_{j:\, X_j \in \mathrm{ECO}} x_{ijt}^{w_{ijt}}}_{\mathrm{ECO}_{it}} \underbrace{\prod_{j:\, X_j \in \mathrm{SOC}} x_{ijt}^{w_{ijt}}}_{\mathrm{SOC}_{it}} \underbrace{\prod_{j:\, X_j \in \mathrm{ENV}} x_{ijt}^{w_{ijt}}}_{\mathrm{ENV}_{it}}$$

Weights are determined based on the Benefit of Doubt (BoD) approach (Cherchye *et al.*, 2007; Zhou *et al.*, 2010), i.e., by maximizing the score for each region. Precisely, for each pair (i, t), the weights $w_{i1t}, \dots, w_{ijt}, \dots, w_{ipt}$ are determined by solving the problem:

$$\max \prod_{j=1}^{p} x_{ijt}^{w_{ijt}} \quad \text{subjected to:}$$

$$\begin{cases} \prod_{j=1}^{p} x_{kjl}^{w_{ijt}} \leq e \quad k = 1, \dots, n; \quad l = 1, \dots, T \\ w_{ijt} \geq 0 \qquad j = 1, \dots, p \end{cases}$$

The composite SUS takes value between 1 and *e*, while its logarithm, which ranges between 0 and 1, is equal to the weighted sum of the log values of basic indicators and to the sum of the log scores in ECO, SOC and ENV:

$$\log \mathrm{SUS}_{it} = \sum_{j=1}^{p} w_{ijt} \log x_{ijt} = \log \mathrm{ECO}_{it} + \log \mathrm{SOC}_{it} + \log \mathrm{ENV}_{it}$$

Therefore, the relative contribution to the composite (importance) of each sustainable dimension for region i at time t is equal to the log score in such dimension, while the relative contribution of the j-th basic indicator to the composite for region i at time t is given by:

$$r_{ijt} = \frac{w_{ijt} \log x_{ijt}}{\sum_{k=1}^{p} w_{ikt} \log x_{ikt}}$$

In order to avoid excessively low or high weights, we added further constraints bounding the relative importance of each basic indicator between 3% and 12%, i.e., $0.03 \leq r_{ijt} \leq 0.12$ $\forall\, i, j, t$.

In order to assess the change over time of the composite SUS and its dimensions ECO, SOC and ENV, we adopted the mobility index proposed by Giambona and Vassallo (2014), which is the mean of first order changes in rank weighed by first order changes in score. Therefore, the mobility index accounts not only for the absolute change of a region, but also for its relative change with respect to the other regions.

3. **Results**

Figure 1 shows the trajectories of the composite SUS (in red) and its dimensions ECO (in blue), SOC (in orange) and ENV (in green) by region in the period 2010-2021.

It can be noted that the trend of sustainability is stable or has a moderate growth rate for most regions, and that no region shows a definitely decreasing trend. The lowest SUS values are recorded for Emilia-Romagna and Tuscany in 2011 and 2012, even if these two regions have increased considerably their score after 2012, but without reaching the average level of sustainability of the other regions. Instead, Veneto and Trentino-South Tyrol maintain the highest SUS values from 2011 to recent years (2018 and 2020, respectively), where these two regions have come through a definite decrease. Other regions with a high level of sustainability are Apulia, Piedmont, Sardinia and Sicily, which also have a stable pattern across the considered period with no relevant declines. Instead, some regions reach a high level of sustainability only in recent years, like Campania, Lazio and Lombardy. For what concerns the importance of sustainable dimensions, the first rank is always occupied by either the economic or the environmental one, with the social dimension ranked third or tied for second place in all regions.

Figure 2 displays the average score in sustainability and in its dimensions versus the mobility index. From Figure 2a, it can be noted that Lombardy and Abruzzo show both a high average score in sustainability and a high improvement of it over time. Instead, Emilia Romagna, Tuscany, Aosta Valley and Umbria have the lowest average score and also show little or no improvement over time, while Veneto, Trentino-South Tyrol, Sardinia and Basilicata have a high average score but also a high worsening of it over time. For the economic dimension (Figure 2b), the region with the best compromise between average score and growth is Tuscany, while Aosta Valley exhibits the lowest average score together with a negative growth. For the social dimension (Figure 2c), Tuscany still shows the best compromise between average score and growth, while the worst compromise (low average score and high decrease over time) is owned by Sardinia and Abruzzo. For the environmental dimension (Figure 2d), Abruzzo and Aosta Valley show both a high average score and a high improvement of it over time, while Emilia-Romagna shows the lowest average score with an almost null growth, and Basilicata exhibits the highest average score but with a decreasing trend.

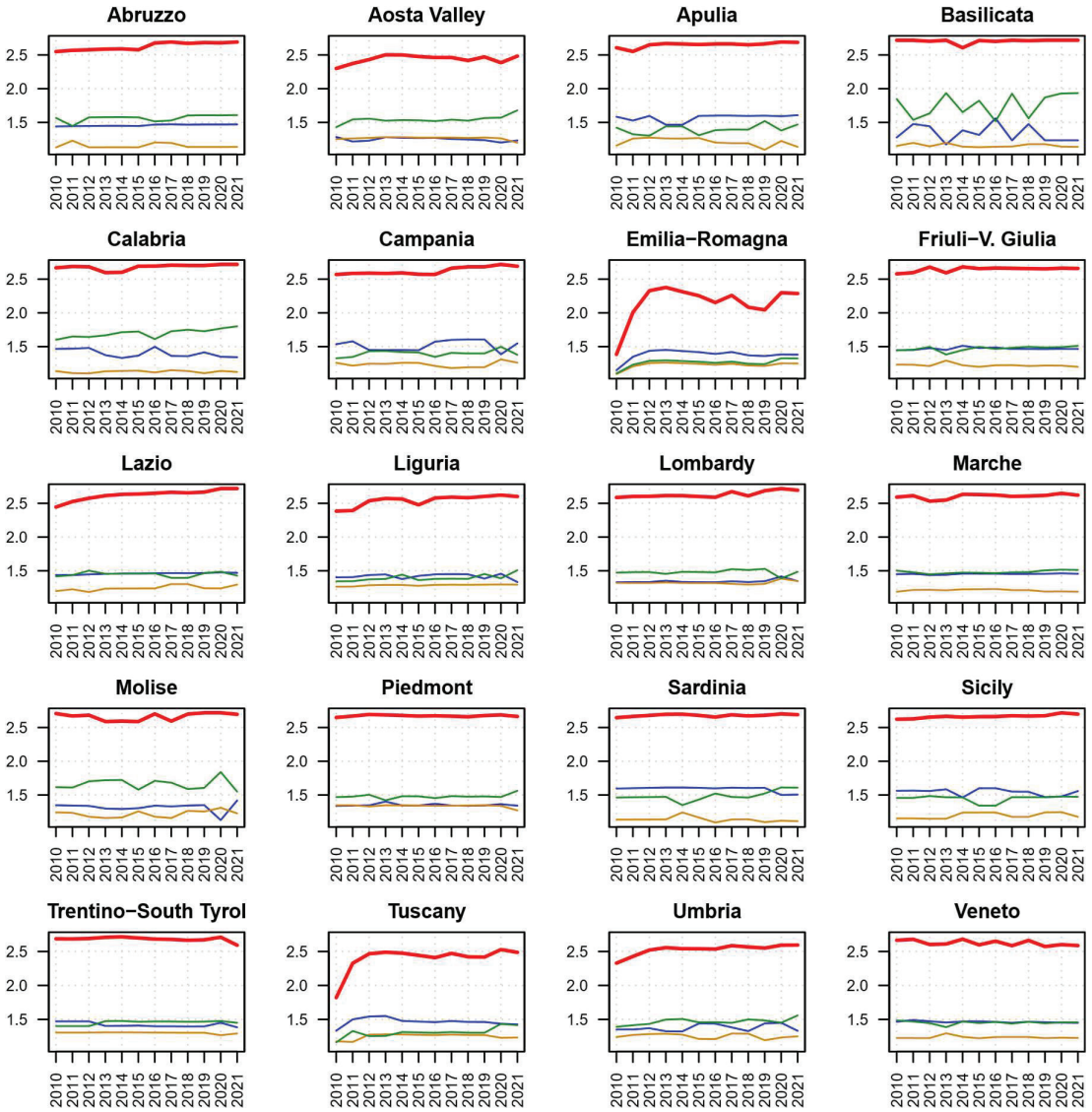


Figure 1: Trajectories by region. SUS: red, ECO: blue, SOC: orange, ENV: green.

Based on the estimated relative importance, basic indicators contributing the most to the composite are ECO3, ECO2, ENV2, SOC1 and ECO1, with an average importance between 8.9 and 11.3%, followed by SOC2, ENV7, ENV3 and ECO4, with an average importance between 8.4 and 9.0%. The other basic indicators have an average importance between 3.1% and 4.7%. For what concerns sustainable dimensions, the environmental one provides the greatest average contribution to the composite (41.0%), followed by the economic (37.4%) and the social (21.6%) dimension.

Our choice to constrain the relative importance of basic indicators in the interval [0.03, 0.12] was motivated by the fact that lower bounds less than 0.03 lead to an excessive number of benchmark regions or to an unjustifiable relative importance for several basic indicators, while lower bounds greater than 0.03 or upper bounds lower than 0.12 do not give solution, and upper bounds greater than 0.12 result ineffective.

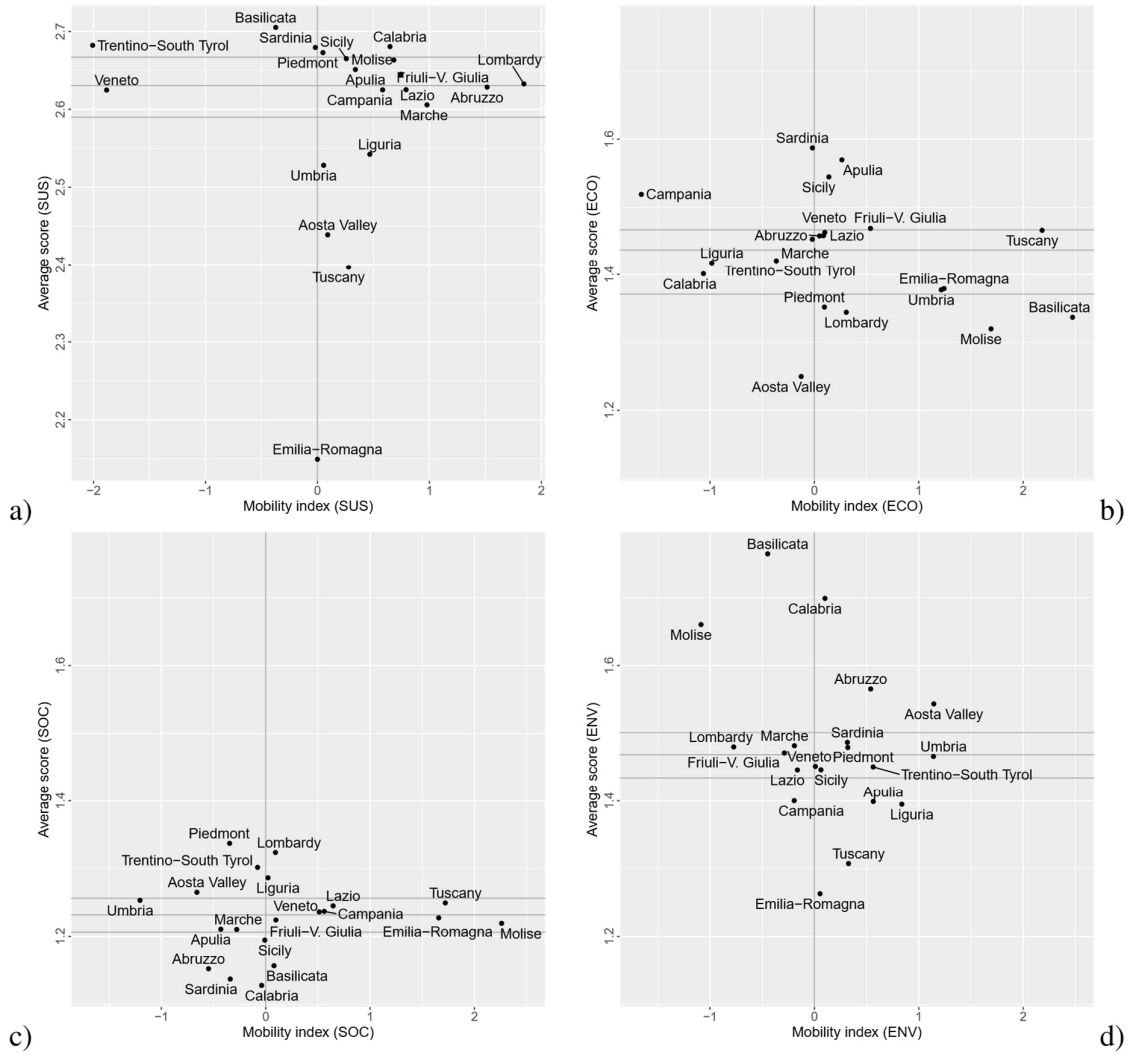


Figure 2: Average score versus mobility index by region (2010-2021). Horizontal lines indicate quartiles. a) Average SUS versus mobility; b) Average ECO versus mobility; c) Average ENV versus mobility; d) Average SOC versus mobility.

4. Conclusions

In this paper we have proposed a composite indicator to assess the sustainability of tourism in Italian regions. Our composite has the following strengths: (i) it considers a reasonably long temporal interval (12 years) allowing to assess not only the level of sustainability but also its trend, (ii) it has a two-level structure accounting for the contribution of economic, social and environmental dimensions of sustainability, and (iii) its weighting scheme permits to account for policy priorities without relying on subjective opinions.

A first limitation of our proposal relies in the limited number of indicators, which was mainly due to the difficulty of collecting complete data on a relevant number of indicators and time points. A remarkable gap in our set of indicators is the absence of information on employment (e.g., the regional productive specialization), which undoubtedly deserves attention in future developments.

A second limitation is represented by the low level of geographical detail, which could be deepened by considering provinces as units of observation, instead of regions. However, focusing on provinces could entail further issues of data availability, for instance we could end with fewer indicators measured on fewer years.

Future work could be also directed towards the improvement of the weighting scheme. Currently, traditional data envelopment analysis (DEA) is employed, but it has the inconvenient to generate excessively low or high weights, thus requiring a priori proportional constraints. In the future, we plan to use stochastic DEA techniques, which could provide more robust results due to their ability to account for stochastic best practice frontier.

References

Blancas, F., Lozano-Oyola, M. (2022). Sustainable tourism evaluation using a composite indicator with different compensatory levels. *Environmental Impact Assessment Review*, **93**, 106733.

Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T. (2007). An introduction to 'benefit of the doubt' composite indicators. *Social Indicators Research*, **82**, pp. 111-145.

Giambona, F., Vassallo, E. (2014). Composite indicator of social inclusion for European countries. *Social Indicators Research*, **116**, pp. 269-293.

Greco, S., Ishizaka, A., Tasiou, M., Torrisi, G. (2019). On the methodological framework of composite indices: a review of the issues of weighting, aggregation, and robustness. *Social Indicators Research*, **141**, pp. 61-94.

Magrini, A., Giambona, F. (2022). A composite indicator to assess sustainability of agriculture in European Union countries. *Social Indicators Research*, **163**, pp. 1003-1036.

Munda, G., Nardo, M. (2009). Noncompensatory/nonlinear composite indicators for ranking countries: a defensible setting. *Applied Economics*, **41**, pp. 1513-1523.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., Giovannini, E. (2005). *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD Statistics Working Papers.

OECD. (2008). *Handbook on Constructing Composite Indicators. Methodology and User Guide*. OECD, Paris, FR.

Saisana, M., Tarantola, S. (2002). *State-of-the-Art. Report on Current Methodologies and Practices for Composite Indicator Development*. Joint Research Center, European Commission, Brussels, BE.

UNWTO (2022). *Statistical Framework for Measuring the Sustainability of Tourism (Sf-Mst)*. World Tourism Organization, United Nations, Madrid, ES.

Zhou, P., Ang, B. W., Zhou, D.Q. (2010). Weighting and aggregation in composite indicator construction: A multiplicative optimization approach. *Social Indicators Research*, **96**, pp. 169-181.

doi.org/10.26398/asaproc.0042

Introducing a new algorithm for weighted triadic census: An application to passing network

Riccardo Ievoli[a], Lucio Palazzo[a], Roberto Rondinelli[b], Giancarlo Ragozini[b]
[a] Department of Chemical, Pharmaceutical and Agricultural Sciences, University of Ferrara, Ferrara, Italy.
[b] Department of Political Sciences, University of Naples Federico II, Naples, Italy.

1. Introduction

The analysis of tactics and strategies of teams represents an interesting challenge, from a statistical point of view, in the large field of sports analytics. Considering team sports and the particular case of football, relevant information can be retrieved using the distribution of passes between players, usually available for each team at the end of the match, and understandable as a weighted and directed network. Thus, network analysis (NA) should be applied to retrieve and/or identify useful information concerning the different degrees of cooperation of the football teams. A seminal contribution can be found in Grund (2012), which used descriptive statistics of the network to predict the football outcome. More recently, a set of network summary measures computed from the passing distribution has been used to model the probability of winning the game (Ievoli et al., 2021), while passing network indicators have been involved in Bayesian hierarchical models to model the football outcome in terms of scored goals or the difference in goals (Ievoli et al., 2023). From a slightly different perspective, 15 tactics have been identified starting from the passing distributions of Italian teams (competing in the "Serie A") by using a two-phase clustering technique (Diquigiovanni and Scarpa, 2019). A review of some contributions concerning passing networks and tactics analysis in football can be found in Caicedo-Parada et al. (2020).

Football passing networks exhibit some features: they involve a small and (generally) fixed number of nodes but may also show a certain level of sparseness. The number of interactions between players may be influenced by the position on the pitch, the individual skills and the team cooperation attitudes and, finally, team strategies and/or style of play. In NA, these features can be investigated through properties such as density, transitivity, and presence of clusters. These topological properties can usually be investigated through the triad census analysis. A "triad" involves the interaction between three nodes, e.g., the relationships between three football players. Considering weighted and directed networks, sixteen triadic configurations, i.e., "isomorphism classes", can be defined and their distribution is denoted as "triad census" (Wassermann and Faust, 1994; Faust, 2010).

Håland et al. (2020) (and references therein) highlight the importance of sequences of three passes (denoted as flow motifs) to evaluate the style of play of football teams. Starting from the probabilistic properties of the triad census (Holland and Leinhardt, 1978), recent research proposes to test the triadic structure of football teams to assess the deviation from the randomness of the observed triad census (Palazzo et al., 2023). Furthermore, the authors also redesigned the Dirichlet-multinomial test to verify the distance between a specific triadic distribution (of a team in a competition) and possible reference patterns. The work considers three consecutive seasons of the UEFA Champions League (UCL) to study the observed distribution of triads in football (Palazzo et al., 2023). However, one of the drawbacks of this approach is that the conventional triad census works only for unweighted networks. For this reason, the information regarding the degree of connection between players, i.e., the number of passes given or received, is totally omitted.

A possible solution is represented by the "network peeling" algorithm (Rondinelli and Palazzo, 2023) evaluating not only the presence of a particular isomorphism class but also the intensity of the relationships. The output of this algorithm consists of (for each passing network) a matrix with

dimensions layers × isomorphism classes. At this stage, two possible developments arise. Firstly, the information obtained at the end of the algorithm should be summarised through summary measures indicators that may also have meaningful interpretations in football analytics. Secondly, a comparison with the conventional triad census for binary networks is needed to understand the difference between weighted and binary triad census. Given these premises, the present paper aims to propose a set of descriptive measures that can summarize the new weighted triad census (wTC) and are also useful for comparing football teams according to their level of cooperation. Finally, conditional logistic regression models are performed to investigate and compare the explanatory power of the weighted and unweighted triad census.

The paper is organised as follows: Section 2. illustrates the NA framework and the "peeling" algorithm. Section 3. includes the measures of wTC while the empirical application, in Section 4., concerns 192 passing networks coming from the group stage of the 2018-2019 UCL. Remarks and possible advances are discussed in Section 5.

2. Background: triad Census and network Peeling

In this section, we briefly introduce the features of the football passing networks and the triad census for binary networks. Then, we illustrate the new "peeling" algorithm (Rondinelli and Palazzo, 2023).

The passes of a team played in a football match can be expressed as a weighted and directed network $\mathcal{G}$ which is a triple $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathcal{W})$, where $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$ is the set of the N nodes (or vertices), i.e, the football players (usually N = 11), $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of directed edges (or links). Each edge $e_{ij} \neq e_{ji}$ is an ordered couple indicating a relation from v_i to v_j; $\mathcal{W}: \mathcal{E} \rightarrow \mathbb{N}$, while $\mathcal{W}(e_{ij}) = w_{ij} \in \mathbb{N}$ is the number of times the relation from v_i to v_j is observed. The generic directed edge e_{ij} is present when (at least) a pass from player v_i to to player v_j occurs, while w_{ij} represents the weight, i.e., the count of their total passes. The passing network is usually available in the form of an adjacency matrix and can be represented through a graph, see, e.g., Table 1 and Figure 1 in Ievoli et al. (2021).

Considering this framework, a subset composed of three nodes is defined as a "triad", representing the minimal structural group in a network. In binary and directed networks there are 16 possible different relational structures, also called isomorphism classes. These isomorphism classes are depicted in Figure 1 within their conventional labels, from the so-called "003", presenting no edges between nodes, to the fully connected triad "300", involving six links. The distribution of triad census in football passing networks is summarized in Table 4 and Figure 4 in Palazzo et al. (2023). The mentioned work also proposes two formal tests based on the observed triad census for binary networks to test the presence of possible play styles. Two main drawbacks arise from this approach: the test against a random structure, based on the U|MAN assumption (Holland and Leinhardt, 1978) appears to have meagre rejection rates considering a set of observed passing networks, while the test based on the Dirichlet-multinomial distribution needs the specification of a particular null hypothesis (i.e., a referenced style of play). In addition, while there is a relevant amount of information about the type of links between subgroups, the intensity of the relationships is not considered by the binary triad census.

From a descriptive perspective, the first advance of this approach is the introduction of a "weighted" triad census through the novel "network peeling" algorithm (Rondinelli and Palazzo, 2023), consisting of a nested sequence of binary sub-networks in which edges are "peeled out" each time of a unit value. The binary triad census is then computed for each of these subnetworks. The final output, for each passing network of a football team in a match, is represented by a matrix with the following dimensions: rows are equal to the highest number of passes (given or received) between players and columns are equal to the number of isomorphism classes (i.e., 16 in our case). The procedure is summarized through expressions 1-3 and Algorithm 1 in Rondinelli and Palazzo (2023).

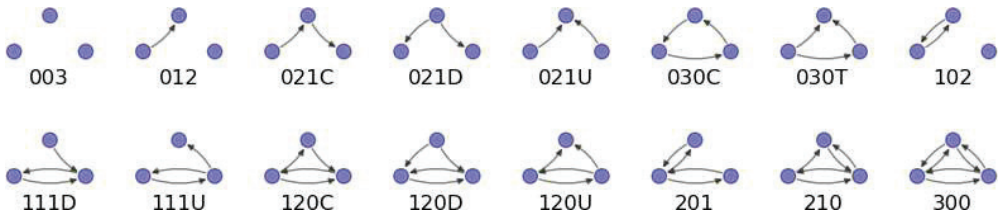


Figure 1: List of the 16 isomorphism classes of the triad census.

3. Measures of weighted triad census

The core of the peeling algorithm consists of "slicing" a weighted network in many binary layers allowing the count of the isomorphism classes to measure the weighted triad census. To summarise the information of each slice and compute a weighted version of the triad census, different statistics can be computed. We address such statistics with the term wTC. Let $TC_{(t)} = (TC_{1,(t)}, \dots, TC_{16,(t)})$ be the empirical distribution of the triad census at the t-th slice, for all the 16 isomorphism classes, as described in Rondinelli and Palazzo (2023). We point out that for each count of passes between players (t), a binary network of their relationship is defined. We define the proposed wTC statistics as follows:

$$\text{Sum:} \quad wTC_i^s \sum_t TC_{i,(t)} \tag{1}$$

$$\text{Mean:} \quad wTC_i^m = \frac{1}{T} \cdot WTC_i^s \tag{2}$$

$$\text{Weighted Mean:} \quad wTC_i^{wm} = \frac{\sum_t TC_{i,(t)}}{\sum_t n_{(t)}} \tag{3}$$

$$\text{Mean Proportion:} \quad wTC_i^p = \sum_t \frac{TC_{i,(t)}}{n_{(t)}} \tag{4}$$

$$\text{Median:} \quad wTC_i^{me} = med_t\{TC_{i,(t)}\} \tag{5}$$

for each isomorphism class $i = 1, \dots, 16$, where T is the number of peeling steps to count the possible weighted triads of a network and $n_{(t)} = \sum_i TC_{i,(t)}$.

The first indicator is the sum of the triad counts of each slice, while wTC_i^m and wTC_i^{wm} are the mean and the weighted mean statistics, respectively. The equation (4) is the average of the proportions of a triad in each slice and the last indicator refers to the median of the triad census, for each isomorphism class.

4. Empirical application

Data for the empirical application come from the group stage of UCL in the season 2018-2019.

The source is the official UEFA website (www.uefa.com) and the data include 192 passing networks, grouped in 6 networks per team (32), each of them referring to a match (for a total of 96 matches). Some descriptive statistics of the passes per team can be observed in Palazzo et al. (2023) (lower panel of Figure 3).

A comparison between football networks is carried out through the correlation analysis of the ranking for each isomorphism class. Figure 2 shows the Spearman correlation coefficients between the wTC measures and the binary triad census. Firstly, a low correlation is observed concerning the class "003" of all wTC measures and the binary case, on the contrary, there is a strong correlation concerning the most connected triad, namely "300". The sum statistic appears to be highly correlated with the binary case for what concerns the most connected triads, especially the "030C" and "030T" and from "120C" on. This feature appears to be weakly replicated in the mean proportion case, although it shows a near-to-zero correlation in the less connected triads ("012", "021C", "021D", "021U", "111D" and "111U"). The mean and weighted mean indicators of the isomorphism classes are the most correlated with the binary case, indicating a coherence of the information retrieved by the two statistics. Among the considered wTC statistics, the median is the indicator presenting the most deviations from the classical triad census. It is possible to recognize in Figure 2 a generally low correlation among the isomorphism classes, with a weak negative correlation between those that represent highly connected triads (i.e., "111D", "111D", "111U", "111D", "120C", "120D" and "120U").

Figure 2: Ranking correlations between the proposed measures of weighted triad census (wTC) and the binary triad census along the 16 isomorphism classes.

The first possibility to investigate the difference between the information provided by weighted and unweighted triads is to show their potential to explain the football outcome, here defined as the probability of winning a match. Thus, the target variable assumes a value equal to 1 if the match is won by the team, and 0 otherwise (draw or defeat). To this end, a conditional logistic regression (CLR) is performed on the 2018-2019 UCL data to understand whether and which isomorphism classes of bTC and wTC can be used in modelling the football outcome. The CLR appears more suitable than conventional logistic regression since each team (in each match) is naturally "matched" with its opponent. A single control variable (Home) is used assuming value 1 if the team plays at home and 0 otherwise, while all the 16 isomorphism classes are included in their bTC or in the wTC (in the case of the arithmetic mean as the summary measure) version. To avoid multicollinearity and redundancy issues, we also carried out a backward selection of variables based on the Akaike information criterion (AIC).

Table 1 shows the output of the CLR models in terms of odds ratios (OR, i.e., the exponential of the coefficients) and p-values of the z-test. Considering the bTC, only home (with a positive effect on the probability of winning the match) and the modality "210" (with a negative effect) appear statistically significant in the full model, while fully connected triads ("300", with a positive effect on the probability of winning) and modality "030C" (with a negative effect) become statistically significant in the restricted model (bTC-Back). Indeed, (binarized) cyclical

relationships can be viewed as a lack of mutual assistance in a subgroup of three players. Regarding models using wTC modalities as covariates, the positive effect of the classes "102" and "021D" arises both in the full and the restricted model. A reasonable interpretation suggests that successful teams can express more reciprocal exchanges of passes ("102") and players that may split the directed passes ("021D"), especially in the higher layers (i.e., binary networks involving a higher number of passes). Thus, mutual dyads and directed relationships may be relevant in a weighted perspective. Furthermore, a larger positive effect of "300" is observed in these models concerning those including bTC, while modalities "201" and "210" are associated with a decrease in the probability of winning the match. The Bayesian Information Criterion (BIC) and the conventional in-sample performance measures (accuracy, sensitivity, and specificity) show the better performance of the models including the wTC-based covariates.

Table 1: Results of CLR applied to 2018-2019 UCL matches.

	bTC-Full		bTC-Back		wTC-Full		wTC-Back	
Variable	OR	p-value	OR	p-value	OR	p-value	OR	p-value
Home	2.674	0.041 (*)	2.139	0.046 (*)	2.080	0.117	2.063	0.061 (.)
"012"	0.981	0.667	1.022	0.164	0.953	0.591		
"021D"	1.112	0.319			3.722	0.089 (.)	2.774	0.050 (*)
"021U"	0.977	0.903			0.773	0.677		
"021C"	1.018	0.845			0.740	0.500	0.622	0.100 (.)
"102"	1.028	0.190			1.251	0.037 (*)	1.207	0.003 (**)
"111D"	1.080	0.179	1.065	0.123	1.228	0.598		
"111U"	0.981	0.594			0.956	0.850		
"030T"	0.920	0.541			0.701	0.758		
"030C"	0.588	0.109	0.588	0.048 (*)	0.114	0.189	0.120	0.148
"120D"	0.976	0.746			0.955	0.949		
"120U"	0.987	0.899			1.671	0.424		
"120C"	1.001	0.985			0.930	0.904		
"201"	0.973	0.158	0.974	0.141	0.704	0.108	0.760	0.032 (*)
"210"	0.925	0.041 (*)	0.929	0.013 (*)	0.504	0.041 (*)	0.601	0.021 (*)
"300"	1.035	0.103	1.039	0.007 (**)	1.822	0.019 (*)	1.713	0.005 (**)
BIC	122.9		86.0		117.4		84.6	
In-sample Accuracy	0.87		0.85		0.89		0.87	
In-sample Sensitivity	0.82		0.82		0.83		0.82	
In-sample Specificity	0.90		0.87		0.93		0.90	

Statistical significance; (.): p <0.1; (*): p<0.05; (**); p <0.01.

5. Concluding remarks

In real cases, the differences between the weights of the edges can be relevant to defining the underlying structure given by the isomorphism classes, especially in the case of weighted and dense networks.

The peeling of the network is a possible solution to deal with such cases. The algorithm peels out a weighted network in many binary layers allowing the count of the isomorphism classes to measure the weighted triad census. As shown through the case study, it proves particularly helpful for small networks, even if it can be easily extended to medium and large ones. Indeed, preliminary results of conditional logistic regression models suggest that wTC-based information may be more appropriate with respect to bCT to model the football outcome.

Further advances should regard an extensive simulation study to compare the results of bTC and wTC, even considering different data-generating processes for the networks. In addition, an in-depth evaluation of the information lost due to the use of the binary triad census instead of the weighted triad census should be developed. This will be also carried out considering more

sophisticated statistical models including in-match statistics along with the bTC- or wTC-based information and computing out-of-sample performance indicators.

References

Caicedo-Parada, S., Lago-Penas, C., Ortega-Toro, E. (2020). Passing networks and tactical action in football: A systematic review. *International Journal of Environmental Research and Public Health*, **17**(18), pp. 1–17.

Diquigiovanni, J., Scarpa, B. (2019). Analysis of association football playing styles: An innovative method to cluster networks. *Statistical Modelling*, **19**(1), pp. 28-54.

Faust, K. (2010). A puzzle concerning triads in social networks: Graph constraints and the triad census. *Social Networks*, **32**(3), pp. 221-233.

Grund, T. U. (2012). Network structure and team performance: The case of English Premier League soccer teams. *Social Networks*, **34**(4), pp. 682-690.

Håland, E. M., Wiig, A. S., Hvattum, L. M., Stalhane, M. (2020). Evaluating the effectiveness of different network flow motifs in association football. *Journal of Quantitative Analysis in Sports*, **16**(4), pp. 311-323.

Holland, P. W., Leinhardt, S. (1978). An omnibus test for social structure using triads. *Sociological Methods & Research*, **7**(2), pp. 227-256.

Ievoli, R., Palazzo, L., Ragozini, G. (2021). On the use of passing network indicators to predict football outcomes. *Knowledge-Based Systems*, **222**, 106997, pp. 1-10.

Ievoli, R., Gardini, A., Palazzo, L. (2023). The role of passing network indicators in modelling football outcomes: an application using Bayesian hierarchical models. *AStA Advanced in Statistical Analysis*, **107**(1-2), pp. 53-71.

Palazzo, L., Ievoli, R., Ragozini, G. (2023). Testing styles of play using triad census distribution: an application to men's football. Journal of *Quantitative Analysis in Sports*, **19**(2), pp. 125-151.

Rondinelli, R., Palazzo, L. (2023). A first proposal of the triad census for weighted networks: An application to football, in A. Bucci, A. Cartone, A. Evangelista, A. Marletta, eds. *Book of Short papers 11th International Conference IES 2023 Statistical Methods for Evaluation and Quality: Techniques, Technologies and Trends (T3)*. Il Viandante, Chieti (IT), pp. 90-95.

Wasserman, S., Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge.

doi.org/10.26398/asaproc.0043

Exploiting consumer sentiment in volume sales forecasting: case studies

Domenica Fioredistella Iezzi [a], Roberto Monte [b]
[a] Department of Enterprise Engineering Mario Lucertini, Tor Vergata University of Rome, Italy
[b] Department of Civil Engineering and Computer Science Engineering, Tor Vergata University of Rome, Italy

1. Introduction

The impact of word of mouth on sales has been well known since before the arrival of the Internet and the explosion of the social media phenomenon. Paul Lazarsfeld's studies have highlighted that media messages can be further mediated by informal opinion leaders who intercept, interpret, and disseminate what they see and hear in the personal networks where they are inserted (Katz et al., 2006). The advent of the Internet with the introduction of digital word of mouth has increased and accelerated the assessment of messages about products and services. Social media, in particular, has dramatically amplified the echo of traditional word of mouth (Huete-Alcocer, 2017).

In the summer of 2023, we witnessed media phenomena linked to the publication of receipts on social media, highlighting the high prices of products or services in tourist locations. This information had a boomerang effect with large shares on the net and strongly negative comments. According to Yang (2017), a negative or positive attitude of customers towards a product or a service will influence customers' future purchase intentions.

This paper aims to identify the best forecast of sales of two goods using state-space models with signals produced by consumer feelings (Iezzi and Monte, 2022, 2023; Basili et al., 2023). We measure the consumer sentiment score to track attitudes expressed in tweets. We apply different approaches: from the Valence Aware Dictionary and sEntiment Reasoner (VADER) rules-based model, using the estimation of negative, positive, and compound feelings (Hutto and Gilbert, 2014), to machine learning-oriented techniques relying on logistic regression (Prabhat and Khullar 2017, to deep learning Word2Vec[1] (Mikolov et al., 2013). We want to investigate which approach might generate a better signal for predicting the sales of the goods. To forecast sales volumes, we use different models and compare the results: the ETS-AAA model[2] (Additive error, Additive slope component in the local trend, Additive seasonality), the ETS-ANA Model (Additive error, No additive slope component in the local trend, Additive seasonality), and Multivariate Autoregressive State-Space (MARSS) models (Harvey, 1990, Hyndman et al. 2002). In particular, to implement the MARSS models, which account for consumer feelings, we have collected two corpora from Twitter and built some sentiment signals with R and Python software.

Section 2 presents data and models; Section 3 presents the main findings; Section 4 concludes.

[1] Word2Vec is a group of models that are used to produce word embeddings. These models are two-layer neural networks trained to reconstruct words' linguistic contexts.

[2] The ETS models, by R. Hyndman and his coworkers (e.g., Hyndman et al., 2002), are a well-known family of time series models, consisting in the state-space form of the Holt-Winters models, having, in general, an error component (E), a local trend with an additional slope component (T), and a seasonal component (S).

2. Data and models

The two corpora collected from Twitter concern Toyota Camry US (165,000 tweets) and Gentilini Osvego biscuits (1,016 tweets). We consider Toyota Camry tweets from June 1, 2009, to December 31, 2022, and Gentilini Osvego biscuits tweets from January 1, 2013, to December 31, 2020. We use the Twitter Academic Research Product Track based on our academic profile (courtesy of Twitter Developer) to scrape Twitter time series [3]. Gentilini Osvego biscuits tweets are in Italian, but we translate them using the library "googleLanguageR", an R package allowing speech-to-text transcription, neural net translation, and natural language processing via the Google Cloud machine learning set. The choice to implement a supervised machine learning model has imposed the need to introduce a labeled dataset for training it. Between different possible options, we have decided to use Sentiment140, a dataset of 1.6 millions of positively and negatively labeled tweets. The reasons are essentially due to the following: the corpus of tweets collected, mainly concerning user opinions related to brands, products, or topics in general, was conceived precisely for classifying sentiment on Twitter.

We use the ETS-ANA and ETS-AAA models, introduced by Hyndman et al. (2008), as benchmarks. These can be written as follows:

$$\text{ETS-ANA} \begin{cases} y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t, \\ \ell_t = \ell_{t-1} + \alpha\varepsilon_t, \\ s_t = s_{t-m} + \gamma\varepsilon_t; \end{cases} \qquad \text{ETS-AAA} \begin{cases} y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t, \\ \ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t, \\ b_t = b_{t-1} + \beta\varepsilon_t, \\ s_t = s_{t-m} + \gamma\varepsilon_t, \end{cases}$$

where y_t is the value of the time series of interest at time t, the hidden variable ℓ_t [resp. b_t, resp. s_t] is the *local level* [resp. *slope*, resp. *seasonality*] of the Holt-Winters decomposition of y_t, and the variable ε_t represents the innovation term at time t with variance σ^2. The parameter σ^2 is determined together with the parameters λ, β, γ, and the initial states of the model in the estimation procedure. The only difference between the ETS-ANA and ETS-AAA models is the lack of the slope component in the former. Referring to the ETS-AAA model, our idea is to generalize the model in the context of State-Space models by attributing the role of consumer sentiment for the product to the slope variable b_t. Naturally, we cannot expect to observe this variable, which is kept as a hidden variable, while we assume to observe a proxy, which is the sentiment signal built from time to time. Accordingly, in this paper, we study a state space model in which the state and observation equations take the following forms:

$$\text{state equations} \begin{cases} \ell_t = \beta_{\ell,\ell}\ell_{t-1} + \beta_{\ell,b}b_{t-1} + \sigma_{\ell,\ell}w_t^{(l)}, \\ b_t = \beta_{b,b}b_{t-1} + \sigma_{b,b}w_t^{(b)}, \\ s_t = s_{t-m} + \sigma_{s,s}w_t^{(s)}, \end{cases} \tag{1}$$

$$\text{observation equations} \begin{cases} y_t = \beta_{y,\ell}\ell_t + \beta_{y,b}b_t + \beta_{y,s}s_t + \sigma_{y,y}w_t^{(y)}, \\ z_t = \beta_{z,b}b_t + \sigma_{z,z}w_t^{(z)}. \end{cases}$$

Here, y_t is still the observed sales volume time series of the product, and z_t represents the observed sentiment scores on the hidden sentiment variable b_t. To add more flexibility to our model compared to the ETS-AAA model, we introduce the additional parameters $\beta_{\ell,\ell}, \ldots, \beta_{z,b}$, Moreover, we introduce independent innovations $w_t^{(l)}, \ldots, w_t^{(z)}$, with variances $\sigma^2_{\ell,\ell} \ldots, \sigma^2_{z,z}$, respectively. All these parameters, together with the initial states of the models, are estimated in a recursive procedure using the functions of the *MARSS R* library (Holmes et al., 2003).

[3] see https://developer.twitter.com/en/blog/product-news/2021/enabling-the-future-of-academic-research-with-the-twitter-api

3. Discussion

To test our model, we consider the Toyota Camry US monthly sales volume time series from June 1, 2009, to December 31, 2022, and the Gentilini Osvego biscuits monthly sales volume from January 1, 2012, to December 31, 2020. We apply the ETS-AAA and ETS-ANA models to these time series and monthly sales volume logarithms as a benchmark and, as a further benchmark, we also use a univariate MARSS model, which is much closer to the ATS-AAA model since it does not account for the sentiment signal. The large difference in the scale of the monthly sales volume and sentiment signals has suggested the opportunity to use the monthly sales volume logarithm time series. To measure the forecasting accuracy in the models, the dataset was divided into two parts: training set (90%) and test set (10%). The results of our comparative analysis are summarized in Tables 1-2. We consider the negative VADER, logistic regression, and Word2Vec as the sentiment signals. Figures 1-2 show the Toyota Camry sales volume and the Gentilini Osvego monthly sales volume logarithm as forecasted by bivariate MARSS models accounting for the sentiment signal.

Figure 1 highlights the existence of some shared peaks and troughs between the observations of each year. March, May, and August have the highest sales over the years, while February and November have the lowest. Possible anomalous values are also evident, such as those recorded in April 2020 and August 2009, the first of which is undoubtedly due to the restrictive measures imposed worldwide after the outbreak of the COVID-19 pandemic.

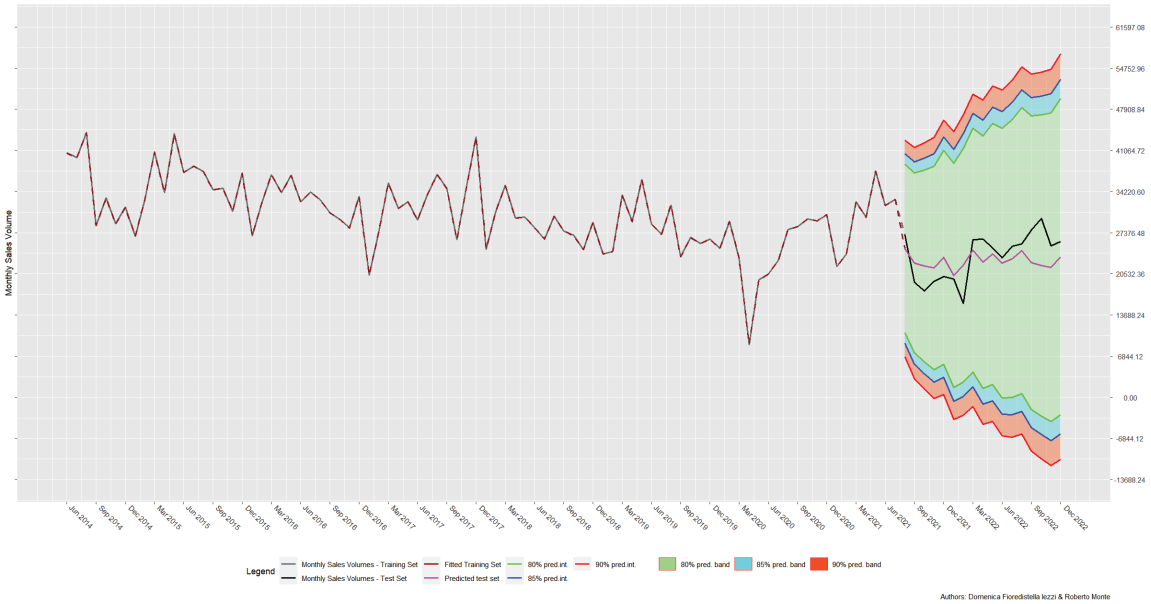


Figure 1: **Toyota Camry US - MARSS bivariate model with monthly sales volumes and Word2Vec sentiment observed signals**

From Tables 1-2, the accuracy measures[4] MAPE, SMAPE%, MASE, and RMSSE show that MARSS bivariate models outperform all univariate benchmark models. In this context, RMSE and MAE are less reliable. The comparison of the information measures is less indicative because the structures and the numbers of parameters in the bivariate and univariate models differ.

[4]Accuracy measures: logLik=log-likelihood, AIC=Akaike information criterion, AICc=Akaike information criterion corrected, BIC=Bayes Information Criterion, RMSE=Root Mean Square Error, MAE=Mean Absolute Error, MAPE=Mean Absolute Percentage Error, SMAPE=Symmetric mean absolute percentage error, MASE=Mean Absolute Scaled Error, RMSSE= Root Mean Square Error.

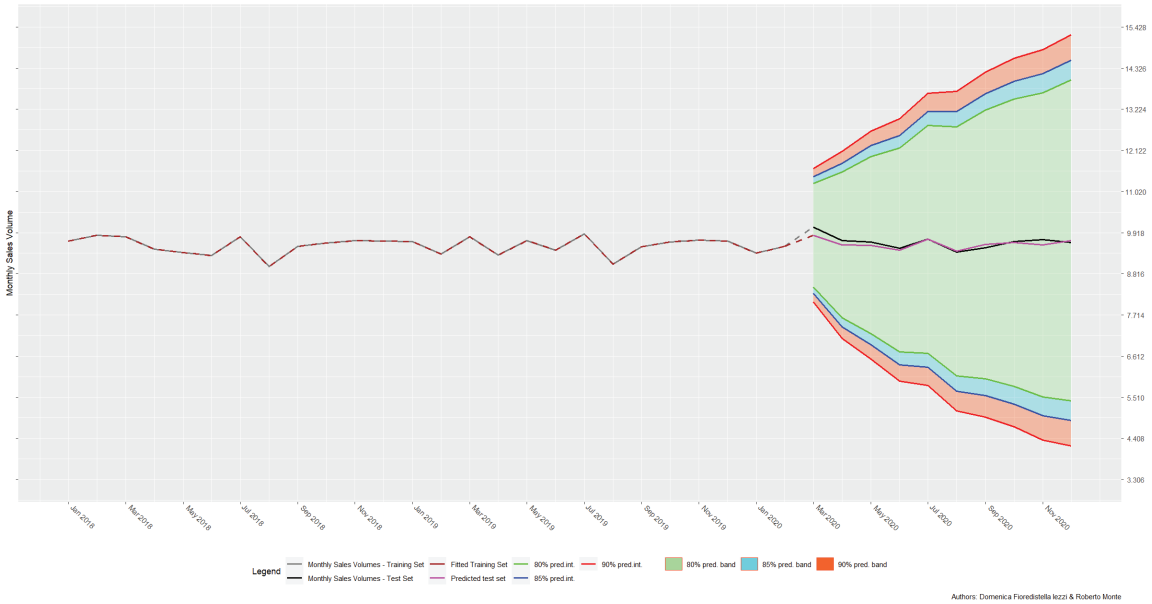


Figure 2: **Gentilini Osvego biscuits - MARSS bivariate model with monthly sales volumes logarithm and negative VADER sentiment observed signals**

Models	logLik	AIC	BIC	AICc	RMSE	MAE	MAPE	SMAPE%	MASE	RMSSE
ETS-ANA (y)	-1591.434	3212.869	3257.623	3216.561	9440.055	8610.602	39.914	15.902	1.669	1.389
ETS-AAA (y)	-1592.126	3218.252	3268.973	3223.033	10809.330	10241.120	46.714	18.321	1.985	1.590
MARSS (y) - No Sent.	-1567.278	3154.555	3184.319	3156.185	7291.802	6087.463	28.754	11.831	1.180	1.073
MARSS (y) - Neg. VADER	-1562.234	3148.469	3184.272	3150.814	4599.050	3608.202	17.271	7.811	0.699	0.676
MARSS (z) - Neg. VADER	-1562.234	3148.469	3184.272	3150.814	0.026	0.018	59.089	24.169	1.156	0.971
MARSS (y) - LR	-1789.955	3621.910	3653.713	3624.255	3947.773	3119.797	14.460	6.925	0.605	0.580
MARSS (z) - LR	-1789.955	3621.910	3653.713	3624.255	0.118	0.062	9.790	5.319	0.818	0.990
MARSS (y) - Word2Vec	-1435.063	2894.125	2929.928	2896.471	3625.778	3064.401	13.633	6.741	0.594	0.533
MARSS (z) - Word2Vec	-1435.063	2894.125	2929.928	2896.471	0.076	0.036	7.971	3.235	0.977	0.992

Table 1: **Information parameters and accuracy measures for Toyota Camry US monthly sales volumes: ETS-ANA model, ETS-AAA model, MARSS univariate, and MARSS bivariate with negative VADER, logistic regression (L.R.), and Word2Vec sentiment signals**.

Models	logLik	AIC	BIC	AICc	RMSE	MAE	MAPE	SMAPE%	MASE	RMSSE
ETS-ANA (y)	-34.444	98.887	135.703	105.744	0.209	0.168	1.724	0.873	0.743	0.755
ETS-AAA (y)	-36.170	106.341	148.065	115.341	0.227	0.189	1.939	0.983	0.836	0.818
MARSS (y) - No Sent.	-128.888	277.777	302.320	280.710	0.122	0.085	0.065	0.437	0.376	0.438
MARSS (y) - Neg. VADER	48.060	-72.121	-42.669	-67.847	0.101	0.081	0.829	0.416	0.358	0.364
MARSS (z) - Neg. VADER	48.060	-72.121	-42.669	-67.847	0.043	0.029	-	43.208	0.788	0.891
MARSS (y) - LR	-5.606	35.212	64.664	39.486	0.132	0.086	0.877	0.442	0.381	0.477
MARSS (z) - LR	-5.606	35.212	64.664	39.486	0.045	0.040	6.984	3.596	0.630	0.525
MARSS (y) - Word2Vec	-55.288	134.576	164.028	138.850	0.105	0.068	0.6901	0.347	0.300	0.380
MARSS (z) - Word2Vec	-55.288	134.576	164.028	138.850	0.120	0.088	15.913	7.276	0.825	0.816

Table 2: **Information parameters and accuracy measures for Gentilini Osvego biscuits monthly sales volumes logarithm: ETS-ANA model, ETS-AAA model, MARSS univariate, and MARSS bivariate with negative VADER, logistic regression (L.R.), and Word2Vec sentiment signals.**

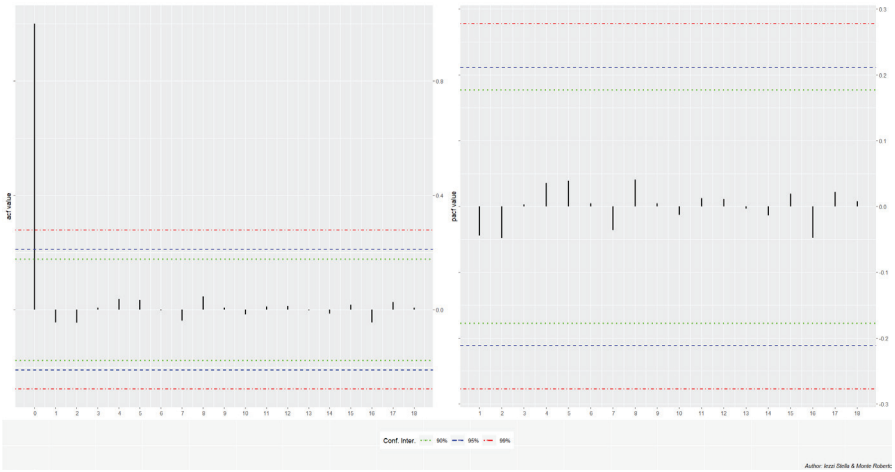


Figure 3: **Autocorrelogram and partial autocorrelogram of the residuals of the bivariate MARSS model for the Osvego Gentilini monthly sales volumes logarithm and negative Vader sentiment**

Figure 1 illustrates the MARSS bivariate model with monthly sales volumes and Word2Vec sentiment observed signals for the Toyota Camry US. So far, this appears to be the best model for our case study. The fitted monthly sales volumes are so close to the states in the training set (it is almost impossible to distinguish the training and fitted set in the plot) that one could think of a typical case of overfitting. However, the predicted monthly sales volumes are so close to the test set that overfitting is unlikely. Figure 2 describes the MARSS bivariate model with the monthly sales volumes logarithm and negative VADER sentiment observed signals for the Gentilini Osvego biscuits. As observed in other research (Iezzi and Monte, 2023; Xiaolin et al., 2019), negative sentiment provides a better signal than the others (positive and compound) possible with VADER. The same considerations as in Figure 1 apply, with even stronger emphasis due to the logarithm transformation, which corrects the difference in scale of the observed signals. In this case, we also provide the plot of the autocorrelogram and partial autocorrelogram of the residuals (see Figure 3), which renders the suspect of overfitting unlikely. The logarithmic transformation of sales volumes generally improves the performance of the models because it tends to reduce the disproportion between the signals. It also manages to reduce sensitivity to extreme deviations by reducing the effect of extreme values (outliers), thus making the data distribution less influenced by them.

4. Conclusions

The results highlight that using a sentiment signal may improve the models' predictive capabilities. This advocates using data from customers' word of mouth. The different models tested for sentiment quantification (VADER, Logistic Regression, and Word2Vec) all show good results with a slight advantage for VADER and Word2Vec. The reason might be that VADER, a model-based approach built on social media, also allows the quantification of emoticons and symbols, capturing the nuances of sentiment in tweets well. Word2Vec is a neural network model that converts words into vectors of real numbers so that semantically similar words are represented by neighboring vectors in vector space. This approach helps to analyze and understand the meaning of words in a specific context. However, it is essential to note that Word2Vec may not fully capture the irony or sarcasm present in tweets, and its effectiveness depends on

the quantity and quality of training data. Given those results in future tests, other more advanced deep learning models such as LSTM (Long Short-Term Memory) and BERT (Bidirectional Encoder Representations from Transformers) will be applied. The recurring structure of LSTMs makes them suitable for capturing temporal relationships and dependencies in text, which can be crucial in sentiment analysis. BERT captures bidirectional contexts of words, considering both directions of the context in which a word appears. This allows for a better understanding of the context of words, improving sentiment analysis.

References

Basili, R., Croce, D., Iezzi, D.F., Monte, R. (2023). The role of BERT in neural network sentiment scoring for time series forecast. In *Proceedings of the Statistics and Data Science Conference*, P. Cerchiello, A. Agosto, S. Osmetti, A. Spelta (eds.) pp. 55-60. PAVIA, Pavia University Press.

Harvey, A.C. (1990). *The Econometric Analysis of Time Series*, 2nd Edition, MIT Press Books, The MIT Press, Cambridge, Massachusetts.

Holmes, E.E, Ward, E.J., Scheuerell, M.D., Wills, K. (2023). MARSS: Multivariate Autoregressive State-Space Modeling. *R Package Version 3.11.8*.

Huete-Alcocer, N. (2017). A Literature review of word of mouth and electronic word of mouth: Implications for consumer behavior. *Frontiers in Psychology* **8** pp. 1-4.

Hutto, C., Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*. **8**(1) (May 2014), pp. 216-225.

Hyndman, R., Koehler, A., Snyder, R., Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting* **18**(3), pp. 439-454.

Hyndman, R.J., Akram, M., Archibald, B.C. (2008). The admissible parameter space for exponential smoothing models. *Annals of the Institute of Statistical Mathematics* **60**(2), pp. 407-426.

Iezzi, D.F., Monte, R. (2022). Sales forecast and electronic word of mouth: the power of feelings. In *Proceedings of the 16th International Conference on statistical analysis of textual data (JADT22)* **2**, pp. 489-494. M. Misuraca, G. Scepi, M. Spano (eds.)

Iezzi, D.F., Monte, R. (2023 - in press). Social media effects on sales: consumer sentiment in a state-space model. In *New Frontiers in Textual Data Analysis* (Book Series: Studies in Classification, Data Analysis, and Knowledge Organization) M. Misuraca, G. Giordano (eds.), Springer.

Katz, E., Lazarsfeld, P.F., Roper, E., (2006). *Personal Influence: The Part Played by People in the Flow of Mass Communications* (1st ed.). Routledge.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems 26 (NIPS 2013)*.

Prabhat, A., Khullar, V., (2017). Sentiment classification on big data using Naïve Bayes and logistic regression, *2017 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, pp. 1-5.

Xiaolin, Li, Chaojiang, Wu, Fen, M. (2019). The effect of online reviews on product sales: A joint sentiment-topic analysis, *Information & Management*, **56**(2), pp. 172-184.

Yang, F.X. (2017). Effects of Restaurant Satisfaction and Knowledge Sharing Motivation on eWOM Intentions: the Moderating Role of Technology Acceptance Factors. *Journal of Hospitality & Tourism Research*. **41**, pp. 93-127.

doi.org/10.26398/asaproc.0044

Assessing the impact of green technology innovation, environmental tax and natural resources towards ecological sustainability in OECD countries: The role of digitalization

Aamir Javed[a], Agnese Rapposelli[b]
[a] Department of Management and Business Administration, "G. D'Annunzio" University of Chieti-Pescara, Pescara, Italy.
[b] Department of Economic Studies, "G. D'Annunzio" University of Chieti-Pescara, Pescara, Italy.

1. Introduction

Many countries are working relentlessly for mitigating the adverse effects of climate change, land degradation, desertification and other associated human environmental distortions. The increasing awareness of the dangerous effects of greenhouse gases (GHGs) on people and ecosystems, as well as the introduction of carbon tax, innovations, and energy-saving and efficient technologies, are some of the specific strategies targeted at reducing carbon dioxide (CO_2) emissions (a major contributor to greenhouse gases) and other greenhouse gases. Despite the efforts done to mitigate this global menace, the Intergovernmental Panel on Climate Change report shows an increase in CO_2 emissions from 9434.4 to 34,649.4 million tons between 1961 and 2011 (IPCC 2014). The Energy International Agency also shows that CO_2 emissions increased from 29,714.2 to 33,444.0 million tons between 1999 and 2017. In supporting these statistics, the Paris Agreement, signed in 2015, reveals that CO_2 global emissions grew by 1.3% between 2006 and 2016, and 1.6% in 2017, thus constituting about 81% of greenhouse gases.

Environmental pollution has a multi-dimensional effect on the ecological system; hence, the proxy used for environmental quality has remained mixed. Even if CO_2 emissions are extensively investigated in the existing literature and remain the core of international climate change agreements, there are other important factors, such as deterioration in the quality of soil, forest, and water, which are facing severe ecological threats (Ansari et al., 2021). These environmental factors represent an integral part of the ecosystem.

Hence, an ecological footprint indicator based on the concept of carrying the capacity of ecosystem is an important issue in the ecological system. Moreover, green innovation (which includes different techniques, such as waste recycling, energy-saving, pollution prevention and environmental management) provides benefits to consumers and organizations and helps reduce adverse environmental effects (Razzaq et al., 2023; Javed et al., 2023). Besides, renewable energy consumption represents one of the major determinants in reducing the level of CO_2 emissions. Furthermore, environmental taxes - aiming at the taxation of carbon emissions thus contributing to environment protection by internalizing negative externalities in the form of environmental pollution - represent the most effective policy tool for reducing GHGs emissions (Bashir et al., 2020).

The aim of this study is to investigate the impact of green innovation, environmental taxes, digitalization, natural resources and renewable energy consumption, along with economic growth and trade openness, on the ecological footprint for OECD countries during the period 1994-2018. This study contributes to the literature in several ways. First, we selected the OECD economies for our analysis. OECD countries are the largest emitter of GHGs in the world.

Nowadays, they emit about 35% of global energy-related CO_2 emissions: while decreasing in OECD America and OECD Europe, CO_2 emissions from energy use continue to grow in the OECD Asia-Oceania region (OECD, 2019). Therefore, there is still room for improvement on the issue of reducing environmental harm in the OECD countries.

Second, in contrast to existing studies, which have mainly employed CO_2 emissions as an indicator of environmental quality, this study uses the ecological footprint - a more comprehensive indicator of environmental accounting - as a proxy for the quality of the environment. Finally, the relationship between the aforementioned variables (green innovation, environmental taxes, digitalization, natural resources, renewable energy consumption, economic growth and trade openness) and the ecological footprint is analyzed by employing a second-generation cointegration technique, the cross-sectional augmented autoregressive distributed lag model (CS-ARDL). Remarkably, to our best knowledge, the current study is the first work to employ this approach to test the short- and long-run impact of these indicators on ecological footprint.

The paper is organized as follows. Section 2 introduces the methodology used and describes the data, Section 3 presents the results obtained and Section 4 concludes.

2. Data and methodology

This research aims at investigating the impact of green technology innovation, digitalization, environmental taxes, natural resources along with renewable energy consumption, trade openness and economic growth on the ecological footprint in OECD countries for the period 1994-2018. The above variables are listed and described in Table 1. To avoid heteroscedasticity and ensure robust vector results the variables are expressed in logarithmic form.

Table 1: **Description of variables**

Variables	Description	Source
EFP	Ecological footprint (global hectares per person)	GFN
GTI	Green technology innovation (number of patents related to environmental technologies on total patents)	OECD
DIG	Sum of mobile and fixed telephone subscriptions per 100 inhabitants	WDI
ET	Environmental taxes (% of GDP)	OECD
NRR	Natural resources rent (% of GDP)	WDI
REN	Renewable energy consumption (% of total final energy consumption)	WDI
TO	Sum of exports and imports (% of GDP)	WDI
GDP	Gross domestic product (constant 2015 US$) per capita	WDI

To examine the impact of the considered explanatory factors on the ecological footprint, we specify the following empirical model:

$$lnEFP_{it} = \beta_o + \beta_1 lnGTI_{i,t} + \beta_2 lnDIG_{i,t} + \beta_3 lnET_{i,t} + \beta_4 lnNRR_{i,t} + \beta_5 lnREN_{i,t} + \beta_6 TO_{i,t} + \beta_7 lnGDP_{i,t} + \varepsilon_{it} \quad (1)$$

where EFP, GTI, DIG, ET, NRR, REN, TO and GDP denote the ecological footprint, green technology innovation, digitalization, environmental taxes, natural resources rent, renewable energy consumption, trade openness and economic growth, respectively.

This study employs the CS-ARDL model to examine the short and long-term impacts of the independent variables on the explained variable. Unobserved common factors impacting all sections may make panel data extremely vulnerable to cross-section dependency; if there is evidence of cross-sectional dependence in the data, regression analysis findings may be inaccurate. To address the presence of slope heterogeneity and cross-sectional dependence the best estimator is the CS-ARDL one. More specifically, the CS-ARDL approach employs a dynamic common correlated effects estimator (Chudik and Pesaran, 2015) to address the aforementioned issues. This model helps estimate the dynamic common correlation effects while also accounting for time dynamics, cross-sectional heterogeneity, and cross-country dependence in the errors (Chudik et al., 2016).

We specify our CS-ARDL model as follows:

$$EFP_{i,t} = \sum_{i=o}^{p_u} \alpha_{i,t} EFP_{i,t-1} + \sum_{i=0}^{p_v} \delta_{i,t} V_{i,t} + \varepsilon_{it} \tag{2}$$

$$EFP_{i,t} = \sum_{i=o}^{p_u} \alpha_{i,t} EFP_{i,t-1} + \sum_{i=0}^{p_v} \delta_{i,t} V_{i,t} + \sum_{i=0}^{p_w} \gamma_i \overline{W}_{t-1} + \varepsilon_{i,t} \tag{3}$$
$$\overline{W}_{t-1} = (\overline{EFP}_{t-1}, \overline{V}_{t-1})$$

$$\hat{\theta}_{CD-ARDL.i} = \frac{\sum_{I=0}^{Pv} \hat{\delta}_{Ii}}{1 = \Sigma = 0} \tag{4}$$

Furthermore, the mean group is defined as follows:

$$CIPS = \hat{\theta}_{MG} = \frac{1}{N} \sum_{i=1}^{N} \hat{\delta}_i \tag{5}$$

Based on this, the short-term coefficient approximation is reported as:

$$\Delta EFP_{i,t} = \theta_i \left[EFP_{i,t-1} - \theta_i V_{i,t-1} \right] - \sum_{i=o}^{p_{u-1}} \alpha_{i,t} \Delta_i EFP_{i,t-1} + \sum_{i=0}^{p_v} \delta_{i,t} \Delta_i V_{i,t} + \sum_{i=0}^{p_w} \gamma_i \overline{W}_{t-1} + \varepsilon_{i,t} \tag{6}$$

where

$$\Delta_i = t - (t-1) \tag{7}$$

$$\hat{\tau}_i = -\left(1 - \sum_{i=o}^{p_u} \alpha_{i,t}\right) \tag{8}$$

$$\hat{\theta}_i = \frac{\sum_{i=o}^{Pv} \hat{\delta}_{i,t}}{\hat{\tau}_i} \tag{9}$$
$$\hat{\theta}_{MG} = \frac{1}{N} \sum_{i=1}^{N} \hat{\delta}_i.$$

3. Empirical results

Before estimating the model, we apply some preliminary tests. Table 2 reports the results obtained after testing the presence of slope coefficients heterogeneity. The results indicate that both the parameters $\tilde{\Delta}$ and $\tilde{\Delta}$ adjusted are significant and provide evidence that the slope coefficients are heterogenous in the considered panel. Table 3 lists the results of the cross-sectional dependence analysis obtained by applying Breusch and Pagan (1980) LM test, Pesaran (2015) scaled LM test and Pesaran (2007) CD test. These results provide evidence against the null hypothesis of no cross-sectional dependence.

Table 2. Slope heterogeneity test results

Test	Statistic	Prob.
$\tilde{\Delta}$	12.340***	0.000
$\tilde{\Delta}$ adjusted	14.651***	0.000

*** indicates significance at 1% level.

Table 3. Cross-sectional dependence test results

Test	Statistic	Prob.
Breusch-Pagan LM	1460.250***	0.000
Pesaran scaled LM	34.640***	0.000
Pesaran CD	12.445***	0.000

*** indicates significance at 1% level.

We tested the variables included in our model for stationarity by implementing the 2nd-generation unit root test (Pesaran, 2007), which is more powerful in dealing with slope heterogeneity and panel cross-sectional dependence. The results, listed in Table 4, indicate that the variables considered are stationary at the first difference. Hence, the stationarity of all variables allows us to examine the long-run elasticities of the variables under analysis. Moreover, the results of Westerlund cointegration test (Westerlund, 2017), shown in Table 5, confirm a stable, long-term cointegration among the variables used.

Table 4. Unit root test results

Variables	Intercept and trend	
	I(0)	I(1)
EFP	-2.654	-5.658***
GTI	-2.560	-4.562***
DIG	-3.145	-6.125**
ET	-2.654	-2.689***
NRR	-2.350	-3.145***
REN	-1.658	-6.456**
TO	-3.567	-4.506***
GDP	-1.526	-2.256***

Note: *** and ** indicate significance at 1% and 5% level, respectively.

Table 5. Westerlund cointegration test results

Statistic	Value	Z-value	Robust P-value
Gt	-4.568***	-3.268	0.000
Ga	-3.159***	-2.564	0.005
Pt	-8.105***	-2.309	0.000
Pa	-5.456*	-3.325	0.072

Note: *** and * indicate significance at 1% and 10% level, respectively.

The results of the CS-ARDL model are presented in Table 6. First, they demonstrate that green technology innovation and environmental taxes reduce the level of ecological footprint, thus suggesting that any increase in these variables promotes the environmental sustainability of OECD economies. These results are in line with the findings shown by Liu et al. (2022), Feng et al. (2022) and Dogan et al. (2022). Additionally, by focusing on digitalization and renewable energy consumption we observe a negative relationship with

ecological footprint in both the short and long run, too. These findings are in line with previous literature (Adekoya et al., 2022; Zhang et al., 2022) and imply that increasing digitalization and sustainable energy use would mitigate the environmental deterioration for OECD countries. Finally, our results show that natural resources rent, economic growth and trade openness positively influence the pollution. This imply that higher income, natural resources extraction and trading activities in the underlying economies represent important factors in OECD countries' economic degeneration.

Table 6. CS-ARDL results

	Long-run estimates				Short-run estimates			
	Coef.	Std. E.	Z	P >Z	Coef.	Std. E.	Z	P >Z
GTI	-0.035***	0.006	-2.70	0.009	-0.010**	0.055	-1.25	0.039
ET	-0.223**	0.013	-1.21	0.040	-0.042*	0.252	-0.56	0.085
DIG	-0.442***	0.305	-0.30	0.000	-0.325**	0.056	-1.26	0.024
NRR	0.023**	0.011	3.13	0.043	0.019**	0.008	2.20	0.028
REN	-0.054**	0.044	-2.29	0.015	-0.032***	0.719	-1.12	0.004
GDP	18.066***	0.652	3.52	0.000	10.455***	1.210	3.15	0.000
TO	0.440***	0.074	0.50	0.000	0.325***	2.541	1.66	0.005
ECM	-	-		-	-0.461***	0.044		0.000

Note: ***, ** and * indicate significance at 1%, 5% and 10% level, respectively.

4. Conclusion

This work investigates the ecological sustainability in OECD countries by employing a second-generation cointegration methodology, the cross-sectional augmented autoregressive distributed lag model. We examined the impact of green technology innovation, environmental taxes, digitalization, natural resources and renewable energy consumption on the ecological footprint for the period 1994-2018. The results reveal that green technology innovation, environmental taxes, natural resources and renewable energy consumption are the most impacting factors of ecological sustainability in OECD countries. Besides, natural resources rent, trade openness and economic growth significantly harm the environmental quality.

The empirical findings of this study suggest some important policy actions for the achievement of ecological sustainability in OECD countries by developing green technology innovation, implementing environmental tax and restoring natural resources. More specifically, OECD countries authorities must plan more carefully the implementation of fiscal incentives for green energy investment while putting more effort into ensuring favourable conditions for encouraging clean energy investors. These actions will significantly boost the actualization of crucial goals for the development of renewable energy and targets for a sustainable environment via a lower carbon emission for each country in the OECD group. This will not only motivate new investments in clean energy, but it will also encourage existing investors in non-renewable energy sources, such as coal and fossil fuel sources, to gradually shift their attention toward green energy investments.

References

Adekoya, O.B., Oliyide, J.A., Fasanya, I.O. (2022). Renewable and non-renewable energy consumption–Ecological footprint nexus in net-oil exporting and net-oil importing

countries: Policy implications for a sustainable environment. *Renewable Energy*, **189**, pp. 524-534.

Ansari, M.A., Khan, N.A. (2021). Decomposing the trade-environment nexus for high income, upper and lower middle income countries: What do the composition, scale, and technique effect indicate? *Journal of Econometrics*, **121**, 107122.

Bashir, M.F., Ma, B., Shahbaz, M., Jiao, Z. (2020). The nexus between environmental tax and carbon emissions with the roles of environmental technology and financial development. *Plos One*, **15**(11), pp. 1-20.

Breusch, T.S., Pagan, A.R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. *The Review of Economic Studies*, **47**(1), pp. 239-253.

Chudik, A., Pesaran, M.H. (2015). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Ecological Indicators*, **188**(2), pp. 393-420.

Chudik, A., Mohaddes, K., Pesaran, M.H., Raissi, M. (2016). Long-run effects in large heterogeneous panel data models with cross-sectionally correlated errors, in *Essays in Honor of Aman Ullah* (*Advances in Econometrics, Vol. 36*) eds. G. Gonzales-Rivera, R. Carter Hill and T.H. Lee, Emerald Group Publishing Limited, Bingley (UK), pp. 85-135.

Doğan, B., Chu, L.K., Ghosh, S., Truong, H.H.D., Balsalobre-Lorente, D. (2022). How environmental taxes and carbon emissions are related in the G7 economies? *Renewable Energy*, **187**, pp. 645-656.

Feng, S., Chong, Y., Yu, H., Ye, X., Li, G. (2022). Digital financial development and ecological footprint: Evidence from green-biased technology innovation and environmental inclusion. *Journal of Cleaner Production*, **380**, 135069.

IPCC (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* IPCC, Geneva, Switzerland.

Javed, A., Rapposelli, A., Khan, F., Javed, A. (2023). The impact of green technology innovation, environmental taxes, and renewable energy consumption on ecological footprint in Italy: Fresh evidence from novel dynamic ARDL simulations. *Technological Forecasting and Social Change*, **191**, 122534.

Liu, C., Ni, C., Sharma, P., Jain, V., Chawla, C., Shabbir, M.S., Tabash, M.I. (2022), Does green environmental innovation really matter for carbon-free economy? Nexus among green technological innovation, green international trade, and green power generation. *Environmental Science and Pollution Research*, **29**(45), pp. 67504-67512.

OECD (2019). Environment at a glance: Climate change, in *Environment at a Glance: Indicators*. OECD Publ., Paris, https://doi.org/10.1787/ac4b8b89-en

Pesaran, M.H. (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics*, **22**(2), pp. 265-312.

Pesaran, M.H. (2015). Testing weak cross-sectional dependence in large panels. *Econometric Reviews*, **34**(6-10), pp. 1089-1117.

Razzaq, A., Fatima, T., Murshed, M. (2023). Asymmetric effects of tourism development and green innovation on economic growth and carbon emissions in Top 10 GDP countries. *Journal of Environmental Planning and Management*, **66**(3), pp. 471-500.

Westerlund, J. (2007). Testing for error correction in panel data. *Oxford Bulletin of Economics and Statistics*, **69**, pp. 709-748.

Zhang, Q., Shah, S.A.R., Yang, L. (2022). Modeling the effect of disaggregated renewable energies on ecological footprint in E5 economies: Do economic growth and R&D matter? *Applied Energy*, **310**, 118522.

doi.org/10.26398/asaproc.0045

The organizational well-being inside Italian academic institutions measured through the ANAC questionnaire

Sara Leoni[a], Bruno Bertaccini[b], Alessandra Petrucci[b]
[a] School of Psychology, University of Florence, Italy.
[b] "G. Parenti" Department of Statistics, Computer Science, Applications, University of Florence, Italy.

1. Introduction

The evaluation of organizational well-being has gained increasing importance in management policies in recent years, as the elements characterizing this construct influence fundamental aspects of work life within a company. Examples of such aspects include motivation, a sense of belonging to the organization, and the sharing of values and proposed objectives (Avallone and Paplomatas, 2005; Dolan, 2006). Companies have therefore begun to invest time and resources in analyzing this construct, making it a fundamental objective of their interventions. At the national level, this interest is expressed through Legislative Decree 81/2008 (known as the Consolidated Act on the protection of health and safety in the workplace), which introduces the obligation to assess work-related stress risks in the context of occupational health. For this purpose, organizational well-being has been assessed through various measurement tools, among which the most commonly used are psycho-attitudinal tests, professional interest questionnaires, and personality questionnaires.

In the specific case of public administrations (see Avallone and Bonaretti, 2003 for specific insights on the topic), Legislative Decree 150/2009 has assigned internal evaluation bodies (OIV) the task of conducting surveys on organizational well-being[1]. The National Anti-Corruption Authority (ANAC) fits into this context by introducing a questionnaire for assessing organizational well-being in public administrations, focusing on the dimensions of organizational well-being, the degree of alignment with the evaluation system, and the assessment of one's immediate superior[2].

This article presents a study aimed at validating this questionnaire by administering it to the administrative and technical staff of the University of Florence. The article is structured as follows: Section 2 provides a detailed description of the ANAC questionnaire, followed by Section 3, which outlines the statistical model employed for data analysis. In Section 4, the obtained results are illustrated. Some concluding remarks complete the work.

2. The ANAC questionnaire

The questionnaire consists of 85 items and encompasses three distinct macro-dimensions: the well-being (WB), the degree of alignment with the evaluation system (EV), and the assessment of one's immediate supervisor (SU). The analysis of organizational well-being aims to emphasize and analyze employees' perceptions of the standards imposed by organizations and any variations

[1] For further information, see https://www.miur.gov.it/benessere-organizzativo

[2] National Anti-Corruption Authority (ANAC) - Independent Commission for the Evaluation, Transparency and Integrity of Public Administrations, *Modelli per la realizzazione di indagini sul personale dipendente volte a rilevare il livello di benessere organizzativo e il grado di condivisione del sistema di misurazione nonché la rilevazione della valutazione del proprio superiore* gerarchico, in application of the art. 14, paragraph 5, of Legislative Decree no. 150 of 27 October 2009 https://www.anticorruzione.it/portal/rest/jcr/repository/collaboration/Digital%20Assets/anacdocs/Attivita/ArchivioStorico/BenessereOrganizzativo/ModelliDiIndagine/Documento_Benessere_Organizzativo2.docx

of these standards in relation to the desired condition. The degree of alignment with the evaluation system, on the other hand, aims to identify personnel agreement with the evaluation procedures mandated by the organization. Finally, the assessment of the immediate supervisor allows for establishing employees' approval regarding personnel management and the performance improvement implemented by their superiors.

Thus, the 85 items are distributed as follows: the first dimension (WB) is divided into eight sub-areas of investigation: the work environment, discrimination, equity in my organization, career and professional development, my job, my colleagues, the context of my work, and the sense of belonging. The second dimension (EV) is divided into three sub-areas of investigation: work performance, my performance, and system functioning. Meanwhile, the third dimension (SU) is divided into three sub-areas of investigation: my supervisor and my growth, my supervisor and equity, my supervisor and the evaluation system. For each item, the participant must indicate, using a 6-point Likert scale, whether and to what extent they agree (from "strongly disagree" to "strongly agree") with the statement provided.

The last section of the questionnaire is devoted to collect some supplementary demographic and occupational information for each employee, such as gender, contract type, age range, years of service, and qualification.

3. The statistical model

The validation process for the ANAC questionnaire utilized structural equation models (SEMs; Bollen, 1989) due to its psychometric structure. Specifically, we conducted a hierarchical (3-level) confirmatory factor analysis (CFA, see Jöreskog, 1969). Through this model, the questionnaire items serve as indicators for 14 first-level latent variables: A - work environment; B – discrimination; C - equity in my organization; D - career and professional development; E - my job; F - my colleagues; G - context of my work; H - sense of belonging; L - organizational performance; M - my performance; N - system functioning; O - my supervisor and my growth; P - my supervisor and equity; Q - my supervisor and the evaluation system. The variability of these first-level factors is thus explained by three second-level latent factors (organizational well-being WB, level of familiarity with the evaluation system EV, assessment of the hierarchical superior SU), which, in turn, are explained by the overarching latent construct of organizational well-being (OWB).

The CFA model, as outlined by the sections the ANAC questionnaire, is depicted in the Figure 1.

The aforementioned model was estimated, using the maximum likelihood approach, on data collected from a survey administered to the administrative and technical staff of the University of Florence in April 2022. Conventional fitting measures, including the Tucker and Lewis index (TLI), the comparative fit index (CFI), and the root mean square error of approximation index, (RMSEA), were employed to assess the goodness of fit of the model (Steiger, 1990). As commonly known, RMSEA values $< .06$ and CFI/TLI values $> .90$ generally indicate a relatively good fit to the data (Xia and Yang, 2018). Modification indices were also used to evaluate slight adjustments in the model structure (Hu and Bentler, 1999).

All the estimates were derived using the R package *lavaan*, version 0.6-12 (Rosseel, 2012).

4. Results

The Florentine University chose a non-mandatory administration of the ANAC questionnaire. Nevertheless, in many cases, the distribution of respondents, observed for the socio-demographic and occupational variables collected (taken individually), can be considered a good approximation of the one characterizing of the target population.

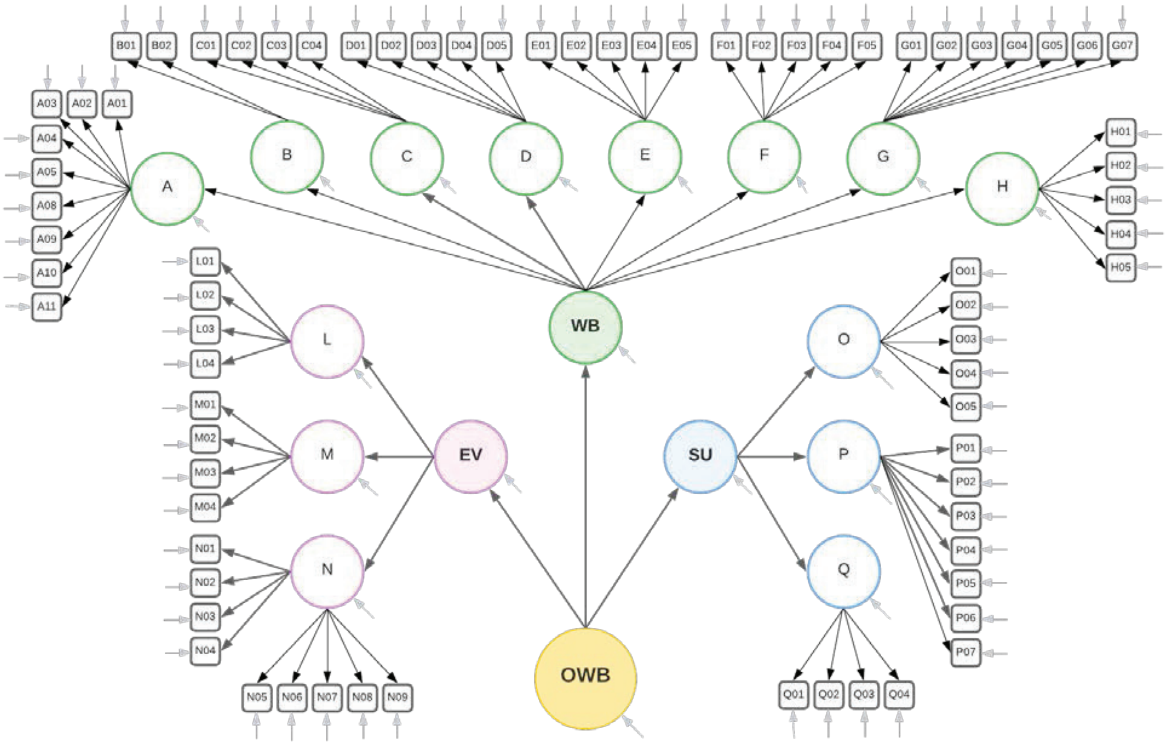


Figure 1: Latent constructs underlying the ANAC questionnaire

The survey participants amounted to 781 individuals (approximately half of the target population), of whom 545 were female. Among them, 424 worked in the central administration, while 357 were assigned to departments. Regarding age, 8 respondents were under 30 years old (the 1% of the respondents), 140 (17.9%) were aged between 31 and 40 years, 203 (26%) were between 41 and 50 years old, 276 (35.3%) were between 51 and 60 years old, and 154 (19.7%) were over 60 years old. Additionally, 703 employees belonged to job categories B, C and D (excluding functional unit responsibilities), while 78 are managers or fell under job categories EP or category D, responsible for functional units. Lastly, concerning tenure, 65 employees were in their first year of service at the university, 174 had between 2 and 5 years of service, 70 had 6 to 10 years, 132 had 11 to 20 years, and 340 had more than 20 years of service.

The direct application of the aforementioned model to the data collected yielded unsatisfactory fitting measures compared to the threshold values that such indices should assume in case of good fit (first row of Table 1).

Table 1: Goodness of fit of the estimated models (CFI, TLI and RMSEA)

Estimated models	CFI	TLI	RMSEA
ANAC model estimated on collected data	0.8382	0.8327	0.0704
Modified ANAC model (correlations between items added)	0.8866	0.8821	0.0614
Modified ANAC model (with some items correlations added, applied to a limited subset of respondents)	0.9123	0.9089	0.0502

This result is essentially attributable to the following two elements: the formulation of the questionnaire, which might present unclear items (or even duplicates according to the respondent's

interpretation), and the respondent's behavior (considering that the University has established a mandatory response for all items once the survey has started).

Through the analysis of modification indices, it was possible to highlight the presence of significant correlations among certain items that, if estimated, contribute to an overall improvement in the overall fit. The fitting measures derived from the model estimation incorporating covariances among certain items show an undoubtedly improved fit, although it remains somewhat unsatisfactory (second row of Table 1).

At this point, the presence of forms filled out in a less attentive (or even random) manner was examined, which could clearly have a negative impact on the overall fit of the model. To ascertain the presence of such forms and measure their effects on the model's fit, a dedicated algorithm (in R language) was developed. This algorithm was designed to arrange the questionnaires based on their proximity to the hypothesized model. The algorithm progresses through these steps:

1) initially, 10,000 samples are evaluated from all collected questionnaires (sampling without replacement), with a size of 400 (deemed as the minimum number sufficient for estimating all model parameters);
2) the sample with the highest fitting indices is selected (giving priority to CFI);
3) among all remaining observations, a forward logic is followed, where at each step, the observation that provides the maximum increase in CFI compared to the previous step is included.

Upon completion, the algorithm presented, in ascending order, the list of observations based on their proximity to the hypothesized model. Thus, through the algorithm, it was possible to understand that the point of maximum fit (in terms of CFI index) could be achieved with about the first 550 observations (Figure 2).

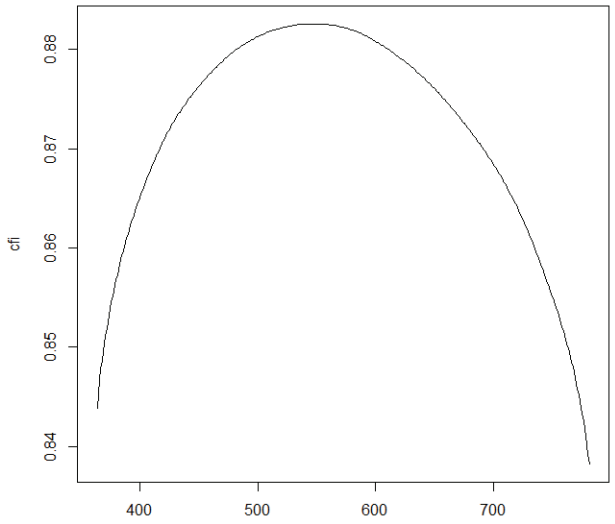


Figure 2: Sequence of the CFI fitting values measured at each step of the algorithm progress

Figure 3 displays the record layout (first part) of the last forms identified by the algorithm as carelessly filled or, in some cases, even randomly. Considering only the first 550 “ordered” observations and retaining some covariances due to the imperfect formulation of certain items, the final model estimation was carried out. The fit indices obtained from this final estimation suggest a favourable model fit to the data (last row of Table 1).

A01	A02	A03	A04	A05	A08	A09	A10	A11	B01	B02	C01	C02	C03	C04	D01	D02	D03	D04	D05	E01
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
6	6	6	6	6	1	1	1	1	6	6	6	6	6	6	6	6	6	5	6	6
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
5	4	4	4	4	5	5	4	4	5	5	4	4	4	4	5	5	5	4	4	4
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
5	5	5	5	5	5	5	5	5	5	4	4	3	4	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	3	3	3	3	3	4

Figure 3: Record layout (first part) of the forms that produce a decreasing of the CFI values at the end of the algorithm

Furthermore, using the functionalities of the *lavaan* package, it was feasible to reconstruct the estimated latent scores for each respondent in relation to the second-level constructs (WB, EV, SU) and third-level construct (OWB). Figure 4 depicts the distribution of these scores (normalized with respect to the mean). It is noteworthy to observe how the distribution of scores related to the overall well-being is essentially analogous to that of the WB construct.

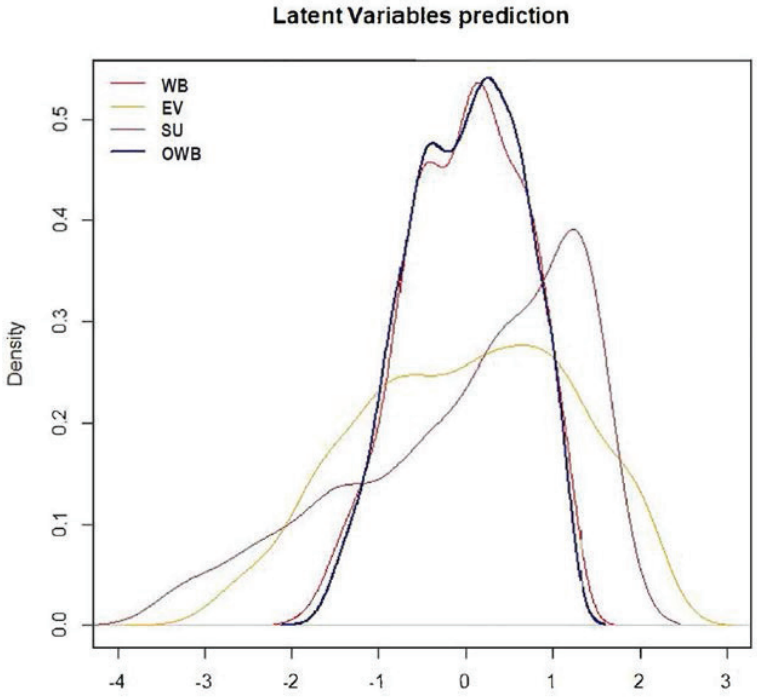


Figure 4: Scores distributions of second (WB, EV, SU) and third (OWB) levels constructs (normalized with respect to the mean)

5. Conclusions

This study was aimed to measure the perceived level of organizational well-being within a university institution. The assessment was conducted by applying a 3-levels confirmatory factor analysis to the data collected in April 2022 through the administration of the questionnaire model proposed by the National Anti-Corruption Authority to the administrative and technical staff of the University of Florence.

The fitting of the CFA model, which clearly reflects the structure of the questionnaire, did not prove to be satisfactory with the data collected through its administration. This is due to both the

presence of correlations among the proposed measurement items and to a significant number of "anomalous" filled forms, i.e., forms containing inconsistent or random responses (deviating from the hypothesized model).

These issues underscore the risks associated with the design of surveys of this kind. Firstly, there is the potential for participants, when faced with various propositions of similar concepts (or those perceived as such) within questionnaires comprising a high number of items, to respond in a less attentive, hurried, or even random manner. Each item should solely capture the dimension to which it was assigned in a manner that adequately describes the latent construct for which it was designed, avoiding the risk of undermining the questionnaire's validity. In other words, correlations between items (especially across different sections of the questionnaire) should be few and clearly justifiable during the analysis.

Regarding the presence of "anomalous" completions, reference can be made to the phenomenon of “social desirability”, which is the tendency for participants to respond in a socially acceptable rather than truthful manner. In a work context, this tendency might be encouraged by the hierarchical organization, which in some cases can lead to the fear of being judged and potentially sanctioned by the superior. Another factor that can be considered is workers' scepticism toward the organizational context. In this case, employees display a lack of confidence in the possibility of improving their level of work well-being, thereby refraining from helping the company identify any issues. These two explanations fit into the same framework where, on one hand, employees do not fully grasp the necessity of measuring organizational well-being, and on the other hand, they struggle to provide an honest assessment due to the fear of potential subsequent repercussions.

In addition to the aforementioned considerations, it can be noted that the administration mode imposed by the University of Florence could also be responsible for random or inattentive completions. Specifically, allowing respondents the freedom to participate or not in the administration leads to self-selection within the population, impacting the representativeness of the sample by selecting only those who are interested in completing the questionnaire or feel morally obliged due to the phenomena explained earlier.

References

Avallone F., Bonaretti M. (2003). *Benessere organizzativo. Per migliorare la qualità del lavoro nelle amministrazioni pubbliche*. Rubbettino Editore.

Avallone F., Paplomatas A. (2005). *Salute Organizzativa. Psicologia del benessere nei contesti lavorativi*. Raffaello Cortina Editore.

Dolan, S. L. (2006). *Stress, Self-Esteem, Health and Work*. Palgrave Macmillan, London.

Bollen, K. A. (1989). *Structural Equation with Latent Variables*. Wiley, New York.

Hu, L., Bentler, P. (1999). Cutoff criteria for fit index in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, **6**, pp. 1-55.

Jöreskog, K.G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika,* **34**, pp. 183–202.

Rosseel Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, **48**(2), pp. 1-36.

Steiger, J.H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research,* **25**, pp. 173–180.

Xia, Y., Yang, Y. (2018). RMSEA, CFI, and TLI in structural equation modelling with ordered categorical data: The story they tell depends on the estimation methods. *Behavioral Research Methods*, **51**, pp. 409-428.

doi.org/10.26398/asaproc.0046

Economic consequences of COVID-19 and war in the utilities industry: How price increase affects customers

Gabriele Liberati[a], Martina Corsini[a]
[a] CFI Group, Milan, Italy.

1. Introduction

In the last three years, players in the utility market have had to deal with an increasingly complex context due to unprecedented scenarios, facing two international crises within a relatively short timeframe: the COVID-19 health emergency first and then the energy crisis due to the conflict in Ukraine.

In this context, the attention and consciousness of consumers towards contracts and tariffs is strongly increasing; thus, making it necessary to pay attention to the trust between companies and users. In fact, this paper will propose that there is a causal link between the loss of trust in companies and the decline in satisfaction, with a decrease in ratings especially for that segment of the population that attributes blame for the increase in prices on electricity suppliers.

Exogenous variables, such as the energy price and the increase in raw material costs, affects the evaluation of service offered. Therefore, surveying consumer sentiment requires a flexible model that accurately measures the individual components of satisfaction and their impact. Using CFI modelling, we can verify the level of user satisfaction over time and the weight that each component has on it; thus, predicting future behaviours through the creation of indicators that measure trust and word of mouth.

As we will demonstrate, customers are the most precious asset to companies, especially for the utility market, in which there is little leeway on product offered and it is essential to bet on caring services.

1.1. Covid-19 health emergency and lockdown

In 2020, lockdowns and the consequent social distancing caused a slowdown, if not a halt, in production activities. Consequently, national energy demand experienced a contraction in the first seven months of 2020, adjusted by 8.4% compared to the corresponding period in 2019 (Terna, 2020), a phenomenon only partially offset by residential demand. Data indeed show a compensatory effect from residential demand due to people spending more time at home, accounting for 20% of the total (Mediobanca, 2020).

1.2. Conflict in Ukraine and energy crisis

While 2020 was characterized by a collapse in electricity prices, the conflict in Ukraine in 2022 reversed this trend. From the mid-2021 onwards, energy prices exhibited unprecedented volatility, with significant spikes. Starting from the second quarter of 2022, consumers and businesses were exposed to the effects of the energy crisis, facing substantial price hikes. This phenomenon reached its peak in August of the same year, marking a new historical record in the wholesale energy market price index (Gestore dei Mercati Energetici, 2022).

2. Study: sample, models and methodology

The Utility Customer Satisfaction Index (UCSI) survey conducted by CFI Group stems from the need to monitor satisfaction within the Italian utilities market during a particularly delicate

period for both companies and individuals. The 2022 edition of the study involved a sample of 1611 customers of which 1002 residential customers and 609 business customers, using various energy providers distributed throughout the national territory. Interviews were conducted between September 28th and October 14th, 2022, with respondents from both the deregulated market and the regulated market.

The UCSI survey of September 2022 revealed that more than 67% of respondents declared their spending on gas and energy to be increasing, with a peak for business clients of more than 70%. In addition to the objective price increase, the great media clamor affected the bad perception of customers with respect to energy crisis. Initially, the increases have been contained for whom with a fixed-rate contract stipulated when the costs of energy were lower. In fact, in September 2022, 48% of respondents had fixed-rate tariff. However, in 2022 the opportunity to sign these kind of contracts decreased due to unilateral choice of energy suppliers to replace fixed-rate contracts with flexible-rate contracts.

The UCSI multi-utilities survey in September 2023 revealed that 37% of respondents with fixed-rate contracts had experienced a tariff renewal in the past year. The research also found that in September 2022, 64% of utilities sector customers had searched for electricity and gas service offers online in the last 6 months. This percentage rose to 74% for business customers in the deregulated market. The scenario is complex, characterized by high mobility (the average customer tenure in the market is 2.2 years according to CFI Group) and unpredictable evolutions. Therefore, an analytical model that evaluates the significance of customer care is necessary in a sector like utilities, governed by supranational power dynamics with seemingly limited maneuvering room.

2.1. Implications on satisfaction: American Customer Satisfaction Index (ACSI) methodology

The relationship between customers and suppliers forms the basis of trust and loyalty. This relationship, along with the resulting behaviours, is reliably expressed through the customer satisfaction index, which is one of the best predictive indicators of market trends and company performance.

The proposed analysis is based on the ACSI methodology, which has been measuring consumer satisfaction on a national scale since 1994. It covers 46 industries grouped into 10 sectors, encompassing over 400 companies and government agencies, accounting for about 60% of the US gross domestic product. ACSI is the only unified measurement system for perceived product and service quality satisfaction across all sectors of activity in the United States. The distinctive feature of the ACSI methodology is its patented analysis system, capable of explaining the cause-and-effect relationship between perceived quality and customer/user/citizen satisfaction.

The underlying technology of ACSI identifies key drivers that influence satisfaction and calculates their impact on loyalty and financial performance (ACSI, 2023).

For customer satisfaction analyses, CFI Group employs the Partial Least Squares (PLS) approach, based on the NIPALS algorithm developed by Claes Fornell in 1994. This method aims to minimize the likelihood of statistical errors and enhance measurement accuracy. This is achieved through the creation of latent variables that synthesize the shared variance of multiple individual indicators, all related to the same phenomenon, such as customer ratings on expectations, quality, value, satisfaction, and future behaviours.

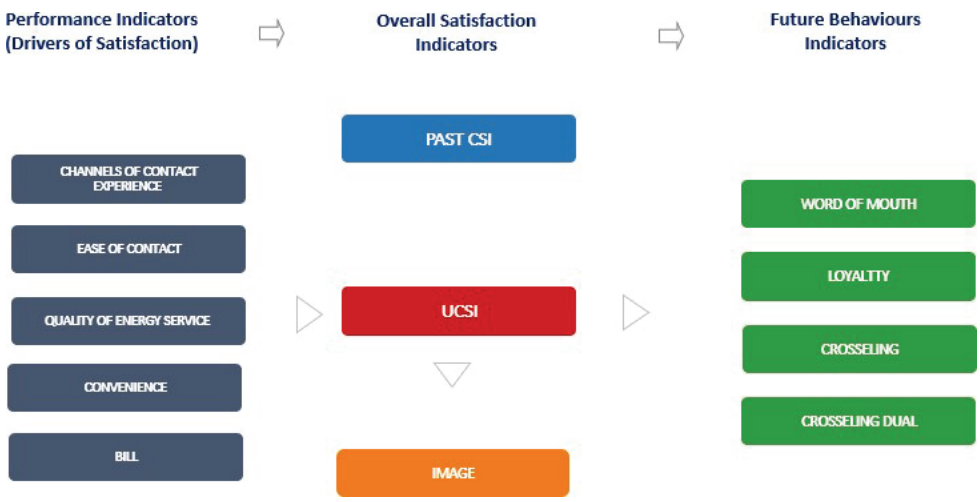


Figure 1: UCSI multi-utilities survey CFI 2022: the model

3. Survey results

The survey revealed that despite the complexity of the situation and the significant media attention to price hikes, the satisfaction index for consumers surveyed in September 2022 stood at 72 points. This value aligns with international ACSI findings concerning utilities sector operators in the US market and serves as the reference threshold for sector operators' satisfaction.

Table 1: Performance indicators from UCSI multi-utilities survey CFI 2022

	Score	St dev	Valid N	Direct impact
UCSI	72,0	22,6	1608	
Quality of services offered	74,6	20,6	1587	1,20
Satisfaction of channels for interacting with supplier	76,6	21,8	1522	0,96
Ease of contact with supplier	75,5	21,6	1505	0,07
Evaluation of bill	73,9	23,9	1592	0,58
Economic convenience of services offered	72,9	25,1	1577	2,22

Table 2: Future behaviours from UCSI multi-utilities survey CFI 2022

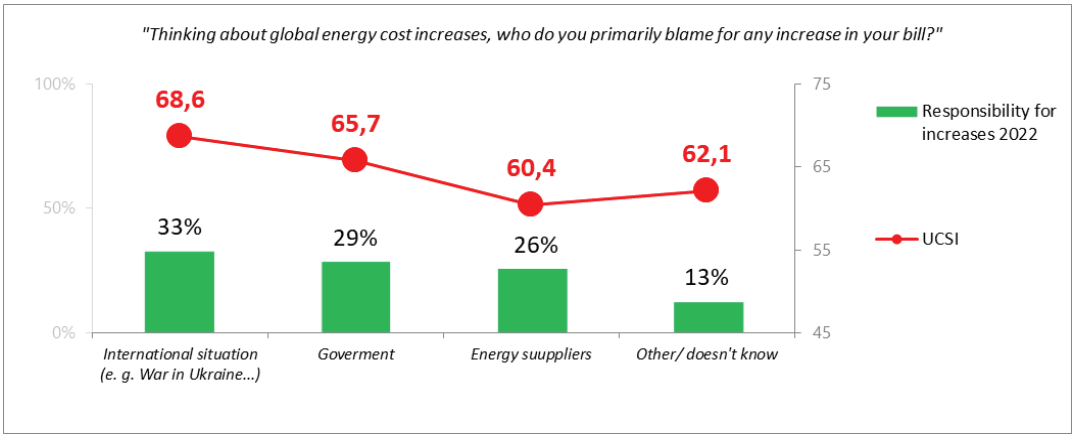
	Score	St dev	Valid N	Direct impact
Image	72,5	24,2	1597	4,53
Loyalty	75,4	27,3	1556	4,37
Word of mouth	72,5	25,1	1556	4,92
Cross selling	58,5	29,4	1482	3,40
Cross selling dual	47,6	29,0	438	3,48

The global satisfaction index for the market was then compared among various companies operating in the Italian sector. Out of the ten major operators examined, it emerged that despite the previously listed challenges, only 2 out of 10 operators achieved results below the sector's average. At first glance, this might suggest that these increases didn't have a direct impact on the overall satisfaction levels for their provider. However, delving into the data reveals a different reality: During the most acute phases of the 2022 energy crisis, the debate surrounding the "responsibilities for the crisis" contributed to generating mistrust and confusion among many users.

For further exploration of this topic, a portion of the sample (those who rated economic

convenience negatively) was asked to whom they would attribute the main responsibilities for the increases in energy prices. The majority of respondents (33%) attributed the blame for the price hikes to the complex international situation, 29% blamed the government and its inability to implement effective countermeasures, while only 26% identified their own energy provider as directly responsible for the increases.

Figure 2: Data analysis from UCSI multi-utilities survey CFI 2022: responsibility for price increases



Excluding those who provided other responses or did not express an opinion (13%), less than 3 out of 10 respondents considered their own energy provider responsible for the price hikes. Analysing the satisfaction index across the various clusters (government responsibility, international situation, own provider), a clear difference in satisfaction levels among the three examined clusters can be noticed.

Those who attributed the responsibility for the increases to external factors (such as the international situation) rated their energy provider with a significantly higher score (69), compared to those who believed that the provider had a direct responsibility for the energy price increase (60).

4. Concluding remarks

The 2022 energy crisis led to an increase in energy prices, causing higher costs for families and businesses and generating dissatisfaction among consumers who faced higher bills and difficulties in their daily expenses. At the same time, utility companies had to deal with operational challenges due to the crisis, such as ensuring energy supply in a more difficult supply context and managing an increase in customer assistance demands. We have seen that the satisfaction index for the utilities sector in 2022 had not yet been affected by the effects of the crisis. However, the long-term impact of the billing increases could resurface, questioning the positive results achieved so far. For companies in the sector, it is necessary to improve transparency and communication with customers regarding the challenges the industry is facing and the actions being taken. Without proper attention to communication and image, the portion of customers attributing the responsibility for the increases to the current provider could increase, dragging down overall satisfaction.

References

American Customer Satisfaction Index (2023). *ACSI Energy Utilities Study 2022-2023* (https://www.cfigroup.it/wp-content/uploads/2023/04/23mar_UTILITIES-Report.pdf)

CFI (2023). ACSI in breve (https://www.cfigroup.it/approfondimenti/acsi2021/)
Gestore dei Mercati Energetici (2022). Dati di sintesi MPE-MGP-riepilogo (https://www.mercatoelettrico.org/it/)
ISTAT (2020). *Rapporto annuale 2020 – La situazione del Paese.* Roma, pp. 22-30.
Mediobanca (2020). Briefing note: Impatto della pandemia COVID-19 sui dati del 1Q 2020 del FTSE-MIB (industria e servizi). *Economic Bullettin,* **3***,* p. 8.
Monaco G., Ferraro B., Marotta F. (2017). Utilizzo di modelli causali per la Metropolitana di Roma (https://www.cfigroup.it/wp-content/uploads/2017/12/2009Brescia_Universit%C3%A0deglistudi.pdf)
Terna (2020). Rapporto mensile sul sistema elettrico – Luglio 2020, p 20.

doi.org/10.26398/asaproc.0047

Inflation effects on the Italian household living conditions at the time of SARS-Cov2

Caterina Marini[a], Vittorio Nicolardi[a]
[a] Department of Economics and Finance, Aldo Moro University of Bari, Bari, Italy.

1. Introduction

Inflation is the change in the general level of prices of goods and services. Its effects on the living conditions of households and the production potentialities of firms are worldwide recognised as an important issue. Policymakers pay a great attention to inflation because of its direct consequences on the economy. Furthermore, the various international circumstances that may provoke general economic instability are matter for the analysis of the reasons behind the inflation trend and magnitude. For these reasons and many others, inflation is a crucial part of the political agenda of all countries in the world and its control is one of the most important objectives of the monetary policies.

In this paper, we focus our analysis on how the cost of living of the Italian households was affected by the inflation trend over a period of seven years starting from 2015, encompassing the SARS-Cov2 period. Therefore, the effects of the consumer price index (CPI, hereafter) trend on the Italian household living conditions across the income and consumption distributions are under investigation. In particular, we categorise Italian households according to size and expenditure and construct the inflation measure for household expenditure and income groups. Our approach is new in the Italian academic literature and highlights results comparable to the official CPI statistics and easily interpretable in the context of the inflation economic analysis.

The international scientific literature relating to the inflation experience of households is vast, and the Italian references are also valuable of attention. Baldini (2005) studied the distribution of inflation rate across Italian households over the period 1986-2004, which embraced also the year 2002 of the monetary changeover to euro, and highlighted that for many households the real living standard progressively worsened.

More recently, the attention is focused on the CPI capacity to effectively grasp the effects that inflation produces on the cost of living of households considering that the latter experience expenditure choices that are differentiated mainly based on their income and their propensity to maintain unchanged their own well-being. As Kaplan and Schulhofer-Wohl (2017) for the United States and Jacobs et al. (2014) for Australia underlined, we need to investigate how inflation behaves at the household level to understand how inflation affects household living conditions, and the CPI as yielded by the national institutes of statistics is showing many faults in this sense. The first, and maybe the most important for our analysis, is the aggregate measure of inflation, referred to a fixed basket of goods and services with unchanged composition and quality among various groups of households and the average prices for the same, that the CPI represents.

The heterogeneity in inflation rate among households is hard to be understood where different approaches are not investigated to reflect the real household concerns about the cost-of-living inflation. Kints and Breunig (2021) examine the consumer price inflation experience of Australian households across the income distribution and achieve relevant results in terms of analysis, suggesting an approach that is worthy of attention for other country specific empirical works.

2. Data and method

For our analyses we used three main data sources: household expenditure and income by household size and quintile, and the prices of goods and services purchased by households. The

data source is the Italian household budget survey (HBS, hereafter) that refers to the final consumption expenditure of the Italian households. Expenditure according to the classification of individual consumption by purpose (COICOP) and household size are provided by HBS. The European Union Survey on Income and Living Conditions (EU-SILC) provides data on income per household size and the distribution of households by income quintiles. Both sources partitions data also according to social, economic and territorial characteristics of households. The CPI data (base-year 2015) are provided by the consumer price survey (CPS, hereafter) and they are also yielded per COICOP and territorial characteristics.

Data we used are per year and referred to the period 2015 – 2021. Therefore, the period under investigation includes the SARS-Cov2 lockdowns and the first period of the energetic market crisis, and in our analysis the consequences of their effects on the Italian household living conditions are heavy and unquestionable.

At this stage of the work, we used only official data as provided by the online ISTAT data warehouse. Therefore, we could not afford some data that are necessary to delineate the complete analysis of the relationship between household income and inflation. In fact, we use the full and detailed information on expenditure by household size and quintile, but we defined a roundabout way method to connect the household expenditure and CPI trend with household income.

The cornerstone of our method consists of yielding the household expenditure time series deflated by applying the CPI by COICOP to determine in real terms expenditure by household size and quintiles. Afterwards, an implicit CPI (I_CPI, hereafter) was calculated to define the change in the household purchasing power over the period analysed. We computed the I_CPI per household size and quintiles based on the ratio of household expenditure data at current prices and those as deflated above. The I_CPI highlights the inflation effects disaggregated per COICOP 2-digits divisions by household size and quintiles and, therefore, describes the consumption composition at the household level based on size and the belonging expenditure quintile when data are appropriately arranged in tables or graphs. In our analysis, we consider the aggregated I_CPI per household size and expenditure quintiles (Table 1)

Table 1. I_CPI per household size and expenditure quantiles

	HSize					HExpenditure quantile				
Year	1	2	3	4	5 +	Q1	Q2	Q3	Q4	Q5
2015	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2016	0.994	0.994	0.995	0.995	0.995	0.993	0.994	0.994	0.995	0.995
2017	1.008	1.009	1.009	1.010	1.009	1.008	1.009	1.009	1.009	1.009
2018	1.025	1.025	1.024	1.023	1.023	1.024	1.024	1.024	1.024	1.024
2019	1.033	1.033	1.031	1.030	1.029	1.032	1.031	1.031	1.031	1.032
2020	1.019	1.021	1.021	1.021	1.022	1.019	1.020	1.020	1.021	1.021
2021	1.058	1.056	1.053	1.051	1.051	1.056	1.055	1.054	1.054	1.055

Finally, the aggregate I_CPI was applied to the income distribution by household size, obtaining the equivalent household income (EHI, hereafter) in real terms (Table 2), to determine exclusively the household expenditure potentialities based on income and the changes of purchasing power over time. At this stage, we could not replicate the analysis in terms of household income quintiles because of the unavailability of the public data necessary for this step.

Based on the data we computed, the distribution of inflation across households highlights a differentiated effect based on the household size, expenditure quintile and EHI, as expected, but the outcome is surprisingly original in the analysis of inflation over the period considered.

Table 2. Equivalent household income and changes in household purchasing power (per cent values), per household size

	Equivalent household income					Changes in household purchasing power				
	HSize					HSize				
Year	1	2	3	4	5 +	1	2	3	4	5 +
2015	21977	35631	42883	45750	46526	-	-	-	-	-
2016	22488	36094	43546	46295	45926	0.64	0.62	0.52	0.49	0.47
2017	23397	37739	44588	46702	49049	-0.79	-0.87	-0.92	-0.95	-0.94
2018	23114	37764	44611	48053	51658	-2.39	-2.44	-2.36	-2.25	-2.26
2019	23761	39663	48039	50188	53489	-3.18	-3.15	-3.04	-2.93	-2.84
2020	23554	39264	47672	50098	51138	-1.82	-2.05	-2.05	-2.02	-2.10
2021	24713	40761	49564	53122	59682	-5.45	-5.33	-5.06	-4.84	-4.81

In Italy, over the period 2015–2021, the inflation trend was almost steady until 2020, moderately swinging with fairly narrow values aligned with the inflation rates tolerated by the European monetary policy. It is starting from 2021 that the inflation rate sharply increased, suggesting a change that was occurring in the general level of prices, confirmed in 2022, whose reasons are mainly consequences of the international contexts. In fact, we have to bear in mind that 2020 and 2021 are the two years of the SARS-Cov2 lockdowns, which paralysed the global economy, and the beginning of the Ukraine-Russia crisis, which provoked uncertainty about the international markets.

Considering that the inflation rates differ across households because of their different consumption habits based on their size and income, the analysis of the I_CPI highlights that the Italian households in the top expenditure quintile and the households of single person or 2 people, in terms of expenditure and income, experienced the highest inflation rate. In particular, as shown in Table 1, the large families of 4 or 5 and more people experienced less advantages from the negative inflation rate at the beginning of the period considered (2016) comparing with the small households that enjoyed an inflation rate slightly more advantageous. In fact, expenditure in real terms for the first increased less than for the second.

Contrariwise, towards the end of the period the situation reversed, and the large families were more protected by the inflation rate than the little, alongside the analytical evaluation of expenditure as described before. However, it is important to notice that in 2020 the largest households suffered more in terms of inflation then all the others. If we consider that the pandemic lockdowns caused a sharply increase of prices for some goods and services more than others, the disaggregated analysis for COICOP highlights that effectively the large families spent more in terms of money for some goods, such as shoes and cloths, and food and non-alcoholic beverages than the other typologies of households. The analysis of EHI per household size is strictly connected with the analysis of expenditure in broad terms, as seen before. In fact, as expected, Table 2 shows how families of 4 or 5 and more people experienced inflation effects on their purchasing power more disadvantageous than the families of single person or 2 people at the beginning of the period analysed, while the situation reversed towards the end with the only exception of the year 2020, as expected.

Furthermore, the entire period is characterised by reductions in purchasing power, sometimes numerically relevant, with the only exception of 2016 that experienced a negative inflation rate and, therefore, a gain in purchasing power for all families. The latter was more remarkable for the smaller families. It is important to underline the different inflation effects on both expenditure and income that we worked out through the application of the general CPI, as published by ISTAT. Table 3 shows the differentials between the changes of purchasing power by household size computed with our I_CPI and the ISTAT aggregate CPI. Our analytical method grasps the real inflation effects on

EHI by size through the application of the I_CPI as described above, while the official CPI tends to misestimate the value. In particular, we demonstrate that the estimation error is higher for the small families than for the large ones. In 2016, for instance, all households experienced a higher gain of purchasing power (because of the negative inflation rate) than that estimated by the official CPI (0.10%) and, specifically, the families of 4 or 5 and more people experienced a narrow outcome (0.38%, on average) comparing with the result for the smaller families (0.53%, on average).

Table 3. Differentials between I_CPI and ISTAT CPI estimates of the changes in household purchasing power. Per cent values.

	HSize				
Year	1	2	3	4	5 +
2015	-	-	-	-	-
2016	0.54	0.52	0.41	0.39	0.37
2017	0.30	0.22	0.17	0.14	0.15
2018	-0.15	-0.20	-0.12	-0.01	-0.02
2019	-0.37	-0.34	-0.22	-0.11	-0.02
2020	0.83	0.60	0.59	0.63	0.54
2021	-1.01	-0.88	-0.60	-0.37	-0.33

Another important aspect is the inflation effects on households of three people, in terms of both expenditure and income: medium size families are those moderately affected by the inflation rate for the whole period considered.

The results are different if we analyse the distribution of households by expenditure quintiles. In fact, household expenditure patterns differ across all household consumption quintiles and the impact of inflation on households can significantly differ according to different COICOP classes. In particular, we discovered that inflation had a permanent negative effect for the whole period on households belonging to the top expenditure quintile. However, since 2018 also the lowest-consumption households started experiencing higher inflation effects that worsened at the end of the period under investigation. The disaggregated analysis by COICOP highlights that in 2021 the necessary goods and services were those more affected by inflation. In fact, inflation differences among the household expenditure quintiles were also driven by the composition of the consumption basket of households that differs across quintiles. For instance, *Housing, Water, Electricity, Gas and Other Fuels* and *Food and non-alcoholic beverages* represent the main components in the basket of the households in the bottom quintile and this is the reason why these households particularly suffered in 2021. The living condition gap increases as inflation increases. Therefore, an income increase for low-income households is frustrated by the inflation increase.

This paper has still a lot of work to investigate when more detailed and disaggregated data will be available.

References

Baldini, M. (2005). Inflation inequality in Italy. Centre for the Analysis of Public Policies (CAPP) 0008, Università di Modena e Reggio Emilia, Dipartimento di Economia "Marco Biagi".

Kaplan, G., Schulhofer-Wohl, S. (2017). Inflation at the household level. *Journal of Monetary Economics*, **91**, pp. 19-38.

Jacobs, D. et al. (2014). *Inflation and the Cost of Living*. Reserve Bank of Australia, Bulletin--March 2014.

Van Kints, M., Breunig, R. (2021). Inflation variability across Australian households: Implications for inequality and indexation policy. *Economic Record*, **97**, pp. 1-23.

doi.org/10.26398/asaproc.0048

The short-run employment effects of Covid-19

Andrea Ugo Marino
Istat (National Institute of Statistics of Italy), Genoa, Italy

1. Introduction

Beyond representing a dramatic global health crisis, the Covid-19 related pandemic has brought also about an unprecedented economic challenge. A growing literature is studying the complex dynamics and wide-ranging implications of this unprecedented shock. In particular, several papers have analysed its labour market effects.

The present note is meant to contribute to this literature by presenting some new evidence on Italy, one of the worst-hit countries by the pandemic, where the number of employed people declined by more than 3% in 2020. The employment rate decreased as well (from 44.8% in 2019 to 43.5%). The reduction in the participation rate was even stronger (from 49.8% to 48.0%), thereby determining a reduction of the unemployment rate (from 9.9% to 9.3%). In this study we focus on employment only.

Our starting point is a contribution by Cerqua and Letta (2022, henceforth CL), who analysed the short-run employment effects of the pandemic upon the 610 Italian local labour markets (LLMs) during the year 2020. CL use machine learning methodologies to obtain a counterfactual estimate of employment (i.e. the employment level we would have observed, had the pandemic not occurred). Deviations between the observed and estimated counterfactual levels measure the pandemic-related employment change.[1] Then, through a regression trees analysis, they investigate which factors are correlated to employment variations. Their results indicate that territorial differences in employment losses were mainly associated to pre-existing labour market fragilities. Surprisingly, CL find also that local differences in excess mortality, which they consider "*as a measure of the spread of Covid-19*", play no role in explaining differences in employment changes. Similarly, the present analysis focuses on LLMs in 2020.

However, in order to test the robustness of their findings, it departs from CL in the following way: 1) it relies on the Italian National Institute of Statistics' (Istat) labour force survey rather than on administrative data (the *Unioncamere Business Register*), which do not capture "informal" employment; 2) it uses a more conventional measure of excess mortality, computed by comparing deaths in 2020 to the average number of deaths in preceding years (whereas CL use estimates obtained by Cerqua et al., 2021, through machine learning techniques); 3) it uses standard panel data methods in counterfactual estimation; 4) it performs an association analysis through quantile regression, to analyse whether and how the relevance of different factors change across the distribution of the outcome variable. The next section deals with estimating the counterfactual employment level, Section 3 presents quantile estimates, Section 4 briefly summarizes.

2. Estimating the counterfactual 2020 employment level

Istat data provide us with the *observed* value of our variable of interest, the employment level in 2020 in each local labour market. In order to estimate the *counterfactual* employment level, we rely on panel data methods. The underlying idea is that the employment level of area i in the year t, $L(i, t)$, may be accurately predicted by a set of k covariates available in the period t-1, plus a set of time invariant regressors capturing regional and economic specialization effects, a year dummy

[1] CL provide a rich discussion and bibliography about the problems of counterfactual estimation in the context of the Covid-19 related economic shock. See also Jerez et al. (2023).

to catch period effects and finally an individual-specific term meant to tackle the "omitted variable bias" problem potentially arising when some factors influencing the outcome variables are unobserved or unobservable. The regression model may be written as follows:

$$(1)\quad L(i,t) = \alpha + \sum_{s=1}^{k} \beta_s Z(i,s,t-1) + \sum_{s=2}^{G} \gamma_s G(i,s) + \sum_{s=2}^{R} \lambda_s R(i,s) + \delta_t Year(t) + v(i) + u(i,t) \quad t = 2,3,..T$$

where $Z(i,s,t\text{-}1)$ is the value of a continuous covariate s at time t-1 in the area (or LLM) i, $G(i,s)$ is a dichotomous variable indicating whether area i is specialized in the economic "Group" s, $R(i,s)$ is another dummy variable indicating whether area i is located in Region s (regions are defined at the NUTS 2 level), $Year(t)$ is a period dummy, $v(i)$ represents the individual-specific effect, $u(i,t)$ is an idiosyncratic error term and T is the total number of periods. The set of the continuous covariates includes productivity (value added per worker), the size of the population aged 15 years and more, the number of local units, turnover (total sales at the LLM level) and the share of people employed in services. Economic reasoning suggests that labour demand should be higher in larger, richer and more productive areas but lower in markets with a higher degree of tertiarization. Measures of specialization come from an Istat classification associating each LLM to a "Group", based upon the sector distribution of workers and urbanization characteristics.[2]

Running the model for T-1 periods yields a vector of parameter estimates, which are used to form a prediction of the dependent variable at time T. Since the continuous covariates are available since 2015, only 4 periods are available to predict the counterfactual employment level in 2020. Two main estimators, the so called "fixed effects" (FE) and "random effects" (RE) estimators are commonly used to evaluate the parameters of a panel data specification like that in equation (1).[3] The criterion followed to decide which estimator is likely to yield better predictions in our context (and also a way to get a preliminary evaluation of the predictive ability of the model) has been estimating equation (1) on a shorter period, 2015-2018, thereby obtaining a prediction for the 2019 employment level to be compared to the actually observed value. The RE estimator has turned out to be preferable in terms of forecast accuracy due to a much lower mean square prediction error. By comparing the actual 2019 employment levels to the RE predictions, Figure 1 shows that the model does really exhibit a satisfactory predictive power.

Table 1 reports RE estimates of equation (1) over the years 2015-2019; these are used to form a prediction on the employment level in 2020 (for the sake of space, only coefficients of the continuous covariates are reported).[4] All estimated parameters have the expected sign and are statistically significant, the only exception being the coefficient of productivity (probably, due to a problem of collinearity).[5] These estimates allow us to compute the difference between the observed and predicted value of the employment level in 2020, Λ, which -like in CL- is interpreted as "*the causal impact of the Covid-19 shock*". Figure 2 provides a map of how the value Λ is distributed across the country. It may be seen that, similarly to what found in CL, the difference between the observed and predicted 2020 employment level is actually *positive* in a

[2] All the data and metadata concerning Italian LLMs provided by Istat and used here may be accessed at the website https://www.istat.it/it/informazioni-territoriali-e-cartografiche/sistemi-locali-del-lavoro. These are the groups: Not specialized and not urban, Highly specialized urban areas, Moderately specialized urban areas, Urban ports, Not specialized urban areas, Touristic vocation, Agricultural vocation, Textile and clothing, Leather, Machinery, Wood and furniture, AgriFood, Glasses, Transportation, Metals, Construction, Petrochemical and Pharmaceutical.

[3] The RE approach assumes the individual-specific effect to be a random term not correlated to the other covariates. Consequently, it is subsumed into the error term and parameters are estimated via generalized least squares (GLS), given a set of hypotheses concerning the covariance matrix of the composite error term. The FE approach allows for correlation between the individual effect and the other predictors. Consistent estimates come from an OLS transformed model, in which observations of all variables are replaced by their deviations from the respective individual means. See Baltagi (2005) for a presentation and discussion of the two approaches.

[4] The entire set of these and other mentioned -but not reported- results is available upon request.

[5] Indeed, productivity is highly correlated to total turnover (the correlation coefficient being higher than 0.7) and would be significant if dropping market size from the specification. I have decided to keep both variables in the model, since they represent distinct concepts. Anyway, dropping either variable would yield small changes in predictions.

minority of LLMs, suggesting that job retention policies adopted during the pandemic may have actually exerted an upward impact on labour demand in some areas.

Table 1. Predicting 2020 employment levels. Panel estimates on 2015-2019 data.

Dep. Variables	*Regressors*				
	Productivity $_{(t-1)}$	***Population 15 years and more*** $_{(t-1)}$	***Turnover*** $_{(t-1)}$	***Local units*** $_{(t-1)}$	***Share of employed in services*** $_{(t-1)}$
Employment level $_{(t)}$	.00931	.8945***	.0100766**	.0900983***	-.0516172**
	(.00811)	(.0234)	(.0048)	(.02173)	(.0239)

NB. All variables measured in logs. Random effects (RE) estimates. Heteroskedastic consistent standard errors reported in parentheses.
The set of the explanatory variables includes also (unreported) time, regional effects and a constant term. N. obs. = 2440; **: pval.<0.05; ***: pval.<0.01

Figure 1 Predicted versus actual employment levels in 2019

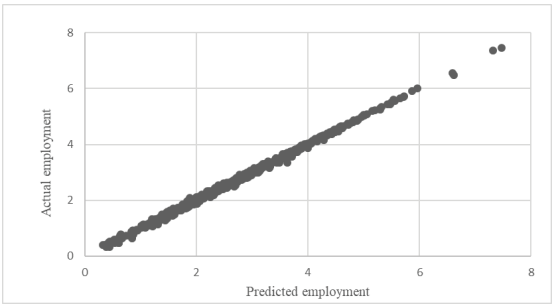
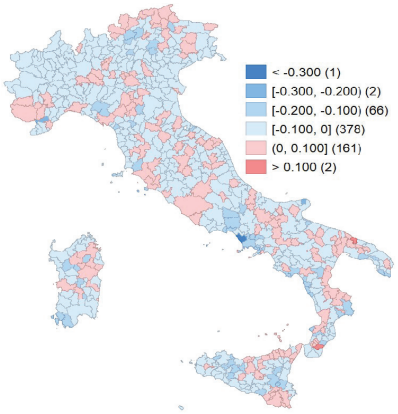


Figure 2. Differences between observed and counterfactual employment levels across Italian Local Labour Markets (LLMs) in 2020



3. Association analysis

In this section we investigate which socio-economic factors better predict differences in Λ, i.e. regional differences in employment changes. Both OLS and quantile regression methods have been used to ask such a question. For the sake of space, we report only estimates from (spatial)

quantile regressions, as one may suspect that the relevance of various factors in explaining differences in the employment effects of the pandemic significantly change across the distribution of Λ. Anyway, insights provided by OLS are broadly similar. The spatial quantile model estimated here is that proposed in Murakami and Seya (2019). By denoting with $q(\tau,\Lambda)$ the τ-th quantile of Λ, it may be written as follows:

$$(2)\quad q(\tau,\Lambda_i)=\alpha+\sum_{j=1}^{M}\beta_{j;\tau}x(i,j)+f\left(\tau,s(i)\right)+\varepsilon(i,\tau) \qquad \text{with} \quad \tau=(0.1;0.2;...;0.9)$$

where $x(i, j)$ is the value of an explanatory variable j in area i, $f(t, s(i))$ is a spatial process included to correct for possible positive spatial dependence and $\varepsilon(i, \tau)$ is the error term.[6]

For the sake of comparison, we include into the vector all the variables used in CL's association analysis. The only difference is that the measure of excess mortality used here is the ratio of deaths in 2020 to the average 2015-19 number of deaths (as in the joint Istat-ISS report, 2020). By exploiting information available at the LLM level from the 2011 population census, we add to the covariates the following variables (measured as a share of the total population): foreign residents, women, people aged 19 years or less, graduates. Finally, we add the 2019 population level and the same discrete variables used earlier to catch regional and specialization effects. Information about the continuous covariates used in the analysis is contained in Table A1 in the Appendix.

For the sake of space, only results concerning the continuous variables are presented in Figure 3. This reports point estimates for the continuous variables along with 95% confidence bands based on bootstrapped standard errors. Let us briefly comment the main results. First, it may be seen that the loss in employment levels was higher in areas with higher excess mortality (recall that we consider this as a proxy of the degree of spread of the pandemic). The size of this correlation is bigger in absolute value among LLMs which experienced the largest employment losses. While this result stands in marked contrast to CL's conclusions, other findings are broadly consistently with theirs. In particular, the estimates suggest that employment losses were heavier in labour market characterized by pre-pandemic fragility. To see this, note e.g. that the estimated coefficient of the 2019 unemployment rate is negative and significant across the entire distribution of the dependent variable. As in CL's results, employment contractions were larger in areas featuring a higher risk of social aggregation, as measured by the variable "*High_agg_lev*" (such activities are "*those that involve contact with other subjects in addition to the company's workers (e.g. catering, entertainment, hospitality)*"). Also, LLMs characterized by higher shares of young people and women suffered higher losses. Unreported results concerning the categorical variables suggest that the pandemic exerted more pronounced adverse effects in some "industrial districts" like those specialized in textile goods and those supplying furniture. Moreover, northern regions suffered higher employment losses on average.

3. Summary and conclusions

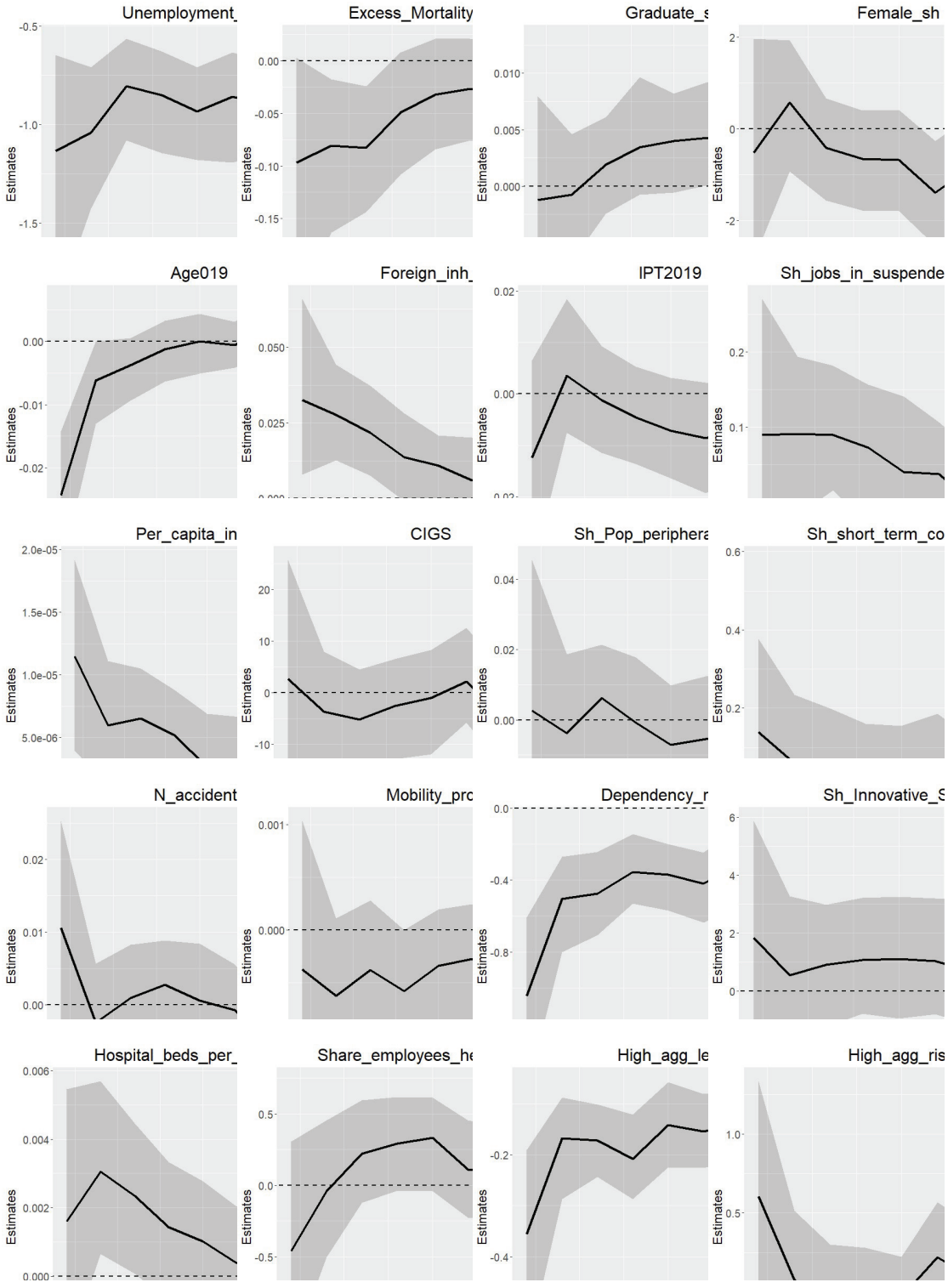
This paper analyses the effects of Covid-19 in Italian local labour markets, by providing counterfactual estimates of 2020 employment levels and by investigating which factors explain differences between observed and counterfactual employment. Our estimates are consistent with previous evidence in that pre-existing weaknesses played a role in explaining changes in labour demand in the aftermath of the pandemic.

According to CL, this "*calls for a place-based approach in the policy response to the crisis*"

[6] More in detail, *f(t, s(i))* is a "Moran eigenvector spatial filter". Its calculation requires the preliminary computation of a centred *N***N* proximity matrix based upon the coordinates of the centroids of the *N* areas. A subset of the eigenvectors of this matrix, those corresponding to positive eigenvalues (and thus useful to explain positive spatial dependence) are extracted from the proximity matrix. The spatial process is a linear combination of these eigenvectors (see again Murakami and Seya, 2019 for further details). Accounting for spatial dependence is particularly important in the present analysis, given the quite asymmetric spread of the pandemic across the country.

(i.e. policies taking into account the "*territorial profile […] of local economic systems*") instead of top-down national policies.

Figure 3 Quantile regression results



Results presented here do not contradict such a claim, but –in contrast to CL– suggest that the employment effects of the pandemic were not independent of its degree of spread (as proxied by excess mortality). We believe this implicitly underscores the importance of the positive economic externalities stemming from the public provision of health services and the need of efficient regional health systems really meeting targets defined at the national level.

References

Baltagi, B.H. (2005). *Econometric Analysis of Panel Data*. Chichester, West Sussex, England.

Cerqua, A., Di Stefano, R., Letta, M., Miccoli, S. (2021). Local mortality estimates during the COVID-19 pandemic in Italy. *Journal of Population Economics,* **34**(4), pp. 1189–1217.

Cerqua, A., Letta M. (2022). Local inequalities of the COVID-19 crisis. *Regional Science and Urban Economics*, **92**.

Istat, ISS (2020). *Decessi per il complesso delle cause*. *Periodo gennaio-novembre 2020*. Italy.

Jerez, M., Montealegre-Luna, A., Garcia-Hiernaux, A. (2023). A counterfactual analysis of the impact of the 2008 and 2020 crises on Spanish employment. *Applied Economic Analysis*, **31**(92), pp. 109-125

Murakami, D., Seya, H. (2019). Spatially filtered unconditional quantile regression. *Environmetrics*. **30**(5), e2556.

APPENDIX

Table A1 Description of continuous variables included into the association analysis

Variable	*Description*	*Source*
Graduate_share	Population share holding a university degree in 2011	(a)
Female_sh	Female population share in 2011	(a)
Age019	Population share aged 0 to 19 years in 2011	(a)
Foreign_inh_sh	Foreign residents in 2011	(a)
lPT2019	Total population size in 2019	(a)
Excess_Mortality_2020	(Total deaths in 2020) / (Average number of deaths in the period 2015-2019)	(b)
Sh_jobs_in_suspended_activities	The 2017 share of jobs in activities suspended in March 2020 by the Italian Government	(c)
Per_capita_income	The amount of money earned per person in 2019	(c)
CIGS	Share of firms registered in the Business Register with employees in CIGS (aver. 2015-2018)	(c)
Sh_Pop_peripheral_areas	Share of population living in areas defined by Istat as peripheral or ultra-peripheral (january 1, 2020)	(c)
Sh_short_term_contracts	Share of employees with temporary contracts divided by the number of employees (in October 2015)	(c)
N_accidents	The number of road accidents with injuries to persons divided by resident population * 10,000 (2019)	(c)
Mobility_proxy	Index of intra-LLM mobility (2011)	(c)
Dependency_ratio	(People aged 0-14 or 65 and more) / People aged 15-64 (january 1, 2020)	(c)
Sh_innovative_SME	The share of innovative start-ups in the Business Register (aver. 2016-2019)	(c)
Hospital_beds_per_1000_inh	Number of hospital beds divided by resident population * 1000	(c)
Sh_employees_health_care	% of jobs in the NACE 2-digit sectors 'human health activities' and 'residential care activities' (2019)	(c)
High_agg_level	Share of employees exposed to a medium-high or high risk of social aggregation (2019)	(c)
High_agg_risk	Share of employees exposed to a medium-high or high integrated risk (2019)	(c)
Unemployment rate	Unemployment rate in 2019	(c)

Sources: (a) = Istat; (b) own calculations from Istat data; (c) Cerqua and Letta (2022) dataset

doi.org/10.26398/asaproc.0049

Labour between languages, memories, services and innovation: A new perspective for highlighting migrants' skills to host communities

Cristina Martelli, Maria Flora Salvatori, Adham Kahlawi
Department of Statistics, Computer Science and Applications, University of Florence, Florence, Italy

1. Introduction

The role of language within the labour market realm is pivotal for various stakeholders, including labour market participants and scholars from diverse fields such as statistics, economics, and vocational training. The collective aspiration revolves around establishing a dynamic, responsive system that addresses the evolving requirements of workers, companies, communities, and territories. This imperative necessitates achieving a shared understanding.

At the same time, scholars seek narratives, memories, and sources that can adequately capture the intricate and interconnected nature of labour dynamics. In an ideal scenario, the records stemming from management processes, conveyed through specific languages and outcome of their use, should be harmonised and amenable to complex statistical analysis as well.

A pragmatic view reveals a stark divergence from these ideals. The landscape of labour is marked by a multitude of languages, each possessing its own attributes that don't always neatly align with one another. This linguistic diversity is exemplified by terms such as ISCO[1], ESCO[2], and O'NET[3], which represent just a fraction of the linguistic frameworks in play. In Italy, the panorama expands to include regional skill repertoires, the Atlante Lavoro[4], Istat's ISCO specifications at the fifth-digit level, the Unioncamere coding standard for Excelsior surveys[5], and even language coined by the Italian Funzione Pubblica to delineate digital skills. [6] Moreover, private recruitment agencies have developed proprietary languages, often protected by copyright, that further compound the overall complexity.

In this intricate tapestry, natural language holds newfound prominence, particularly in management systems striving to foster seamless interactions with users. The allure of natural language lies in its ability to allow users to express their needs with simplicity and directness. Challenges arise when attempting to align natural language descriptions, such as those from companies and workers, in processes like job matching, wherein the responsibility of matching often falls on natural language processing (NLP) and artificial intelligence (AI) mechanisms.

Harmonizing this babel of idioms employs the time-tested approach of *big data*, leveraging NLP and AI protocols to uncover shared meanings across divergent semantics. Yet, the costs and efforts involved can be substantial, and the outcomes don't invariably fulfil the desired granularity and precision. For instance, Lightcast data[7], a trove of labour market insights drawn from job vacancies, sometimes grapples with delivering the level of skills detail crucial for in-depth regional labour analyses.

This study adopts an alternative standpoint by delving into the reutilization of administrative

[a] Department of Statistics Computer Science Applications, University of Florence, Florence, Italy.
[1] https://www.ilo.org/public/english/bureau/stat/isc
[2] https://esco.ec.europa.eu/en/classification
[3] https://esco.ec.europa.eu/en/classification
[4] atlantelavoro.inapp.org
[5] https://excelsior.unioncamere.net/sites/default/files/documenti/Nota_Metodologica.pdf
[6] https://www.syllabus.gov.it/syllabus/
[7] https://lightcast.io/

data, posing the question: Can we transcend the post hoc big data model, notorious for its expense and information loss, in favour of an ex-ante approach? This approach, labelled as *great data* concept within this work, revolves around crafting administrative data seamlessly tailored to administrative processes, inherently amenable to statistical utilization (Bender et al., 2023; Tenopir et al., 2020). ESCO, the European classification of skills, emerges as a potential solution—a semantic and linguistic instrument designed to harmonize the depth of natural language with the precision inherent in coding systems.

The structure of this study is as follows: Section 2 delves into the characteristics of the European language of work, scrutinizing its applicability in both management services and statistical analyses. Subsequently, Section 3 introduces the functions and attributes of ESCO's new interface, designed to enhance its utility in employment support services. Moving forward, Section 4 presents a practical application grounded in the ESCO framework and its new interface, aimed at identifying the skill sets migrants possess upon entering Italy. Finally, Section 5 encapsulates the study's conclusions and outlines future prospects.

2. The European labour language ESCO between natural language and standard coding

ESCO, short for the European Skills, Competences, Qualifications, and Occupations, serves as a multilingual classification system within the framework of the Europe 2020 strategy. Its primary purpose is to systematically categorize and delineate skills, competences, qualifications, and occupations relevant to the European Union's labour market and educational landscape. One key advantage of ESCO is its ability to illustrate complex relationships between these various concepts.

Developed by the European Commission starting in 2010, ESCO is an extension of the ISCO (International Standard Classification of Occupations) scheme. Its continuous improvement process hinges on input from expert groups and outcomes from European projects, such as the "Blueprint for Sectoral Cooperation on Skills."[8]

ESCO offers several advantageous features for crafting job descriptions:

I. Precise coding: ESCO employs a well-defined coding system that uniquely identifies each individual job element, ensuring unambiguous representation.
II. Each encoded ESCO element is accompanied by a rich and complete description in natural language to be used, in conjunction with other ESCO objects, to compose complex narratives such as CVs or job offers.
III. Multilingual descriptions: Each ESCO object, featuring a standard code and natural language description, is available in all 27 European languages plus Arabic and Ukrainian.
IV. Relational structure: ESCO elements are interconnected, enabling the description of business processes associated with skills. This capability sets it apart from other coding systems (De Smedt et al., 2015).

The simultaneous presence of standard coding and descriptions in natural language helps to find a trade-off between the need to describe oneself in natural language and the need for statistical analysis and job matching to count on rigorous coding. Also, consider another important aspect: many language-based solutions aimed at job matching tend to generate clusters and classifications based on the language used by companies and workers. These classifications are essentially expressions of the language itself and often suffer from limitations, such as flattening concepts into more conventional and frequently used ones.

In contrast, ESCO follows a different approach. International groups of experts and participants in European projects meticulously analyse European production processes, with a keen focus on emerging, evolving, or declining processes. They identify the skills required for these processes and provide comprehensive descriptions of these skills and their interconnections. Consequently, ESCO stands as a representation of the most cutting-edge developments in this field. It excels in describing

[8] https://ec.europa.eu/social/main.jsp?catId=1415&langId=en

the labour market and skills with a level of detail, relevance, and innovation awareness that traditional coding languages based on narrative elaboration typically struggle to achieve.

However, ESCO encounters certain operational challenges in fulfilling this role, primarily associated with its existing interface. While it proves highly effective for navigating between objects, including the ability to switch reference languages, it falls short when it comes to the tasks of saving, selecting, and utilizing intermediate navigation steps. These limitations inherent in the current interface hinder the ease of incorporating ESCO into text composition and creating labour-related administrative resources that adhere to ESCO standards.

This paper, therefore, introduces and deliberates on a novel ESCO interface that offers the following functionalities:

I. Intermediate step storage: It allows users to save intermediate steps of their ESCO navigation, akin to the familiar tools employed for online shopping.
II. Personal profile creation: Users can construct a personalized profile for either a worker or a company, described in terms of ESCO objects.
III. CV and job offer generation: Building upon this profile, users can effortlessly compose a CV or a job offer. Specifically, in the case of a worker, the new ESCO interface streamlines the automatic generation of a European-format CV in all the languages pertinent to fully harnessing the worker's potential.

3. The new ESCO object manipulation information system: Structure and functionalities

To overcome ESCO drawbacks in terms of usability and to make it suitable to produce great data, we have projected and built a new interface for ESCO, and we have introduced this new instrument in a concrete process of work intermediation: the production of European CVs of migrants in Italy.

The project's goal was to develop a management information system tailored for reception and shelter centres, as well as for migrants. The system's primary purpose was to facilitate migrants in assessing and showcasing their skills within a comfortable and culturally familiar linguistic environment, benefiting both cultural mediators and migrants themselves.

The overarching objective was to streamline the process of recognizing migrants' professional competencies. This recognition process aimed to culminate in the provision of a European-format CV in languages relevant to the individual migrant's preferences.

Simultaneously, in line with the great data paradigm, the project focused on gathering data concerning migrants' skills at the point of their arrival in Italy. This data, which is largely uncharted, could prove invaluable in enhancing job placement opportunities and accurately guiding the perceptions of host communities.

3.1. The relationship between natural and standard codification of the contents: inverting the sense of the treatment process

It is a commonly held belief that structured languages, much like any coding system, should possess greater adaptability to support the typical processes involved in job seeking or job offering. Adhering strictly to a rigid system, rather than describing what stakeholders genuinely seek, is often deemed impractical.

In practical applications, the information processing procedure that underpins job matching usually commences with a narrative presented in natural language. Subsequently, these texts are subjected to NLP protocols to generate a standardized coding representation.

However, when working with ESCO, this conventional process can be reversed due to the highly detailed content descriptions within ESCO's coding structure. Advanced linguistic techniques can be employed to pinpoint and select the relevant "building blocks" for composing a text that simultaneously possesses the qualities of a straightforward description while maintaining

the precision of coded data. This approach transforms the treatment process by utilizing linguistic assets to identify the essential text elements, effectively bypassing the coding phase in the process.

3.2. The new ESCO interface: Elements and functionalities

Shopping cart

After logging in, the system allows you to navigate ESCO using all the available languages. The user navigates using a search field adopting the preferred language in a context where the usual distinction between occupation and skill/knowledge has become transparent. The same password allows for exploring both ESCO environments. The system's response to keywords depends, in the classic ESCO interface, on the application's context (occupation or skills). While remaining valid at the architectural system level, this distinction is transparent to the user, who, otherwise, may need to know the difference between the two concepts.

Profile

The ESCO navigation phase led to storing in the shopping cart all the interesting objects, which resulted, at first analysis, attractive to the user. The system now allows using a new environment, the profile area, in which the user evaluates all the items in the cart and selects the most appropriate to represent and describe the personal/professional or business reality.

CV/ Job offer

The profiling area is the starting point for generating both the workers' CVs and the companies' job offers: starting from the profiling area, which describes, by speaking "the ESCO language", the reality of the worker or the company, numerous different texts can be derived, in which the skills and knowledge offered or sought for specific contexts will be selected. A worker will choose from time to time which elements to highlight to meet the needs of a particular employer: a company will do the same when building the job offer for a specific need.

ESCO-LIKE objects.

The skills and knowledge offered or sought may not be included in the ESCO repertoire. In that case, the system asks the user to describe (through a descriptive template identical to the one adopted by ESCO) the lacking elements. After being evaluated, the information system includes these ESCO-LIKE elements in the basket of objects to describe one's skills. These additional elements, which are kept distinct in terms of archiving and denotation, are then communicated to ESCO, which can thus collect functional elements for updating the system from its use in the field. This way, adopting this information system provides feedback similar to that obtained through specialised surveys such as those organised in EU Blueprint projects.

4. Experiences currently in progress: Skills assessment of migrants accommodated in some Tuscan shelters

The previously described information system is currently in use in the context of a FAMI project[9] (called PARAGRI[10]) aimed at actions to combat illegal hiring in agriculture and at assessing the skills of migrants upon their entry into Italy: while it is relatively known what migrant people do in the Italian labour market nothing is known of what they could have done based on the skills they already have upon their arrival in Italy. The exploratory actions currently underway within FAMI-PARAGRI show a range of knowledge and skills which, if known, could be valuable for the needs of Italian companies. ESCO presents its information in all European languages and Arabic; the Ukrainian language has also recently been added. This ESCO multilingual information service has been taken over and enhanced by the new information system and has proved to be extremely important in assessing the skills of migrant people. In field activities, the operators of the shelter institutions ask the migrant to write the first keywords in their language, and from that point,

[9] https://www.interno.gov.it/it/temi/immigrazione-e-asilo/fondi-europei/fondo-asilo-migrazione-e-integrazione-fami
[10] https://paragri.it/

the browsing and saving in the shopping cart begin. Operator and migrant switch between Italian and the language chosen by the migrant to be fully aware of the selected and archived contents. The new system has expanded the range of languages made available to the migrant, also adding the Sinti, Urdu and Pashtu languages, which are those spoken by migrants arriving from the so-called "Balkan route". These are no-official translations, automatically generated through advanced translation tools. Before proposing them to the users, they were tested by doing a triangulation from Italian, the new languages and English. We have automatically translated an ESCO element described in Italian into one of the additional languages, and then we have again automatically translated it into English. The quality of the translation from Italian to English was excellent, as also confirmed by the migrants who are experiencing it in the field.

In a common service scenario, a foreign migrant who primarily speaks only his/her mother tongue is being aided by an Italian shelter assistance operator to assess their skills. Typically, the operator does not speak the migrant's language and may not be well-versed in the migrant's specific skills from their home country. Thanks to the structured ESCO skills description framework, the migrant can describe their profile in their own language, and the operator can assist by translating it into Italian, suggesting additions and potential enhancements. This system facilitates communication between the two parties, making full use of the ESCO terminology and ultimately generating a European-style CV in all relevant languages.

4.1. Matching labour supply and demand in the new ESCO based information system

The tools described in the previous paragraphs are also available to companies that offer jobs. In this case the company uses keyword research to compose an offer to be published.

In the case of the PARAGRI project, the companies created and published their offers mainly in Italian, but the linguistic features of the ESCO-based system also allow for the preparation of offers in different languages.

The system is designed to directly match CVs with offers: In the new context, the matching is not based on an NPL linguistic analysis but by leveraging the proximity of the ESCO nodes used in the texts we intend to compare (Le Vrang et al., 2014).

This approach, therefore, on the one hand, allows pushing the description to the high level of detail typical of the ESCO language; on the other, it will enable the evaluation of semantic distances based on the distance between the nodes (Riedel et al., 2023). This gap can eventually be overcome by the learning outcomes of training activities aimed at bridging the gap.

It's worth noting that the system is also adaptable for use on numerous supply-and-demand meeting platforms that are accessible to both job seekers and employers. In such cases, CVs and job listings are simply uploaded to these platforms and processed according to the standard linguistic practices employed by these matching systems.

This approach serves a dual purpose: it enables the elaboration of job descriptions to achieve the high level of detail characteristic of the ESCO language while also facilitating the assessment of semantic distances based on the proximity of nodes. Any distance between the nodes of CV and job offers can be addressed through training activities aimed at bridging the gap.

4.2. The statistical information system produced by the migrant support service

The work conducted within the Paragri project at migrants' shelter centres extends beyond offering services solely to migrants and businesses. It has also generated valuable statistical resources, enabling the analysis of migrants' competency profiles upon their arrival in Italy.

Utilizing the ESCO language in service provision has offered a significant advantage. It obviated the need for a subsequent coding phase for administrative records. These administrative archives possess inherent statistical utility, facilitating the immediate generation of summary data, benefiting localities and institutions alike.

In this context, we are progressing toward an enhanced rendition of traditional open data. This evolution empowers statistical users with readily accessible, ready-to-use data that can seamlessly

complement and be benchmarked against the most prominent official statistical sources.

5. Conclusion and future work

The advancements in artificial intelligence and natural language processing have ushered in a wave of technical innovation that is heavily integrated into all facets of big data technologies. One common application is the conversion of narratives expressed in natural language into structured coding systems. These technical and methodological strides also underpin the approach outlined in this work and are instrumental in the Paragri project.

In this context, the project objective is twofold. First, it aims to enhance navigation within the intricate network of ESCO nodes, simplifying the process of accessing information. Secondly, it seeks to modernize the traditional user password search, making it more user-friendly and efficient. The challenge of bridging the gap between the demand for and supply of digital skills, along with accurately defining these skills, is a focal point for numerous intriguing initiatives and projects geared towards aiding companies in their innovation endeavours. Frequently, these proposals rely on specially crafted languages tailored to the task. Within the scope of the Paragri project, an examination of this facet revealed a noteworthy observation: from a descriptive standpoint, there exists no distinct specificity for digital skills. Instead, the skills typically sought after by companies are comprehensively documented and accounted for.

Another pivotal application on the horizon for upcoming releases involves strengthening the transition from identifying skill gaps to facilitating relevant training. This will be achieved through the development of dedicated chatbots, which will play a central role in guiding individuals towards the training they need to bridge those skill gaps.

References

Bender, S., Blaschke, J., Hirsch, C. (2023). Statistical data production in a digitized age: The need to establish successful workflows for micro data access. In *Advances in Business Statistics, Methods and Data Collection* (pp. 519–536). John Wiley & Sons, Ltd. https://doi.org/https://doi.org/10.1002/9781119672333.ch22

De Smedt, J., le Vrang, M., Papantoniou, A. (2015). ESCO: Towards a Semantic Web for the European Labor Market. *Ldow@ Www*.

Le Vrang, M., Papantoniou, A., Pauwels, E., Fannes, P., Vandensteen, D., De Smedt, J. (2014). ESCO: Boosting job matching in Europe with semantic interoperability. *Computer*, **47**(10), pp. 57–64. https://doi.org/10.1109/MC.2014.283

Riedel, J., Neumann, J., Stark, L. (2023). Standardized skill description: How to match the supply and demand of job skills. *Open Research in Progress--Reports of the Center for Open Digital Innovation and Participation*, **1**(1).

Tenopir, C., Rice, N.M., Allard, S., Baird, L., Borycz, J., Christian, L., Grant, B., Olendorf, R., Sandusky, R.J. (2020). Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLoS ONE*, **15**(3), pp. 1–26. https://doi.org/10.1371/journal.pone.0229003

doi.org/10.26398/asaproc.0050

Frailty, activity participation, and welfare: An analysis using SHARE data

Rossella Miglio, Chiara Puglisi, Rosella Rettaroli, Giulia Roli, Francesco Scalone
Department of Statistical Sciences, University of Bologna, Bologna, Italy.

1. Introduction

Population ageing has become a pressing issue in the European Union, where the share of older individuals (aged 65 or more) is projected to reach almost 30% of the total population by 2050 (Eurostat 2019). Frailty represents a condition of increased vulnerability among older people, implying a worsening of the risk of adverse health outcomes following losses in one or more domains of the human functioning (Bessa et al., 2018). Thus, a proper understanding of the determinants of frailty can support the identification of policies for healthy ageing, further providing information on the role that institutions can play to this end.

Despite this, the role played by some specific factors in influencing elderly's frailty, both at the micro and at the macro-level, is still to be fully understood. At the micro-level, a key missing piece concerns the role of social networks and activity participation in affecting frailty trajectories. While a great deal of scholarly attention has been devoted to the effect of these aspects on health and survival (Wenger, 1997; Berkman et al., 2000; Jenkin et al., 2018; Morrow-Howell et al., 2003), only few studies have explored the relationship between social aspects and engagement in various kinds of activities and frailty in Europe, such as Etman et al., (2015) and Watts et al. (2017). More recently, Miglio et al. (2023) have attempted to fill this gap, by using SHARE data on 13 European countries to explore how social connectedness and activity participation predict individual frailty trajectories. Their results revealed that both aspects play a significant role in predicting transition to pre-frailty and frailty and risk of death, with activity participation resulting as a much more powerful determinant. However, in their definition of activity participation the authors include widely different types of activities, among which "socially intensive" activities, entailing relevant levels of social interaction, such as doing charity work, participating to sport and other kinds of clubs, and "individual" activities, such as reading and playing number and word games, typically performed in isolation. This leaves open the question of whether there is any difference in the role that these distinct set of activities play in predicting frailty and, if so, which ones matter the most.

Thus, our goal is to complement Miglio et al. (2023), by comparing the effect of participating to "socially intensive" and "individual" activities on frailty. Understanding this aspect can help elucidate on whether it is the social doing and the interpersonal interaction entailed in the more socially intensive activities that can decrease non-frail individuals' likelihood of becoming pre-frail and frail and their risk of death, or whether the entertainment and the intellectual engagement associated with activities such as reading and playing puzzle games are also relevant aspects affecting frailty trajectories.

In addition, we further contribute to the literature by investigating the role of a macro-level characteristic whose impact on frailty has not been thoroughly studied, namely countries' welfare regime. Considering differences in aspects like access to healthcare, generosity and coverage of pension systems, and strength of social safety nets among different welfare regimes, individuals from countries belonging to distinct welfare systems might differ in their frailty trajectories.

2. Data and methods

We rely on data from the 4th (2011) and 5th (2013) waves of the Survey on Health, Aging and Retirement in Europe (SHARE) (Börsch-Supan 2022a 2022b), a longitudinal survey that provides

information on health, socioeconomic status and social life of community-dwelling individuals aged 50 or older in European countries. Out sample includes 8,677 men and 8,616 women aged 60 and older, across 13 European countries.

To assess the impact of different kinds of activities, we rely on some specific questions that SHARE asks survey subjects about the frequency of their engagement with them. In doing so, we distinguish between four "socially intensive" and three "individual" activities. The socially intensive ones include voluntary/charity work, educational or training courses, sport/social/other kinds of clubs, and political/community-related organizations. The individual activities include reading books, magazines, and newspapers, doing word or number games and playing cards or chess. We coded participation to each activity as 1 if the individual engaged with the activity at least once a month in the previous 12 months and 0 otherwise. We then constructed two different scales, one for each group, by counting the number of activities individuals took part to in the previous 12 months, so that the higher the individual's score in each scale, the greater the variety of activities survey subjects participated to. The Social Activities Scale ranged between 0 and 4, while the range of the Individual Activities Scale was 0-3.

To investigate the role of welfare arrangements on frailty, we employ Eikemo et al.'s (2008) categorization, classifying countries into Social-democratic/Scandinavian (Denmark and Sweden), Bismarckian/continental (Austria, Belgium, France, Germany, the Netherlands, and Switzerland), Mediterranean/southern (Italy and Spain), and eastern/post-communist (Czech Republic, Estonia, Poland, and Slovenia).

Our frailty measure is Romero-Ortuno and Kenny's (2012) SHARE Frailty Instrument (SHARE-FI), a pre-calculated, population-representative and gender-specific frailty index based on Fried et al.'s (2001) phenotype model. With such index, frailty is assessed based on the presence of five symptoms, that is, unintended weight loss, weakness, slowness, self-reported exhaustion, and low physical activity. Our dependent variable thus classifies individuals into non-frail, pre-frail, and frail. We further included "deceased" as a possible outcome.

Considering the different incidence of frailty among men and women, we stratified our analysis by sex, estimating separately for men and women a multinomial logistic regression model on a pooled sample of all analyzed countries[1] to explore how characteristics assessed in wave 4 influenced frailty in wave 5. We control for social network scale, age, marital status, working status, country-specific wealth quintiles, number of chronic conditions, verbal fluency, and EURO-D scale of depression.

3. Results

We provide the results of our multinomial regression models Table 1, in the form of relative risk ratios (RRR).

Results of our models reveal that, for both men and women, engagement with "social" activities protects against the risk of moving from a state of non-frailty to one of pre-frailty and frailty, even if with different intensity. Specifically, among both men and women, participating to social activities significantly decreases the likelihood of becoming frail. However, the magnitude of the effect is different for men and women: while among men each additional score in the social activities scales is associated to an RRR of 0.65, among women the RRR is equal to 0.74, signaling a relatively smaller effect. Rather, participating to individual activities seems to only affect men's risk of becoming pre-frail, and not women's. Gender-related differences are also observed when analysing the likelihood of moving from a state of non-frailty to one of frailty. Individual activities appear to only affect the transition to frailty among women, while no effect is observed among men.

[1] Considering our interest towards investigating the role of welfare regimes, we did not include country-fixed effects in our models, as doing so would have entailed collinearity issues. However, results concerning the role of social and individual activities in affecting frailty are robust to an alternative specification of our models, including country-fixed effects rather than welfare regimes effects.

Table 1: Estimate of regression parameters from the multinomial logistic regression models (s.e. in brackets)

	Men			Women		
	Pre-frail	**Frail**	**Deceased**	**Pre-frail**	**Frail**	**Deceased**
Social Activities Scale	0.83***	0.65***	0.86	0.83***	0.74*	0.54***
	(0.04)	(0.05)	(0.07)	(0.03)	(0.12)	(0.05)
Individual Activities Scale	0.91*	0.90	0.95	0.94	0.80**	0.78
	(0.04)	(0.09)	(0.05)	(0.03)	(0.07)	(0.12)
Social Networks Scale	0.85***	0.96	1.05	0.93	0.97	1.05
	(0.04)	(0.06)	(0.08)	(0.03)	(0.07)	(0.14)
Age (Cont.)	1.06***	1.11***	1.10***	1.06***	1.09***	1.10***
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Wealth Quintile (Ref. 1st)						
2nd	1.23	1.07	0.87	0.92	0.75	0.65
	(0.23)	(0.24)	(0.21)	(0.06)	(0.12)	(0.23)
3rd	1.13	1.02	0.66**	0.86*	0.65*	0.85
	(0.12)	(0.14)	(0.10)	(0.06)	(0.13)	(0.33)
4th	0.95	0.52**	0.71	0.80**	0.43***	0.94
	(0.10)	(0.13)	(0.18)	(0.06)	(0.07)	(0.33)
5th	0.89	0.57***	0.70	0.72***	0.44**	0.89
	(0.11)	(0.09)	(0.21)	(0.06)	(0.13)	(0.33)
Married/in relationship (Ref. Single)	0.89	1.12	0.64	1.01	1.10	0.67
	(0.10)	(0.21)	(0.15)	(0.06)	(0.06)	(0.14)
Currently working (Ref. Not working)	0.75	0.62	0.73	0.90	0.43***	0.84
	(0.12)	(0.16)	(0.16)	(0.10)	(0.10)	(0.17)
N. of chronic conditions	1.25***	1.25***	1.09**	1.21***	1.41***	0.96
	(0.03)	(0.01)	(0.04)	(0.02)	(0.09)	(0.09)
Verbal Fluency score	0.99	0.97	0.96**	0.99***	0.95***	0.97
	(0.01)	(0.02)	(0.01)	(0.00)	(0.01)	(0.03)
EURO-D scale of depression	1.15***	1.17*	1.09	1.16***	1.20***	1.11
	(0.04)	(0.07)	(0.05)	(0.03)	(0.05)	(0.08)
Welfare regime (Ref. Scandinavian)						
Continental	1.20	1.03	0.78	1.11	0.95	0.45*
	(0.12)	(0.47)	(0.22)	(0.11)	(0.22)	(0.17)
Mediterranean	1.86***	1.59	1.26	1.99***	2.63***	0.73
	(0.27)	(0.71)	(0.31)	(0.15)	(0.65)	(0.30)
Post-communist	1.77***	1.51	2.14**	1.26*	1.51	0.78
	(0.20)	(0.67)	(0.51)	(0.13)	(0.37)	(0.30)

Source: Authors' elaborations on SHARE data. *Notes:* 95% C.I. *** $p<0.001$, ** $p<0.01$, * $p<0.05$.

Gender-related differences also emerge when it comes to the mortality outcome, as while for non-frail women participating to social activities significantly decreases the probability of death (RRR = 0.54, $p < 0.001$), for men this effect is not observed. In addition, for both men and women, engaging in "individual" kinds of activities does not significantly impact the likelihood of death.

As for the other variables included in the model, our results are in general consistent with previous studies on the determinants of health in later life, highlighting a positive association between number of chronic conditions and frailty (Mangin et al., 2023), as well as a positive association between depression and frailty (Soysal et al., 2017).

In terms of effects of welfare regime of the country where the individual lives on frailty, relevant differences emerge between men and women: a welfare-related effect is observed, among men, for the transition to pre-frailty, so that men in countries with Mediterranean and post-communist welfare regimes are more likely to become pre-frail, relatively to men in Scandinavian countries. In addition, a welfare effect is observed for the risk of death, with men in post-communist countries displaying a 1.14 higher RRR ($p < 0.01$), compared to men in Scandinavian countries. Conversely, significant differences emerge in the frailty trajectories experienced by non-frail women from different welfare regimes, for all outcomes considered. Specifically, non-frail women in Mediterranean welfare states (*i.e.*, Spain and Italy) appear to be significantly more likely to successively become pre-frail (RRR = 1.99, $p < 0.001$) and frail (RRR = 2.63, $p < 0.05$) relative to women in Scandinavian countries. Similarly, women in post-communist welfare regimes, too, appear to be more likely to become pre-frail than their Scandinavian counterparts. In addition, significant differences are observed among non-frail women from continental and Scandinavian welfare regimes, with the former displaying significantly lower relative risk of death than the latter.

4. Conclusions

This study aimed at examining and comparing how participating to "social" and "individual" activities affects frailty at older ages. In addition, we also aimed at understanding differences in the prevalence of frailty among men and women from different welfare regimes. To do so, we employed SHARE data on community-dwelling individuals from 13 European countries, aged 60 and older.

Our findings suggest that both participating to socially intensive activities and "individual" activities are aspects that affect frailty trajectories, with the former, often entailing a great deal of interpersonal interaction, seemingly exerting a stronger effect. Some of these effects, however, were stratified across gender. In particular, among both non-frail men and women, participation to activities such as doing voluntary work and participating to community-related organizations and sport, social, or other kinds of clubs protects against the risk of becoming pre-frail and frail, and for women, against the risk of death, too. The impact of engagement with individual activities, on the other hand, appears to only affect transition to pre-frailty, for men, and to frailty among women.

All in all, these findings are suggestive of a different role of the two groups of activities examined, so that activities requiring a certain degree of networking and interaction appear to affect all the considered outcomes, albeit differently among men and women, while engaging with activities such as doing crossword puzzles, number games, and playing cards, affects differently pre-frailty and frailty for men and women. There are various hypotheses as to what might produce these differences. For instance, the role theory emphasizes the key impact of activities like volunteering and participating to various kinds of clubs, thanks to the active engagement these entail. This active engagement can be especially important for elderly individuals who face the end of their lifelong responsibilities as employees, caregivers, or partners, due to life events like retirement and widowhood (Anderson et al., 2014).

In terms of differences in frailty trajectories among different welfare-regimes, we further found evidence of higher mortality risks among men in countries characterized by post-communist welfare regimes, compared with men from Social-democratic/Scandinavian countries. This is consistent with findings of a stagnation in older people's (particularly men's) life expectancy in Eastern European countries (Karanikolos, Adany and McKee, 2017), a pattern that is widely different from that observed in Social-democratic/Scandinavian countries, typically displaying life expectancy levels that exceed the global average (Knudsen et al., 2019).

We also highlighted higher likelihood of developing pre-frailty and frailty among non-frail

women in Mediterranean welfare regimes, compared to women in Scandinavian ones. Possible explanations for this result include the high degree of familialism characterizing Mediterranean welfare states (Misra and Moller, 2005), determining a heavy responsibility of the family for their members' wellbeing. This can be particularly detrimental for the health of elderly women, given the gendered expectations seeing women bearing the greatest caregiving responsibilities for family members (Zygouri et al., 2021). As such, women in familialistic societies are all the more likely to end up bearing a disproportionate caregiving burden, providing care for grandchildren and spouses. This can increase their stress and fatigue, negatively impacting their health and thus determining an increase in the likelihood of becoming pre-frail and frail, relative to women from countries characterized by Scandinavian welfare regimes. Indeed, in such systems, the state tends to assume an active role in providing comprehensive social support, via policies and public services that ensure the well-being and equality of its citizens, such as healthcare, long-term care, pension and housing programs, all aspects that partially relieve the individual of often burdening care responsibilities and concurrently improve healthcare outcomes among elderly women, thanks to a widespread access to quality medical care and support services.

Acknowledgements

This research was funded by the 2020 PRIN-Research Project of National Relevance "SOcial and health Frailty as determinants of Inequality in Aging (SOFIA)". We acknowledge co-funding from European Union – Next Generation EU, in the context of The National Recovery and Resilience Plan, Investment Partenariato Esteso PE8 "Conseguenze e sfide dell'invecchiamento", Project Age-It (Ageing Well in an Ageing Society).

This paper uses data from SHARE Waves 4 and 5. The SHARE data collection has been funded by the European Commission, DG RTD through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812), FP7 (SHARE-PREP: GA N°211909, SHARE-LEAP: GA N°227822, SHARE M4: GA N°261982, DASISH: GA N°283646) and Horizon 2020 (SHARE-DEV3: GA N°676536, SHARE-COHESION: GA N°870628, SERISS: GA N°654221, SSHOC: GA N°823782, SHARE-COVID19: GA N°101015924) and by DG Employment, Social Affairs & Inclusion through VS 2015/0195, VS 2016/0135, VS 2018/0285, VS 2019/0332, and VS 2020/0313. Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C, RAG052527A) and from various national funding sources is gratefully acknowledged (see www.share-project.org).

References

Anderson, N.D., Damianakis, T., Kröger, E., Wagner, L.M., Dawson, D.R., Binns, M.A., Bernstein, S., Caspi, E., Cook, S.L., and The BRAVO Team. (2014). The benefits associated with volunteering among seniors: A critical review and recommendations for future research. *Psychological Bulletin*, **140**(6), pp. 1505–1533.

Baxter, S., Blank, L., Cantrell, A., Goyder, E. (2021). Is working in later life good for your health? A systematic review of health outcomes resulting from extended working lives. *BMC Public Health*, **21**, pp. 1-11.

Berkman, L.F., Glass, T., Brissette, I., Seeman, T.E. (2000). From social integration to health: Durkheim in the new millennium. *Social Science and Medicine,* **51**(6), pp. 843-857.

Bessa, B., Ribeiro, O., Coelho, T. (2018). Assessing the social dimension of frailty in old age: A

systematic review. *Archives of Gerontology and Geriatrics,* **78**, pp. 101–113.
Börsch-Supan, A. (2022a). Survey of health, ageing and retirement in Europe (SHARE) Wave 4. Release version: 8.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w4.800
Börsch-Supan, A. (2022b). Survey of health, ageing and retirement in Europe (SHARE) Wave 5. Release version: 8.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w4.800
Eikemo, T.A., Bambra, C., Joyce, K., Dahl, E. (2008). Welfare state regimes and income-related health inequalities: A comparison of 23 European countries. *The European Journal of Public Health*, **18**(6), pp. 593-599.
Etman, A., Kamphuis, C.B.M., van der Cammen, T.J.M., Burdorf, A., van Lenthe, F.J. (2015). Do lifestyle, health and social participation mediate educational inequalities in frailty worsening? *The European Journal of Public Health*, **25**(2), pp. 345–350.
Eurostat (2019). *Ageing Europe - Looking at the Lives of Older People in the EU*. EU, Stat. book.
Fried, L.P., Tangen, C.M., Walston, J., Newman, A.B., Hirsch, C., Gottdiener, J., Seeman, T., Tracy, R., Kop, W.J., Burke, G., McBurnie, M.A., Cardiovascular Health Study Collaborative Research Group. (2001). Frailty in older adults: Evidence for a phenotype. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, **56**(3), pp. M146-156.
Jenkin, C.R., Eime, R.M., Westerbeek, H., van Uffelen, J.G.Z. (2018). Sport for adults aged 50+ years: Participation benefits and barriers. *Journal of Aging and Physical Activity*, **26**(3), pp. 363–371.
Karanikolos, M., Adany, R., McKee, M. (2017). The epidemiological transition in Eastern and Western Europe: A historic natural experiment. *The European Journal of Public Health*, **27**(suppl_4), pp. 4-8.
Knudsen, A.K., Allebeck, P., Tollånes, M.C., Skogen, J.C., Iburg, K.M., McGrath, J.J., ... Øverland, S. (2019). Life expectancy and disease burden in the Nordic countries: Results from the Global Burden of Diseases, Injuries, and Risk Factors Study 2017. *The Lancet Public Health*, **4**(12), e658-e669.
Mangin, D., Lawson, J., Risdon, C., Siu, H.Y.H., Packer, T., Wong, S.T., Howard, M. (2023). Association between frailty, chronic conditions and socioeconomic status in community-dwelling older adults attending primary care: a cross-sectional study using practice-based research network data. *British Medical Journal Open*, **13**(2), e066269.
Miglio, R., Puglisi, C., Rettaroli, R., Roli, G., Scalone, F. (2023). Frailty in Europe: The role of social networks and activity participation. Manuscript submitted for publication.
Misra, J., Moller, S. (2005). Famililism and welfare regimes: Poverty, employment, and family policies. *LIS Working Paper Series*, *399.*
Morrow-Howell, N., Hinterlong, J., Rozario, P.A., Tang, F. (2003). Effects of volunteering on the well-being of older adults. *The Journals of Gerontology: Series B*, **58**(3), pp. S137–S145.
Romero-Ortuno, R., Kenny, R.A. (2012). The frailty index in Europeans: Association with age and mortality. *Age and Ageing*, **41**(5), pp. 684–689.
Soysal, P., Veronese, N., Thompson, T., Kahl, K.G., Fernandes, B.S., Prina, A.M., ... Stubbs, B. (2017). Relationship between depression and frailty in older adults: A systematic review and meta-analysis. *Ageing Research Reviews,* **36***,* pp. 78-87.
Watts, P., Webb, E., Netuveli, G. (2017). The role of sports clubs in helping older people to stay active and prevent frailty: A longitudinal mediation analysis. *International Journal of Behavioral Nutrition and Physical Activity*, **14**(1), 95.
Wenger, G.C. (1997). Social networks and the prediction of elderly people at risk. *Aging and Mental Health*, **1**(4), pp. 311–320.
Zygouri, I., Cowdell, F., Ploumis, A., Gouva, M., Mantzoukas, S. (2021). Gendered experiences of providing informal care for older people: A systematic review and thematic synthesis. *BMC Health Services Research,* **21**, pp. 1-15.

doi.org/10.26398/asaproc.0051

Radiometric indices to evaluate total green areas in the major Italian urban centres

Stefano Mugnoli, Alberto Sabbi
DIPS-DCAT-ATA, ISTAT (National Institute of Statistics), Rome, Italy

1. Introduction[1]

Remote sensing is a branch of applied science aimed at acquiring qualitative and quantitative information by investigating objects without direct contact. This can be achieved through sensors installed on various platforms such as planes, satellites, and drones, which can measure electromagnetic wave radiation emitted or reflected by the objects under study.

For several decades, satellite-based earth observation has been a well-established procedure for monitoring our planet and conducting valuable surveys to study various environmental and territorial aspects, including vegetation condition, water pollution, hydrogeological instability, land cover, and soil consumption (Chiocchini et al., 2018).

The advantages of remote sensing images are significant, starting with the ease of acquiring territorial information compared to other techniques. Furthermore, the continuous availability of images allows for the study of phenomena that would be otherwise impossible to investigate using alternative methods.

The parameters detected by sensors are electromagnetic in nature, including radiation emitted, phase, polarization, and amplitude of the electromagnetic field. These parameters collectively determine the so-called 'spectral signature' of all detected objects. The spectral signature of an object is essentially its unique response to incident radiations at different wavelengths. Therefore, knowledge of the spectral signature of an object enables its unique identification.

One of the most extensively studied 'objects' in remote sensing is vegetation. Numerous spectral indices have been developed by specialists to highlight various aspects of vegetation cover, such as water stress, biomass quantification, fire damages, etc. This entire process is grounded in the behaviour of chlorophyll (Figure 1) concerning its ability to absorb light radiation at various wavelengths.

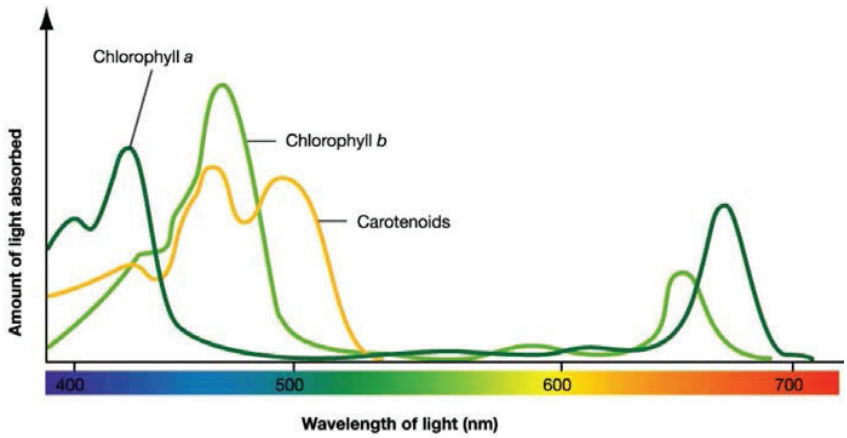


Figure 1 – Absorption spectrum of plant pigments[2]

In our work, we began with high-resolution remotely sensed images (AGEA Orthophotos with pixel resolutions of 20 and 50 cm), focusing on some of the most commonly used

[1] Even if the paper was devised by both the authors, S. Mugnoli wrote Sections 2 and 3; A. Sabbi wrote Sections 1 and 4.
[2] https://ressources.unisciel.fr/tp_virtuels/Pigment_Extraction_Lab/co/module_Virtual%20Experiment_1.html

vegetation indices (Xue et al., 2017; Pristeri et al., 2021) are calculated in order to extract statistics linked to the total vegetation cover in the major Italian urban centres.

In the following paragraphs, clarifications are provided on the importance of quantifying greenery in urban areas; basic information on vegetation indices; the significance and challenges of analyzing the obtained image (the histogram threshold problem); and an assessment of the results obtained.

2. Significance of Quantifying Total Vegetation Cover in Urban Areas

Measuring the approximate green cover in urban areas represents crucial data for the analysis and development of various indicators linked to different aspects of city life. For instance, the 'quality of life' is often closely tied to the availability of public and private green spaces (parks, gardens, historic villas, sports facilities, etc.).

Furthermore, environmental quality is intricately connected to the presence and health of vegetation cover in a given location. Beyond these macroscopic aspects influenced by vegetation, there are numerous other, perhaps more cryptic but no less important factors (such as air and water quality, biodiversity, environmental impact assessment, soil consumption, etc.).

The proposed statistical analysis aims to provide data as a fundamental tool for further investigation into the dynamics that regulate large cities. The reliability of the methodology, solidified by numerous experiments and studies, ensures a high quality of the obtained results. Additionally, the ease of repeating index calculations using straightforward algorithms represents a significant advantage. Thus, each set of orthophotos can be compared at different times.

3. Vegetation indices

The point of reference is the set of very high spatial resolution ortho-images released by AGEA (Agency for the Agricultural Supply), characterized by a 20 cm pixel on the ground for urban areas and 50 cm for extra-urban areas.

These images cover the entire Italian territory over a three-year period, and starting from 2012, they have been provided to ISTAT with data in four spectral bands:

Red (R) λ 650 nm;
Green (G) λ 550 nm;
Blue (B) λ 450 nm;
Near Infrared (NIR) λ 800 nm.

Through straightforward Python scripts, we can compute the following indices:

NDVI (Normalized Difference Vegetation Index)[3]: It is certainly the most widely used vegetation index. Its conventional formula, which is based on the behavior of chlorophyll a and b, is:

$$\mathbf{NDVI} = (NIR-RED) / (NIR+RED).$$

Based on all the information provided, NDVI utilizes the RED wavelength as the absorption channel and the NIR as the reflection channel.

To better understand the phenomenon, the ENDVI (Enhanced Normalized Difference Vegetation Index)[4] was also calculated, but it did not yield significant results.

Considering the above, we obtain continuous images (float type), and as such, these files need to be re-classified to extract only the pixels related to areas covered by vegetation. To achieve this, it is necessary to thoroughly analyze the histogram of the obtained image to determine the

[3] For more information: https://it.wikipedia.org/wiki/Normalized_Difference_Vegetation_Index
[4] For further information https://maxmax.com/endvi.htm

threshold value beyond which there is a high probability that the pixel is 'green'.

4. The threshold problem

Extracting a vegetation index from an image, in our case, a mosaic that represents the entire urban area, results in a continuous raster file where it is possible to identify four land cover classes. To better illustrate this concept, the following figures show how the pixels of the vegetation index image are grouped inside the image histogram (Figure 3). The reference image, containing all four 'land cover' classes, is displayed below (Figure 2).

Figure 2 – Image R, G, B of a portion of an urban area in Rome (Castel Giubileo, Villa Spada, Serpentara, Colle Salario)

It is clear that it is quite difficult to detect the exact value of the pixel that uniquely identifies a 'green pixel.' Therefore, we decided to classify the vegetation index image using the Isodata classification algorithm[5] of the ERDAS Imagine© (version 2022) software, by setting to four the number of the classes.

The last cluster is composed of green pixels.

5. First results

Table 1 reports the initial results obtained [6]. The numbers presented in Table 1 highlight the variability that exists among Italian localities. It appears that the residents of Padua and Reggio Calabria are more fortunate than those in Turin and Naples due to a higher amount of green urban space in square meters per capita.

The obtained results can be related to others associated with macro-land cover classes. Considering that in major urban centres, as delineated in the ISTAT BTs (Territorial Basis produced by the Italian National Institute of Statistics), agricultural land is virtually non-existent and water bodies are easily quantifiable, greenery can be correlated with anthropic soil.

That said, there seems to be a correlation between the area of the inhabited locality and the 'anthropic' and 'green' components within it. As expected, their relationship is inverse: the smaller the area of the inhabited locality, the greater the anthropic/green ratio.

[5] ERDAS IMAGINE uses the ISODATA algorithm to perform an unsupervised classification.
Click http://localhost:8080/imaginehelp/html/#/home/unsupervised_isoframe/10/11 to find out more

[6] The population data refer to the number of resident inhabitants by locality published by ISTAT, with the reference year 2011

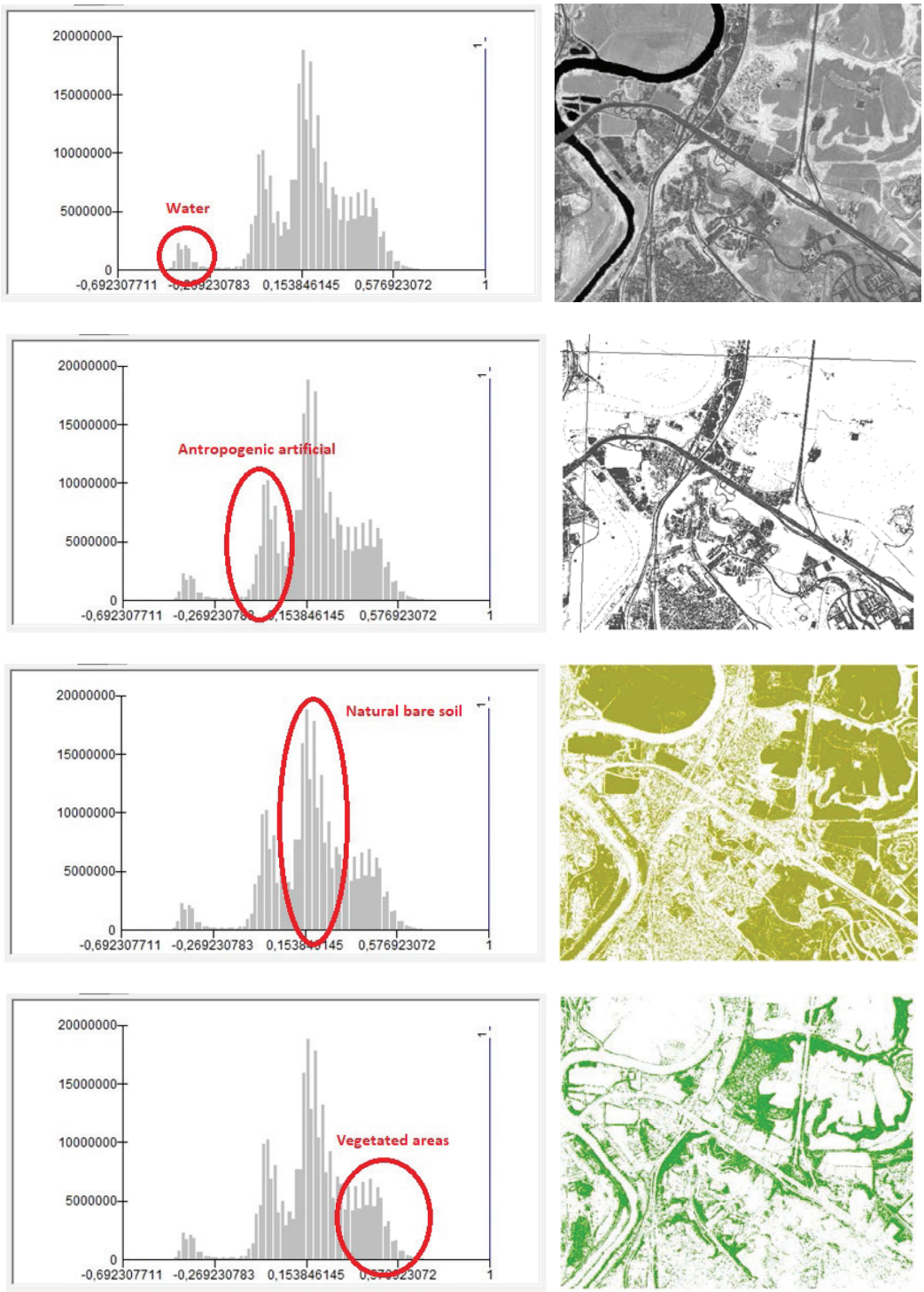


Figure 3 - Histogram partition of the image shown in Figure 2, related to the 'land cover' classes.

Table 1: Statistics about total vegetation areas obtained applying radiometric indices

Urban area	Av. green area (Ha)	Av. m² per capita	Flight year	Green area (Ha)	Flight year	Green area (Ha)	Flight year	Green area (Ha)	Flight year	Green area (Ha)
Turin	2.119,8	24,3			2015	1.834,8	2018	2.404,7		
Milan	3440,5	27,8	2012	3.315,1	2015	3565,9				
Verona	1055,9	48,1			2015	1111,7	2018	1000,1		
Padua	1664,5	81,7			2015	1658,8	2018	1670,2		
Venice-Me	1072,6	72,6	2012	1.045,8	2015	1054,5	2018	1117,4		
Trieste	726,6	38,8	2011	767,1	2014	716,1	2017	696,7		
Genoa	1780,1	30,6	2010	1.726,9			2016	1833,2		
Bologna	2843,5	77,6	2011	2.789,6	2014	2892,2	2017	2827,3	2020	2864,7
Florence	1244,9	35,6					2016	1386,7	2019	1103,1
Rome	9854,9	42,5			2014	10012,8	2017	9697,1		
Naples	2372,8	24,7	2011	2.473,2	2014	2522,9	2017	2276,2	2020	2218,7
Bari	1047,9	37,7	2010	1.053,6	2013	996,5	2016	1148,0	2019	993,4
Reggio Cal	1331,2	78,7	2012	1.258,3	2015	1404,0				
Messina	655,9	29,8			2013	553,9	2016	697,1	2019	716,8
Catania	764,7	26,5			2013	790,6	2016	738,8		
Palermo	1911,2	29,4	2010	1.859,0			2016	1963,4		
Sassari	563,4	61,3							2019	563,4
Cagliari	473,8	32,7					2016	459,2	2019	488,5

Source: ISTAT elaboration on ISTAT and AGEA data

The future developments of this experimentation will undoubtedly focus on two fundamental aspects: 1) Seeking a method as objective as possible to identify the threshold that best discriminates 'green' pixels; 2) Quantifying the areas within urban centres that can represent greenery most significantly for social life, both in terms of actual use of green spaces and their perception. Technically, this latter point is addressed through the correction of the salt-and-pepper effect in the classified images.

6. Conclusion

Certainly, the issues related to the calculation of green space within urban areas are complex and depend on various factors. However, image processing techniques can significantly contribute to addressing these problems.

To this end, ISTAT has recently established a working group, consisting of thematic and IT experts, to experiment with deep learning techniques for the automated extraction of specific objects from orthophotos.

References

Chiocchini R., Ferrara A., Mugnoli S. (2018). Quantificazione e analisi territoriale delle infrastrutture verdi sulla base di indicatori statistici: risultati della sperimentazione per l'area del IX Municipio del comune di Roma Capitale, in *Centro di Ricerca sui consumi di suolo, Rapporto 2018*, INU Edizioni Srl, pp. 156-162.

Xue J., Boafeng S. (2017). Significant remote sensing vegetation indices: A review of developments and applications, *Journal of Sensors*, **2017**, pp. 1-17.

Pristeri, G., Peroni, F., Pappalardo, S.E., Codato, D., Masi, A., De Marchi, M. (2021). Whose urban green? Mapping and classifying public and private green spaces in Padua for spatial planning policies. ISPRS *International Journal of Geo-Information*, **10**, 538. https://doi.org/10.3390/ijgi10080538.

doi.org/10.26398/asaproc.0052

Translation and validation of the Italian version of the academic entitlement questionnaire (AEQ) and learning orientation – grade orientation scale (LOGO II)

Mariaconcetta Nieli[a], Simone Di Zio[b], Lara Fontanella[b], Chiara Berti[a]
[a] Department of Psychological, Health and Territorial Sciences, "G. d'Annunzio" University of Chieti–Pescara, Italy
[b] Department of Legal and Social Sciences, "G. d'Annunzio" University of Chieti–Pescara, Italy

1. Introduction

Academic entitlement (AE) and learning and grade orientation (LOGO) are two important topics in educational psychology.

Until a few decades ago, education and learning were considered opportunities for a student's enrichment and the relationship with the teacher was an opportunity for growth. Instead, more recently, a student as customer paradigm has been consolidated in a context in which universities are competitive market players offering services (Bunzel, 2007). As a result, this paradigm has fostered the beginning of an instrumental and utilitarian approach that may distance students from learning and personal growth (Bunce et al., 2017; Marginson, 2012). This new cultural context may encourage negative attitudes towards education, particularly academic entitlement, and grade orientation.

AE is defined as the tendency of students to expect academic success while investing minimal effort and limiting their responsibility for the educational experience (Chowning and Campbell, 2009; Kopp et al., 2011; Singleton-Jackson et al., 2010). The effects of AE are, to date, widely known and recognized as a problem in the education system. AE involves several negative behaviours, including hostility in relationships between colleagues, conflict between students and teachers, aggression, and low self-control. Students with high scores on AE also underperform in terms of academic achievement (Wasieleski et al., 2014; Reysen et al., 2020).

Research into AE started in the late 1990s. Morrow (1994) defined AE as a change in the value of education which has become a mere assessment of success regardless of the quality of learning produced. Systematic study of the construct began in the early 2000s. Achacoso (2002) defined AE as the students' belief according to which they consider themselves rights holders independent of their own performance and personal involvement.

The first validated academic entitlement questionnaire was presented by Greenberger et al. in 2008. Authors conceptualized AE in just one dimension: the expectation of reward in response for modest commitment and a demanding attitude towards teachers. Subsequently, Chowning and Campbell (2009) proposed a two-dimensional model of AE: externalized responsibility and entitled expectation. Externalized responsibility reflects students' tendency to neglect their own responsibility for their educational experience; entitled expectations reflect the expectation of students to be entitled to a series of rights and protections from teachers and the entire academic body.

Kopp et al. (2011) presented a more complete model of AE that maps five different domains: the right to learn with the least effort, the need for someone to take charge of education, the need to have an adequate teacher to whom any problems in learning are attributed, the right to have control over class policies, the assurance of results as the student pays tuition fees.

The second theoretical construct that has attracted psychologists' attention is the learning and grade orientation. This different approach to learning was first described by Eison et al. (1986) who suggested that university students display a different educational style. This

difference is traced in their different learning style orientation towards learning or towards the grade. This different approach reflects on the involvement that the student experiences during learning. Therefore, students who stand out experience the university classroom as a context in which to experiment with new information and ideas significant for them - both personally and professionally - while there are students who consider university life as a necessary evil to obtain a degree or a diploma. These markedly different perspectives were labelled learning orientation (LO) and grade orientation (GO), respectively. Eison (1986) also reports that learning-oriented students are more emotionally stable, confident, imaginative, self-reliant, and relaxed than grade-oriented students. These students also have less debilitating test anxiety, and they are more collaborative and participatory.

Currently, AE and learning and grade orientation have not yet been investigated within the Italian context. Our study aims to translate and validate the Italian version of the two tools: Kopp's academic entitlement questionnaire (AEQ) and Eison's learning orientation - grade orientation scale (LOGO II). In this paper we focus on the validation of the AEQ scale.

2. Materials and methods

Sample and survey

Two hundred and fifty-four Italian students (female: 84%) completed an online survey on Alchemer. The average age of participants was 23.8 (SD = 4.9). Participants were briefed on all the relevant aspects of the study according to the ethical standards of the Declaration of Helsinki. After accepting the consent form, they completed their demographic variables (age, gender, university, degree course and if they are either on time or have fallen behind the planned number of study years).

Academic entitlement questionnaire (AEQ)

The original academic entitlement questionnaire (AEQ), developed by Kopp et al. (2011), measures the construct of AE. The 26 items reflect the five dimensions of AE: the right to learn with minimal effort, the need for someone to take charge of education, the need to have an adequate teacher to whom any problems in learning are attributed, the right to have control over class policies, the assurance of the result as the student pays tuition fees. Items were answered on a 7-point Likert scale for agreement ranging from 1 (strongly disagree) to 7 (strongly agree). We removed the item "if I have a family vacation scheduled, I should be able to make up work that I miss" from the original scale because it does not fit well in the Italian academic context, appearing implausible. In Italian degree courses that require attendance, it is possible to tot up a certain percentage of absences without incurring in unpleasant consequences for the student's academic career. In addition, participation in the exam session is self-managed by the students who decide during which session to take the exam.

IRT models for scale validation

The process of creating reliable measurement tools relies heavily on the concept of construct validity, which ensures that the structure of the instrument aligns with the intended constructs it aims to measure. Factor analysis, either exploratory (EFA) or confirmatory (CFA), is commonly used to establish this form of validity. EFA helps identify the underlying structure of the test without predefining the number or nature of the factors, while CFA tests a predefined factor structure based on theoretical or empirical knowledge. If the results of factor analysis support the expected relationships between items and factors, it strengthens the argument that the test effectively measures the intended constructs.

In fields like social and behavioural sciences, where data are often categorical, item factor analysis (IFA) offers an appropriate alternative to traditional linear factor analysis. IFA models can be found in both structural equation models (SEM) and item response theory (IRT) literature and also in this context the approaches can be classified either as exploratory or confirmatory. To assess the construct validity of the translated scales, we consider a Bayesian multidimensional item response theory (MIRT) model (Beguin and Glas, 2001). More

specifically, given a test consisting of *K* ordered categorical variables and assuming *M* latent traits, we implement the two-parameter normal ogive (2PNO) formulation of the multidimensional graded response model given by

$$P(X_{ik} = c|\boldsymbol{\theta}_i, \boldsymbol{\alpha}_k, \boldsymbol{\gamma}_k) = \Phi\left(\boldsymbol{\alpha}_k'\boldsymbol{\theta}_i - \gamma_{k,c-1}\right) - \Phi\left(\boldsymbol{\alpha}_k'\boldsymbol{\theta}_i - \gamma_{k,c}\right) \quad (1)$$

where X_{ik} is the observed response of person *i=1,...,N* to item *k=1,...,K*; *c = 1,..., C* denotes the category of the ordered response scale, and Φ is the standard normal cumulative distribution function. Therefore, the probability of responding a certain category *c* depends on the *M*-dimensional vector of the unobserved latent trait scores $\boldsymbol{\theta}_i$, the *M*-dimensional vector of item discrimination parameters or factor loadings $\boldsymbol{\alpha}_k$ and on the (*C*-1)-dimensional vector of ordered category thresholds $\boldsymbol{\gamma}_k$. In the IRT literature, the latent traits are known as person parameters, while the discriminations and thresholds are referred to as item parameters. The factorial structure of the model is represented by the matrix **A** containing the discrimination parameters.

To confirm the construct validity of the developed scales, we began by conducting an exploratory analysis in which all coefficients in the factor loading matrix were estimated. We employed a Bayesian approach to detect a simple structure, using the method suggested by Fontanella et al. (2019). This involved incorporating a sparsity-inducing prior for the factor loadings during estimation to achieve a sparse structure. For this analysis, we exploited the Matlab code provided by Fontanella et al. (2019). Once the sparse structure was identified, we proceeded to estimate the parameters of the confirmatory model using a Bayesian estimation procedure. For the confirmatory analysis, the authors utilized their own MATLAB script.

3. Results

The dataset under examination refers to 24 items, excluding the item "professors should provide their lecture notes online". This exclusion was made due to the fact that 94% of respondents selected the highest category when answering this particular item. Furthermore, we recoded the original 7-point scale in 5 categories, collapsing the first two categories and the last two, in order to ensure a more reliable estimation of the category threshold parameters (Andrich, 1995).

Exploratory IRT model

For determining the appropriate number of factors, we conducted an exploratory analysis that encompassed a range of IRT models with dimensions varying from 1 to 5. The model comparison (Table 1), obtained through the *mirt* package in R (Chalmers, 2012), indicated a preference for a 4-dimensional solution instead of the 5-dimensional solution of the original scale. We labelled the four dimensions as entitlement, voice, attribution of responsibility, and student as customer.

Table 1 Model comparison for dimensionality assessment

Number of factors	AIC	BIC	logLik	$\chi 2$	df	p-value
1	14571.8	14993.9	-7165.9			
2	14520.1	15023.1	-7117.1	97.7	23	0
3	14493.4	15073.8	-7081.7	70.7	22	0
4	14489.8	15144.1	-7058.9	45.6	21	0.001
5	14512.4	15237.0	-7050.2	17.5	20	0.622

Figure 1 displays the hypothesised structure of the AEQ scale and the retrieved structure, along with the discrimination parameter posterior estimates and their 95% credible interval for each factor.

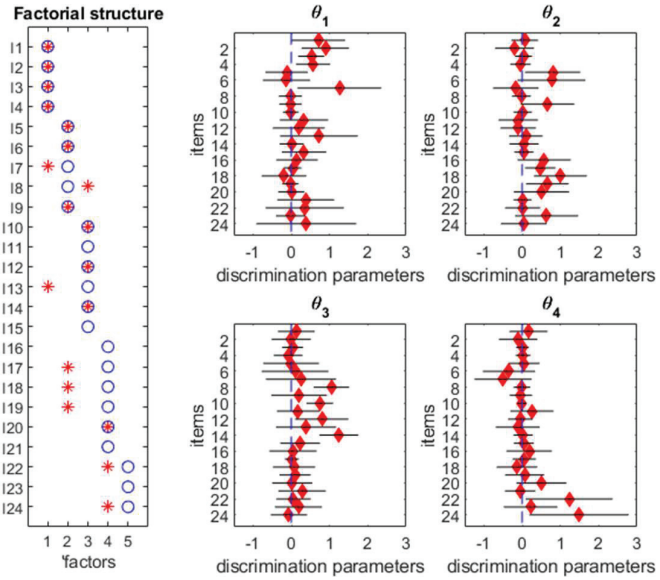


Figure 1: Hypothesised (*o*) and retrieved (∗) factorial structure and posterior estimates of the discrimination parameters (♦). The lines represent 95% credible intervals.

Table 2: Model comparison for dimensionality assessment

Latent trait		$\hat{\alpha}$	95%	CI
θ_1	ITEM1: If a professor does not cover material in class, I should not be expected to learn it.	0.83	0.60	1.09
	ITEM2: Professors should only lecture on material covered in the textbook and assigned readings.	0.60	0.42	0.81
	ITEM3: If I don't do well on a test, the professor should make tests easier or curve grades.	1.14	0.84	1.58
	ITEM4: I focus on learning what is necessary to satisfy the requirements, but no more.	0.49	0.31	0.67
	ITEM5: Professors should not expect me to complete work or study for tests over school breaks.	0.61	0.43	0.82
θ_2	ITEM6: I should have input into how my classes are taught.	0.77	0.54	1.05
	ITEM7: Professors must be entertaining for me to learn.	0.66	0.46	0.87
	ITEM8: Because students are the ones who take classes, they know best what good teaching is.	0.87	0.64	1.14
	ITEM9: I should be given the opportunity to make up a test, regardless of the reason for the absence.	0.58	0.38	0.82
θ_3	ITEM10: Because it is the professor's job to help me learn, if I do not do well, it is the professor's fault.	1.54	1.13	2.32
	ITEM11: It is the professor's responsibility to make it easy for me to succeed.	1.10	0.84	1.43
	ITEM12: If I cannot learn the material for a class from lecture alone, then it is the professor's fault when I fail the test.	0.86	0.63	1.10
	ITEM13: The professor is responsible for how well I do in class.	0.75	0.56	0.96
θ_4	ITEM14: Because I pay tuition, I deserve passing grades.	1.99	1.19	2.85
	ITEM15: Because I pay tuition, I expect to pass the class and get credit.	1.59	1.14	2.19
	ITEM16: I'm paying for my classes, so I should be able to skip class without a grade penalty.	0.83	0.60	1.11

Confirmatory IRT model

The findings from the exploratory analysis reveal a lack of alignment between the composition of the original scale and that of the translated version and 8 out of the 24 analysed items do not load significantly into any of the identified factors. However, considering the

estimated discrimination parameters obtained through a confirmatory approach (Table 2), the retrieved dimensions can be fairly interpreted.

The retrieved dimensions are entitlement, voice, attribution of responsibility, student as customer. The dimension of entitlement includes items from 1 to 5 and it is referred to the right to learn with the least effort. The second factor contains items from 6 to 9 and it regards the voice and the control that students require to have in academic politics. Third dimension, items from 10 to 13, is the attribution of responsibility which switches from students to professors. Last dimension, item from 14 to 16, is referred to the students as customer.

The estimated threshold parameters and the distribution of individuals along the latent traits' continuum are represented in the person-item map (Figure 2) where the selected items are sorted according to ascending order of their location values computed as the mean of the corresponding estimated thresholds. The location parameters can be interpreted as the level of disagreement with the corresponding item. More specifically, the higher the location parameter, the lower the consensus of the students with regard to the related statement.

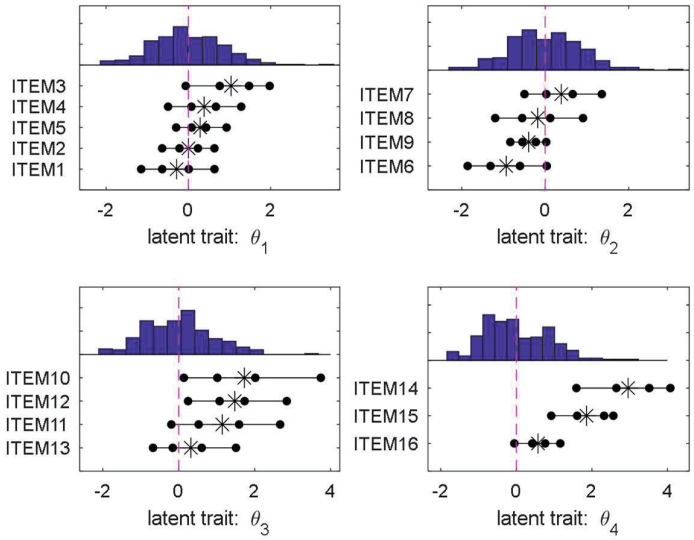


Figure 2: Person-item maps for the identified latent traits. Upper panel of each plot: estimated person parameter distribution. Lower panel of each plot: estimated thresholds (•) and location (*) parameters. The dashed line represents the latent trait mean.

It is worth noting how the items for which there is a very low level of agreement are related to the dimensions representing the student as customer paradigm and the attribution of responsibility. Higher agreement can be observed for the dimensions of entitlement and voice.

4. Conclusion

This study aims to validate the Italian version of AEQ. Our results show that AEQ has a good construct validity, even though the questionnaire we validated differs from the original scale in two aspects. First, we did not include two items in our analysis, and we recoded the scale using a 5-point scale compared to the original 7-point scale. The main difference was the factorial structure of the questionnaire. The original questionnaire has five dimensions, whereas the factor structure recovered in our questionnaire includes four dimensions (i.e., entitlement, voice, attribution of responsibility and student as customer). This structural difference can be attributed to the different cultural context in which AE was studied.

AE is a widespread construct in Western culture but has been little studied in Europe. As far as we know, this is the first application in Italy, and a more thorough study is needed to investigate how cultural influences act in explaining this attitude.

One limitation of our study is the sample size. Further studies are needed to consider more representative samples of the student population in terms of demographic characteristics (i.e., geographical area and degree course).

References

Achacoso, M.V. (2002). "What do you mean my grade is not an A?" An investigation of academic entitlement, causal attributions, and self-regulation in college students. Doctoral dissertation. Retrieved from http://repositories.lib.utexas.edu/handle/2152/2486.

Andrich, D. (1995). Models for measurement, precision, and the non-dichotomisation of graded responses. *Psychometrika*, **60**(1), pp. 7-26.

Béguin, A.A., Glas, C.A. (2001). MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika*, **66**, pp. 541-561.

Bunce, L., Baird, A., Jones, S.E. (2017). The student-as-consumer approach in higher education and its effects on academic performance. *Studies in Higher Education*, **42**(11), pp. 1958-1978.

Bunzel, D.L. (2007). Universities sell their brands. *Journal of Product and Brand Management*, **16**(2), pp. 152-153.

Chalmers, R.P. (2012). Mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, **48**(6), pp. 1-29.

Chowning, K., Campbell, N.J. (2009). Development and validation of a measure of academic entitlement: Individual differences in students' externalized responsibility and entitled expectations. *Journal of Individual Difference*, **101**(4), pp. 982-997.

Eison, J.A., Pollio, H.R., Milton, O. (1986). Educational and personal characteristics of four different types of learning-and grade-oriented students. *Contemporary Educational Psychology*, **11**(1), pp. 54-67.

Fontanella, L., Fontanella, S., Valentini, P., Trendafilov, N. (2019). Simple structure detection through Bayesian exploratory multidimensional IRT models. *Multivariate Behavioral Research*, **54**(1), pp. 100-112.

Greenberger, E., Lessard, J., Chen, C., Farruggia, S.P. (2008). Self-entitled college students: Contributions of personality, parenting, and motivational factors. *Journal of Youth and Adolescence*, **37**, pp. 1193–1204.

Kopp, J., Zinn, T., Finney, S., Jurich, D. (2011). The development and evaluation of the academic entitlement questionnaire. *Measurement and Evaluation if Counseling and Development*, **44**, pp. 105–131.

Marginson, S. (2012). Including the other: Regulation of the human rights of mobile students in a nation-bound world. *Higher Education*, **63**, pp. 497-512.

Morrow, W. (1994). Entitlement and achievement in education. *Studies in Philosophy and Education*, **13**, pp. 33–47.

Reysen, R.H., Degges-White, S., Reysen, M.B. (2020). Exploring the interrelationships among academic entitlement, academic performance, and satisfaction with life in a college student sample. *Journal of College Student Retention: Research, Theory and Practice*, **22**(2), pp. 186-204.

Singleton-Jackson, J.A., Jackson, D.L., Reinhardt, J. (2010). Students as consumers of knowledge: Are they buying what we're selling? *Innovative Higher Education*, **35**, pp. 343-358.

Wasieleski, D.T., Whatley, M.A., Briihl, D.S., Branscome, J.M. (2014). Academic entitlement scale: Development and preliminary validation. *Psychology Research*, **4**(6), pp. 441-450.

doi.org/10.26398/asaproc.0053

Efficiency of Italian courts: Insights from stochastic frontier analysis

Eugenia Nissi [a], Annalina Sarra [b], Adelia Evangelista[b], Tonio Di Battista [b]
[a]Department of Economics, University G. d'Annunzio, Chieti Pescara, Italy;
[b]Department of Philosophical, Pedagogical and Economic-Quantitative Sciences, University G. d'Annunzio, Chieti Pescara, Italy.

1. Efficiency of Italian judicial system

The efficiency of courts is an essential aspect of the justice system, referring to their ability to achieve their objectives with minimal resources. Efficient courts are crucial for ensuring that justice is delivered in a timely and cost-effective manner, as well as for maintaining public trust in the legal system.

Measuring court efficiency is a complex task that requires consideration of various factors, such as the size of the court, the number of cases processed, and the quality of judicial decisions. The efficiency of Italian courts has been a subject of debate and criticism. There are several factors that contribute to this perception:

- Backlog of cases: Italian courts have been plagued by a significant backlog of cases, leading to delays in justice. This backlog is often attributed to complex procedures and lengthy legal processes;
- Lengthy trials: Trials in Italy can be considerably lengthy, causing delays in delivering judgments. This is due to factors such as examinations of evidence, cross-examinations, and the involvement of multiple parties.
- Bureaucracy and administrative issues: The Italian judicial system has been criticized for its bureaucratic nature, which can slow down the process of resolving cases. Administrative issues, such as a lack of resources and procedural inefficiencies, contribute to delays.
- Limited resources: Italian courts often face resource constraints, including a shortage of judges and courtrooms. This can further exacerbate delays in case resolution.
- Complex legal system: Italy has a complex legal system with multiple tiers and jurisdictions, which can lead to confusion and prolonged legal proceedings.

Numerous investigations have been undertaken to evaluate the functioning of the Italian legal system. An overview of the studies carried out on this topic is provided in a table located in the Appendix.

Continuing in this vein of research, this paper delves into an exploration of Italian court efficiency through the lens of stochastic frontiers. Our inquiry initiates with an elucidation of efficiency, specifically as the aptitude of a court to maintain an equivalent level of output with diminished inputs or to amplify output with unaltered inputs. Our measurement approach entails the estimation of a stochastic frontier model, which facilitates the dissection of each court observed output into two discernible segments. The first component is deterministic, attributed to input factors, while the second component is stochastic, representing the unaccounted-for portion of the output.

The paper is organized as follows: Section 2 describes the Stochastic Frontier Analysis while Section 3 contains the data and the empirical results. Finally, some concluding remarks are given in Section 4.

2. Stochastic frontier model

A stochastic frontier model is a statistical and econometric framework used to measure and analyze production or cost inefficiency in various industries and contexts. Its utility becomes particularly pronounced when the aim is to evaluate the performance of a decision-making entity (such as a corporation, agricultural operation, or government body) concerning an ideal or optimally efficient benchmark. Stochastic production frontier models were introduced by Aigner

et al. (1977) and Meeusen et al. (1977). Since then, stochastic frontier models have become a popular subfield in econometrics. A comprehensive introduction can be found in Kumbhakar et al. (2000).

Let us go over the basics of the stochastic frontier problem. Suppose that a producer has a production function

$$y_i = f(x_i, \beta) \quad (1)$$

In a perfectly efficient world, the $i-th$ firm would produce $y_i = f(x_i, \beta)$ without any errors. Stochastic frontier analysis (SFA), however, recognizes that inefficiency exists, meaning that each unit may produce less than its full potential.

In a more detailed context, we express the unit's output as: $y_i = f(x_i, \beta)\varepsilon_i$. Here, ε_i represents the efficiency level of unit i, and it must fall within the interval (0, 1]. If ε_i equals 1, the unit is operating optimally with the given production function $f(x_i, \beta)$. However, when ε_i is less than 1, the unit is not fully utilizing its inputs x_i in line with the production function's capabilities. Since the output is strictly positive (i.e., $y_i > 0$), we assume that the degree of technical efficiency is also strictly positive (i.e., $\varepsilon_i > 0$).

The presence of random shocks affecting the output means that we can represent it as follows: $y_i = f(x_i, \beta)\exp(v_i)$. When we take the natural logarithm of both sides, we get: $\log(y_i) = \log(f(x_i, \beta)\log(\varepsilon_i) + (v_i)$. Given the presence of k inputs and the assumption of a log-linear production function, we can introduce a variable u_i defined as the negative natural logarithm of ε_i. This results in the equation:

$$\log(y_i) = \sum_{j=1}^{k} \beta_j \log x_{ji} + v_i - u_i \quad (2)$$

The rationale for this specification in economic terms lies in the acknowledgment that the production process is influenced by two distinct and economically significant sources of randomness: statistical variability, denoted as v_i, and technical inefficiency, denoted by u_i, representing how far decision-making unit i is from the frontier (due to factors like mismanagement, under- utilization of resources, etc.). The random error term v_i and the inefficiency term u_i are assumed to be independently and identically distributed (iid) with normal and half-normal distribution, respectively, i.e.

$v_i \sim iid\ N(0, \sigma_v^2)$ (3)

$$u_i \sim iid\ N^+(0, \sigma_u^2) \quad (4)$$

3. Case study

To investigate the activity of courts in Italy, we utilized a unique dataset provided by the Italian Ministry of Justice. This dataset includes annual data for the year 2021, encompassing information on the number of pending cases, incoming cases, resolved cases, serving judges, and auxiliary staff members of 140 Italian ordinary courts. The analysis considers four primary input variables: the number of judges (J), administrative employees (A), incoming cases (I), and average duration (measured in days, denoted as D). These inputs contribute to the determination of the output, represented by the resolved cases (S). Therefore, the equation representing the general production function is:

$$Ln(S) = Ln(J) + Ln(A) + Ln(I) + Ln(D) + v_i - u_i \quad (5)$$

In the model all the courts are assumed to operate efficiently. (In)Efficiency among Courts is admitted by subtracting a truncated normal stochastic component from the error component of the classical regression model. The objective is to estimate the parameters of Eq.5 and the distributions of u_i and v_i to understand the level of inefficiency, sources of inefficiency, and the best-fitting frontier function. We utilized Stata commands to estimate the parametric Stochastic Frontier model. It is worth noting that *Stata's frontier module* accommodates three different stochastic frontier models, each with its unique parameterization of the inefficiency term. Additionally, it can be used for both stochastic production and cost frontier models. An overview of the distribution and central tendencies of efficiency scores derived from SFA for Italian courts is presented in Table 1.

Table 1: Descriptive statistics of SFA efficiency scores

Variable	Obs	Mean	Std. Dev.	Min	Max
SFA Eff. Scores	140	0.656	0.171	0.135	0.885

As mentioned previously, the dataset includes 140 observations, and the mean efficiency score is approximately 0.656, with a standard deviation of approximately 0.171. Efficiency scores range from a minimum of 0.135 to a maximum of 0.885, illustrating the variability in efficiency across the examined units. Table 2 provides a comprehensive overview of the estimated coefficients and associated statistical information for various variables in the stochastic frontier model. These variables include LnJudges, Lnadministrative, Ln Duration (days), and LnIncoming, along with their respective coefficients, standard errors, z-scores, p-values, and 95% confidence intervals.

Additionally, it presents statistics for the parameters $ln\sigma^2$ and $ln\sigma^2$, as well as scale parameters (σ_v, σ_u, σ^2) and the λ parameter. This detailed breakdown offers a valuable insight into the model estimation for the total number of resolved cases.

Table 2: Estimation of stochastic frontier: Total number of resolved cases

	Coef.	Std. Err.	z	$P > \|z\|$	[95% Conf.	Interval]
LnJudges	-0.39	0.14	-2.76	0.006	-0.67	-0.11
Lnadministrative	0.44	0.12	3.62	0.000	0.20	0.68
Ln Duration	1.18	0.08	2.15	0.03	0.01	0.35
LnIncoming	0.87	0.03	23.36	0.00	0.80	0.95
ln $\sigma^2 v$	-2.47	0.23	-10.84	0.00	-2.91	-2.02
ln$\sigma^2 u$	-1.70	-0.27	-6.35	0.00	-2.23	-1.18
σ_v	0.29	0.33			0.23	0.36
σ_u	0.43	0.06			0.33	0.55
σ^2	0.27	0.04			0.18	0.35
λ	1.47	0.08			1.31	1.61

We find out that there exists a negative relationship between the number of judges and the overall number of resolved cases: this effect is statistically significant (pvalue< 0.01). The results indicate that the inclusion of more judges in the judicial system may have a temporary adverse impact on productivity. Several factors, such as training duration, procedural changes, heightened complexity, altered priorities, and limited resources, can contribute to this effect. Consequently, there might be an initial decrease in the number of cases resolved.

The coefficient associated to the administrative staff, registering at 0.44, reveals a statistically significant association with the quantity of resolved cases. In simpler terms, as the number of administrative employees increases, there is a corresponding rise in the quantity of resolved cases. This is why an increase in administrative staff can enhance case management efficiency, enabling the judicial system to handle a greater caseload and resolve cases more quickly.

Furthermore, the duration of cases also has a significant positive influence on the number of resolved cases. In other words, as cases persist over a longer period, a greater number of cases are successfully resolved. This can be attributed to the fact that longer durations often allow for more thorough examination, deliberation, and adherence to legal procedures, increasing the likelihood of successful case resolutions. Additionally, longer case durations may encourage parties involved to seek settlements or alternative dispute resolutions, further contributing to a higher number of resolved cases.

The empirical findings also highlight a robust and positive relationship between the number of incoming cases and the number of resolved cases. This relationship is based on the fundamental principle that a larger influx of incoming cases requires increased effort from the

judicial system to address and resolve them. Consequently, it is expected that a higher number of incoming cases would lead to a proportional increase in the number of resolved cases.

The stochastic outcomes derived from Equation 2, recognizing that actual production processes are susceptible to random disruptions or inaccuracies, indicate that the model may not comprehensively address certain sources of randomness or unexplained variation. Additionally, it suggests that there are factors contributing to inefficiency that the current parameterization does not fully capture.

Finally, we conclude by examining the variability in σ_u and σ_v, which offers insights into the consistency of inefficiency across units and the degree of variation in the stochastic component, respectively. Additionally, we explore the impact of σ^2 on observed variations and emphasize the critical role played by the λ parameter, highlighting its significance in shaping our understanding of the total resolved cases. The relatively low variability in σ_u indicates that the inefficiency component is somewhat consistent across units, while σ_v shows moderate variation in the stochastic component. The σ^2 estimate suggests that the composite error term contributes relatively little to observed variations. Lastly, the λ parameter significant influence underscores the importance of the production function shape in understanding the total number of resolved cases.

Table 3: Overview of selected studies on Italian judicial system

Author	Key Observations
Carmignani 2004	This study is aimed at assessing the influence of judicial enforcement efficiency on a company's financial framework.
Felli et al. 2007	This research offers a framework for understanding the "demand for justice" in Italy. The authors utilize a microeconomic model to examine the decisions made by litigants, taking into account the effects of the judicial, legal, and economic systems on their choices. This analysis has the potential to uncover instances of opportunistic behavior.
Carmignani et al. (2010)	This paper explored the connection between the quantity of lawyers and civil litigations within Italian provinces from 2000 to 2005. The empirical findings indicated a positive link between the number of lawyers and litigation, prompting an inquiry into the potential existence of a causal relationship between these variables.
Bripi et al. (2011)	The authors, in collaboration with the Bank of Italy, have extensively researched and evaluated the quality and efficiency of public services in Italy. Their analysis of the justice system, both internationally and at the Italian macro-level, has uncovered notable regional disparities and considerable delays, aligning with previous study findings.
D'Amico et al. 2012	D'Amico and Manca's investigation in 2012 scrutinizes the 2010 proposed reform of the Italian judicial system using multi-state models. The transitions between states are assumed to be influenced by a semi-Markov process. The introduction of the reform is considered a holistic system maintenance policy. The authors evaluate the reform's effects by computing the system's transition probabilities, failure rates, and odds ratios of failure rates.
Peyrache and Zago (2012)	This research evaluates the inefficiency of the Italian judicial system, introducing a methodology that dissects overall inefficiency into technical, size, merger, and reallocation inefficiencies. The findings highlight technical inefficiency as a noteworthy factor, suggesting that, on average, courts could effectively handle almost 40 % more cases by embracing best practices.
Marselli et al. (2004)	The authors pioneered a non-parametric analytical model for the Italian judicial context, using Data Envelopment Analysis (DEA; Charnes et al. (1978)) to pinpoint and quantify inefficiencies, identify effective reduction methods, and estimate potential backlog reduction through more efficient practices.
Antonucci et al. (2011)	Antonucci et al. (2011) employed the DEA model to assess the efficiency of 26 Italian courts of appeal and their proximity to the production frontier. They compared the number of civil and criminal proceedings concluded in 2008 for each court with the allocated resources, including magistrates and costs for wiretapping.
Finocchiaro-Castro et al. (2014)	The authors conduct a thorough two-stage analysis of technical efficiency in Italian judicial districts (individual courts). Their methodology attributes inefficiency to factors associated with the demand for justice and advocacy, and they use the algorithm developed by Simar et al. (2007) to improve estimate accuracy. Unlike most studies in this field, they utilize the DEA technique to examine the justice sector. In the initial phase, they employ a bootstrap procedure for unbiased technical efficiency estimates, and in the second phase, they use a semi-parametric technique.
Ippoliti et al. 2020	Ippoliti et al. (2020) also applied the DEA method, utilizing a two-stage approach inspired by Simar et al. (2007). The first stage involved calculating efficiency scores using this algorithm, while the second stage focused on analyzing factors contributing to inefficiency.

4. Concluding remarks

This investigation was aimed to offer insights into the intricate dynamics influencing the effectiveness of the Italian judicial system. It is worth evaluating the progress and benefits of the proposed methodology in relation to the perspectives presented in earlier studies. Previous research primarily relied on microeconomic models, interpretative frameworks, and DEA models to address inefficiencies in Italian courts. In contrast, our study employs a comprehensive SFA model, offering efficiency estimates and dissecting observed output into deterministic and stochastic components. This enhances our understanding of challenges in the Italian judicial system, surpassing prior limitations. While there is a nuanced interplay between various factors influencing the number of resolved cases, including judges, administrative staff, case duration, and incoming cases, our analysis has provided insights into their relationships and potential impacts. These findings can inform policy decisions and strategies aimed at enhancing the efficiency of the judicial system.

Regarding future investigations in this area, it could be beneficial to examine and refine the spatial variations within the SFA model. This can be accomplished by dividing the inefficiency factor into three distinct components: the first component linked to the unique spatial characteristics of the jurisdiction in which each individual court functions, the second component associated with the particular production attributes, and the third component representing the residual error term.

References

Aigner, D., Lovell, C., Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, **6**1, pp. 21–37.

Antonucci, L., Crocetta, C., d'Ovidio, D.F., Toma, E. (2020). Valutazione dell'efficienza amministrativa del sistema giudiziario tramite Data Envelopment Analysis. *Annali del Dipartimento di Scienze Statistiche Carlo Cecchi. Università degli studi di Bari Aldo Moro. Cleus*, **X**, pp. 281 - 296.

Bripi, F., Carmignani, A., Giordano, R. (2011). La qualità dei servizi pubblici in Italia. *Questioni di Economia e Finanza, Occasional Paper Banca d'Italia*, **84**.

D'Amico, G., Manca R. (2012). A system maintenance approach to analyzing the Italian judicial system reform proposal of 2010. *Socio-Economic Planning*, **46**, pp. 205–215.

Carmignani, A. (2004). Funzionamento della giustizia civile e struttura finanziaria delle imprese: il ruolo del credito commerciale. *Bank of Italy. Economic Research and International Relations Area*, **487**.

Carmignani, A., Giacomelli, S. (2010). Too many lawyers? Litigation in Italian civil courts. *Bank of Italy. Temi di Discussione*, **745**.

Chambers, R.G., Chung, Y., Färe, R. (1978). Benefit and distance functions. *Journal of Economic Theory*, **70**(2), pp. 407-419.

Charnes, A., Cooper, W.W., Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, **2**(6), pp. 429–444.

Falavigna, G., Ippoliti, R., Manello, A., Ramello, G. (2015). Judicial productivity, delay and efficiency: A directional distance function (DDF) approach. *European Journal of Operational Research*, **240**,3, pp.592-601.

Fantini, P., Giacomelli, S., Palumbo, G., Volpe, G. (2011). La litigiosità presso giudici di pace: fisiologia e casi anomali. *Questioni di Economia e Finanza, Occasional Paper Banca d'Italia*, **92**.

Felli, L. Tria, G., Londono-Bedoya, D., Solferino N. (2007). The "demand for justice" in Italy: Civil litigation and the judicial system, in *Italian Institutional Reforms. A Public Choice Perspective*. F. Padovano, R. Ricciuti eds., Springer-Verlag, Berlin-Heidelberg, (DE), pp. 155–177.

Finocchiaro-Castro, M., Guccio, C. (2014). Searching for the source of technical inefficiency in Italian judicial districts: An empirical investigation. *European Journal of Law and Economics*, **38**(3), pp.369 -391.

Ippoliti, R., Tria G. (2020). Efficiency of judicial systems: Model definition and output estimation. *Journal of Applied Economics*, **23**1, pp. 385 - 408.

Kumbhakar, S.C., Lovell, C.A.K. (2000). *Stochastic Frontier Analysis*. Cambridge: Cambridge University Press.

Marchesi, D. (2003). *Litiganti, avvocati e magistrati. Diritto ed economia del processo civile*. Il Mulino.

Marselli, R., Vannini, M. (2004). L'inefficienza tecnica dei distretti di corte di appello italiani: aspetti metodologici, benchmarking e arretrato smaltibile. *CRENOS Working Paper 2004*, **745**.

Meeusen, W., Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, **18**2, pp. 435 - 444.

Peyrache, A., Zago, A. (2012). Large courts, small justice!: The inefficiency and the optimal structure of the Italian justice sector. *Centre for Efficiency and Productivity Analysis (CEPA), Working Paper Series*.

Simar, L., Wilson, P. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, **136**,1, pp.31-64.

doi.org/10.26398/asaproc.0054

Do early freshmen graduate earlier than late ones? Enrolment promptness as an indicator of academic success

Marco Novarese[a], Paolo Chirico[a], Viviana Di Giovinazzo[b]
[a] DIGSPES, Alessandria, University of Eastern Piedmont, Italy
[b] Dept. of Sociology and Social Research, University of Milano-Bicocca, Italy.

1. Introduction

Much of the existing literature on students' academic careers builds on studies concerning procrastination and enrolment patterns. Research on the effect of academic procrastination on student performance is abundant (Solomon and Rothblum 1984; Tice and Baumeister 1997; Ariely and Wertenbroch 2002; Schouwenburg, 2004; Howell et al. 2006; Burger at al. 2011; Bisin and Hyndman 2020), but findings are mixed (see, for example, Rotenstein et al. 2009). The literature on student enrolment patterns is also rich. This has been directed towards better understanding enrolment intensity (Stratton et al. 2008) and the relevance of gender and socio-economic factors (Brunello and Winter-Ebmer 2003; Bozick and DeLuca 2005; Rowan-Kenyon 2007). Much of the enrolment pattern literature relates to North American colleges and universities and is focused mainly on the transition between secondary studies and university.

In the Italian setting, little research has been done. Considering an earlier phase of the student's career, Azzolini et al. (2018) examine the effect of asset-building programmes for socio-economically deprived Italian high-school children on university enrolment and performance over the first academic year. In an exploratory analysis, Novarese and Di Giovinazzo (2013) show that date of enrolment can predict some important aspects of a student's career. Students who are quick to enrol have a lower probability of dropping out during their first academic year, and show both higher grades and a greater probability of graduating. On the contrary, those who enrol close to the deadline have a greater probability of withdrawing within the first year of studies.

De Paola and Scoppa (2015) consider the relations between university enrolment patterns and students' grades from the perspective of procrastination. Examining student enrolment patterns, they discovered that progress during the first two years of university career tends to be negatively correlated with delays in initial enrolment. They also found that late enrolment is positively correlated with the decision to drop out. However, in a related study, De Paola and Gioia (2017) find that students' impatience (determined using responses to a questionnaire concerning hypothetical intertemporal choices) is also correlated with the likelihood of their dropping out. The relationship between procrastination and impatience is explored in Reuben et al. (2015).

In this study we consider both promptness and procrastination in enrolment, with a view to understanding which factors determine these two opposed patterns of behaviour. Beyond that, the goal is to use enrolment data in order to develop a model/procedure that allows us to determine the effect of both promptness and procrastination on the risk of failure to graduate in freshmen. In Section 2 the sample under study is described, and a profile of early and late enrollers is made; in Section 3 a predictive model is tested to identify freshmen at risk; the last section concludes.

2. Early and late freshmen at the University of Eastern Piedmont

The sample under study concerns 7,267 freshmen in the Bachelor degree programmes[1] of the

[1] More specifically, three-year Bachelor's degrees.

University of Eastern Piedmont (UPO) in the years 2001-2010. The freshmen meet the following requirements: (i) enrolled in open-access degrees; (ii) aged under 25 at the time of enrolment; (iii) not transferring from another degree. Freshmen in restricted-access degree programmes (typically in the health/nursing area) are excluded because these courses have short enrolment periods, which are generally respected. For similar reasons, freshmen transferring from another degree programme are excluded. The over-24s are excluded because they are working students whose career is significantly affected by the time they can to devote to study. We focus on first-level (Bachelor) degrees, because enrolment in second-level degrees (Master's) is determined by the date on which the first-level degree is obtained.

In the decade in question, the period for regular enrolment at the University of Eastern Piedmont lasted 9-10 weeks. After this, there was an "overtime" in which enrolment was possible upon payment of an additional fee. Table 1 shows the distribution of enrolments in the regular period and in the overtime period. Since the length of the regular period was not the same every year, this length has been standardized to 1 for each year, so each time interval of width 0.1 corresponds to approximately one week.

Table 1. Enrolment time of UPO's freshmen in the years 2001-2010

time interval	frequency	% freq	cumulative
0.0 - 0.1	156	2.2%	2.2%
0.1 - 0.2	131	1.8%	4.0%
0.2 - 0.3	173	2.4%	6.3%
0.3 - 0.4	317	4.4%	10.7%
0.4 - 0.5	509	7.0%	17.7%
0.5 - 0.6	686	9.4%	27.1%
0.6 - 0.7	915	12.6%	39.7%
0.7 - 0.8	1121	15.4%	55.2%
0.8 - 0.9	1438	19.8%	74.9%
0.9 - 1.0	1520	20.9%	95.9%
>= 1.0	301	4.1%	100.0%

Note: each interval of width 0.1 corresponds to approximately one week.

We can see that only a small percentage of freshmen enrol in the first week (2.2%) and around 10% in the first 4 weeks. Then the weekly enrolments increase until the peak is reached in the last week. There may be various reasons for such behaviour: (i) enrolment appears as a "heavy" choice, probably the first really important choice that these students have to face, a choice that frightens, and is perhaps not entirely a mature choice; (ii) the enrolment period generally begins in the middle of summer, when future freshmen are on vacation and very few are eager to enrol; (iii) enrolment involves an initial payment, which students may wish to postpone.

Based on enrolment time, each freshman is classified according to the following four categories: (i) "*early birds*" for enrolment in the first week of the regular period; (ii) "*ordinary*" for enrolment after the first week and before the last regular week; (iii) "*procrastinators*" for enrolment in the last week; (iv) "*latecomers*" if the student enrols after the regular deadline but before the end of the "overtime" period. Table 2 reports some data on the careers of these categories. The rate of graduation among early birds is almost twice that of latecomers and about 50% higher than that of procrastinators. Moreover, the percentage of dropouts within the first year is significantly lower in early birds (16.4%) than in the other categories: it is almost a third of that of latecomers, who present a dramatic dropout rate: 48.5%! Why are early birds more successful than procrastinators and latecomers? Are they more motivated or do they have superior skills?

If we look at the high-school education of the freshmen (Table 3), we see that the early birds mainly come from classical or scientific Italian high schools, which are considered the most suitable for continuing studies at university; they have obtained a final high school grade of between 80 and 100 in 53.3% of the cases. On the other hand, procrastinators and latecomers have lower

percentages of classical or scientific high school students and higher percentages in the low-average grade classes. It is reasonable to deduce that: the share of "good" students is higher among early birds than among procrastinators and latecomers: good students generally decide in advance which degree to enrol in and do not like wasting time. The deduction is confirmed by the box-plots in Figure 1.

Table 2. Enrolment time and graduation of UPO's freshmen in the years 2001-2010.

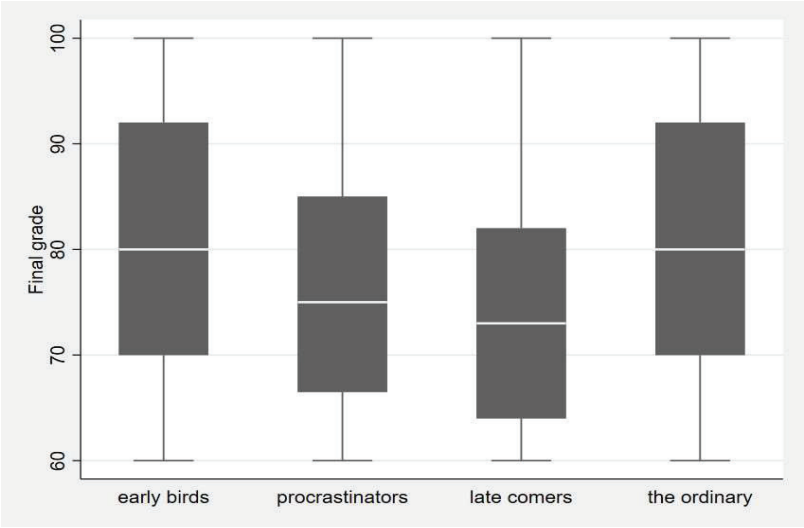
Carrier status	early birds	the ordinary	procrastinators	latecomers	All
not graduating	24.2%	35.0%	50.3%	60.5%	39.6%
- dropout1	*16.4%*	*25.9%*	*40.1%*	*48.5%*	*25.9%*
graduating	75.8%	65.0%	49.7%	39.5%	60.4%
- regular	*60.6%*	*50.3%*	*35.2%*	*29.6%*	*45.9%*
Total	100.0%	100.0%	100.0%	100.0%	100.0%

Note: dropout1=dropout within the first course year"; regular=" graduate within the regular duration of the course".

Table 3. Enrolment time and high school of origin (with final grade).

High School	early birds	the ordinary	procrastinators	latecomers	all
Classical-Scientific	43.0%	34.4%	32.1%	26.2%	33.7%
Technical	33.3%	38.2%	33.1%	36.5%	36.8%
Professional	5.5%	7.2%	10.3%	8.6%	8.0%
Other high school	18.2%	20.2%	24.4%	28.5%	21.5%
Final h.s. grade:					
60-69	20.6%	22.6%	31.9%	37.9%	25.5%
70-79	26.1%	25.5%	27.9%	26.2%	26.1%
80-89	24.2%	21.9%	19.7%	20.6%	21.4%
90-100	29.1%	30.1%	20.5%	15.3%	27.0%
Total	100.0%	100.0%	100.0%	100.0%	100.0%

Figure 1. Box-plot of the high-school final grade in the enrollment categories.



However, the observed differences between freshman categories do not seem to fully explain why early birds, on average, enjoy a better student career. A logistic regression was performed to better understand the relationship between enrolment time and career success (Table 4).

With the exception of the final grade of the high-school diploma, the other explanatory variable

of the model is a dummy variable[2]. In particular, the variable *non-isee* detects who does not present the ISEE certificate at the enrolment, which allows the student to obtain a reduction in university fees. Some freshmen do not present the certificate because their financial state exceeds the limit for having the fee reduction, but others do not present it due to carelessness or misinformation. The variables *economics, law, sciences, humanities* indicate the fields of the degree programmes and are useful for controlling for the difficulty of the corresponding degrees relative to a benchmark degree in political science (which is excluded from the model in order to avoid the collinearity problem).

Ceteris paribus (high-school education, degree difficulty, etc.), early birds are more likely to graduate than "the normal" students: the odds ratio between the early birds and the ordinary is 1.65. This result reflects what the early birds have in comparison to "ordinary" students: stronger motivation and enthusiasm. Conversely, procrastinators and latecomers, as well as non-isee freshmen, are less likely to graduate because they generally have less motivation and conviction.

Table 4. Logistic regression results; dependent binary variable

	B	SE(b)	z	p	exp(b)
Constant	-1.781	0.247	-7.203	0.000	0.168
early birds	0.502	0.204	2.459	0.014	1.653
procrastinators	-0.545	0.063	-8.614	0.000	0.580
latecomers	-0.793	0.139	-5.715	0.000	0.452
non-isee	-1.367	0.083	-16.480	0.000	0.255
high school final grade	0.069	0.002	27.833	0.000	1.071
classical-scientific high school	1.158	0.063	18.347	0.000	3.183
field degree: - economics	-0.615	0.106	-5.811	0.000	0.540
- law	-1.052	0.131	-8.008	0.000	0.349
- sciences	-0.970	0.112	-8.690	0.000	0.379
- humanities	-0.629	0.114	-5.533	0.000	0.533
male gender	-0.120	0.059	-2.035	0.042	0.887

Note: Dependent variable is the career status (1=graduating; 0=not graduating).

3. Random forest classification for "failure" risk

The enrolment data allow us to have an idea of what the student's career will be like: we can use this data to identify students at "failure risk", i.e. at a risk of not graduating, and monitor their career. For this goal there are several statistical tools known in the literature: *logit* and *probit* regression (Stratton et al., 2008), discriminant analysis and classification trees (Rai et al., 2014). In what follows, we briefly illustrate an application of a "machine learning" algorithm which in recent years has been the object of growing interest and widespread use: the random forest classification (RFC, see Breiman, 2001). An RFC algorithm classifies each new unit by taking the most frequently occurring classification provided by an ensemble of "reduced" classification trees (Breiman et al., 1984). Each reduced tree is generated using a random subset of units and explanatory variables. This procedure generally assures lower out-of-sample classification errors than the ordinary classification trees[3].

The RFC algorithm was "trained" using a random sample of 5,100 freshmen extracted from the set of 7,267 freshmen under study[4]; The same variables were used as in the logistic regression described in the previous section. Then the algorithm was used to classify the 2,167 "out-of-sample" freshmen (Table 5).

The correct classification rate of student that fail to get the degree, i.e. "not graduating", is

[2] These variables are equal to one when the student satisfies the condition of the name variable, zero otherwise.
[3] For more details, see Breiman (2001) and chapter 15 in Hastie et al. (2009).
[4] FRC was performed using the Stata module RFOREST (Schonlau and Zou, 2023) with 300 iterations, 3 explicative variables and unlimited deep.

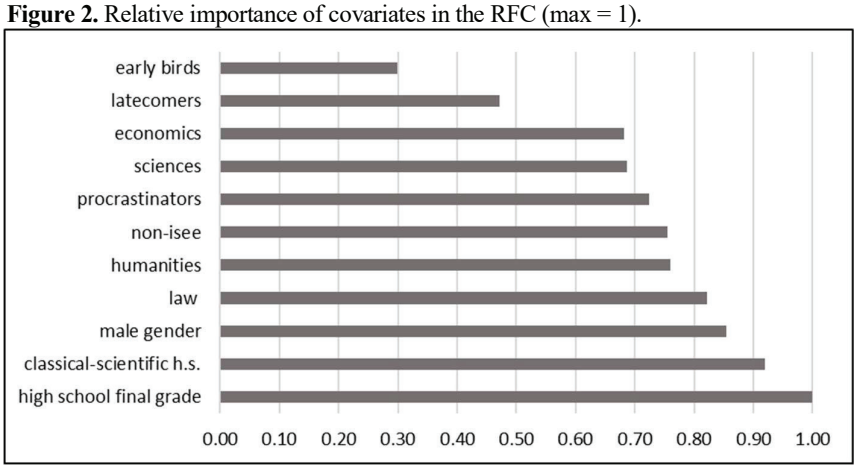
56.8%, which can be considered acceptable at the outset. The RFC classification procedure can be periodically repeated by adding the progressive data of the student's career, such as the number of exams and credits achieved after each exam session. This form of monitoring would certainly provide more and more accurate classifications of the cases at risk.

Table 5. Prediction results of a "random forest" classification (OOB error = 31.6%).

Career status Actual:	Predicted: not graduating	graduating	Total	% of correct predicted
not graduating	497	378	875	56.8%
graduating	280	1,012	1,292	78.3%
Total	777	1,390	2,167	69.6%

Finally, Figure 2 reports the relative importance of the covariates (i.e. the importance of each covariate compared to the most important one) in classifying the freshmen. In this case, "early birds" is the least important covariate. This result does not mean that being early birds (i.e. being more motivated and enthusiastic) is not relevant for obtaining a degree, but that this variable is not particularly useful for classifying freshmen as graduates and non-graduates: it could not be because the early birds are only 2.2% of freshmen. Same explanation for "latecomers", who make up only 4.3% of freshmen.

Figure 2. Relative importance of covariates in the RFC (max = 1).



4. Conclusion

Using data on student registration behaviour and their academic career, this paper showed that freshmen who register early also graduate early because they have better high-school education (i.e. attend a better high-school and achieve a higher final grade), motivation to study and enthusiasm.

No doubt, more research is required in order to strengthen our results, for example, by comparing our data with those on students registered at other universities. However, we believe this study is an important first step for devising aids to prevent dropouts and support students along their academic career.

Acknowledgements This work was supported by funding from the University of Eastern Piedmont.

References

Ariely, D., Wertenbroch, K. (2002). Procrastination, deadlines and performance: self-control by precommitment. *Psychological Science*, **13**, pp. 219–224.

Azzolini, D., Martini, A., Romano, B., Vergolini, L. (2018). Affording college with the help of asset

building: first experimental impacts from Italy. *Economic Letters*, **169**: pp. 27–30.
Bisin, A. and Hyndman, K. (2020). Present-bias, procrastination and deadlines in a field experiment. *Games and Economic Behavior*, **119**, pp. 339–357.
Bozick, R., De Luca, S. (2005). Better late than never? Delayed enrollment in the high school to college transition. *Social Forces*, **84**(1), pp. 531-554.
Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. (1984). *Classification and Regression Trees.* Wadsworth, Belmont, (CA).
Breiman, L. (2001). Random forests. *Machine Learning*, **45**, pp. 5-32.
Brunello, G., Winter-Ebmer, R. (2003). Why do students expect to stay longer in college? Evidence from Europe. *Economics Letters*, **80**(2), pp. 247-253.
Burger, N., Charness, G., Lynham, J. (2011). Field and online experiments on self-control. *Journal of Economic Behavior & Organization*, **77**, pp. 393–404.
De Paola, M., Scoppa, V. (2015). Procrastination, academic success and the effectiveness of a remedial program. *Journal of Economic Behavior & Organization*, **115**, pp. 217–236.
De Paola, M., Gioia, F., (2017). Impatience and academic performance. Less effort and less ambitious goals. *Journal of Policy Modeling*, **39**, pp. 443–460.
Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York.
Howell, A.J., Watson, D.C., Powell, R.A., Buro, K., (2006). Academic procrastination: The pattern and correlates of behavioural postponement. *Personality and Individual Differences*, **40**, pp. 1519–1530.
Novarese, M., Di Giovinazzo, V. (2013). *Promptness and Academic Performance*. https: //mpra.ub.uni-muenchen.de/49746/1/MPRA_paper_49746.pdf.
Rai, S., Saini, P., Jain, A. K. (2014). Model for prediction of dropout student using ID3 decision tree algorithm. *International Journal of Advanced Research in Computer Science & Technology*, **2**(1), pp. 142-149.
Reuben E., Sapienza, P., Zingales, L. (2015). Procrastination and impatience. *Journal of Behavioral and Experimental Economics*, **28**, pp. 63–76.
Rotenstein, A., Davis, H.Z., Tatum, L. (2009). Early birds versus just-in-timers: the effect of procrastination on academic performance of accounting students. *Journal of Accounting Education*, **27**(4), pp. 223-232.
Rowan-Kenyon, H.T. (2007). Predictors of delayed college enrollment and the impact of socioeconomic status. *The Journal of Higher Education*, **78**(2), pp. 188-214.
Schonlau, M., Zou, R.Y. (2020). The random forest algorithm for statistical learning. *The Stata Journal*, **20**(1), pp. 3-29.
Schouwenburg, H.C. (2004). Trait procrastination in academic settings: An overview of students who engage in task delays, in *Counselling the Procrastinator in Academic Settings*, eds H.C. Schouwenburg, C. Lay, T. Pylchyl, and J. Ferrari, American Psychological Association, Washington (US), pp. 3-13.
Solomon, L.J., Rothblum, E.D. (1984). Academic procrastination: frequency and cognitive-behavioral correlates. *Journal of Counseling Psychology*, **31**(4), pp. 503-509.
Stratton, L.S., O'Toole, D.M., Wetzel, J.N. (2008). A multinomial logit model of college stopout and dropout behavior. *Economics of Education Review*, **27**(3), pp. 319-331.
Tice, D. M., Baumeister, R. F., (1997). Longitudinal study of procrastination, performance, stress, and health: the costs and benefits of dawdling. *Psychological Science*, **8**: 454–458.

doi.org/10.26398/asaproc.0055

How direct business official surveys change following the availability of new sources and new technologies

Pasquale Papa, Paola Bosso, Silvana Curatolo
Istat, Rome, Italy

1. Introduction

Recent literature on data collection techniques in official statistics points out a wide convergence towards a multisource approach (De Waal et al., 2020; Salemink et al., 2020). This approach provides for integration of traditional direct survey sources with alternative sources based mainly on big-data (Bender et al., 2022) – such as sensor data, satellite images, scanner data, web sites, data communications – and administrative data. Indeed, many national statistical institutes are making an effort to move from single-source statistics to multisource statistics.

In line with this trend, during the last years, Istat started a process of convergence towards a multisource approach to business statistics. This trend is associated with a review of the role of direct surveys in order to take on agile characteristics, and tends to prioritize the overall quality of the process and sustainability for the statistical system. In fact, the traditional data collection system, mainly based on direct surveys, shows growing impatience and acceptability problems.

In this context, the focus of this paper concerns effectiveness of CAWI technique supported by a centralized contact center service in business surveys and the necessary adjustments for integration into a multisource data collection system. In fact, the agile approach proceeds in parallel with exploiting use of alternative sources both in structured and unstructured forms.

2. Contribution of centralized contact center in design of "agile" surveys

The centralized contact center includes two main services supporting data collection of business surveys: assistance activities (inbound service) and reminder campaigns to non-respondents (outbound service). The services are centralized, multi-channel and managed in outsourcing by specialized suppliers. The contact center service is a flexible tool for modulating agile and personalized data collection strategies based on the different profiles of users.

In business surveys, CAWI (Computer Assisted Web Interviewing) mode represents a compromise between quality and cost. CAWI mode is adapted to needs of business surveys, which often require the collaboration of multiple facilities and roles within the enterprises. It also ensures high coverage due to high digitization of enterprises. One criticality, as pointed out in literature, is the absence of an interviewer, which could negatively affect the response rate and the completeness of questionnaires. A contact center supporting the CAWI single-mode, with compilation assistance (inbound service) and reminders for the recovery of total and partial non-responses (outbound service), can promote response accuracy and improve the response rate in absence of the interviewer.

3. Permanent business census

In the context of multi-purpose statistical surveys for the permanent census of enterprises (carried out by Istat in the period November 2022 - March 2023), this work aims to identify operational guidance aimed at improving quality, agility and sustainability of data collection processes in a multisource integrated management perspective. This process involves a sample of about 278,000 companies, articulated into seven segments, and a complex questionnaire made up of mainly qualitative variables and articulated in two forms (short and long). A motivation for the choice of this surveys is a participation rate lower than expected.

The survey was carried out adopting a CAWI mono-technique data collection, with the support of a specialized inbound and outbound professional contact center and a web-portal for the users, a technique generally adopted for all business surveys conducted by Istat. The objective was to identify criteria for optimizing this technique in relation to agile data acquisition.

For data collection purposes, the sample was divided into seven segments, according to the variables dimension, previous registration to web portal, new entry in the web portal, questionnaire type (long or short form). A detailed description of the identified segments follows.

Large economic units at national level (500+ employees) registered to web portal, with long questionnaire (1,612 companies), 2. Nationally relevant economic units (250-500 employees) registered to web portal, with long questionnaire (2,282 companies), 3. Medium economic units (20-250 employees) registered to web portal, with long questionnaire (62,521 enterprises), 4. Small economic units (10-20 employees) registered to web portal, with long questionnaire (29,753 companies), 5. Other economic units (10 and more employees) not registered to web portal, or new portal with long questionnaire (32,394 companies), 6. Micro economic units (less than 10 employees) registered to web portal, with short questionnaire (44,054 companies), 7. Micro economic units (less than 10 employees) not registered to web portal, or new portal with short questionnaire (105,788 companies). The following Figure 1 reports response rates for each segment of the sample. While in segments 1, 2, 3, 4 ,6 the participation rate was acceptable, according to the standards of the other structural business surveys, the participation found in segments 5 and 7 turned out unsatisfactory, respectively at 28 and 36 per cent. These two segments represent the focus of our analysis, with the aim of drawing general considerations.

Figure 1. Per cent response rates by segments of the permanent business census 2022.

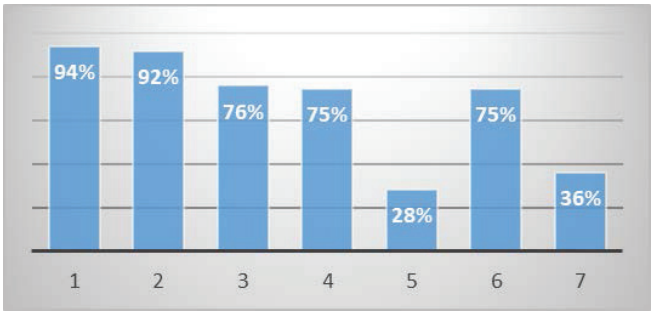


Figure 2. Per cent response rates by segments and registration to web portal state (on the left) and employee class (on the right).

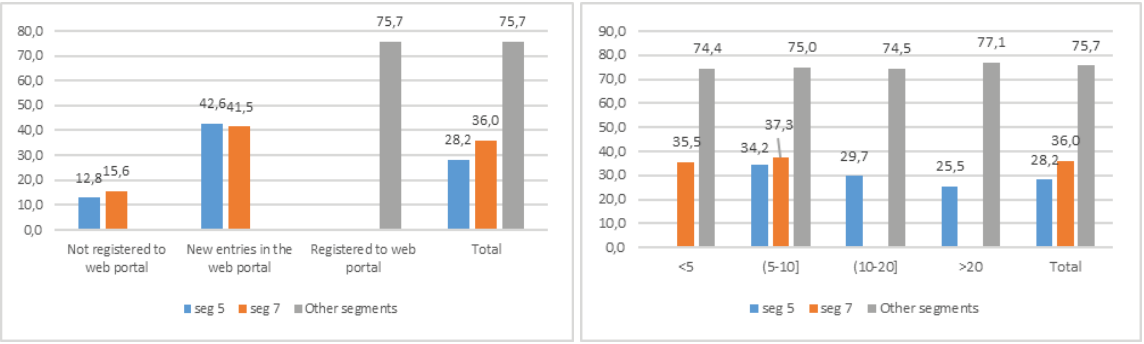


Figure 2 points out the very low participation rates in segments 5 and 7 registered for companies already enabled to access the web portal in the past, but not yet signed in (12.8 and 15.6 per cent respectively) and relatively low for companies newly enabled to the portal (just over 40 per cent). Moreover, companies belonging to segment 5 already enabled to access the portal in the past, but not signed in, show a reduction in the participation rate as the size, measured in terms

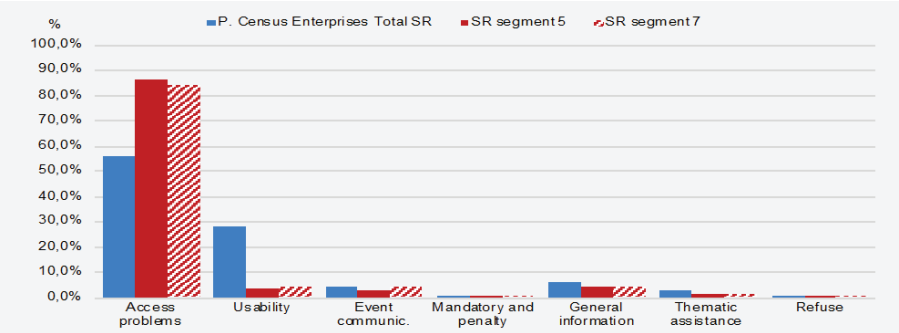
of employees, increases. For these companies, the application of penalties for units having 20 employees or more does not represent a real incentive to participate. This subset is problematic as made up of companies characterized by very low motivation and willingness to collaborate and it represents the first area of attention. As already observed, the set of companies new to the portal also registers very low participation rates, also constituting a specific area of attention.

3.1 Role and efficiency of the inbound service in critical segments

The permanent census developed in total 32.822 service requests. The requests were in 60% of cases by telephone and in 40% by email channel. Average time to handle a request was approximately 7 minutes for telephone support and approximately 5 minutes for e-mail or certified e-mail support. Generally, main reasons of assistance requests from businesses concerned *access problems* (54.3%) and *usability of data acquisition system and the questionnaire* (28.4%). In 90% of cases, contact center operators (I level) resolved directly requests while Istat referents (II level) managed the remaining 10%.

Moving on to the analysis of most critical segments, it emerges that compared to the total number of cases, the problems of segments 5 and 7 focus almost exclusively on *access to the acquisition systems* that is losing or forgetting login credentials and resetting passwords (more than 80% of cases). Massive presence of new companies in the portal in the two segments in question (Figure 3) explains this evidence.

Figure 3. Per cent respondents to the permanent business census 2022, by reason of service requests



Despite the high degree of resolutions at the first level, segments 5 and 7 show a higher rate of cases forwarded to Istat referents (II level). This evidence depends on the presence of complex access problems, as first registrations to the system. Finally, the analysis of the access channels of the requests shows for segments 5 and 7 a wider use of the asynchronous channel, email and certified email. For these business profiles, the asynchronous channel could therefore play an important role in customized outbound and recovery strategies (Figure 4).

3.2 Role of the outbound service in critical segments

The outbound contact center service concentrated on the last two months of the data collection period. It involved 27.2 percent of units in the sample. Only 32.6 per cent of these units had a useful contact[1]. The response rate after telephone reminder was 43.1 per cent. The analysis of most critical segments shows a use of the service below average, respectively to 18.6 per cent in segment 5 and 22.4 per cent in segment 7. Lower percentages resulted for these two segments also in useful contacts rates and in response rates after the reminder. These results pointed out difficulties not only in participation (column e) of these units but also in attempt to make a contact (column d), both of which are significantly lower than average of the sample.

In order to analyze potential effects of the service on response rates in most critical segments, a simulation was hypothesized which provides for the extension of the reminder to all non-

[1] A useful contact is a contact attempt with a definitive outcome.

responsive units (Table 2). The simulation shows (column f) that this strategy would have led to an increase in response rates in the two segments of 3.9 per cent and 2.8 per cent, respectively.

Figure 4. Resolution of service requests (on the left). Channel used by enterprises (on the right). Permanent business census 2022 *(per cent values)***.**

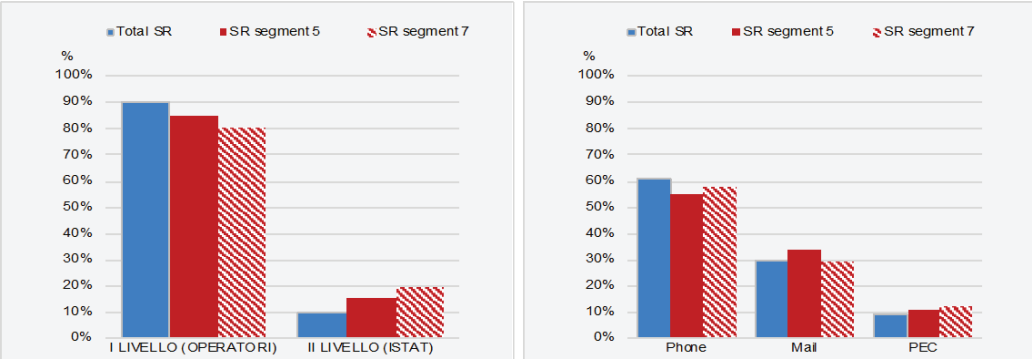


Table 1. General features of outbound telephone reminder service by segment

Segment	Units (a)	Outbound units (%) (b)	Useful contacts (%) (c)	Response rate for Outbound units (%) (d)	Response rate after the reminder (only useful contacts) (e)
1	1,612	62.7	67.4	91.0	58.4
2	2,282	57.5	63.9	87.3	58.8
3	62,521	27.8	48.3	69.3	49.3
4	29,753	30.8	38.2	58.9	47.5
5	**32,394**	**18.6**	**24.8**	**26.7**	**27.0**
6	44,053	38.8	32.1	57.0	40.1
7	**105,788**	**22.4**	**18.0**	**29.8**	**31.5**
Total	**278,403**	**27.2**	**32.6**	**50.1**	**43.1**

Table 2. Simulation of extension of outbound to all non-responsive units of critical segments.

Segment	Units (a)	Spontaneous respondents (%)(b)	Useful contacts (%)(c)	Respondents following a useful contact (%) (d)	Actual response rates (e)	Simulation of extension of outbound service (f)
5	32,394	7,537	24.84	6.7	28.2	32.1
7	105,788	31,059	17.99	5.7	36.0	38.8

4. Results

Application of CAWI data collection technique supported by a specialized contact center service ensures adequate levels of businesses participation in most segments of the sample but not in all. It leaves uncovered some sets of companies that require additional interventions. The analysis of profiles of most critical segments, suggests some targeted solutions.

a) *In weaker segments, there is a clear issue of accessibility to web portal.*

Possible solutions could be the provision of an "enhanced" inbound assistance service and implementation of generalized access methods such as SPID (Public Digital Identity System) or CNS (National Service Card). An enhanced inbound activity consists in focusing attention and resources on requests for assistance coming from most critical segments. Generalized login credentials useful for the various administrative and tax obligations can be of great support. Outbound service collecting any critical issues can also be useful for solving access problems.

b) *Poor understanding and limited practice of participating in Istat surveys.*

The weaker segments mainly include small businesses, which are involved for the first time in official statistical surveys. In the face of a new fulfillment, the company must make an effort

in the initial phase to understand nature of request, use of resources to collect information, management of data acquisition systems, schedule timing, etc. In order to facilitate the required compliance, it is necessary to review the timing, methods and tools connected to the initial involvement of businesses in Istat surveys, providing for personalized and targeted communication methods. Furthermore, specific support is necessary to manage this initial entry phase in the context of Istat surveys.

c) Motivation to participate and awareness of the role required.

The analysis of the segments points out a subset of companies that are scarcely motivated to collaborate as they tend to consider statistical compliance just as a direct cost in terms of time and resources employed. For these companies, awareness of the social role they play is also very limited. This issue requires actions aimed at consolidating Istat authority and awareness about convergence programs towards multisource approaches in management of official surveys. In order to encourage the motivation of companies to participate, it may be important to provide forms of returning statistical information to the companies involved in the surveys. The role of communication campaigns aimed specifically at motivating companies included in critical segments can assume importance in resolving this critical issue.

d) Role of obligation and penalties.

The permanent census provided for penalties for companies having at least 20 employees. Within the critical segments, penalties did not seem to play a significant role in motivating companies to participate. More precisely, within segment 5, that is the only including companies subject to penalties, the participation rate was significantly lower: it is 25.5 per cent for companies with penalties and 30.3 per cent for those without penalties. As already shown in Figure 3, this effect is marked for companies already enabled to access the web portal in the past, but that are still not registered. These specific categories of companies require a more incisive way of conveying information and enhancing assistance on the risks of penalties.

e) Effectiveness of outbound reminder campaigns.

The analysis of the response rate by segments suggests a reorientation of the resources employed on the reminder activity and notably the outbound telephone ones, with a specific concentration on the response probabilities of the subsets of businesses included in critical segments. It is advisable to limit reminders to companies belonging to segments with a high probability of response and concentrate them on the most critical segments. In order to optimize the reminders campaigns the use of adaptive techniques based on response probabilities recorded during the data collection processes is advisable (Bellini et al., 2023).

g) High samples size.

In addition to causing burden for respondents, large samples make data collection process expensive, impersonal and massive, negatively affecting the awareness of respondents and ultimately response rates. Often the high size also derive from oversampling procedures linked to the low response rates expected in some domains. An effective reduction of the sample sizes can derive from the progress of the convergence process towards a multisource approach.

5. Conclusions

In line with the mainstream trends, Istat started to converge towards a multisource approach to business statistics. This trend is associated with a review of the role of direct surveys in order to take on agile characteristics. In fact, with particular regard to business statistics, the traditional data collection system, mainly based on direct surveys, shows growing impatience by involved companies and acceptability issues. This work aimed to identify some operational guidance aimed at improving the quality, agility and sustainability of the data collection process, in a multisource integrated management perspective.

Major issues in terms of participation rates concerned the segments of units having 10 and more employees not registered to web portal, or new portal with long questionnaires and units less than 10 employees not registered to web portal, or new portal with short questionnaires.

Companies already enabled to access the web portal in the past, but never registered, recorded response rates critically low, while companies newly enabled to the portal showed relatively low rates. Moreover, companies already authorized to access the portal in the past but never registered, recorded a reduction in the participation rate as the size in terms of employees increased. For these companies, the application of penalties, starting from the threshold of 20 employees, did not produced the expected effects. This subset is particularly problematic as made up of companies characterized by very low motivation and willingness to collaborate and it represents a specific attention area.

The analysis confirms the relevant role of the centralized inbound and outbound contact center service in the convergence strategy towards an agile data collection process aimed at efficiency and compliance with user needs. The service may be considered an effective support to the CAWI technique as it helps to sustain response rates and reduce bias related to total non-responses.

It also pointed out the issues encountered in the critical segments and the possible solutions to be implemented to reconcile the needs of agility of surveys with that of ensuring adequate levels of quality, in line with the trends in place at all national statistical offices. As an example, the results of a simulation which extends the telephone outbound reminder to all non-responsive units of the critical segments shows a sensible increase in response rates for these segments. At the same time, the use of this service on other segments can be reduced. Results obtained have wide margins of extension to other business surveys conducted by Istat.

Author contributions

P. Bosso: Sections 2, 3.1; S. Curatolo: Sections 3, 3.2, 4d; P. Papa: Sections 1, 4, 5.

References

Bavdaž M., Snijkers G., Sakshaug J. W., Brand T., Haraldsen G., Kurban B., Willimack D.K. (2020). Business data collection methodology: Current state and future outlook. *Statistical Journal of the IAOS,* **36**(3), pp. 741-756.

Bellini G., Monetti F., Papa P. (2020). The impact of a centralized data collection approach on response rates of economic surveys and data quality: The Istat experience. *Statistika, Statistics and Economy Journal,* **100**(1), pp. 1804-8765.

Bellini G., Ballin M., Di Paolo, G.G., Filiberti, S. La Rocca, A. Papa, P. (2023). Pursuing data quality in the multi-purpose statistical survey for the permanent census of enterprises and the adaptive re-contact strategies, *UNECE Expert Meeting on Statistical Data Collection, Rethinking DC, 12 - 14 June 2023.*

Bender, S., Sakshaug J.W. (2022). Data sources for business statistics: What has changed? *The Survey Statistician*, **85**, pp. 10–18.

Biemer, P.P., de Leeuw, E.D., Eckman, S., Edwards, B., Kreuter, F., Lyberg, L.E., Tucker, N.C., West, T.B. (2017). *Total Survey Error in Practice*. John Wiley & Sons Inc, New York.

Binci S., Monetti F., Papa P. (2019) Centralised data collection: Effects of a new administrative penalties provision procedure in business short-term surveys *EESW19, 6th European Establishment Statistics Workshop*, 24-27 September 2019, Bilbao.

Bosso, P., Curatolo, S., Papa, P. (2023). An agile approach to direct official surveys, *UNECE Expert Meeting on Statistical Data Collection, Rethinking Data Collection,12 - 14 June 2023*.

De Waal, T., Van Delden, A., Scholtus, S. (2020). Multi-source statistics: Basic situations and methods. *International Statistical Review,* **88**(1), pp. 203-228.

Istat, (2022). Smart statistics from big data, Available at https://www.istat.it/en/analysis-and-products/smart-statistics-from-big-data (accessed 10 June 2023).

Salemink I., Dufour S., van der Steen M., (2020) A vision on future advanced data collection, *Statistical Journal of the IAOS,* **36**, pp. 685–699.

doi.org/10.26398/asaproc.0056

Health differences across population subgroups: Exploring inequalities through bibliometric techniques

Anna Paterno [a], Maria Gabriella Grassia [b], Thaís García Pereiro [a], Rocco Mazza [a]
[a] Department of Political Science, Aldo Moro University of Bari, Bari, Italy
[b] Department of Social Science, Federico II University of Naples, Naples, Italy

1. Introduction and research aims

Several studies have identified the presence of a socio-economic gradient on both health outcomes and health care access in almost every country of the world. This gradient is acknowledged as a downward slope on differences in health outcomes or health care access across population subgroups and defined as health inequalities. Health inequalities can be guessed as outcomes of a complex combination of individuals' characteristics and their chances to access resources. The former regards the conditions and experiences in which individuals are born, live and work—such as their gender, education, income, social networks, migratory status, and social relations. The latter, instead, deals with their real chances to access resources and decision-making processes, in terms of access to social protection (such as affordable child services or housing, sickness and unemployment protection; etc.), access to quality health services and prevention measures; access to healthy housing and settlement, or access to financial and non-financial services, among others.

Discrimination most often affects women, older people, people with disability, or are based on ethnicity or sexual identity. Discrimination has often a personal basis but may also be caused by social bodies, and this could imply that whole population receive inferior services or has difficulties in accessing health services, which, in turn, could prevent them from enjoying healthier lives.

Previous bibliometric analyses regarding the scientific production on health inequalities have been focused either on a specific region or country or groups of countries (Benach de Rovira, 1995; Almeida-Filho et al., 2003; Ritz et al., 2010), or on health systems or reforms (Macias-Chapula, 2002; Yao et al., 2014). Other recent literature focused on citation practices, together with the most productive authors and journals in health inequalities (Bouchard et al., 2015); north-south gaps in research and international collaborations (Cash-Gibson et al., 2018); and citation space and roles of several factors on health structure (Collyer and Smith, 2020).

As health is a fundamental human right, identifying health inequalities and its main drivers remains essential to achieve health equity. Health equity is achieved when everyone can attain their full potential for health and wellbeing. Research on health inequality is one of the principal sources of knowledge for policy and planning in aged and multicultural modern societies (WHO, 2019).

The main purposes of this paper are, first, to describe the temporal evolution of the amount of academic production focused on health inequalities/disparities/equity during an extended period 1991-2022 and, second, to identify its main research topics and map the specific roles of these topics within the academic production.

2. Data collection and methodology

In this paper we employed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) model proposed by Moher et al. (2009) to retrieve scientific publications (Figure 1). PRISMA offers a comprehensive framework that outlines the criteria for selecting articles in a systematic literature review, ensuring transparent and reproducible selection processes through three distinct phases: the identification of the queries for the search of the documents, the screening of the texts collected through specific filters and the application of these and then the inclusion of the papers in the corpus to be analyzed.

To retrieve the articles, we accessed the Web of Science (WoS) indexing database, specifically the expanded Science Citation Index (SCI) and the Social Science Citation Index (SSCI) WoS Core Collections maintained by Clarivate Analytics. The query used for document search in the WoS database

was the following: health AND inequalit* OR health AND inequit* OR health AND disparit*. The tag searches for the query terms in the titles, abstracts, and keywords fields of indexed documents. Quotation marks were used to retrieve records with exact term sequences, while asterisks served as wildcards to capture term variations.

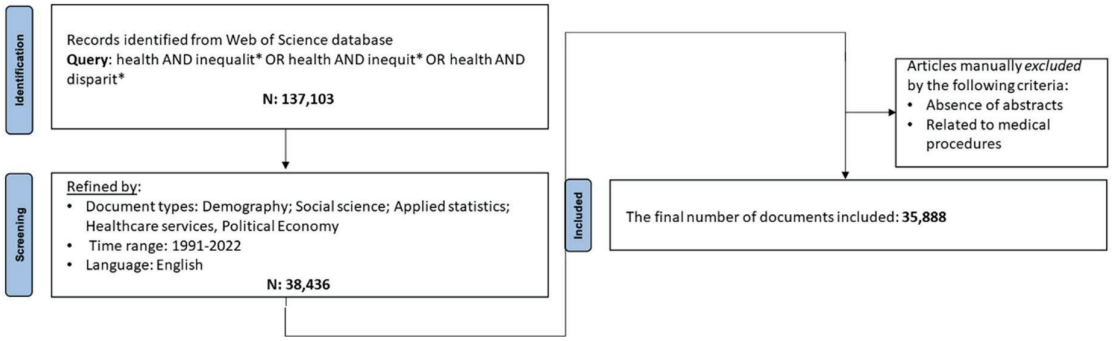
Data were collected in early July 2023. We refined the search by selecting only articles, proceedings papers, review articles, and book chapters published in English between 1991 and 2022, based on the relevance of their content. We collected bibliographic data, including titles, abstracts, author names, keywords, and cited references. The documents were exported to PlainText format and screened by two selectors to include only relevant and coherent documents. After excluding records without abstracts and those focusing on detailed medical procedures or practices, a total of 35,888 documents were retrieved.

To conduct the analysis, we employed the bibliometrix R open-source package (Aria and Cuccurullo, 2017; Aria et al. 2022), which facilitates quantitative research in scientometrics and bibliometrics. We utilized bibliometric analysis to examine the conceptual structure of publications within a specific scientific field, enabling the generation of clusters that provide a comprehensive overview of the research in the field (Borner et al., 2003). To explore the conceptual structure, we performed two complementary analyses: co-occurrence network analysis and thematic mapping. These approaches facilitated the identification of relationships among terms, key research themes, and their development. The degree of similarity between publications was determined by the extent of shared keywords, indicating their association within the same research field. Co-occurrence network analysis (Wang et al., 2019) specifically captured themes represented by sets of terms extracted from documents, such as author and journal keywords. This technique quantified the frequency of term co-occurrence in the document collection and normalized the results using the association index proposed by Van Eck and Waltman (2009). The resulting co-occurrence matrix was represented as an undirected weighted network.

Community detection, performed using the Walktrap algorithm (Pons and Latapy, 2006), identified strongly linked groups of terms sharing common characteristics or playing similar roles within the network. We employed thematic mapping, a two-dimensional representation of network findings proposed by Cobo et al. (2011), to visualize the identified themes. The x-axis represented Callon centrality, indicating the level of significance that a theme holds within a research field, while the y-axis represented Callon density, reflecting the degree of theme development (Callon et al., 1983). This combination of measures facilitated the identification of four types of topics based on their location on the map. The first quadrant of the map represents highly significant and well-developed motor themes. The second quadrant encompass isolated or niche themes with limited external links, resulting in low centrality and limited importance for the broader research field. The third quadrant captures emerging or declining themes, indicating weak or marginal development. The fourth quadrant identifies basic and transversal themes that cut across different research areas.

Each theme was represented as a network cluster on the map, with the bubble name indicating the word with the highest occurrence within the cluster, and the bubble size representing the proportion of word occurrences within the cluster. This way, our study effectively mapped the conceptual structure of the collected scientific documents, thereby revealing significant research topics and trends in the field of health inequalities.

Figure 1: Data retrieval. PRISMA chart for health query.



3. Preliminary results

The information extracted from the utilized library allowed us to outline an overview of the analyzed corpus. There are 4,640 documents, a relatively high number considering the applied filters which might be suggesting a high number of journals are interested in the theme. The average number of citations per document amounts to approximately 28, indicating a prominent activity in the community about this field. The total number of authors is 79,534, with 4,089 documents being single authored.

Figure 2 displays the growth of scientific production over the examined time interval for this analysis, with an overall annual growth rate of 20.46%. The graph shows that the growth in scientific production has significantly growth recently. In fact, until 2008 the number of published papers was 5,007 and, since 2009, in the last 14 years, more than 30 thousand papers on this subject have been published.

Figure 2: Frequency of the scientific production about health, 1991-2022

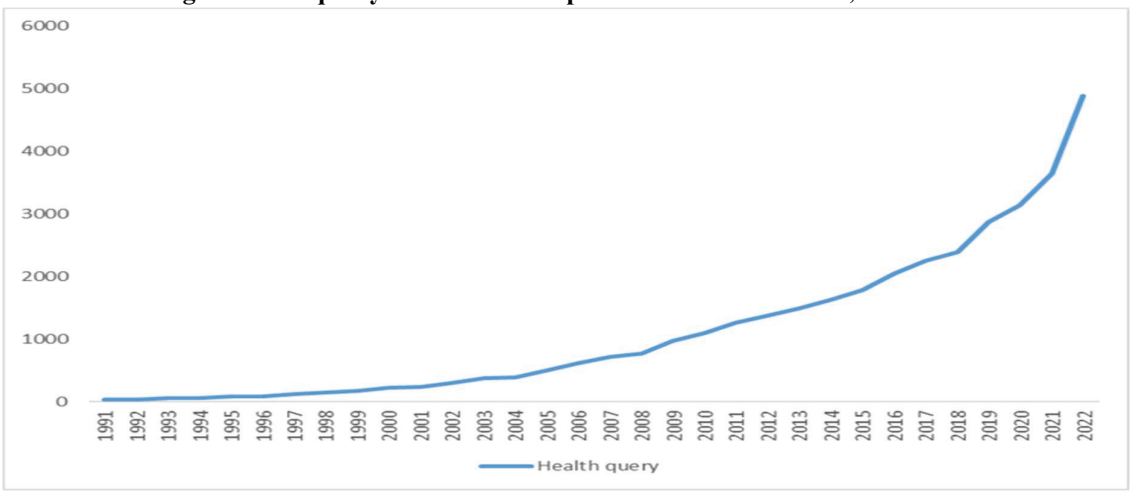


Figure 3: Thematic map of the scientific production on health

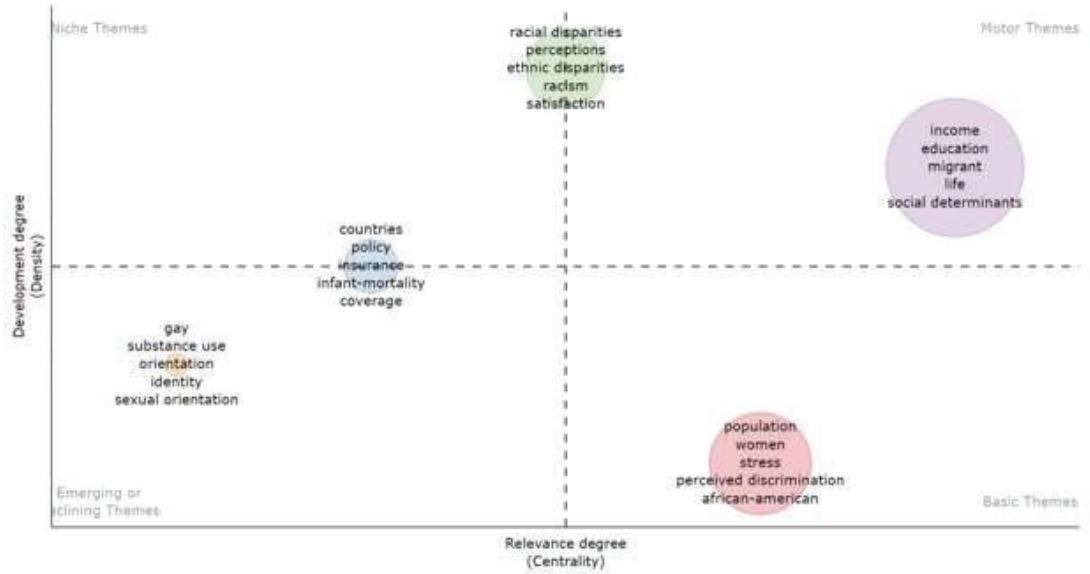


Figure 3 shows the five themes extracted from the papers; the area of a cluster represents the frequency of the cluster. To support the interpretation of results we included Table 1, which displays the values describing network measures that are calculated by this method. The cluster with the highest frequency (35%) is labelled with the keywords "income, education, migrant, life, social determinant". It encompasses motor themes, with high density and centrality. In the second cluster (23%) we find the labels "population, women, stress, perceived discrimination, African-American", which represent basic themes with a transversal dictionary, low density but high centrality. The third cluster (18%) with the labels "racial disparities, perceptions, ethnic disparities, satisfaction" is difficult to classify because it is

between motor and niche themes, it is very high in density but is ranked third by degree. The same difficulty is encountered when analyzing the fourth cluster (14%) labelled "countries, policy, insurance, infant-mortality, coverage". This cluster is between niche and emerging or declining themes. The last cluster (10%) with the keywords "gay, substance abuse, orientation, identity, sexual orientation" seems to represent emerging or declining themes. To clearly identify what the cluster is actually representing, it would be necessary to analyze the time trend through a longitudinal analysis.

Table 1: Network measures for each cluster.

N.	Cluster	Callon Centrality	Callon Density	Rank Centrality	Rank Density	Cluster Frequency
1	population, women, stress, perceived discrimination, African-American	0.033	0.37	4	1	23%
2	countries, policy, insurance, infant-mortality, coverage	0.017	0.40	2	3	14%
3	racial disparities, perceptions, ethnic disparities, satisfaction	0.020	0.47	3	5	18%
4	income, education, migrant, life, social determinant	0.035	0.41	5	4	35%
5	gay, substance abuse, sexual orientation, identity	0.014	0.38	1	2	10%

4. Conclusion and future research

This study has confirmed that socioeconomic inequalities in health mostly has socio-demographic determinants (gender, education, income, migratory status, sexual orientation), but are also related to health care systems coverage. Also, it highlighted the mediation role of health systems, which can act as buffers for inequalities, while supporting the reduction not only of the gap observed across groups of the population but also of its consequences at both the private and public spheres.

Translating the results of these studies into policies remains a main challenge. It emerges the need to look for better and innovative forms of knowledge integration between researchers, policy makers and stakeholders. Once health inequalities were measured (identified, described and analyzed) and the corresponding actions and strategies developed by national and local governments and international organizations, then, it is time to measure its impact.

Further research on this subject -through a bibliometric approach- must deal not only with the identification of the main policy guidelines and interventions proposed and evaluated by the scientific production but also with its evolution over time and across countries.

References

Almeida-Filho, N., Kawachi, I., Filho, A.P., Dachs, J.N.W. (2003). Research on health inequalities in Latin America and the Caribbean: Bibliometric analysis (1971-2000) and descriptive content analysis (1971-1995). *American Journal of Public Health*, **93**(12), pp. 2037-2043.

Aria, M., Cuccurullo, C. (2017). *bibliometrix*: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, **11**(4), pp. 959-975.

Aria, M., Cuccurullo C., D'Aniello L., Misuraca M., Spano M. (2022). Thematic analysis as a new culturomic tool: The social media coverage on COVID-19 pandemic in Italy. *Sustainability*, **14**(6) pp. 36-43.

Benach De Rovira, J. (1995). Análisis bibliométrico de las desigualdades en salud en España (1980-1994). *Gaceta Sanitaria*, **9** pp. 251-264.

Borner, K., Chen, C., Boyack, K.W. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology*, **37**(1), pp. 179-255.

Bouchard, L., Albertini, M., Batista, R., De Montigny, J. (2015). Research on health inequalities: a bibliometric analysis (1966-2014). *Social Science & Medicine*, **141**, pp.100-108.

Callon, M., Courtial J.P., Turner W.A., Bauin S. (1983). From translations to problematic networks:

An introduction to co-word analysis. *Social Science Information*, **22**(2), pp. 191- 235.

Cash-Gibson, L., Rojas-Gualdron, D.F., Pericas, J.M., Benach, J. (2018). Inequalities in global health inequalities research: A 50-year bibliometric analysis (1966-2015), *PloS one*, **13**(1): e0191901.

Cobo, M. J., Lopez-Herrera A. G., Herrera-Viedma E., Herrera F. (2011). Science mapping software tools: Review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology*, **62**(7), pp. 1382-1402.

Collyer, T.A., Smith, K.E. (2020). An atlas of health inequalities and health disparities research: "How is this all getting done in silos, and why?". *Social Science & Medicine*, **264**, pp. 113-330.

Macias-Chapula, C., Mijangos-Nolasco, A. (2002). Bibliometric analysis of AIDS literature in Central Africa. *Scientometrics*, **54**(2), pp. 309-317.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., The PRISMA Group (2015). Linee guida per il reporting di revisioni sistematiche e meta-analisi: il PRISMA Statement. *Evidence*, **7**(6): e1000114.

Pons, P., L atapy M. (2006). Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*, **10**(2), pp. 191-218.

Van Eck, N., Waltman L. (2009). How to normalise co-occurrence data? An analysis of some well-known similarity measures. *Journal of the American Society for Information Science and Technology*, **60**, pp. 1635-1651.

Wang, H., Zhao Y., Dang B., Han P., Shi X. (2019). Network centrality and innovation performance: The role of formal and informal institutions in emerging economies. *Journal of Business & Industrial Marketing*, **34**(6), pp. 1388-1400.

World Health Organization (2019). *Thirteenth General Programme of Work, 2019-2023: Promote Health, Keep the World Safe, Serve the Vulnerable (No. WHO/PRP/18.1)*. World Health Organization.

Yao, R., Zhang, W., Evans, R., Cao, G., Rui, T., Shen, L. (2022). Inequities in health care services caused by the adoption of digital health technologies: scoping review. *Journal of Medical Internet Research*, **24**(3): e34144.

doi.org/10.26398/asaproc.0057

Analysis of urban dynamics and tourists' behaviour using functional data clustering

Selene Perazzini[a], Rodolfo Metulini[b], Maurizio Carpita[a]
[a] DMS Statlab, Department of Economics and Management, University of Brescia, Brescia.
[b] Department of Economics, University of Bergamo, Bergamo.

1. Introduction

The United Nations has recognized the importance of sustainable transportation for achieving the so-called Sustainable Development Goals. In this context, the objective is to enhance transportation efficiency, urban air quality, and public health by reducing greenhouse gas emissions in urban contexts. In order to effectively manage traffic flows, it is crucial to gain insight into the travel behaviours of individuals visiting the cities. In what follows, we aim to investigate the presence of various individual profiles in the central area of the municipality of Brescia.

This research centres around a notable event called "Bergamo-Brescia capitale italiana della cultura 2023" ("Bergamo-Brescia Italian capital of culture 2023") (BGBS23). This initiative promotes projects and activities that celebrate the Italian cultural heritage, thereby encouraging tourism growth and investments. As Italian capitals of culture typically attract a higher number of visitors, transportation might be impacted. Hence, we focus on the first semester of 2023 and analyse the temporal crowding dynamics of five user profiles in the historic city of Brescia. Of particular interest are occasional visitors, who represent the tourist component linked to the event. This study addresses two aspects highlighted in the Italian National Recovery and Resilience Plan - Next Generation Italia 2021-2026: smart infrastructures for sustainable mobility (Mission 3: Infrastructure for sustainable mobility) and sustainable tourism (Mission 1: Digitisation, innovation, competitiveness and culture; Component 3: Tourism and culture 4.0).

To capture crowd movements, we utilize a source of mobile phone data known as "mobile phone density" (MPD), which measures the average number of individuals simultaneously present in a specific census area (ACE) within a given time interval. The use of this type of data in studying people's dynamics is well-established in the literature (Metulini and Carpita, 2021, 2023; Perazzini et al., 2023) due to its comprehensive spatial-temporal details. In fact, mobile phone data has emerged as a valuable resource for the study of people's presence and movements. Nevertheless, their potential for traffic management has yet to be fully explored. Indeed, mobile phone data holds the potential to replace expensive traffic monitoring systems in the future. Leveraging technological advancements, mobile phone-based tools might offer cost-effective alternatives to traditional traffic monitoring systems in the future, facilitating more efficient and sustainable urban transportation.

We examine five user-profiles and employ functional data clustering to identify sets of days exhibiting similar crowding patterns for each profile. For this purpose, we apply two functional data clustering methods, namely k-means alignment (k-mA) and model-based functional data clustering (M-BC), to the 181 observed days (from January to June 2023) for each profile. We then compare the results of these methods and evaluate the outcomes using extrinsic evaluation metrics.

The paper is organized as follows: Section 2 describes the database and the functional data clustering methodologies applied; Section 3 presents the results; Section 4 concludes.

2. Data and methods

The MPD data used in this study were released by TIM company and generously provided to us by the Municipality of Brescia. The database contains information about the average number of

individuals simultaneously present in each ACE of the municipality of Brescia within an hour. In order to generate the MPD data, TIM processes the signals gathered by the antennas and disseminates them across the region in accordance with a spatial model. This process inherently introduces a degree of approximation in the data. Consequently, the data may exhibit diminished precision when considering very small geographical areas, leading TIM to advocate the use of MPD data for regions encompassing at least 400 meters on each side. Consequently, MPD data is typically not made available for areas smaller than ACEs. Moreover, it is essential to highlight that the signals collected by TIM are exclusive to TIM subscribers, and TIM subsequently extrapolates this information to the broader population. This extrapolation also introduces a measure of approximation. Unfortunately, the data does not come with specific measures quantifying these approximations.

Given our focus on the BGBS23 event, we specifically concentrated on the historic centre of the city, where notable monuments and places of artistic and cultural interest are located, namely ACEs 12 and 14. For this area, we have data on the 181 days from January 1st to June 30, 2023, pertaining to five profiles of individuals: residents, commuters, intra-regional travellers, extra-regional travellers, and foreigners. These profiles were defined by TIM based on the areas where the detected individuals mainly spend their nights and working days, as well as on the nationality associated with their SIM card. Figure 1 illustrates that these five profiles exhibit distinct crowding behaviours, and consequently, we analysed them separately.

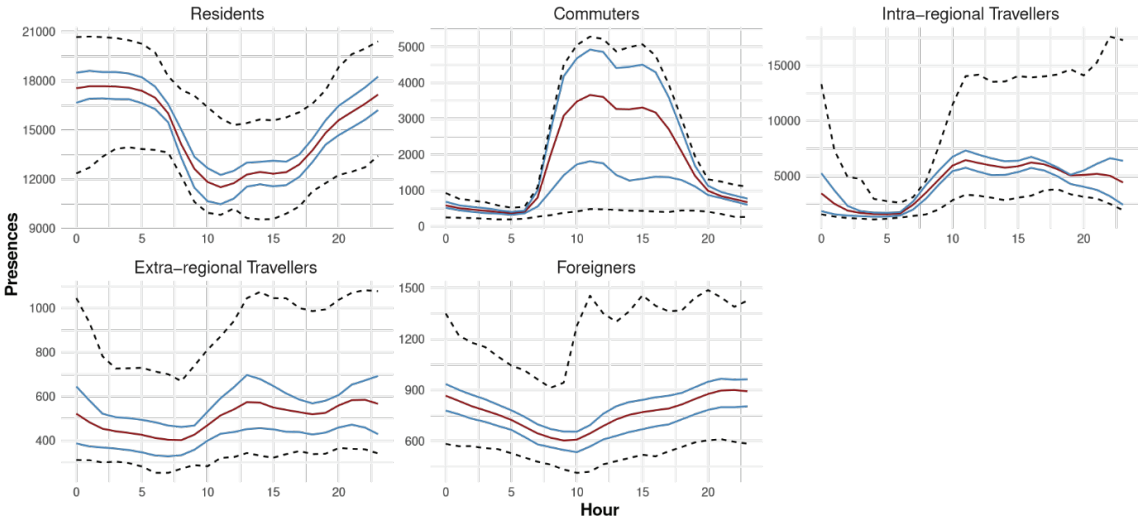


Figure 1: Summary of the set of 181 daily curves (January - June 2023) of mobile phone density for five user profiles. The lines represent the mean (red), the first and third quartiles (blue), and the minimum and maximum (black dotted) presences observed. Note that axes' values are allowed to vary among the profiles.

Each profile's data in the database forms a matrix with 181 columns representing days and 24 rows representing hours. Thus, each observed day can be regarded as a realization of a stochastic process observed at 24 time points. As a result, the MPD data can be treated as functional data. To identify sets of days with similar crowding dynamics, we employed functional data clustering.

The scientific literature offers various clustering methods for functional data. In our study, we concentrate on methodologies designed to estimate proper functional curves, thus excluding the use of "raw data" methods. The remaining clustering algorithms can be categorised as per the classification presented in Jacques and Preda (2013). They fall into two main categories: "distance-based" methods, which rely on adapting traditional distances to functional data, and "filtering" methods, which approximate the curves using basis functions and conduct clustering based on the

coefficients of the basis expansion.

To investigate the dynamics of daily traffic, we employ two clustering algorithms, one from each of these categories, with the objective of comparing the resulting clusters and assessing whether these methods yield similar outcomes. The applied algorithms are: the distance-based "k-means alignment" (k-mA) method proposed by Sangalli et al. (2010), and the filtering method known as "model-based functional data clustering" (M-BC) introduced by Bouveyron et al. (2015).

In the k-mA method, curves are clustered based on an index that measures the dissimilarity between the functional data and a set of warping functions defining the transformation of the abscissa (i.e., the 24 hours). The dissimilarity index used in this study is given by $\left|\frac{x}{|x_1|} - \frac{x}{|x_2|}\right|$, where $|\cdot|$ denotes the norm, and x_1 and x_2 represent two functions. The warping function is defined as $h(t) = mt + q$ with $m \in \mathbb{R}^+$ and $q \in \mathbb{R}$.

On the other hand, the M-BC assumes that the observed set of curves $x_1(t), x_2(t), \ldots x_n(t)$ is generated by an unknown stochastic process $X(t) = \sum_{r=1}^{p} \gamma_r(X) v_r(t)$, where $v_r(t)$ are a set of basic functions with p fixed and known, and $\{\gamma_1(X) \ldots \gamma_p(X)\}$ is a random vector. The M-BC clusters the curves into K homogeneous groups using a discriminative functional mixture model. For this study, Fourier basis functions were employed as $\{\gamma_1(X) \ldots \gamma_p(X)\}$, considering the periodic nature of the data, and p = 9 was chosen based on the sum of root mean squared errors between the process realizations x_i and the smoothed estimated curves evaluated at discrete points in time.

The analyses were conducted using the R packages fdacluster and funFEM for k-mA and M-BC respectively. For each profile, the same number of clusters K was set for the two clustering methodologies to facilitate result interpretation. The optimal value K has been chosen according to the following procedure. First, we imposed a minimum requirement of 10 days per cluster. Then, the optimal number of clusters was separately evaluated for the two methods. For the k-mA method, K_{k-mA} was chosen as the number of clusters associated with a significant reduction of the total dissimilarity index compared to a clustering solution with one less cluster. For the M-BC, K_{M-BC} was selected based on the BIC. Hence, we set the same number of clusters as $K = \min(K_{M-BC}, K_{k-mA})$.

3. Results and discussion

We obtained four clusters for commuters, intra-regional travellers, and foreigners and five clusters for residents and extra-regional travellers. The composition of these clusters in terms of the days of the week is depicted in Figure 2. The figure highlights a significant discrepancy between the two clustering methods. While all clusters constructed using the k-mA method comprise days from all the weekdays, the M-BC tends to group some days together. Specifically, for most profiles, the M-BC identifies one or two clusters consisting mainly of Saturdays and Sundays.

As far as the month's composition of the clusters, most of the clusters appear to equally collect curves pertaining to each of the 6 observed months. Only in the M-BC applied to foreigners and residents emerge some evidence of distinct habits between winter and spring-summer.

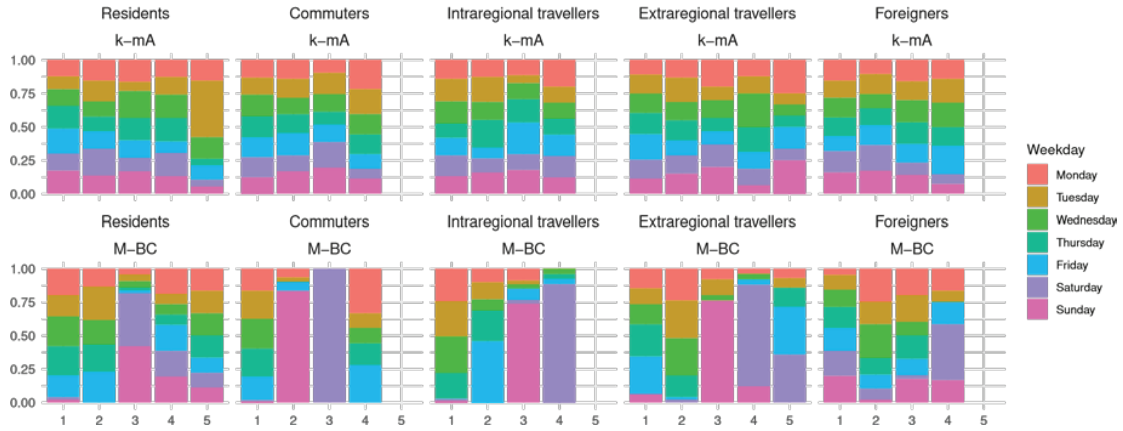


Figure 2: Weekday composition of clusters obtained with the k-mA and M-BC per profile.

To further investigate the differences between the two methods, we computed some extrinsic evaluation metrics. Specifically, we employed the Adjusted Rand Index (ARI) (Hubert and Arabie, 1985) and the Fowlkes-Mallows Index (FMI) (Fowlkes and Mallows, 1983). These metrics are commonly used to assess the similarity or agreement between two clustering applications and are robust to varying cluster sizes, making them suitable for scenarios with imbalanced cluster sizes. Additionally, they account for chance agreement, which refers to the level of agreement that could occur by chance alone between two clustering applications.

The ARI compares the agreement between pairs of data points in both clustering results, considering both true positive and true negative agreements. It ranges from -1 to 1, where 1 indicates perfect agreement, 0 indicates random clustering, and negative values suggest that the agreement is worse than random. The FMI is similar to the ARI but has a slightly different normalization. It ranges from 0 to 1, where 1 indicates perfect agreement, and 0 indicates no agreement between the two cluster techniques. Table 1 presents the values of the two evaluation metrics for each profile of individuals. According to both indices, the two algorithms produce substantially different clusters, with an extremely poor agreement, especially for the ARI.

Table 1: Adjusted Rand index (ARI) and Fowlkes-Mallows index (FMI) comparing the clusters obtained with the k-mA and the M-BC for the five individuals' profiles.

	Residents	Commuters	Intra-regional tr.	Extra-regional tr.	Foreigners
ARI	-0.00	-0.01	-0.01	0.00	0.00
FMI	0.23	0.35	0.29	0.26	0.29

Finally, we conducted a comparison of the clusters obtained for the different profiles. In this respect, we refer to the FMI due to its ability to provide reliable evaluations even when the compared cluster techniques have varying numbers of clusters. The FMI values are presented in Tables 2 and 3 for the k-mA and the M-BC respectively. Regarding the k-mA clustering, the five profiles show little agreement. This suggests that the crowding dynamics of the profiles exhibited substantially different patterns throughout the 181 days of the six-month period. Conversely, the M-BC revealed some similarities among some profiles. The FMI indicates that there are certain resemblances in the clustering of days for commuters and intra- and extra-regional travellers. However, the clusters of intra- and extra-regional travellers display less similarity. Additionally, some weaker agreement was found between commuters and residents. Overall, these findings suggest a weak relationship between the movements of the five profiles on different days. The presence of dissimilar clustering patterns highlights the diverse and distinct crowding dynamics

observed in the city for the various groups during the period of analysis.

Table 2: FMI comparing the k-mA clusters obtained for the five individuals' profiles

	Residents	Commuters	Intra-regional tr.	Extra-regional tr.
Commuters	0.30			
Intra-regional tr.	0.30	0.35		
Extra-regional tr.	0.27	0.28	0.28	
Foreigners	0.26	0.28	0.29	0.28

Table 3: FMI comparing the M-BC clusters obtained for the five individuals' profiles

	Residents	Commuters	Intra-regional tr.	Extra-regional tr.
Commuters	0.52			
Intra-regional tr.	0.42	0.70		
Extra-regional tr.	0.40	0.60	0.49	
Foreigners	0.33	0.37	0.32	0.30

4. Conclusion

In this study, we examined the temporal crowding dynamics of different profiles of individuals in the central area of the municipality of Brescia during the "Bergamo-Brescia Italian capital of culture 2023" event. By analysing mobile phone density data, we identified distinct clusters for residents, commuters, intra-regional and extra-regional visitors, and foreigners.

The clustering of days based on the presence of different profiles of city users holds significant relevance for practical implications in the domains of traffic management and tourism strategies in Brescia. For example, the identification of days exhibiting analogous patterns of commuters' or visitors' presence forms a foundational element for the customisation of traffic management strategies. During high-traffic days, traffic control measures can be implemented, accompanied by the allocation of additional resources and meticulous planning to enhance transportation efficiency. On peak tourism days, public transportation services can be augmented, and additional personnel can be deployed at tourist attractions. On off-peak days, resources can be efficiently redistributed to meet heightened demand in specific areas or reserved for forthcoming high-traffic periods. Tourism authorities can customise marketing campaigns and special events by examining days characterised by similar visitor patterns. This approach can attract tourists during less crowded periods. Moreover, tourist facilities have the capacity to offer tailored services on busier days, such as timed entry arrangements, guided tours, or extended opening hours to accommodate the influx of visitors. Conversely, on quieter days, the focus can shift towards delivering personalised experiences and providing incentives to attract visitors.

The clustering results obtained using the k-means alignment and model-based functional data clustering methods showed substantial disagreement. The extrinsic evaluation metrics confirmed the significant differences between the results of the two approaches, highlighting the need for careful consideration when choosing a clustering method for similar analyses in the context of urban transportation and cultural events. It is noteworthy that, when comparing the two methods, the ARI

attains values close to zero and lower than the corresponding FMI. This occurrence deserves to be further explored (Warrens and van der Hoef, 2022).

Furthermore, the potential reasons for the disparities observed in the results obtained from the two clustering methods are intricate and warrant in-depth investigation. While the distinct characteristics of these clustering algorithms and the rationale underpinning their approaches inherently play a role in generating these distinctions, the main determinants remain non-trivial. We defer these matters for future analysis. At last, the observed dissimilarities in crowding patterns underscore the diverse nature of travel behaviours among the different profiles throughout the study period.

Acknowledgement This study has been developed for the European Union (EU) and Italian Ministry for Universities and Research (MUR), National Recovery and Resilience Plan (NRRP), within the project "Sustainable Mobility Center (MOST)" 2022-2026, Spoke N 7 "CCAM, Connected networks and Smart Infrastructures", CUP D83C22000690001.

References

Bouveyron, C., Come, E., Jacques, J. (2015). The discriminative functional mixture modelˆ for a comparative analysis of bike sharing systems. *The Annals of Applied Statistics*, **9**(4), pp. 1726–1760.

Fowlkes, E.B., Mallows, C.L. (1983). A method for comparing two hierarchical clusterings. *Journal of the American Statistical Association*, **78**, pp. 553–569.

Hubert, L., Arabie, P. (1985). Comparing partitions. *Journal of Classification*, **2**, pp. 193–218.

Jacques, J., Preda, C. (2013). Functional data clustering: A survey. *Advances in Data Analysis and Classification*, **8**, 231–255.

Metulini, R., Carpita, M. (2021). A spatio-temporal indicator for city users based on mobile phone signals and administrative data. *Social Indicators Research*, **156**(2), pp. 761–781.

Metulini, R., Carpita, M. (2023). Modeling and forecasting traffic flows with mobile phone big data in flooding risk areas to support a data-driven decision making. *Annals of Operations Research*, pp. 1–26, online first.

Perazzini, S., Metulini, R., Carpita, M. (2023). Statistical indicators based on mobile phone and street maps data for risk management in small urban areas. *Statistical Methods & Applications*, online first.

Sangalli, L. M., Secchi, P., Vantini, S., Vitelli, V. (2010). K-mean alignment for curve clustering. *Computational Statistics & Data Analysis*, **54**(5), pp. 1219–1233.

Warrens, M.J., van der Hoef, H. (2022). Understanding the adjusted Rand index and other partition comparison indices based on counting object pairs. *Journal of Classification,* **39**, pp. 487–509.

doi.org/10.26398/asaproc.0058

The dynamics of collaboration between researchers of the Italian Institute of Technology: An ERGM approach

Sara Preti [a], Enrico di Bella [b]
[a] Department of Economics, University of Genoa, Genoa, Italy.
[b] Department of Political and International Sciences, University of Genoa, Genoa, Italy.

1. Introduction, data and descriptive analysis

It has been proven that collaboration between authors has positive effects on research, not only for individuals but also for organizations or countries, with benefits that imply greater productivity and innovation. The scientific collaboration network analyzed in this paper is the *Italian Institute of Technology* (IIT) researchers' network. The IIT is one of the primary Italian research centers, publicly financed, it receives approximately 100 million euros per year from the State; additionally, it secures supplementary resources through its involvement in competitive and commercial projects. Currently, the IIT allocates 80% of its total workforce to four main research domains (RDs), Computational sciences, Life technologies, Nanomaterials, and Robotics.

In this study we analyze the collaborative network of researchers at the IIT, focusing on co-authorship as a formal manifestation of intellectual cooperation in scientific research (Acedo et al., 2006). We examine a network considering the authors' research field and socio-demographic characteristics, including gender, nationality, and seniority, that may affect their propensity to form a collaborative relationship. In a graph, the nodes are the 1,469 researchers employed in the Institute's headquarters in 2020, and an edge between two nodes occurs if they have written a paper together. Furthermore, we study the connections between researchers belonging to different RDs, since interdisciplinary research collaboration became necessary to produce advanced knowledge and accelerate innovation.

Finally, an *exponential random graph model* (ERGM) is used to analyze the dynamics of collaboration between authors based on their social and demographic profiles and the structural features of the network. ERGM models the probability of link formation between nodes within a network by studying the configuration of connections between researchers. Currently, link formation represents one of the most promising areas within knowledge and innovation networks (Resce et al., 2022).

The analysis may assess the relevance of sociodemographic characteristics in making collaborative relationships and provide an opportunity for individual researchers to self-evaluate their propensities to cooperate with other scientists. Furthermore, the findings of this study may support the implementation of research policies that may enhance the community organization.

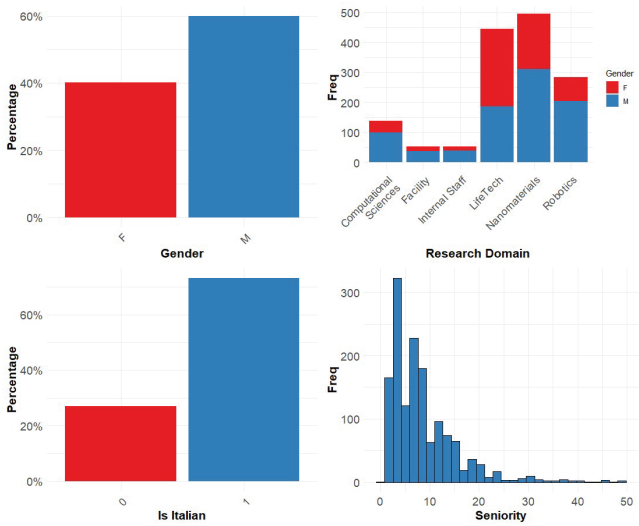
This paper collects information on all the documents published by the IIT researchers, even those published before their hiring at the scientific institute[1]. The starting data source is the full list of researchers with IIT affiliation downloaded using the statistical software *R* through the package called *rscopus* that permitted the extraction of data from the bibliographic database Scopus. This information has been integrated with data provided directly by the Institute, encompassing socio-demographic attributes and variables related to the job position covered by the authors within the organization.

After a cleaning process to address issues related to author name disambiguation[2], we merged

[1] For this reason, the period of interest goes from 1971 to 2020.
[2] The most frequent author name disambiguation problems are the following: surname and given name swapped; full name versus first name use; unicode conversion issue; typos in the author's name.

these datasets and extracted, through the *rscopus* package, the information on all the papers published by each author during their career. Finally, we realized a network starting from these data.

Figure 1: Sample overview of researchers of the Italian Institute of Technology



After the creation of the co-authorship network, we performed a descriptive analysis of the researchers' characteristics. Figure 1 shows a brief overview of our sample. It is worth noting the distribution of the research domains[3] per gender: in some RDs, gender distribution is proportional to the total number of females and males, while others exhibit a gender prevalence.

Since interdisciplinary cooperation fosters innovation, the following section undertakes a community detection analysis to assess the level of heterogeneity within the research communities of the IIT. Finally, to quantify the strength of these intra-group effects and understand how the sociodemographic characteristics of each researcher affect the propensity to form collaborative relationships, an ERGM was implemented.

2. Methods and results

The Institute presents four main research domains, along with the facilities and the internal staff, leading to the classification of researchers into six distinct clusters. Information on the RDs can be used to evaluate *homophily*[4] through the *modularity* measure[5]. In addition to this domain-based node classification, the network's structure can be used to identify an alternative clustering. For graph partitioning, we applied the Girvan-Newman betweenness algorithm that computes the edge betweenness for each link[6], progressively eliminating those with the highest values to detect clusters, as illustrated in Figure 2. Modularity measures are notably high in both partitions: 0.53 for RD-based clusters and 0.65 for the divisive hierarchical clustering algorithm. Consequently, we can assert that the collaboration network of the IIT shows domain-based homophily.

[3] In addition to the four main research areas, we analyzed the group of facilities that are lines of support for the main RDs, and the internal staff.

[4] Homophily refers to the tendency of a node to form links with other nodes sharing similar properties. It is a first demonstration of how the context of a social network can change the formation of the link within it.

[5] Mathematically, the modularity is the fraction of the edges that fall within a cluster *c* minus the expected fraction if edges were randomly distributed. This measure, like a correlation coefficient, ranges between -1 and 1.

[6] *Edge betweenness centrality* is a measure describing the number of shortest paths that go through an edge between pairs of nodes in a graph.

Figure 2: Functioning of the Girvan-Newman betweenness algorithm

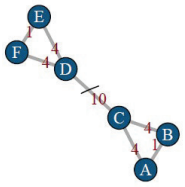


Figure 3: Giant component of the IIT co-authorship network 2006-2020. Note: vertices of the same color belong to the same cluster

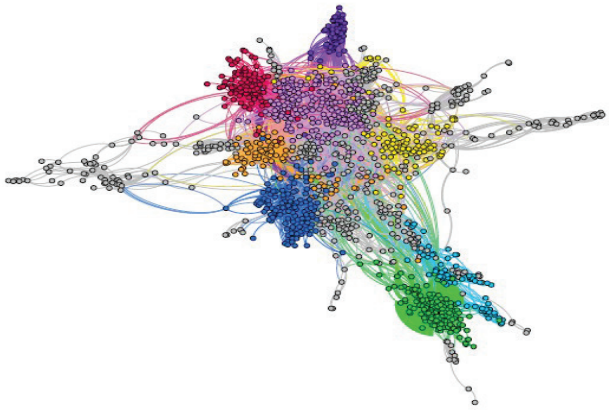


Table 1: Contingency table of two partitions

Cluster	*Computational Sciences*	*Facility*	*Internal Staff*	*LifeTech*	*Nanomaterials*	*Robotics*	*Total*
1	5	15	2	12	183	4	221
3	28	4	8	139	4	8	191
19	3	1	1	1	0	129	135
2	2	0	0	10	85	1	98
6	29	6	0	42	6	2	85
10	1	5	13	27	2	32	80
8	1	14	8	0	0	46	69
9	1	77	3	9	40	0	60
5	1	0	0	58	0	0	59
7	0	0	0	53	2	0	55
4	0	0	3	0	40	0	43
13	0	0	1	1	39	2	43
18	0	0	1	35	0	0	36
15	1	0	1	0	32	1	35
12	0	0	0	0	4	29	33
26	0	0	0	1	20	0	21
22	13	0	0	0	0	0	13
24	0	0	3	2	0	7	12
14	0	0	0	0	0	11	11
28	10	0	0	0	0	0	10

Table 1 reports the results of a two-cluster analysis, confirming researchers' tendency to collaborate within the same domain. The largest cluster mainly comprises Nanomaterials researchers, the second LifeTech, and the third Robotics. Computational sciences authors show wider distribution but a strong tendency to cooperate with the Life tech domain, as exemplified in clusters 3 and 6. Figure 3 illustrates the largest network communities identified by the hierarchical algorithm.

We hypothesized that the attributes of the authors who make up our graph vertices may affect their propensity to form ties. The probability of a tie's existence could be determined by the structural features of the graph. To test these hypotheses and identify the extent to which each variable influences the probability of a link between two authors, we performed an ERGM. This model probabilistically explains network links, clarifying the specific structure over alternative configurations. Equation 1 outlines the general model form:

$$\Pr(Y = y|X) = \frac{\exp(\theta' g(y, X)}{k(\theta)} \quad (1)$$

where y is a network realization in the set of all possible networks Y; $g(y, X)$ is a vector of statistics and additional covariate information about the network; θ is a vector of coefficients for those statistics; $k(\theta)$ represents a normalizing constant that ensures that Equation 1 is a legitimate probability distribution.

The *R* package *ergm*, from the *statnet* suite was used for ERGMs fitting (Hunter et al., 2008). For analysis purposes, four models were developed: two simpler models having as regressors only the sociodemographic attributes of each author and their interaction terms (e.g., Sex*RD or Sex*Seniority)[7], and two models including network structure statistics, such as subgraphs or triadic closure, in addition to the variables of the simplest models[8].

In what follows, we describe the results of the better-fit model according to the log-likelihood and Akaike's information criterion (AIC) values and the implementation of some goodness-of-fit diagnostics for ERGMs[9]. This model controls not only for exogenous attributes but also for endogenous ones.

Table 2 reports the main results, including estimated coefficients, their standard errors, and significance levels. ERGMs operate on a logic analogous to that of a logistic regression model, predicting a binary outcome, i.e., the existence of a link, based on selected predictor regressors that are supposed to explain the observed outcome. The model includes a term for the overall number of edges in y several terms related to nodal properties, and a network statistic variable (*GWESP*). The nodal covariates represent homophily terms, reflecting the growing body of empirical research highlighting the influence of node similarity on collaborative relationship formation (Lü et al., 2009). The *nodematch* and *absdiff* commands for homophily terms permit comparison of log-odds to a reference point or a baseline, as in a conventional logistics model.

Table 2: Estimated coefficients and standard errors for ERGM parameters

Exponential Random Graph Model	
Edges	-35.90(0.83)***
GWESP	24.81(0.64)***
Nodematch Sex (F)	-0.08(0.02)***
Nodematch Sex (M)	0.07(0.01)***
Nodematch RD Internal Staff	0.38(0.13)**
Nodematch RD Computational	0.95(0.04)***
Nodematch RD Facility	1.23(0.07)***
Nodematch RD LifeTech	0.77(0.02)***
Nodematch RD Nanomaterials	0.69(0.02)***
Nodematch RD Robotics	0.56(0.03)***
Nodematch Nationality (Other)	0.06(0.01)***
Nodematch Nationality (IT)	0.23(0.02)***
Absdiff Seniority	0.00(0.00)***
AIC	76538.81
Log likelihood	-38256.40
*** p <0.001; ** p<0.01; * p<0.05	

The baseline (*Edges* in Table 2) represents a collaborative relationship between two authors

[7] These models, which encompass solely nodal covariates as regressors are defined as *dyadic-independence models* since links and their values are stochastically independent given the model parameters.

[8] Conversely, these models are referred to as *dyadic-dependence models* since the presence or absence of a link lead to changes in the network statistics. These models are more complex and require the *Markov chain Montecarlo* (MCMC) algorithm instead of the *Maximum likelihood* for the parameter estimates.

[9] The goodness-of-fit diagnostics consist of performing graphical goodness-of-fit tests. The approach entails simulating a large number of potential networks using estimated model parameters via an MCMC algorithm. Subsequently, various structural measures are computed from these graphs. Typically, these measures encompass degree counts, edgewise shared partner counts, and minimum geodesic distances since these measure phenomena such as sociality, clustering, transitivity, and centrality. Finally, the goodness-of-fit involves comparing the distributions of network statistics from the simulated graphs to the observed graph.

with differing sex, nationality, RD membership, and identical career lengths (Seniority).

Firstly, it is evident that all the regressors hold statistical significance, with RD playing a more relevant role than other variables, aligning with community detection results: Collaboration is more likely to occur among researchers within the same domain. Furthermore, the model highlights a higher probability of collaboration among facilities or computational sciences researchers than in the other RDs. Regarding sex, holding other variables constant, the likelihood of a cooperative relationship between two males surpasses that between a male and a female or between two females; this suggests that females are less likely to cooperate. For nationality attribute, Italian authors are around 1.26 times (in fact, $exp(0.23) \approx 1.26$) more likely to connect in the network (ceteris paribus). Finally, a larger difference in seniority enhances the likelihood of collaboration, making co-authorship more probable between a junior and a senior researcher than between two juniors or two seniors.

Up to this point, we have seen that two authors can match based on exogenous attributes (socio-demographic characteristics); however, it is necessary to consider endogenous attributes (network structural features) as well. For instance, local clustering may arise from actors forming collaborative relationships with shared partners. To represent the network structure more accurately, and since it is well known that triangles abound in social networks, the model also includes a term for transitivity, the *geometrically weighted edgewise shared partnership* (GWESP) statistic. GWESP is a function that uses a curved exponential family form to represent the shared partner distributions[10]. This variable represents a term for triadic closure, or network transitivity, according to which the presence of a link between actors i and j (Y_{ij}=1), and between actors j and k (Y_{jk}=1) increases the probability of a social relationship between actors i and k (Y_{ik}=1$) (Levy et al., 2018) since previous relationships can influence the formation of links in the cluster (Capone et al., 2020). The GWESP statistic is given by the following equation:

$$GWESP(y;\alpha) = e^{\alpha} \sum_{k-1}^{n-2} \{1 - (1 - e^{-\alpha})^k\} \, ESP_k(y) \tag{2}$$

where α is a decay parameter, the higher its value, the slower the decay will be[11]; $ESP_k(y)$ represents the number of edges between two vertices in y with exactly k shared (edgewise) partners; n is the number of nodes in the network; $n-2$ is the maximum number of edgewise-shared partners for any pair of vertices in the network.

In terms of network structure, the magnitude of the triangle statistic coefficient indicates a non-trivial transitivity effect. This implies that a link that closes a triangle is now more likely to occur than one link that does not close a triangle. Therefore, the IIT researchers' collaborative network reflects a triadic structure, indicating a social proximity process where links form between acquaintances and actors connected through an intermediary (Capone et al., 2020).

3. Conclusions

Nowadays, it is essential to maintain ongoing assessments of research centers, encompassing the evaluation of scientific output, inter-researcher collaboration, and the level of openness relative to national and international peers. For this reason, the present study focuses on the Italian Institute of Technology, examining collaboration within the center and across its research domains.

We verified the assumption according to which researchers from the same domain tend to write papers together. However, we also identified a pattern of collaboration between different domains; for instance, the Computational science domain demonstrates a strong inclination to cooperate with the Life Tech research domain. The transversal impact of the Computational

[10] For details on the curved exponential-family models for graphs see Hunter and Handcock (2006) and Hunter (2007).
[11] In the present work, we set a fixed decay parameter $\alpha = 0.25$ following common ERGM modeling practice.

sciences' RD on the other IIT's scientific activities represents a major asset for the research of the Institute since interdisciplinary cooperation could lead to innovative research outputs and outcomes and potentially drive progress in specific scientific areas. The findings provide individual researchers with a means to assess their cooperative inclinations and may inform the development of research policies with potential implications for organizations' communities.

Furthermore, the paper contributes to the literature on network formation. Our results show gender differences in the propensity to collaborate, also searching for potential differences across research fields. Precisely, females are less likely to cooperate within the Institute. The empirical findings of the study offer valuable support to policymakers in designing and evaluating the effectiveness of interventions intended to favor gender equity in collaboration. In addition, the results concerning the network structure variable (GWESP statistic) show that there is also evidence for a non-trivial transitivity effect, and this result highlights the relevance of the role played by trust and previous relationships in the formation of a collaborative relationship. This culture of trust should be extended to the research community to facilitate more effective and productive collaboration.

It is fitting to acknowledge certain limitations in this study. Firstly, co-authorship data is just one indicator of scientific collaboration; not all collaborations result in publications (Sampaio et al., 2016). Secondly, our data collection relied solely on a bibliometric source, suggesting our dataset doesn't cover the entire IIT collaborative network. In fact, other databases may contain publications in different journals or languages, contributing to a more comprehensive representation (Abbasi et al., 2011).

To extend our research, we intend to include additional variables in the ERGM, such as the position covered by researchers within the IIT, to confirm the fact that junior and senior researchers are more likely to cooperate than two juniors or seniors and demonstrate that triangles are not formed exclusively by male researchers in top positions.

References

Abbasi, A., Hossain, L., Uddin, S., Rasmussen, K.J. (2011). Evolutionary dynamics of scientific collaboration networks: Multi-levels and cross-time analysis. *Scientometrics,* **89**(2), pp. 687-710.

Acedo, F.J., Barroso, C., Casanueva, C., Galán, J.L. (2006). Co-authorship in management and organizational studies: An empirical and network analysis. *Journal of Management Studies,* **43**(5), pp. 957-983.

Capone, F., Zampi, V. (2020). Proximity and centrality in inter-organisational collaborations for innovation: A study on an aerospace cluster in Italy. *Management Decision,* **58**(2), pp. 239-254.

Hunter, D.R., Handcock, M.S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics,* **15**(3), pp. 565-583.

Hunter, D.R. (2007). Curved exponential family models for social networks. *Social Networks,* **29**(2), pp. 216-230.

Hunter, D.R., Handcock, M.S., Butts, C.T., Goodreau, S.M., Morris, M. (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software,* **24**(3), nihpa54860.

Levy, M.A., Lubell, M.N. (2018). Innovation, cooperation, and the structure of three regional sustainable agriculture networks in California. *Regional Environmental Change,* **18**, pp. 1235-1246.

Lü, L., Jin, C.H., Zhou, T. (2009). Similarity index based on local paths for link prediction of complex networks. *Physical Review E,* **80**(4), 046122.

Resce, G., Zinilli, A., Cerulli, G. (2022). Machine learning prediction of academic collaboration networks. *Scientific Reports,* **12**, 21993.

Sampaio, R.B., Fonseca, M.V.D.A., Zicker, F. (2016). Co-authorship network analysis in health research: Method and potential use. *Health Research Policy and Systems,* **89**(2), pp. 1-10.

doi.org/10.26398/asaproc.0059

Sustainability in business practice: Price and value

Ercolino Ranieri[a], Andrea Guizzardi[b]
[a] Xenia S.p.A. Società Benefit, Guardiagrele (CH), Italy.
[b] Department of Statistical Science, University of Bologna, Bologna, Italy.

1. Introduction

Sustainability has become an essential value for businesses, not just an ethical choice. Legislators, finance institutions, customers and markets are pushing companies to be ESG (environmental, social and governance) compliant. Rules both mandate (e.g., mandatory waste separation) and incentivize (e.g., tax breaks for energy-efficient structures) sustainability. Banks condition access to credit on compliance with ESG parameters, while finance is giving importance to these parameters when assessing creditworthiness or selecting investments. Customers, especially younger generations already seek "green" products and services with specific sustainability characteristics. Even market and competitive condition are forcing companies to choose sustainability to differentiate oneself from competitors and enhance attractiveness to talent in the job market.

Sustainability is indeed a value for a hotel, but it also comes with a cost, as it generally entails higher expenses related to raw materials (management and logistics), staff (training and salaries) and a worse net financial position. Moreover, it is challenging for a hotel to justify charging a higher price for a sustainable stay, especially when compared to many eco-labeled products (such as an electric car) that already have the support of an increasing proportion of customers, despite their higher price (compared to a diesel car). This draws a distinction between services and products, where it is easier to turn the higher price of a sustainable product into a customer benefit than it is for a service.

Focusing on services, it is even more evident that sustainability is a holistic concept that cannot be limited to the environmental dimension and, therefore, to the letter E of the acronym ESG, even though, today, the topics of climate change, energy and pollution are prevalent in the common feeling. There is an equally powerful movement of attention towards social and ethical as well as deontological topics that is similarly characterizing the new industrial relations between workers and companies, between consumers and products/services and between economy, finance and the community at large. These elements are no longer disjointed, and it is not possible, for those who wish to achieve sustainability goals, to focus on just one of them. That is why we believe it is important to revisit the most recent literature on sustainability in the accommodation industry from the perspective of a medium-sized hotel chain that has recently completed the journey to become a (certified) Benefit corporation. The idea is to present environmental, social, ethical, and deontological topics under two main themes: the cost of sustainability (its price) and the value of sustainability.

This holistic approach encompasses the concept of "benefit for all," as a synthesis of a regenerative paradigm of business (not only reducing the impact on the environment but also producing positive effects, by creating shared value in the society as well as regenerating the biosphere) as opposed to an extractive one (where a business uses more economic, social and environmental inputs than it is able to produce as its output while operating.

By combining literature and field experience, we will try to understand if the value of sustainability will outweigh the costs or, at least, the profit expectations, which is a crucial aspect for any enterprise. Our thesis is that, for a medium-term period, additional costs will outperform additional revenues thus it may still be possible to witness a contrast between accommodation

enterprises that strive for sustainability and those that consider it merely a formal requirement. However, there will come a time when there is no room for those who have not built sustainable models. Accommodation businesses will soon no longer have the choice of whether or not to be sustainable; they will have to be.

2. The value

Competitive advantage

Due to the severe competition among hotels, developing a personalized image has become critical in the positioning schemes. There is a strong relationship between green practices adopted by hotels and the green image of hotels. Hameed et al. (2022) find that the green image of a hotel (built on environment-friendly operations) positively impacts consumer green satisfaction and green trust, and these two variables have significant influence over the green word-of-mouth and hotels' physical image, as it emerges from digital platforms (Galati et al., 2021). 4 and 5-star (luxury) hotels are considered the most sensitive market segment. Many studies report a positive influence of green image on guest's overall online evaluation, guest's revisit-intentions and their intention to pay a premium price, stronger among Gen Z customers (Thomas, 2022).

Bianco et al. (2023), considering data from 251 certified hotels located in Florida, United States showed that certified hotels can increase their average daily rate and revenue per available room (RevPAR) - compared to competitors - through a first-mover advantage. If a competitor is certified, the best response is to become certified as well but the response effect performance as second movers are better than late movers. They also showed that certifications can create competitive advantage, mostly for small-size luxury and upper-upscale hotels. Large hotels with 300–500 rooms had the lowest gains from certification.

Price premium

Price is a pivotal variable of hotel's ability to increase its economic performance (Guizzardi et al., 2019). However, although thousands of non-market valuation studies have been published and a plethora of studies have been conducted on consumers' willingness to pay a premium for staying at green hotels or for supporting green initiatives in the hotel sector, only a limited number of studies have estimated the monetary value of the premium. A recent systematic literature review (Damigos, 2023) – partially – fills this gap exploring the premium that consumers would be willing to pay to stay at a green hotel as results from 22 articles from 13 countries (published in the range 2007-2022). The article reveals that the percentage of customers willing to pay more for a stay at a green hotel varies between 5.4% and 91.6%, depending on the country being investigated and the valuation scenario. The average value is 53.2% (sd = 26.5%). Those who are unwilling to pay more typically argue that the financial benefits derived from energy and resource savings are sufficient, and therefore, no additional support from customers is necessary. Consequently, they believe that there should be no price difference between green and traditional hotels, or they expect a discount to incentivize their engagement in environmental practices.

According to Damigos (2023), consumer preferences, both revealed and stated, indicate an acceptable premium ranging from 4% to 6%. However, this varies depending on the specific sustainability action. Energy efficiency interventions, water consumption reduction, water reuse, waste recycling, and effective management policies have a premium range of 4.5% to 5.2%. The impact of renewable energy supply is more variable, ranging from 1% to 5%. Several factors positively influence the premium that consumers are willing to pay. These include star rating and age, with both customers staying in more luxurious hotels and younger customers showing a greater willingness to pay for green hotels or initiatives (although few studies did not find statistical significance). Income level is also a contributing factor. Nationality and environmental awareness do not show clear evidence of their influence, with some studies finding significance while others do not. However, there is a general consensus that educational level, number of children, and marital status have no influence on the willingness to pay a premium.

Market value and finance

Being sustainable means being transparent, and transparency is an essential value for investors. It is not a minor aspect of sustainability, yet it is often overlooked. This is why ESG issues have gained significant importance, even within the European Community, and they are considered on par with financial reporting in terms of significance and decision-making influence, especially in large companies. It is not surprising that a recent study, involving a sample of major publicly traded hotels in the United States, discovered a positive correlation between sustainability certifications and the market value of hotels (Bernard and Nicolau, 2022).

3. The costs (sustainability is difficult to sustain)

Sustainability also entails costs that the tourism sector will not easily absorb. It involves integrating new elements into existing business models, requiring investments in production, governance, and social commitment that often bring about revolutionary changes to the corporate context. The transition towards sustainability in the production process necessitates actions like eliminating plastic usage or sourcing from local suppliers, which radically change the way both core services (like rooms) and ancillary services (like breakfast or spa) are delivered. Significant investments in innovation, research (for materials and processes), and training (for personnel) are required, leading to inevitable cost increases as we discuss in the following.

Production and management costs

Studies conducted by various eco-labeling agencies claim that green certifications can lead to cost savings in energy, water, maintenance, and waste management, suggesting a reduction in production costs (Chi et al., 2022). However, an aspect that is often overlooked (even in the scientific literature), is that the investments required to achieve these cost-reductions can be substantial, posing a challenge in terms of financial sustainability, especially when interest rates are high. While implementing energy efficient measures, such as a low consumption boiler may help to save costs (depending on exogenous conditions, e.g. fuel prices), it will surely either increase the volume of passive interest or decrease the volume of active interest. Furthermore, when a hotel adopts energy-saving systems, soap dispensers, low lighting, and portion control at breakfast to reduce food waste, it faces a significant increase in costs associated with sourcing suppliers, implementing the breakfast setup, and acquiring the necessary raw materials.

The challenge presented by these costs is to maintain profitability at a steady level. Certified hotels are likely to raise their prices, considering that customers are willing to pay more for hotels that embrace sustainable initiatives (Casado-Díaz et al., 2020). However, Robinson et al. (2016) have reported insignificant effects on occupancy, as the broader market segment satisfied by sustainability certificates can be offset by reduced demand from price-sensitive customers reducing RevPAR. Nevertheless, Bianco et al. (2023) found that being the first mover in attaining a certification within a competitive set can lead to higher occupancy rates and positive RevPAR.

The main issue lies in the disparity between products and services. Eco-labeled products already enjoy growing support from customers. We are aware that organic products, which are implicitly sustainable, come at a higher cost, yet this has not hindered the positive trend in the market, with an increasing display of these products on department store shelves. However, when it comes to hotels, the higher price is not perceived in the same way as the price difference between an electric car and a diesel car. That is, it is easier to translate into a customer benefit the higher price of a sustainable product than a sustainable service. Finally, we highlight that the impact of additional governance, compliance and transparency costs (e.g., the obligation to draw up sustainability reports) are uncovered by the literature.

Labor cost

The transition toward productive work opportunities with fair income, secure workplaces, personal development prospects, freedom of expression and association, and equal treatment for both genders represents an important cost, especially in the tourism sector, known for its significant share of informal labor-intensive work, facing challenges that hinder the achievement

of decent work, such as extended working hours, low wages, limited social protection, and a persistent gender pay (Santos, 2023).

In terms of aiming to construct a sustainability-oriented workforce, hotels have two options: (1) sustainable recruitment, or (2) upgrading the sustainable human capital providing critical related awareness, learning, training, and development to the existing employees. Focusing on sustainable recruitment is more cost-effective than providing sustainable knowledge and training to the existing employees (Salama et al., 2022) but not always possible. Sustainability requires rotations in positions (including top positions), gender and wage equality and/or smart-working and flexibility in workforce utilization that do not always fit easily into business' development plans. Many studies suggest that sustainable HRM (human resources management) adopts similar pillars to those of the mainstream HRM to define its practices: job design practices, recruitment and selection, training and development, performance appraisal and management, compensation and rewards (Anlesinya and Susomrith, 2020).

Sustainable job design involves determining the responsibilities and qualifications of a position to meet both the environmental goals of the hotel organization and the desires of responsible employees. It considers the technology used to achieve sustainable environmental performance.

Sustainable recruitment integrates environmental dimensions into employment policies and strategies. It should provide information about job requirements, nature, and type to select suitable individuals with relevant experience, skills, and behaviors. However, hiring staff with capabilities, skills, orientations, and green preparations is not easy. Job interviews should assess the potential fit of candidates with the organization's sustainable programs considering candidates' social orientation as a primary criterion for job vacancies (Luu, 2021).

Sustainable training and development practices need understanding market, governmental, and societal requirements. They focus on green innovations, new technologies, efficient products, and changes in business models (Tang et al., 2018). These practices include workplace workshops, job rotations, socially responsible management models, staff welfare training, health and safety measures, work regulations, equality, and employee rights. Proper training is crucial for hotels to convey their sustainable image and to foster socially accountable and concerned workforce.

Assessing staff performance involves promoting environmental learning and aligning performance criteria with sustainable standards (Mathews, 2018). Incorporating sustainability into performance appraisal ensures a positive image. Hotel managers should establish sustainable targets and procedures, assess incidents, practice social responsibility, and communicate sustainability strategy in daily operations (Zaki et al., 2018).

Sustainable reward management motivates employees and contributes to corporate sustainability. Salary bonus, extra incentives, tipping can reward sustainability criteria achievements. Some organizations use non-financial rewards (e.g., recognition, motivation, social incentives, and honors), instead of financial compensation leveraging on organization-wide labeling that raises employees' awareness of sustainability achievements.

Motivation and incentives are crucial HRM practices that link employee and organizational interests, supporting environmental management and innovation. Career advancement is a prime motivator for new hotel employees, and career management is increasingly recognized by Goh and Okumus (2020). Trusting employees and highlighting corporate growth plans increase understanding and commitment. Sustainable promotion involves the not easy task to help employees plan their careers based on their capabilities and organizational needs.

Social commitment ("costs")

Considering sustainability as a social commitment goes beyond mere presence in the community and reducing the undesirable use of social resources. It involves assessing firms' implementation of social goals, such as working conditions, health and safety, employee relationships, diversity, human rights, community engagement, with the aim of maximizing

enduring socio-economic and environmental development for social well-being.

Social commitment in the accommodation industry is often associated to social inclusion, highlighting that employing people with disabilities represents an opportunity to attract and retain socially responsible customers and customers with disabilities, as it acknowledges and reflects the diversity of the consumer market, leading to improved customer relations and higher consumer loyalty (Miethlich and Oldenburg, 2019). These authors claim that further research is needed regarding managers' attitudes in the hospitality and tourism industry. Madera et al. (2022) show that concerns (i.e., costs) arise when employing individuals with mental or physical disabilities in roles where face-to-face communication with employees is required, core tasks must be completed, and the success of the business is heavily reliant on service excellence (e.g., in high-priced luxury segments). Employing people with disabilities does not have a negative impact on guest experiences (Gröschl, 2013) but to the best of our knowledge, there are no studies quantifying the associated cost (if any). Moreover, no literature exists regarding the consequences of employing individuals with challenging backgrounds (e.g., former prisoners or former drug addicts) in terms of guest complaints or any additional costs/investments required for guest safety.

Furthermore, social inclusion also entails the development of products that are accessible to individuals with various forms of disabilities, including those related to physical conditions, age, or religious and ethical choices. The employment of people with disabilities can have a positive impact on a company's success in creating products and services tailored to the needs of individuals with disabilities (Lindsay et al., 2018). Guests with disabilities specifically choose businesses that offer a barrier-free environment and accommodating facilities, where their disabilities are not highlighted by the employees. Additionally, the barrier-free environment is also appreciated by families with young children or older adults (Gröschl, 2013).

Philanthropy is another "cost" associated with social commitment, especially for businesses that are not profitable. While the credit system is geared towards evaluating whether to grant financing for the purchase of machinery that (allegedly) increases productivity, it is not yet ready to finance a project aimed at recovering plastic from the sea to improve the destination's image. Even more serious are the consequences if the "philanthropic" company becomes subject to insolvency proceedings because, in such cases, donations can be considered distractions.

4. Conclusion: sustainability will become an essential requirement at any price

In this paper, we combine the experience of a medium-sized Benefit hotel chain with the existing (though limited) literature on the impact of sustainability certificates on hotel firms' financial performance and competitive positioning. We demonstrate that sustainability entails cost and prices increase in the short term, but only a short-sighted perspective would view it as a problem as sustainability will become an essential requirement, and consumers will shape their expectations based on perceived value rather than price. In fact, in the medium term, the benefits of employing a socially responsible workforce will materialize in the form of a higher attractiveness for new talent. Moreover, social commitment will be seen as an investment that offer positive returns in terms of the chain's image, customer loyalty, and the ability to influence customer behavior, encouraging additional cost-savings via more responsible resource utilization.

That is why accommodation businesses will soon have no choice but to embrace sustainability; it will become a necessity, not an option.

References

Anlesinya, A., Susomrith, P. (2020). Sustainable human resource management: a systematic review of a developing field. *Journal of Global Responsibility*, **11(**3), pp. 295-324.

Bernard, S., Nicolau, J. L. (2022). Environmental certification and hotel market value. *International*

Journal of Hospitality Management, **101**, 103129.
Bianco, S., Bernard, S., Singal, M. (2023). The impact of sustainability certifications on performance and competitive action in hotels. *International Journal of Hospitality Management,* **108**, 103379.
Casado-Díaz, A.B., Sellers-Rubio, R., Rodriguez-Sanchez, C., Sancho-Esper, F. (2020). Predictors of willingness to pay a price premium for hotels' water-saving initiatives. *Journal of Travel & Tourism Marketing*, **37**(7), pp. 773-784.
Chi, C.G., Chi, O.H., Xu, X., Kennedy, I. (2022). Narrowing the intention-behavior gap: The impact of hotel green certification. *International Journal of Hospitality Management*, **107**, 103305.
Damigos, D. (2023). How much are consumers willing to pay for a greener hotel industry? A systematic literature review. *Sustainability,* **15**(11), 8775.
Galati, A., Thrassou, A., Christofi, M., Vrontis, D., Migliore, G. (2021). Exploring travelers' willingness to pay for green hotels in the digital era. *Journal of Sustainable Tourism (*10.1080/09669582.2021.2016777).
Goh, E., Okumus, F. (2020). Avoiding the hospitality workforce bubble: Strategies to attract and retain generation Z talent in the hospitality workforce. *Tourism Management Perspectives*, **33**, 100603.
Gröschl S. (2013). Presumed incapable: Exploring the validity of negative judgments about persons with disabilities and their employability in hotel operations. *Cornell Hospitality Quarterly*, 54, pp. 114-123.
Guizzardi, A., Pons, F.M.E., Ranieri, E. (2019). Competition patterns, spatial and advance booking effects in the accommodation market online. *Tourism Management,* **71**, pp. 476-489.
Hameed, I., Hussain, H., Khan, K. (2022). The role of green practices toward the green word-of-mouth using stimulus-organism-response model. *Journal of Hospitality and Tourism Insights*, **5**(5), pp. 1046-1061.
Lindsay, S., Cagliostro, E., Albarico, M., Mortaji, N., Karon, L. (2018). A systematic review of the benefits of hiring people with disabilities. *Journal of Occupational Rehabilitation*, **28**(4), pp. 634- 655.
Luu, T.T. (2021). Socially responsible human resource practices and hospitality employee outcomes. *International Journal of Contemporary Hospitality Management*, **33**(3), pp. 757-789.
Madera, J.M., Taylor, D.C., Barber, N.A. (2020). Customer service evaluations of employees with disabilities: The roles of perceived competence and service failure. *Cornell Hospitality Quarterly*, **61**(1), pp. 5-18.
Mathews, J. (2018). Implementing green management in business organizations. *IUP Journal of Business Strategy*, **15**(2), pp. 46-72.
Miethlich, B., Oldenburg, A.G. (2019). How social inclusion promotes sales: An analysis of the example of employing people with disabilities. *Journal of Marketing Research and Case Studies*, **2019**, ID 463316.
Salama, W., Nor El Deen, M., Albakhit, A., Zaki, K. (2022). Understanding the connection between sustainable human resource management and the hotel business outcomes: Evidence from the green-certified hotels of Egypt. *Sustainability*, **14**(9), 5647.
Santos, E. (2023). From neglect to progress: Assessing social sustainability and decent work in the tourism sector. *Sustainability*, **15**(13), 10329.
Robinson, S., Singh, A.J., Das, P. (2016). Financial impact of LEED and energy star certifications on hotel revenues. *The Journal of Hospitality Financial Management*, **24**(2), pp. 110-126.
Tang, G., Chen, Y., Jiang, Y., Paillé, P., Jia, J. (2018). Green human resource management practices: scale development and validity. *Asia Pacific Journal of Human Resources*, **56**(1), pp. 31-55.
Thomas, G. (2022). Corporate social responsibility as a sustainable business practice: A study among generation Z customers of Indian luxury hotels. *Sustainability,* **14**(24), 16813.
Zaki, K., Qoura, O. (2019). Profitability in Egyptian hotels: Business model and sustainability impact. *Research in Hospitality Management*, **9**(2), pp. 89-97.

doi.org/10.26398/asaproc.0060

Wellbeing of Italian provinces through a multi-group approach

Pasquale Sannino[a], Cristina Davino[b], Stefania Taralli[c]
[a] University of Macerata
[b] Department of Economics and Statistics, University of Naples Federico II
[c] Istat - Directorate for analysis and enhancement of social and demographic statistics and for the information needs of the national recovery and resilience plan

1. Introduction

For over thirty years now, a widely accepted view is that gross domestic product (GDP) is an inadequate measure of a society's progress. Among the many arguments related to its inadequacy, the main one is the idea that GDP, although it can be used to some extent as a proxy for economic growth, is unable to capture everything that is not strictly economic (Stiglitz et al., 2010). In Italy, the BES project (from the Italian "Benessere Equo e Sostenibile," meaning Equitable and Sustainable Wellbeing), was started in 2010 by Istat and Cnel (Italian National Council for Economics and Labor), and it was aimed at measuring the progress of Italian society and providing a comprehensive and multidimensional statistical framework to monitor societal wellbeing, going beyond the purely economic measurements provided by GDP. To describe wellbeing in Italy, 12 domains were identified encompassing objective and subjective aspects: Health, Education and training, Work and life balance, Social relationships, Economic wellbeing, Politics and institutions, Safety, Subjective wellbeing, Landscape and cultural heritage, Environment, Innovation research and creativity, Quality of services.

The BES is updated annually, providing a complete overview of Italy's societal wellbeing. Considering the complexity, abstract nature, and multidimensionality of BES, its accurate measurement requires a wide range of statistical measures, while for the composition of the information provided by individual indicators, it is useful to adopt a tailored technique. In this paper we propose the use of partial least squares - path modeling (PLS-PM) (Wold, 1985). Specifically, a hierarchical model is constructed, with BES as the higher-order construct and other domains situated hierarchically beneath it, contributing to its measurement. The sudden onset of the Covid-19 pandemic, affecting the entire Eurozone and particularly vulnerable countries like Italy, has unveiled new needs and inequalities, exacerbating pre-existing concerns in the country.

The aim of this paper is to compare the pre-pandemic situation with the situation during the pandemic, attempting to quantify the Covid-19 effects on the BES and its components. To achieve this, two models are constructed: one for 2019 and one for 2020 (with the same indicators measured in two different years), followed by a comparison of path coefficients between the two years, so to assess potential changes. To achieve this goal, we refer to the limited literature on PLS-PM for longitudinal data (Roemer 2016) adjusting some approaches typically used in multigroup approach (MGA) (Sarstedt et al., 2011). The models are constructed and measured at small-region level (Nuts3), using the BES at local level indicators (Istat 2022) because more granular data allow a better appreciation of wellbeing levels changes and differences[1].

2. Methods

The estimation of the model was preceded by a thorough preliminary analysis with the aim of identifying those indicators that are available for the two selected years and suitable for the methodology used (for details see the concepts of assessment and validation in PLS-PM (Sarstedt

[1] https://www.istat.it/it/benessere-e-sostenibilità/la-misurazione-del-benessere-(bes)/il-bes-dei-territori

et al., 2019). Figure 1 shows the model structure: we employed 36 indicators for each year, divided into 12 sub-blocks and 10 domains. The domains are further grouped into two major blocks (*Outcome* and *Context* domains) in accordance with the model proposed by Davino et al. (2018). There are 10 domains used because *Subjective wellbeing* is not measured at the sub-regional level, while no indicators were available for Innovation research and creativity updated to 2020.

The objective of this paper is to use PLS-PM in longitudinal studies and to apply a multigroup approach to quantify the effects of the Covid-19 crisis on the BES levels in the different domains at local level. The application of PLS-PM in longitudinal studies is relatively limited in the literature, despite it being a highly useful technique for assessing changes in constructs over time, due to its advantageous methodological features such as construct prediction, suitability for small sample sizes, and handling highly complex models (Henseler et al., 2009; Roemer, 2016).

After estimating a separate model for each year, a multigroup analysis is performed where each of the two models represents one group. Specifically, a non-parametric test, the "Bootstrap MGA" is implemented to understand if the coefficients of the model in Figure 1 (called path coefficients in the PLS-PM terminology) between the two years are significantly different from each other. This approach compares each bootstrap estimates from one group with all the other bootstrap estimates of the same parameter in the other group, and then counts how many times the estimate from the first group is greater than the second. The result is a one-tailed test probability, indicating the probability that the estimate from the first group is greater than the second one. We run the bootstrapping procedure with 5,000 subsamples (Hair et al., 2011).

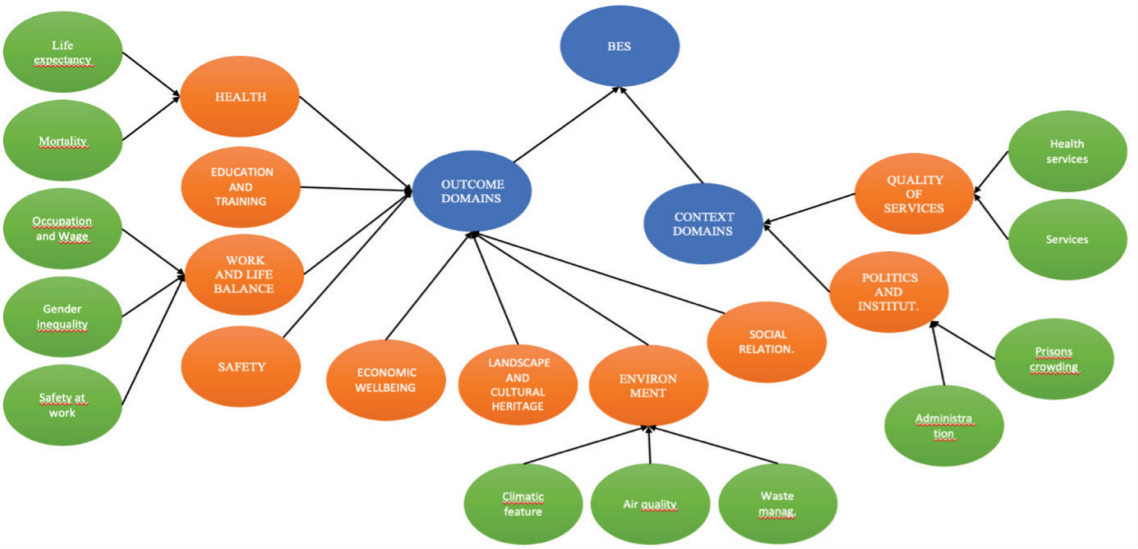


Figure 1: Structure of the model

3. Results

Table 1 presents the path coefficients for 2019 and 2020. When moving from one year to another, the hierarchical order of domains changes only slightly: *Work and life balance* remains the domain with the highest impact in both 2019 and 2020, followed by *Education and training*, *Economic wellbeing*, and *Environment*. When comparing the two years, it is interesting to note the substantial decrease in the role of *Health* dimension, which essentially drops by about half. Notably, this decrease occurs in conjunction with the Covid-19 pandemic. After *Health*, the dimension that varies the most is *Environment*. While the change in *Health* is negative due to the pandemic's effects, the change in *Environment* is positive. During the pandemic there was a reduction in waste production, because of lockdown measures that forced commercial activities to close.

Also, we implemented a multigroup approach to see if the path coefficients between the two years are significantly different from each other. Table 2 reports the differences between the path coefficients with their corresponding p-values (in the third column) from the Bootstrap-MGA test. The two domains we mentioned earlier, *Health* and *Environment*, are the only ones that change significantly (at a 5% significance level). However, it may be considered that the BES indicators do not all have the same temporal variability, indeed many indicators have a certain latency period, meaning they need to be monitored for a longer period to observe the changes they record.

Table 1: Path coefficient values in both years

Latent variables		Path coefficients 2020		Path coefficients 2019	
Outcome	*Health*	0.049	0.882	0.133	0.891
	Education and training	0.280		0.275	
	Work and life balance	0.330		0.304	
	Economic wellbeing	0.169		0.153	
	Social relationships	0.080		0.077	
	Safety	0.131		0.119	
	Landscape and cultural heritage	0.061		0.051	
	Environment	0.176		0.119	
Context	*Politics and institutions*	0.280	0.159	0.295	0.143
	Quality of services	0.878		0.851	

Table 2: Results of Bootstrap-MGA test

Latent variables	Differences (2020 - 2019)	p-value
Health	-0.085	0.000
Education and training	0.005	0.812
Work and life balance	0.027	0.246
Economic wellbeing	0.016	0.282
Social relationships	0.004	0.844
Politics and institutions	-0.015	0.898
Safety	0.012	0.354
Landscape and cultural heritage	0.010	0.637
Environment	0.057	0.024
Quality of services	0.027	0.749
Outcome domains	-0.009	0.777
Context domains	0.016	0.659

Table 3: Average of Health for the different years and Italian geographic area

Geographic area	Health 2019 (average)	Health 2020 (average)
North-West	0.059	-0.876
Noth-Est	0.677	0.461
Center	0.548	0.774
South and islands	-0.748	-0.139

Taking this latency period into account is crucial for correctly interpreting variations in the data within the context of pandemic conditions and to avoid possible incorrect or hasty interpretations of the changes observed. Looking at the territorial heterogeneity, it is possible to quantify the Covid effect by considering the averages of the *Health* score in the two years (Table 3). This allows us to appreciate how much the provinces in the North-West of Italy have experienced significant reductions in *Health* scores between 2019 and 2020.

4. Conclusion

Since the objective of this paper was to quantify the effects of the pandemic on regional inequalities via the different BES domains, we have observed, as expected, that the most significant change between 2019 and 2020 occurred in the domain of *Health*. The pandemic had a substantial effect on *Health* differences among Italian provinces, as evidenced by the significant reduction in the path coefficient of *Health* on the *Outcome* domains, which is equal to 0.085. Levels of *Health* worsen in those provinces that in 2019 had the highest *Health* levels. As a result of this downward convergence, the contribution of the *Health* domain to wellbeing differences at local level reduced. These data reflect a range of challenges and negative changes that the pandemic has had on people's wellbeing, including the geographical spread of the virus itself during 2020.

It may be acknowledged that the dimensions of the BES had different response times to the pandemic shock. Therefore, for a more complete study, it is appropriate to analyze the changes in the complex structure of the BES and the evolution of regional differences in wellbeing even after 2020 year (which we are starting to do from the second half of 2023, as soon as the updates are available). In summary, this study highlights the importance of modelling data in a way that is sensitive to the specific vulnerabilities and response capacities of the units of analysis (in our case, Italian provinces) so that we have accurate indications to guide policy as well.

References

Davino, C., Dolce, P., Taralli, S., Vinzi, V. E. (2018). A quantile composite-indicator approach for the measurement of equitable and sustainable well-being: A case study of the Italian provinces. *Social Indicators Research*, **136**, pp.999-1029.

Hair, J.F., Ringle, C.M., Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice*, **19**(2), pp.139-152.

Henseler, J., Ringle, C.M., Sinkovics, R.R. (2009). The use of partial least squares path modeling in international marketing. In *New Challenges to International Marketing*. Emerald Group Publ. Ltd.

Jacobs, N., Hagger, M.S., Streukens, S., De Bourdeaudhuij, I., Claes, N. (2011). Testing an integrated model of the theory of planned behaviour and self-determination theory for different energy balance-related behaviours and intervention intensities. *British Journal of Health Psychology*, **16**(1), pp.113-134.

Roemer, E. (2016). A tutorial on the use of PLS path modeling in longitudinal studies. *Industrial Management Data Systems*, **116**(9), pp.1901-1921.

Sarstedt, M., Hair Jr, J.F., Cheah, J.H., Becker, J.M., Ringle, C.M. (2019). How to specify, estimate, and validate higher-order constructs in PLS-SEM. *Australasian Marketing Journal*, **27**(3), pp.197-211.

Sarstedt, M., Henseler, J., Ringle, C. M. (2011). Multigroup analysis in partial least squares (PLS) path modeling: Alternative methods and empirical results. In *Measurement and Research Methods in International Marketing* (pp. 195-218). Bingley: Emerald Group Publishing Limited.

Stiglitz, J.E., Sen, A., Fitoussi, J.P. (2010). *La misura sbagliata delle nostre vite. Perché il PIL non basta più per valutare benessere e progresso sociale*. Rizzoli, Milano.

Wold, H. (1985). Partial least squares. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of Statistical Sciences*. New York: Wiley.

doi.org/10.26398/asaproc.0061

Investigating the link between teachers' burnout and technology through the network psychometrics analysis: A survey of educators from diverse schools and grade level

Annalina Sarra[a], Lara Fontanella[b], Maila Pentucci[a], Giovanna Cioci[a], Ida Esposito[a]

[a] Department of Philosophical, Pedagogical and Economic-Quantitative Sciences, University "G. d'Annunzio" of Chieti-Pescara, Italy

[b] Department of Legal and Social Sciences, University "G. d'Annunzio" of Chieti-Pescara, Italy

1. Introduction

The era of information technology has brought about swift technological progress, significantly influencing our daily lives. The widespread and rapid adoption of technology has compelled people to adopt a digital lifestyle, which can impact their professional productivity and psychosocial well-being (Erten et al., 2020). In the education context, the incorporation of technology has undergone changes over time, with teachers emphasising the necessity for specialised training to effectively utilise technology in the classroom. Several studies have shown that there is a connection between the well-being of teachers and that of their students. Therefore, it has been established that, given the contagious nature of teachers' well-being, it should be ensured and protected over time. This is not only in the interest of the teachers themselves but also of the entire educational system.

It is widely demonstrated in the literature that the introduction of technology in education can evoke mixed responses among teachers, with some facing difficulty in managing external demands, increasing the risk of burnout. The burnout syndrome observed in caring professions, particularly among health, social workers, and educators, is characterised by prolonged exposure to emotionally taxing work conditions resulting in physical and emotional exhaustion, depersonalization, and a low sense of personal accomplishment (Maslach et al., 1981). Research on the impact of technology use by teachers suggests that there are both potential risks and benefits. On the one hand, extensive technology use may contribute to burnout due to increased workload, lack of training, and difficulties in managing technology-related issues. On the other hand, technology can provide advantages such as greater work flexibility, personalized learning opportunities for students, and support for teachers' professional development. The relationship between technology use and teachers' well-being and job satisfaction is complex and varies across different studies.

Driven by the interest to further explore these topics, we conducted a study specifically aimed at investigating the link between burnout and the use of technology by teachers. To achieve this, we reached out to both support teachers and teachers from various disciplines in primary, lower secondary, and upper secondary schools across Italy. To gather data from teachers residing in different regions of the country, we employed a web-survey as our primary data collection method. Through a network psychometrics analysis (Epskamp et al., 2012), we checked if three group-teachers from primary, middle and high school differ in terms of the relationship between burnout and knowledge and pedagogical competencies about technology and its effective use during classroom activities. The remainder of the paper is organised as follows. After this introduction, Section 2, gives insights about participants, instruments and methodology for data analysis. Section 3, presents the main results of the empirical analysis carried out to investigate the link between teachers' burnout and technology, while the final Section offers the main conclusions from the paper.

2. Materials and methods

2.1 Participants

In this research, 430 Italian teachers were involved. After obtaining the initial approval from their school principals, we reached out to the teachers through social networks. We gathered data from schools situated throughout the entire national territory. Out of the 430 participants, 90% are female. The majority of teachers are aged between 46 and 55 years old (34.5%) with only 3 participants below the age of 25 and 6 individuals above the age of 65. We collected data in schools located across the whole national territory. Campania, Lombardy, Abruzzo, and Lazio are the most represented regions. When considering the school level where they teach, 26.7% of the teachers are employed at the primary school level (1st–5th grade, N = 115, 98.3% females), 48.8% at the middle school level (6th–8th grade, N = 214, 90.1% females), and 23.5% at the high school level (9th–13th grade, N = 101, 86.8% females).

2.2 Instruments

To evaluate the burnout levels among teachers, we utilised the Italian version of the Copenhagen Burnout Inventory (CBI) (Kristensen et al., 2005), which was validated by Avanzi et al. (2013). This inventory consists of 19 self-reported items, and respondents rated their experiences on a 5-point Likert scale. The CBI measures three dimensions of burnout: personal burnout, work-related burnout, and student-related burnout. The personal and work-related subscales focus on emotional and physical exhaustion, both in personal and workplace domains. The third subscale specifically examines the experience of professional fatigue concerning students. Furthermore, we administered eight ad hoc items specifically designed to assess teachers' interactions with technology during lessons. These items covered the use of devices in the classroom, reflections on whether they use technologies excessively for work-related reasons, efforts to reduce technology usage, perceptions of how technologies may affect the conduct of lessons, and beliefs about how devices can aid students' understanding of explanations. Additionally, teachers were asked about their opinions regarding the presence of devices in the classroom, including considerations about removing devices if possible, and the potential impact of personal and collective device usage on students' attention. Items description is listed in Table 1.

2.3 Psychometric Networks

Psychometric network analysis, introduced as a distinct methodological approach by Epskamp et al. (2012), represents a novel method for modeling the interrelationships between psychological constructs and their items in a measurement instrument. In contrast to conventional psychometrics, which treats items as mere indicators of some latent constructs, network analysis considers both items and constructs as interconnected nodes within a network. In this context, nodes represent various psychological variables such as mood states, symptoms, or attitudes, while the links between nodes signify the underlying statistical relationships that necessitate estimation. As a result, this distinctive feature presents novel challenges for statistical inference, differentiating it from traditional social network analysis, where links between nodes are readily observable.

In this study, our emphasis lies on a specific network model known as the pairwise Markov random field (PMRF; Lauritzen, 1996). PMRF belongs to the category of undirected network models where variables are depicted as nodes linked by edges. These edges indicate the strength of the conditional association between two variables, considering the influence of all other variables within the network. The prevailing approach to building a psychological network is by estimating a network of partial correlation coefficients (McNally, 2016; Borsboom et al., 2013).

Table 1: Teacher Burnout and Technology Interaction Items

Item	
How often do you feel tired?	B1
How often do you feel physically exhausted?	B2
How often do you feel emotionally exhausted?	B3
How often do you think, "I can't take it anymore"?	B4
How often do you feel worn out?	B5
How often do you feel weak and susceptible to illness?	B6
Is your work emotionally exhausting?	B7
Do you feel burnt out because of your work?	B8
Does your work frustrate you?	B9
Do you feel worn out at the end of the working day?	B10
Are you exhausted in the morning at the thought of another day at work?	B11
Do you feel that every working hour is tiring for you?	B12
Do you have enough energy for family and friends during leisure time?	B13
Do you find it hard to work with students?	B14
Do you find it frustrating to work with students?	B15
Does it drain your energy to work with students?	B16
Do you feel that you give more than you get back when you work with students?	B17
Are you tired of working with students?	B18
Do you sometimes wonder how much longer you will be able to work with students?	B19
Do you use technologies during classroom lessons?	D1
Have you thought about using technologies excessively for work-related reasons?	D2
Have you tried to reduce your usage of technologies?	D3
Do technologies (devices and digital tools) hinder the conduct of the lesson?	D4
Do devices help students understand explanations?	D5
If it were up to me, I would remove devices from the classroom.	D6
Does the presence of personal and collective devices negatively affect students' attention?	D7
Have various devices facilitated interactions with students' families?	D8

These networks are also referred to as Gaussian graphical models (GGMs; Lauritzen, 1996). In a GGM, the main idea is to represent each variable as a node in the graph, and the absence or presence of edges between nodes indicates conditional independence or dependence between the corresponding variables. An edge between two nodes implies that the two variables are conditionally dependent, given all the other variables in the model. The variables are assumed to follow a multivariate normal (Gaussian) distribution. This means that the joint distribution of all variables can be fully characterised by their mean vector and covariance matrix. From a formal standpoint, let us denote by $\mathbf{Y}$ a vector containing n responses from a random subject, and $\mathbf{y}$ as a specific realisation of this vector. We will make the assumption that $\mathbf{Y}$ follows a normal distribution with a mean vector $\boldsymbol{\mu}$ and a variance-covariance matrix $\boldsymbol{\Sigma}$. Particular significance assumes the variance-covariance matrix Σ as it captures both the marginal and conditional relationships among variables in $\mathbf{y}$. In the GGMs, we do not directly model $\boldsymbol{\Sigma}$ itself, but instead, we focus on modeling the standardised elements of

$$\boldsymbol{K} = \Sigma^{-1} \tag{1}$$

Each element in the standardised precision matrix represents the partial correlation between two variables, controlling for all other variables in the model:

$$\omega_{\mathrm{ij}} = \frac{-k_{ij}}{\sqrt{k_{ii}k_{ij}}} \tag{2}$$

If the partial correlation is zero (no edge), then there is no link between the corresponding variables in the network. As recommended by Epskamp et al. (2017), to achieve a clearer modeling and interpretation of the network parameters, the inverse variance-covariance matrix can be framed in this form:

$$\Sigma = \Delta(I - \Omega)^{-1}\Delta, \tag{3}$$

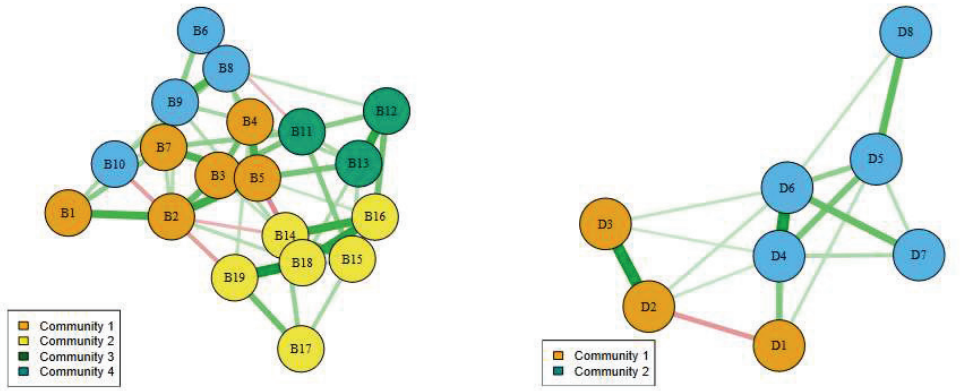
where Ω contains partial correlation coefficients on off-diagonal elements and zeroes on the diagonal, and Δ is a diagonal scaling matrix. Estimating the GGM involves estimating the precision

matrix. Various estimation methods can be used. More details can be found in Isvoranu et al. (2022). Since we are working with responses evaluated on a 5-point Likert scale, a potential approach involves building a model in which a GGM is hypothesized to underlie a collection of ordered categorical responses. Drawing from Epskamp et al.'s (2018) methodology, we employed polychoric correlations as input for GGM estimation tools that operate on a correlation matrix.

3. Results

Psychometric network analyses were performed on partial correlation matrices derived from both the burnout inventory and items related to teachers' technology interactions during lessons. It is worth noting that prior to conducting the analyses, the items B10, D1, D5, and D8 were reversed. The analysis employed the *qgraph*, the *bootnet*, and *igraph* packages (Isvoranu et al., 2022), within the R programming environment. The networks generated were subsequently visualised using the qgraph package.

The psychometric network analysis was carried out in accordance with the steps described by Epskamp et al. (2018). We combined the data into three group-teachers' responses, representing primary, middle, and high school levels. We initially constructed distinct networks for each of the three datasets and for each scale that was taken into account. The nodes in these networks corresponded to the variables listed in Table 1, with then interconnections between nodes depicted by partial correlation coefficients. Then, we utilised the Network Comparison Test (NCT; van Borkulo et al., 2021) from the R NetworkComparisonTest package to assess whether the underlying structure could be consistent across these three groups. Results of the test for both burnout scale and the investigation of teacher's device interactions suggest no significant disparities in the three network structures. Consequently, we merged the datasets for all subsequent analyses. For each scale, we also determine a partition of nodes. To this end, we perform a community detection analysis. The results of *louvain* community detection algorithm (Blondel et al., 2008) are displayed in Figure 1. The *louvain* algorithm supported the selection of 4-cluster solution for the burnout network. The colors employed in the plot correspond to the community identifiers



(a) Burnout-items **(b)** Technology (devices)–related items.

Figure 1: Network of the nodes listed in Table 1, with communities estimated using the *louvain* algorithm.

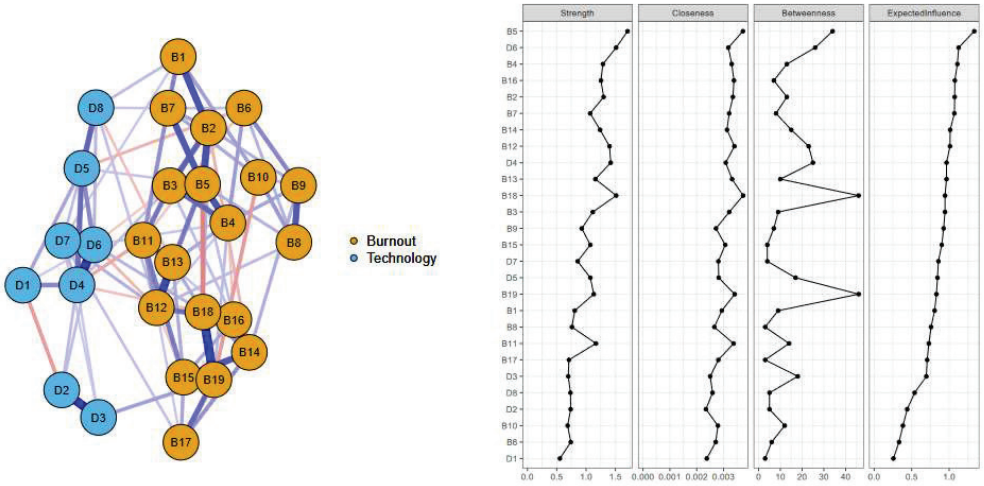


Figure 2: Estimated global network (a) and centrality indices of the nodes ordered by expected influence (b)

The initial cluster, encompassing items B1:B4 and B7, corresponds to the underlying construct referred to as "personal burnout" in the CBI by Kristensen et al. (2005). On the other hand, item B6 was situated within a different cluster, as demonstrated in Figure 1. Predictably, item B1displays a significant connection with B2. Additionally, we observe a strong correlation between item B5 and both B2 and B4. The second *louvain* community accurately captures the student-related burnout dimension of the CBI. Strong positive edges are found in most of the items of this cluster. On the other hand, the algorithm has divided the initial work-related burnout dimension of the CBI into two separate communities, which can be interpreted post-hoc as: "EmotionalStrain" (Community 3) and "Physical Exhaustion and Energy Depletion" (Community 4).

Figure 1 also provides a two-cluster solution for the technology network. The first community, containing the items D1:D3, can be termed "Technology utilisation and management", whereas the second community is associated with "Technology influence on Teaching and Engagement." To explore the connection between teacher's burnout and technology interaction, we complemented the analysis by constructing a psychometrics network that incorporates both dimensions. Figure 2 illustrates the network configuration, where the edges are represented by partial correlation coefficients. A careful examination of Figure 2 uncovers an adverse association between the item that explores the perception of teachers' fatigue during their work hours (B12) and desire to remove devices from the classroom if given the choice (D6). Likewise, negative relationships are observed between the items addressing potential obstacles due to technology usage in lessons (D4) and emotional exhaustion (B11). D4 also exhibits a negative correlation with B12.

Additionally, for each node, we present in Figure 2 its estimated strength (direct connections to other nodes), its closeness (indirect connections to other nodes), its betweenness (importance in linking other nodes together), and its expected influence (potential influence on other nodes within the network). Notably, concerning betweenness and closeness, items D1 and D2 were positioned at the periphery of the network. Conversely, the items B5 and B18 were among the most central nodes with respect to strength and betweenness. In other words, these nodes had strong direct connections to other nodes. Regarding the expected influence metric, it is evident that B5, D6, B4, B16 and B2 are the items that significantly influence the overall network dynamics.

4. Concluding remarks

Our psychometric network analysis has unveiled a consistent foundation for both teachers' burnout and technology interaction. Unlike structural equation modelling, which relies on predefined relationships grounded in theory, psychometrics network analysis allows for a more data-driven exploration, revealing patterns and connections not initially hypothesized. The centrality indices provide valuable insights into overall network dynamics, guiding future investigations into teachers' exhaustion, career longevity concerns, and energy depletion. Notably, adverse associations were discovered between indicators of teachers' fatigue and their inclination to remove devices from the classroom. Additionally, negative correlations emerged between technology-related obstacles and emotional exhaustion.

In conclusion, these findings contribute to a deeper understanding of the complex interplay between teachers' well-being and their interactions with technology in an educational setting.

References

Avanzi, L., Balducci, C., Fraccaroli, F. (2013). Contributo alla validazione italiana del Copenhagen Burnout Inventory (CBI) [A contribution to the Italian validation of the Copenhagen Burnout Inventory (CBI)]. *Psicologia della Salute*, **2**, pp. 120–135.

Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics*, P10008.

Borsboom, D., Cramer, A.O.J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, **9**, pp. 21–121.

Epskamp, S., Cramer, A.O., Waldorp, L.J., Schmittmann, V.D., Borsboom, D. (2012). Qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, **48**(4), pp. 1–18.

Epskamp, S., Borsboom, D., Fried, E.I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, **50**, pp. 195–212.

Epskamp, S., Fried, E.I. (2018). A tutorial on regularized partial correlation networks. Psychological Methods, **23**, 617.

Epskamp, S., Rhemtulla, M., Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika*, **82**, pp. 904–927.

Erten, P., Özdemir, O. (2020). The Digital Burnout Scale Development Study. Inonu University. *Journal of the Faculty of Education*, **21**(2), pp. 668–683.

Isvoranu, A.M., Epskamp, S., Waldorp, L., Borsboom, D. (2022). *Network Psychometrics with R: A Guide for Behavioral and Social Scientists* (1st ed.). Routledge. https://doi.org/10.4324/9781003111238

Kristensen, T.S., Hannerz, H., Högh, A., Borg, V. (2005). Questionnaire–A tool for the assessment and improvement of the psychosocial work environment. *Scandinavian Journal of Work, Environment & Health*, **31**, pp. 438–449.

Lauritzen, S. L. (1996). *Graphical Models*. Clarendon Press, Oxford, United Kingdom.

Maslach, C., Jackson, S.E. (1981). The measurement of experienced burnout. *Journal of Organizational Behavior*, **2**, pp. 99–113.

McNally, R.J. (2016). Can network analysis transform psychopathology. *Behaviour Research and Therapy*, **86**, pp. 95–104.

van Borkulo, C.D., Boschloo, L., Kossakowski, J., Tio, P., Schoevers, R. A., Borsboom, D., Waldorp, L.J. (2021). Comparing network structures on three aspects: A permutation test. *Psychological Methods*. doi: 10.1037/met0000476

doi.org/10.26398/asaproc.0062

Scientific abstracts analysis to uncover Ukrainian-Russian war insights: A text mining-based approach

Leonardo Scarso[a], Marco Novelli[b], Francesco Saverio Violante[a]
[a] Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy
[b] Department of Statistical Sciences, University of Bologna, Bologna, Italy

1. Introduction

Occurrences such as the COVID-19 pandemic and the conflict between Russia and Ukraine are two of the main events that impact the global population. Their outcomes may pertain to diverse areas: ranging from the economic sphere, which is a common occurrence, to healthcare and education, which are less frequently areas of concern. Nevertheless, these incidents are subjected to continuous analysis in order to acquire valuable knowledge and assist policymakers in making decisions and preventing further aggravations.

Differently from social media, scientific papers could be described as a type of 'trustworthy evidence'. Built on the main idea of being short but interesting and effective resumes of a research, they are based on and recorded after the verification of the occurrence. Additionally, the phenomenon has been thoroughly scrutinized, focusing on a specific aspect, to yield findings that convey something novel to the readers, which could be beneficial in the future. For this reason, publications are a 'gold miner' for acquiring knowledge about a given fact Dos Santos (1996).

Analysis related to abstracts could convey, for instance, to the identification of hot topics and key authors of the statistical community Ji (2016). Another example of how this type of short textual data could be analyzed is to verify whether their structure, in terms of words and sentences, has been re-used in articles corpus Atanassova (2016).

In this article, we focus on the analysis of scientific abstracts related to the Ukrainian- Russian war. Specifically, we examine articles published from February 24th, 2022, to April 15th, 2023. These articles were obtained through Google Scholar and PubMed searches using a query containing the words 'war' and 'Ukraine'. Our goal is to classify the articles based on their main themes, employing text mining techniques and implementing a model-based clustering strategy that relies on a mixture of cosine distance-based density functions, as developed in Anderlucci (2019). Specifically, we hope that the results will aid in comprehending the research direction regarding war-related subjects and the insightful findings that can inform decision-making processes.

2. Methodology

Within this particular scenario, involving concise textual data related to abstracts on the Ukrainian-Russian war, a wide range of statistical models could be employed. However, conventional models rooted in natural language processing and machine learning methodologies may be inadequate for the research objective. This inadequacy stems from the fact that we are dealing with an unsupervised problem characterized by a substantial degree of sparsity. Additionally, the absence of prior knowledge about topics and their number further complicates matters, making it challenging to dispose of a potential training set in a supervised context.

Moreover, due to the brevity of the considered documents, where only few words exhibit positive frequency, opting for a framework that assumes a singular association between documents and topic is recommended.

For these reasons, in this work we employ a model-based clustering technique using mixtures of cosine based distance densities, originally proposed in Anderlucci (2019) to create a 'taxonomy' of statistical papers. This approach was chosen due to the high sparsity of the document-term matrix (over 99%) that stores the information from the abstracts. The cosine distance is the most reliable and suitable metric in this case because it focuses solely on the nonzero entries.

This allows for an accurate measurement of dissimilarity between two documents based on their subject matter. More specifically, we consider a mixture of k cosine distance-based density functions:

$$f(y;\xi,\lambda) = \sum_{i=1}^{k} \pi_i \psi(\lambda) e^{-\lambda d(y,\xi_i)},$$

where:

- π_i, $i=1,...,k$ are the positive mixture weights, such that $\sum_{i=1}^{k} \pi_i = 1$
- $\boldsymbol{\xi_i}$, $1,...,k$ are the component varying centroid vectors;
- $d(\boldsymbol{y}, \boldsymbol{\xi_i})$ is the cosine distance between the p-dimensional document y and the respective centroid $\boldsymbol{\xi_i}$, that is:

$$d(y,\xi_i) = 1 - \frac{\sum_{h=1}^{p} y_h \xi_{ih}}{\sqrt{\sum_{h=1}^{p} y_h^2} \sqrt{\sum_{h=1}^{p} \xi_{ih}^2}},$$

  with y_h and ξ_{ih} denoting the frequency of word h in $\boldsymbol{y}$ and $\boldsymbol{\xi}_i$, respectively;
- $\psi(\lambda)$ is the normalizing constant, depending on the precision parameter λ, with $\lambda > 0$, which acts as a scaling factor of $d(\boldsymbol{y},\boldsymbol{\xi_i})$.

Parameters estimation is performed via EM algorithm, initialized with the solution of the spherical k-means based on the cosine distance Dhillon and Modha (2001).

The employed method proved to work well with abstracts of scientific papers; compared with distance-based methods (e.g. spherical k-means, partition around medoids Kaufman (2009), hierarchical clustering, to name a few), it is more flexible and it is equipped with coherent criteria to choose the optimal number of clusters. Within the model-based clustering framework, more sophisticated models like latent Dirichlet allocation Blei et al. (2003) and topic models Blei (2012) assume that more than a topic is present within a single document. However, such an assumption could have been realistic if whole articles were considered; given that our documents are very brief, it is very hard and certainly imprecise to disentangle many subtopics. Moreover, the problem of choosing k would have remained.

3. Data analysis

In this section of the article, we will discuss the conducted investigation and the approach that was applied, starting with a concise description of the collected data.

Data set description

The built data set contains n = 1174 observations, which are scientific articles published from February 24th, 2022 to April 15th, 2023, related to the ongoing war between Ukraine and Russia. The extracted attributes were: title of article, principal author, DOI, type of article, publisher, year and abstract. Data are collected from two different sources:

- *Google scholar*: web scraping is used with the aim of obtaining all the outcomes derived from the query containing both 'war' and 'Ukraine' terms, with a total of 892 articles with abstracts.
- *Pubmed*: the biomedical articles database enables to extract the articles information, which is a result of the aforementioned query, admitting to gather additional 282 observations.

Most of the articles were published in 2022 (67.2%), also because the considered timestamp includes a larger number of months with respect to the following year. The most frequent observations for the *type* attribute are *journal-article* and *book-chapter* (890 and 140, respectively).

Besides the abstract, which is the primary variable of interest, we have also retrieved the title as it is also a valuable source of textual data to draw from. The textual information has been subjected to pre-processing in advance to obtain fewer but still pertinent words, by excluding special characters, punctuation marks and stop words. Moreover, as part of data processing, stemming was carried out in order to reduce inflected or derived words to their root word stem.

A further filtering step has been performed: the three most commonly occurred words, 'Ukraine', 'war' and 'Russia' have been removed as their high frequencies could skew the analysis. In particular, these words could be assigned to each cluster, without providing any additional information on the topic. The same procedure was applied to all up to the three- character tokens, the numeric ones, and all the words that occurred in the entire data set less than 16 times. The rationale for this decision is based on the objective of mitigating the sparsity of the document-term matrix. As a result, by excluding around half of the columns (leading to a decrease from 14,000 to 7,700 words), the sparsity was reduced from 99.41% to 99.26%. Given our objective of identifying words that play a crucial role in distinguishing homogeneous groups and exhibit differential presence across the documents, we proceed with an additional feature selection step using entropy. Words exhibiting entropy greater than the mean value have been retained, as they make the greatest contribution to the identification of homogeneous groups. Consequently, our analysis will ultimately include 660 words, with a sparsity of 98.30%.

We propose a strategy that consists on assigning each document to a topic by employing a mixture of cosine distance-based density functions on re-weighted term frequencies which characterize our short text data. This methodology showcases its efficacy in capturing clusters and their respective subject areas. Moreover, text mining techniques have been applied to present model results.

4. Results

With the objective of selecting the optimal number k of clusters that effectively capture the heterogeneous distribution of scientific abstracts pertaining to the Ukrainian-Russian war, the EM algorithm is executed for a range of k values, spanning from 2 to 20. The Akaike information criterion (AIC) is employed to determine the most appropriate number of clusters.

Clusters homogeneity was therefore computed through the cohesion index C_i, which relies on the cosine distance metric:

$$C_i = \sqrt{1 - \bar{d}_i^2}, \quad 0 \leq C_i \leq 1.$$

with $\bar{d}_i$ denoting the average cosine distance between all papers within cluster i. A higher value of C_i indicates a higher level of homogeneity within the i-th cluster.

Another metric applied to evaluate the quality of clusters is the Average Silhouette Width (ASW) Rousseeuw (1987). It provides a measure of how well-separated clusters are, indicating the appropriateness of the cluster assignments. The silhouette width s_j is calculated for each observation j in the data set:

$$s_j = \frac{b_j - a_j}{max(a_j, b_j)}, \quad j = 1, \ldots, n$$

with a being the average distance from the j-th data point to all other points in the same cluster and b the average distance from the data point j to all points in the nearest cluster. For each group, the mean of these values is then computed to obtain the ASW.

$$s_j = \frac{b_j - a_j}{max(a_j, b_j)}, \quad j = 1, \ldots, n$$

with a being the average distance from the j-th data point to all other points in the same cluster and b the average distance from the data point j to all points in the nearest cluster. For each group, the mean of these values is then computed to obtain the ASW.

The best clustering, identified by $k = 7$ groups, shows a Pearson-Gamma value of 0.34 and a Dunn index of 0.047. Table 1 presents information on each identified cluster, including the number of papers within each cluster, n_i, the cohesion index, C_i, the Average Silhouette Width, ASW_i, and the most representative paper. For the latter, information about authors, title, year of publication, and the corresponding journal is

provided. The paper refers to the article having the lowest cosine distance in relation to the cluster centroid; this type of information could aid in determining the cluster's topic.

Cluster	Paper	n_i	C_i	ASW$_i$
1	Hassen, B., T., El Bilali, H, *Impacts of the Russia-Ukraine War on Global Food Security: Towards More Sustainable and Resilient Food Systems*?, 2022, Foods	66	0.83	0.35
2	Irtyshcheva, I., Kramarenko, I., Sirenko, I., *The economy of war and postwar economic development: world and Ukrainian realities*, 2022, Baltic Journal of Economic Studies	178	0.63	0.07
3	Kvyetnyy, R., Bisikalo, O., Palamarchuk, Y., Storchak, V., *JetIQ electronic ecosystem at the service of the university in the country that has undergone armed aggression*, 2023, Problems and Perspectives in Management	66	0.77	0.29
4	Raik, K., Blockmans, S., Dandashly, A., Noutcheva, G., Osypchuk, A., SuslovTackling, A., *The Constraints on EU Foreign Policy towards Ukraine: From Strategic Denial to Geopolitical Awakening*, 2023, JOINT Papers	149	0.76	0.25
5	Jaeger, F.N., Berger, C., Buettcher, M., Depallens, S., Heininger, U., Heller, Y., Kohns Vasconcelos, M., Leforestier, B., Pellaud, N., Relly, C., Trück, J., von Overbeck Ottino, S., Wagner, N., Ritz, N.; Migrant Health Reference Group of Paediatrics Switzerland; Paediatric Infectious Disease Group in Switzerland (PIGS), *Paediatric refugees from Ukraine: guidance for health care providers*, 2022, Swiss Med Wkly	201	0.63	0.13
6	Shariati Feizabadi, M., Delgado, F., Pormennati, M., *Deterrent Role of Sport in Military War (Case Study: Netnography of Russia-Ukraine Conflict)*, 2022, Journal of New Studies in Sport Management	435	0.41	-0.01
7	Boubaker, S., Nguyen, N., Trinh, V.Q. and Vu, T. *Market reaction to the Russian Ukrainian war: a global analysis of the banking industry*, 2023, Review of Accounting and Finance	79	0.71	0.19

Table 1: Most representative papers, size and cohesion index for each identified cluster.

Groups dimension greatly varies, with clusters 1, 3, and 7 having the lowest size, containing 66, 66, and 79 papers, respectively. These groups are among the most homogeneous ones, as reported in their respective cohesion index values, as well as the ASW ones, indicating well- defined groups. Cluster 1 demonstrates the highest cohesion with a value of 0.83, followed by cluster 3 with a value of 0.77; cluster 7, comprising 79 abstracts, achieves a cohesion index of 0.71. Cluster 6 could be considered as the residual group, as its C_i value of 0.41 and the negative ASW value suggest.

The cluster with the highest combination of number of papers and cohesion is Cluster 4, which consists of 149 papers with a cohesion index of 0.76. Clusters 2 and 5 follows, with 178 and 201 papers, respectively and C_i value of 0.63.

While the characteristic paper hints at cluster themes, wordclouds can offer a more comprehensive influence on researchers' academic paths related to the war.

The wordcloud plot showcased in Figure 1 illustrates the 20 most commonly used stemmed words within each recognized cluster. The size of the wordclouds is proportional to the number of articles, and the size of individual words corresponds to their frequency within the groups. Furthermore, the arrangement of the clusters is determined through the implementation of multidimensional scaling.

$$s_j = \frac{b_j - a_j}{max(a_j, b_j)}, \; j = 1, \ldots, n$$

with a being the average distance from the j-th data point to all other points in the same cluster and b the average distance from the data point j to all points in the nearest cluster. For each group, the mean of these values is then computed to obtain the ASW.

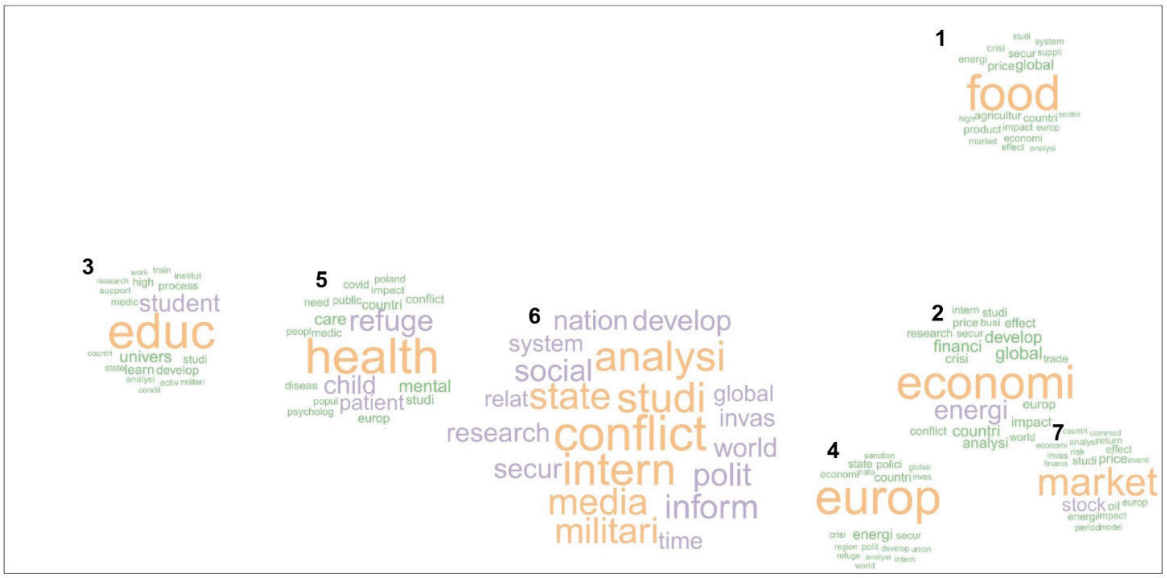


Figure 1: Wordcloud plot of the topics identified in Ukranian-Russian war related articles.

Food scarcity stands as one of the most crucial arguments linked to war, as reported in cluster 1. The war hits Ukraine's vital agriculture, particularly wheat cultivation, causing widespread economic crises. Moreover, in cluster 2, abstracts emphasize economic concerns, specifically energy-related issues.

Cluster 3 addresses the examination of educational aspects as a research topic, highlighting a significant focus on the involvement of students and the potential impacts on education, with particular attention given to universities.

The exploration of geopolitical and diplomatic outcomes resulting from crises takes center stage in cluster 4, where an array of papers extensively analyze the repercussions. These scholarly works centered on examining the complexities of how crises can reshape the geopolitical landscape, offering valuable insights for future investigation.

Another significant factor to take into account is the healthcare dimension. It is a well- established fact that hospitals have been subjected to attacks since the beginning of the war, aggravating the situation for the people, especially children. This represents the central theme identified within cluster 5.

Furthermore, news related to the war have been instantly spread on a global scale through social media platform. This dimension, in conjunction with its effects and the consequential impact on the internet, comprises the subject of cluster number 6.

Lastly, cluster number 7 is about the ensemble of global and stock markets' reactions. The concept of this group is closely aligned with the one of cluster 2: indeed, it is worth noting that the application of multidimensional scaling is crucial in highlighting the semantic proximity among clusters.

4. Conclusion

In this article, we present a case study on how scientific abstracts can be evaluated to uncover the primary addressed arguments related to a specific topic. By adopting a mixture of cosine distance-based density functions and employing tools for textual analysis, the influence of the Russia-Ukraine conflict is deepened by the scientific community. Our results reveal the specific areas in which researchers aimed to provide additional answers and extend support. The central theme of the discussion lies on the conflict's economic and financial implications; particular attention was directed towards the healthcare sector, which suffered and continues to suffer notable effects.

However, several limitations of our analysis need to be highlighted. Firstly, the choice of terms used in the queries for selecting the sample of abstracts can impact both the results and the conclusions drawn. Depending on the specificity or ambiguity of the terms, different sets of documents could be gathered,

leading to different clusters. A second crucial factor to be mindful of is the choice of language. In this work, by taking into account that English is the predominant language in the scientific community, we focused only on abstracts written in that language. However, it is essential to acknowledge the potential impact of considering other languages, such as Ukrainian and Russian. These languages may provide valuable insights and perspectives that could enhance the comprehensiveness and diversity of the study. By extending our analysis to include multiple languages, the possibility of capturing a broader range of scientific contributions and ensuring a more representative understanding of the research landscape is opened up.

Finally, as a prospective avenue for further development in our analysis, the incorporation of bigrams or even trigrams deserves consideration. By exploring them, our analysis may enrich the semantic representation of the data, providing a more detailed perspective on the subject matter.

References

Anderlucci, L., Montanari, A., Viroli, C. (2019). The importance of being clustered: uncluttering the trends of statistics from 1970 to 2015. *Statistical Science*, **34**(2), pp. 280-300.

Atanassova, I., Bertin, M. and Larive`re, V. (2016). On the composition of scientific abstracts. *Journal of Documentation*, **72**(4), pp. 636-647.

Blei, D.M., Ng, A.Y., Jordan, M.I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research,*
3(Jan), pp. 993-1022.

Blei, D.M. (2012). Probabilistic topic models. *Communications of the ACM*, **55**(4), pp. 77-84.

Dos Santos, M. (1996). The textual organization of research paper abstracts in applied linguistics. *Text & Talk*, **16**(4), pp. 481-500.

Dhillon, I.S., Modha, D.S. (2001). Concept decompositions for large sparse text data using clustering. *Machine Learning.*, **42**(1-2), pp. 143-175.

Ji, P., Jin, J. (2016). Coauthorship and citation networks for statisticians. *Annals of Applied Statistics*, **10**, pp. 1779-1812.

Kaufman, L., Rousseeuw, P.J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley &Sons.

Rousseeuw, P.J. (1987). Silhouettes, a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, **20**, pp. 53-66.

doi.org/10.26398/asaproc.0063

Youth and science: A randomized survey experiment on adolescents' trust in scientific institutions

Elena Sestini[a], Alessio Muscillo[a], Gabriele Lombardi[b], Francesca Garbin[c], Paolo Pin[a,d]

[a] Department of Economics and Statistics, University of Siena, Siena, Italy.
[b] "Giuseppe Parenti" Department of Statistics, Computer Science, and Applications, University of Florence, Florence, Italy.
[c] Charles River Associates (CRAI), 80802 Munich, Germany.
[d] Bocconi Institute for Data Science and Analytics, Bocconi University, Milan, Italy;

1. Introduction

Public erosion of trust in science is a controversial topic for everyday life, as well for research advancement, especially when experts misconduct becomes visible. People are very sensitive to a large number of factors which contribute to the weakening of public trust. Increasing lack of transparency has resulted in quality uncertainty, which represents a detrimental factor for people ability to trust scientific outcomes (Vazire, 2017).

As we have experienced so far with the rise of COVID-19 pandemic, trust is needed to cope with emergency situations, but is also under threat due to it (Jennings et al., 2021). Trust plays a pivotal role in the effectiveness of government (OECD, 2017), enabling greater compliance with measures and vaccine adoption. In short, the public needs science (Goldenberg, 2022), and the scientific community depends on strong public support for its endeavours.

The present analysis aims at measuring the level of trust in scientific and medical institutions in a very challenging period, since answers were collected at the beginning of 2022 when COVID-19 still represented a global emergency. The study presents a randomized survey experiment designed to verify whether adolescents have been affected by the "replication crisis" and to discover possible ways to stimulate trust among young people. We decided to restrict our attention to this age range since we strongly support the theory that trust and other beliefs mainly modify when young people pass through early adolescence, facing evolving skepticism, self-consciousness and higher risk behaviours.

Several contributions have been devoted to trust in science evaluation, especially with the advent of the COVID-19 when trust represented a determinant factor of measures compliance. Anvari and Lakens (2018) examined whether informing people about three aspects of the replication crisis, namely replication failures, criticisms of questionable research practices, and reforms, affects how much people trust research. Despite this branch of research intensified with the rise of the pandemic, it remained less common for adolescents.

Kennedy et al. (2011) conducted a systematic review to assess the level of teenagers trust regarding health information on social media, discovering a complex interplay between trust in social media platforms, other users and health content.

Overall, very few studies have been devoted to the analysis of adolescents' trust in scientific institutions. Thus, we aim to provide a unique contribution to previous literature in the field, analysing a new dataset by the means of a randomized survey experiment on Italian high school individuals. The experiment presents two separate treatments: first, believing that trust in a person is able to facilitate operations that lead to benefits for the trustor (Freeman et al., 2020), we evaluate the reaction to either favourable or pessimistic medical response when some health issues emerge. Secondly, we focus on the impact that different key figures generate on young individuals trust on specific health recommendations, with a special attention to the influence

exercised by doctors, parents and friends on adolescents' beliefs. The study also assesses trust in public institutions like schools, political system and judicial bodies.

To address all these issues it is important to stress the way institutions communicate with adolescents and young adults. Research underlines the importance of implementing strategies to reduce the barriers in raising and discussing health-related concerns (Kim and White, 2018), as well enhancing transparency, necessary to sustain trust in health authorities and impede the spread of conspiracy beliefs (Sheluchin et al., 2020). Recent findings support the need for targeted communication strategies to improve vaccination rates (Muscillo et al., 2023).

The work is articulated in five sections: we offer a review of past contributions, then we describe the experiment design and data collection process, and in the end we present the methodology and discussion of empirical results

2. Experimental approach and methodology

The article offers a quantitative research design on adolescents trust, with data collected through a randomized survey experiment in the period February 2022 to April 2022. The experiment was performed online with Qualtrics and it was administered to Italian high school individuals thanks to the collaboration the no-profit foundation Fondazione Mondo Digitale (FMD), after the approval of the Ethical committee of the University of Siena. Schools self-selected their participation to the project, but single students were randomly assigned to the treatment.

The survey includes the following sections: after an introduction and consent request, it presents demographic questions (i.e. gender, age, region, school, parents' education) and information concerning habits and healthy lifestyle (i.e. alcohol, smoke, fast-food and mobile phone addiction). The core part is based on two different treatments: the "Personal Experience Treatment", designed to test the adherence to expert's suggestions concerning healthy lifestyle, and the "authority treatment", designed to evaluate the influence of different key figures on adolescents opinion. We also include three trust items (i.e. trust in science, trust in medicine and trust in future advances). Answers are collected on a 10-point Likert scale from "Strongly disagree" to "Strongly agree".

For the "Personal experience treatment" individuals have to imagine a dialogue with a doctor where after some days of strong stomach ache, they decide to ask for an expert opinion; each individual is required to assign a score from 0 to 10 to indicate how much he is convinced by the expert's diagnosis. The sample is randomly divided in two groups: on the one hand (T1) respondents receive a comforting response without any additional exams, while on the other (T2), they receive a pessimistic response, with the doctor recommending additional exams to investigate the cause of the problem.

Next, we set up another treatment to test whether different authorities are able to affect the way teenagers accept health recommendations, and we call it the "Authority treatment". In this setting, we provide an information concerning the negative impact of social network abuse on mental health, and respondents are randomly assigned to one of the three groups: they receive the information either from a doctor (T1), or from friends (T2) or from parents (T3). The goal is to discover which key figure is more effective in persuading with adolescents on health issues.

We build a randomized control trial to identify the main factors able to influence adolescents confidence and behaviours and we began with two hypothesis:

First hypothesis: "Trust in science is compatible with higher support of experts' suggestions", as trust is associated to larger measures compliance (Devine et al., 2021).

Second hypothesis: "Respondents who receive health recommendation by a doctor are more inclined to trust the suggestion rather than those who are advised by parents" (Fox et al., 2022).

The empirical analysis includes two different frameworks: the evaluation of trust and the estimation of the average treatment effect. For the measurement of trust we adopted an *ordered probit* model, since we use ordered variables going from 0 to 10 as dependent. Here we performed

three regressions: we worked with "trust in science", "trust in medicine" and "trust in future advances" respectively as dependent variables. We adopted demographic characteristics as controls.

For the second part of the analysis, the one concerning the average treatment effect, we chose a relatively simple linear model:

$$Y = \propto + T\beta + X\gamma + Z\delta + \varepsilon$$

where T represents the matrix of treatments (either "personal experience treatment" or "authority treatment"); X is a matrix of demographic controls and includes all the observable characteristics available from the survey (i.e. gender, age, school type, father and mother education), while Z represents a matrix of trust items (i.e. trust in science, trust in future scientific advances and trust in medicine). The dependent variable is represented by individual responses to the questions asked in the treatments: "response acceptance" and "response rejection", for the first treatment; how much they are going to follow the advice received, for the second treatment designed.

Since we are interested in the effect size, in the experimental framework the nature of the dependent variable does not have any implication and the outcome is unbiased and consistent (Gomila, 2021). For robustness, we also repeated the analysis by using an ordered probit model, which we do not discuss now since it provided the same outcomes. In addition, we performed the non-parametric Mann-Whitney U test, whose null hypothesis supports equal distribution of characteristics of different groups.

2. Empirical analysis and results

The sample under analysis consists of N=1,433 individuals who attend high schools in Italy: the majority attend lyceum (62%), with the rest split among professional institute (10%) and technical institute (28%). The sample consists of 608 males and 768 females, while the 4% prefers not to declare it. The geographical distribution is heterogeneous, with a higher concentration of participants in Piedmont, Lazio and Lombardy regions. Also the family background is quite heterogeneous, characterized by higher frequency for graduated mothers (29%) than graduated fathers (23%). We also collected some personal information about people habits (i.e. smoke, alcohol, phone and fast food addiction).

Overall, adolescents show a positive perception of scientific and medical institutions. We asked them how much they trust each of the trust item and answers have been collected on a 10-point Likert scale from "Strongly Disagree" to "Strongly Agree". Results exhibit high confidence in science (mean = 7.329, SD = 2.268) as well in medicine (mean = 7.069, SD = 2.114). The highest score has been recorded for trust in future discoveries (mean = 7.592, SD = 2.181), despite respondents display some concerns about the consequences of boundless advances of scientific research (mean = 5.240, SD = 2.780).

Due to the nature of the dependent variables, we adopted an ordered probit model to evaluate which controls have an impact on trust items. There is lower probability that females reach high scores of trust in science, in medicine and in future research outcomes. On the other hand, adolescents attending lectures at lyceum tend to have higher probability to trust science and medicine. Adolescents around 18 years old tend to trust more science and future advances. Moreover, the willingness to continue with scientific and medical studies produces a strong significant effect, increasing the likelihood of selecting higher scores of trust in science, in medicine and in future developments.

Concerning the treatments responses we checked for distributional differences with a χ^2 test: we rejected the null hypothesis of equality in distribution of the "personal experience treatment" (p-value= 0.024) and we strongly reject the null of the "authority treatment" for what concerns doctor's advice and parents' advise (p-value= 0.004), instead, we weakly reject

the null for doctor's advice and friends' advise (p-value= 0.079) and we did not observe any difference among parents and friends affecting individuals behaviours.

Table 1: Personal experience treatment regression (standard errors in parentheses)

	Dependent variable:			
	Response acceptance		Response rejection	
	(1)	(2)	(1)	(2)
T2 (negative feedback)	0.223∗	0.023**	1.455***	1.476***
	(0.116)	(0.117)	0.161	(0.158)
Female	-0.275∗∗	-0.136	0.314*	0.231
	(0.134)	(0.124)	(0.186)	(0.183)
Lyceum (school)	1.342*	0.066	-0.162	0.534
	(0.784)	(0.694)	(1.035)	(1.031)
Mother's ed. high school	0.478*	0.196	-0.265	-0.098
	(0.228)	(0.228)	(0.341)	(0.247)
Trust in science		0.067**		-0.102**
		(0.032)		(0.047)
Trust in future advances		0.112***		0.019
		(0.033)		(0.048)
Trust in medicine		0.305***		-0.142***
		(0.030)		(0.043)

*** p< 0.01, ** p< 0.05, * p<0.10

For the estimation of the average treatment effect, we performed a linear regression, taking as dependent the individual response recorded a 10-points Likert scale ("response acceptance" and "response rejection respectively"). Regression outcomes concerning the "personal experience treatment" are presented in Table 1. For each regression all the control variables (i.e. gender, age, school, parents' education) have been considered. In column 1 we consider as explanatory variables the treatment dummy and demographic characteristics: we can observe a stronger tendency to reject the expert's recommendation when they receive a negative feedback; in addition, the effect is larger for females.

For robustness we performed a new regression checking also for trust items, and results presented in columns (2) concerning the effect of the treatment are consistent with previous findings. Moreover, we observe that trust items play a very fundamental role in response acceptance. Higher trust in science, in medicine and in future scientific advances produces a positive effect in response acceptance, while it generates an opposite effect in response rejection. From these results, we can conclude that young individuals are in general more prone to reject an health recommendation when it includes additional controls, and that at the same time, having higher trust in scientific institutions is a key predictor of medical experts' recommendations.

Table 2 reports the results for "authority treatment". The dependent variable is the individual response to the question "How much are you going to follow the healthy advice concerning social network abuse and its consequences on mental stability?" delivered either by a doctor, or friends or parents. Column 1 includes only the treatment dummy and exhibits a negative statistical significant effect of both friends (p-value = 0.001) and parents (p-value = 0.000) on doctor's advice. Column 2 collects regression results adding all demographic controls previously specified, while Column 3 also includes trust items. Both outcomes confirm previous findings, keeping high statistical significance: individuals who receive a suggestion from friends and parents are less likely to follow it rather than those who are recommended by a doctor. In addition, in Column 2 we observed a lower tendency to trust doctor's advice for

females; in Column 3 instead we get a strong positive effect of trust in science and trust in medicine in following doctor's recommendation. We also employed an Ordered Probit model or both treatments analysis and it delivered very similar results.

Table 2: Authority treatment regression (standard errors in parentheses)

	Dependent variable		
	(1)	(2)	(3)
Friends	-0.589***	-0.536***	-0.566***
	(0.180)	(0.179)	(0.175)
Parents	-0.701***	-0.656***	-0.632***
	(0.180)	(0.180)	(0.175)
Female		-0.356**	-0.234
		(0.169)	(0.166)
Trust in science			0.115***
			(0.042)
Trust in future advances			0.051
			(0.043)
Trust in medicine			0.206***
			(0.039)

*** p< 0.01, ** p< 0.05, * p<0.10

For robustness, we relaxed some assumption and we adopted the Mann-Whitney U test to check for treatment effects. The non-parametric test confirmed the outcomes obtained with previous regressions: for "Personal Experience Treatment" we find out a strong significance difference in distributions among the two groups (p-value = 0.000).

Concerning the "Authority Treatment" we cannot accept the null hypothesis of no difference in distribution of values among those receiving the news from doctor and from friends (p-value = 0.001); even much stronger difference among doctor and parents (p-value = 0.000); no significant difference among friends and parents' advice.

2. Conclusions

With the current analysis we aim to provide valuable results to the branch of literature dedicated to adolescents' trust, which has not been extensively explored thus far. The experiment offers interesting results concerning the degree of trust in science and in medicine among young individuals, as well it provides useful insights on how effectively foster trust in public institutions in young generation, relevant for policy recommendations.

What emerges from the study is that Italian adolescents, although the large inconvenience caused by the outbreak of the pandemic, are very confident in the role played by scientific research and strongly support public institutions like hospitals, research centres and school. In policy terms, higher investments should be devoted to the education system and to those activities committed to enhance trust building, given the massive effects that public trust generates in the society.

References

Anvari, F., Lakens, D. (2018). The replicability crisis and public trust in psychological science. *Comprehensive Results in Social Psychology*, **3**(3), pp. 266–286.

Devine, D., Gaskell, J., Jennings, W., Stoker, G. (2021). Trust and the coronavirus pandemic: What are the consequences of and for trust? an early review of the literature. *Political Studies Review*, **19**(2), pp. 274–285.

Fox, K.R., Bettis, A.H., Burke, T.A., Hart, E.A., Wang, S.B. (2022). Exploring adolescent experiences with disclosing self-injurious thoughts and behaviors across settings. *Research on Child and Adolescent Psychopathology*, **50**(5), pp. 669-681.

Freeman, J.L., Caldwell, P.H., Scott, K.M. (2020). The role of trust when adolescents search for and appraise online health information. *The Journal of Pediatrics*, **221**, pp. 215–223.

Goldenberg, M.J. (2022). Public trust in science. *Interdisciplinary Science Reviews*, **48**(9), pp. 1–13.

Gomila, R. (2021). Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis. *Journal of Experimental Psychology: General*, **150**(4), pp. 700.

Jennings, W., Stoker, G., Valgarsson, V., Devine, D., Gaskell, J. (2021). How trust, mistrust and distrust shape the governance of the covid-19 crisis. *Journal of European Public Policy*, **28**(8), pp. 1174–1196.

Kennedy, K.J., Mok, M.M.C., Wong, Y.W.M. (2011). Developing political trust in adolescents: Is there a role for schools? In B.R. Curtis (Ed.). *Psychology of Trust* (pp. 137-156). New York: Nova Science Publishers.

Kim, B., White, K. (2018). How can health professionals enhance interpersonal communication with adolescents and young adults to improve health care outcomes?: systematic literature review. *International Journal of Adolescence and Youth*, **23**(2), pp. 198–218.

Muscillo, A., Lombardi, G., Sestini, E., Garbin, F., Tambone, V., Campanozzi, L.L., Pin, P. (2023). Adolescents' opinions on covid-19 vaccine hesitancy: Hints toward enhancing pandemic preparedness in the future. *Vaccines*, **11**(5), pp. 967 https://doi.org/10.3390/vaccines11050967.

OECD (2017). *Trust and Public Policy: How Better Governance Can Help Rebuild Public Trust*. OECD, Paris.

Sheluchin, A., Johnston, R.M., van der Linden, C. (2020). Public responses to policy reversals: The case of mask usage in Canada during covid-19. *Canadian Public Policy*, **46**(S2), pp. S119–S126.

Vazire, S. (2017). Quality uncertainty erodes trust in science. *Collabra: Psychology*, **3**(1):1

doi.org/10.26398/asaproc.0064

Data science to support decision makers

Sandro Stancampiano
Istat, Rome, Italy

1. Introduction[1]

This paper aims to present a research that extracts information from reviews published on Google Maps. People use this application in order to discover new places and experiences. When people explore locations on Google Maps, they see content provided by other users – such as reviews and photographs – to get a better understanding of those places. On Google Maps, you can read and write reviews. Comments and opinions are voluntary. Google does not pay reviewers to add reviews to Google Maps. We assume that people write only if they genuinely want to share their point of view about a location. Furthermore, Google enforces a strict policy that aims to remove deceptive content, which can lead to misinformation, fake engagement, and misrepresentation. These removal measures help maintain reviews that are relevant, helpful, and trustworthy.

Through writing reviews, people function as human sensors, providing information based on their observations. With a vast number of reviews being published daily, attempting to read them all becomes an impossible task. This is where text mining comes into the picture.

Text mining enables us to convert unstructured data into actionable knowledge, which both administrators and citizens can utilize to make informed decisions. Textual information published on the Internet can serve as a significant foundation for projects aimed at alleviating statistical burden.

When analysing reviews published on Google Maps, we take into consideration not only the actual content (which includes topic and sentiment) but also other factors such as the user's rating, the timestamp of the review, its popularity (determined by the number of likes), and, of course, the specific location. For instance, when examining reviews about the Colosseum in Rome, we compare all the aforementioned data points. Our approach involves topic modeling to identify the main subjects being discussed. Subsequently, we search for pertinent text documents and compile a comprehensive report that can be utilized by public administrators.

2. Methodology

Topic modeling presents an unsupervised algorithmic approach for exploring collections of short texts. Among these algorithms, latent Dirichlet allocation (LDA) stands out, driven by two core principles: 1) each document comprises a blend of various topics, and 2) each topic consists of a mixture of words (Silge and Robinson, 2022).

In his paper, Blei outlines the objective of topic modeling: the automated discovery of topics within a document collection. He defines a topic as a distribution over a fixed vocabulary (Blei, 2011). This statistical methodology acknowledges the inherent presence of multiple topics in documents.

For this study, our objective is to determine the topics of discussion among tourists in their comments and to assess the utility of their observations. We have gathered a collection of reviews about the Colosseum spanning from 2019 to the first semester of 2023. Our focus centers on analysing 735 reviews from the year 2022.

The analysis consists of four phases:

[1] The views expressed in this paper are those of the author and do not necessarily reflect the policies of Istat (Italian National Institute of Statistics).

1. Initially, a corpus of reviews is formed using the *quanteda*[2] package, and pre-processing operations are executed.
2. Following this, the LDA algorithm is employed to create a topic model. LDA objects from the R package are utilized for this purpose.
3. Subsequently, an exploration of word-topic probabilities and document-topic probabilities is undertaken.
4. Lastly, reviews with the highest probability of belonging to each identified topic are selected, based on the outcomes of the preceding step.

3. Results

In our analysis, we used LDA, which estimates each topic as a combination of various words. It also models each document as a blend of these topics. Hence, a document is not limited to a single topic, but rather comprises a mixture of multiple topics (Blei et al., 2003).

We have identified four distinct topics in the discussions and linked them with their respective ratings and likes. The topics are distinct and cover various aspects of the Colosseum. They range from management and organization of the visit to Colosseums, to expressions of feelings about it, and historical notes related to it. This variety suggests a comprehensive analysis covering practical, emotional, and historical perspectives.

Table 1 provides a detailed breakdown of our analysis, with the column "Number of documents (N. Doc) with Gamma > 0.9" specifically indicating the number of documents strongly related to each identified topic. The gamma (γ) parameter estimates what proportion of words in a document are attributed to a specific topic. For instance, if a document (d1) has a gamma value of 0.1 for topic 1 and 0.9 for topic 2, it is more associated with topic 2 than topic 1. Therefore, we infer that documents with a gamma value exceeding 0.9 for a particular topic are predominantly about that topic. The columns titled "average likes" and "standard deviation likes" quantify the level of engagement or popularity of each topic among the audience. The columns labelled "average rating" and "standard deviation of rating" represent the mean rating and the variability of ratings of a given topic, respectively. These metrics offer insight into the overall visitor satisfaction and the consensus level among the reviews for each topic.

The "expressing feelings" topic has the highest number of strongly related documents (100), suggesting it might be a popular aspect among discussant. The topic records the lowest average likes (0.5), despite its high number of related documents and high average rating. This suggests that while visitors frequently discuss their feelings and rate their overall visit experience highly, this aspect may not be as engaging or resonate as strongly with the broader audience. The average number of likes (0.5) is notably lower compared to other topics, which could indicate a lower level of engagement of these reviews among readers.

In contrast, "historical notes" has significantly fewer documents (26), which could imply it is a less represented topic. The limited number of reviews could also explain the lack of variability in the ratings. This topic shows exceptionally high appreciation with uniformly perfect ratings. However, the moderate and varied likes suggest that these reviews, similar to those in the "expressing feelings" topic, might not strongly engage other users. Among the documents showing a per-document-per-topic probability (referred to as 'gamma') of over 0.9 for a particular topic, the average rating undergoes an increase from 4.5 to 5. The average rating shows high appreciation for all topics.

A slight decrease in the average rating was observed between "historical notes" (topic 4) and "management and organization of the Colosseum" (topic 1). The average rating declined from 5 in topic 4 to 4.5 in topic 1, on a 1 to 5 scale. This decrease of half a point could suggest the presence of problematic issues in the documents associated with topic 1. Moreover, a standard deviation of around 0.9, while not excessively high, indicates a certain variety in opinions. Not all

[2] See https://quanteda.io for a complete description of quanteda R package.

reviews are uniformly positive. This variability could suggest the presence of specific aspect within the topic that do not fully satisfy some visitors.

The data for "management and organization of the Colosseum" indicates significant visitor engagement, as reflected by 93 documents with a high topic relevance ('gamma' value over 0.9). This aspect of the Colosseum experience received an average rating of 4.5 and an average of 1.1 likes, suggesting it is fairly well regarded among visitors.

While some topics are popular in terms of the number of reviews (such as topic 3), they do not always correspond to high engagement as measured in 'likes.' This could suggest that, although some subjects are commonly discussed, they may not necessarily evoke strong emotional or social reactions, such as 'likes.'

The greater variability in ratings and 'likes' in topics 1 and 2 might reflect broader differences in visitors' experiences and expectations regarding both the organization and management of the Colosseum and the visit itself.

Table 1: Topic analysis *vs.* rating and popularity

Topic	Description	N.Doc γ > 0.9	Avg ratings	Avg likes	Std Dev ratings	Std Dev likes
1	Management and organization of the Colosseum	93	4.5	*1.1*	*0.9*	*5.2*
2	Management and organization of the visit	84	4.9	*1.2*	*0.5*	*5.8*
3	Expressing feelings	100	4.9	*0.5*	*0.4*	*1.3*
4	Historical notes	26	5	*0.9*	*0.0*	*2.4*

Table 2: Topic 1 - Management and organization of the Colosseum

Review	Rating	Likes
Ovviamente ne vale la pena, prenotare online è necessario. State attenti a false guide e truffatori tutto intorno il Colosseo	5	0
Purtroppo non sono riuscita ad entrare, bisognava prenotarsi almeno 1/2 giorni prima... ma del resto come quasi tutti i punti culturali il weekend, anche se ad ingresso gratuito... suppongo siano restrizioni preventive per il covid quindi capisco	4	5
4 stelle perché il monumento ne merita 10 ma l'organizzazione è tristemente pessima. Varco d'entrata piccolo e ti respingono anche se arrivi come espressamente indicato entro 30 minuti dall'orario indicato. Praticamente convinse presentarti all'orario e basta. Ho trovato in molti punti del percorso più calca che nella metropolitana. Ho fatto altra coda per comprare l'audio guida (15 minuti di coda) ed è carina	4	1
Il posto non si deve neanche commentare, peccato per la gestione prenotazioni, un pò difficoltosa online, il sito fa pena! Potrebbero scrivere che non fanno più entrare nessuno da quando manca un'ora alla chiusura.	3	0
Gestito malissimo. Biglietto solo online quindi non acquistabile sul posto. File interminabili per l'accesso considerato che rispetto all'enorme flusso di turisti dovrebbero esserci almeno tre varchi. Insopportabile poi l'impertinenza di pseudo guide turistiche che offrono ad ogni passo presunte visite guidate e in alcuni casi anche biglietti introvabili online	1	3

The data suggest that while certain aspects of the Colosseum, like expressed sentiments and historical notes, are highly rated, others, such as its organization and management, elicit a wide range of opinions. This indicates potentially critical areas where improvements could significantly

enhance the overall visitor experience. Moreover, engagement measured in 'likes' does not always align with the frequency or positivity of the reviews, indicating diverse dynamics in how visitors interact with online reviews. However, the ratings also imply there might be room for improvement in this area to further enhance overall visitor satisfaction.

In Tables 2, 3 and 4, we show the most significant comments representing each subset within the four identified topics, as per our assessment. The first two groups share a similarity in their focus but approach the subject from distinct perspectives. The former group holds potential value for cultural heritage administrators (Table 2), whereas the latter group offers insights beneficial for other tourists seeking to enhance their visit experience (Table 3).

Table 3: Topic 2 - Management and organization of the visit

Review	Rating	Likes
Esperienza unica. La guida Simona simpaticissima e molto molto brava!! Ho capito tutto ciò che è stato spiegato, fantastico	5	0
È assolutamente da visitare…quanta storia.... fantastico	5	0
Visitato di notte (da fuori). Meraviglioso!	4	1
Il Colosseo è Roma Visitarlo almeno una volta nella vita è d'obbligo.	5	2
Piccolo consiglio: se si è di fretta non consiglio di visitarlo internamente, non offre la stesss magnificenza di fuori.	4	1

Table 4: Topic 3 - Expressing feelings

Review	Rating	Likes
Quando lo vedi rimani a bocca aperta ... Questa sarà la quinta volta ma la sensazione è sempre la stessa ... Imponente maestoso qualcosa di meravigliosamente bello .... Roma	5	1
Imponente, magico, incredibile. Uno se non IL punto di riferimento della città di Roma, dell'Italia e del Mondo. […]. Ancora più affascinante quando illuminato, di notte. Da vedere e rivedere, da ammirare. Tappa OBBLIGATORIA visitando la Città Eterna.	5	10
Roma e i suoi monumenti è bella tutto l' anno, il Colosseo sempre magico	5	0
Come recensire il Colosseo? Se andate a Roma non potete non visitarlo, un luogo straordinario e unico al Mondo, ricco di storia, molto più impressionante dal vivo di quanto non appaia nelle foto!	5	0
Oltre alla classica passeggiata interna giornaliera, ammirate questa meraviglia la sera l'atmosfera è stupenda.	5	0

Within the first subset, tourists express that there are significant issues with online booking, site organization, and dealing with unauthorized activities (false guides and scammers). On the other hand, the second group of reviews (Table 4) highlights tourists' recommendations for guided tours, noting the clear and comprehensive explanations provided during such tours.

The reviews categorized within the third group ("expressing feelings") delve into aspects such as strolls, sentiments and emotions. In these reviews, thoughts and personal reflections play a pivotal role. On the other hand, the fourth group adopts a didactic approach: Upon initial inspection, it appears that these tourists have studied the subject prior to (or possibly after) their visit. It is conceivable that they compose their comments while seating at their desks, referencing sources like Wikipedia or a Colosseum guide to provide historical and architectural insights.

Table 5: Topic 4 - Historical notes

Review	Rating	Likes
Uno dei luoghi più belli ed importanti da visitare a Roma. Situato nel centro della città di Roma, è il più grande anfiteatro del mondo. In grado di contenere un numero di spettatori stimato tra 50 000 e 87 000 unità, è il più importante anfiteatro romano, nonché il più imponente monumento dell'antica Roma che sia giunto fino a noi. […] La struttura esprime con chiarezza le concezioni architettoniche e costruttive romane della prima Età imperiale, basate rispettivamente sulla linea curva e avvolgente offerta dalla pianta ellittica e sulla complessità dei sistemi costruttivi. Archi e volte sono concatenati tra loro in un serrato rapporto strutturale.	5	0
È il simbolo di Roma l'eccellenza dei monumenti più visitati	5	0
È una delle costruzioni più spettacolari ed imponenti che ci siano. Situato al centro della città di Roma il Colosseo ne è il simbolo indiscusso. È il più grande Anfiteatro di romano al mondo. E' stato inserito fra le 7 meraviglie del mondo moderno nel 2007 e questo aiuta a ricordare la grandezza e i successi che l'impero romano era riuscito a raggiungere	5	4
È il monumento simbolo della Romanità, è stato l'ombelico del mondo romano, il centro della Storia, ce ne sono passate di storie tra queste mura.	5	0

4. **Deciphering tourist feedback: key conclusions**

We have embarked on an exploration of topic models as a solution for effectively managing an extensive repository of reviews focused on the Colosseum. The application of the LDA technique holds the potential to uncover genuine insights from these reviews, shedding light on various points of interest. This same methodology can be adapted to investigate, condense, and visually represent collections of reviews pertaining to other sites of significance.

The significance of topic modeling in this context cannot be overstated. It serves as a filter that sieves through the avalanche of reviews, enabling decision-makers to uncover prevalent sentiments, recurring topics, and emerging concerns. This distilled perspective eliminates the arduous task of manually sifting through an overwhelming volume of reviews, providing a macroscopic yet nuanced understanding of visitor impressions and concerns. Linking to our previous discussion, the document on the investigation initiated by the Italian Competition Authority (AGCM) into the ticket sales for the Colosseum Archaeological Park sheds light on issues related to topic 1[3].

The investigation addresses the difficulty consumers face in purchasing tickets from the official retailer, CoopCulture. Consumers struggle to buy tickets from the official seller due to bulk purchases by secondary sellers using automated systems. In topic 1, reviews feature complaints about ticket prices and difficulties in purchasing them online. This is echoed by recent media reports on the issue of secondary ticketing[4], which has also been addressed legislatively (Legislative Decree of March 7, 2023, No. 26, implementing EU Directive 2019/2161). These tickets are then resold at higher prices on platforms like Musement and GetYourGuide, often bundled with extra services. This practice prevents consumers from accessing standard-priced tickets, aligning with the complaints of high prices and online purchasing difficulties discussed in topic 1.

Moreover, topic modeling's capacity to adapt and evolve over time ensures its relevance in a dynamic landscape. This adaptability ensures that decision-makers remain informed about

[3] https://www.agcm.it/media/comunicati-stampa/2023/7/PS12603

[4] https://tg24.sky.it/economia/2023/07/19/bagarini-online-colosseo-antitrust
https://roma.repubblica.it/cronaca/2023/07/18/news/colosseo_biglietti_secondary_ticketing_antitrust_indagine_finanza_concorrenza-408160957/

shifting perceptions, allowing them to adjust strategies and prioritize interventions in alignment with visitor expectations.

In summary, topic modeling stands as an indispensable tool in the pursuit of comprehending the diverse tapestry of reviews about the Colosseum on platforms like Google Maps. It not only makes the seemingly insurmountable task of review analysis feasible but also enhances the decision-making process by providing a succinct and holistic overview of visitor sentiments. By harnessing the power of topic modeling, we transcend the limitations of information overload, ensuring that the voice of each tourist contributes meaningfully to the preservation, enhancement, and understanding of this iconic cultural heritage site.

Given the paramount importance of preserving Italy's cultural heritage, we recognize the necessity to harness every available resource for managing our historical legacy. Text mining emerges as a valuable component in this process, offering increased efficiency and the potential for seamless collaboration with complementary tools (Stancampiano, 2023). As part of our ongoing efforts, we are extending our study to encompass additional cultural assets, with the aim of providing administrators with the means to optimize their decision-making procedures.

References

Blei, D.M., Ng A.Y., Jordan M.I. (2003). Latent Dirichlet allocation. *The Journal of Machine Learning Research*, **3**, pp 993-1022.

Blei, D. (2011*). Introduction to Probabilistic Topic Models*, Communications of the ACM. 55.

Silge J., Robinson, D. (2017). *Text Mining with R: A Tidy Approach (1st. Ed.).* O'Reilly Media, Inc.

Stancampiano, S. (2023), Text mining e turismo culturale: l'analisi testuale delle recensioni, in *Il turismo culturale in Italia: analisi territoriale integrata dei dati*, Istat, Letture Statistiche: Territori, Roma, pp.116-130.

doi.org/10.26398/asaproc.0065

The scientific production of social frailty: A bibliometric analysis

Agostino Stavolo[a], Giulia Cavrini[b], Marina Marino[a]
[a] Department of Social Science, Federico II University of Naples, Naples, Italy.
[b] Department of Education, Free University of Bozen, Bozen, Italy.

1. Introduction

In the 20th century, Western countries underwent a significant demographic transformation characterized by increased life expectancy and declining fertility rates (Cisotto et al., 2022). This resulted in an aging population, notably prominent in Italy, which became the oldest country in Europe.

The concept of frailty emerged as a major public health concern, referring to a condition of physical vulnerability and the negative health consequences associated with aging. There is an ongoing debate regarding the nature of frailty, with some models emphasizing physical aspects while others highlight its multidimensional nature (De Vries et al., 2011). Integrated models recognize frailty as a dynamic biopsychosocial phenomenon, considering the accumulation of deficits across physical, psychological, and social domains. Social frailty specifically refers to a lack of engagement in social networks and the perceived absence of connections and support (Andrew et al., 2008). While physical and psychological frailty have been extensively explored, social frailty remains relatively understudied (Bunt et al., 2017).

To enhance our understanding of this subject, it is crucial to map the existing scientific knowledge within the field. This involves conducting a comprehensive review of the relevant literature, synthesizing previous research findings, and identifying potential areas for future investigation. Bibliometrics, a statistical measurement-based methodology, serves as a valuable tool to facilitate this review process, providing a systematic, transparent, and replicable approach to assess scientific production, researchers, and scientific activities (Cuccurullo et al., 2016). Bibliometric methods have been widely applied across various research domains to evaluate the impact of the field itself, specific researchers, papers, and journals. They offer insights into knowledge inputs, research gaps, emerging trends, and future opportunities.

In this study, we employ a prominent approach to highlight both structural and cognitive patterns within the research field, namely science mapping (Aria et al., 2020). Science mapping allows us to identify, visualize, and analyse the themes and trends in a research domain, both synchronically and diachronically, without repetition (Cobo et al., 2011). With this approach, we aimed to map the social frailty research field and assess its scientific productivity.

2. Methodology

To retrieve the scientific production, we employed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) model, as illustrated in Figure 1 (Liberati et al., 2009). PRISMA offers a comprehensive framework that outlines inclusion and exclusion criteria for article selection in a systematic literature review, ensuring transparency and reproducibility in the selection process. The Web of Science (WoS) indexing database, initially launched by the Institute for Scientific Information (ISI) and currently maintained by Clarivate Analytics, was used to retrieve the articles. Specifically, the Science Citation Index Expanded (SCI expanded) and the Social Science Citation Index (SSCI) within the WoS core collections were employed for this purpose.

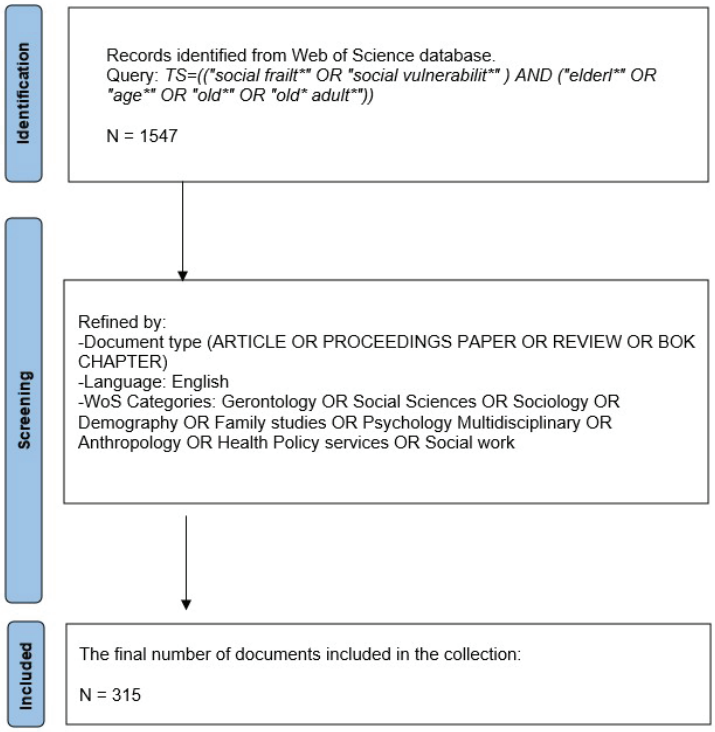


Figure 1: PRISMA diagram flow

The query used for documents search is as follow: TS=(("social frailt*" OR "social vulnerabilit*") AND ("elderl*" OR "age*" OR "old*" OR "old* adult*")). The TS tag (Topic) searches for terms from the query in the titles, abstracts, and keywords fields of documents indexed on WoS. Quotation marks and asterisks serve as wildcards commonly used in search queries. Quotation marks retrieve records that contain the exact term sequence in the TS fields, while the asterisk allows for finding various forms of the associated term.

The collection of documents was gathered in early July 2023. To refine the search, only Articles, Proceedings papers, Review articles, and Book chapters in English published from 1998 to 2022 were selected. Bibliographic data, including titles, abstracts, author names, keywords, and cited references, were collected. The documents were exported to PlainText format. Documents without abstracts and those related to the medical and psychiatric fields were excluded as the goal is to focus solely on productions in the social field, resulting in a total of 315 retrieved documents. The bibliometrix R open-source package (Aria and Cuccurullo, 2017) was utilized to perform bibliometric analyses on the entire collection.

Specifically, this study utilized bibliometric analysis to examine the conceptual structure of publications in a specific scientific field. The analysis involved two main techniques: co-occurrence network analysis and thematic mapping. The co-occurrence network analysis involved constructing a co-occurrence matrix to quantify the frequency with which terms appeared together in the documents (Wang et al., 2019). The resulting matrix was represented as a weighted network, where nodes represented terms and edges denoted associations between linked terms. Community detection (Fortunato, 2010) was performed using the Walktrap algorithm (Pons and Latapy, 2006).

Thematic mapping was used to visually represent these identified themes on a bi-dimensional map (Cobo et al., 2011). The x-axis measured Callon centrality, indicating the significance of a theme within the research field, while the y-axis measured Callon density, indicating the level of theme development (Callon et al., 1983). This mapping approach allowed for the categorization of themes into four types based on their location on the map. The first quadrant of the map, located

in the upper right, consists of motor themes that are highly significant and well-developed, as indicated by their high centrality and density. The second quadrant, located in the upper left, represents isolated or niche themes that are highly developed internally but lacked significant external links, resulting in low centrality and limited importance for the broader research field. The third quadrant, located in the lower left, comprises emerging or declining themes with low centrality and density, indicating weak or marginal development. The fourth quadrant, located in the lower right, identifies basic and transversal themes with high centrality and low density, representing principal issues cutting across different research areas.

Each theme is represented as a network cluster, with the bubble name reflecting the word with the highest occurrence value within the cluster and the bubble size indicating the proportion of the cluster's word occurrences. This comprehensive analysis provided insights into the conceptual structure of the field and the relationships between different research themes.

3. Main findings

Starting from 1,547 documents, several filters were applied to refine our collection to 315 publications published from 1998 to 2022. As we can see in Table 1, the documents were published by 1355 authors across 172 distinct sources, encompassing both journals and books. The average number of citations per document published within the reference timespan (1998-2022) stands at 22.13. The publication's annual growth rate is calculated at 7.43%. As shown in Figure 2, scientific research pertaining to social fragility has been experiencing a notable upward trend, with notable peaks in 2014 and 2020, marked by a significant increase in literature. However, it is from the year 2021 onward that we observe a remarkable surge in research activity concerning this topic.

Table 1: Basic information about the collection

OVERVIEW	
Timespan	1998:2022
Sources (journals, books, etc)	172
Documents	315
Annual growth rate %	7.43
Document average age	5.36
Average citations per doc	22.13
References	14967
DOCUMENT CONTENTS	
Keywords Plus (ID)	985
Author's keywords (DE)	921
AUTHORS	
Authors	1355
Authors of single-authored docs	29
AUTHORS COLLABORATION	
Single-authored docs	30
Co-authors per Doc	4.73
International co-authorships %	24.76%
DOCUMENT TYPES	
Article	284
Book chapter	5
Proceedings paper	16
Review	10

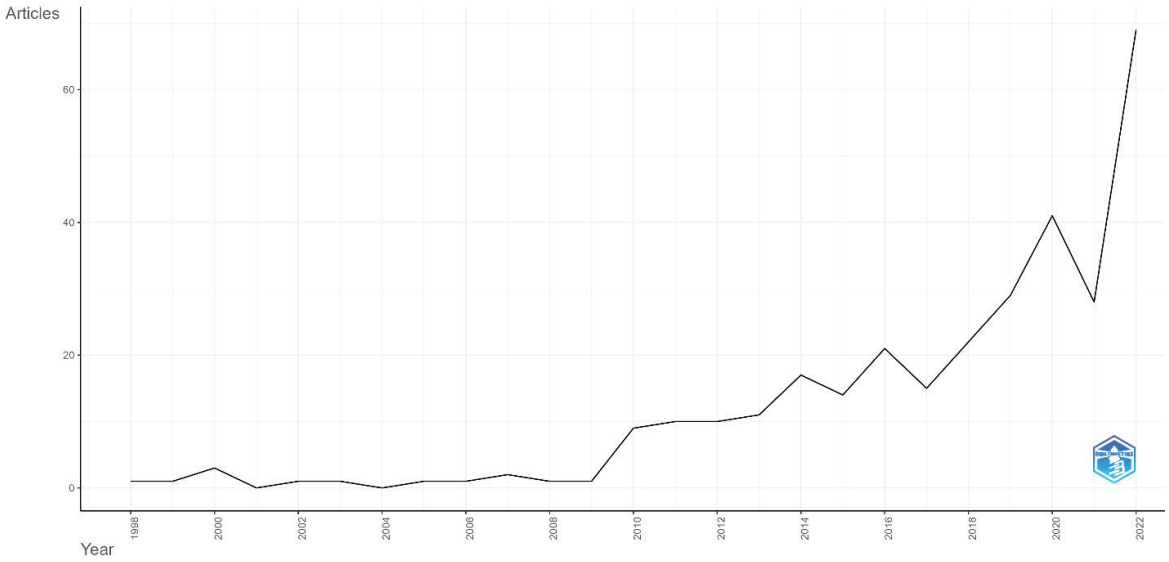


Figure 2: Annual scientific production from 1998 to 2022

To extract and analyse the concepts present in the collected scientific documents, we performed a thematic map (Figure 3). Keywords Plus (IDs) were used as the unit of analysis to build the co-occurrence matrix and, consequently, generate the thematic map. IDs represent words or phrases that commonly appear in the titles of article references to identify and categorize articles according to their content within the database.

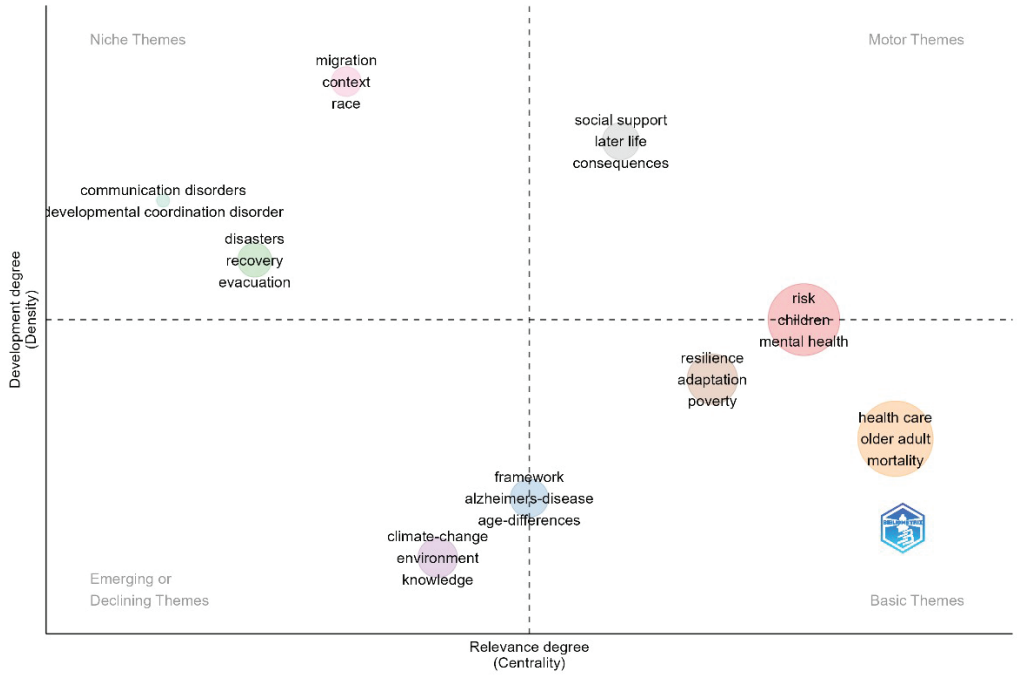


Figure 3: Thematic map on social frailty

One noteworthy topic, identified as a core theme, centres on the social implications of social support during adulthood, encompassing a wide range of dimensions. Researchers analyse the impact of social support on aspects such as social integration, community involvement, and social

connectedness. They investigate how social networks and support systems contribute to social capital and social cohesion within communities. Additionally, studies explore the impact of social support on interpersonal relationships, including family dynamics, friendships, and romantic partnerships. The primary themes in scientific production on social support in adulthood revolve around mental health in older people and the issues of poverty and social disparities. A substantial body of research has been conducted to understand the mental health challenges faced by older individuals, such as depression, anxiety, cognitive decline, and dementia. Researchers strive to pinpoint risk factors, assess their influence on well-being, and formulate effective interventions. Additionally, studies focus on socioeconomic factors contributing to poverty among older individuals, investigating the consequences on health, quality of life, and overall social well-being. Strategies and interventions are developed to address poverty and promote social equity among older adults (Hayajneh and Rababa, 2022).

Within the emerging theme, a compelling cluster within the scientific output is dedicated to climate change. This area of research recognizes the pressing necessity to address environmental challenges and their consequences on human health and well-being. Scholars are investigating the impacts of climate change on physical and mental health, identifying populations that are particularly vulnerable, and exploring strategies to mitigate these effects (McDermott-Levy et al., 2019). This line of inquiry contributes to the broader understanding of the interconnectedness between environmental sustainability and public health.

Finally, niche topics represent specialized areas of research within the broader field. Studies focusing on the migration of older people explore patterns, motivations, and challenges associated with older adults relocating from one place to another (Bradley and Van Willigen, 2010). Moreover, research on the social and psychological consequences of communication and coordination disorders during adulthood examines the impact of disorders like aphasia or apraxia on social interactions, daily functioning, and quality of life in older adults (Heine and Browning, 2002).

4. Conclusions and future developments

The paper presented a preliminary analysis of recurring themes within scholarly production on the topic of social fragility in the psycho-social field. Overall, the scientific production in the field of social support in adulthood encompasses a wide range of themes. Researchers investigate the physical, psychological, and social dimensions of aging, with particular emphasis on mental health, poverty, Alzheimer's disease, climate change, migration, and communication disorders. The goal was to generate knowledge that informs policies, interventions, and support systems to promote the well-being and overall quality of life for older adults.

Although the topic remains relatively underexplored, forthcoming endeavours will be directed towards conducting a longitudinal bibliometric analysis to investigate the evolution of social fragility topics over time.

References

Andrew, M.K., Mitnitski, A.B., Rockwood, K. (2008). Social vulnerability, frailty and mortality in elderly people. *PloS one*, **3**(5), e2232.

Aria, M., Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, **11**(4), pp. 959-975.

Aria, M., Misuraca, M., Spano, M. (2020). Mapping the evolution of social research and data science on 30 years of Social Indicators Research. *Social Indicators Research*, **149**, pp. 803-831.

Bradley, D.E., Van Willigen, M. (2010). Migration and psychological well-being among older adults: A growth curve analysis based on panel data from the health and retirement study, 1996-2006.

Journal of Aging and Health, **22**(7), pp. 882-913.
Bunt, S., Steverink, N., Olthof, J., Van Der Schans, C.P., Hobbelen, J.S.M. (2017). Social frailty in older adults: A scoping review. *European Journal of Ageing*, **14**, pp. 323-334.
Callon, M., Courtial, J.P., Turner, W.A., Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, **22**(2), pp.191-235.
Cisotto, E., Meli, E., Cavrini, G. (2022). Grandparents in Italy: Trends and changes in the demography of grandparenthood from 1998 to 2016. *Genus*, **78**(1), pp. 1-22.
Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E., Herrera, F. (2011). Science mapping software tools: Review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology*, **62**(7), pp. 1382-1402
Cuccurullo, C., Aria, M., Sarto, F. (2016). Foundations and trends in performance management. A twenty-five years bibliometric analysis in business and public administration domains. *Scientometrics*, **108**, pp. 595-611.
De Vries, N.M., Staal, J.B., Van Ravensberg, C.D., Hobbelen, J.S.M., Rikkert, M.O., Nijhuis-van der Sanden, M. (2011). Outcome instruments to measure frailty: a systematic review. *Ageing Research Reviews*, **10**(1), pp. 104-114.
Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, **486**(3-5), pp. 75-174.
Hayajneh, A.A., Rababa, M. (2022). The association of frailty with poverty in older adults: A systematic review. *Dementia and Geriatric Cognitive Disorders*, **50**(5), pp. 407-413.
Heine, C., Browning, C.J. (2002). Communication and psychosocial consequences of sensory loss in older adults: Overview and rehabilitation directions. *Disability and Rehabilitation*, **24**(15), pp. 763-773.
Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gøtzsche, P.C., Ioannidis, J.P., Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Annals of Internal Medicine*, **151**(4), pp. 65-94
McDermott-Levy, R., Kolanowski, A.M., Fick, D.M., Mann, M.E. (2019). Addressing the health risks of climate change in older adults. *Journal of Gerontological Nursing*, **45**(11), pp. 21-29.
Pons, P., Latapy, M. (2006). Computing communities in large networks using random walks. *Journal of Graph Algorithms Appl*ication, **10**(2), pp. 191-218.
Wang, H., Zhao, Y., Dang, B., Han, P., Shi, X. (2019). Network centrality and innovation performance: The role of formal and informal institutions in emerging economies. *Journal of Business & Industrial Marketing,* **34**(6), pp.1388-1400.

doi.org/10.26398/asaproc.0066

Critical issues in the delivery of information letters to the households of the population census in Italy: A data-driven approach to territorial monitoring

Raffaella Succi[a], Alessandra Nuccitelli[b]
[a] Data Collection Directorate, Istat, Genoa, Italy
[b] Data Collection Directorate, Istat, Rome, Italy

1. Introduction

In the Italian permanent population and housing census (PPHC), the postal delivery of information letters to the households plays a crucial role in achieving an adequate overall response rate and, above all, a satisfactory spontaneous response rate through CAWI (Computer Assisted Web Interviewing). The delivery of information material can be monitored in the field using a data-driven approach based on quality control tools (Jans *et al*., 2013). Indeed, the availability of data concerning the delivery outcomes at the postcode level makes it possible to apply statistical procedures to monitor the phenomenon in detail across the country in near real-time.

This work proposes a procedure to control the delivery of the information material relating to the PPHC carried out by the Italian Statistical Institute (Istat). The paper is organized as follows. Section 2 describes the role of the information letter in the context of the PPHC. Section 3 introduces the data available to control the delivery of information material and reports some results from the data collection process in 2022, in terms of delivered letters and survey outcomes. Then, the monitoring procedure is presented (Section 4), focusing on the proposed indicators and the analysis-of-means (ANOM) method, which is used to analyse them graphically at the postcode level and to uncover any problematic situations that need to be addressed quickly. With reference to the 2022 PPHC, some examples of ANOM charts are given in Section 5, while the main results concerning the most critical areas for the delivery of information letters are discussed in Section 6. Finally, some conclusions are drawn (Section 7).

2. The role of the information letter

In Italy, the PPHC is conducted annually by Istat, integrating information from administrative sources with data from two sample surveys: an area survey (AS) and a list-based survey (LS). The AS, initially conceived to estimate the coverage errors in the population register, is currently being re-designed; the LS is carried out on a sample of households to collect data for variables that cannot be replaced (or that can only be partially replaced) by administrative registers.

In this paper, the focus is on the data collection process of the LS in the second cycle of the PPHC, which started in 2022 at the end of the first cycle (2018-2021).

Data are collected using a mixed-mode approach involving CAWI, CAPI (Computer Assisted Personal Interviewing), and CATI (Computer Assisted Telephone Interviewing). In a first phase of the survey, households can either fill in the questionnaires themselves via the web spontaneously, or go to the municipal collection centres (MCCs), or call the contact centre (CC). In a second phase, after the period of spontaneous response, the enumerators visit the non-respondent households and conduct interviews mainly in CAPI mode.

An information letter is sent in advance to all the sampled households to inform them about the purpose of the census, the modes of participation, the contacts for information and support in completing the questionnaire, the deadline, the obligation to respond, and the consequences in case of no participation in the survey. In particular, the information letter contains a unique identification code as well as precise instructions on how to access and return the online questionnaire.

In addition, during the survey two reminders are mailed to the non-responsive households.

Information letters and remainders are sent as standard bulk mail with notification service by a postal carrier. If the postman fails to deliver the letter, he is required to provide the reason for this: untraceable, deceased, moved, or unknown addressee; refused delivery; wrong, insufficient, or non-existent address.

3. Relationship between sent letters and survey outcomes

Although the delivery of information letters and remainders is a process outside the control of Istat and the data collected by the postmen may have reliability problems, it is worth noting that these data are very important to:

- monitor the progress of the delivery of information letters and reminders;
- provide information on the first contact with the households during the initial phase of the survey, when interviewers are not yet in the field.

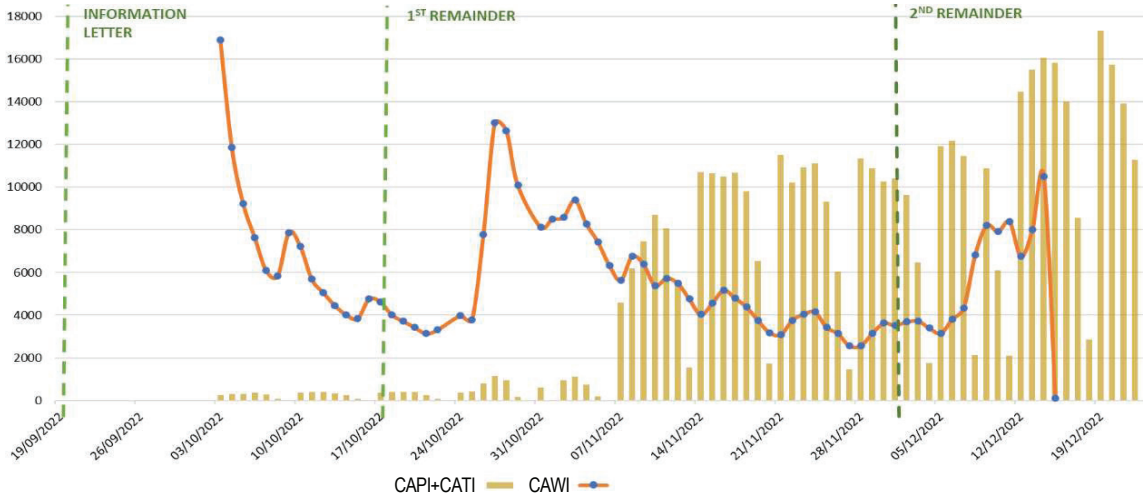
In 2022, the LS showed a certain decrease in the spontaneous CAWI response rate compared to 2021 (from 51.9% in 2021 to 48.2% in 2022). Since the spontaneous response depends on receiving information letters and reminders, it is important to monitor their progress and identify any critical points. After the survey is completed, the delivery information can be compared with the findings in terms of number of respondents by data collection mode.

The postal service makes data available on deliveries at the household level for the following variables: questionnaire code, delivery date, delivery outcome, postcode, municipality, province, and region. On the other hand, the Istat web-based management system (SGI) supporting the survey process provides information for counting households by response mode, out-of-scope households, and non-respondents.

For the 2022 LS, approximately one million information letters in Italian were sent, with a delivery rate of 94.4%. One week after the letters were sent, 93% of them were delivered. The main reason for the non-delivery is an unknown or untraceable addressee (4.5% of the total amount of letters sent).

Figure 1 shows the close relationship between the mailing of the information material and the number of CAWI responses. Timely sending the information letter and the first reminder proved to be crucial to ensure a good CAWI response in the first phase of the survey (from October 3 to November 7), during which only spontaneous responses were allowed.

Figure 1. Sending of letters and responses (CAPI+CATI and CAWI) by date



Source: Data from the postal service and SGI, 2022

Table 1 indicates that 41% of the households to whom neither the information letter nor the first

reminder could be delivered were out-of-scope for the survey, 26.8% were CAPI or CATI respondents, 23.5% were non-respondents, and only 3.8% responded via CAWI (some households called the CC or MCC to ask for their Internet credentials).

Table 1. Percentage of households by survey outcome and number of delivered letters

SURVEY OUTCOME	NUMBER OF DELIVERED LETTERS		
	0	1	2
CAWI response	3.8	17.5	39.5
CAPI or CATI response	26.8	39.6	48.9
Other type of contact	4.9	4.0	1.9
Out-of-scope	41.0	23.4	3.8
No contact	23.5	15.5	5.9
Total	100.0	100.0	100.0

Source: Data from the postal service and SGI, 2022
Note: Only the information letter and the first remainder are considered

4. The monitoring procedure

The monitoring procedure for the LS of the PPHC is mainly aimed at understanding for which postcodes the delivery of the information material is critical and what actions must be taken to contact the households in the problematic areas. The procedure provides three types of output:

1) indicators on deliveries, defined by exploiting all available information and calculated at the most detailed territorial level allowed by the data (postcode[1]);
2) ANOM charts, to display the indicators and highlight any anomalous values;
3) a tabular report listing only those postcodes for which anomalous values of the indicators are highlighted.

As concerns 1), two indicators are defined, one based on the data provided by the postal service and the other obtained by integrating the information provided by the postal service with that collected through the census:

a) *non-delivery rate*, *i.e.* the number of missed deliveries divided by the number of letters sent to the households;
b) *non-reception rate*, *i.e.* the number of delivered letters, declared as not received by the households, divided by the number of completed questionnaires.

Rate (*a*) is sufficient to monitor the delivery of the information material – information letters or remainders – at the territorial level in near real-time, whereas indicator (*b*) can only be used for analytical purposes at the end of the survey to better investigate the phenomenon of non-deliveries. Specifically, the non-reception rate makes it possible to verify the concordance between what is declared by the households and what is reported by the postal service.

With regard to 2), ANOM *p*-charts are used to display and analyse the above indicators[2]. The ANOM is a graphical and statistical method for simultaneously comparing k treatment means with their overall mean (Nelson *et al.*, 2005; Jayalath *et al.*, 2021). In a typical ANOM chart, the treatment means are plotted as deviations from the central line (*i.e.* the overall mean) and can be compared with upper and lower decision limits – respectively, UDL and LDL[3] – to identify which treatments are significantly different from the overall mean. In this context, the postcodes corresponding to indicator values falling above the UDL are considered flagged.

In addition to the above-mentioned indicators and charts, the monitoring procedure automatically produces a tabular report listing only the flagged postcodes, together with information

[1] There are approximately 4,600 postcodes in the country. Designed to make the postal service more economical and efficient, they have the advantage – compared to other territorial entities (such as municipalities) – of being more homogeneous in terms of population size and therefore more suitable for monitoring purposes.
[2] The ANOM charts are implemented using SAS/QC software (SAS Institute Inc., 2018)
[3] The UDL and LDL are determined by a specified significance level (α), *i.e.* the probability that, under the null hypothesis of no treatment difference, at least one of the response summary statistics exceeds the decision limits.

and statistics useful for understanding the reasons for the non-deliveries (*e.g.*, frequency of non-deliveries by type of delivery outcome) and deciding on the type of action to be taken (*e.g.*, alerting the postal service or contacting the MCCs for targeted field inspections and feedback on the households and/or addresses to which the non-deliveries relate).The absolute values of the variables used to build the indicators are also reported to take in due account those alerts based on many units.

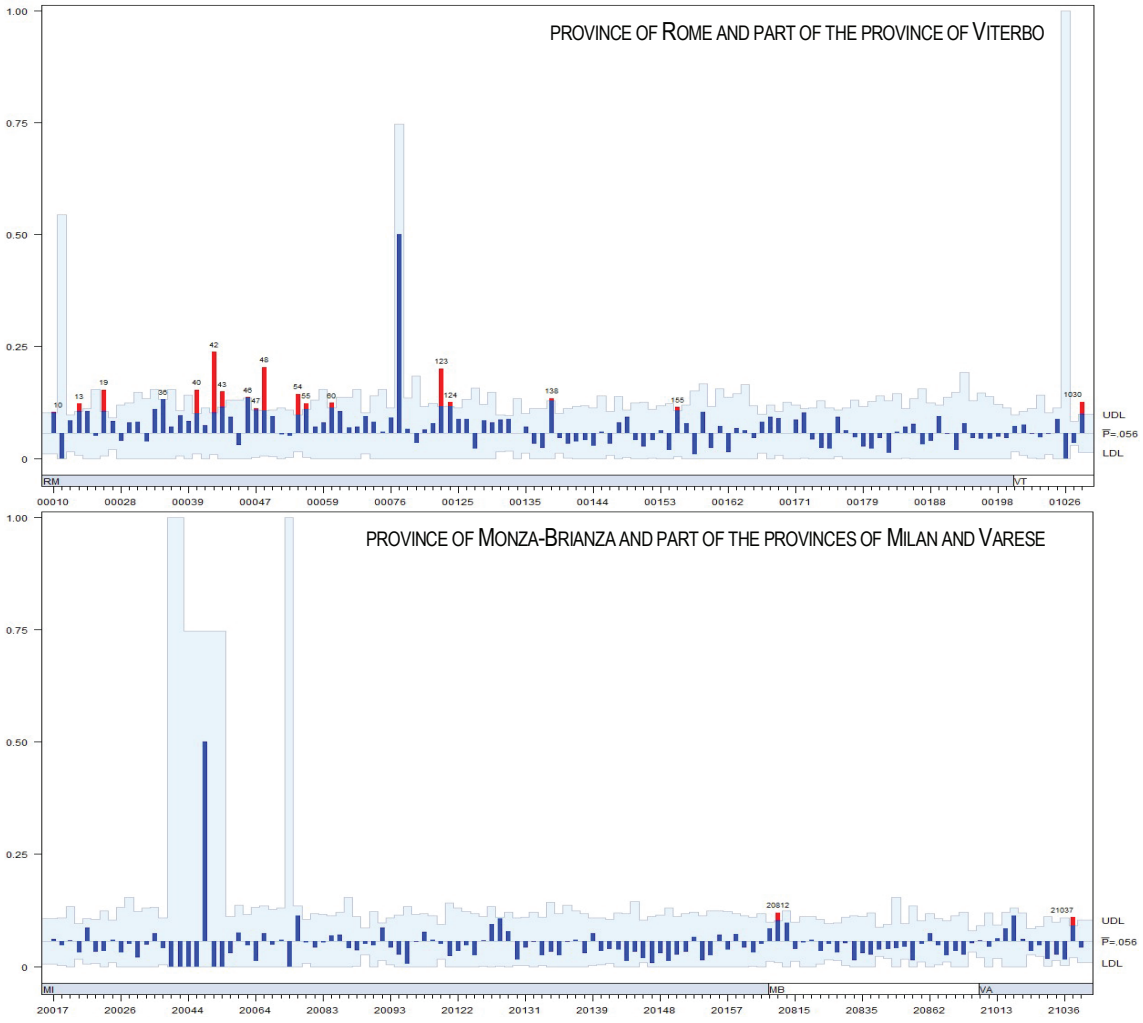
5. Some charts for the non-delivery rate

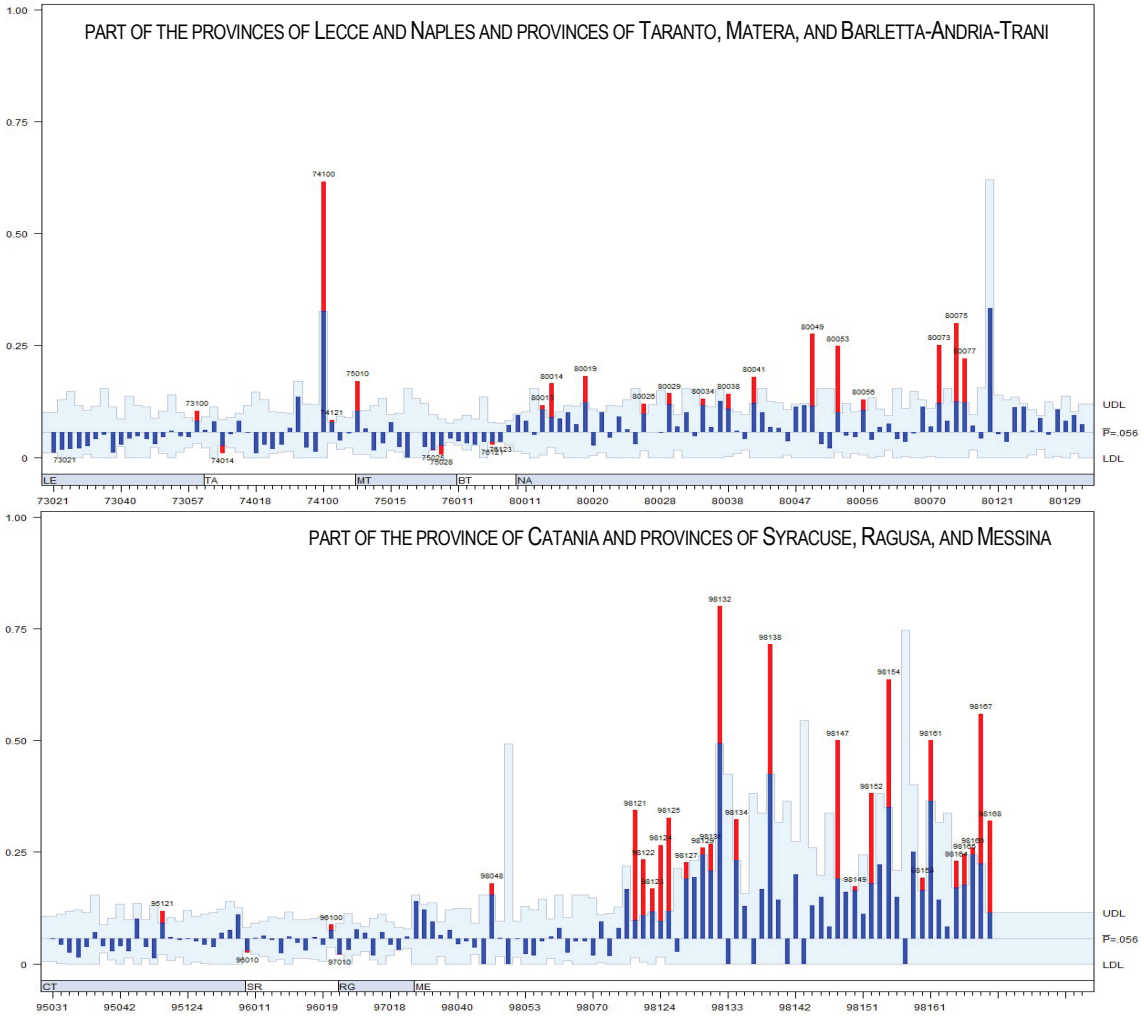
Some ANOM *p*-charts for indicator (*a*) are presented below (Figure 2), with reference to the information letters sent in Italian (in the following, the provinces of Aosta and Bolzano are excluded from the analysis). The overall mean is $\bar{P}$=0.56 and the UDL and LDL correspond to α=0.05.

The first panel of Figure 2 shows the *p*-chart for the province of Rome. Outliers can be observed especially for postcodes relating to certain suburbs of the city (*e.g.*, Tor Sapienza) or to peripheral municipalities of the province where the presence of foreign immigrants is significant.

The province of Milan does not record any anomalous values (second panel). This is also the case for many other areas in northern Italy. On the other hand, the third *p*-chart, in which there are several critical values in the province of Naples and only one particularly high outlier in the province of Taranto, can be considered representative of the situation in other parts of southern Italy.

Figure 2. ANOM *p*-charts by postcode





Source: Data from the postal service, 2022

Finally, a very critical situation emerges in the province of Messina, where particularly severe outliers can be observed for various postcodes (fourth panel). A thorough examination of the available information showed that most of the anomalies can be attributed to incorrect addresses.

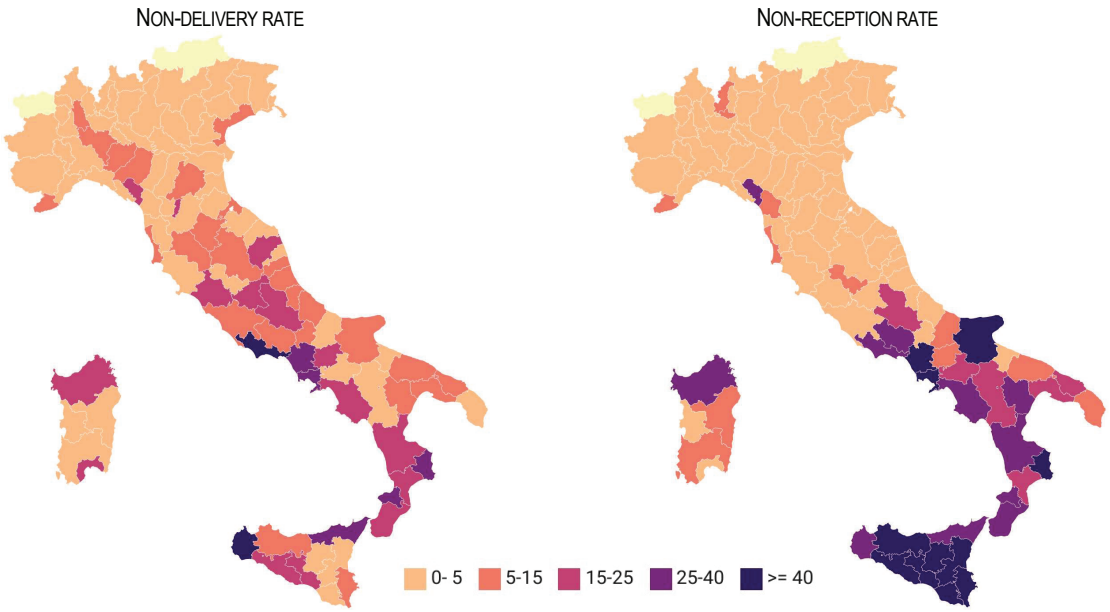
6. Summary results

At the national level, undelivered information letters amount to 5.6%, while those delivered, but declared as not received by the households, are 9.9%.

The results from the ANOM charts for both indicators (*a*) and (*b*) are given in a very summarised form, showing the percentage of flagged postcodes for each province, regardless of the alert values (Figure 3). With reference to (*a*), Latina and Trapani turn out to be the provinces with the highest percentages (above 40%). They are followed, in descending order of criticality, by some southern provinces (Caserta, Naples, Crotone, Vibo Valentia, and Messina) with flagged postcodes between 25% and 40%. Northern Italy presents the least problematic situation.

With regard to (*b*), the concordance between what is declared by the households and what is reported by the postal service is very low in Sicily and in several southern provinces – particularly in Caserta, Naples, Foggia, and Crotone – while almost all the northern provinces and part of the central ones do not show any alarming discrepancies. These results suggest that in those areas where the delivery is more problematic, the data reported by the postal service might be less accurate.

Figure 3. Provinces by the percentage of flagged postcodes for indicators (*a*) and (*b*)



Source: Data from the postal service, 2022 *Source*: Data from the postal service and the LS, 2022
Note: Provinces of Aosta and Bolzano are excluded from the analysis

7. Conclusions

The monitoring procedure for the LS of the PPHC is mainly aimed at understanding for which postcodes the delivery of the information material is critical and, in these cases, what actions must be taken to contact the sampled households. Specific indicators are defined and calculated at a very detailed territorial level (postcode) using data provided by the postal service or collected through the census. In addition to being directly related to the delivery process, the postcodes are more homogeneous than other areas in terms of population size and are therefore more suitable for monitoring purposes. The ANOM charts, which are used to display the proposed indicators, helps balance cost and thoroughness of monitoring activities by using statistical principles to differentiate potentially problematic cases from those that vary naturally around a process average. In this way, fieldwork supervisors and survey managers are guided in making targeted interventions on the critical postcodes only, without spending time exploring false alarms.

The monitoring procedure should be used as part of a close cooperation among postal service, survey managers, and MCCs to find the best solutions to reach the households to whom the non-deliveries relate. This approach improves the data collection process, providing useful elements for planning subsequent editions of the census or any other survey on households or individuals.

References

Jayalath, K.P., Turner, J. (2021). Analysis of Means (ANOM) concepts and computations. *Applications and Applied Mathematics: An International Journal* (AAM), **16**(1), pp. 75-96.

Jans, M., Sirkis, R., Morgan, D. (2013). Managing data quality indicators with paradata based statistical quality control tools: the keys to survey performance, in *Improving Surveys with Paradata: Analytic Uses of Process Information*, ed. F. Kreuter, John Wiley & Sons Inc., Hoboken, (NJ), pp. 191-229.

Nelson, P.R., Wludyka, P.S., Copeland, K.A.F. (2005). *The Analysis of Means: A Graphical Method for Comparing Means, Rates, and Proportions*. ASA-SIAM Series on Statistics and Applied Probability, SIAM, Philadelphia, ASA, Alexandria, (VA).

SAS Institute Inc. (2018). *SAS/QC® 15.1 User's Guide*. SAS Institute Inc., Cary, (NC).

doi.org/10.26398/asaproc.0067

The essential elicitation set for approximating a multi-attribute utility function in agronomy: A Gaussian process approach

Federico Giordano Susi[a], Lorenzo Valleggi[b], Alessandra Micheletti[c], Federico Mattia Stefanini[c]

[a] Department of Mathematics Federigo Enriques, University of Milan, Italy.
[b] Department of Statistics, Computer Science, and Applications "Giuseppe Parenti", University of Florence, Italy;
[c] Department of Environmental Science and Policy, Faculty of Science and Technology, University of Milan, Italy;

1. Introduction

Decision-making processes for crop protection may benefit of elicitation procedures both to exploit expert's degree of belief about future events and to define preference schemes on consequences of an intervention by means of a utility function. When many attributes are involved in the evaluation of consequences, the elicitation of utilities based on pairwise comparisons between alternatives may become very time-consuming. Moreover, the process of eliciting preferences can be susceptible to cognitive biases, leading to variations in how individuals express their preferences. Various factors, including the context in which preferences are assessed, the elicitation method employed, and individuals' inherent inconsistencies in decision-making, may affect elicited preferences (Lichtenstein et al., 2006).

Agronomists may face difficulties in accurately eliciting utility values in multi-attribute settings if the assumption of utility independence (French et al., 2000) is not tenable. For example, when an agronomist needs to decide whether or not to treat a vineyard for a phytopathogen, the multi-attribute utility value of the decision cannot be elicited by independently considering the utility of a particular level of disease incidence and the utility of the environmental impact of the treatment. It follows that the precise form of the utility function may become hard to elicit, in particular within specific regions of the joint sample space of attributes.

In this paper, we propose a Gaussian process-based method to address the mentioned elicitation problem, finding the minimal set of utility values needed to approximate a multi-attribute utility function for similar applications, namely the decision-making process regarding the treatment of vineyards against Plasmopara viticola, the pathogen responsible for downy mildew and a major concern for grapevine cultivation (Wong et al., 2001).

2. Method

In this section, we focus on the application of Gaussian processes (GPs) for fitting a multi-attribute utility function. We describe the GP model and detail the parameter estimation procedure for the correlation function.

2.1. Background on the multi-attribute utility function

The multi-attribute utility function elicited by Valleggi and Stefanini (2022) was used as "expert" to find the utility values for each combination of environmental summary and the probability of healthy leaves in the utility space. The rationale behind the approach is natural: if the utility of the environmental attribute is low due to the high impact of a phytosanitary treatment then a low value of multi-attribute utility will follow. The multi-attribute utility function here

playing the role of "expert" is

$$U(f_{s,b,t+1}, \tilde{\pi}_{i,j}) = k_1 U_1(f_{s,b,t+1}) + k_2 U_2(\tilde{\pi}_{i,j}) + k\ k_1\ k_2\ U_1(f_{s,b,t+1}) \cdot U_2(\tilde{\pi}_{i,j})$$

where s_{t+1} is a score that classifies the degree of cleanness of soil after chemical treatment (including derived side products), $\Omega_{s_{t+1}} \in \{1,2,\dots,5\}$, where $s_{t+1} = 1$ for the worst state after 10 years from treatment, and $s_{t+1} = 5$ for the cleanest case after 10 years; b_{t+1} is a biodiversity score to classify the degree of biological diversity, $\Omega_{b_{t+1}} \in \{1,2,\dots,5\}$, thus $b_{t+1} = 1$ refers to the worst state of biological diversity after 10 years from treatment and $b_{t+1} = 5$ is the best diversity class after 10 years from treatment; $f_{s,b,t+1} - ((s_{t+1} + b_{t+1})/2 - 1)/4$, with $\Omega_{s,b} - [0,1]$ is a value function averaging and rescaling biodiversity and soil scores was considered as an environmental summary of the future state; j, stands for each combination of the environmental scenario for pathogen growth, $\Omega_j \in \{Useful, Not\ Useful\}$, $a_i, i \in \{0,1\}$ for the decision; $\tilde{\pi}$ is the probability to have healthy leaves in the grapevine. The sub-utility functions U_1 and U_2 are Beta cumulative distribution functions (Valleggi and Stefanini, 2022, for further details).

2.2. Regression model-based on GPs

In this work, a non-parametric regression model based on a GP is fitted using the GPfit R package (MacDonald, et al., 2015). Let's consider a real-valued Gaussian stochastic process denoted as $y(\cdot)$ in the context of a p-dimensional input domain $\mathcal{X}$. This process follows a GP ($\mathcal{GP}$) model, characterized by a mean function $\mu(\cdot)$ and a covariance function $\sigma^2 c(\cdot,\cdot)$ which governs its behavior. The variance is represented by σ^2, and $c(\cdot,\cdot)$ is the correlation function. For any given input $x_i \in \mathcal{X}$ where $i = 1, \dots, n$, the corresponding outputs $(y(x_1), \dots, y(x_n))^{\mathrm{T}}$ collectively follow a multivariate normal distribution. In mathematical terms,

$$[(y(\mathrm{x}_1), \dots, y(\mathrm{x}_n))^{\mathrm{T}} | \boldsymbol{\mu}, \sigma^2, \mathbf{R}] \sim \mathcal{MN}((\mu(\mathrm{x}_1), \dots, \mu(\mathrm{x}_n))^{\mathrm{T}}, \sigma^2 \mathbf{R}). \quad (1)$$

Here, $\mathbf{R}$ represents the correlation matrix with each entry (i,j) reflecting $c(\mathrm{x}_i, \mathrm{x}_j)$, $\sigma^2\mathbf{R}$, is the variance, and $\boldsymbol{\mu}$ is defined as $(\mu(\mathrm{x}_1), \dots, \mu(\mathrm{x}_n))^{\mathrm{T}}$. In our study, we chose to use the Matern correlation (2), with roughness parameter $\alpha = \frac{5}{2}$, among all possible choices because it has been found to be highly reliable for both nearby and distant spatial input, primarily due to its asymptotic behavior,

$$c_l(d_l) = \left(1 + \frac{\sqrt{5} d_l}{\gamma_l} + \frac{5 d_l^2}{3\gamma_l^2}\right) \exp\left(-\frac{\sqrt{5} d_l}{\gamma_l}\right) \quad (2)$$

where d_l stands for any of the $|x_{il} - x_{jl}|$ and the parameter γ_l, also known as the range parameter, controls the relationship between input based on their distance.

One of the most used parameter estimation methods (MacDonald et al., 2015) is focused on efficiency, as it improves upon the gradient algorithm by an initialization step in which a random sample of points is drawn using a uniform distribution on the sample space. To explore the parameter space (Θ), the authors proposed a parametrization of the kernel to consider a limited subspace ($\Theta_0 \subset \Theta$), in order to initialize the procedure through a homogeneous set of parameters scattered in space. Then, using a multi-start L-BFGS-B algorithm starting from an initial set of parameters, the minimum negative log-likelihood is found for each parameter, and only a subset is selected based on the performance of the gradient algorithm. Then a k-means clustering is applied on the subset of parameters, and 2 clusters are found. In order to find the global minimum, the initialization procedure is repeated for each cluster in order to find only two minima and then compare them. The corresponding γ associated with this minimum represents the estimated parameter value for the regression model.

2.3. Iterative utility values elimination method

The purpose of the method was to identify a minimal set of utility values for model training. Each utility value for the training set was obtained using the multi-attribute utility function described in Section 2.1 that corresponds to a question during the elicitation phase. Among them, 49 points are evenly spaced on a grid, while the remaining 11 points are chosen near the known utility values of (0, 0) and (1, 1), defined as 0 and 1, respectively. The method is based on sequentially removing these utility values from the training set and then retraining the GP with a reduced training set. The goal is to assess the GP's ability to predict the removed value, the removing stops when the prediction error becomes too high. The method is described in detail below.

The method begins by initializing an error threshold $\varepsilon > 0$ at the start. It then randomly selects a point $\tilde{\mathbf{x}} \subset \mathbf{x}^{\mathcal{D}}$. The selected point $\tilde{\mathbf{x}}$ is then removed from $\mathbf{x}^{\mathcal{D}}$, and the model is retrained using the modified training set. Thus the predictive error is evaluated using a mean squared error (MSE) metric. Next, the method compares the calculated predictive error against the predetermined threshold value ε. If the predictive error is below the threshold, it confirms the elimination of $\tilde{\mathbf{x}}$ from the training set. This means that the model achieves sufficient accuracy without the selected point. On the other hand, if the predictive error exceeds the threshold, point $\tilde{\mathbf{x}}$ is reintroduced back into the training set. By doing so, the method acknowledges that retaining this point improves the model's accuracy. The error MSE0 associated with the model fitted using the entire initial training set of 60 utility values (Figure 1, right panel) is measured to be 1.9×10^{-7}. To further refine the model and reduce computational complexity, we decided to employ an error threshold (ε) set to two orders of magnitude higher than the initial error, resulting in $\varepsilon = \text{MSE}^0 \times 10^2$. After applying the method, we obtained a refined configuration of points in the final training set.

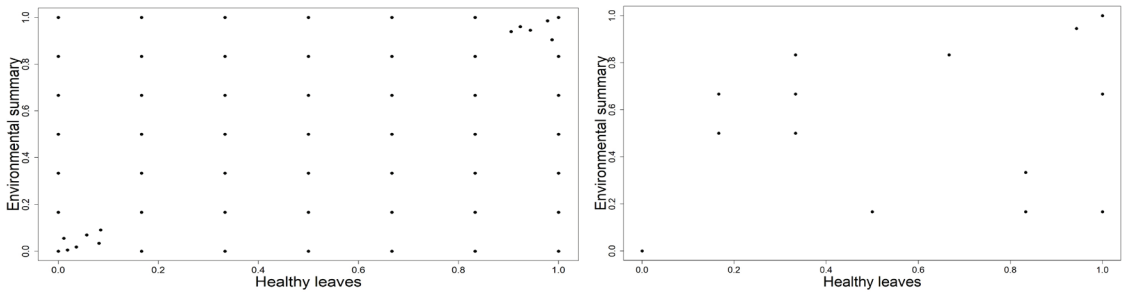


Figure 1: **Initial 60 utility values as training set (left) and final configuration of utility values obtained after applying the elimination method with 14 points (left).**

In order to assess the performance of the GP method in replicating the utility surface after eliminating utility values, iso-utility curves were generated. These curves allow for a comprehensive comparison between the original target function and the fitted model, thereby evaluating the capability of the GP method.

3. Results

The elimination method was run until the MSE became too high, starting from the *60* initial utility values, it was found that only *14* utility values were needed (Figure 1) to obtain a good approximation of the real utility function. Moreover, the iso-utility curves reported in the contour plot and the 3D surface of the real utility function (Figure 2, right panel) compared with the approximation obtained with GP (Figure 2, left panel) confirm that result. Indeed, iso-utility curves have a similar trend, also the GP approximation respects the basic assumption of having low global utility when the utility of the environmental attribute is low regardless of the utility of

the disease attribute.

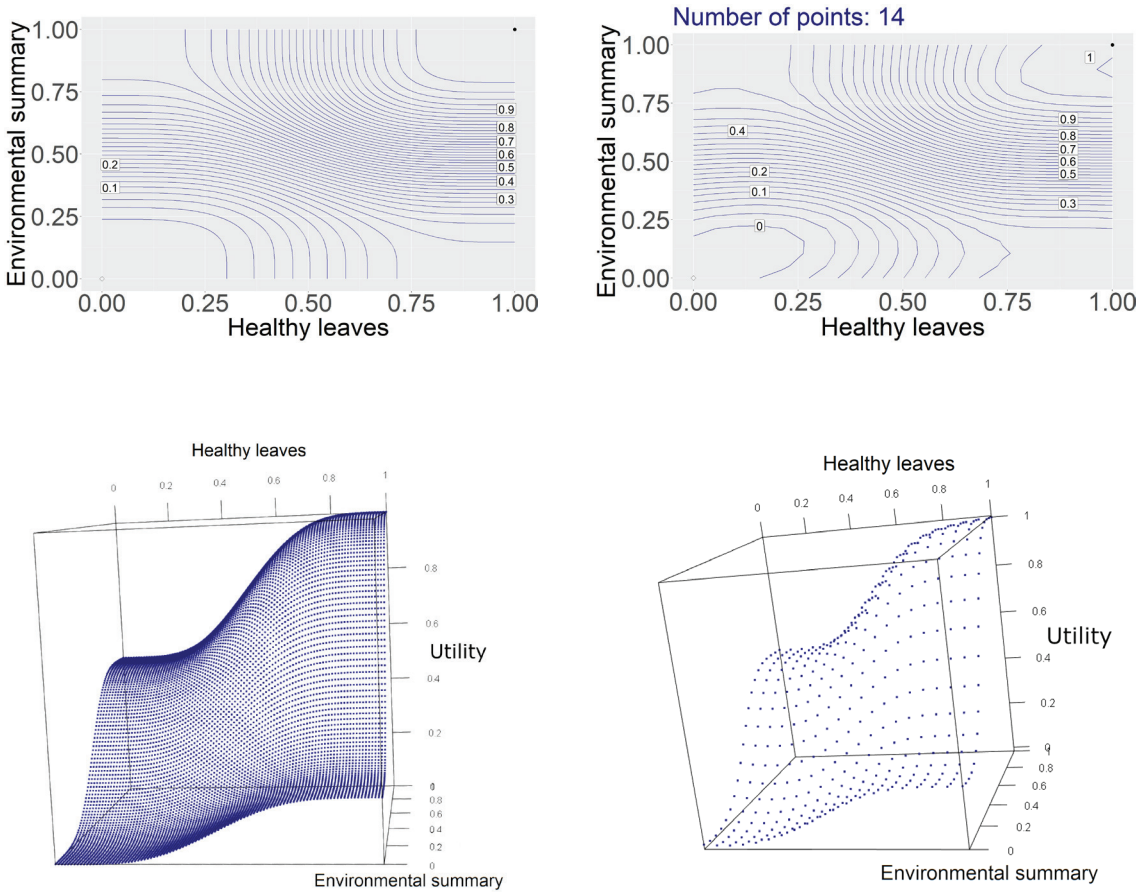


Figure 2: **Comparison panel of the utility function (1) and the fitting results** (The top-left plot displays the contour lines of the original utility function, while the top-right plot showcases the contour lines of the fitted model. The bottom-left plot presents the 3D surface representation of the utility function, and the bottom-right plot depicts the 3D surface representation of the fitted model)

4. Discussion and conclusion

The Gaussian process (GP) is widely recognized as a powerful method for solving interpolation problems, thanks to the flexibility of this class of model. Its remarkable flexibility enables it to effectively capture the functional forms of complex functions. This characteristic makes it particularly well-suited for decision theory, especially in the context of preference elicitation problems (Guo et al., 2010), where the challenge of handling the inconsistency of elicited preferences has to be addressed. By selecting the appropriate correlation function, the GP ensures the differentiability of the estimated utility function, while also delivering low predictive errors compared to other parametric approaches (Gu et al., 2020).

In this work, a GP was defined to approximate a multi-attribute utility function (MAUF) in a problem domain made by the decision to treat a vineyard against *Plasmopara viticola*. Two attributes were considered, the probability of healthy leaves after treatment and the environmental impact due to the treatment. The goal was to find the minimum set of utility

values to be elicited by the expert to still have a good approximation of the MAUF. Despite that GPs appear to be a promising approach to face the above challenges, the quality of results using GPs for this purpose may depend on the choice of estimation method. The initial step of this study was focused towards the determination of the most suitable estimation method. A comparison of the performances was conducted among a maximum likelihood estimator, the estimation of the posterior distribution of the parameters, and the gradient-based algorithm (Gu et al., 2018). To assess the effectiveness of these methods the Mean Squared Error (MSE) was used as the evaluation metric for this application. The simulation results have shown that the gradient algorithm-based method outperforms the others in terms of a lower MSE (data not shown).

Notwithstanding the general robustness of both methods, one using the maximum likelihood and the other the posterior distribution, with respect to the conformation of the utility surface, empirical findings indicate that for this particular application the gradient-based algorithm yields superior results. Nevertheless, it is essential to ensure that the tuning parameter is appropriately set to achieve optimal performance. The GP method demonstrates strong capability in accurately reproducing the utility surface of the MAUF, while maintaining the fundamental assumption of the decision process. However, it is important to note that this approach's limitation lies in its ability to approximate the MAUF specifically as done in this study.

Future research could explore whether the same GP specification can effectively approximate MAUFs with varying degrees of expert belief, provided they share the same domain. This investigation would shed light on the generalizability and robustness of the GP approach across different expert beliefs within the same domain.

References

French, S., Rios Insua, D. (2000). *Statistical Decision Theory*. Kendall's Library of Statistics 9. Oxford University Press Inc., New York.

Gu, M., Wang X., Berger J. (2018). Robust Gaussian stochastic process emulation. *The Annals of Statistics*, **46.6A**, pp. 3038 - 3066.

Gu, M., Bhattacharjya, D., Subramanian, D. (2020). GaSPing for utility. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**.

Guo, S., Sanner, S., Bonilla, E.V. (2010) Gaussian process preference elicitation. In *Advances in Neural Information Processing Systems*, **23** (Eds. J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel and A. Culotta), https://papers.nips.cc/paper_files/paper/2010.

Keeney, R.L., Raiffa, H. (1993). *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*. Cambridge University Press.

Lichtenstein, S., Slovic, P. (2006). *The Construction of Preference*. Cambridge University Press.

MacDonald, B., Ranjan, P., Chipman, H., (2020). GPfit: An R package for fitting a Gaussian process model to deterministic simulator outputs. *Journal of Statistical Software*, **64**:1548-7660.

Valleggi, L., Stefanini, F. M. (2022). On the utility of treating a vineyard against Plasmopara viticola: A Bayesian analysis. *ASA Conference, Data-Driven Decision Making*, Genoa University Press – Firenze University Press (Eds. E. Di Bella, L. Fabbris., C. Lagazio), pp. 233-237.

Wong, F.P., Burr, H.N., Wilcox, W.F. (2001). Heterothallism in Plasmopara viticola. *Plant Pathology*, **50**, pp. 427–432

doi.org/10.26398/asaproc.0068

The actual effect of the Sars-Cov-2 vaccine in Italy and the Istituto Superiore di Sanità's erroneous evaluation of it

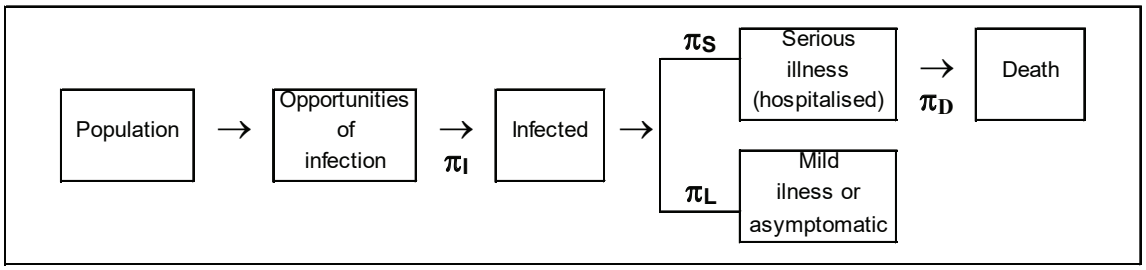
Massimo Tozzi
Statistical researcher, freelance

1. Introduction

This paper aims to discuss the Covid-19 vaccines actual effects to contrast the infection consequences[1]. In particular, we highlight that the estimates released by the Istituto Superiore di Sanità (ISS), the Italian national health institute, are affected by a statistical error in the calculation of the relative risk of a vaccinated person compared to a non-vaccinated one.

According to statistical theory (Armitage, 1971, pp. 414-420), the risk of incurring a certain event is given by the ratio of the number of observed events to the number of risk-exposed subjects and, in case of infectious disease, we have three different connected events: infection, serious illness with hospitalisation, death (see also Figure 1). After all, the risk of incurring a certain event is nothing else than its probability and the basic way to calculate a probability is exactly the above mentioned ratio.

Figure 1: The epidemic chain



In Figure 1, π represents the probability, and then the risk, for a subject belonging to that population to pass from a previous status to the following one. To quantify π:

- π_I (probability of infection) = infected people / population with the possibility to be infected;
- π_S (probability of serious illness and hospitalisation) = hospitalised people / infected people;
- π_L (probability of mild illness or to be asymptomatic) = 1 - π_S
- π_D (probability of death) = dead people / hospitalised people.

Unfortunately, in this Covid-19 epidemic it is very hard to determine exactly the number of people really risk-exposed to be infected, so we can only approximate the probability with:

- π_I = infected people / entire population,

although it could be more correct to approximate the infection risk with the ratio of the confirmed infections to the number of subjects who took the diagnostic test (although various problems persist even in this case) but these data are not always available.

[1] The paper is an abstract from an organic study, yet to be published, concerning all the aspects of the Sars-Cov-2 epidemic in Italy in the period between late February 2020 and late August 2022. The study was based on the data released by various public authorities (Ministry of Health, Civil Protection; Istituto Superiore di Sanità).

Moreover, to compare the risk within different categories of people, in our case vaccinated or not-vaccinated people, we can build the relative risk simply relating the risk of a category to the risk of the other one.

The ISS used another indicator to evaluate the risk, say, the incidence rates per 100,000 inhabitants of infected people, hospitalised people and dead people to the whole population, thus causing an inevitable distortion of the estimates, in particular concerning the risk to be hospitalised and the risk to die. Using the incidence rates that consider the whole population equally exposed to risk of dramatic events in the same way, implies, unreasonably, that a *not infected* subject could be seriously sick or even die for the topical disease.

If the risks are incorrect, the relative risk that, comparing non-vaccinated with vaccinated people, identifies the vaccine's effect is incorrect as well.

2. The consequences of a wrong choice

In what follows, we argue that the analysis conducted by the ISS on vaccines effects did not concern the drug's actual contribution in contrasting the epidemic's consequences in Italy but was confined to describing its accounting aspect, and these two phenomena, despite looking similar, are not the same.

To better understand all the consequences of considering the incidence rate as a risk indicator for Covid-19 related events instead of the correct one, we present a numeric example using some ISS data.

Since early July 2021, the weekly report of the ISS identifies the total of the infected subject in the previous 31 days within the population 12-year old and above, distinguishing them according to vaccination status (whether and when they received it) and age group. Following these subjects in time, the ISS could trace their course and the consequent end of the infection, clearly identifying how many of the infected in that period were seriously sick and hospitalised in a Covid ward, how many in a ICUs (intensive care unit) and how many died. A kind of basic survival analysis.

Moreover, it needs to be pointed out that, for the ISS weekly data, the reference periods for a SARS-CoV-2 infection's diagnosis, for hospitalisation in Covid wards and ICU and for deaths are different between each other due an average delay of two weeks for hospitalisation in Covid wards and ICU and three weeks for deaths (see ISS Extended Reports – Methodological note, Section 4.3).

For example, let use the ISS November 2021 data (precisely, the October 29 - November 28, 2021 period[2]) focusing, for the sake of brevity, on only two categories: non-vaccinated subjects and subjects who completed the full vaccination course for less than 150 days, both within the older age group (80 and more year old). The basic data are reported in Table 1.

With these data we can calculate the risk for each event within the non-vaccinated group and vaccinated group and the relative risk given by their ratio, using both the "correct method" and the ISS incidence rates in order to compare them (Table 2).

Starting from the first event, the infection, we must remember that we don't have the current number of risk-exposed people but only the population size, so it is necessary to settle for the accounting indication released by the ISS incidence rates, causing the results to be definitely questionable[3].

[2] The examined data were released in the following ISS reports: 1) ISS Extended Report released on December 3 2021, which supplied the data on the period's average population and infected subjects; 2) ISS Extended Report released on December 17 2021, which supplied the data on the subjects hospitalised in Covid wards and ICU; and 3) ISS Extended Report released on December 24 2021, which supplied the data on deaths.

[3] Indeed, it is plausible to believe that a vaccinated subject could somehow feel safer than a non-vaccinated one and thus be less likely to take a test, narrowing the number of tested cases among the vaccinated subjects. To make it more extreme, there could even be a scenario of a higher incidence among the non-vaccinated subjects contemporarily to a number of tested cases among the vaccinated ones so much lower to the non-vaccinated that

Table 1: Population aged 80 and above, newly infected and their outcome in the October 29 - November 28 2021 period by vaccination status

	Population aged 80 and above	Newly infected	Hospitalised in Covid wards	Hospitalised in ICU	Deceased	Total people Hospitalised (1)
Vaccination Status						
non vaccinated	221,018	1,927	847	48	399	1,294
full vax.cycle < 120 dd.	420,620	853	219	13	82	314

(1) The total number of hospitalised subjects includes deceased subject as well, because they were hospisalised before their death

The correct risk of hospitalisation for an infected subject is the ratio between the total number of hospitalised people and the total number of infected ones:
1,294 / 1,927 = 67.1% for non-vaccinated subjects
314 / 853 = 36.8% for vaccinated subjects
and their ratio is equal to 1.82. So, the risk for non-vaccinated people of serious illness with hospitalisation is 1.82 times higher than vaccinated people (the estimates for the risk within the hospitalised in Covid wards and the hospitalised in ICU are obtained proportionally).

To estimate the risk to die for hospitalised people we need to relate the number of deceased subjects to the total number of hospitalised people obtaining:
399 / 1,294 = 30.8% for non-vaccinated subjects
82 / 314 = 26.1% for vaccinated subjects
and their ratio is equal to 1.18. So, the risk for hospitalised non-vaccinated people of death is 1.18 times higher than hospitalised vaccinated people.

Using instead the ISS incidence rates per 100,000 inhabitants with the same data we obtain all the risks easily relating each number of observed events to the number of population in both groups and the results are shown in Table 2. From Table 2 we can draw that:

- even using a statistically correct analysis of the basic ISS data, the vaccine played a key role in contrasting the various consequences of the epidemic;
- the ISS estimates present a considerable overestimation of such role, assigning to it the greatest effect in contrasting the patients' death, while the opposite actually occurred.

This overestimation is the first remarkable consequence of the ISS methodology, but from this another consequence arose, maybe worse: incorrect media information, which altered the population's perception of what was happening, providing them with a feeling of greater protection than the actual one. Just to make an example, because of the ISS overestimation of the vaccine effect, starting from November 2021, all major Italian media channels provided the information that strictly confined the epidemic within the non-vaccinated population, and this was not true. And it is likely that many vaccinated people with this incorrect information in their mind started less safe behaviours and were exposed to a higher risk of becoming infected, seriously sick or even die.

Indeed, relying on the ISS official data, in the period between December 2021 and August 2022 there was a significantly greater increase in all events (infection, hospitalisation and death) among the vaccinated people compared to the increase among the non-vaccinated ones and in those nine months 85% of the newly infected, 73% of the hospitalised and 70% of the registered

a probability of infection opposite to what the incidence states, that is to say higher among the vaccinated subjects rather than the non-vaccinated ones, would result. A very unlikely but yet possible situation, clarifying once more the difference between the two indicators. On the other hand, such a high value of an incidence 4.3 times as much for the non-vaccinated subjects compared to the vaccinated ones, as shown in Table 2, suggests that the vaccine did actually provide a good protection against the possibility of infection. However, with these data, we are not able to quantify it exactly.

deaths was all vaccinated people. The interested reader can compare the ISS extended reports from the release on 5 January, 2022 to that on 21 September, 2022.

Table 2: Risks to be infected, hospitalised or died for non-vaccinated and vaccinated people within the population aged 80 and above in the October 29 - November 28 2021 period and their relative risks. Comparison between the suggested method and the ISS rates

	Newly infected (1)	Hospitalised in Covid wards	Hospitalised in ICU	Deceased	Total people Hospitalised
	Risks according with the correct statistical method				
non vaccinated	871.9	63.5	3.6	30.8	67.2
full vax.cycle < 120 dd.	202.8	34.7	2.1	26.1	36.8
relative risk (2)	4.3	1.8	1.7	1.2	1.8
	Risks according with the ISS incidence rates x 100,000 inh.				
non vaccinated	871.9	383.2	21.7	180.5	585.5
full vax.cycle < 120 dd.	202.8	52.1	3.1	19.5	74.7
relative risk (2)	4.3	7.4	7.0	9.3	7.8

(1) Risks are the same in both methods because it's impossible have the actual number of risk exposed people but only the number of entire population
(2) Risk of non vaccinated / risk of vaccinated

Using the data weekly released by the ISS and applying the correct statistical method, it was possible to reconstruct a brief story of what actually happened between July 2021 and August 2022.

The first important observation is the highest population's adherence to the vaccination campaign. In late July 2021, seven months after the start of vaccination campaign, about 60% of the Italian population aged 12 and above had received at least one vaccine dose (24% was waiting for the second while 36% had already completed the full cycle).

In August and September more than 10.5 million new administrations were registered, thus bringing, in late September 2021, the amount of the Italian population aged 12 and above which had received the vaccine to 80% (about 72% had completed the full cycle while 8% was waiting for a second dose). Such values rose even higher within the population aged 80 and above, with 92% of them having completed the full cycle and further 2% waiting to complete it, the result of the priority that was given from the beginning to elderly and frail subjects. On the other hand, the participation in the vaccination campaign of the population aged 12-39 was lower but, nevertheless, a consistent one, with 55% of subjects having completed the full cycle and 16% waiting to do so.

Basically, at the end of summer 2021 the threshold generally considered as the inferior limit for the so-called "herd immunity" had been largely overcome and the vaccine's protection in this first period resulted definitely high compared to the possibility of infection, with a risk 80% lower for subjects with a full cycle of vaccination compared to non-vaccinated ones (although the comparison was based on incidence rates) and not very dissimilar between age groups. [4]

In this first period the protection against the probability of getting seriously ill and being hospitalised was still rather high, with subjects who completed a full cycle of vaccination

[4] To highlight the advantages of vaccinated subjects instead of the disadvantages of non-vaccinated ones, in what follows, the relative risk is constructed as the ratio of the risk of vaccinated subjects to the risk of non-vaccinated ones, completely equivalent to the reverse ratio employed in the earlier November 2021 example.

presenting a risk 60% lower than non-vaccinated ones, even though in this case the vaccine's effect resulted higher for the younger population compared to the one aged 80 and above.

Lastly, the vaccine's protection against the probability of death resulted to be more contained, with subjects who completed a full cycle of vaccination presenting a risk only 15-20% lower than non-vaccinated ones, with a behaviour not univocal among the different age groups on top of it.

Already in mid-October 2021 it was clear that the protection given by the vaccine was diminishing with time and with the distancing from the moment of its administration. For this reason, the ISS itself introduced a temporal watershed in order to separate the subjects who completed the full vaccination cycle, a threshold firstly set at six months which was reduced to five in the following November and to four in December 2021.

Essentially, the vaccine's effects started to vanish four to six months after the cycle had been completed: The data recorded in the months of November and December 2021 fully confirm this phenomenon under every aspect, with more recently vaccinated subjects presenting a much lower risk compared to subjects whose administration of the vaccine was more dated.

In the following months the tendency of the vaccine's protection to decrease as the administration date grows apart progressively continued to the point that, for the subjects who completed only the first full vaccination cycle, the vaccine's protection in late August 2022 was almost none concerning the probability of infection and death, while it was still at around 20-30% concerning the risk of getting seriously ill.

Going back to late November 2021, it was decided to start a new vaccination campaign for an additional third dose, which had not been previously foreseen, to contrast this now too obvious phenomenon. The administration of this further dose, which was indeed spread faster than the first two doses,[5] clearly curbed this "de-vaccination" process, thus bringing the vaccine's protection back to its summer 2021 levels. Moreover, the third dose demonstrated a substantial difference to the first two, registering a higher level of efficiency within the elder age groups.

However, the effect of the third dose started to diminish as the administration date grew apart as well, to the point that already in early summer 2022 the vaccine's protection against infection was almost equal to zero even in the subjects with all three vaccine doses.

On the other hand, the third dose's protection against the probability of getting seriously ill reported a less considerable decrease compared to the first two doses, settling, still in summer 2022, on risk values for the vaccinated subjects equal to about half of the risk values for non-vaccinated ones, while in the same period the protection against the probability of death for hospitalised subjects was not above 25%.

3. Conclusion

In conclusion, a crucial question arises: How is it possible that an established organisation like the ISS made such a trivial statistical error? What need was there to inflate the outcomes, positive nevertheless, of a drug necessarily completed in a hurry to contrast an illness previously unknown but capable of provoking the most serious global health crisis of the last one-hundred years?

Of course, we do not know the answers. Statistical theory provides many indicators to describe and analyse phenomena and each of them provides specific information about the phenomenon itself. The incidence rate gives an overall evaluation regarding the evolution of the observed phenomenon as it is a more refined way of counting the events that gradually occur (Colton, 1974). Therefore, employing the incidence rate to analyse the epidemic's trend represented a precise direction in the overall evaluation of the epidemic itself, a direction which

[5] Already by late March 2022 70% of the Italian population aged 12 and above had received the vaccine's third dose, 86% in the age group above 80 years old.

had as primary objective to track the evolution in time of the sustainability of the overall impact which the epidemic was having on the population's social and economic life alone, only taking into account its accounting aspect, a substantial “damage toll” of the epidemic.

And by this time this hypothesis is much more preferable than believing it possible that was only a trivial statistical error, moreover weekly replayed so long.

References

Armitage, P. (1971). *Statistical Methods in Medical Research.* Blackwell Scientific Publications.

Colton, T. (1974). *Statistics in Medicine*. Little, Brown and Company, Boston.

doi.org/10.26398/asaproc.0069

New public management: Performance analyses of Italian justice court

Rosanna Verde[a], Maria Gabriella Grassia[b], Matteo Mazziotta[c], Rocco Mazza[d], Simone Paesano[b]
[a] Luigi Vanvitelli University of Campania
[b] Federico II University of Naples
[c] ISTAT
[d] Aldo Moro University of Bari

1. Introduction

The new public management (NPM) is an approach to managing public service organizations used in government agencies, both at the local and national levels. In essence, the NPM can be defined as a set of principles and practices aimed at improving the efficiency and effectiveness of public organizations by adopting techniques from the private sector (Gomes & Mendes, 2022). The NPM objectives, as outlined in Pollitt's taxonomy (1997), encompass eight interconnected components: cost reduction, budget limitations, transparency in resource allocation, disaggregation of traditional bureaucratic organizations into separate agencies, decentralization of management authority within public agencies, separation of service provision from procurement, introduction of market-like mechanisms, performance management with goals, indicators, and results, and a shift in public employment from tenure-based, nationally standardized pay and conditions to group contracts, performance-related pay, and local determination of pay and working conditions. Additionally, the NPM emphasizes service quality, setting standards, and customer responsiveness. The NPM also finds specific application in the judicial domain, reflecting the general principles previously defined. Indeed, its implementation aims to enhance public sector management by focusing on performance, bureaucratization, results-based accountability, and efficiency, employing a managerial approach that emphasizes techniques from the private sector. In the judicial system, the application of NPM has led to the adoption of new management practices and processes, such as performance measurement, benchmarking, and customer orientation.

Another aspect to consider is the need to build trust in the legal system. As a crucial institution in any democratic society, the legitimacy of the court system depends on the public trust. Therefore, the application of NPM must focus on increasing public confidence in the court system. This requires particular attention to transparency, accountability, and responsiveness to the public's needs (Langbroek, 2011). The implementation of this approach has resulted in significant improvements in the efficiency and effectiveness of the legal system. For instance, the digitalization process has led to faster and more efficient case processing (Zhou, 2022).

Another improvement stemming from the application of NPM in the legal system is the use of performance measurement and benchmarking. These practices have contributed to enhancing the quality of judicial services by measuring performance and comparing it with benchmarks, using predetermined criteria that are relevant, reliable, and tailored to the unique challenges and complexities of the judicial system (Volacu, 2018). Through this approach, it is possible to identify areas for improvement and implement changes to optimize the quality of judicial services, thus better meeting the needs of the public (Medeiros, 2015).

As NPM is increasingly emerging as a management reality, including in the legal field, this paper aims to apply an existent method to formalise a new composite indicator for measuring the performance and efficiency of Italian judicial offices.

2. Methodology

This paper aims to propose a new way of synthetically evaluating the performance of the procedural processes in Italian courts, considering both dynamic and static components, which are commonly referred to in literature as flow trends and efficiency (Vecchi, 2018). Specifically, the focus is on the courts of Santa Maria Capua Vetere and Matera, this choice is due to the desire to consider two realities with a high differential between input workloads, so that flows of "comparable" nature can be placed under analysis. To achieve these goals, we chose to apply the adjusted Mazziotta Pareto index (AMPI), a non-compensatory composite indicator, to elementary performance indicators collected through a key performance indicator (KPI) approach. The AMPI allows the synthesis of multiple indicators into a single indicator for the civil sections of the mentioned courts, covering the period 2013 - 2022 for the Santa Maria Capua Vetere court, and 2020-2022 for the Matera court. The data were collected by the Office of the Process (UPP), which plays a central role as unit dedicated to data collection and KPI development. All the analyses were conducted using the R programming language.

As the name of the synthetic index suggests, the AMPI is an adjustment of a previously formulated method. It involves a standardization process for individual indicators, making them independent of variability by normalizing them with a mean of 100 and a standard deviation of 10 at the reference time *t* (Mazziotta and Pareto, 2016), the values thus obtained will be within, approximately, in the range (70; 130) (Mazziotta and Pareto, 2011). Moreover, all indicators are given equal weights, allowing only relative temporal comparisons. AMPI allows absolute comparisons over time, achieved through a different type of normalization. However, this comes at the expense of slightly reduced robustness, as the type of normalization used does not equalize variances as in non-adjusted MPI but makes them like each other.

Regarding the type of normalization used for AMPI, as mentioned earlier, it involves scaling individual indicators based on two "goalposts," which represent the minimum and maximum values within the possible range of each variable for all periods and time units.

To efficiently calculate the goalposts for this type of application, a reference value is used (e.g., the average value of a specific year). Let's define the following terms:

- Inf_{xj}= the lower value of indicator j for the entire period considered;
- Sup_{xj} = the upper value of indicator j for the entire period considered;
- Rif_{xj} = the reference value of indicator j (e.g., the average of a specific year).

the goalpost are:

$$Min_{xj} = Rif_{xj} - \Delta_{xj}$$
$$Max_{xj} = Rif_{xj} + \Delta_{xj}$$

where:

$$\Delta_{1xj} = Sup_{xj} - Rif_{xj}$$
$$\Delta_{2xj} = Rif_{xj} - Inf_{xj}$$
$$\Delta_{xj} = (\Delta_{1xj} + \Delta_{2xj})/2$$

In this case, the values of r_{ij} will be approximately within the interval 70-130[1]. Once the goalposts are calculated, the application proceeds as follows:
Given a matrix $X = \{x_{ij}\}$ with n rows (units) and m columns (indicators), the normalized matrix $R = \{r_{ij}\}$ is computed using the following formula:

$$r_{ij} = \frac{(x_{ij} - Min_{xj})}{(Max_{xj} - Min_{xj})} 60 + 70$$

In this formula, x_{ij} represents the value of indicator j for unit i, while Min_{x_j} and Max_{x_j} are the reference endpoints for indicator j. If indicator j has a negative polarity, the formula is

[1] According to the Bienaymé-Cebycev theorem, the terms of the distribution inside the interval (70; 130) constitute at least 89 percent of the total terms of the distribution (Mazziotta & Pareto, 2011).

complemented with respect to 200. In both cases, the range of the matrix $R = \{r_{ij}\}$ is computed with the formula: (130 - 70).

Consequently, the range of normalized values is set to (70, 130).

$$MPI_i^{+/-} = M_{ri} \pm S_{ri}CV_i \qquad \textit{where:} \qquad CV_i = M_{ri}/S_{ri}$$

The AMPI is then calculated using the generalized form, denoted as $MPI_i^{+/-}$: Where M_{r_i} and S_{r_i} represent the mean and standard deviation of the normalized values for unit i, and cv_i is the coefficient of variation for unit i. The sign ± depends on the type of measured phenomenon. For an increasing or positive composite index, indicating positive variations of the phenomenon (e.g., well-being), $MPI_i^{-} is\ used$. Conversely, for a decreasing or negative composite index, indicating negative variations of the phenomenon (e.g., poverty), $MPI_i^{+}$is used. This approach penalizes units with imbalanced values of normalized indicators by using the product (S_{r_i} cv_i) to favor units with a better balance among different indicators (penalization effect), while considering the average effect (M_{r_i}). AMPI shows the position of each unit with respect to the "goalposts."

3. Construction of the composite indicator.

The indicators considered for developing the synthetic index to evaluate the combined realizability and productive efficiency of the civil section in the courts of Santa Maria Capua Vetere and Matera were selected following a thorough literature analysis (Vecchi, 2018) and correlation analysis among the indicators. Specifically, the selected indicators are as follows:

1. Incoming Workload (CLI): It represents the ratio between new procedures registered during the period and the total procedures at the beginning of the reference period. This indicator expresses the load of new incoming work (Indicator Polarity: Negative).
2. Work Completed (LE): It represents the quantity of work completed on all open cases during the reference period, indicating the disposal of workload during that time (Indicator Polarity: Positive).
3. Turnover Index (IR): It relates the number of cases completed to the number of new cases received. When IR is greater than one, it means that the “topic” manages to handle more cases than it receives (Indicator Polarity: Positive).
4. Turnover Rate (TT): It relates the number of cases completed to the final pending cases. Also known as the rotation index, it describes the tendency of the case flow to renew (Indicator Polarity: Positive).
5. Average Pending Cases (GM): It calculates the ratio between the sum of final pending cases and the sum of initial pending cases, over the sum of new cases and completed cases. This "stock formula" provides an estimate of the days the case remained pending in the workflow (Indicator Polarity: Negative).

It is essential to emphasize that the selection of these indicators, aiming to provide an overall view of the phenomenon, was also driven by the intention to include both static dimension indicators (Incoming Workload and Work Completed) and dynamic dimension indicators (Turnover Rate, Average Pending Cases, and Turnover Index) to ensure a balanced synthesis from this perspective as well.

The Adjusted MPI_i^{-} index is used since the composite index is "positive," meaning that increasing values of the index correspond to positive variations in the court's efficiency. The "goalposts" are the observed minimum and maximum values for each indicator during the examined period. After polarizing the individual indicators, they all have a positive polarity, as high values of the variables are considered "good" regarding the court's efficiency.

Before examining the analysis, it is essential to note that the data considered for the cases are aggregated by “topic”. These generally correspond to the registers used by the official

databases of the Ministry and adopted by the Ministry's dedicated portal for monitoring the quality of justice.

4. Preliminary results

The use of this composite index (AMPI) allows us to monitor the absolute variations over time. These absolute variations are represented through a multiple interconnected line chart. As a result, Figures 1 and 2 will display the overall synthetic trend for each considered "topic", distinguishable based on the attached legend.

Figure 1 show the trend of each "topic" concerning the civil section of the Santa Maria Capua Vetere court, covering the period from 2013 to 2022. On the other hand, Figure 2 presents the trends related to the Matera court, with reference to the period 2020-2022. In this way, readers can observe the trends for each section within an open-ended range varying between 70 and 130.

Meanwhile, Tables 1 and 2 contain all the values related to the application of AMPI for each year considered, for both courts.

The combination of these two tools will enable us to comprehend the overall trend of each "topic" individually and in relation to others. This approach allows for a comprehensive overview of the efficiency levels of the respective court.

YEARS	CONTENTIOUS AFFAIRS	AGRICULTURAL DISPUTES	WORK	SOCIAL SECURITY AND ASSISTANCE	SUMMARY SPECIAL PROCEEDINGS	VOLUNTARY JURISDICTION AFFAIRS
2013	87.979	94.987	90.379	100.430	82.258	98.583
2014	95.263	103.867	91.909	102.111	86.143	99.431
2015	92.899	96.559	93.878	96.947	84.878	97.819
2016	97.934	102.340	101.282	92.073	104.118	101.092
2017	97.427	107.887	100.758	91.997	106.908	100.739
2018	96.804	103.758	100.188	92.904	106.316	102158
2019	96.122	100.980	98.384	90.794	105.300	100.874
2020	93.656	95.013	99.937	91.277	104.161	100.783
2021	98.554	116.694	99.175	95.251	112.604	102.036
2022	99.579	105.679	99.692	91.233	108.177	103.224

Table 1: *ere tribunal.*

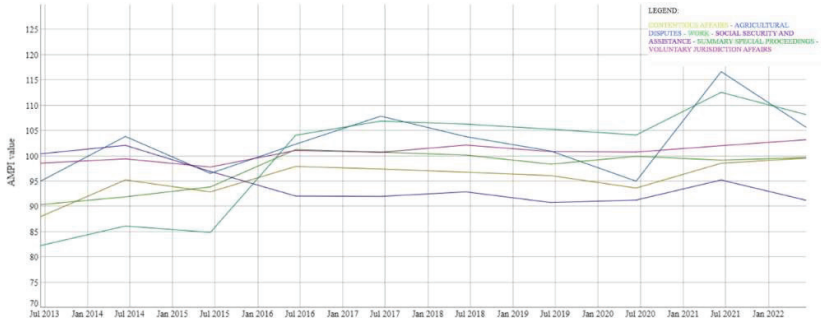


Figure 1: Application of the AMPI on the civil section of the Santa Maria Capua Vetere court.

The paragraph describes the efficiency trends of individual "topic" within the Civil section of the Santa Maria Capua Vetere court:

- "Contentious affairs": Shows a consistently increasing trend with small inflections in 2015 and 2020. The overall positive variation between 2013 and 2022 is more than 12 points, indicating significant improvement.
- "Agricultural disputes": Exhibits a less stable trend. It shows positive variation from 2013 to 2014, negative variation in 2015, followed by efficiency improvement until 2017. Subsequently, there is a decline until 2020, followed by a strong positive spike in 2021 (21 points). Finally, it decreases in 2022, still showing an overall improvement of just over 10 points.

- "Work": Demonstrates a growing efficiency trend from 2013 to 2016, with a positive variation of 10 points during this period. It stabilizes afterward, except for a slight negative variation in 2019 compared to 2016.
- "Social security and assistance": This is the only "topic" that records a decrease in efficiency throughout the considered period, with a negative variation of just over 9 points between 2013 and 2022. The trend is consistently declining, mainly between 2014 and 2016, and remains stable for the remaining period.
- "Summary special proceedings": Shows the most significant efficiency improvement within the considered period, with a positive variation of 26 points between 2013 and 2022. The main peak occurs between 2015 and 2016, with a positive variation of 20 points, indicating the highest acceleration in the last decade.
- "Voluntary jurisdiction affairs" Presents the most stable trend among all sections, with a positive variation of 5 points between 2013 and 2022, accumulated mainly from 2015 to 2022.

In summary, it can be asserted that, except for the Welfare and Assistance section, the Civil section of the Maria Capua Vetere court has experienced a significant increase in efficiency in the last decade.

As for the Civil section of the Matera court, due to a lack of sufficient data points, the analysis is relatively limited compared to the previous section. Nevertheless, some information can be highlighted, as shown in Table 2 and Figure 2.

Regarding the Contentious affairs topic, there is a gap of approximately 7 points between 2020 and 2023. As for the Voluntary Jurisdiction cases, there are practically the same efficiency points in 2020 and 2022, but there is still a surplus of 5 points in 2021. The efficiency of the Labor "topic" has seen an addition of 5 efficiency points in recent years. Unfortunately, the Welfare and Assistance "topic" shows a small gap, and in 2021, the efficiency dropped by six points. Lastly, concerning the Special Summary Proceedings, it can be argued that it has the worst trend among those considered, as the gap between 2020 and 2022 exceeds 10 points.

YEARS	CONTENTIOUS AFFAIRS	VOLUNTARY JURISDICTION AFFAIRS	WORK	SOCIAL SECURITY AND ASSISTANCE	SUMMARY SPECIAL PROCEEDINGS
2020	96.867	100.513	92.930	107.570	107.559
2021	90.893	104.079	91.991	99.334	102.853
2022	89.085	99.659	95.370	102.305	97.506

Table 2: Tabulated values by year of the synthetic indicator, Civil section of the Matera tribunal.

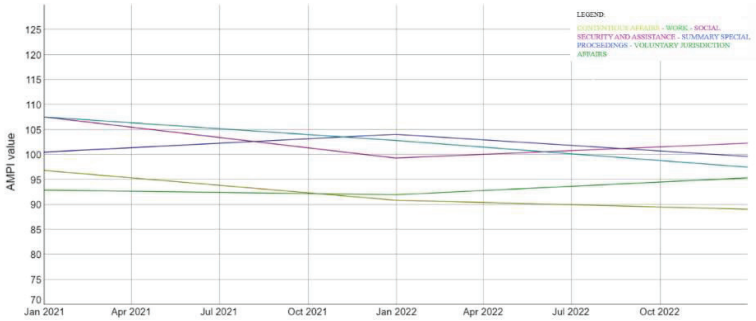


Figure 2: Application of AMPI on the civil section of the Matera court.

5.Conclusions and future investigation

The purpose of this article was to demonstrate the possibility of a subsequent application after considering the individual trends of the indicators. Through these analyses, it becomes

feasible to create an overall picture of the trends in a synthetic and non-compensatory manner. These processes also allow for the identification of absolute variations over time, enabling a concise comparison between the various years within the reference period. One of the primary challenges inherent in the construction of composite indices involves the selection of a methodology facilitating temporal comparisons.

The paper introduces a modified version of the MPI designed to enable temporal comparisons within a non-fully compensatory framework. The newly proposed index operates by rescaling individual indicators within a range of (70; 130), where 100 serves as the midpoint between the two goalposts associated with each indicator. This approach is characterized using a function (the product $S_{ri}\ CV_i$) to penalize the units with unbalanced values of the normalized indicators. The 'penalty' is based on the coefficient of variation and is zero if all the values are equal. The purpose is to favour the units that, means being equal, have a greater balance among the different indicators. Therefore, the adjusted MPI is characterized by the combination of a 'mean effect' (M_{ri}) and a 'penalty effect' ($S_{ri}\ CV_i$). To evaluate absolute performance changes over time, it is essential to identify the minimum and maximum values for each indicator across all the years under consideration, or alternatively, these values may be externally determined by the researcher.

Furthermore, the use of this method, in the context of future in-depth analyses, could lead to the establishment of cross-sectional rankings, meaning the comparison of multiple courts. As this synthetic indicator is inherently non-compensatory, it facilitates such applications, making the efficiency differences among different courts visible.

The objectives set for future investigations pertaining to the intention of comprehensively encompassing the entire population residing in the Italian peninsula, thus establishing a veritable national ranking of efficiency concerning judicial offices. This endeavour could also lead to the definition of specific efficiency standards to pursue.

References

Vecchi, G. (2018). *La valutazione delle performance negli uffici giudiziari come strumento di politica pubblica.* Milano: FrancoAngeli.

Gomes, P., Mendes, S. (2022). Organizational context, use of performance management practices and their effects on organizational performance: An empirical look at these interrelationships. *International Journal of Productivity and Performance Management,***72**(8), 10.1108/IJPPM-01-2022-0014.

Langbroek, A. (2011). Courts and judges and the media. *International Journal for Court Administration*, **4**(1). 10.18352/ijca.57.

Mazziota, M., Pareto, A. (2013). Methods for constructing composite indices: One for all or all for one. *Rivista Italiana di Economia Demografia e Statistica*, **LXVII**, pp.67-80.

Mazziotta, M., Pareto, A. (2016). On a generalized non-compensatory composite index. *Social Indicators Research*, **127**(3), pp.983–1003.

Mazziotta, M., Pareto, A. (2020). *Gli indici sintetici*. Torino: G. Giappichelli Editore.

Medeiros, F. (2015). Social and organizational impacts of the electronic processing system of the Brazilian superior tribunal of justice. *Journal of Information Systems and Technology Management*, **12**(2). doi:https://doi.org/10.4301/s1807.

Volacu, A. (2018). Justice, efficiency, and the new public management. *Australian Journal of Public Administration*, **77**(3), pp. 404–414. doi:10.1111/1467-8500.12263.

Zhou, H., Young, Y. (2022). Design and application of entrusted judicial expertise system under the background of "internet plus public judicial service". In *Proceedings of the 2022 2nd International Conference on Education, Information Management and Service Science (EIMSS 2022s*, Atlantis Press.

doi.org/10.26398/asaproc.0070

New scenarios for measuring household deprivation in sub-municipal areas

Flavio Biasciucci[a], Giancarlo Carbonetti[a], Raffaele Ferrara[a],
Matteo Mazziotta[a], Valeria Quondamstefano[a],
[a] Istat, Rome

1. Introduction

The availability of accurate and timely data at the sub-municipal level represents a great opportunity to improve and enrich statistical knowledge of the main demographic, economic and social transformations affecting local communities. In this regard, thanks to the data of the Italian Permanent Population Census and the information sources that feed the Istat Registry System, it will be possible to offer decision-makers a broad information base to better set policies at the local level: urban requalification interventions, social policies in favour of the weaker segments of the population, programming and planning acts.

This work provides an opportunity to present a project recently launched at Istat, in collaboration with a number of large municipalities, to study the phenomenon of family deprivation and the factors characterising it, and to identify areas within the municipal territory where the highest concentrations of cases of such deprivation are observed. In these areas it will be possible for municipalities to implement family-friendly policies and assess their effects through spatial and temporal comparisons. In this regard, some results of a preliminary study conducted on a number of municipalities are presented with the aim of assessing the feasibility and replicability of a methodological path of analysis and the potential of the project results.

2. The importance of sub-municipal data for policy makers

The municipalities, especially the larger ones, present strong heterogeneities within them in the distribution of economic and socio-demographic phenomena.

Data with a high level of spatial detail, such as those related to the census enumeration area, are needed by local authorities to access funding through participation in regional, national or international calls, or to meet specific knowledge needs about cities.

Istat is the only institution able to respond to the strong demand from users for statistical data with a high territorial detail (Carbonetti *et al.*, 2023a). In particular, municipalities require such data for policy planning at the local level, e.g. to contrast educational poverty, to plan interventions to support the economy, to implement labour policies or specific housing policies.

The possibility of identifying the areas where criticalities are most concentrated makes it possible to better target intervention actions so as to maximise the results expected from the use of financial resources. The possibility of proposing studies and analyses to research the areas in which policy efforts should be concentrated is a great opportunity that Istat wishes to seize as a challenge in thematic and methodological research and as an opportunity to return useful information to the municipalities.

3. Possible spatial analyses

For a number of years now, Istat has been on a path of innovation that has seen, on the one hand, a change in census strategy (Falorsi, 2017) and, on the other, the use of administrative source data for the definition of the integrated registry system. This path will lead to an increase in the availability and quality of data at the sub-municipal level. In particular, the use of data with high representativeness, accuracy and timeliness, which can be updated annually and geo-coded up to the enumeration areas, will expand the possibilities for spatial analyses of the main demographic, social and economic phenomena of the resident population.

The availability of very fine spatial datasets will allow, in continuity with the past, the conduct of studies on spatial differences within municipalities, the spatial concentration of specific population groups (residential segregation) and, in addition, the definition of interpretative models of the phenomena of interest that also take into account spatial autocorrelation. The availability of annual data series offers, in perspective, the possibility of studying and forecasting changes in the spatial dimension of the phenomena of interest, of monitoring the spatial dynamics of phenomena and their changes over the years.

Finally, sub-municipal areas with high concentrations of critical cases of the phenomenon under consideration can be identified, which can be monitored over the years to assess any changes in both the value of the reference indicators and the geographical limits, attributable to specific interventions implemented by local authorities (Carbonetti *et al.*, 2023b).

4. An experiment on the study of household deprivation

In this context, an exercise conducted on three municipalities - Bologna, Florence and Palermo - is illustrated with the aim of demonstrating the feasibility of the analysis path and the possible results, with reference to the study of the phenomenon of household deprivation.

4.1 Data and measures of household socio-economic deprivation

Istat annually acquires many archives from different sources, possibly exploiting a wide variety of information in the study of the main socio-demographic and economic phenomena at high territorial detail. The sources and archives used in this contribution represent only a selection of the information potential available to Istat for analyzing the socio-economic deprivation of households. The sources used were INPS (Wage Guarantee Fund beneficiary Archives, Citizenship Income, Pensioners and Pension Benefits Register), INAIL (Temporary Workers Archive), MIUR (Register of Students in Schools), ACI (Vehicle Registration Database), and the Tax Agency (Land Registry of Real Estate Units).

The integration of data from the different such archives and some socio-demographic information (age and employment status) from Istat's Permanent Population Census made it possible to identify and construct 7 elementary indicators to be used later in the calculation of a single synthetic index of family deprivation. The indicators were calculated for the municipalities of Bologna, Florence and Palermo at the territorial level of Functional Areas[1] (FAs) and Enumeration Areas[2] (EAs). Some summary data are shown in Table 1.

Table 1: Census population, number and average population of FAs and EAs of Bologna, Florence and Palermo (Italian Permanent Census 2021; 2011 Territorial Bases).

	Census population	Number and average population of FAs		Number and average population of EAs(*)	
Bologna	387,842	90	4,309	2,070	187
Florence	361,619	72	5,022	1,957	185
Palermo	635,439	55	11,553	2,704	235

(*) EAs refer to the 2011 Istat Territorial Bases; only EAs with population are considered in the table.

Below is a list of the indicators and how they are calculated:

Ind1 - *Population aged 25-64 with low educational qualifications (%)* - Ratio of the population aged 25-64 with low level of education to the total population aged 25-64.

Ind2 - *Individuals aged 15-64 who are non-employed and non-pensioners with a citizenship income (%)* - Ratio of the population aged 15-64 who are non-employed, non-pensioners and with a citizenship income, to the total population aged 15-64.

[1] They are defined by municipalities for various purposes (administrative, functional, and statistical) as an aggregation of contiguous EAs and constitute a partition of the municipal territory. Larger municipalities have more than one sub-municipal subdivision. In this work the finest subdivisions were considered: "Elementary Areas" for Bologna; "Statistical Areas" for Florence; "First Level Units" for Palermo.

[2] For the analysis at the EA level, the 2011 Territorial Bases were considered as the 2021 Territorial Base will be released at the end of 2023.

Ind3 - *Individuals aged 67 and over without home ownership and without cars (%)* - Ratio of the population aged 67 and over without home ownership and without cars to the total population.
Ind4 - *Employment rate age 20-64 (%)* - Ratio of the employed population age 20-64 to the total population age 20-64.
Ind6 - *Individuals aged 15-29 not working and not enrolled in any course of study (%)* - Ratio of the population aged 15-29 not working and not studying to the total population aged 15-29.
Ind6 - *Individuals aged 15-64 in the Wage Guarantee Fund (%)* - Ratio of the population aged 15-64 who have received the Wage Guarantee Fund to the total population aged 15-64.
Ind7 - *Temporary workers aged 15-64 (%)* - Ratio of temporary workers aged 15-64 to the total population aged 15-64.

The indicators described above all have a positive polarity (see footnote 3) except for the Ind4 which has a negative polarity.

4.2 The Adjusted Mazziotta-Pareto Index

In order to synthesize the individual indicators in a unique measure, the Adjusted Mazziotta-Pareto Index (AMPI) is used. It is a partially non-compensatory composite indicator based on a standardization of the individual indicators, at the reference time, that makes the indicators independent from the unit of measure (De Muro *et al.*, 2011). Therefore, all the individual indicators are assigned equal weights and also absolute time comparisons are allowed (Mazziotta and Pareto, 2016). In fact, a re-scaling of the individual indicators in the range (70; 130) according to two 'goalposts' is proposed, i.e., a minimum and a maximum value which represent the possible range of each variable for all time periods and for all units. The base value=100 is taken from the reference municipality. The steps for computing the AMPI are given below (Mazziotta and Pareto, 2016).

Given the matrix X={xij} with n rows (units) and m columns (individual indicators), we calculate the normalized matrix R={rij} as follow:

$$r_{ij} = \frac{(x_{ij} - \text{Min}_{x_j})}{(\text{Max}_{x_j} - \text{Min}_{x_j})} 60 + 70$$

where x_{ij} is the value of the indicator *j* for the unit *i* and Min_{x_j} and are the 'goalposts' for the indicator *j*. If the indicator *j* has negative polarity[3], the complement of the formula aforementioned with respect to 200 is computed. Denoting with M_{r_i} e S_{r_i}, respectively, the mean and the standard deviation of the normalized values of the unit *i*, the generalized form[4] of the AMPI is given by:

$$\text{AMPI}_i^{+/-} = \text{M}_{r_i} \pm \text{S}_{r_i} \text{cv}_i$$

where $\text{cv}_i = \text{S}_{r_i} / \text{M}_{r_i}$ is the coefficient of variation of the unit *i* and the sign ± depends on the kind of phenomenon to be measured. The 'price' to pay for having final scores comparable over time is that individual indicators with different variability are aggregated. However, normalized indicators in an identical range have much more similar variability than original ones. For the mathematical properties of AMPI see Mazziotta and Pareto, 2016. In order to compare AMPI with a compensatory method, the arithmetic mean of the standardized values (using the same procedure of AMPI) is computed: in this way, the only differences between the methods are due to the function of synthesis (Mazziotta and Pareto, 2017).

[3] The indicator's "polarity" is the sign of the relation between the indicator and the phenomenon to be measured (+ if the individual indicator represents a dimension considered positive and - if it represents a dimension considered negative).
[4] It is a generalized form since it includes 'two indices in one'.

4.3 Some results

We report here the summary data of the indicators illustrated in Section 4.1 and the value of the AMPI (Section 4.2), calculated for the three municipalities considered (Bologna, Florence, and Palermo), with reference to both FAs (Table 2) and EAs (Table 3). It should be noted that for the calculation of the AMPI, the data referring to the entire municipal territory for each municipality was taken as the base value (equal to 100). The highest values of the AMPI show the most critical deprivation situations with respect to the municipality taken into consideration, conversely, the lowest values show a low number of deprivation cases.

Table 2: Minimum, median, and maximum of indicator values (simple and AMPI) calculated for the FAs of Bologna (BO), Florence (FL), and Palermo (PA) (Istat, 2021).

	Minimun			Median			Maximum		
	BO	FL	PA	BO	FL	PA	BO	FL	PA
Ind1	0.0	12.8	12.6	21.4	22.1	43,7	60.0	43.3	75.7
Ind2	0.0	0.0	1.0	0.6	0.6	6.6	12.4	2.9	15.1
Ind3	0.0	2.0	2.4	3.6	4.1	4.0	29.4	7.9	5.0
Ind4	49.1	67.1	34.5	78.9	77.4	53.3	95.7	85.0	65.7
Ind5	0.0	5.9	14.0	13.2	13.7	27.7	100.0	27.8	48.3
Ind6	0.0	0.0	2.1	4.0	5.5	3.8	26.7	9.1	5.1
Ind7	0.0	0.0	0.3	3.0	1.6	0.8	13.6	3.5	1.1
AMPI	*91.7*	*87.9*	*75.8*	*100.1*	*99.9*	*102.0*	*124.5*	*124.4*	*121.3*

Table 3: Minimum, median, and maximum of indicator values (simple and AMPI) calculated for the EAs of Bologna (BO), Florence (FL), and Palermo (PA) (Istat, 2021).

	Minimun			Median			Maximum		
	BO	FL	PA	BO	FL	PA	BO	FL	PA
Ind1	0.0	0.0	0.0	17.4	21.7	40.0	100.0	100.0	100.0
Ind2	0.0	0.0	0.0	0.0	0.0	5.0	33.3	33.3	100.0
Ind3	0.0	0.0	0.0	3.1	3.3	3.4	100.0	100.0	100.0
Ind4	0.0	0.0	0.0	79.8	77.7	55.2	100.0	100.0	100.0
Ind5	0.0	0.0	0.0	10.6	11.6	24.1	100.0	100.0	100.0
Ind6	0.0	0.0	0.0	3.5	5.0	3.1	100.0	100.0	100.0
Ind7	0.0	0.0	0.0	2.1	1.2	0.0	77.8	50.0	25.0
AMPI	*94.3*	*93.7*	*89.8*	*99.6*	*99.7*	*99.6*	*125.1*	*124.4*	*123.0*

Figures 1, 2 and 3 show the AMPI values, classified according to the deciles of the distribution, for each of the three municipalities examined. In this regard, Figures 1a, 2a and 3a refer to FAs while Figures 1b, 2b and 3b refer to EAs. From the visual analysis of the FAs maps, one can easily identify the most critical sub-municipal areas (those with red textures) for which the highest AMPI values were obtained. Analysing the EAs map, the high level of granularity of the data favours a greater detail of information that leads to the identification of several critical areas[5] where, from the AMPI values, the highest concentrations of deprivation cases are observed. Comparing the FAs and EAs maps for each municipality, two types of critical areas can be identified in a similar way for the three municipalities considered:

a) "known" areas (circled in blue) that fall within the most critical FAs;
b) "hidden" areas (circled in yellow) that fall within non-critical FAs and emerge because the spatial detail of the analysis has increased (by enumeration areas).

[5] The boundaries of these areas will be the result of an aggregation procedure of contiguous and homogeneous EAs with respect to a predefined AMPI threshold value; the determined areas will also have to respect certain constraints of geographical and/or demographic size.

Figure 1a: map of AMPI values on Bologna FAs.

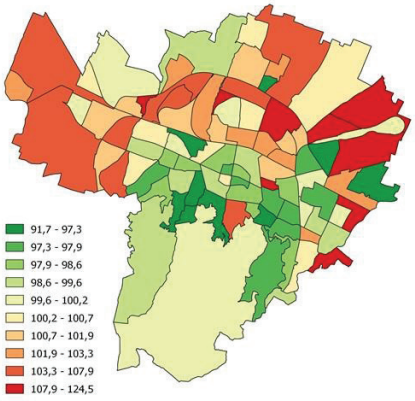


Figure 1b: map of AMPI values on Bologna EAs and indication of potential "critical areas".

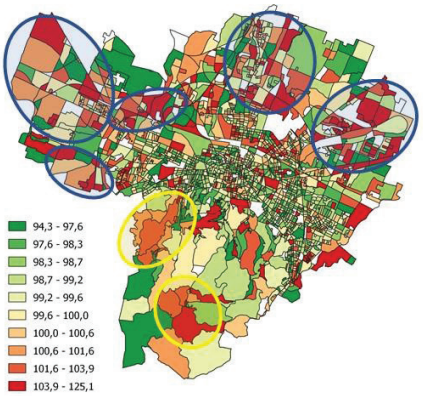


Figure 2a: map of AMPI values on Florence FAs.

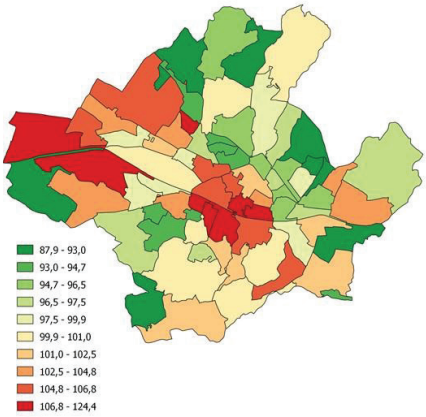


Figure 2b: map of AMPI values on Florence EAs and indication of potential "critical areas".

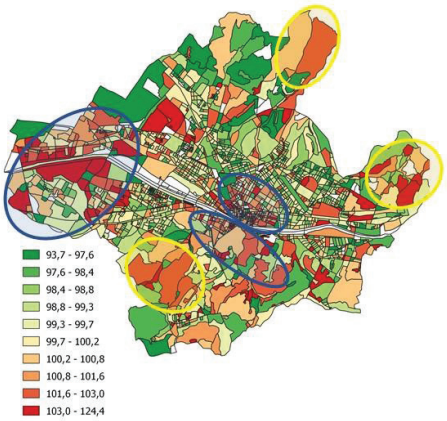


Figure 3a: map of AMPI values on Palermo FAs.

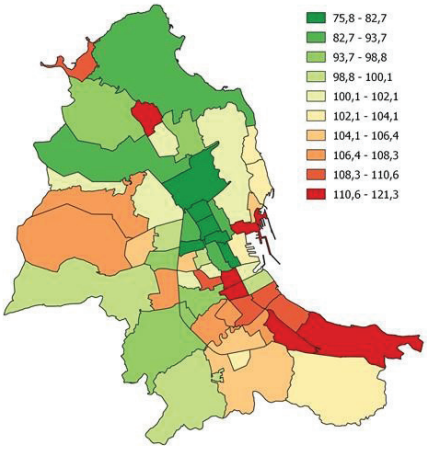
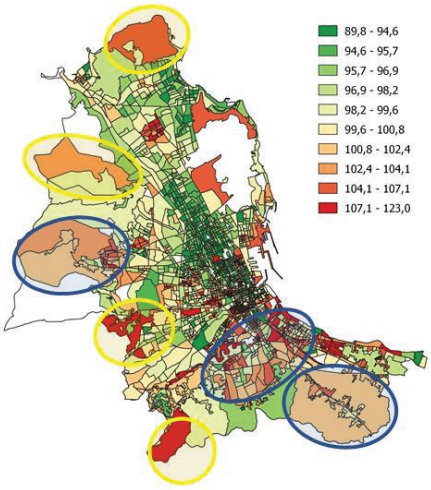


Figure 3b: map of AMPI values on Palermo EAs and indication of potential "critical areas".



5. Perspectives

The project represents a great opportunity to exploit both the results of the Italian Permanent Census and the administrative archives that feed into Istat's Registry System, offering different possibilities for reading the territory. In this work the focus has been on the study of household deprivation, but there are several phenomena relating to individuals and households for which there is a strong interest in conducting territorial analyses similar to the one reported, on an experimental basis, in this work.

The task is only just beginning and requires various study, research and design activities concerning:

- definition of deprivation and characterization of its components;
- analysis of sources and construction of simple indicators;
- choice of the summary indicator of the simple indicators;
- determination of the algorithm for the identification of critical sub-municipal areas;
- design of the results dissemination platform.

The coordination of these activities by Istat and the cooperation with some of the largest Italian municipalities is a strength of the project in several respects. Since we operate within the National Statistical System, there is a guarantee of the quality of the results and compliance with the principles of official statistics. Istat's conduct allows the use of annual data series with high territorial detail, both on a census basis and from administrative sources, homogeneous for all municipalities; this will also offer the possibility of spatial and temporal analyses.

The most relevant aspect, however, is the involvement of the municipalities who know the territory in which they operate and can provide the correct keys to interpreting the results that will be obtained.

The expected results will define a new model for territorial analysis of demo-social and economic phenomena of individuals and households and will constitute a fundamental tool for local policy choices and for maximising the effect of the financial resources involved.

References

De Muro, P., Mazziotta, M., Pareto, A. (2011). Composite Indices of Development and Poverty: An Application to MDGs. *Social Indicators Research*, 104, pp.1-18.

Falorsi, S. (2017). Census and Social Surveys Integrated System. Note by the National Institute of Statistics of Italy, presented at the UNECE/Eurostat Group of Experts on Population and Housing Censuses, Nineteenth Meeting, Geneva, Switzerland, 4–6 October 2017.

Carbonetti, G., Ciccarese, A., Roncati, R. (2023a). An analysis of the demand for sub-municipal data from the Population and Housing Census. Review of Official Statistics, Rome, Istat *(under evaluation)*.

Carbonetti, G., Dardanelli, S., De Matteis, G., Ferrara, R., Verrascina, M. (2023b). Le potenzialità di analisi a livello sub-comunale con l'offerta informativa del Censimento Permanente della Popolazione e delle Abitazioni. Note by the National Institute of Statistics of Italy, presented at the Sis-Aisp, Population Days 2023, Rome, Italy 1–3 February 2023.

Mazziotta, M., Pareto, A. (2017). Synthesis of indicators: the composite indicators approach. In: Complexity in Society: From Indicators Construction to their Synthesis. *Social Indicators Research Series*. Filomena Maggino Editors, Springer, pp.159-191.

Mazziotta, M., Pareto, A. (2016). On a Generalized Non-compensatory Composite Index for Measuring Socio-economic Phenomena. *Social Indicators Research*, 127 (3), pp.983-1003.

doi.org/10.26398/asaproc.0071

Sustainability consciousness and well-being: Results from an Italian survey

Silvia Bacci[a], Bruno Bertaccini[a], and Anna Pettini[b]

[a]Dep. of Statistics, Computer science, Applications "G. Parenti", University of Florence, Italy
[b]Dep. of Economics, University of Florence, Italy

1. Introduction

The concept of sustainable development is more and more pervasive in today's society and there is now a general consensus that it is an attempt to bring about global change by striking a balance between economic prosperity, social progress and environmental protection.

Many indices have been defined to measure the sustainability of a system and to guide policy actions towards sustainability, however measures needed to assess the degree of public awareness of the full meaning and importance of sustainability are very few. The first psychometric instrument developed to measure the individuals' sustainability consciousness is represented by the Sustainability Consciousness Questionnaire (SCQ), recently developed by Gericke et al. (2019). This questionnaire collects information about people's attitudes, knowledge and behaviour in the economic, social and environmental domains and, thus, it allows to construct latent variables that help to see in detail how strongly and in what form respondents' sustainability consciousness is expressed. The questionnaire was written in English and designed for Swedish people, thus a validation is necessary to use it in other countries and cultures. The authors of the present contribution validated (in a previous work just submitted) the Italian version of the SCQ on a sample of Italian people collected before the Covid-19 outbreak.

In this contribution, we aim at testing the robustness of the validated version of the Italian SCQ using a sample of data collected after the first lockdown. Moreover, in the light of the recent literature concerning the association between subjective well-being and some aspects of sustainability consciousness (mainly related to the environmental dimension; Ojala, 2013; Suárez-Varela et al., 2016) we enrich the original questionnaire adding a section devoted to the measurement of the self-perceived well-being, in order to investigate whether and to what extent the sustainability consciousness is affected by the subjective well-being. For these aims, we implement a structural equation model that accounts for the relations among observed items and latent constructs and among latent constructs.

The remaining part of the contribution is structured as follows. Details on the questionnaire and the statistical approach are provided in Sections 2. and 3, respectively. Preliminary results are illustrated in Section 4. Section 5. concludes with some final remarks.

2. The questionnaire

The questionnaire adopted is composed of two main parts: one concerns the measurement of the sustainability consciousness and the other one is about the measurement of the self-perceived well-being.

The sustainability consciousness may be conceptualised as an unobservable (i.e., latent) construct that relies on three specific underlying latent constructs: Knowingness, Attitudes and Behaviour. Knowingness denotes what people recognise as the necessary characteristics of sustainability; Attitudes refer to personal beliefs towards sustainability, and Behaviour reflects what individuals do in relation to sustainability. In turn, each of these constructs can be disentangled in three further dimensions: an environmental dimension, a social dimension, and an economic dimension. Hence, a hierarchy of latent constructs results that are ordered on three levels: at the top level the overall Sustainability Consciousness (SC), which is decomposed at a lower level in sustainability knowingness (K-SUS), sustainability attitudes (A-SUS), and sustainability behaviour (B-SUS), which in turn at the bottom level decompose, respectively, in K-ENV, K-SOC, and K-ECO that contribute defining the knowingness of sustainability; A-ENV, A-SOC, and A-ECO that contribute defining the attitudes towards sustainability; and B-ENV, B-SOC, and B-ECO that contribute defining the sustainable behaviour.

This theoretical framework inspired the original version of the Sustainability Consciousness Questionnaire (SCQ; Gericke et al., 2019). The SCQ is composed of 49 items that reduce to 27 in the validated short form. Items are measured on a Likert type scale with five ordered categories. The original SCQ is written in English language and designed for Swedish people.

The self-perceived well-being is measured through 7 items defined on a Likert type scale with ten ordered categories. The items are defined following the definition of "subjective well-being" provided by ISTAT in the "Indagine Multiscopo sulle Famiglie: Aspetti della Vita Quotidiana" (see at `https://www.istat.it/it/archivio/91926`).

3. Statistical method

The validation of the Italian version of the SCQ was performed by ourselves in a previous study along the same lines of Gericke et al. (2019). Relations among the first-level latent constructs and the observed items as well as relations among latent constructs at first-, second-, and third levels were analysed and tested on the basis of Structural Equation Models (SEM; Bollen, 1989; Bollen et al., 2008). More precisely, we performed a Hierarchical (or Higher-order) Confirmatory Factor Analysis (CFA; Jöreskog, 1969), assuming that the variability of the 9 latent constructs of the first level (i.e. K-ENV, K-SOC, K-ECO, A-ENV, A-SOC, A-ECO, B-ENV, B-SOC and B-ECO) is explained by the 3 second-level latent constructs (i.e. K-SUS, A-SUS, B-SUS), whose variability is in turn explained by the global Sustainability Consciousness (SC). In the present contribution, we focus on the short form of the validated SCQ because it reaches a better fit with respect to the long form, as already pointed out by Gericke et al. (2019). It is worth outlining that the short form of the SCQ was built according to a data-driven strategy, by selecting the three items with the highest factor loadings for each first-level latent construct and adjusting for the covariances between observed items whenever necessary.

The Italian short form of SCQ was validated on a sample of 614 Italian university students before the beginning of the Covid-19 pandemic (wave 1; Model 1). Now, we test it on a new sample of data collected on 359 university students after the first lockdown due to Covid-19 (wave 2; Model 2). For this aim, we calculate and compare traditional fit measures, such as the Tucker and Lewis Index (TLI; Tucker and Lewis, 1973), the Comparative Fit Index (CFI; Bentler, 1990), and the Root Mean Square Error of Approximation index (RMSEA; Steiger, 1990). Furthermore, we enrich the analysis extending the Hierarchical CFA model to a SEM adding the measurement of the self-perceived well-being and its regression on the sustainability consciousness (Model 3).

All models at issue are estimated through the maximum likelihood approach using the R package `lavaan`, version 0.6-12 (Rosseel, 2023).

4. Results

Table 1 displays the goodness of fit measures of the three estimated models, whereas Figure 1 shows the structural and measurement parts of Model 3, that is, the model that integrates SC and WB. The standardised regression coefficients are displayed on the arrows linking the latent constructs, whereas the standardized factor loadings (linking latent constructs to observed items) are omitted for the sake of parsimony.

Table 1: Goodness of fit of SEMs: CFI, TLI, and RMSEA

Model	Description	CFI	TLI	RMSEA
Model 1	Italian short form of SCQ validated on wave 1	0.949	0.942	0.036
Model 2	Italian short form of SCQ tested on wave 2	0.950	0.943	0.044
Model 3	Italian short form of SCQ + WB (on wave 2)	0.973	0.970	0.033

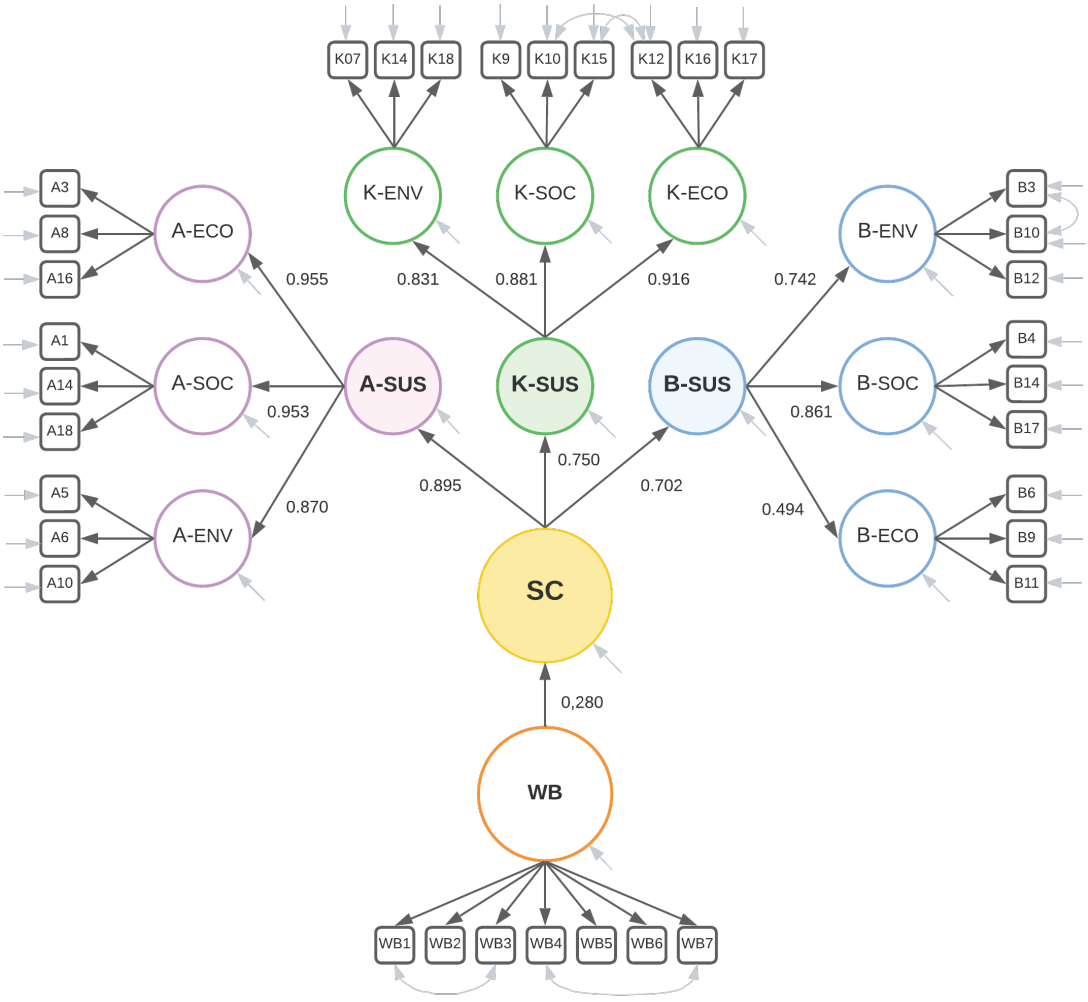


Figure 1: Structure of the latent constructs SC and WB with estimated factor loadings and regression coefficients.

Looking at Table 1, the factorial structure of the short form of the questionnaire validated in the original proposal is confirmed for the Italian version (Model 1), since CFI and TLI reach satisfactory values with a RMSEA definitely lower than 5%. It is worth to outline that the structure of the questionnaire in the Italian version was validated on the first wave (pre-pandemic survey), but it appears robust with data collected on the second wave (post-pandemic survey; Model 2), being the fit measures aligned with those of the first wave. Moreover, the introduction of the section concerning the self-perceived well-being (Model 3) provides evidence for a further improvement.

Looking at Figure 1, at the highest hierarchical level the global SC mainly affects the attitudes towards sustainability (A-SUS; standardised coefficient equal to 0.895) and at a minor extent the dimensions related to knowledge (K-SUS; 0.750) and behaviour (B-SUS; 0.702). At the second level of the hierarchy, A-SUS and the K-SUS contribute similarly to the social and economic dimensions, while the effect on the environmental dimension is slightly lower. Differently, B-SUS influences primarily the social behaviour (standardised coefficient equal to 0.861), followed at a certain distance by the environmental behaviour (0.742), whereas the effect on the economic behaviour is definitely lower (0.494). This last result partly contrast with what observed in the Swedish data, where B-SUS influences economic behaviour at the same extent than the social one. Finally, we observe a statistically significant effect of perceived well-being on the sustainability consciousness (estimated regression coefficient equal to 0.280).

5. Conclusions

In the present study we investigated the robustness of the Italian short form of the Sustainability Consciousness Questionnaire (Gericke et al., 2019), which has just been validated by ourselves in a previous study. Through structural equation models, the theoretical model validated on a pre-pandemic sample of university students was compared with the same model estimated on a new sample of data collected after the first lockdown due to Covid-19 pandemic. Results from the second wave are coherent with those from the first wave, with models estimated on the two samples achieving the same fit. In addition, the questionnaire was enriched with a part concerning the measurement of self-perceived well-being.

The estimated model shows that people express their sustainability consciousness mainly into the attitudinal component, more than into knowledge and, mostly, behaviour. This result is coherent with other studies carried out on a sample of Swedish and Taiwanese students (Berglund et al., 2020) and of Spanish students (Vegel, 2021). Moreover, self-perceived well-being has a positive and significant effect on the sustainability consciousness.

Work in progress is focusing on generalising the structural equation model to assess the effects of well-being on the single dimensions of the sustainability consciousness and to exploit in more detail possible changes in the relations between latent constructs intervening after the pandemic outbreak; individual covariates (e.g., gender, age) will be also taken into account.

References

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletiny*, 107:238–246.

Berglund, T., Gericke, N., Boeve-de Pauw, J., Olsson, D., and Chang, T.-C. (2020). A cross-cultural comparative study of sustainability consciousness between students in taiwan and sweden. *Environment, development and sustainability*, 22:6287–6313.

Bollen, K. A. (1989). *Structural Equation with Latent Variables*. Wiley, New York, USA.

Bollen, K. A., Rabe-Hesketh, S., and Skrondal, A. (2008). Structural equation models. In

BoxSteffensmeier, J. M., Brady, H., and Collier, D., editors, *The Oxford Handbook of Political Methodology*, pages 432–455. Oxford University Press, Oxford.

Gericke, N., Boeve-de Pauw, J., Berglund, T., and Olsson, D. (2019). The sustainability consciousness questionnaire: The theoretical development and empirical validation of an evaluation instrument for stakeholders working with sustainable development. *Sustainable Development*, 27:35–49.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34:183–202.

Ojala, M. (2013). Coping with climate change among adolescents: Implications for subjective well-being and environmental engagement. *Sustainability*, 5:2191–2209.

Rosseel, Y. (2023). Latent variable analysis.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25:173–180.

Suárez-Varela, M., Guardiola, J., and González-Gómez, F. (2016). Do pro-environmental behaviors and awareness contribute to improve subjective well-being? *Applied Research Quality Life*, 11:429–444.

Tucker, L. R. and Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38:1–10.

Vegel, A. (2021). Are university students conscious about sustainability? In Bele, D. and Weis, L., editors, *Sustainable Development in a Modern Knowledge Society*, pages 41–53. Ljubljana School of Business, Ljubljana.