Statistica Applicata - ITALIAN JOURNAL OF APPLIED STATISTICS
Vol. 34 n. 1 January - April 2022

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 72 34 29 04). Advertising: RCE Multimedia, Piazza Bagnoli 19 – 80124 Napoli, Italy (phone and fax +39 0812303416); email: rcemultimedia@yahoo.it.
Rules for manuscript submission: http://sa-ijas.stat.unipd.it.
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. RCE Multimedia 1990-Sept. 6, 2020. Associazione per la Statistica Applicata from Sept. 7, 2020.

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/)

MICRO & MACRO MARKETING (www.mulino.it/edizioni riviste/issn/1121-4228)

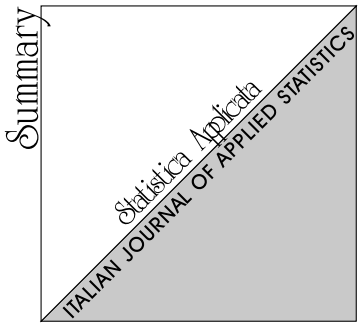
SINERGIE – Rivista di Studi e Ricerche (www.sinergiejournal.it)

Published: July 2022

RCE Multimedia

Piazza Bagnoli, 19 – 80124 Napoli Phone - Fax +39 081 230 34 16 +39 3471706215

email: rcemultimedia@yahoo.it



Volume 34 number 1

ON ZERO INFLATED GENERALIZED POISSON-SUJATHA DISTRIBUTION AND ITS APPLICATIONS

Samuel Adewale Aderoju[1]

Department of Statistics and Mathematical Sciences, Kwara State University, Malete, P.M.B. 1530, Ilorin, Kwara State, Nigeria.

Isaac Adeola Adeniyi

Department of Mathematical Sciences, Federal University Lokoja, P.M.B. 1530, Lokoja, Kogi State, Nigeria.

Abstract: *A new model called Zero Inflated generalized Poisson-Sujatha distribution (ZIGPSD) is proposed in this paper. Some characteristics of the model were derived and the maximum likelihood method was used to obtain the estimators of the parameters numerically. Through simulation and application to two datasets, the goodness-of-fit of the ZIGPSD was examined in comparison with the zero inflated Poisson (ZIP), Poisson and the new generalized Poisson-Sujatha (NGPS) models. The results showed that the ZIGPSD provides better fit compared to the other existing models considered in some cases of count data with excess zeros.*

Keywords: *Zero inflated Poisson, Excess zeros, Goodness-of-fit, Moments, Maximum Likelihood*

1. INTRODUCTION

The Poisson distribution is generally considered as the standard model for modeling count data but overdispersion and excess zeros than expected from the Poisson distribution are common problems in modelling count data, which the model cannot handle. Johnson and Kotz (1969) were the first to define a mixture Poisson distribution that accounted for excess zeros in the data (Sirichantra & Bodhisuwan, 2017). However, in some cases, overdispersion is a function of excess zero count or none of it in the data. The count data with excess zeros are common in various fields like physical, natural, biological and social sciences as well as in engineering and agriculture. Regular discrete distribution may fail to fit such data either because of zero inflation or over- or underdispersion (Aryuyuen et al., 2014). There is increased interest in zero inflated distribution to account for extra zeros, which are

[1] Corresponding author: Samuel Adewale Aderoju: E-mail: samuel.aderoju@kwasu.edu.ng

common in count data

"The zero inflated distribution assumes that the observed data are the result of a two-part process that generates structural zeros and a process that generates random counts" said Aryuyuen et al., (2014). They further expressed that the distribution can be simply defined as follows:

$$P(X = x|\omega) = \omega\pi_0(x) + (1 - \omega)f(x|\lambda), \quad (1)$$

where ω is the extra proportion of zeros, *X* is the count variable, *f(x)* is the pmf of *X* with parameter λ, and π_0 *(x)* = 1 if *x* = 0; otherwise π_0 *(x)* = 0.

In this paper, we derived zero inflated generalized Poisson-Sujatha (ZIGPS) distribution by compounding zero inflated Poisson (ZIP) distribution with Tesfay and Shanker's proposed "another two-parameter Sujatha distribution (ATPSD)", Tesfay and Shanker (2019). Some characteristics of the ZIGPS distribution and its application are shown as well.

The rest of this paper is organized as follows: in section 2 the new zero inflated model is presented and some characteristics of the model are derived in section 3. The Maximum likelihood estimation (MLE) method to estimate the parameters of the model are discussed in section 4. Simulation studies are presented in section 5 while section 6 contains the application of the model to two real data sets. The conclusion is presented in section 7.

2. THE ZERO INFLATED GENERALIZED POISSON-SUJATHA DISTRIBUTION

The zero inflated Poisson distribution having probability mass function (pmf) as follows:

$$h(X = x) = \begin{cases} \omega + (1 - \omega)e^{-\lambda}, & if\ \ x = 0 \\ (1 - \omega)\dfrac{\lambda^x}{x!}e^{-\lambda}, & if\ \ x = 1, 2, \ldots \end{cases}, \quad (2)$$

where $0 \leq \omega \leq 1$ and $\lambda > 0$

Suppose the parameter λ in ZIP distribution is a random variable and it follows the ATPSD, such that $\lambda \sim$ *ATPS* (α, β) that with the probability density function (pdf):

$$g(\lambda) = \frac{\beta^3}{\beta^2 + \alpha\beta + 2\alpha}(1 + \alpha\lambda + \alpha\lambda^2)\, e^{-\beta\lambda}; \quad \lambda > 0, \beta > 0, \alpha \geq 0, \quad (3)$$

where β is a scale parameter and α a shape parameter (Tesfay and Shanker, 2019). Obviously, for $\alpha = 0$ and $\alpha = 1$ the ATPSD reduces to exponential distribution and Sujatha distribution respectively.

Definition 1: Suppose a random variable *X* is said to be of zero inflated generalized Poisson-Sujatha distribution (ZIGPSD) if:

$$X|\lambda \sim ZIP(\lambda)$$

$$while \;\; \lambda|\alpha,\beta \sim ATPSD(\alpha,\beta)$$

for $\lambda > 0$, $\beta > 0$ and $\alpha \geq 0$. Hence, we denote the pmf of $X \sim ZIGPS(\alpha, \beta, \omega)$ is given by:

$$f(x) = \begin{cases} h(X=0)g(\lambda), & if \;\; x = 0 \\ (1-\omega)h(X \geq 1)g(\lambda) & if \;\; x = 1,2,\ldots \end{cases} \tag{4}$$

where, $0 < \omega < 1$ is the inflation parameter.

Theorem 1: If $X \sim ZIGPS(\alpha, \beta, \omega)$ be a Zero inflated Generalized Poisson-Sujatha distribution, then the pmf is:

$$f(x) = \begin{cases} \omega + (1-\omega)\dfrac{\beta^3}{\beta^2+\alpha\beta+2\alpha}\left(\dfrac{\alpha(\beta+3)+(\beta^2+2\beta+1)}{(\beta+1)^3}\right), & if \;\; x = 0 \\ (1-\omega)\dfrac{\beta^3}{\beta^2+\alpha\beta+2\alpha}\left(\dfrac{\alpha(x^2+\beta+3)+(\beta+4)\alpha x+(\beta^2+2\beta+1)}{(\beta+1)^{x+3}}\right), & if \;\; x = 1,2,\ldots \end{cases} \tag{5}$$

where $\beta > 0$, $\alpha \geq 0$ and $0 \leq \omega \leq 1$.

***Proof of part 1*:** [if X = 0]: Suppose $X|\lambda \sim ZIP(\lambda)$ *and* $\lambda|\beta, \alpha \sim ATPSD\ (\beta, \alpha)$ then the pmf of conditional random variable X is given as:

$$f_1(x) = \int_0^\infty h(X=0)g(\lambda)d\lambda\,,$$

where $\omega + (1-\omega)h(X=0)$ is ZIP when $x = 0$ and $g(\lambda)$ is ATPSD. Therefore,

$$f_1(x) = \int_0^\infty \omega + (1-\omega)e^{-\lambda}\frac{\beta^3}{\beta^2+\alpha\beta+2\alpha}(1+\alpha\lambda+\alpha\lambda^2)\,e^{-\beta\lambda}d\lambda$$

$$= \omega + (1-\omega)\frac{\beta^3}{(\beta^2+\alpha\beta+2\alpha)}\int_0^\infty e^{-\lambda(\beta+1)}\,(1+\alpha\lambda+\alpha\lambda^2)\,d\lambda$$

$$= \omega + (1-\omega)\frac{\beta^3}{(\beta^2+\alpha\beta+2\alpha)}\left[\int_0^\infty e^{-\lambda(\beta+1)}\,d\lambda + \alpha\int_0^\infty \lambda e^{-\lambda(\beta+1)}\,d\lambda + \alpha\int_0^\infty \lambda^2 e^{-\lambda(\beta+1)}\,d\lambda\right]$$

$$= \omega + (1-\omega)\frac{\beta^3}{(\beta^2+\alpha\beta+2\alpha)}\left[\frac{0!}{(\beta+1)^1}+\frac{\alpha}{(\beta+1)^2}+\frac{2\alpha}{(\beta+1)^3}\right]$$

$$= \omega + (1-\omega)\frac{\beta^3}{(\beta^2+\alpha\beta+2\alpha)}\left[\frac{\alpha(\beta+3)+(\beta^2+2\beta+1)}{(\beta+1)^3}\right]$$

$$\therefore f_1(x) = \omega + (1-\omega)\frac{\beta^3}{(\beta^2+\alpha\beta+2\alpha)}\left[\frac{\alpha(\beta+3)+(\beta+1)^2}{(\beta+1)^3}\right]; \quad if\ x=0 \tag{6}$$

Proof of part 2 [if $X \geq 1$]: Suppose $X|\lambda \sim ZIP(\lambda)$ *and* $\lambda|\beta, \alpha \sim ATPSD(\beta, \alpha)$ then the pmf of conditional random variable X is given as:

$$f_2(x) = \int_0^\infty \omega + (1-\omega)h(X \geq 1)g(\lambda)d\lambda\,,$$

where ω + (1- ω) h ($X \geq 1$) is ZIP when $x > 0$ and (g)λ) is ATPSD. Therefore,

$$f_2(x) = \int_0^\infty (1-\omega)\frac{\lambda^x}{x!}e^{-\lambda}.\frac{\beta^3}{\beta^2+\alpha\beta+2\alpha}(1+\alpha\lambda+\alpha\lambda^2)\,e^{-\beta\lambda}d\lambda$$

$$= (1-\omega)\frac{\beta^3}{x!\,(\beta^2+\alpha\beta+2\alpha)}\int_0^\infty \lambda^x e^{-\lambda(\beta+1)}\,(1+\alpha\lambda+\alpha\lambda^2)\,d\lambda$$

$$= (1-\omega)\frac{\beta^3}{x!\,(\beta^2+\alpha\beta+2\alpha)}\left[\int_0^\infty \lambda^x e^{-\lambda(\beta+1)}\,d\lambda + \alpha\int_0^\infty \lambda^{x+1}e^{-\lambda(\beta+1)}\,d\lambda + \alpha\int_0^\infty \lambda^{x+2}e^{-\lambda(\beta+1)}\,d\lambda\right]$$

$$= (1-\omega)\frac{\beta^3}{x!\,(\beta^2+\alpha\beta+2\alpha)}\left[\frac{x!}{(\beta+1)^{x+1}}+\frac{\alpha(x+1)!}{(\beta+1)^{x+2}}+\frac{\alpha(x+2)!}{(\beta+1)^{x+3}}\right]$$

$$= (1-\omega)\frac{\beta^3}{(\beta^2+\alpha\beta+2\alpha)}\left[\frac{1}{(\beta+1)^{x+1}}+\frac{\alpha(x+1)}{(\beta+1)^{x+2}}+\frac{\alpha(x+2)(x+1)}{(\beta+1)^{x+3}}\right] \tag{7}$$

$$= (1-\omega)\frac{\beta^3}{(\beta^2+\alpha\beta+2\alpha)}\left[\frac{\alpha(x^2+\beta+3)+(\beta+4)\alpha x+(\beta+1)^2}{(\beta+1)^{x+3}}\right]$$

$$\therefore f_2(x) = (1-\omega)\frac{\beta^3}{(\beta^2+\alpha\beta+2\alpha)}\left(\frac{\alpha(x^2+\beta+3)+(\beta+4)\alpha x+(\beta+1)^2}{(\beta+1)^{x+3}}\right); \quad x = 1, 2, \ldots,$$

where $0 < \omega < 1$, $\beta > 0$ and $\alpha \geq 0$. Therefore, the full pmf of ZIGPSD is

$$f(x) = \begin{cases} f_1(x), & if\ x = 0 \\ f_2(x), & if\ x = 1, 2, \ldots \end{cases}$$

That is,

$$f(x) = \begin{cases} \omega + (1-\omega)\dfrac{\beta^3}{\beta^2+\alpha\beta+2\alpha}\left(\dfrac{\alpha(\beta+3)+(\beta+1)^2}{(\beta+1)^3}\right), & if\ x = 0 \\ (1-\omega)\dfrac{\beta^3}{\beta^2+\alpha\beta+2\alpha}\left(\dfrac{\alpha(x^2+\beta+3)+(\beta+4)\alpha x+(\beta+1)^2}{(\beta+1)^{x+3}}\right), & if\ x = 1,2,\ldots \end{cases}$$

Note that when $\omega = 0$ equation (5), reduces to new generalized Poisson-Sujatha (NGPS) distribution developed by Aderoju (2020).
Probability mass plots of ZIGPS distribution for particular values of ω, α and β are given in Figure 1 below.

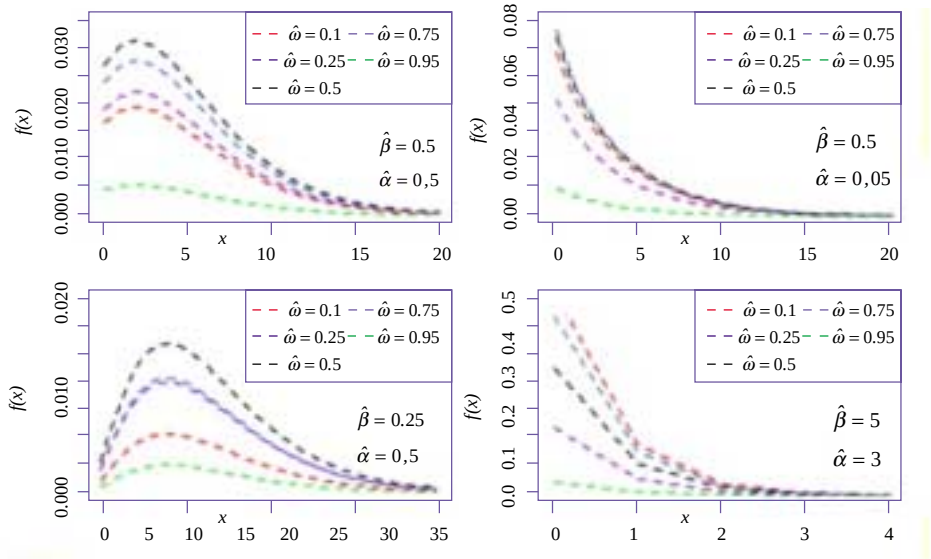


Fig. 1: The pmf of ZIGPS distribution with specified parameters

3. MATHEMATICAL PROPERTIES

The r[th] factorial moment $E[(X)_r]$ of the zero inflated generalized Poisson-Sujatha distribution with random variable X, if is given as:

$$\mu'_r = \frac{\beta^3}{\beta^2+\alpha\beta+2\alpha}\int_0^{\infty}(1-\omega)\sum_{x=1}^{\infty}\left[(x)_r\frac{\lambda^x}{x!}e^{-\lambda}\right](1+\alpha\lambda+\alpha\lambda^2)\,e^{-\beta\lambda}d\lambda$$

$$= \frac{\beta^3}{\beta^2+\alpha\beta+2\alpha}\int_0^{\infty}(1-\omega)\lambda^r\sum_{x=1}^{\infty}\left[(x)_r\frac{\lambda^{x-r}}{x!}e^{-\lambda}\right](1+\alpha\lambda+\alpha\lambda^2)\,e^{-\beta\lambda}d\lambda$$

Note that $(X)_r = X(X-1)(X-2)\cdots(X-r+1)$, therefore,

$$\mu'_r = \frac{\beta^3}{\beta^2+\alpha\beta+2\alpha}\int_0^{\infty}(1-\omega)\lambda^r\sum_{x=1}^{\infty}\left[\frac{\lambda^{x-r}}{(x-r)!}e^{-\lambda}\right](1+\alpha\lambda+\alpha\lambda^2)\,e^{-\beta\lambda}d\lambda$$

Let $z = x-r$,

$$\therefore \mu'_r = \frac{\beta^3}{\beta^2+\alpha\beta+2\alpha}\int_0^{\infty}(1-\omega)\lambda^r\sum_{x=1}^{\infty}\left[\frac{\lambda^z}{z!}e^{-\lambda}\right](1+\alpha\lambda+\alpha\lambda^2)\,e^{-\beta\lambda}d\lambda$$

$$\sum_{x=1}^{\infty}\left[\frac{\lambda^z}{z!}e^{-\lambda}\right] = 1, \quad therefore,$$

$$\mu'_r = \frac{\beta^3}{\beta^2+\alpha\beta+2\alpha}(1-\omega)\int_0^{\infty}\lambda^r(1+\alpha\lambda+\alpha\lambda^2)\,e^{-\beta\lambda}d\lambda$$

$$= (1-\omega)\frac{\beta^3}{(\beta^2+\alpha\beta+2\alpha)}\left[\frac{r!}{\beta^{r+1}}+\frac{\alpha(r+1)!}{\beta^{r+2}}+\frac{\alpha(r+2)!}{\beta^{r+3}}\right]$$

$$= (1-\omega)\frac{\beta^3}{(\beta^2+\alpha\beta+2\alpha)}r!\left[\frac{1}{\beta^{r+1}}+\frac{\alpha(r+1)}{\beta^{r+2}}+\frac{\alpha(r+2)(r+1)}{\beta^{r+3}}\right]$$

$$\therefore \mu'_r = (1-\omega)r!\left[\frac{\beta^2+\alpha(r+1)(\beta+r+2)}{\beta^r(\beta^2+\alpha(\beta+2))}\right], \qquad r = 1,2,\ldots \tag{8}$$

To obtain the corresponding r^{th} factorial moments for the proposed distribution, the values of $r = 1, 2, 3, \ldots$ will be substituted into (8). Below are results of the first four moments of this distribution. When $r = 1, 2, 3, \ldots$ and 4:

$$\mu_1' = (1-\omega)\frac{\beta^2 + 2\alpha(\beta+3)}{\beta(\beta^2 + \alpha(\beta+2))}$$

$$\mu_2' = (1-\omega)\frac{2[\beta^2 + 3\alpha(\beta+4)]}{\beta^2(\beta^2 + \alpha(\beta+2))}$$

$$\mu_3' = (1-\omega)\frac{6[\beta^2 + 4\alpha(\beta+5)]}{\beta^3(\beta^2 + \alpha\beta + 2\alpha)}$$

$$\mu_4' = (1-\omega)\frac{24[\beta^2 + 5\alpha(\beta+6)]}{\beta^4(\beta^2 + \alpha\beta + 2\alpha)}$$

Hence, the variance (σ^2) of the distribution is obtained as:

$$\sigma^2 = \mu_2' - (\mu_1')^2 + \mu_1'$$

$$= (1-\omega)\frac{(\beta^2 + \alpha(\beta+2))\left[2(\beta^2 + 3\alpha(\beta+4)) + \beta(\beta^2 + 2\alpha(\beta+3))\right] - (1-\omega)(\beta^2 + 2\alpha(\beta+3))^2}{\beta^2(\beta^2 + \alpha(2+\beta))^2}$$

The coefficient of variation is given as

$$CV = \frac{\sqrt{(1-\omega)\left[(\beta^2 + \alpha(\beta+2))\left[2(\beta^2 + 3\alpha(\beta+4)) + \beta(\beta^2 + 2\alpha(\beta+3))\right] - (1-\omega)(\beta^2 + 2\alpha(\beta+3))^2\right]}}{(1-\omega)(\beta^2 + 2\alpha(\beta+3))}$$

The probability generating function (PGF) of the distribution can be expressed as follows:

$$\eta_X(t) = \sum_{x=0}^{\infty} t^x f(x)$$

$$= \omega + (1-\omega)\frac{\beta^3}{\beta^2 + \alpha\beta + 2\alpha}\left[\frac{(\beta+1-t)^2 + \alpha(\beta+1-t) + 2\alpha}{(\beta+1-t)^3}\right] \qquad (9)$$

This PGF can also be used to obtain moments. For example, the first momentis given as $\eta_X'(t=1)$, where

$$\eta'_X(t) = \frac{d\eta_X(t)}{dt} = (1-\omega)\left(\frac{1-2t+2\beta+t^2-2t\beta+\beta^2+2\alpha(4-t+\beta)}{(\beta+1-t)^4(\beta^2+\alpha\beta+2\alpha)}\right).$$

Hence, $$\mu = \eta'_X(t=1) = \frac{d\eta_X(t)}{dt}\bigg|_{t=1} = (1-\omega)\frac{\beta^2+2\alpha(\beta+3)}{\beta(\beta^2+\alpha(\beta+2))}$$

as obtained earlier.

4. MAXIMUM LIKELIHOOD ESTIMATES OF THE PARAMETERS

The likelihood function of the ZIGPSD is:

$$\begin{aligned} L(\omega,\alpha,\beta) = \prod_{i=1}^{n}\Bigg[\omega + (1-\omega)\frac{\beta^3}{\beta^2+\alpha\beta+2\alpha}\left(\frac{\alpha(\beta+3)+(\beta^2+2\beta+1)}{(\beta+1)^3}\right)\Bigg] \\ + \Bigg[(1 \\ - \omega)\frac{\beta^3}{\beta^2+\alpha\beta+2\alpha}\left(\frac{\alpha\left(x_i^2+\beta+3\right)+(\beta+4)\alpha x_i+(\beta^2+2\beta+1)}{(\beta+1)^{x_i+3}}\right)\Bigg] \end{aligned} \tag{10}$$

The log-likelihood function of the *ZIGPS*(α, β, ω) can be expressed as follows:

$$\begin{aligned} \mathcal{L} = \sum_{i=1}^{n} log\Bigg[\omega + (1-\omega)\frac{\beta^3}{\beta^2+\alpha\beta+2\alpha}\left(\frac{\alpha(\beta+3)+(\beta^2+2\beta+1)}{(\beta+1)^3}\right)\Bigg] \\ + \sum_{i=1}^{n} log\Bigg[(1 \\ - \omega)\frac{\beta^3}{\beta^2+\alpha\beta+2\alpha}\left(\frac{\alpha\left(x_i^2+\beta+3\right)+(\beta+4)\alpha x_i+(\beta^2+2\beta+1)}{(\beta+1)^{x_i+3}}\right)\Bigg] \end{aligned} \tag{11}$$

The estimates of the parameters in the nonlinear equation (11) can be obtained by numerical optimization using “optim” or “nlm” functions in the R software (R Core Team, 2021).

5. SIMULATION STUDY

In this section, a simulation study to examine the goodness-of-fit performance of the ZIGPS is presented. A maximum likelihood estimation scheme was implemented within the R environment to carryout model fitting for the ZIGPS model. The ZIGPS is compared with the Poisson, ZIP and NGPS distributions. The assessment

is done under various settings of sample sizes and proportion of excess zeros. All simulations were carried out in the R environment.

5.1 SIMULATION SETTING

The simulation settings for the three cases considered are defined as follows.

Case 1: The data are from a Poisson distribution with parameter λ. We considered cases of $\lambda = 5, 50$ at sample sizes of n = 20, 50, 200.

Case 2: The data are from a ZIP distribution with parameters $\lambda = 10$ and $\omega = 0.1$, 0.2, 0.5, 0,8. We considered sample sizes of n = 50, 200.

Case 3: The data are from a ZIGPS distribution with parameters $\alpha = 0.03$, $\beta = 0.8$ and $\omega = 0.1, 0.2, 0.5, 0.8$. We considered sample sizes of n = 50, 200.

5.2 SIMULATION RESULTS

The performance of the methods were evaluated over 100 replications of each case discussed above. The evaluation criteria are: Loglikelihood (Loglik), Akaike information criterion (AIC) and the Bayesian information criterion (BIC). The AIC and BIC are defined as $AIC = -2L + 2p$ and $BIC = -2L + p\,(\log n)$, where L is the loglikelihood, p is the number of parameters to be estimated for the model and n is the number of observations. Tables 1-3 summarizes the means of Loglik, AIC and BIC over 100 replications. It should be noted that higher Loglik and lower AIC and BIC indicate better fit (Hastie, et al., 2001; Adeniyi et al, 2018).

Table 1 presents the results for case 1 where the data are Poisson distributed. The results indicate that at $\lambda = 5$, the performance of Poisson and ZIP are similar across all the sample sizes considered. The performance of the ZIGPS is the least as expected since the data do not contain excess zeros.

The results for case 2, where the data follow the ZIP distribution, are presented in Table 2. Under this setting, the ZIP produced the best performance followed by the ZIGPS across the various levels of excess zeros and sample sizes. The least performance, as expected, was produced by the Poisson distribution as the data were zero-inflated.

Table 3 presents the results for case 3, where the generated data are from the ZIGPS distribution.

Tab. 1: Means of for Loglik, AIC and BIC over 100 replications for Poisson distributed data (CASE 1).

n	λ	Model	Loglik	AIC	BIC
20	***5***	**Poisson**	-43.95	89.90	90.89
		ZIP	-43.81	91.62	93.61
		NGPS	-48.07	100.14	102.13
		ZIGPS	-48.20	102.40	105.39
	50	**Poisson**	-67.28	136.56	137.55
		ZIP	-67.28	138.56	140.55
		NGPS	-87.42	178.84	180.83
		ZIGPS	-92.94	191.89	194.88
50	***5***	**Poisson**	-109.27	220.54	222.45
		ZIP	-109.16	222.32	226.14
		NGPS	-120.31	244.62	248.45
		ZIGPS	-120.36	246.71	252.45
	50	**Poisson**	-167.94	337.87	339.78
		ZIP	-167.94	339.87	343.70
		NGPS	-218.30	440.60	444.43
		ZIGPS	-232.77	471.54	477.28
200	***5***	**Poisson**	-441.32	884.65	887.95
		ZIP	-441.05	886.11	892.70
		NGPS	-482.28	968.57	975.16
		ZIGPS	-482.46	970.91	980.81
	50	**Poisson**	-675.08	1352.16	1355.46
		ZIP	-675.08	1354.16	1360.75
		NGPS	-873.39	1750.77	1757.37
		ZIGPS	-969.52	1945.04	1954.94

Tab. 2: Means of for Loglik, AIC and BIC over 100 replications for Zero-inflated Poisson (ZIP) distributed data (CASE 2).

n	ω	Model	Loglik	AIC	BIC
50	*0.1*	**Poisson**	-160.49	322.98	324.89
		ZIP	-130.54	265.08	268.9
		NGPS	-151.08	306.17	309.99
		ZIGPS	-147.57	301.13	306.87
	0.2	**Poisson**	-188.11	378.22	380.13
		ZIP	-127.21	258.41	262.24
		NGPS	-150.74	305.49	309.31
		ZIGPS	-142.16	290.32	296.06
	0.5	**Poisson**	-234.73	471.46	473.37
		ZIP	-98.03	200.05	203.88
		NGPS	-133.59	271.18	275.01
		ZIGPS	-108.30	222.61	228.34
	0.8	**Poisson**	-182.87	367.74	369.65
		ZIP	-50.23	104.45	108.28
		NGPS	-92.65	189.30	193.13
		ZIGPS	-57.74	121.48	127.22
200	*0.1*	**Poisson**	-649.59	1301.18	1304.47
		ZIP	-524.88	1053.77	1060.36
		NGPS	-605.85	1215.69	1222.29
		ZIGPS	-591.02	1188.03	1197.93
	0.2	**Poisson**	-766.26	1534.52	1537.82
		ZIP	-510.97	1025.93	1032.53
		NGPS	-604.16	1212.32	1218.92
		ZIGPS	-568.17	1142.34	1152.24
	0.5	**Poisson**	-944.54	1891.07	1894.37
		ZIP	-395.70	795.41	802.00
		NGPS	-535.13	1074.26	1080.86
		ZIGPS	-433.96	873.92	883.81
	0.8	**Poisson**	-745.72	1493.45	1496.74
		ZIP	-202.37	408.73	415.33
		NGPS	-376.10	756.20	762.80
		ZIGPS	-235.75	477.5	487.39

Tab. 3: Means of for Loglik, AIC and BIC over 100 replications for Zero Inflated generalized Poisson-Sujatha distribution (ZIGPSD) distributed data (CASE 3).

n	ω	Model	Loglik	AIC	BIC
50	*0.1*	**Poisson**	-91.60	185.20	187.11
		ZIP	-82.07	168.14	171.96
		NGPS	-78.90	161.80	165.62
		ZIGPS	-78.36	162.71	168.45
	0.2	**Poisson**	-90.08	182.16	184.08
		ZIP	-77.23	158.46	162.28
		NGPS	-75.02	154.05	157.87
		ZIGPS	-73.59	153.19	158.92
	0.5	**Poisson**	-73.42	148.83	150.74
		ZIP	-56.27	116.55	120.37
		NGPS	-59.03	122.07	125.89
		ZIGPS	-54.61	115.23	120.96
	0.8	**Poisson**	-43.87	89.74	91.66
		ZIP	-28.90	61.81	65.63
		NGPS	-35.08	74.16	77.98
		ZIGPS	-28.35	62.71	68.44
200	*0.1*	**Poisson**	-373.29	748.57	751.87
		ZIP	-332.34	668.67	675.27
		NGPS	-320.03	644.06	650.66
		ZIGPS	-318.57	643.14	653.04
	0.2	**Poisson**	-358.67	719.34	722.64
		ZIP	-310.75	625.5	632.09
		NGPS	-300.47	604.94	611.53
		ZIGPS	-296.97	599.94	609.84
	0.5	**Poisson**	-294.83	591.66	594.95
		ZIP	-229.11	462.22	468.81
		NGPS	-236.72	477.44	484.04
		ZIGPS	-220.71	447.42	457.32
	0.8	**Poisson**	-171.65	345.3	348.6
		ZIP	-114.04	232.09	238.69
		NGPS	-138.09	280.18	286.78
		ZIGPS	-110.83	227.66	237.55

6. APPLICATIONS TO REAL LIFE DATASETS

To examine the goodness-of-fit of the ZIGPSD in modelling zero inflated as well as overdispersed count data sets, we use two real data sets. The first data referred to the counts of cysts from 111 steroid-treated kidneys (McElduff, et al., 2010; Kumar & Ramachandran, 2019) while the second data were the number of

Mammalian Cytogenetic dosimetry Lesions in Rabbit Lymphoblast induced by *lymphoblast streptonigrin* (NSC-45383), exposure- 60 μg/kg provided by Shanker and Fesshaye (2016). The model evaluation in this section was based on the chi-squared (χ^2) goodness of fit test to compare between observed (O_i) and expected (E_i) values of data as well as the AIC and BIC:

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \quad \sim \quad X^2_{n-p}$$

where p is the number of parameters and is the number of classes after the data have been grouped into a frequency distribution.

The observed mean and variance are 1.486 and 7.07027 respectively, which indicate presence of overdispersion. Moreover, the expected zero counts in the data are 25, that is, $111 \times e^{-1.486} = 25.104$ while there are 65 zeros in the data, which indicate presence of excess zero. With these evidences, it is clear that the Poisson model is not appropriate for this data set.

Going by the rule proposed by Lawal (1980), that expected value can be as small as $\frac{r}{d^{3/2}}$ (where r is the number of expected values less than 3 and $\boldsymbol{d}$ is the degree of freedom under such model) without violating the X^2 assumption. Hence, the minimum expected values required from the ZIGPSD for this dataset 1 will be $\frac{7}{9^{3/2}} = 0.2592$; only 7 expected values are less than 3 and the degrees of freedom *df* = 9 Hence, since there is no expected value less than 0.259, there is no need to collapse cells. However, the minimum expected values from Poisson, ZIP and NGPS distributions are 0.2192, 0.1897 and 0.2214 respectively. Therefore, cells less than the values were collapsed as shown in Table 1 and the degree of freedom adjusted appropriately.

Based on the computed values of X^2 statistic given in Table 4, it is obvious that the ZIGPSD fits the data well (p-value = 0.7675) while ZIP (pvalue < 0.0001) and NGPS (pvalue < 0.0001) distributions failed to fit the data at 5% significant level. It is worthy of note that Kumar & Ramachandran (2019) fitted zero inflated Hermite (ZIH), zero inflated generalized Poisson (ZIGP), Negative Binomial (NB) and zero inflated Negative Binomial (ZINB) distributions among others to this same data set. Their results showed that even ZINB distribution failed woefully at fitting the data, only ZIH distribution slightly fit at 5% level of significance (having pvalue = 0.0914).

Tab. 4: Observed and expected frequencies from fitted models to the two data set on counts of cysts from 111 steroid-treated kidneys

X (count)	Observed Freq.	Poisson	ZIP	NGPS	ZIGPS
0	65	25.1	65.0	45.3	65.0
1	14	37.3	5.1	26.4	13.1
2	10	27.7	8.9	15.6	9.2
3	6	13.8	10.3	9.3	6.6
4	4	$7.1 = \begin{cases} 5.1 \\ 1.5 \\ 0.4 \\ 0.1 \\ 0.0 \\ 0.0 \\ 0.0 \\ 0.0 \\ 0.0 \end{cases}$	8.9	5.6	4.7
5	2		6.2	3.4	3.4
6	2		3.6	2.1	2.5
7	2		1.8	1.3	1.8
8	1		0.8	0.8	1.3
9	1		$0.4 = \begin{cases} 0.3 \\ 0.1 \\ 0.0 \\ 0.0 \end{cases}$	0.5	1.0
10	1			$0.7 = \begin{cases} 0.3 \\ 0.2 \\ 0.2 \end{cases}$	0.7
11	2				0.5
12	1				1.2
Total	111	111	111	111	111
MLE		$\hat{\lambda} = 1.4865$	$\hat{\lambda} = 3.4760$ $\hat{\omega} = 0.5724$	$\hat{\alpha} = 0.0311$ $\hat{\beta} = 0.8074$	$\hat{\alpha} = 0.0263$ $\hat{\beta} = 0.5280$ $\hat{\omega} = 0.4162$
Loglik		-263.2722	-191.866	-184.8609	-172.8206
X^2 ***(P-value)***		104.8565 (<0.0001)	76.6887 (<0.0001)	35.0950 <0.0001	5.720 (**0.7675**)
df		3	**7**	7	9
AIC		528.5444	387.7321	373.7219	**351.6413**
BIC		526.5444	383.7321	374.6917	**359.7699**

Tab. 5: Observed and expected frequencies from fitted models to the data set on number of Mammalian Cytogenetic dosimetry Lesions in Rabbit Lymphoblast induced by streponigrin (NSC-45383), *exposure - 60 µg/kg.*

X (count)	Obs. Freq	Poisson	ZIP	NGPS	ZIGPS
0	413	374.0	413.0	408.6	413.0
1	124	177.4	116.0	130.1	124.1
2	42	42.1	52.1	42.0	42.1
3	15	6.6	15.6	13.7	14.4
4	5	$0.9 = \begin{cases} 0.8 \\ 0.1 \\ 0.0 \end{cases}$	3.5	4.5	4.9
5	0		0.6	1.5	1.7
6	2		0.2	0.6	0.8
Total	601	601	601	601	601
MLE		$\hat{\lambda} = 0.4742$	$\hat{\lambda} = 0.8990$ $\hat{\omega} = 0.4725$	$\hat{\alpha} = 0.1472$ $\hat{\beta} = 2.4154$	$\hat{\alpha} = 0.2514$ $\hat{\beta} = 2.4160$ $\hat{\omega} = 0.0702$
Loglik		-582.6775	-559.5806	-556.4111	**-556.1823**
X^2 ***(P-value)***		72.1766 (<0.0001)	19.9756 (0.0005)	0.5171 (0.2598)	3.5273 (**0.3172**)
df		3	4	4	3
AIC		1167.355	1123.161	1116.822	**1112.365**
BIC		1169.247	1119.161	1116.406	**1112.365**

The observed mean and variance are 0.4742 and 0.7397 respectively, which indicate presence of overdispersion. Moreover, the expected zero counts in the data are 374, that is, $601 \times e^{-0.4742} = 374$ while there are 413 zeros in the data, which indicate excess zeros. With these evidences, it is clear that Poisson model is not appropriate for this second data set.

Similarly, based on the values of X^2 statistic in Table 5, the results shown that ZIGPSD (with pvalue = 0.3172) and NGPSD (pvalue = 0.2598) fits the data well while Poisson (pvalue < 0.0001) and ZIP (pvalue = 0.0005) distributions failed to fit the data at 5% level of significance.

Moreover, the minimum expected values required from the ZIGPSD for this dataset will be $\frac{2}{3^{3/2}} = 0.3849;$ only two expected values are less than 3 and the *df* = (7-3-1) = 3. So, there is no need to collapse cells except for Poisson model - where we collapse cells X = 4, 5 and 6.

7. CONCLUSION

In this paper we proposed the zero inflated Generalized Poisson-Sujatha (ZIGPS) distribution and derived some of its mathematical characteristics. Maximum Likelihood Estimates of the parameters through direct maximization of the log-likelihood function is proposed and implemented numerically using the R software. A simulation study was used to examine the goodness-of-fit performance of the ZIGPS distribution. Results from simulation study revealed that the ZIGPS - distribution is a good alternative for modelling count data with excess zeros.) = Application of the model was made to two real data sets. It has also been shown that the proposed model fits the two data sets well at 5% significant level. Note that the model performs extremely well in the two data sets, which makes it a model to reckon with in modeling zero inflated count data that is highly overdispersed. R codes used in this work are available on request from the authors.

ACKNOWLEDGMENTS

The authors greatly appreciate the anonymous reviewers for their helpful comments and suggestions towards the revision of this paper.

REFERENCES

Adeniyi, I.A., Yahya, W.B., & Ezenweke, C.P. (2018). A Note on Pharmacokinetics Modelling of Theophylline Concentration Data on Patients with Respiratory Diseases. *Turkiye Klinikleri Journal of Biostatistics*, *10*(1), 27-45. doi:10.5336/biostatic.2017-58451.

Aderoju, S.A (2020). A New Generalized Poisson-Sujatha Distribution and its Applications. Applied Mathematical Sciences, 14(5), 229–234. https://doi.org/10.12988/ams.2020.914185

Aryuyuen, S., Bodhisuwan, W. and Supapakorn T., (2014), Zero Inflated negative binomial-generalized exponential distribution and its applications; J. Sci. Technol. 36 (4), 483-491.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer.

Johnson, N. and Kotz, S. (1969). *Discrete distributions, Distributions in statistics*. Houghton Mifflin,

Kumar, C.S and Ramachandran, R. (2019): On some aspects of a zero-inflated overdispersed model and its applications, *Journal of Applied Statistics, DOI: 10.1080/02664763.2019.1645098*

Lawal, H. B. (1980), Tables of percentage points of Pearson's goodness-of-fit statistic for use with small expectations, *Applied Statistics*, **29** (1980), 292–298. https://doi.org/10.2307/2346904

McElduff, F., Cortina-Borja, M., Chan, S.K. and Wade, A. (2010). When t-tests or Wilcoxon-Mann-Whitney tests won't do, Adv. Physiol. Educ. 34 , pp. 128-133

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Shanker R and Fesshaye H (2016). "Size–Biased Poisson–Sujatha Distribution with Applications." *American Journal of Mathematics and Statistics*, **6**(4), 145–154.

Sirichantra C.'and Bodhisuwan W. (2017).Parameter Estimation of the Zero Inflated Negative Binomial Beta Exponential Distribution.

AIP Conference Proceedings 1905, 050041. https://doi.org/10.1063/1.5012260

Tesfay, M and Shanker R (2019). Another Two-Parameter Sujatha Distribution with Properties and Applications. J*ournal of Mathematical Sciences and Modelling, 2* (1), 1-13. DOI: http://dx.doi.org/10.33187/jmsm.416628

Winkelmann, D. R. (2000). *Econometric Analysis of Count Data;* Springer Berlin Heidelberg.

LEARNING ONLINE: REMOTE TEACHING AND UNIVERSITY STUDENTS' ENGAGEMENT

Margherita Russo[1], Fabrizio Alboni, Isabella Morlini

Dipartimento di Economia Marco Biagi, Università di Modena e Reggio Emilia, Italy

Pasquale Pavone

EMBED, Sant'Anna Pisa, Italy

Laura Sartori

Dipartimento di Scienze Politiche e Sociali, Università di Bologna, Italy

Abstract. *The COVID-19 pandemic has had a dramatic impact on many dimensions of living and working conditions, and uncertainties about the developments that we shall still face in the near future. This paper analyses the implications of a forced overnight push to online teaching. Drawing upon an online survey conducted during the 2020 lockdown by the University of Modena and Reggio Emilia, this article describes students' living and studying conditions revealed by a large set of open and closed questions. The survey provides significant information on the students' real off-campus conditions, crucial data for the multidimensional analysis by combining non-parametric multivariate analysis of closed questions with textual analyses. It offers important indications about the most useful tools for inclusive teaching across thematic areas and highlights the main difficulties that emerged during the lockdown. Reflections on advantages and disadvantages, strengths and weaknesses in the innovative learning environment set up overnight are offered at a policy level.*

Keywords: *Covid-19, Online teaching, Student satisfaction, Student engagement.*

INTRODUCTION

After China, in Italy and suddenly all over the world, the health emergency brought about by COVID-19 triggered a radical and rapid change in university life. In March 2020, remote teaching became the rule almost overnight in all Italian universities, with drastic consequences for the length and depth of the organisation of both students and lecturers. Classes, examinations and laboratories were suddenly reorganised by a collective effort that benefited from previous experimentation and

[1] Corresponding author Margherita Russo, email: margherita.russo@unimore.it

innovation in teaching methods. Not only did the technical and remote delivery of all classes lie at the core of the challenge, but also the array of tools and practices concerning emergency teaching (Hodges et al., 2020) that could build on existing practices. Here, the goal is to illustrate the lessons learnt by listening to students, building on a survey tool setup to evaluate students' feelings and socioeconomic living conditions. The online survey conducted by the University of Modena and Reggio Emilia (Unimore) investigated the material and socioeconomic conditions of its students during the lockdown, how they coped with the pandemic and how the remote teaching system affected organisation of their studies (methods and plans). In short, we sought to evaluate the consequences of social and physical distancing – as a practice imposed in order to cope with the pandemic.

The specific open and closed questions formulated in the questionnaire enabled us to investigate the following issues:

- What material and study conditions characterise those students who participated in the survey during the COVID-19 emergency period and what the main issues throughout this period were (Section 3.1).
- What the effects of the online teaching system on the students' plans were (Section 3.2).
- What drivers explain the degree of satisfaction with the online teaching system (Section 3.2).
- How many the typical profiles (and their characteristics) among students are (Section 3.3).
- Which new policy measures and actions can be implemented in order to help students and to increase their level of satisfaction after the emergency period (Section 5).

We have investigated these issues with descriptive analyses as well as by means of supervised and unsupervised multivariate methods. The paper is organised as follows: Section 2 describes the survey and the data. Section 3 reports the main results obtained performing: simple descriptive statistics (Section 3.1), random forest models implemented to explain the effects of the online teaching system (Section 3.2), multivariate analyses sought to identify homogeneous groups of students (Section 3.3), textual analysis on open questions (Section 3.4). Section 4 discusses the free text answers collected through the survey: beyond the emergency condition, it provides a fuller picture of the inequalities that students face during their academic lives while Section 5 concludes with some considerations for future university policies. Annexes present detailed results. The database is available online at http://dx.doi.org/10.25431/11380_1272288.

2. THE SURVEY

The survey covered 27,792 students, categorised by: department, type of study (bachelor's degree, master's degree, 5-year or 6-year degree), status in completing or not completing university examinations in the prescribed time (the status of regular/not regular is a peculiar characteristic of the Italian university system), ECTS credits, gender. The rate of complete answers was 19.2%, we therefore obtained a sample of 5,341 students.

The participation rate greatly varies by area and department, with a higher participation of students enrolled in the Society and Culture syllabus. Considering the type of degree course, the year of enrolment, the credits achieved and the gender, the response percentages in the sample are similar to the percentages of those enrolled students in the population only for some variables (see Table 1). We did not decide to weight the observations.

The online survey was carried out utilising Survey Monkey, with an individual link sent to the students. The questionnaire consisted of 36 questions grouped into four sections: (i) general information on trips home triggered by the emergency, living conditions and ongoing problems and changes; (ii) the organisation of study with respect to the teaching materials available, the timing and methods of the organisation of study; (iii) distance learning, with the focus on attendance and satisfaction, specific difficulties, conditions of concentration and interest, aspects that were missing and those that were appreciated, open questions on strengths and weaknesses of distance learning and suggestions and proposals; (iv) information on internships and working conditions. Annex 1 reports labels and short descriptions of variables used in the analyses.

Tab. 1: Frequency distributions of respondents and not-respondents considering the post-stratification variables

		Enrolled students (n)	*Completed survey (n)*	*Enrolled students (%)*	*Completed survey (%)*	*Response rate (%)*
Total		27792	5341	100	100	19.2
Area	Health	2434	478	8.76	8.95	19.6
	Life	1906	306	6.86	5.73	16.1
	Science	1859	262	6.69	4.91	14.1
	Society and Culture	14199	3310	51.09	61.97	23.3
	Technology	6956	981	25.03	18.37	14.1
	Erasmus	438	4	1.58	0.07	0.9
Current degree course	L - Bachelor's Degree	17757	3287	64.40	61.68	18.5
	LM - Master's Degree	5636	1132	20.44	21.24	20.1
	LM5 - Master's Degree 5 years	2747	597	9.96	11.20	21.7
	LM6 - Master's Degree 5 years	996	309	3.61	5.80	31
	Erasmus	438	4	1.59	0.08	0.9
Year of enrolment	1	7777	1214	27.98	22.73	15.6
	2	5239	1264	18.85	23.67	24.1
	3	7367	1321	26.51	24.73	17.9
	4	3030	677	10.90	12.68	22.3
	5	4170	818	15.00	15.32	19.6
	6	208	47	0.75	0.88	22.6
	N/A	1	0	0.00	0.00	0
Quartiles of archivied credits	Q1	8418	1134	30.29	21.23	13.5
	Q2	6889	1440	24.79	26.96	20.9
	Q3	6645	1525	23.91	28.55	22.9
	Q4	5336	1236	19.20	23.14	23.2
	N/A	504	6	1.81	0.11	1.2
Gender	Female	14379	3511	51.74	65.74	24.4
	Male	12973	1825	46.68	34.17	14.1
	Not Responding	440	5	1.58	0.09	1.1

3. RESULTS

3.1 DESCRIPTIVE ANALYSIS

We first summarise some important features that emerged from the univariate descriptive analyses of the dataset.

The surprise of the lockdown did not greatly affect the need of students to change their accommodation: 73.4% of students were living with their families, a tiny 15% needed or wanted to go back to their hometowns (mainly in distant regions such as Apulia, Basilicata or Sicily), and 11.6% remained in their community, student housing or apartment. Remote teaching had an impact on their studying conditions as did the digital resources they were able to use, such as electronic devices, internet connections and private spaces, to consult and browse class materials. Whilst a slight majority of students (53.4%) had their own private room where they could follow remote teaching and study, almost one fourth had to share the room or the computer with siblings, and more than one tenth could use a workstation for only a tiny fraction of the day (2-3 hours). Although the vast majority of respondents had a personal computer, there were 25% that did not: they asked a friend or relative to borrow their computer or were forced to use a smartphone. Whilst this could undermine the quality of remote teaching, it proved to be a major problem when it came to taking examinations. Overall, the Internet connection was not a problem except for 14% of respondents. The message is clear: the infrastructures that campus facilities offer are crucial for equalising the chances of studying effectively.

How remote teaching affected studying is revealed by the difficulties that emerged regarding self-organisation and the situation at home. A large 40% of respondents reported encountering major difficulties in self-organising their studying at home (listening, taking notes, assimilating content) and 9% (with a strong gender bias) declared that they had to take time away from their studying and instead prioritise domestic chores and responsibilities. Although one fourth declared that they were not in any way affected by the remote teaching, one fifth declared that they studied more hours and felt less prepared, while 15% needed more time to use the recorded classes. The impossibility of studying with colleagues in the library was felt as a barrier and a problem in the personal organisation. As discussed more thoroughly in Section 5, these findings demonstrate that not only material conditions, but also emotional and relational ones are crucial in defining a (more or less effective) study ecosystem.

Overall, the remote teaching system proved to be a positive experience for the majority of the students: 25% gave a score of 8 (on a scale from 1 to 10) whilst 50%

gave a median 6.4 score. Interestingly, students declared that their main difficulties were the following: 70% spent more time browsing and studying compared to the offline teaching experience; 60% felt that the absence of engagement with the lecturer was a problem; 60% had major difficulties in concentrating for personal (lack of motivation, no peer contact) or structural reasons (juggling the multiple teaching tools available, poor home material conditions).

3.2 A STUDY OF THE EFFECTS OF ONLINE EDUCATION ON STUDENTS' STUDY PLANS

The aim of the analysis is to evaluate if variations in students' study plans depend on university career features as well as on students' satisfaction with the online teaching.

We adopted the random forest algorithm (Breiman, 2001) with the target variable d20 "Compared to the plans you had for this semester, indicate where you are with your preparation in the various subjects" and input variables regarding living conditions, ongoing problems and changes, organisation of studies, distance learning (in Annex 1)[2]. The original target variable was a categorical variable with 5 modalities and, in order to prevent imbalanced data problems, two substantially negligible classes were excluded from the analysis (Table 2). The random forest model had, overall, a good predictive performance: the error rate was about 33%.

Tab. 2: Frequency distribution of modalities of variable D20

	Number	*% of respondents*	*% of used answers*
I am keeping abreast of all subjects	1197	22.4	23.5
I'm behind in some subjects	2827	52.9	55.4
I'm behind in all of my subjects	1080	20.2	21.2
I decided not to prepare any exams	23	0.4	
Other (specify)	214	4.0	
Sum	5341	100.0	100.0

When using a random forest algorithm, it is possible to rank the predictor variables according to their importance. Different indices can be used to determine the importance of the explicative variables in explaining the target variable. More specifically, we used a multi-way importance plot that combines some of these indices.

[2] Open questions of those sections are not included in this analysis.

In the case of qualitative variables, such as question d20, the plot uses: i) Mean Decrease Accuracy, which expresses how much accuracy the model loses by excluding each variable; ii) Mean Decrease Gini, which is a measure of how each variable contributes to the homogeneity of the nodes and leaves in the resulting random forest; iii) Mean Minimal Depth, based on the minimum depth of a variable in a tree, i.e. the depth of the node that splits on that variable and is closest to the root of the tree.

The higher the value of Mean Decrease Accuracy or Mean Decrease Gini score is, the higher the importance of the variable in the model is, and, in a random forest, if the Mean Minimal Depth is low then many observations are divided into groups based on this variable. Figure 1 (multi-way importance plot) shows the most important variables, whose impact can be analysed with the partial dependence plots (Figure 2).

The partial dependence plot (PDP) shows the marginal effect that one or two features have on the predicted outcome of a machine learning model (Friedman, 2001). The PDP makes it possible to highlight whether there is a dependency between the target variable and a set of input features of interest and to show whether the relationship between the target and a feature is linear, monotonic or more complex. The output is a plot, in which we have in abscissa the values of the input feature and in ordinate, in the case of classification problems (as in the case of question d20), the probability of belonging to a class of the target variable, while, in the case of a continuous target variable (d23), the value estimated by the model at the different values of the explanatory variable.

We observe the following impacts on the target variable d20 (student's performance in preparing the various subjects):

– being able to organise (d25_5: I can't organise my daily study activities effectively) impacts on the performance;
– as expected, proceeding equally with all the subjects (*d21_1: I am progressing equally on all the subjects taught in this semester*) positively impacts on keeping abreast of all subjects;
– with respect to distance learning (*d22: Are you following the distance learning activities provided by your study programme?*), there is a direct relationship between the amount of distance learning and keeping abreast of all subjects; there is an inverse relationship with those who are behind all, while the relationship is not monotonous with those who are behind by some subjects;
– lessons accumulation (*d24_1: The accumulation of lessons is creating difficulties for me*) has a direct relationship to those who are behind on everything, and obviously not for those who are keeping abreast of all subjects;

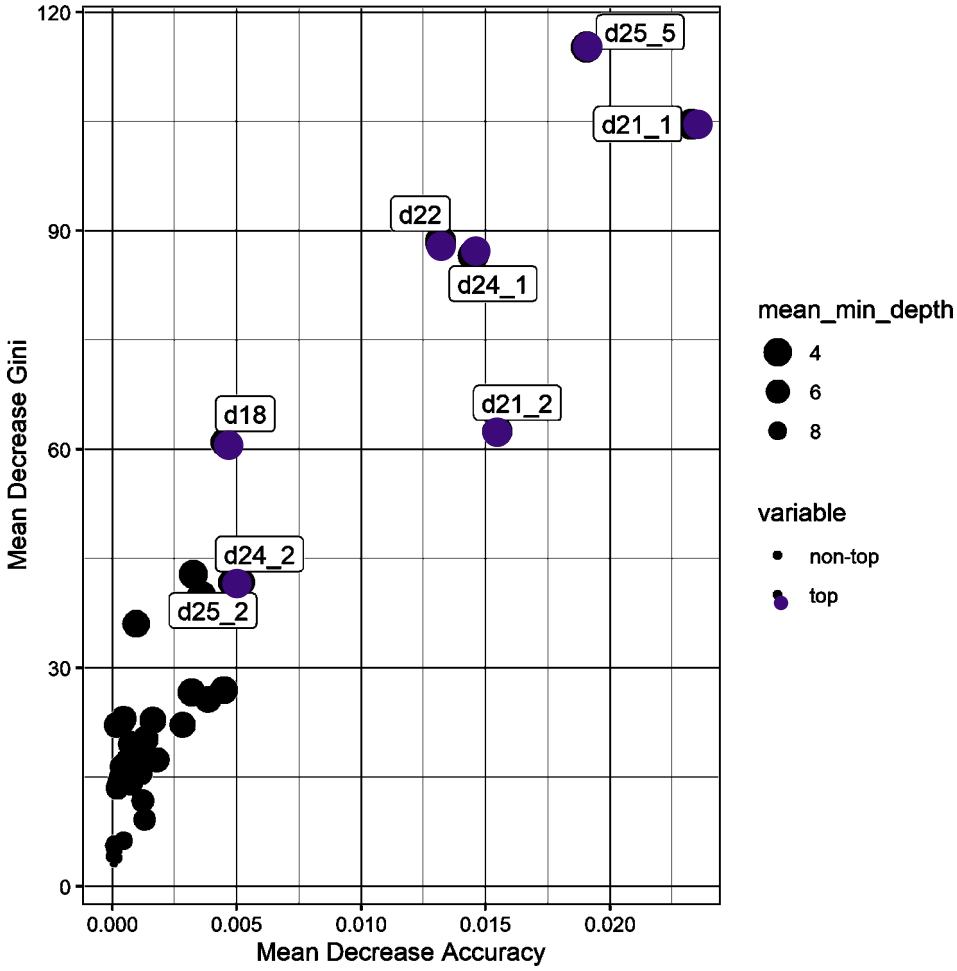


Fig. 1: Multi-way importance plot [d20]

Legend: top variables (blue) are those in the highest eight positions of the new variable importance that is obtained by combining the three criteria, i.e. the mean decrease accuracy, the mean decrease Gini and the mean minimal depth (mean_min_depth); non-top variables (black) are all remaining variables. In the mean minimal depth, the size of dots indicates the mean of the minimal depths (1, 2, …) in all trees. The lowest values are the closest to the root of the tree and, therefore, if the size of a variable is low, then many observations are divided into groups based on this variable.

– as expected, those who have prioritised some subjects (*d21_2: I have given priority to certain subjects taught in this semester*) are behind on some subjects;
– how students rearranged the study (*d18: How did you rearrange your studying times?*) seems to have an effect on falling behind in some subjects, while not seeming to affect keeping abreast or falling behind on all subjects.

In general, from the partial dependence plots (Figure 2) we observe that the most important input variables listed above mostly influence students that are behind in some subjects compared with the plans they had for the semester[3].

The analysis of the aspects influencing the level of satisfaction with the online teaching system can be considered a first step in understanding how to improve the online experience. For this further analysis, we consider variable d23 as the target

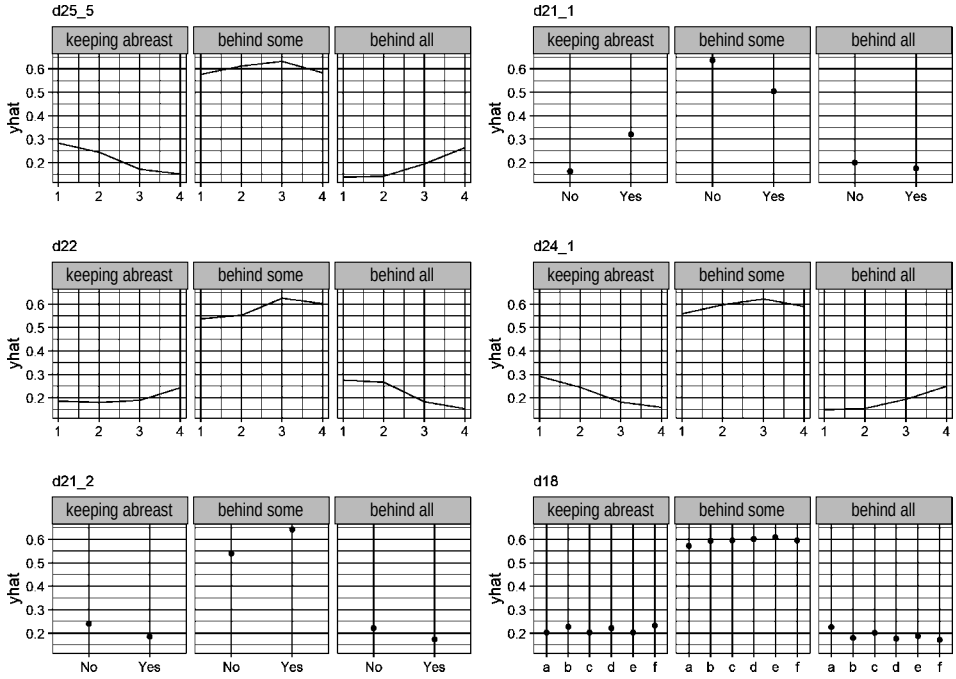


Fig. 2: Partial dependence plot [d20 vsd25_5, d21_1, d22, d24_1, d21_2, d18]

Legend: variables d25_5, d22, d24_1 are in a Likert scale (1-4) with the following modalities: 1 Not at all | 2 Barely | 3 Enough | 4 Very much
d21_1 and d21_2 are dichotomy variable (yes/no)
Variable d18 has the following modalities:

a Other (specify)
b I am studying as I used to, in terms of hours
c I study less because online lessons require less effort
d I study more because online classes require more effort
e I study more, but my level of preparation seems to be lower
f I study more, because I don't have to leave

[3] The central category ("behind some") is the most frequent in the dataset, so there is a higher probability of belonging to this class, however the objective of the PDPs is to visualise the effect induced by the changes in explanatory variables (d25_5, d21_1, d22, d24_1, d21_2, d18) on the target variable d20.

variable ("How satisfied are you with your global distance learning experience?", score range 1-10). The regression random forest provides a quite good result in terms of adaptability, with the explained variance equal to 50.24% (the mean of squared residuals is 2.18).

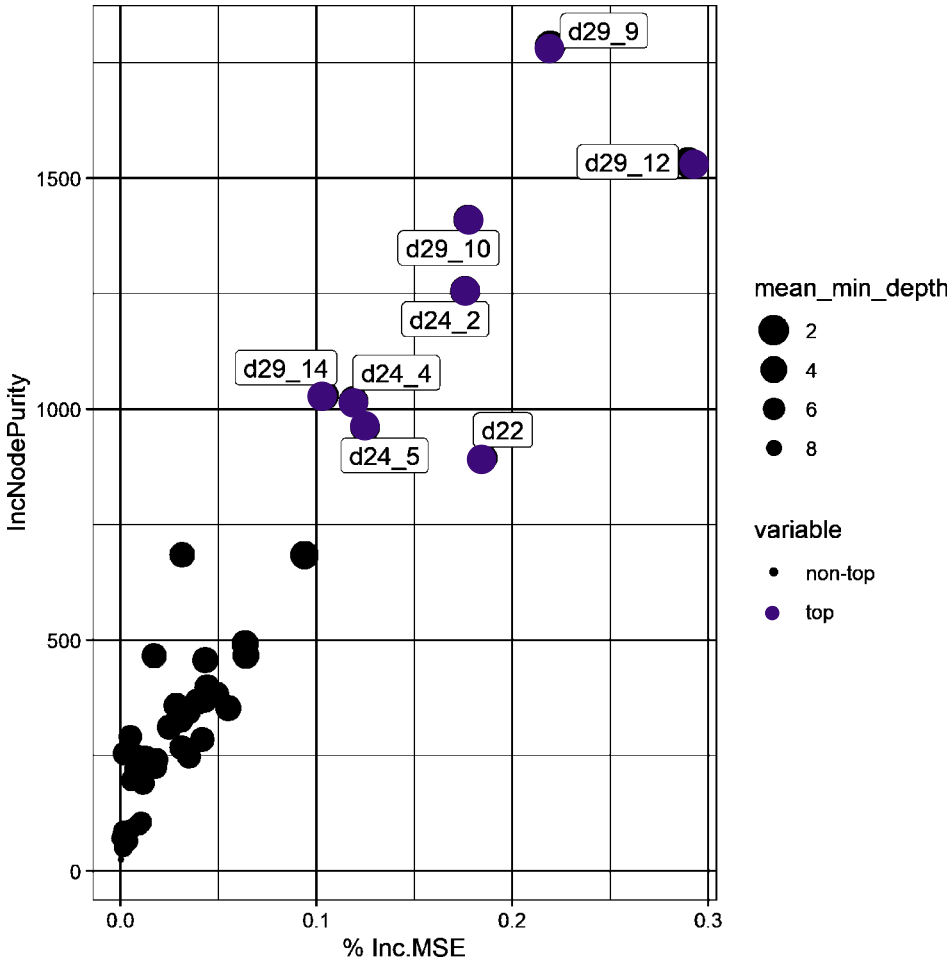


Fig. 3: Variable importance and multi-way importance plot of the dimensions affecting how satisfied students were with their distance learning experience [d23]

Legend: top variables (blue) are those in the highest eight positions of the new variable importance that is obtained by combining the three criteria, i.e. the mean increase of mean squared error (% Inc.MSE), mean node purity increase (incNodePurity) and the mean minimal depth (mean_min_depth); non-top variables (black) are all remaining variables. In the mean minimal depth, the size of dots indicates the mean of the minimal depths (1, 2, …) in all trees. The lowest values are the closest to the root of the tree and, therefore, if the size of a variable is low, then many observations are divided into groups based on this variable.

The different methods utilised for assessing variable importance provide different rankings that can be summarised through a multi-way importance plot (Figure 3). In this case, as the target variable is continuous, the multi-way importance plot was built using: i) Percentage increase in mean square error (%IncMSE), analogous to accuracy-based importance, and is calculated by shuffling the values of the out-of-bag samples, ii) Increase in node purity (incNodePurity), analogous to Gini-based importance, and is calculated based on the reduction in sum of squared errors whenever a variable is chosen to split.

Figure 4 analyses in detail the relationships that link the most important variables (highlighted in Figure 3) to the target variable. From the partial dependence plots we observe substantially linear relationships. The results show that satisfaction

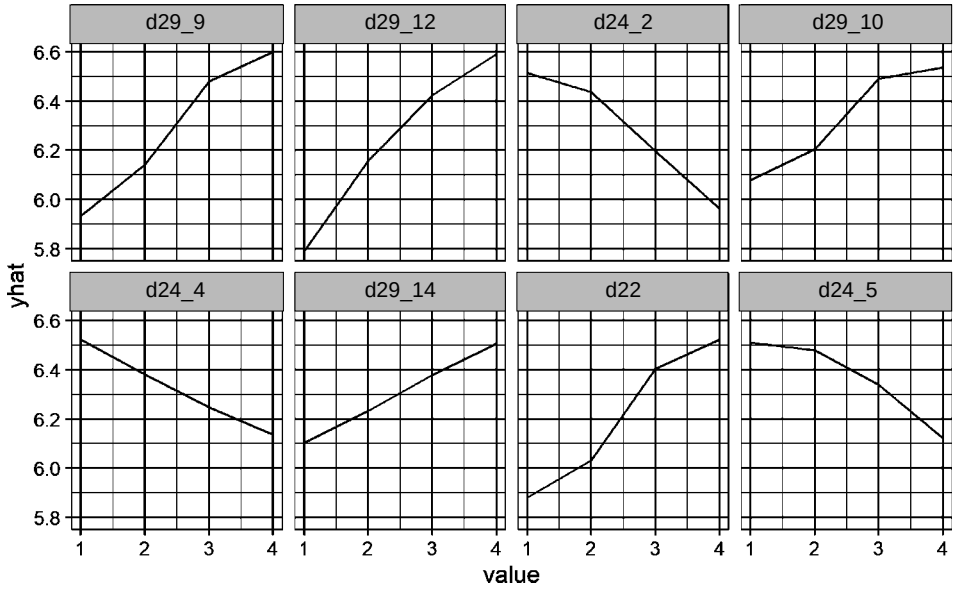


Fig. 4: Partial dependence plot of the top important dimensions affecting how satisfied students were with their distance learning experience [d23]

Legend: all variables are in a Likert scale (1-4) with the following modalities: 1 Not at all | 2 Barely | 3 Enough | 4 Very much

d29_9 The commitment of teachers in teaching and in keeping us updated

d29_12 The readiness and the efforts poured by the University into setting up distance learning, as well as ad hoc initiatives like this one

d24_2 I don't know how to extricate myself among the different recordings

d29_10 The prompt uploading of the recorded lessons

d24_4 I find it difficult to follow the lessons and therefore to take good notes

d29_14 The awareness of having all the resources of the course available at all times and everywhere

d22 Are you following the distance learning activities provided by your study programme?

d24_5 the absence of involvement during the lesson makes it difficult for me to stay focused

increases concomitantly with appreciation for the commitment of lecturers (d29_9 and 29_10), the readiness and efforts made by the university to organise distance learning (d29_12) and to provide all the resources necessary for studying (d29_14), and the share of attendance on distance learning activities (d22). Conversely, satisfaction decreases with respect to three variables: the difficulty in knowing how to extricate oneself between the different recordings increases (d24_2), the difficulty in following the lessons and therefore taking good notes (d24_4) and the absence of involvement during the lesson that makes it difficult for students to stay focused (d24_5).

3.3 DEFINING HOMOGENEOUS GROUPS OF STUDENTS AND TYPICAL PROFILES

The multivariate analysis of the closed questions intends to classify students by considering the combinations of all material conditions and choices related to studies and life organisation. The 48 variables under analysis belong to three main groups: Ongoing problems and changes; Organisation of study with respect to the teaching materials available, the timing and methods of the organisation of study; Distance learning[4]. We applied the Partitioning Around Medoids algorithm (PAM) (Schubert and Rousseeuw, 2019) with the Gower's similarity index. A feature of this algorithm is the independence from the data order unless some of the distances among objects are tied (Kaufman and Rousseeuw, 2005, p. 104), as was the case in our dataset[5]. In order to evaluate the optimal number of groups to be used in PAM, we analysed the silhouette coefficients for an increasing number of partitions (from 2 to 15) (Hennig and Liao, 2013; Rousseeuw, 1987). We choose the 8-cluster partition because all groups have a similar relative importance in terms of frequencies[6].

---

[4] The variables under analysis are: Ongoing problems and changes (d15_7, d15_11), Organisation of study with respect to the teaching materials available, the timing and methods of the organisation of study (d17, d18, d19, d20, d21_1, d21_2, d21_4), Distance learning (d22, d23, d24_1, d24_2, d24_3, d24_4, d24_5, d24_6, d24_7, d24_8, d25_1, d25_2, d25_3, d25_4, d25_5, d25_6, d26, d27, d28_1, d28_2, d28_3, d28_4, d28_5, d28_6, d28_7, d29_1, d29_2, d29_3, d29_4, d29_5, d29_6, d29_7, d29_8, d29_9, d29_10, d29_11, d29_12, d29_13, d29_14). Detailed description of each variable is available in Annex 1. The selection of the 48 variables under analysis excludes all the variables (belonging to the three main groups) that show not significant variability in the dataset. Detailed data on variability can be obtained upon request to the authors.

[5] We see this feature in the similarity matrix. As expected, with a very large number of units and categorical and discrete variables (with few values), many some pairwise similarities in the matrix are equals.

[6] Detailed results can be obtained upon request to the authors.

In what follows, the groups of students will be described with respect to the partitioning variables and to the information on each student's status of her/his university career (Area of studies, Degree course, Year of enrolment, Quartiles of achieved credits) and Gender.[7]

cl-cq1: The Satisfied: *students fully satisfied with the online teaching*. This cluster contains about 8% of the students, mainly female, attending the 3rd year of a 5 or 6-year course, who belong to departments in the area of Life, Health and Science. These students have generally acquired a number of credits far below the median level of the credits acquired by students attending the same course and the same year. They were enthusiastic about online teaching and did not have any problem with e-learning or interpersonal communications. They owned all the necessary information technology tools and stated that they studied more than ever, now that they had more free time and they did not have to spend time in commuting. They did not see any negative aspects as regards relationships with lecturers or classmates and in the organisation of study. Moreover, online teaching did not influence their concentration and their ability to focus.

cl-cq2: The Diligent-but-Detached *students who were compliant but disapproving.* This cluster contains about 14% of the students, mostly female, mainly belonging to departments in the Society and Culture area, who were attending the 4th year and with a number of credits acquired below or equal to the median. They attended online classrooms (only on some topics) and had a good opinion of the online lessons. They were able to deal with the new teaching system, but they were disappointed with some aspects of it.

cl-cq3 The Hostile: s*tudents against the online teaching*. This cluster contains about 11% of the students, mostly males enrolled on a master's degree course in technology with a number of credits acquired far above the median. The main drawback was the absence of interaction between students and lecturers and among students, probably because the area of study requires laboratory experience that cannot be replaced by online teaching. They attended lectures, studied all subjects with more dedication than in the pre-COVID period. Yet, their knowledge and skills were deemed not satisfying.

---

[7] Annex 2 presents data on the composition of each cluster according to the university status and gender. Annex 3 displays the bi-plots of the simple correspondence analysis; Annex 4 displays the boxplots of the students' satisfaction with the distance learning experience (*d_23 How satisfied are you with your global distance learning experience?* graded from 1 to 10), by cluster.

cl-cq4: The Confused: *the overall baffled students*. This cluster contains about 15% of the students, mainly female, belonging to the Science, Health and Life scientific areas and attending the 3rd year of a 5- or 6-years course. In general, these students had a number of acquired credits far below the median level. Even though they appreciated the online teaching system, they did not attend online lessons. They had difficulties with studying at home because they were not able to concentrate. However, they did not miss the organisation of teaching before the COVID period and appreciated some aspects of the online system, such as more free time and the absence of commuting to attend lessons.

cl-cq5: The Enthusiastic: *the passionate and eager students*. This cluster contains about 11% of the students, mainly male, belonging to the Technology and Science areas and attending a master's degree course. They had acquired a number of credits above the median level. They did not have any particular problems, attended lessons online and appreciated online teaching. They studied more than in the pre-COVID period because they had more free time. They were able to concentrate and were able to schedule tasks. The online teaching system gave them incentives to study and attend lessons.

cl-cq6: The Hard-working: *the committed but dissatisfied student*s This cluster contains about 12% of the students, mainly male, belonging to the Technology area and with a number of credits acquired above the median level. They were dissatisfied with the online teaching system because even though they attended lessons and study, they perceived a lower level of knowledge and competence. They thought that online teaching was negative in every respect, but particularly with regard to the interaction with lecturers and between students.

cl-cq7: The Lost: *students who were puzzled*. This cluster contains about 18% of the students, mainly male, belonging to the Society and Culture scientific area, attending the 2nd year and with a number of credits acquired both above and below the median level. They gave a 'pass grade' to online teaching and attended some of the lessons. However, they perceived an inferior level of knowledge and competence and they were not able to keep up with their studies. They did not regret the absence of interaction with lecturers but they missed interactions with other students.

cl-cq8: The Undecided: *students who were uncertain and might drop out*. This cluster contains about 11% of the students, mainly female attending the 1st year in the Society and Culture area and with a number of credits acquired below the median level. They expressed a negative opinion of online

teaching and did not attend lessons. They were not able to keep up with exams, but they stated that the organisation of everyday tasks had not changed; nor had the interactions with students and lecturers been modified in the COVID period. They gave negative scores to all aspects of online teaching, especially about the impossibility of concentrating.

3.4 MULTIVARIATE ANALYSIS ON OPEN-ANSWER QUESTIONS

The non-compulsory open-answer questions are dealt with text mining strategies, according to the specific characteristics of the different sets of answers. Firstly, four corpora were created to organise the free texts answering on *Ongoing changes* (d16)[8], on the *Strengths* (d30)[9] andn*Weaknesses* (d31)[10] of distance learning, and on *Suggestions and proposals* (d32)[11]. Annex 5 reports descriptive statistics of corpora (Fergadiotis *et al*., 2015) produced after their tokenisation (acquisition of text by numerical word indexing)[12]. The corpora are very different in terms of respondent rate and lexical variety. In particular, corpora on *Strengths* and *Weaknesses* of distance learning, and on *Suggestions and proposals* refer to fewer respondents than those engaged in describing the *Ongoing changes*, and a smaller average response length. In the case of the corpus *Ongoing changes*, the significant response rate allows a classification of students based on the types of changes they highlight. In the case of the other three smaller corpora, the analysis is limited to highlighting the themes addressed in each corpus, providing a focused view on the students' perspective over their distance learning experience and its potential for the academic organisation.

In each corpus, the text was lemmatised and all content terms, nouns and adjectives, with at least five occurrences, were selected as keywords. In the case of

[8] Question d16: "Describe how and in what way (positive and negative) the Covid-19 emergency has changed your life in general, apart from your studying activities".

[9] Question d30: "Report up to a maximum of 3 strong points of your distance learning experience".

[10] Question d31: "Report up to a maximum of 3 weak points of your distance learning experience".

[11] Question d32: "Please provide your suggestions and proposals".

[12] In particular, for each corpus, Annex 5 describes: the number of Types, i.e. the number of different words present in the corpus; the number of tokens, i.e. the number of occurrences of the types; the average text length of the answers; the type/token ratio, which provides an assessment of the lexical diversity of the corpus; the number of hapax forms, i.e. the forms that appear only once in the corpus, and their weight in terms of occurrences. Types/token ratio measures the lexical diversity in a corpus based on the relationship between the number of tokens and the number of types. A high index value indicates greater linguistic diversity and thus less repetition of the same words in the corpus. This ratio is highly dependent on the text's length, making it difficult to accurately compare corpora of different sizes.

the corpus *Ongoing changes*, this selection was used to define a Vector Space Model to be analysed with factorial and cluster analysis techniques using SPAD[13] software. In the case of the remaining three corpora, the lemmatised texts were analysed using the Iramuteq[14] software.
In what follows, Section 3.4.1 and 3.4.2 report the analysis carried out on the corpus *Ongoing changes*, Section 3.4.3 illustrates the results obtained on the three other smaller corpora.

3.4.1 CORPUS ONGOING CHANGES – STUDENTS' CLASSIFICATION

The size of the *Ongoing changes* corpus, with 5,607 responses[15] and 234,677 total occurrences, supports an analysis aimed at identifying different groups of students characterised by different needs and different request of changes in online teaching. A textual analysis was applied to the corpus to select the terms of content (nouns and adjectives) and to define a Vector Space Model in which each student response is formalized as a vector in a p-dimensional vector space, spanned by the content terms selected in the vocabulary (Misuraca and Spano, 2020). Then the resultant matrix *Students* $\times$ *Lexicon* [4562 $\times$ 1687] was subjected to correspondence analysis and cluster analysis in order to group the students according to similarities in terms of lexicon expressed in responses (Lebart, Morineau and Piron, 2004; Lebart, Salem and Berry, 1998; Lebart, Morineau and Warwick, 1984; Reynolds *et al.*, 2006). Observations under analysis are 4,562, out of 5,607, since we consider only students having at least five different keywords in their response.

Analysing the dendrogram (see Figure 5, upper panel), we select the partition in five clusters. Figure 5 (lower panel) reports the factorial map (*f1f2*) in which clusters of respondents are highlighted in the different convex hulls, and cluster' centroid has a size proportional to the cluster's instances.

In the factorial map, we observe a polarisation, along the first factor, between clusters cl-txt1 and cl-txt2 on the left, and clusters cl-txt4 and cl-txt5 on the right; while the second factor polarises clusters cl-txt2 and cl-txt5 (bottom) versus cluster cl-txt4 (top). In order to give meaning to such polarisation we need to interpret the themes encompassed in each cluster. Such an interpretation can be done by reading the characteristic dictionaries of each cluster, with terms ranked by their test

---

[13] SPAD is a software dedicated to Data Mining and Predictive analysis that provides a totally graphical and intuitive interface. https://ia-data-analytics.com/data-mining-software

[14] Iramuteq is freely available at http://www.iramuteq.org/

[15] In the analysis of open-answer questions we have used all the available answers, even though the questionnaire was not fully completed.

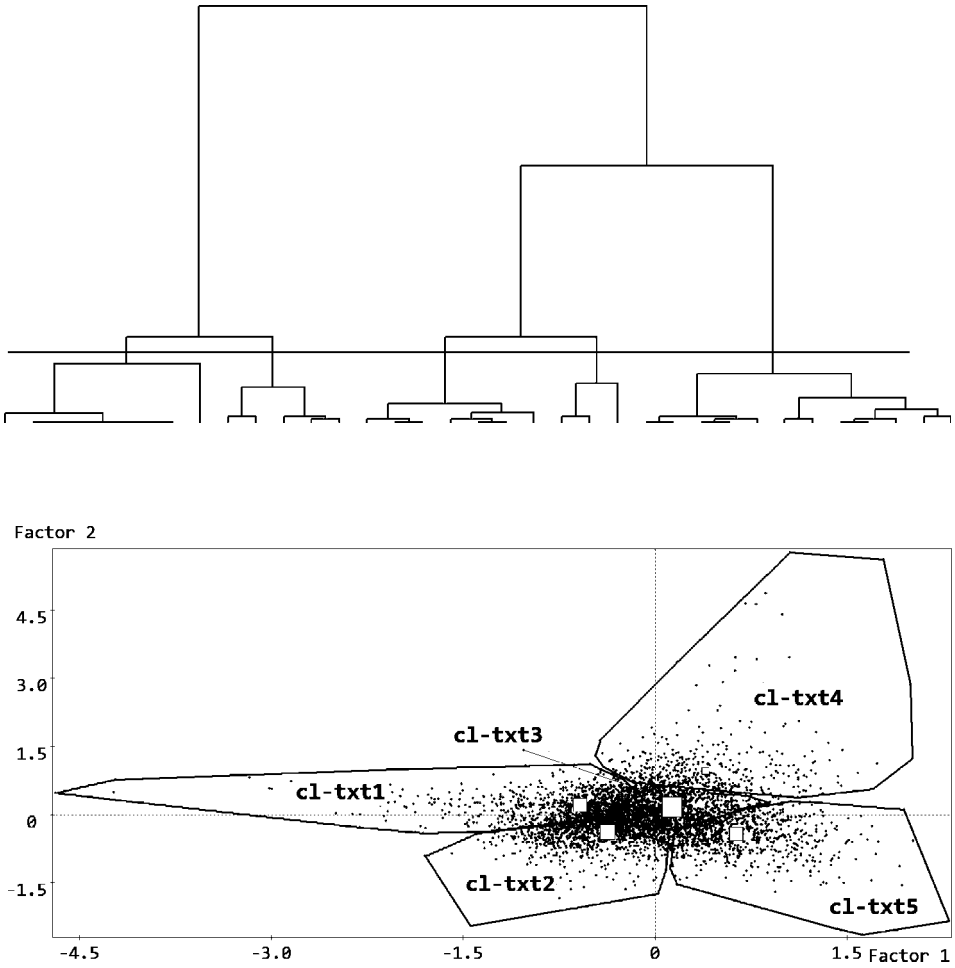


Fig. 5: *Ongoing changes* corpus: cluster analysis dendrogram and factorial map

Legend: dots in the factorial map (lower panel) indicate each student's answer, convex hulls delimitate each cluster and squares indicate cluster centroids, with size proportional to the number of instances in the cluster.

values[16]. The following themes characterise the five clusters of students:

cl-txt1 Students with changes in their personal sphere and in relationships with family members and friends. This cluster contains about 22% of the respondents.

[16] For each cluster, the characteristic terms are available in Annex 6.

cl-txt2 Students who evaluate negative and positive aspects of the new context. This cluster contains about 25% of the respondents.

cl-txt4 Students with problems in their economic and in social and material conditions. This cluster contains about 10% of the respondents.

cl-txt5 Students focusing on the positive and negative novelties of online teaching. This cluster contains about 18% of the respondents.

Although question d16 was asking students to consider changes "aside from their studying activities", two clusters of students hardly separated the issues: in cl-txt3 they exhibit themes intertwining a mix of family life and study conditions, and students in cl-txt5 focus on the big change in their study conditions, i.e., online teaching.

Thanks to the thematic definition of each group of students it is now easy to interpret the latent meaning explained by each factor. The first factor is a composite indicator of students' considerations about the relational sphere and the living and studying conditions. On the left side of the first factor, there are themes focusing on relationships with friends and family members (characterising cl-txt1), and with the external context (characterising cl-txt2); on the right, the focus is on material conditions (cl-txt4) and studying conditions (cl-txt5). The second factor highlights a polarisation of students addressing themes related with the private sphere (top) versus the public sphere (at the bottom). At the centre, students in cluster cl-txt3 address mixed themes concerning the emergency in their personal and public dimensions.

The aim of the question d16 posed to the students was to identify which areas of the students' lives (other than studying) were experiencing the greatest changes. The result obtained through the cluster analysis shows that for some areas (i.e., new context created and online teaching) the effects of the changes are assessed as positive for some students while they are assessed as negative for others. This result provides a more focused hint at the interpretation of changes driven by COVID-19 emergence than can be obtained through a sentiment analysis (Misuraca et al., 2021).

3.4.2 CORPUS ONGOING CHANGES: CORRESPONDENCE ANALYSIS ON STUDENTS' PROFILES ACCORDING TO CLOSED VS. OPEN QUESTIONS

By means of a correspondence analysis, we intersected the classification of students obtained by closed-answer questions (Section 3.3) with the classification obtained from the open questions comprised in the *Ongoing changes* corpus (Section 3.4.1). Observing the biplot of the analysis in Figure 6, we note that students who focused on positive and negative novelties of online teaching (cl-txt5) are strictly associated

with both students who are satisfied with the online teaching system (cl-cq1) and students who are diligent but dissatisfied with it (cl-cq6). Moreover, we see that students with changes in the personal sphere (cl-txt1 and cl-txt2) are, in general, diligent but hostile or indifferent to online teaching (cl-cq2 and cl-cq3). Students with economic and material problems (cl-txt3) are, in general, undecided female students (cl-cq8). Finally, we may note that groups of students who are diligent and enthusiastic about online teaching (cl-cq5) and of confused and unhappy female students (cl-cq4) who are in a halfway position between the personal sphere and the organisation of study and of family life. Groups of "lost" students (cl-cq7) and students with material problems (cl-txt4) are in the centre of the axes.

3.4.3 STRENGTHS, WEAKNESSES, SUGGESTIONS AND PROPOSALS CORPORA

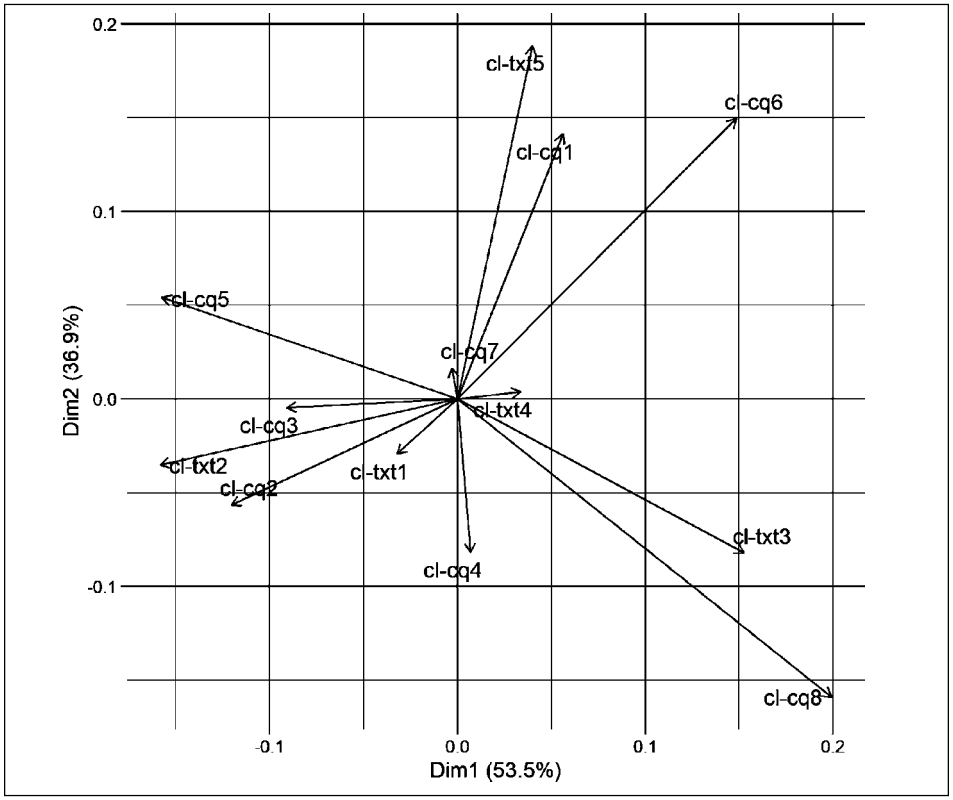


Fig. 6: Biplot of simple correspondence analysis: 8 groups classification with closed-answer questions (cl-cq) and 5 groups classification on the open question ongoing changes (cl-txt)

The other three smaller corpora - S*trengths*, *Weaknesses*, S*uggestions and Proposals* - were analysed in order to characterise the different themes. For these corpora, the ALCESTE[17] method (Reinert, 1983 and 1990; Marpsat, 2010) was used. In general, the characteristics of the sample of respondents to open questions were similar to the characteristics of the sample of respondents to closed ones, with respect to the traits of the population. Females attending a course in the Society and Culture area were over-represented, while students with a number of acquired credits below the first quartile of the credits acquired by students attending the same course and the same year were under-represented. Freshmen as well, were under-represented. For some students, the over-representation was more evident in open answers than in closed ones and this indicates the greater inclination of these students to supply more detailed answers rather than a limited set of closed answers. These students were those with a number of credits below the third quartile of the credits acquired by students in the same year and in the same course and attending the second year of a bachelor degree and the first year of a master's degree.

Unlike the previous cluster analysis, with the ALCESTE method, implemented by Iramuteq, we obtained classifications of the textual segments[18] rather than classifications of the students. Analysing the matrix *Segments* $\times$ *Active forms* of each corpus, the ALCESTE method enabled us to identify 2, 5 and 4 thematic clusters, respectively, in the corpora S*trengths*, *Weaknesses* and *Suggestions and Proposals*[19]. For each of the corpora analysed, the themes around which the answers were focused have been identified. Iramuteq groups into classes the text segments that are similar in terms of content and provides a set of graphic tools with which to visualise the results: dendrogram of the text segment classes, the factorial map, the graph representing the links between the terms in each class. The network graphs (available Annex 8) provide a better focus on the contents of each class[20].

Themes characterising the classes of the three corpora have been synthesised in Tables 3-5.

With regard to the corpus *Strengths* (Table 3), the two classes of themes underscore the fact that the strengths of online teaching mainly refer to flexibility

---

[17] ALCESTE is the acronym of Analyse des Lexèmes Cooccurrents dans les Enoncés Simples d'un Texte. A summary presentation of this method is available in Marpsat (2010).

[18] Segmentation is obtained through fragmentation into phrases. Iramuteq defines phrases by using mainly punctuation and defines as a segment a sequence with a maximum of 40 words.

[19] Annex 7 reports the descriptive statistics of these corpora.

[20] Statistics on the main characteristic terms and graphic visualisation of dendrograms and factorial maps can be obtained from the authors upon request. In addition to visual tools, the disambiguation of terms in the characteristic dictionaries has been verified through the analysis of concordances, available in Iramuteq. A selection of texts for the main characteristic terms in each class is available in Annex 9

in attending lessons (Class 1, especially as concerns Engineering students) and to interactions with peers and lecturers (Class 2, as well as those female students who work and study, attending a bachelor's degree course in the departments of Human Science and Cultural studies and Education). On the contrary, from the *Weaknesses* corpus (five classes listed in Table 4), the lack of interactions and difficulties in attending online lessons were considered one of the weak points in the online teaching, particularly as regards the Engineering students. The other classes of weak points refer to: difficulties in attending online lessons, lack of concentration; the incapability of lecturers to organise courses online (lessons, loaded handouts and organisation of the course internet page); the limits of the technology and the uncertainty related to exams, stages and other teaching activities.

Tab. 3: S*trengths*

Class 1	Flexibility in listening to the recorded lessons and in the organisation of study and less stress (73.33% of the analysed segments). Students in this class were mainly male, attending the first or the second year of a bachelor 's degree course in the Science and Technology areas (particularly, Engineering)
Class 2	More interactions with students and lecturers, positive outcomes of the University for working students, fewer difficulties and speed in the organisation of the emergency period (23.67% of the analysed segments). Students in this class were mainly female (who work and study) attending a bachelor's degree course in the departments of Human Science and Cultural Studies and Education.

Tab. 4: Weaknesses

Class 1	Lack of interactions/discussions with peers and lecturers, absence of library services, less stimuli and engagement (24.07% of the analysed segments). Students in this class were on the whole, not working and attending a course in the Technology area (particularly, Engineering).
Class 2	Difficulties in attending online lessons, lack of concentration (22.81% of the analysed segments). Students in this class were mainly female, in the first or second year of a bachelor's degree course in the Science area (especially, Economics).
Class 3	Ineffective online lessons and lecturers incapable of supplying services online; too many recorded and loaded materials and difficulties in printing slides and handouts (16.84% of the analysed segments). Students in this class were for the most part in their second year of a bachelor's degree course in the Society and Culture area.
Class 4	Limitation of the technology: problems with internet connections and low quality of the sound and the videos loaded (12.03% of the analysed segments). Students in this class were for the most part, attending a course in the Education and Human Sciences department.
Class 5	Uncertainty about examinations, activities in the labs or about internship programs (24.25% of the analysed segments). Students in this class were mainly attending a course in the Education and Human Sciences and Health area.

Observing the characteristics of the different classes of strengths and weaknesses, we may note that it is impossible to identify a single type of beneficiary of the online teaching system. Students may differ in perceiving the strong and weak points of the emergency period in relation to their scientific area, to their family circumstances, and to their employed or not employed status. The analysis confirms that the emergency period reduces the opportunities and exacerbates problems for students with family or economic difficulties.

Furthermore, we can see that working students in the corpus *Suggestions and* P*roposals*, recommended maintaining the online teaching platform in a post-lockdown period (Class 1). Other suggestions refer to possible improvements in the overall organisation such as: clear rules (Class 2); better quality of online lessons (Class 3); time schedules and the Departmental rules about teaching organisations to be strictly observed by lecturers to better organise the home self-study (Class 4).

Tab. 5: Suggestions and Proposals

Class 1	Maintain online teaching also after the end of the COVID-19 emergency period (22.39% of the analysed segments). Students in this class were mainly in the labour market, attending the fifth year of a course in the Health area.
Class 2	Clear rules (26.05% of the analysed segments). Students in this class were mainly female attending a course in the Health area and in the department of Education and Human Sciences.
Class 3	Improve the quality of the online lessons (sounds and availability of the loaded materials) and improve the quality of the internet pages of the courses, especially organisation and layout (19.07% of the analysed segments). Students in this class were mainly part-time workers, attending a course in the Society and Culture area.
Class 4	Observe the lesson time schedule and the Departmental rules about teaching organisations (32.49% of the analysed segments). Students in this class were mainly attending the first or the second year of a bachelor's degree course in Economics and are non-working or occasionally working students.

4. DISCUSSION

Some years ago, Charles Sabel and colleagues (Sabel et al, 2011) commented on the successful PISA results obtained in Finland: the secret of their success was not investment in excellence, but systematic multiple actions intended to help less advantaged students through differential support. In accordance with this perspective of differentiated policy measures, this paper has analysed the impact of COVID-19 emergency distance teaching on university students. The main results of an empirical survey conducted during the emergency highlight the importance of acquiring a shared knowledge about students' material and socioeconomic living

conditions in order to be fully prepared to implement effective university policy measures to address further needs of online teaching. The University of Modena and Reggio Emilia promptly devised a first tool for investigating the strengths and weaknesses during the emergency.

The non-parametric multivariate analyses of the empirical survey supported evidence-based learning from the emergency remote teaching. Within an encompassing framework of enhanced quality of education and personal and collective well-being, we now discuss the overall results, focusing on the factors that could be minimised or supported to create a more equal background able to provide differential support to those more in need.

The local dimension. Unimore benefited from a close connection with the local area that protected its students from sudden and unexpected changes. Overall, the local and territorial vocation of Unimore protected the students from the negative consequences of the emergency.

Infrastructural divide between the home and campus. Clearly, there are stark differences between the home and campus as regards to the material conditions for studying. On-campus facilities furnish an ecosystem full of advantages: electronic devices, high speed Internet, online resources. Quiet and semi-private spaces, support concentration, but above all else, the peer-to-peer relations that the emergency revealed as even more important than usually believed. These facilities are places for aggregation, sociability and critical debate as much as they offer both comfortable spaces for studying and accessing digital infrastructures (such as high-speed internet and laboratories). Many of these conditions are far from being available to students off campus within their homes. Here lies the vicious circle between poor material conditions and weaker university achievements, which calls a differential approach to students (especially when it comes to emergencies such as the one that turned on-campus teaching into remote teaching overnight) to be considered. Even in normal times, a marked infrastructural divide is mitigated by university policies aiming at providing basic conditions for equal opportunities in access to tertiary education. Considering the essential and non-negotiable on-campus character of the institution we call 'university', we are going to face one or two more semesters of distance teaching and students need to fully benefit from better organised classes and electronic materials. Moreover, the institution should be able to recreate the social context and vital relational ties that make the university experience a truly rich and long-running experience.

Remote teaching. Overall, the remote teaching system was positively evaluated across departments and different courses. Burdensome and fatigue are the highlighted key results. Following classes online is more challenging than in a physical room

because it requires greater concentration in addition to continuous monitor watching and headset listening. Students complained that it was harder to listen and take notes in remote settings, especially - needless to say - when private spaces or personal computers were shared and the Internet connection was slow. Over two thirds of the students expressed bewilderment, impatience and dissatisfaction, exacerbated by the sometimes-cumbersome technical procedures necessary to access the remote teaching platform and electronic resources. Three quarters complained of a major change in their organisation, with 40% having problems in organising their more solitary home activity. The lack of social contact and peer-to-peer discussion had a heavily negative influence.

Gender gap. Overall, the empirical analysis points to a significant and unexpected gender gap. Female students did not reveal any significant differences on many indicators (study organisation, concentration, material conditions of devices and spaces), save one: sharing domestic caretaking duties. Problems at home are not personal but depend on family arrangements. Females, even of such a young age, are required to contribute to domestic work to a greater extent than males. They share the caretaking of the elderly, sick or minors with their parents, which negatively affects their study 'self-organisation'. Even among younger women, the topic of an unequal distribution of duties among domestic and non-domestic spaces angrily emerges.

Clusters of students. Students have been grouped into clusters according to their answers to the closed questions and to the open-end questions. Specific combinations of material conditions, organisational choices and characteristics of the studying ecosystem clearly identify eight subpopulations of students when it comes to remote teaching: the Satisfied; the Diligent but detached; the Hostile; the Confused and unhappy; the Enthusiastic; the Hard-working but dissatisfied; the Lost; the Undecided. The extensive descriptions presented in Section 3.4.1 and 3.4.2 clarify the salient traits of these specific groups of students, prompting interesting considerations. Gender better qualifies specific groups: the Satisfied, the Diligent-but-detached, the Confused, the Undecided are groups dominated by women, while men seem to cluster less and are more spread across other groups. It emerges from free texts that female students have more home tasks (including taking care of the elderly, or of remote learning of their younger siblings) than male students. Coupled with the unexpected gap that women encounter at home, gender preferences should be taken into account by future policies in order to create both more flexibility and targeted actions. These clusters have shown themselves l to be important, especially for academic year freshmen, whose first experience of academic life is doomed to be completely virtual.

Relevant topics and suggestions. The automated text analysis furnishes a complementary material that helps in grasping the material and socioeconomic conditions of the students. Interestingly, their spontaneous answers offer insights not foreseeable at the time of the questionnaire's drafting. The main suggestion is about the style of teaching. In the new ecosystem, it is important to innovate and adapt the style of teaching to the remote setting. Online interactions among peers and with the lecturer, a more precise organisation of classes, a balanced homework load, a better audio quality are some practical suggestions.

Projecting into the near future, these suggestions hold true more in general as they intertwine with pre-emergency autonomous actions in terms of innovative teaching methods and transformations in class and curriculum supply. The survey has acknowledged the importance of a closer look at the material off-campus living conditions of the students. Comprehensive knowledge could concretely lead to that 'differential approach' and its subsequent successful results in benefiting the majority of students, not only the excellent ones. Moreover, the efficacy of and satisfaction with the teaching style could also gain from this differential approach and all the changes that the COVID-19 emergency immediately required.

5. POLICY IMPLICATIONS

There are some features of the remote teaching forced by the emergency that are typical of a phase that will be overcome as we write; but there are other elements that are rooted in what was already in previous ecosystem made of living and studying conditions, attitudes and results that were achieved before the COVID-19 outbreak. The remote teaching and services offered to students in the next semesters will have to take these features into account, in order to emerge from this phase without (or with less) dispersion. Moreover, there is a tendency to not concentrate primarily on excellence, but on a more general collective effort to increase the effectiveness of university teaching.

The results of the survey confirm the need to define what specific actions are necessary for different groups of students, departments, years of the course, conditions of study and access to resources essential for academic studying. Four lines of action seem to be essential.

The first, preliminary, action is to ensure that students are equipped with adequate digital devices and internet connections, economically helping those unable to afford such purchases in a timely manner.

The second action concerns flexibility in teaching modes that could facilitate the realignment between study programmes. If teaching in presence remains a

distinctive feature of public universities, online lessons in synchronous or remote mode, facilitate the organisation of teaching and study activities by students. Some universities are developing a blended teaching method. While it is not clear how the best practices for remote teaching will emerge and be concretely configured, there is a potential risk that the more classic onsite academic experience is biased by economic discrimination: only those who can afford travel and accommodation costs will live the traditional full academic ecosystem. It is important in any event, to reflect upon the potentialities of a mixed mode for those who cannot attend in presence.

The third action concerns the potential for improving the general quality of teaching, which can look at both students' experiences and new ways of designing lessons (e.g., peer instructions, flipped classrooms or mixed teaching tools). Designing a distance teaching system requires perhaps even more time than in-presence courses. During the emergency, teachers have had to rapidly revise not only the timing and mode of delivery of lectures, but also themes, as well as modes of interaction with students. A spinoff effect is that many students and teachers have appreciated the potential of digital technologies, that were previously ignored, such as those that allow teaching with small groups of students (in discussion/exercise sessions), the involvement and self-organisation of students for study activities or in-depth study, the exploration of a variety of topics not always able to find defined at the beginning of the course.

The fourth line of action concerns university libraries: they are vital spaces for community bonding and finding study material. This twofold function cannot be entirely shifted online, but at least investing on the online access to library' e-resources is a mandatory step for containing the potential for socio-economic bias among students.

REFERENCES

Breiman, L. (2001). Random forests. *Machine learning*. 45: 5–32.

Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics* 29: 1189–1232.

Hennig, C. and Liao, T.F. (2013). How to find an appropriate clustering for mixed -type variables with application to socio -economic stratification. *Journal of the Royal Statistical Society: Series C (Applied Statistics).* 62: 309–369.

Hodges, C., Moore, S., Lockee, B., Trust, T. and Bond, A. (2020). *The Difference Between Emergency Remote Teaching and Online Learning*. https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning. Last access, 28/01/2022.

Kaufman, L. and Rousseeuw, P.J. (2005). *Finding groups in data: an introduction to cluster analysis*. Wiley series in probability and mathematical statistics. Wiley, Hoboken, N.J.

Lebart, L., Morineau, A. and Piron, M. (2004). *Statistique exploratoire multidimensionnelle*. Dunod, Paris.

Lebart, L., Morineau, A. and Warwick, K.M. (1984). *Multivariate Descriptive Statistical Analysis*. John Wiley and Sons, New York.

Lebart, L., Salem, A. and Berry, L. (1998). *Exploring Textual Data, Text, Speech and Language Technology*. Springer, Netherlands. https://doi.org/10.1007/978-94-017-1525-6.

Marpsat, M. (2010). La méthode Alceste. *Sociologie*. Vol.1.

Misuraca, M., Scepi, G. and Spano, M. (2021). Using Opinion Mining as an educational analytic: An integrated strategy for the analysis of students' feedback. *Studies in Educational Evaluat*ion. 68: 100979.

Misuraca, M. and Spano, M. (2020). Unsupervised analytic strategies to explore large document collections, in Text Analytics. Springer, pp. 17–28.

Reinert, M. (1990). "Alceste" - une méthodologie d'analyse des données textuelles et une application : "Aurelia" de Gerard De Nerval. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*. 26: 24–54. https://doi.org/10.1177/075910639002600103.

Reinert, M. (1983). Une méthode de classification descendante hiérarchique : application à l'analyse lexicale par contexte. *Les cahiers de l'analyse des données.* VIII: 187–198.

Reynolds, A.P., Richards, G., de la Iglesia, B. and Rayward-Smith, V.J. (2006). Clustering Rules: A Comparison of Partitioning and Hierarchical Clustering Algorithms.'*J. Math. Model. Algor*. 5: 475–504. https://doi.org/10.1007/s10852-005-9022-1.

Rousseeuw, P.J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*. 20: 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

Sabel, C., Saxenian, A., Miettinen, R., Kristensen, P.H. and Hautamäki, J. (2011). *Individualized service provision in the new welfare state: Lessons from special education in Finland*. Helsinki: Sitra Studies 62.

Schubert, E. and Rousseeuw, P.J. (2019*). Faster k-Medoids Clustering: Improving the PAM, CLARA, and CLARANS Algorithms. ArXiv:1810.05691* [cs, stat] 11807: 171–187. https://doi.org/10.1007/978-3-030-32047-8_16.

LEARNING ONLINE: REMOTE TEACHING AND UNIVERSITY STUDENTS' ENGAGEMENT

ANNEXES

Annex 1 - Table A1: *Metadata*

Section	Question id	subquestion id	variable selected for partition	Question	*scale of the variable*	MODALITIES of the variables	
Language of compilation							
	d1			Language of compilation	*categoric*	Italiano English	
Gender							
	d2			Genre	*categoric*	Male | Female | I do not answer	
General information on home trips caused by the emergency: place_origin/destination							
	d3			Following the Covid-19 emergency			
		d3a			*string*	I moved to another place (please specify)	
	d4			What is the name of the province where you were staying before the Covid-19 emergency? (If you choose ABROAD, specify the country and place in the box)	*categoric*	ABROAD	111 provinces
		d4a			*string*	ABROAD (specify)	
	d5			What is the name of the municipality where you lived before the Covid-19 emergency?	*string*	Open-Ended Response	
	d6			On what date did you return home or move to another place?	*date*	MM/DD/YYYY	
	d7			What is the name of the region where you are currently living? (If you choose ABROAD, specify the country and place in the box)	*categoric*	ABROAD	111 provinces
		d7a			*string*	ABROAD (specify)	
	d8			What is the name of the municipality where you are currently living?	*string*	Open-Ended Response	
Living conditions							
	d9			How many people are living with you in your current dwelling? (Please enter a number between 0 and 100 and no text in the box)	*numeric*	Open-Ended Response	
	d10			Are you currently living in a studentsâ€™ hall of residence/boarding house?	*yes/no*	yes/no	
	d11			Do you have a space where you can study?	*categoric*	yes, I have a room for my exclusive and only use | yes, it's the room that I share with other family members / roommates | yes, it's the place that is available when I study | no, but I manage to use a dedicated space a few hours a day | no, but there is a space where I can study every now and then even if I cannot reserve it in advance Other (specify)	
		d11a			*string*	Other (specify)	
	d12			Do you have an electronic device (computer, laptop, tablet) that you can use when you study?	*categoric*	Yes, for my exclusive use | Yes, shared with others | No | No, but I can use my personal smartphone | Other (specify)	
		d12a			*string*	Other (specify)	
	d13			How do you judge the quality of your internet connection?	*categoric*	very good | good | average | not satisfactory | very bad	
	d14			Does your internet connection have a limit for maximum usable GB?	*categoric*	No, I use my home fiber or ADSL connection that has no limits | Yes, I use my smartphone connection (less than 30 GB monthly) | Yes, I use my smartphone connection (more than 30 GB monthly) | Other (specify)	
		d14a			*string*	Other (please specify)	

Section	Question id	subquestion id	variable selected for partition	Question	*scale of the variable*	MODALITIES of the variables
Ongoing problems and changes						
	d15			What are the main difficulties you are currently facing? (More than one answer is possible)		
		d15_1			*dichotomic*	I have economic problems that are not connected with Covid-19 emergency
		d15_2			*dichotomic*	I have economic problems due to the Covid-19 emergency
		d15_3			*dichotomic*	I have health problems related to the Covid-19 (my family or roommates)
		d15_4			*dichotomic*	I have health problems that are not related to Covid-19 (my family members or roommates)
		d15_5			*dichotomic*	I have problems due to the condition of my family members, roommates or friends
		d15_6			*dichotomic*	I do not have time to study because I have to take care of my family, live-in-friends
		d15_7	yes		*dichotomic*	I don't understand how to organize my studying activities
		d15_8			*dichotomic*	I don't have a proper internet connection
		d15_9			*dichotomic*	I don't have an electronic device that I can use for my studying activities.
		d15_10			*dichotomic*	I would prefer not to say anything
		d15_11	yes		*dichotomic*	None of these issues
		d15_12			*string*	Other (please specify)
	d16text			Describe how and in what way (positive and negative) the Covid-19 emergency has changed your life in general, apart from your studying activities	*string*	
Organization of study with respect to the teaching materials available, the timing and methods of the organization of study						
	d17		yes	What teaching materials do you have available?	*categoric*	All those indicated at the beginning of the lessons | Some texts | Other (specify which part you are missing, if any)
		d17a			*string*	Other (specify which parts are missing, if any)
	d18		yes	How did you rearrange your studying TIMES? (More than one answer is possible)	*categoric*	I study more because I don't have to leave the house and spend time commuting | I study more because online classes require more effort | I study more, but my level of preparation seems lower | I study the same as before in terms of hours| I study less because online classes require less effort | Other (specify)
		d18a			*string*	Other (please specify)
	d19		yes	Did you change the WAY you study? (More than one answer is possible)	*categoric*	I have not changed my way of studying: I have organized with my colleagues through chat and video | I have not changed my way of studying: I study mostly by myself | I have changed my way of studying: now I study mostly by myself | I have changed my way of studying: I have organized with my colleagues through chat and video (from WhatsApp to zoom, etc) to study together more efficiently than before | I have difficulty finding books and materials to study, because libraries are closed | Other (specify)
		d19a			*string*	Other (specify)
	d20		yes	Compared to the plans you had for this semester, indicate where you are with your preparation in the various subjects	*categoric*	I am basically caught up | I am behind on some| I am behind on all | I had decided not to prepare any exams | Other (specify)
		d20a			*string*	Other (please specify)
	d21			Describe how you are organizing your studying activities among the different subjects.		
		d21_1	yes		*dichotomic*	I am progressing equally on all the subjects taught in this semester
		d21_2	yes		*dichotomic*	I have given priority to certain subjects taught in this semester
		d21_3			*dichotomic*	I've prioritized only one subject in this semester
		d21_4	yes		*dichotomic*	I'm catching up on the preparation of last semesters' exams

Section	Question id	subquestion id	variable selected for partition	Question	*scale of the variable*	MODALITIES of the variables	
		d21_5			*dichotomic*	Differently from what I was doing before, I've decided to take the next exams as a non-attending student	
		d21_6			*dichotomic*	I'm not studying	
		d21_7			*string*	Other (specify)	
Distance learning							
	d22		yes	Are you following the distance learning activities provided by your study programme?	*categoric*		Not at all | Barely |Enough | Very much
	d23		yes	How satisfied are you with your global distance learning experience?	*numeric*	score: 1-10	
	d24			DISTANCE LESSONS: Indicate the extent to which the following issues are difficult for you today.			
		d24_1	yes		*ordinal*	the accumulation of lessons is creating difficulties for me	Not at all | Barely |Enough | Very much
		d24_2	yes		*ordinal*	I don't know how to extricate myself among the different recordings	Not at all | Barely |Enough | Very much
		d24_3	yes		*ordinal*	it takes longer to follow a recorded lesson than it does to follow a lesson inside the classroom	Not at all | Barely |Enough | Very much
		d24_4	yes		*ordinal*	I find it difficult to follow the lessons and therefore to take good notes	Not at all | Barely |Enough | Very much
		d24_5	yes		*ordinal*	the absence of involvement during the lesson makes it difficult for me to stay focused	Not at all | Barely |Enough | Very much
		d24_6	yes		*ordinal*	the absence of direct interaction does not allow for enough explanations	Not at all | Barely |Enough | Very much
		d24_7	yes		*ordinal*	the quality of the connection makes it difficult for me to follow the lesson	Not at all | Barely |Enough | Very much
		d24_8	yes		*ordinal*	I spend a lot of time in front of the screen and I feel fatigued	Not at all | Barely |Enough | Very much
	d25			CONCENTRATION, INTEREST, INTERACTION: Indicate to what extent the following issues prove difficult for you.			
		d25_1	yes		*ordinal*	I'm having a hard time concentrating	Not at all | Barely |Enough | Very much
		d25_2	yes		*ordinal*	I lack the will to study	Not at all | Barely |Enough | Very much
		d25_3	yes		*ordinal*	I'm constantly distracted or interrupted by other family or care needs at home	Not at all | Barely |Enough | Very much
		d25_4	yes		*ordinal*	The many channels of communication used for didactic purposes(chat, email, dolly, etc.) disorient me	Not at all | Barely |Enough | Very much
		d25_5	yes		*ordinal*	I can't organize my daily studying activities effectively	Not at all | Barely |Enough | Very much
		d25_6	yes		*ordinal*	I have found no difference between the way I am studying now and the way I used to study before the emergency	Not at all | Barely |Enough | Very much
	d26		yes	Has your mode of interaction with the teachers changed? Compare your present mode of interaction with your experience before the emergency.	*categoric*		Yes, for the better | Yes, for the worse | It has not changed| With distance learning, any type of interaction has ceased
	d27		yes	Has your mode of interaction with the other students changed? Compare your present mode of interaction with your experience before the emergency.	*categoric*		Yes, for the better | Yes, for the worse | It has not changed| With distance

Section	Question id	subquestion id	variable selected for partition	Question	*scale of the variable*	MODALITIES of the variables	
							learning, any type of interaction has ceased
	d28			What do you lack now that all face-to-face teaching activities have been cancelled?			
		d28_1	yes		*ordinal*	The possibility to ask for explanations while listening to the lesson	Not at all | Barely |Enough | Very much
		d28_2	yes		*ordinal*	The chance to interact face-to-face with my university mates and discuss educational issues	Not at all | Barely |Enough | Very much
		d28_3	yes		*ordinal*	The stimuli given in class that motivate you to study	Not at all | Barely |Enough | Very much
		d28_4	yes		*ordinal*	The weekly schedule of lessons	Not at all | Barely |Enough | Very much
		d28_5	yes		*ordinal*	The possibility of attending face-to-face classes	Not at all | Barely |Enough | Very much
		d28_6	yes		*ordinal*	Face-to-face office hours	Not at all | Barely |Enough | Very much
		d28_7	yes		*ordinal*	The possibility to study in the library	Not at all | Barely |Enough | Very much
		d28_8			*string*	Other (specify the missing element and the ranking in the scale - Not at all, a little, enough, a lot)	
	d29			What do you appreciate about this period, as you cannot attend face-to-face lessons?			
		d29_1	yes		*ordinal*	Increased flexibility of timetables	Not at all | Barely |Enough | Very much
		d29_2	yes		*ordinal*	I have time to devote myself to planning my future and to my hobbies	Not at all | Barely |Enough | Very much
		d29_3	yes		*ordinal*	Zero Journey time	Not at all | Barely |Enough | Very much
		d29_4	yes		*ordinal*	The chance of keeping abreast of all subjects more easily	Not at all | Barely |Enough | Very much
		d29_5	yes		*ordinal*	The possibility to pause the recording and assimilate the concepts before listening again	Not at all | Barely |Enough | Very much
		d29_6	yes		*ordinal*	Increased interaction among students, through many channels	Not at all | Barely |Enough | Very much
		d29_7	yes		*ordinal*	Less stress	Not at all | Barely |Enough | Very much
		d29_8	yes		*ordinal*	The comfort / pleasure of being at home	Not at all | Barely |Enough | Very much
		d29_9	yes		*ordinal*	The commitment of teachers in teaching and in keeping us updated	Not at all | Barely |Enough | Very much
		d29_10	yes		*ordinal*	The prompt uploading of the recorded lessons	Not at all | Barely |Enough | Very much
		d29_11	yes		*ordinal*	The University Staff commitment in providing us with all the help you can get	Not at all | Barely |Enough | Very much
		d29_12	yes		*ordinal*	The readiness and the efforts poured by the University into setting up distance learning, as well as ad hoc initiatives like this one	Not at all | Barely |Enough | Very much
		d29_13	yes		*ordinal*	The possibility to reconcile the studying activities with the burden of domestic and care commitments	Not at all | Barely |Enough | Very much
	d29_14		yes	What do you appreciate about this period, as you cannot attend face-to-face lessons? 121	*ordinal*	The awareness of having all the resources of the course available at all times and everywhere	Not at all | Barely |Enough | Very much
	d29_15				*string*		
d30text				Report up to a maximum of 3 strong points of your distance learning experience.	*string*	Open-Ended Response	
d31text				Report up to a maximum of 3 weak points of your distance learning experience.	*string*	Open-Ended Response	
d32text				Please provide your suggestions and proposals.	*string*	Open-Ended Response	

Annex 2 - Table A2: ***Share of students, by cluster partitioning elaborated on closed questions[1], according to the university status and gender***

	1 Female students satisfied	2 Diligent and detached f	3 Students hostile	4 Confused and unhappy f	5 Students enthusiastic	6 Hard-working but dissatisfied	7 The lost …first-year students	8 Undecided female students	*Sum by row*	1 Female students satisfied	2 Diligent and detached f	3 Students hostile	4 Confused and unhappy f	5 Students enthusiastic	6 Hard-working but dissatisfied	7 The lost …first-year students	8 Undecided female students	*Sum bu column*
share of students by cluster																		
	7.5	**14.5**	**11.3**	**14.9**	**10.6**	**12.3**	**18.4**	**10.6**	***100.0***	**7.5**	**14.5**	**11.3**	**14.9**	**10.6**	**12.3**	**18.4**	**10.6**	***100.0***
Area of study																		
Health	9.3	14.0	9.9	18.4	12.3	10.2	18.2	7.8	*100.0*	11.0	8.6	7.8	11.0	10.2	7.4	8.8	6.6	*8.9*
Science	8.8	11.1	12.2	21.4	11.8	10.7	13.7	10.3	*100.0*	5.8	3.8	5.3	7.1	5.5	4.3	3.7	4.8	*4.9*
Society and Culture	7.4	16.2	11.2	13.3	8.7	12.0	19.5	11.8	*100.0*	60.8	69.4	61.4	55.4	50.8	60.8	65.9	69.5	*62.1*
Technology	6.1	10.4	12.1	15.0	14.7	16.3	17.0	8.4	*100.0*	15.0	13.3	19.7	18.5	25.4	24.5	17.1	14.6	*18.4*
Life	9.9	12.8	11.5	21.1	15.1	6.6	14.8	8.2	*100.0*	7.5	5.1	5.8	8.1	8.1	3.1	4.6	4.5	*5.7*
Sum	*7.5*	*14.5*	*11.3*	*14.9*	*10.7*	*12.3*	*18.4*	*10.5*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*
Degree Course[2]																		
L	7.1	14.5	10.6	14.7	9.8	13.1	19.2	11.1	*100.0*	58.3	61.7	57.7	60.9	56.6	66.2	64.5	64.9	*61.7*
LM	7.8	13.8	14.8	14.0	12.8	12.9	15.6	8.4	*100.0*	22.0	20.3	27.7	19.9	25.6	22.4	18.1	16.9	*21.3*
LM5	7.4	16.4	10.9	16.3	10.9	8.0	17.4	12.7	*100.0*	11.0	12.7	10.8	12.2	11.5	7.4	10.6	13.6	*11.2*
LM6	11.3	13.3	7.4	17.8	11.7	8.7	21.4	8.4	*100.0*	8.8	5.3	3.8	6.9	6.4	4.1	6.8	4.6	*5.8*
Sum	*7.5*	*14.5*	*11.3*	*14.9*	*10.7*	*12.3*	*18.4*	*10.5*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*
Year of enrolment[3]																		
1	8.0	16.5	9.2	11.1	9.3	12.3	20.7	12.8	*100.0*	24.3	26.0	18.4	16.9	19.9	22.8	25.7	27.6	*22.7*
2	6.3	15.1	12.0	11.2	9.1	12.9	22.1	11.4	*100.0*	19.8	24.8	25.2	17.8	20.3	25.0	28.5	25.7	*23.7*
3	7.6	13.1	10.4	21.3	10.9	12.8	15.0	8.9	*100.0*	25.0	22.3	22.7	35.3	25.4	25.9	20.3	20.9	*24.7*
4	8.6	15.5	13.6	10.7	8.9	13.6	21.5	7.7	*100.0*	14.5	13.6	15.3	9.1	10.6	14.1	14.8	9.3	*12.7*
5	8.0	12.3	13.1	19.4	15.1	9.6	11.9	10.8	*100.0*	16.3	13.0	17.7	19.9	21.7	11.9	9.9	15.7	*15.3*
6	2.4	4.8	9.5	19.1	28.6	4.8	19.1	11.9	*100.0*	0.3	0.3	0.7	1.0	2.1	0.3	0.8	0.9	*0.8*
Sum	*7.5*	*14.5*	*11.3*	*14.9*	*10.7*	*12.3*	*18.4*	*10.5*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*	*100.0*

1 Details on cluster partitioning and variable under analysis are discussed in Section 3.3.

2 Degree Course: L: Bachelor Degree; LM: Master Degree; LM5: 5-year Degree; LM6: 6-year Degree.

3 Year of enrolment: 1, 2, 3 are the first, the second and the third year, respectively, of a bachelor's degree, 4 and 5 are the first and the second year of a master's degree or the fourth and the fifth year of a 5-year degree, 6 is the last year of a 6-year degree.

ANNEX 3 - *Simple correspondence analyses between clusters and independent variables: Area of study, Degree course of studies, Year of course, Quartiles of ECTS credits acquired with respect to achievable credits, Gender*

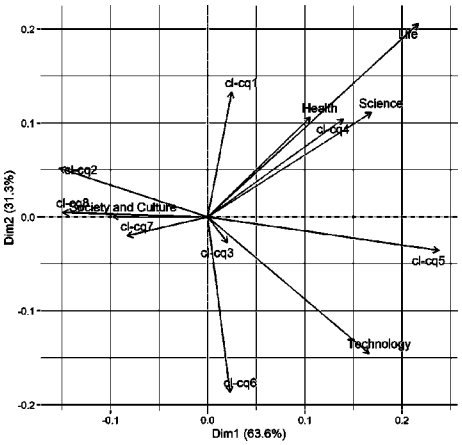


Fig. A3.1: Area of study
Thin arrows are clusters, thick arrows the modalities of the variable area.

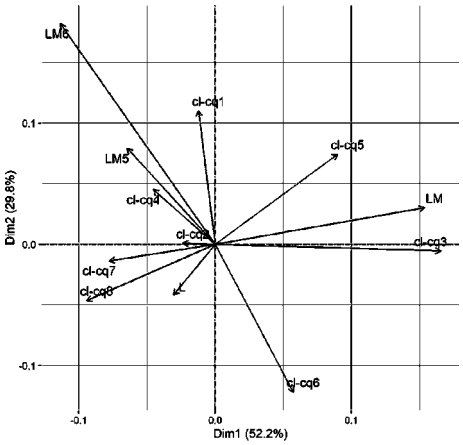


Fig. A3.2: Degree course of studies
Thin arrows are clusters, thick arrows the modalities of the variable course type: L=bachelor's degree, LM=master's degree, LM5=5-year degree, LM6=6-year degree

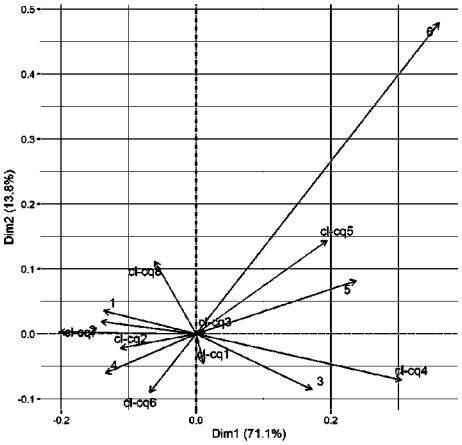


Fig. A3.3: Year of course
Thin arrows are clusters, thick arrows the modalities of the variable course year: 1, 2, 3 are the first, the second and the third year, respectively, of a bachelor's degree, 4 and 5 are the first and the second year of a master's degree or the fourth and the fifth year of a 5-year degree, 6 is the last year of a 6-year degree.

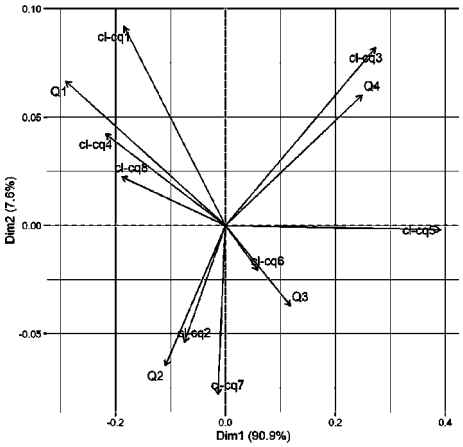
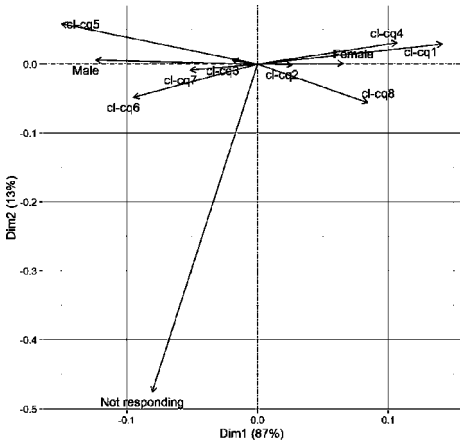


Fig. A3.4: Quartiles of ECTS credits acquired with respect to achievable credits
Thin arrows are clusters, thick arrows the modalities of the variable number of credits acquired with respect to the total achievable credits: Q1=below the 1st quartile, Q2=between the 1st and 2nd quartile, Q3=between the 2nd and 3rd quartile, Q4=above the 3rd quartile.

Fig. A3.5: Gender
Thin arrows are clusters, thick arrows the modalities of the variable gender.



ANNEX 4

d_23 *How satisfied are you with your global distance learning experience?* (graded from 1 to 10)

Clusters ranked by decreasing order of median of answers to question d_23; boxes' width: proportional to the share of students in each cluster

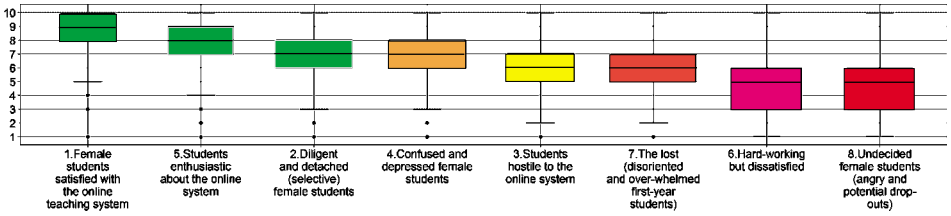


Fig. A4: ***Box plot of satisfaction level of the students with the distance learning experience (d_23), by cluster***

ANNEX 5

Table A5: ***Corpora created on free texts: respondents, types, tokens and measures of lexical diversity by corpus***

Corpus	Number of respondents	Number of types[1]	Number of tokens[2]	Mean of tokens by respondent[3]	Type/ Token ratio[4]	% hapax[5] by token	% hapax[5] by type
Ongoing Changes [q16]. Describe how and in what way (positive and negative) the COVID-19 emergency has changed your life in general, apart from your study activities	5,607	10,858	234,677	41.8	4.6	47	2.2
Strengths [q30]. Report up to a maximum of 3 strong points of your distance learning experience	2,152	3,027	35,779	16.6	8.4	51.5	4.4
Weaknesses [q31]. Report up to a maximum of 3 weak points of your distance learning experience	2,319	4,535	50,443	21.7	8.9	52	4.7
Suggestions and proposals [q32]. Please provide your suggestions and proposals	1,398	4,629	38,277	27.4	12.1	53.5	6.7

1 Number of different words present in the corpus
2 Number of occurrences of the types
3 It provides an assessment of the average text's length of the answers
4 It provides an assessment of the lexical diversity of the corpus. A higher index value indicates greater linguistic diversity and thus less repetition of the same words in the corpus. However, this ratio is highly dependent on the text's length, making it difficult to accurately compare corpora of different sizes.
5 Forms that appear only once in the corpus.

ANNEX 6 - Characteristic dictionaries identified in the Corpus *Ongoing changes*

The Corpus Ongoing changes refers to the free text in question d16-"Describe how and in what way (positive and negative) the COVID-19 emergency has changed your life in general, apart from your studying activities"

- The characteristic dictionaries in the original language (Italian) of each cluster are listed in descending order of test-value. Statistics on the terms characterising each cluster (frequency within the Corpus and within the cluster, t-value and probability) are available from the authors on request

- *cl-txt1 –Students with changes in the personal sphere and in relationships with family members and friends [Sfera personale, relazioni nella famiglia e amici]*
- about 22% of the respondents.

  cose, piccole, importanza, vita, persone, famiglia, scontato, valore, amici, importanti, importante, scontate, libertà, care, gesti, riscoperto, permesso, nonni, piccoli, passeggiata, paura, natura, portato, potere, resa, manca, persona, amiche, prova, mondo, virus, fidanzato, posto, dura, caffè, piacere, genitori, stile, quotidiane, ansia, tranquilli, banali, giardino, persa, spazi, peso, aspetto, privato, sole, negativo, quotidianità.

  [things, small, importance, life, people, family, taken for granted, value, friends, important, taken for granted, freedom, dear, gestures, rediscovered, permission, grandparents, small, walk, fear, nature, brought, power, yield, missing, person, friends, test, world, virus, boyfriend, place, hard, coffee, pleasure, parents, style, daily, anxiety, quiet, trivial, garden, lost, spaces, weight, appearance, private, sun, negative, everyday]
- *cl-txt2 – Students evaluating negative and positive aspects of the new context [Valutazione del contesto (aspetti positivi e negativi)]*
- about 25% of the respondents

  negativi, aspetti, positivi, tempo, attività, sociali, mancanza, amici, fisica, hobby, svago, relazioni, rapporti, lettura, contatti, sociale, contatto, libero, movimento, routine, limitato, studio, attenzione, minor, impossibilità, interazioni, maggiore, interessi, fisiche, noia, sport, limitazioni, gestione, sonno, dedicato, sedentaria, Vita, disposizione, aperta, parenti, passioni, esterno, aria, limitazione, spostamenti, mancanze, monotona, umore, socialità, socializzazione, uscite, giornata, allenamento, tipo, fisico, stimoli, lati, palestra, personali, quotidianità, abitudini, essenziale, sviluppo, persone, nuove, quotidiana, viaggi, eliminazione, film, teatro, manca, umani.[negative, aspects, positive, time, activity, social, lack, friends, physical, hobby, leisure, relationships, reading, contacts, social, contact, free, movement, routine, limited, study, attention, minor, impossibility, interactions, major, interests, physical, boredom, sport, limitations, management, sleep, dedicated, sedentary, Life disposition, open, relatives, passions, outside, air, limitation, movements, lacks, monotonous, mood, sociality, socialisation, outings, day, training, type, physical, stimuli, sides, gym, personal, daily, habits, essential, development, people, new, daily, travel, elimination, film, theatre, lacks, human]
- *cl-txt3 – Students with changes in the organization of the family life and of the study [Organizzazione tra casa e lavoro/studio]*
- about 24% of the respondents casa, lavoro, emergenza, dovuto, situazione, momento, causa, positiva, periodo, universitaria, sorella, biblioteca, regione,

cosa, stato, sanitaria, bloccato, ansia, sola, economici, carriera, fase, concentrazione, studio, scuola, studi, sessione, settimane, credo, universitario, difficile, risultata, sospesi, inizio, economico, possibili, negozio, spazio, genitori, problemi, percorso, bologna, rossa, studente, biblioteche, giorni, fosse, condizione, week-end, data, costretti, famiglie, solito, confusione, vista, maniera, certezza, ritardo, libri, livello, abituato, quarantena, vacanze, giorno, esercizi, unica, lasciato, strutture, mansioni, disagi, silenzio, residenza, esame.[home, work, emergency, due, situation, moment, cause, , positive, period, university, sister, library, region, thing, state, health, stuck, anxiety, lonely, economic, career, stage, concentration, study, school, studies, session, weeks, I think, university, difficult, resulted, suspended, beginning, economic, possible, shop, space, parents, problems, path, bologna, red, student, libraries, days, was, condition, weekend, date, forced, families, usual, confusion, view, manner, certainty, delay, books, level, accustomed, quarantine, holiday, day, exercises, only, left, facilities, tasks, inconvenience, silence, residence, exam]

- *cl-txt4 – Students with problems in economic and living conditions [Problemi materiali e condizioni economiche]*
- *about 10% of the respondents*

  affitto, tirocinio, cassa integrazione, anno, tesi, lavoro, contratto, stipendio, laurea, azienda, pago, studentessa, appartamento, bollette, compagno, spese, figlio, madre, stato, sperimentale, origine, lavoratrice, bimba, tornata, crediti, genitori, educatrice, sede, farmacia, messa, progetto, causa, nido, conseguimento, anni, dovuto, marito, panico, economica, chiamata, fuorisede, arrivata, indeterminato, ritrovata, borsa, tasse, sospeso, laboratorio, attacchi, rischio, padre, interrotto, lavorato, piani, rimasta, lavorativo, soldi, estero, crisi, garantito, momento, ragazzo, scuole, percorso, emergenza, finito, costretta, Italia, studi, future, chiusura, città, magistrale, economico, tutor, dipendente, scuola, economiche, ritrovo, preoccupata.[rent, internship, year, thesis, job, contract, salary, degree, company, pay, student, flat, bills, partner, expenses, child, mother, state, experimental, origin, worker, child, returned, credits, parents, educator, seat, pharmacy, put, project, cause, nest, achievement, years, due, husband, panic, economic, call, out-of-school, arrived, indefinite, found, bag, taxes, suspended, lab, attacks, risk, father, interrupted, worked, plans, stayed, working, money, foreign, crisis, guaranteed, moment, boy, schools, path, emergency, finished, forced, Italy, studies, future, closing, city, master, economic, tutor, employee, school, economic, finding, worried]
- *cl-txt5 – Students focusing on the positive and negative novelties of online teaching [Lezioni e organizzazione della didattica di emergenza]*
- *about 18% of the respondents lezioni, online, esami, professori, registrate,*

lezione, docenti, connessione, prof, modalità, orari, caricate, video, esame, aula, studenti, distanza, corsi, argomenti, disponibili, materie, sessione, dubbi, orali, internet, appelli, materiale, e-mail, professore, frontali, università, docente, registrazioni, appunti, scritti, date, computer, diretta, presenza, orario, calendario, corso, videolezioni, domande, telematica, dolly, studio, diretto, stati, preparazione, distrazioni, estiva, materia, possibilità, didattica, ascolto, organizzazione, utile, risparmio, comunicazione, chiarimenti, mezza, svolte, qualità, fornito, piattaforma, eventuali, registrazione, audio, laboratori, tempi, complicato, risposta, sistema, carico, fissi, pubblicano, utili, classe, carica, notato, spostamenti.[lessons, online, exams, professors, recorded, lesson, teachers, connection, prof, mode, times, uploaded, video, exam, classroom, students, distance, courses, topics, available, subjects, session, doubts, oral, internet, appeals, material, e-mail, professor, frontal, university, teacher, recordings, notes, written, dates, computer, direct, presence, time, calendar, course, videolessons, questions, telematics, dolly, study, direct, states, preparation, distractions, summer, matter, possibility, didactics, listening, organization, useful, saving, communication, clarification, half, carried out, quality, provided, platform, any, recording, audio, labs, time, complicated, answer, system, load, fixed, publish, useful, class, charge, noticed, movements]

ANNEX 7 - Table A7: ***Descriptive statistics refers to the output of the ALCESTE method on the corpora Strengths, Weaknesses, Suggestions and proposals***

	d30: Strengths	d31: Weaknesses	d32: Suggestions and proposals
Number of respondents (texts)	2152	2319	1398
Number of textual segments	2198	2493	1636
Number of forms	3027	4535	4629
Occurrences	35779	50443	38277
Number of lemmas	2026	2822	2986
Active forms	1741	2494	2633
Supplementary forms	241	282	297
Active forms with a frequency >3	618	944	931
Average of forms per segment	16.28	20.23	23.4
Number of classes	2	5	4
Number of classified segments (absolute value)	2129	2227	1505
Number of classified segments (percent value)	96.86	89.33	*91.99*

ANNEX 8 – Network graphs of the terms of each thematic cluster identified by the ALCESTE method in the corpora *Strengths*, *Weaknesses* and *Suggestions and proposals*

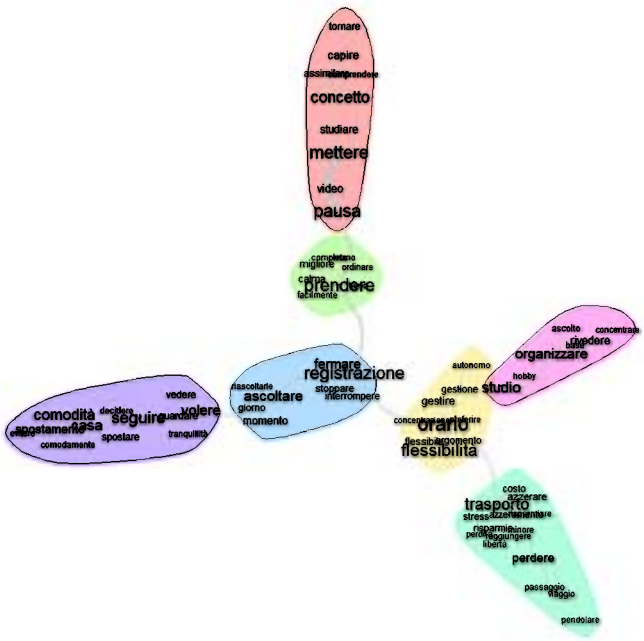


Fig. A8.1: Corpus Strengths – Thematic cluster 1

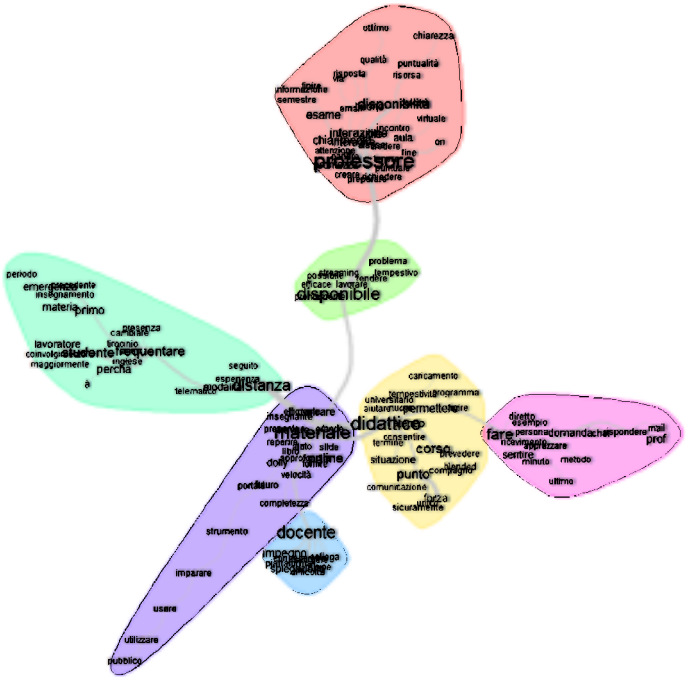


Fig. A8.2: Corpus Strengths – Thematic cluster 2

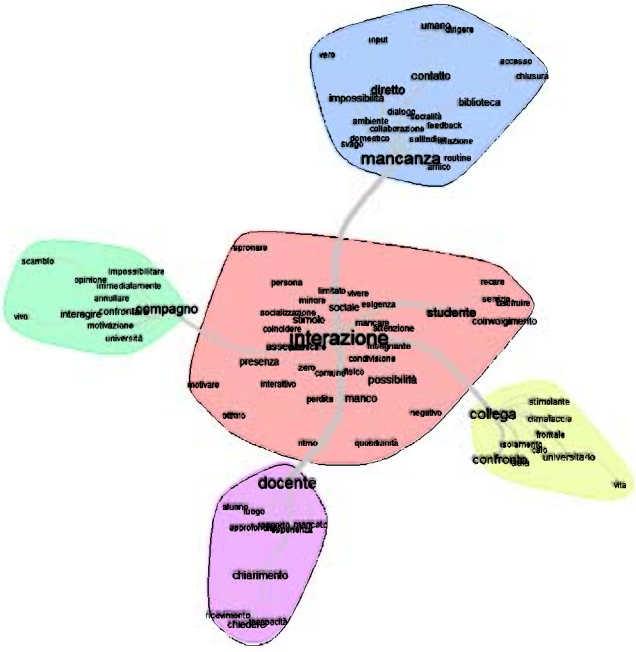


Fig. A8.3: Corpus Weaknesses – Thematic cluster 1

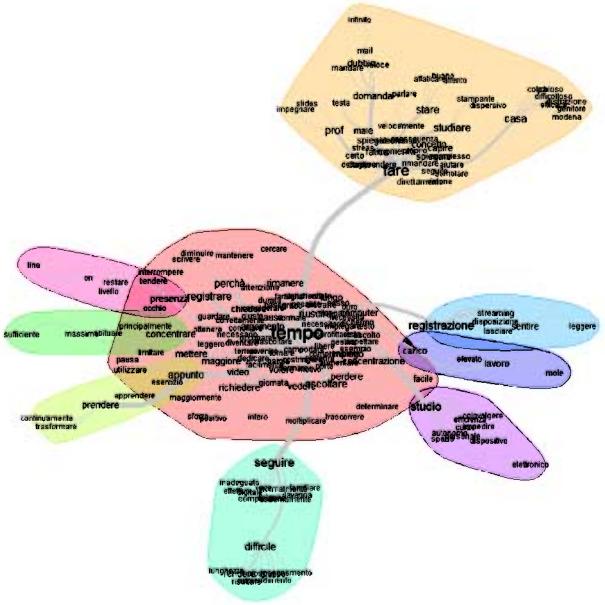


Fig. A8.4: Corpus Weaknesses – Thematic cluster 2

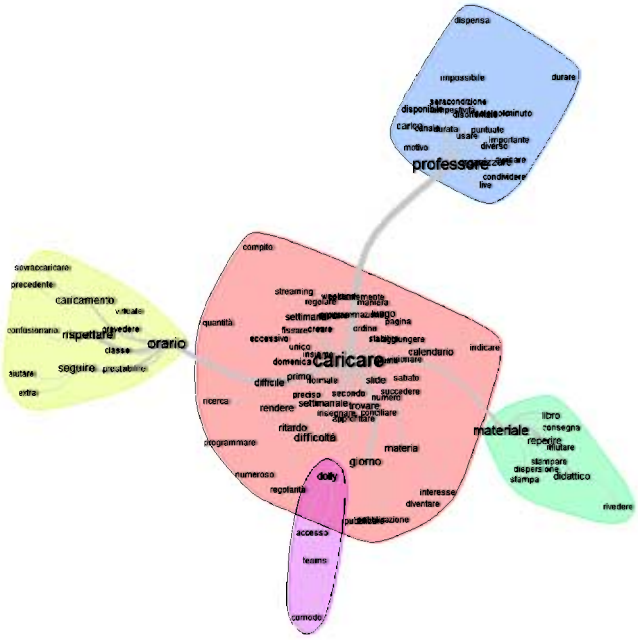


Fig. A8.5: Corpus Weaknesses – Thematic cluster 3

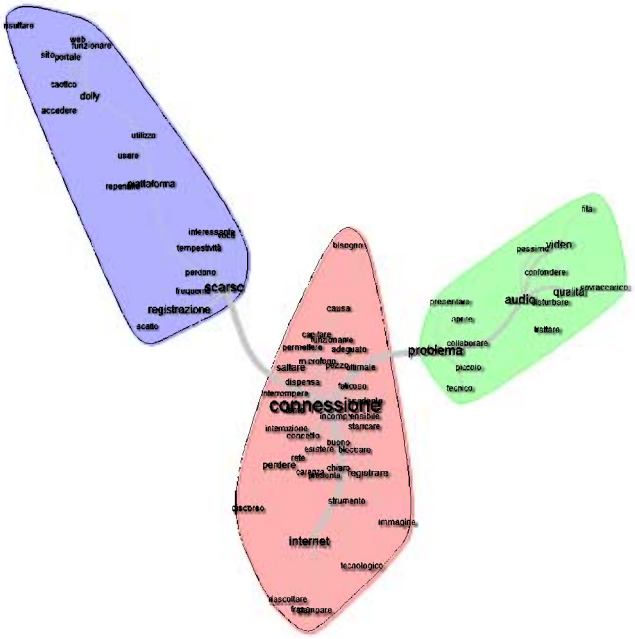


Fig. A8.6: Corpus Weaknesses – Thematic cluster 4

Fig. A8.7: Corpus Weaknesses – Thematic cluster 5

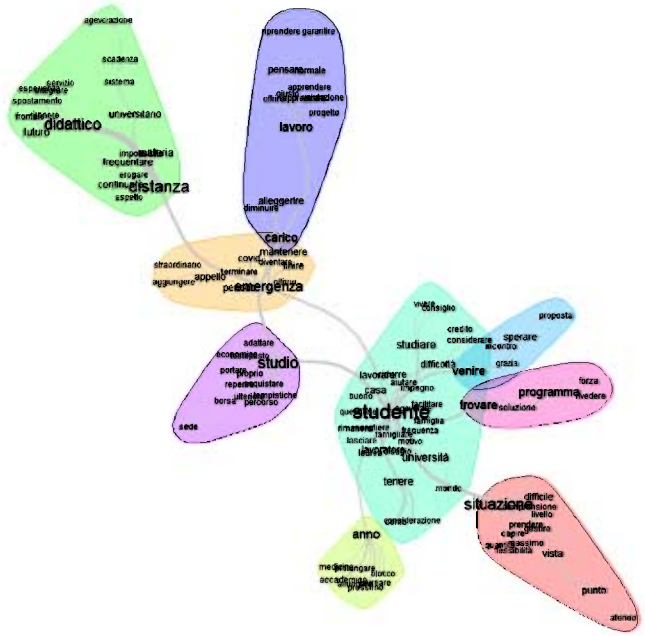


Fig. A8.8: Corpus Suggestions and proposals – Thematic cluster 1

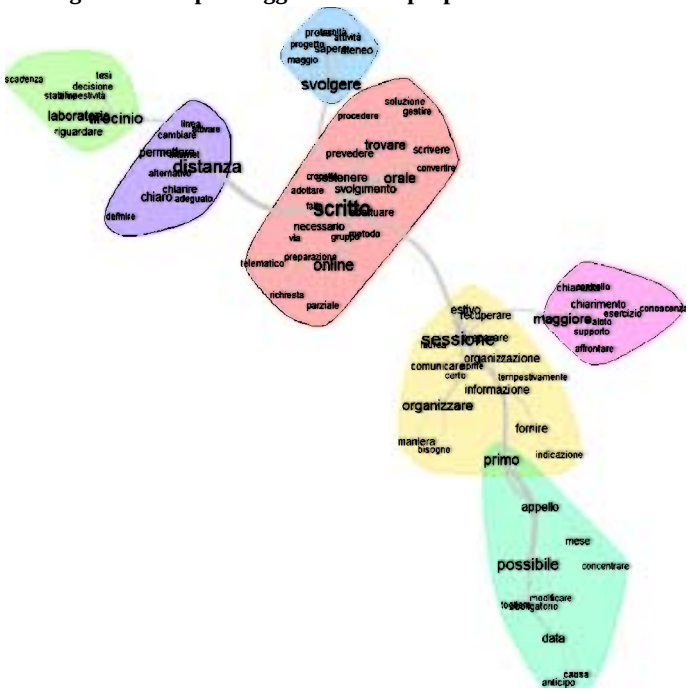


Fig. A8.9: Corpus Suggestions and proposals – Thematic cluster 2

Fig. A8.10: Corpus Suggestions and proposals – Thematic cluster 3

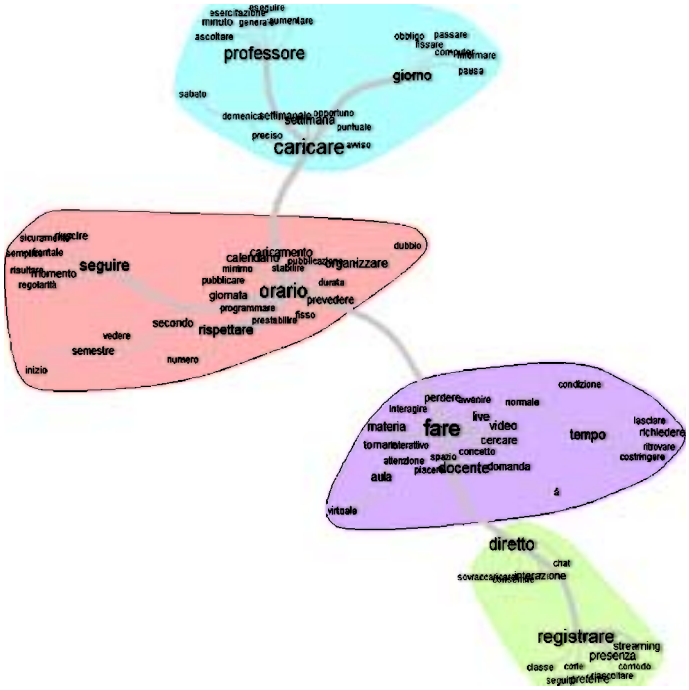


Fig. A8.11: Corpus Suggestions and proposals – Thematic cluster 4

ANNEX 9 – Selection of texts from the open-ended responses (in *Italian* and translated in English) on *Strengths, Weaknesses, Suggestions and proposals* Corpora

- Statistics on the terms characterising each class (frequency within the Corpus and within the class, t-value and probability) are available from the authors on request.
- Respondent ID generated by Survey Monkey
- In italics: original text in Italian

Corpus Strengths

Thematic cluster 1

* *11489097187 organizzazione flessibilità comodità.*
* *11490896693 flessibilità degli orari possibilità di riascoltare le lezioni possibilità di mettersi in pari con le lezioni.*
* *11512070162 grande flessibilità di orari possibilità di tornare su punti precedentemente discussi con molta facilità.*
* *11487505612 videolezioni sempre disponibili.*
* *11509041617 i docenti via mail sono sempre disponibili e anche via skype.*
* *11508704693 i professori sono molto disponibili.*
* *11509740650 1 professori molto disponibili e volenterosi 2 possibilità di bloccare e riascoltare le lezioni a piacimento 3 possibilità di ascoltare le lezioni in qualunque giorno*
* 11489097187 organisation flexibility convenience.
* 11490896693 flexibility of timetables possibility of re-listening to lessons possibility of catching up on lessons.
* 11512070162 great flexibility of timetables possibility to return to previously discussed points very easily.
* 11487505612 video lessons always available.
* 11509041617 teachers are always available via email and also via Skype.
* 11508704693 the teachers are very helpful.
* 11509740650 1 very helpful and willing professors 2 possibility of blocking and listening to the lessons at will 3 possibility of listening to the lessons on any day

Thematic cluster 2

* *11492927952 alcuni professori stanno sfruttando le lezioni online per creare maggiore interazione con noi studenti.*
* *11509485560 ottima organizzazione da parte dell'ateneo, l'impegno dei docenti.*
* *11516285371 maggior interazione col docente durante la lezione*
* *11522970349 organizzazione dei professori e università.*
* *11522685697 organizzazione costanza interazione con i compagni di studio.*
* *11491071763 c'è molta più interazione sia tra docenti che tra i miei colleghi durante la lezione.*
* *11492642993 azzeramento dei costi e tempi di trasporto 2 maggiore interazione e confronto con i compagni di corso*

* 11492927952 some professors are using online classes to create more interaction with us students.
* 11509485560 very good organisation by the university, the commitment of the teachers.
* 11516285371 more interaction with the lecturer during the lesson
* 11522970349 organisation of professors and university.
* 11522685697 organisation constancy interaction with fellow students.
* 11491071763 much more interaction with professors and fellow students during class.
* 11492642993 zero transport costs and times 2 more interaction and confrontation with classmates

Corpus Weaknesses

Thematic cluster 1

* *11489481689 mancanza di interazione diretta con i prof che spesso rispondono alle mail dopo molto tempo e o in modo non consono.*
* *11509716367 zero interazioni con i professori, zero interazioni con i colleghi.*
* *1150833588 manca interazione con gli insegnanti alcuni prof annullano le lezioni manca il confronto tra colleghi di corso.*
* *11489616068 no interazione diretta con i docenti, possibile accumulo delle registrazioni, mancanza di stimoli.*
* *11509785526 difficile interazione e confronto con i compagni chiusura delle biblioteche difficile interazioni con i professori*
* 11489481689 lack of direct interaction with professors who often respond to emails after a long time and or in an inappropriate manner.
* 11509716367 zero interaction with professors, zero interaction with colleagues.
* 1150833588 lack of interaction with teachers some profs cancel lessons lack of confrontation between course colleagues.
* 11489616068 no direct interaction with professors, possible accumulation of recordings, lack of stimuli.
* *11509785526 difficult interaction and confrontation with fellow students closure of libraries difficult interaction with professors*

Thematic cluster 2

* *11489650862 le lezioni a distanza non permettono di essere concentrati al massimo.*
* *11508518450 le lezioni online sono molto più lunghe delle lezioni in classe mi è difficile concentrarmi e studiare in casa la mole di studio è inadeguata alla situazione.*
* *11489926986 lezioni troppo lunghe e quindi difficile stare attenti continuamente, difficile rimanere in pari con tutte le materie da seguire.*
* *11488971315 ci metto più tempo ad ascoltare le lezioni rispetto alla loro durata effettiva.*
* *11492381861 impiego tanto tempo per ascoltare bene una lezione.*
* *11493253390 alcuni professori sono veramente pesanti da ascoltare non riescono a mantenere l'attenzione*
* 11489650862 the distance learning lessons do not allow you to concentrate to the maximum.
* 11508518450 online lessons are much longer than classroom lessons it is difficult for me to concentrate and study at home the amount of study is inadequate for the situation.
* 11489926986 lessons are too long so it's hard to pay attention all the time, it's hard to keep up with all the subjects to follow.

* 11488971315 I take longer to listen to the lessons than the actual duration.
* 11492381861 it takes me a long time to listen well to a lesson.
* 11493253390 some professors are really hard to listen to, they can't keep your attention.

Thematic cluster 3

* *11508329144 non vengono rispettate le durate delle lezioni sebbene sia stato detto che 1h 30 di lezione in presenza equivalga a 45min di registrazione spesso ci troviamo a seguire anche 4h di fila ed è insostenibile.*
* *11492335702 le lezioni registrate richiedono molto più tempo rispetto a quelle in presenza alcuni professori sembrano aver aumentato il carico di studi senza rispettare le tempistiche di una normale settimana di lezioni in presenza la preoccupazione di dover svolgere gli esami online.*
* *11489680249 3 i professori non rispettano il limite dei 45 minuti circa e fanno lezioni di oltre un ora e mezza.*
* *11509719266 troppo carico di materiale da parte dei professori non rispetto degli orari di caricamento delle lezioni.*
* *11512937188 disorganizzazione del corso ogni professore carica in una piattaforma diversa e molti professori non caricano il materiale didattico*
* 11508329144 the duration of the lessons is not respected although it has been said that 1h30 of lessons in person is equivalent to 45 minutes of recording, we often find ourselves following 4 hours in a row and it is unbearable.
* 11492335702 the recorded lessons take much longer than the face-to-face lessons some professors seem to have increased the study load without respecting the times of a normal week of face-to-face lessons the worry of having to take exams online.
* 11489680249 3 teachers do not respect the limit of about 45 minutes and give lessons of more than an hour and a half.
* 11509719266 too much material on the part of the professors not respecting the loading times of the lessons.
* 11512937188 disorganisation of the course each professor uploads on a different platform and many professors do not upload the teaching material.

Thematic cluster 4

* *11490407742 a volte ci sono problemi di connessione e stare davanti al computer tutto il giorno è stancante. .*
* *11510305821 connessione internet non delle migliori.*
* *11519084549 1 problemi di connessione 2 il video a volte si blocca 3 non si sente l audio a volte.*
* *11489617698 stanchezza nel stare davanti al pc per ore problemi di connessione.*
* *11511897073 problemi con la piattaforma dolly nel reperire materiale online.*
* *11489201222 connessione che talvolta salta poca uniformità di sistemi scelti tems dolly skype.*
* *11489729038 carico registrazioni malfunzionamento piattaforma dolly.*
* *11508481309 microfoni mal funzionanti professori che non registrano lezioni ma assegnano libri o dispense.*
* *11496644258 spesso i docenti effettuando le lezioni in diretta non accettano che eventuali dispositivi come il microfono non possano funzionare oppure se funzionano il docente non riesce a sentire gli studenti per problemi tecnici*
* 11490407742 sometimes there are connection problems and being in front of the computer all day is tiring.

* 11510305821 not the best internet connection.
* 11519084549 1 connection problems 2 the video sometimes freezes 3 I can't hear the sound at times.
* 11489617698 tiredness in front of the pc for hours connection problems.
* 11511897073 problems with the dolly platform in finding material online.
* 11489201222 connection that sometimes breaks down lack of uniformity of systems chosen tems dolly skype.
* 11489729038 loading registrations dolly platform malfunction.
* 11508481309 malfunctioning microphones professors who do not record lessons but assign books or handouts.
* 11496644258 often teachers, when they do live lessons, do not accept that devices such as microphones do not work or, if they do work, the teacher cannot hear the students because of technical problems.

Thematic cluster 5

* *11489361109 mancanza di informazioni riguardanti la modalità degli esami le date precise scarsa comunicazione spiegazione delle differenze tra esame da frequentante e non se avessi date certe di esami riuscirei a farmi un piano di studio e organizzarmi*
* *11489748604 ritardo nel fornire informazioni circa gli esami.*
* *11519184370 poche informazioni riguardo gli esami o meglio nessuna.*
* *11534404004 incertezza sulle modalità e tempi degli esami per le materie di questo semestre soprattutto per gli esami scritti che alcuni professori sono convinti di poter svolgere a breve in presenza incertezza sulle modalità di recupero dei tirocini.*
* *11492560383 non vi è modo di sostenere gli esami scritti i professori non hanno ideato soluzioni alternative e la cosa mi mette decisamente in difficoltà.*
* *11508866947 poca chiarezza dei professori su alcuni argomenti pochi chiarimenti riguardo le possibili modalità di esame e assenza di pre appelli poche esercitazioni compiti progetti da fare a casa.*
* 11489361109 lack of information regarding the modality of the exams, the precise dates, poor communication, explanation of the differences between attending and non-attending exams, if I had certain dates of the exams, I would be able to make a study plan and organize myself
* 11489748604 delay in providing information about the exams.
* 11519184370 little or no information about exams.
* Uncertainty about the modalities and timing of the exams for the subjects of this semester, especially for the written exams that some professors are convinced they will be able to carry out shortly in the presence of uncertainty about how to recover the internships.
* 11492560383 there is no way to take the written exams the professors have not devised alternative solutions and this definitely puts me in difficulty.
* 11508866947 lack of clarity of the professors on some topics few clarifications regarding the possible exam modalities and absence of pre-trials few exercises tasks projects to do at home

Corpus Suggestions and proposals

Thematic cluster 1

* *11547817203 spero si possa utilizzare questa modalità anche in un futuro per chi lavora può essere un ottima opportunità.*

* *11508645481 continuare in futuro con la didattica a distanza con eccezione solo per materie obbligatorie e in laboratorio.*
* *11510614138 mantenere la modalità delle lezioni registrate anche in futuro al di fuori dell'emergenza per permettere anche agli studenti lavoratori o con altre problematiche di poter assistere alle lezioni anche da casa.*
* *11508441897 sotto certi aspetti si potrebbe tenere in considerazione la didattica a distanza anche in futuro.*
* *11487998284 bisognerebbe attivare la didattica a distanza sempre in qualsiasi momento dell'anno accademico.*
* *11508514494 continuare la didattica a distanza anche dopo.*
* *11489654484 mi auguro quindi di poter usufruire in futuro di materiale didattico di questa qualità senza rendere vano quello che questa esperienza ha da insegnarci.*
* 11547817203 I hope we can also use this method in the future for those who work, it can be an excellent opportunity.
* 11508645481 to continue with distance learning in the future with the exception of compulsory subjects and laboratories only.
* 11508645481 to continue with distance learning with the exception of compulsory subjects and the laboratory.
* 11508441897 in some respects distance learning could also be considered in the future.
* 11487998284 distance learning should always be activated at any time during the academic year.
* 11508514494 Distance learning should also be continued afterwards.
* 11489654484 I therefore hope to be able to benefit from teaching materials of this quality in the future without making what this experience has to teach us meaningless.

Thematic cluster 2

* *11508312488 stabilire quanto prima le modalità d esame nel dettaglio e comunicarle agli studenti.*
* *11508719487 chiarimenti sulle modalità d esame.*
* *11508380824 chiarire al più presto le modalità d esame. .*
* *11509951754 chiarimenti sulla modalità di esame pronto.*
* *11490071431 dare informazioni riguardo gli esami della sessione estiva se vengono modificate le modalità in modo da adattare lo studio in previsione della prova che verrà somministrata. .*
* *11512994728 dare delle regole di ateneo per quanto riguarda modalità di svolgimento di lezioni e esami che vengano rispettate da tutti i docenti.*
* *11509081868 bisognerebbe rispettare l orario di lezione inoltre sarebbe utile se venissero fornite a noi studenti maggiori indicazioni riguardo le modalità di esame per l imminente sessione ad oggi non è stato detto ancora nulla.*
* 11508312488 establish detailed examination procedures as soon as possible and communicate them to the students.
* 11508719487 clarification of examination procedures.
* 11508380824 clarify the examination procedures as soon as possible.
* 11509951754 clarification of examination mode ready.
* 11490071431 give information about the exams of the summer session if the modalities are changed in order to adapt the study in anticipation of the test that will be administered.
* 11512994728 to give the university rules for the conduct of lectures and exams that are respected

by all teachers.

* It would also be useful if we students were given more information about the examination procedures for the upcoming session, but nothing has been said yet.

Thematic cluster 3

* *11490999323 pagine dolly meno confusionarie.*
* *11508516154 migliorare la piattaforma dolly.*
* *11510487010 fornire libri in pdf sulla pagina dolly poiché non tutti i libri sono di facile reperibilità. .*
* *11489018378 caricare materiale su dolly e non su altre piattaforme nelle lezioni registrate ricordarsi che gli studenti devono prendere appunti e non possono chiedere di ripetere.*
* *11522791588 la piattaforma dolly deve essere migliorata e i professori dovrebbero caricare le lezioni in formati che funzionino in maniera fluida.*
* *11508482205 mettere a disposizione materiale quali libri di testo come file pdf sul portale dolly causa la difficile reperibilità semplificare i canali di comunicazione.*
* *11512060656 migliorare la qualità delle registrazioni.*
* *11508341491 rivedere assolutamente la qualità audio delle registrazioni.*
* *11492627690 migliorare la qualità audio delle lezioni. .*
* *11512228427 migliorare qualità audio e video di alcune lezioni.*
* *11508803630 migliorare la qualità delle lezioni che sono poco comprensibili a causa di problemi tecnici.*
* *11489588693 fornire ai professori attrezzatura adeguata a realizzare registrazioni di buona qualità.*
* 11490999323 less confusing dolly pages.
* 11508516154 improve the dolly platform.
* 11510487010 provide pdf books on the dolly page as not all books are easy to find. .
* 11489018378 upload material on dolly and not on other platforms in recorded lessons remember that students have to take notes and cannot ask to repeat.
* 11522791588 the dolly platform needs to be improved and professors should upload lessons in formats that work fluidly.
* 11508482205 make materials such as textbooks available as pdf files on the dolly portal because of the difficult availability simplify communication channels.
* 11512060656 improve the quality of recordings.
* 11508341491 absolutely revise the audio quality of the recordings.
* 11492627690 improve audio quality of lessons.
* 11512228427 improve the audio and video quality of some lessons.
* 11508803630 improve the quality of the lessons that are not understandable due to technical problems. .
* 11489588693 provide teachers with adequate equipment to make good quality recordings.

Thematic cluster 4

* *11508312612 i professori dovrebbero caricare le lezioni negli orari stabiliti non alle 11 di notte e non il sabato e la domenica perché noi studenti non abbiamo il computer in mano in ogni momento della giornata e dobbiamo organizzarci anche noi con le lezioni e lo studio.*
* *11508530059 penso sia opportuno che i professori si organizzino per caricarci le lezioni in giorni*

prefissati da loro perché ora tutti i giorni occorre salire sul portale per verificare se sono state aggiunte le lezioni e facendo così è impossibile organizzarsi con lo studio.

* *11510153433 le lezioni dovrebbero essere più concise è molto difficile seguire bene una lezione online perché richiede molto più tempo di una lezione normale..*
* *11489588416 cercare di far rispettare il più possibile ai docenti il proprio orario o per lo meno la giornata in cui avrebbero lezione lasciare il tempo materiale di studiare e assimilare i concetti affrontati a lezione.*
* *11508624193 rispettare l orario assegnato.*
* *11489760464 far rispettare un orario preciso per la pubblicazione delle lezioni esattamente come avverrebbe in caso non ci fosse tale emergenza quindi seguendo un orario normale di lezioni.*
* *11508680987 rispettare il caricamento o il sostenimento delle lezioni negli orari previsti dall'orario delle lezioni: ci siamo ritrovati alcuni giorni senza avere nessuna lezione e altri in cui si accumulava talmente tanta roba da aver bisogno di passare tutta la giornata al computer.*
* *11508576980 cercare di far rispettare il più possibile l orario delle lezioni ai professori e cercare di incentivare i professori a svolgere prove intermedie scritte.*
* *11489435884 rispettare il numero di lezioni che si ha a settimana senza approfittare del fatto che siamo casa non à corretto.*
* *11489807647 consiglierei ai professori di rispettare i 45 minuti stabiliti per ciascuna lezione.*
* *11534365431 incentivare caldamente il rispetto delle ore di lezione e la serietà dei professori nel rispettare il numero di ore assegnate senza chiedere di fissarne tante in più.*
* 11508312612 the teachers should upload the lessons at the fixed times, not at 11 o'clock at night and not on Saturdays and Sundays because we students do not have the computer in our hands at all times of the day and we also have to organise our lessons and study.
* 11508530059 I think it would be a good idea for the teachers to organise themselves so that they can upload the lessons on days set by them, because now every day we have to go on the portal to check whether the lessons have been added, and by doing this it is impossible to organise our studies.
* 11510153433 the lessons should be more concise, it is very difficult to follow an online lesson well because it takes much more time than a normal lesson....
* 11489588416 try to make the teachers respect their own timetable as much as possible, or at least the day of the lesson, leaving them enough time to study and assimilate the concepts covered in the lesson.
* 11508624193 respect the assigned time.
* 11489760464 to respect a precise timetable for the publication of the lessons exactly as it would happen in case of no such emergency, therefore following a normal timetable of lessons.
* 11508680987 to respect the uploading or the taking of lessons at the times foreseen by the lesson timetable: we have found ourselves some days without having any lessons at all and others where so much stuff accumulated that we needed to spend the whole day at the computer.
* 11508576980 try to make the teachers respect the lesson timetable as much as possible and try to encourage the teachers to carry out written intermediate tests.
* 11489435884 respect the number of lessons you have per week without taking advantage of the fact that we are at home is not correct.
* 11489807647 I would advise the professors to respect the 45 minutes set for each lesson.
* 11534365431 strongly encourage respect for teaching hours and the seriousness of professors in respecting the number of hours allocated without asking for many more to be scheduled.

DID TRANSITION TO ONLINE TEACHING DURING COVID-19 PANDEMIC AFFECT STUDENTS' PERFORMANCE? EVIDENCE FROM A STATISTICS COURSE

Samantha Leorato[1]

Dept. Economics, Management and Quantitative Methods University of Milan

Abstract. *The COVID-19 pandemic and the lockdown imposed by many countries had a huge impact on social and economic aspects of life, throughout the world. Schools of all order and grades were affected severely and forced to transition to alternative teaching methods. Most universities, in particular, turned to different forms of teaching in-distance. Students and instructors had to face the challenge of adapting to a sudden change in the delivery of courses and exams. This paper tries to address the question of how the unavoidable distress caused by these changes has affected higher education students' performance, using data from a Statistics course in a Master of Science program in the two consecutive years 2019 and 2020. Evidence of a significant Year effect is found, which seems to be ascribable, at least partially, to factors different from student's cohort characteristics.*

Keywords: *Distribution regression, linear mixed-effects models, COVID-19, online exams, higher education.*

1. INTRODUCTION

The COVID-19 pandemic and the consequent lockdown imposed by governments all over the world resulted in schools of all grades and levels to shut and resort to alternative teaching methods and evaluation procedures.

Most of schools, colleges and universities turned to online learning to continue their activity in compliance with health security measures (see Crawford et al. (2020), Toquero (2020), Sahu (2020)). These transitions were a response to an exogenous shock and concerns were raised about readiness of institutions, teachers and students (OECD (2020), Zalite and Zvirbule (2020), Scherer et al. (2021)).

Italy (Lombardy region in particular) was the first, among western countries, to be violently hit by the spread of the virus. Universities and schools in Lombardy suspended in-presence activities in February the 24th. Soon after the initial

[1] Corresponding author: Samantha Leorato, email: samantha.leorato@unimi.it

suspension of teaching, all universities started a progressive transition towards online learning, and the lockdown was extended to the whole Italian territory in March the 9th.

Despite an undeniable effort to provide a prompt response to the emergency, many universities and students faced weeks of uncertainty and re-organization that might have had an impact on students' psychological condition and academic performances. One of the challenges of the transition to e-learning was how to redefine students' evaluation. Whenever possible, many universities suggested to replace written exams with oral examination, while, for courses with large number of attendees, several different procedures were proposed, mostly involving the use of some form of certified "examination platform", with monitoring either performed by instructors themselves or by artificial intelligence.

This paper aims at quantifying the effect of the disruption induced by the pandemic on student's performance. The changes and inconveniences imposed by the COVID-19 have likely caused difficulties for several different reasons including: problems in adapting to the transition to e-learning and to online examination procedures, as well as emotional distress caused by uncertainty, by isolation and health emergency in general. The data used in this analysis do not allow to disentangle all the different ways the pandemic could potentially have affected students' performance. Thus, this paper aims at measuring a *collective Covid effect*, meaning by this "the sum of all Covid-related effects on student's achievements", irrespective of their nature (psychological, cognitive, practical…).

In this respect, this work enters into the emerging literature trying to quantify the effect that COVID-19 had on different aspects of life. Most of the quantitative analyses so far focused mainly on psychological consequences of lockdown on students of different school levels (Sahu (2020), Cao et al. (2020), Odriozola-Gonzàlez et al. (2020), Zimmermann et al. (2021), Lee (2020) among the others), or on student's perception of the transition to online teaching (Murphy et al. (2020)) rather than on student's performance per se.

For what concerns the measurement of learning losses, higher attention has been given to younger students, for example Andrew et al. (2020) and Maldonado and De Witte (2020) focus on primary schools while Dietrich et al. (2021) consider high school pupils. However, the evaluation of the impact on graduate student's achievements is crucial to predict mid-long term effects on their academic performances and on employment prospects (see Pietro et al. (2020)). It is not clear *a priori* whether the expected effect should be positive or negative. On the one hand, the lockdown imposed changes in teaching delivery and in student's assessment method (from in-person to online) that might have produced a negative effect, due

to lack of preparation of the universities.

Further, according to recent studies, online courses themselves are reckoned a negative – if ever significant – effect. On the other hand, although the isolation may have produced a psychological distress that is likely to have negatively affected their performance, the lockdown forced students to stay home and give up other activities, potentially leaving more time for study. For example, Gonzalez et al. (2020), using data from years 2017 to 2020, compare the results of the final evaluation in three different exams offered at Universidad Autónoma de Madrid (Applied Computing, Metabolism, Design of Water Treatment Facilities) and obtain a positive effect of COVID-19 confinement, which they attribute to improved efficiency due to changing the learning strategies towards a more continuous habit. Consistently with this dual mechanism, Aucejo et al. (2020) find highly heterogeneous effects, mostly following socioeconomic differences; lower income students are 55% more likely to have delayed graduation due to COVID-19 than their higher-income peers. Similarly, using a survey-based study of 232 undergraduate and postgraduate students in West Bengal, India, Kapasia et al. (2020) point out that although about 70% of students were able to use digital platforms for learning during the lockdown, most of them had to "face huge challenges in online study", especially those leaving in rural areas or with lower income.

Independently on the extreme situation that led to this sudden transition, online courses and exams have seen an exponential increase over the last decade. As a consequence, their effect on students' achievement have been largely studied since then and some insight can also be drawn from the literature. As mentioned above, several works find a negative or at best no impact of online courses on student's test scores (see for instance Figlio et al. (2013), Alpert et al. (2016); Joyce et al. (2015), Bowen et al. (2014)). These works mostly focus on single courses (in Economics or Statistics) where random assignment of students to online or in-person classes has been implemented, that gives them the characteristics of an experimental design. A more recent comprehensive study takes into account more than 700 courses and 200,000 students from for-profit colleges in US (Bettinger et al. (2017)), and a negative effect of online courses on students grades is found both for the course taken online and for future courses. Data are non-experimental and therefore subject to potential selection bias, which is tackled by the introduction of instrumental variables.

This paper takes into account data on grades from two consecutive academic years (2019 and 2020) of an exam of Statistics from a two year Master program. This specific exam has several advantages. First, focusing on an exam in the area of Statistics is likely to produce stronger effects compared to courses requiring *soft*

skills and it is therefore particularly interesting. The course maintained the same syllabus in the two years and also the structure of the final assessment was almost unchanged: in both cases, it consisted in answering to several questions after running the appropriate statistical analysis on a given dataset, the only difference is that, while in 2019 the exam was held in-presence, in 2020 students did the exam from their home, with online proctoring. Further, the teaching period was January-March in both years, which means that the lockdown and the transition to distance learning occurred halfway through the course in 2020. It must be underlined that, although the first exam session is normally held immediately after the end of the course (late March-early April), the opening of exam sessions in 2020 was deferred to May, because of technical issues. The number of students sitting the first session is more than 70% of students taking the exam at least once in a whole year. Most of the students choosing the first session attended regularly the same year the lessons and thus suffered the transition and experimented the new online examination platform for the first time. For this reason the preliminary focus is on the results attained at first sessions only.

The data used in this paper can be treated as quasi-experimental, since students had no possibility of opting for different courses and just happened to be enrolled in 2019-20. Using data on the results in the Statistics course, this work tries to address the following research questions: did the COVID-19 emergency impact on students performance? In particular: (i) did it have an effect on the expected grade? (ii) Did students experience a higher probability of failure or more generally of underperforming?

Section 2 describes the data and displays some summary statistics. Section 3 presents the main results obtained by estimating different regression models.

Specifically, Section 3.1 focuses on measuring the effect of Covid year on the expected grade, therefore on answering question (i). Although, for the aforementioned reasons, the first exam session is the one where it is most likely to observe a significant Covid effect, it is convenient to use data from all the exam sessions. In fact, it is clear that the difference in performances of students of two consecutive years could be simply due to a cohort effect. This threat to identification is addressed by controlling for a class composition effect, using data from all sessions of the two years (6 sessions per year) and variability between sessions. In particular, this is done through the introduction of the odds-ratio of Italian vs non Italian students attending each session. Further, since the sample includes results from 282 exams for 164 students only (due to retakes), by using a linear mixed-effects model specification, it is also possible to control for unobserved individual effects.

Section 3.2 instead addresses question (ii): the focus is, in this case, in trying

to understand whether the *treated* group (students who took their exam in 2020) experienced different probabilities to fail the exam, or to perform below or higher certain thresholds. The approach followed is to define binary variables associated to different grade levels and to estimate an independent probit model for each one of them. In particular, the chosen levels correspond to the four events: {exam not passed}, {grade below 25-th percentile}, {grade below median} and {grade above 75-th percentile}. With a larger dataset and by defining a finer grid for the thresholds, this approach would enable to compute a semiparametric estimate of the whole conditional distribution of the dependent variable Grade, via the so-called Distributional regression (see Foresi and Peracchi (1995) or Chernozhukov et al. (2013)).

2. DATA

The data cover all students of a Statistics course from an Italian Master who took part to the first exam session in 2019 or 2020 (also extended to all the sessions of both years). Besides the performance at the exam, the dataset includes some demographic information, and their whole academic track record from the enrollment until May 2020.

A descriptive analysis on demographics shows that the percentage of male students sitting in the first session in 2019 and 2020 is 42.7% and 38.8% respectively. The distributions of grades, among those who passed the exam, are quite similar in the right tail (see Figure 1): they both show a bimodality with the higher peak in the last interval (> 28) and the fraction of students taking grades higher than 27 is 0.44 in 2019 and 0.4 in 2020, although the pattern in the left tail tends to differ. A Kolmogorov-Smirnov test for the comparison of the two distributions is not conclusive: the null hypothesis is not rejected if data from first sessions are considered (p-value equal to 0.4 for all grades and equal to 0.86 for grades larger than or equal to 18 only), but the result changes as soon as one or more sessions are added.

The portion of students who failed the exam (including withdrawals) are quite similar if one considers the first sessions, being 0.31 in 2019 and 0.33 in 2020. When all exam sessions are taken into account, though, the difference between years increases, with an odds ratio of failure in 2020 vs 2019 approximately equal to 1.86. The exam outcome is strongly dependent on nationality, with a significantly larger portion of international students failing the first session: about 59.5% of international students failed or withdrew from the first session, vs 21.5% of Italians (see Table 1).

This difference is not mitigated if the exams of the 12 sessions (that however

include retakes) are taken into account: only 37, out of the 109 exams registered as "fails or withdrawals", correspond to Italian students. This gap is confirmed by higher average grades (average of all grades above the passing threshold) and grade point average (GPA) of Italian students. There is also a marked difference in gender (except for GPA), that however tends to disappear when all sessions are considered.

Table 2 summarizes the main descriptive statistics of most of the variables used for the analysis. The summary refers to the 125 students who sat at the first session and didn't withdraw from the exam (for which the variable grade is unobserved), separated by year. The variable ECTS stands for European Credit Transfer and Accumulation System, and, similarly to the GPA, it is computed from all other exams passed until the beginning of the pandemic. Of students taking the exam in the fist session of 2019 or 2020, about 70.4% is Italian (81.1% in 2019 vs 62.5% in 2020). The variable Grade is the final grade assigned. A binary variable is assigned equal to one if the grade is 30 *cum laude* (*A+*). A value of Grade below 18 means failure of the exam.

Tab. 1: Fractions of failed exams, average grade (of passed exams only) and GPA, by year, gender and nationality, on the first session only (withdrawals are counted as "Fails")

	Year		Gender		Nationality	
Result	2019	2020	Male	Female	Non-Italian	Italian
N	58	72	49	81	37	93
Failed	0.31	0.33	0.27	0.36	0.59	0.22
Mean Grade	25.8	20.1	25.8	25.5	24.1	25.9
Mean GPA	26.4	27.6	27.4	26.9	25.5	27.7

Tab. 2: Descriptive statistics first session exams in 2019 and 2020

	Year = 2019			Year = 2020			Total		
Statistic	N	Mean.	St.Dev	N	Mean.	St. Dev	N	Mean	St. Dev
Grade	53	22.547	6.941	72	20.083	8.862	125	21.128	8.165
Female	53	0.585	–	72	0.639	–	125	0.616	–
Italian	53	0.811	–	72	0.625	–	125	0.704	–
Age	53	24.019	1.896	72	24.750	2.915	125	24.440	2.551
ECTS	53	66.962	16.670	72	17.000	7.817	125	38.184	27.678
GPA	53	26.428	1.879	72	27.548	2.616	125	27.073	2.390

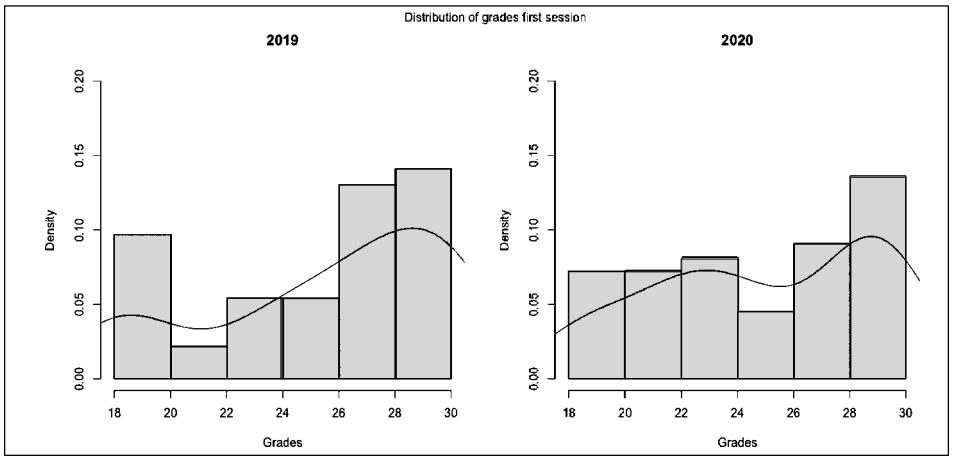


Fig. 1: Histogram and estimated densities of the grades of 2019 and 2020

Table 3 reports the same descriptive statistics for all the exam sessions of the two years. Students with Grade equal to zero were counted as withdrawals. A comparison of the two tables brings out some characteristics of students sitting the first session: on average younger, with higher GPA and with a higher fraction of Italians.

Tab. 3: Descriptive statistics all session exams in 2019 and 2020

	Year = 2019			Year = 2020			Total		
Statistic	N	Mean.	St. Dev	N	Mean.	St. Dev	N	Mean	St. Dev
Grade	109	22.600	6.920	173	20.000	7.250	282	21.000	7.230
Female	109	0.587	–	173	0.63	–	282	0.613	–
Italian	109	0.743	–	173	0.497	–	282	0.592	–
Age	109	24.400	2.390	173	25.400	3.510	282	25.000	3.160
ECTS	109	66.500	18.200	173	19.600	13.300	282	37.700	27.500
GPA	109	26.000	1.940	173	26.900	2.780	282	26.600	2.520

3. MODELS AND RESULTS

3.1 MEASURING THE EFFECT ON THE EXPECTED GRADE

This section is focused on trying to measure the effect of the pandemic on students' average performance. The performance measure is the grade obtained at the exam.

The baseline specification is the linear model of the exam output over the variable Covid, a few demographic controls, including gender, age and a dummy for nationality (Italian or not), and controls for ability. Covid is a dummy equal to one if the exam was taken in 2020.

The general specification has the form

$$Y_i = \beta_0 + \gamma \text{Covid}_i + \beta^T X_i + \varepsilon_i,\ \varepsilon_i \sim iid(0, \sigma^2) \qquad (1)$$

where *Xi* incorporates demographic variables and the combination of one or more of the ability measures defined below. Note that, in the case of a single session, heteroscedasticity robust standard errors are computed for the OLS estimates. Further, the independent errors assumption is no longer invoked when the data used are from multiple exam sessions and repeated observations for the same students occur.

To control for student's academic ability, different measures are proposed, mainly based on the overall mean grades and the number of exams taken in a certain period of time. Specifically, these include the variables GPA, defined as the standardized mean grades taken in all other exams until a certain date, ECTS, the sum of accumulated ECTS until the pandemic (with a correction for seniority). This last variable, however, is omitted because it is dominated by GPA. Other two dummy variables are considered as proxies for students' specific skill in statistics. The first one is FirstYrTakers: a dummy for students who did the exam in their first year. This variable is included because students who find the exam particularly hard to pass are more likely to postpone or repeat it. The second variable, Stats, is a dummy equal to one if the student passed a statistics exam during the bachelor.

Table 4 shows the estimates of some selected linear model specifications. The total number of exams considered is 121: 4 outliers, corresponding to blank exams, were counted as withdrawals and omitted from the sample.

The first column includes the variable of interest (Covid), the dummy for Italian, that was identified as a relevant factor in Section 2, together with GPA and their interaction. The other two models displayed in Table 4, models (2) and (3), differ from (1) by the inclusion of age (not significant) and of the squared GPA (qGPA = $(GPA)^2$) and, in model (3) only, the two dummies FirstYrTakers and Stats.

The interaction between GPA and Italian is highly significant in all models,

Tab. 4: Results from linear models, 1 session only

	Dependent variable: Grade		
	Models		
	(1)	(2)	(3)
Covid	– 3.364***	– 2.340**	–2.087*
	(1.058)	(1.135)	(1.222)
Age	– 0.387	– 0.358	
		(0.360)	(0.367)
Italian	3.889 **	3.120	3.259
	(1.834)	(2.144)	(2.217)
GPA	1.460	– 1.178	– 1.360
	(1.191)	(1.885)	(1.929)
qGPA		– 1.559**	– 1.643**
		(0.696)	(0.717)
FirstYrTakers			1.100
			(2.294)
Stats			1.555
			(2.069)
Italian:GPA	3.774 ***	6.307 ***	6.410 ***
	(1.440)	(2.055)	(2.087)
Constant	20.187***	30.683***	27.295**
	(1.828)	(9.836)	(11.135)
Observations	121	121	121
R^2	0.378	0.428	0.432
Adjusted R^2	0.356	0.397	0.391

Note: *p<0.1; **p<0.05; ***p<0.01

and implies a nationality premium: according to column (1), a unit standard deviation increase in GPA is expected to improve (not significantly) the final grade of non Italian students by less than 1.5 points, while the final grade of Italian students improves by more than 5 points. This *nationality premium* could be a consequence of several factors, among which linguistic and cultural divide (most of the students are not native English speakers, and Italian speaking students are likely to have an advantage in understanding an Italian lecturer), or other difficulties related with studying abroad, such as sharing internet connection and workspace with roommates.

The models in columns (2) and (3) evidence the presence of a nonlinear effect of GPA, the inclusion of which significantly reduces the coefficient of Covid. According to these models, the variable GPA has a nonlinear effect on student's performance. Considering that GPA is standardized with scale 2.29 and center 27.15, from column (3)[2], one can measure the marginal effect of GPA. It is approximately equal to $-1.4 - 3.3 \times GPA$ for non Italian students: this is negative if GPA is larger than 26.2, while it is positive for lower GPA levels. For Italian students, the marginal effect is instead $5 - 3.3 \times GPA$, that is positive for all GPA ≤ 30. The model in column (2) is the result of stepwise selection from all potential covariates, but column (3) shows that, albeit not significant, the inclusion of Stats and FirstYrTaker determines a 10% drop of the coefficient of Covid. Thus, according to the two best models, students in 2020 have experienced a mildly significant (at least 10%) worsening of their performance, corresponding to approximately $2.1 - 2.3$ grade points.

As pointed out in the Introduction, focusing on the first exam session is interesting not only because the large majority of students attends the first exam session but also because it was the very first session following the lockdown in 2020 and the first one with the new online examination procedure. However, a major problem with the results in Table 4 is that they do not allow to disentangle the effect of Covid (that is just a year dummy) from a cohort effect. One thing that emerges from the analysis so far is that Italian students tend to have higher grades and a significant GPA premium relative to non Italians. Nationality-wise, there is a large difference in cohort composition in the two years, with a 19% of non Italian students in 2019 and almost twice (37.5%) in 2020. This difference could determine a cohort effect captured by the variable of interest. It is therefore useful to include in the model the ratio between Italian and non Italian students in each exam session. To do so, it is necessary to widen the analysis by considering a larger number of sessions for each year.

For this reason, the analysis is repeated on the set of all exam sessions of 2019 and 2020. The sessions are 6 each year: the second sessions were in May 2019 and June 2020, the other 4 sessions were in July, September, November and December respectively, in both years.

The number of students who attended the 12 sessions is 164 but, because of fails or retakes, the total number of observations is 309. As in the previous analysis, 21 withdrawals and 6 blank exams are excluded and the final sample consists of 282 exam results for 164 students.

[2] The results from column (2) are the same.

Besides allowing the identification of a cohort composition effect, the larger sample size should give more accurate estimates. We note that it is also possible to observe a mitigation of the Covid effect as time passes, because both students and universities adapted to the challenges of training online and the containment measures progressively loosened.

The structure of the data consists in an unbalanced longitudinal dataset, with repeated measures of the variable Grade throughout (at most) 12 exam sessions. Two approaches are used for the specification and estimation of the effect of Covid on student's performance: in the first case, a linear specification as in (1), for $i = 1, \ldots, 282$, with the inclusion, among the regressors, of the exam session, to account for a *within-year time effect*. In this specification, individual effects are not included, assuming they are captured by the individual specific regressors (like GPA and Italian). Two more variables can be added to the regressors considered in the first session analysis: the first one, Iratio, is the ratio of Italian vs non Italian students in the session and is used to identify cohort composition effect. Since this variable takes at most only 12 distinct values, it prevents the inclusion of time dummies into the model. The second variable, PrevGrade, is observed only for students who re-take the exam and is the grade obtained at the previous exam take.

Provided there are no residual omitted factors, the OLS are consistent and unbiased estimators, but robust standard errors must be computed to account for cross-sectional dependence due to repeated observations of the same units.

A second approach is also considered, that allows for the possibility of a residual individual effect. This effect is assumed to be uncorrelated with the regressors and with the errors and it is modeled through a linear random intercept model:

$$\text{Grade}_{ij} = \beta_0 + \beta_1^{\mathrm{T}} X_{ij} + \gamma_i + u_{ij} \tag{2}$$

where $i = 1, \ldots, 164$ refers to the student, while $j = 1, \ldots, n_i$ and n_i is the number of times student i took the exam in the 12 available sessions. The vector X_{ij} includes regressors time specific (session/year) and unit specific. The errors are Gaussian and conditionally independent on the random effects γ, also Gaussian.

Table 5 displays the results obtained from all sessions with the two approaches: the first two columns refer to the best OLS fits in the case of a linear regression, while columns (3) and (4) present the best fits in terms of AIC and BIC respectively, of the maximum likelihood (ML) estimates of the coefficient in equation (2). Robust standard errors for the OLS estimators are estimated following the approach of Driscoll and Kraay (1998), estimates of the linear mixed models are computed with the R packake lme4 (see Bates et al. (2015)).

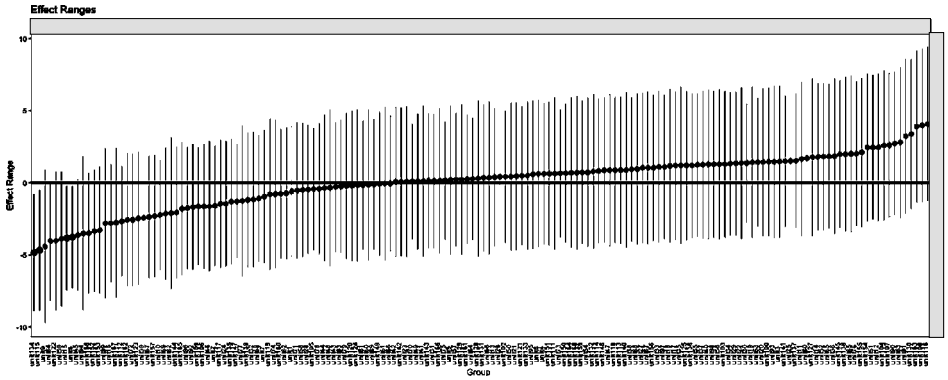


Fig. 2: Simulated random effects from posterior distributions, model (4) of Table 5

The coefficients of Covid in columns (1)–(4) are smaller, in absolute value, than those of Table 4, and range between – 1.46 and – 2. Thus, a student taking the exam in 2020 is expected to be graded at least about 1.5 points less than a student with similar characteristics who took the exam in the same session the year before. Differently from the first session data, not only Female, but also Age and $(GPA)^2$ are omitted because their inclusion does not alter significantly the estimates of the other coefficients, especially of Covid but worsens the fit. In all models displayed, the variable Session is a time variable, ranging from 1 to 6. Using section dummies in models that include Iratio decreases the estimated effect of Covid but increases the variance inflation factor.

Although the estimates of the Covid coefficient in the mixed effects models are larger (in absolute terms) than their OLS counterpart, they suffer a higher variance that induces higher p-value. However, an Anova-like test for random effects is conducted on both time (session) and students' effect and individual random effect is not excluded with $p - value = 0.011$ (see Figure 2).

Differently from Table 4, GPA remains strongly significant also when the interaction with Italian is added, while the interaction term it is only 10% significant in the mixe deffects models. The variables Iratio and PrevGrade are significant in all models and, unsurprisingly, have a positive effect on the expected performance.

Although non significant, the variables Stats and FirstYrTakers, if omitted, cause a significant change of the estimates of Covid, and for this reason the models in columns (2) and (4) should be preferred. In general, whilst some differences are observable, these two models give similar estimates for all the coefficients. This is also highlighted by comparing the average effects plots, from the OLS and the linear mixed effects model (LMEM), in Figure 3.

Tab. 5: Results from OLS and mixed models, all sessions

	Dependent variable: Grade			
			Models	
	OLS		*linear mixed-effects*	
	(1)	(2)	(3)	(4)
Covid	– 1.458*	– 1.743**	– 1.663	– 2.043*
	(0.845)	(0.710)	(1.122)	(1.210)
Session	0.321	0.361*	0.567**	0.621**
	(0.218)	(0.213)	(0.274)	(0.278)
PrevGrade	0.075**	0.064**	0.119***	0.112**
	(0.032)	(0.027)	(0.044)	(0.045)
Italian	4.092***	3.829***	3.919***	3.598***
	(0.745)	(0.905)	(0.936)	(0.997)
GPA	2.261***	2.412***	2.194***	2.358***
	(0.469)	(0.492)	(0.605)	(0.643)
Iratio	0.941**	0.948**	0.841*	0.842*
	(0.367)	(0.371)	(0.472)	(0.471)
FirstYrTakers		– 1.476		– 1.641
		(1.058)		(1.792)
Stats		0.893		0.851
		(0.956)		(1.279)
Italian:GPA	1.427	1.562	1.496*	1.650*
	(1.386)	(1.422)	(0.874)	(0.895)
Constant	16.374***	17.221***	16.574***	17.691***
	(1.311)	(1.572)	(1.835)	(2.761)
Observations	282	282	282	282
R2	0.377	0.380		
Adjusted R^2	0.361	0.359		
Log Likelihood			– 888.250	– 885.137
Akaike Inf. Crit.			1,796.499	1,794.274
Bayesian Inf. Crit.			1,832.918	1,837.977

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

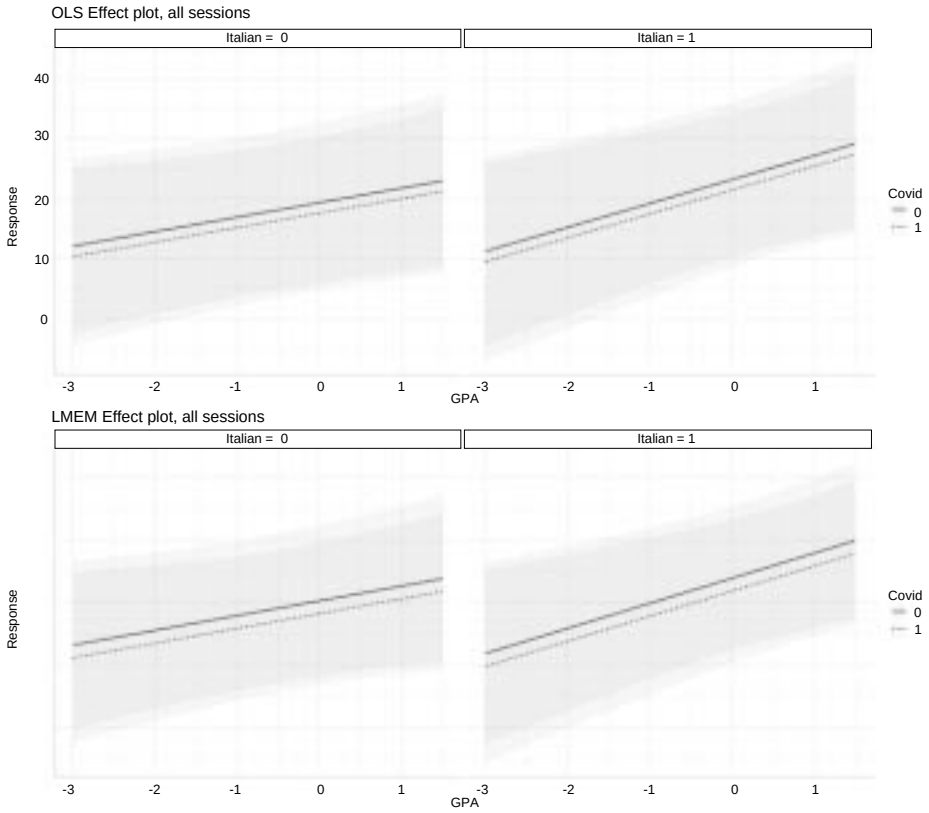


Fig. 3: Estimated marginal means of Grade computed from linear and linear mixed effects model, over GPA, Covid and nationality from models (2) and (4) of Table 5

3.2 MEASURING THE EFFECT ON THE CONDITIONAL DISTRIBUTION

The second question addressed in this work is measuring the effect Covid had on achieving different grade levels. In particular, the interest in this case is in determining whether students in 2020 were more likely to fail the exam or to perform below average, or to gain high scores, relative to 2019 students.

The focus is, therefore, on the estimation of the conditional cumulative distribution function (cdf) at different points of the support. The approach used in this work is to define H thresholds y_h, $h = 1,\ldots,H$ and, for each one of them, estimate a probit model for the binary dependent variable Grade $\leq y_h$:

$$P(Grade_i < y_h \,|\, X_i) = \Phi(X_i^T \beta_h),\ h = 1,\ldots,H,$$

where Φ is the standard Gaussian cdf. The regression vectors β_h are free to vary with the thresholds, and this entails a versatile approach, that could be used to estimate semiparametrically the whole conditional distribution of the dependent variable

(see Foresi and Peracchi (1995) or Chernozhukov et al. (2013)).

This flexible specification permits to account for non-constant effects of one or more covariates on Grade: for example, low-GPA students might suffer a higher effect on the probability of failing the exam, while high-GPA students are more likely to observe a stronger negative effect (if any is observed at all) on the probability to have higher grades.

As mentioned in the Introduction, the chosen thresholds are four ($H = 4$): the first one is the pass level $Grade = 18$, the other three are 25, 27 and 29. These three values correspond to the smallest integers above the first, second and third empirical quartiles of the variable GPA (not standardized), equal to 24.67, 26.9 and 28.6 respectively. Except for the third quartile, for which it is more interesting to estimate the probability $1 - P(Grade \geq 29 \mid X)$, the probabilities considered are all left tail probabilities. Therefore, the expected sign for the coefficient of Covid is positive for the first three thresholds, and negative for the last one.

Table 6 reports the results of the four models, obtained using the whole sample. As in the linear regression analysis, zero grades are treated like withdrawals and the session variable ranges between 1 and 6. The results based on the first session only are available on request. They are coherent with the ones in Table 6, except for the inclusion of the interaction term of Italian with GPA, that is instead omitted in all models of Table 6 because it is irrelevant.

In this case, due to students re-takes, the independence assumption of classical probit regression is implausible, therefore the estimates are obtained using the generalized estimating equation approach, that allows for dependence across units (see Halekoh et al. (2006), Yan and Fine (2004), Yan (2002)).

According to the estimates, the probability of failing the exam did not raise in 2020, once controlling for all other factors, on the contrary the coefficient of Covid is negative, albeit non significant, while there was a significant increase in the probability of grades below the median and below the first quartile (of the grade point average). Being an Italian student overcompensates the year effect, but this effect is smaller and only 10% significant on the probability of attaining an excellent grade. GPA is, unsurprisingly, always significant and has the expected sign, positive in the last column and negative in all other cases, although higher GPA has a smaller and only 10% significant effect on the probability of passing the exam. PrevGrade, on the other hand, has a positive and strong effect only on the probability to pass, clearly because the event of having failed before, that is the reason of almost all retakes, is strongly associated with lower grades. The time variable (Session) always tends to decrease the probability of a lower grade, although its effect is significant only for the Fail event.

Tab. 6: Results from probit models, all sessions

	Dependent variable:			
	Fail	(Grade<25)	(Grade<27)	(Grade ≥ 29)
	(1)	(2)	(3)	(4)
Covid	– 0.118	0.638**	0.543**	– 0.368
	(0.321)	(0.278)	(0.269)	(0.324)
Session	– 0.169*	– 0.049	– 0.016	0.019
	(0.089)	(0.078)	(0.079)	(0.080)
Italian	– 1.010***	– 0.591***	– 0.575***	0.428*
	(0.276)	(0.216)	(0.199)	(0.236)
GPA	– 0.094*	– 0.301***	– 0.286***	0.256***
	(0.054)	(0.051)	(0.049)	(0.066)
Iratio	– 0.190	– 0.166	– 0.170	0.089
	(0.184)	(0.111)	(0.126)	(0.137)
PrevGrade	– 0.057***	– 0.008	– 0.002	0.003
	(0.021)	(0.010)	(0.010)	(0.011)
Stats	– 0.215	0.155	0.418	– 0.335
	(0.322)	(0.333)	(0.298)	(0.346)
FirstYrTakers	– 0.618	0.426	0.358	– 0.315
	(0.489)	(0.410)	(0.411)	(0.477)
Constant	4.010***	8.280***	7.970***	– 7.610***
	(1.360)	(1.310)	(1.240)	(1.740)
Observations	282	282	282	282
Note:				*p<0.1; **p< 0.05; ***p<0.01

In order to have a clearer picture of the actual effects of the variable of interest on the four probabilities, Figure 4 plots the estimated marginal means, computed over the variable Covid (called Year in the figure, dashed line is 2020, solid line is 2019) and also over GPA and nationality (left panels correspond to non Italians, right panels to Italian students). Picture A shows that, irrespective of the year, Italian students Fail the exam with a probability roughly from 25% (high GPA) to 35% (low GPA) lower than non Italian students.

Nationality does not play such a strong role in the other three cases. In general, Italian students with higher GPA have lower probabilities that their grades fall below the threshold relative to their non Italian counterparts. However, for the two intermediate thresholds (Grade below 25 or 27), low GPA Italian students seem

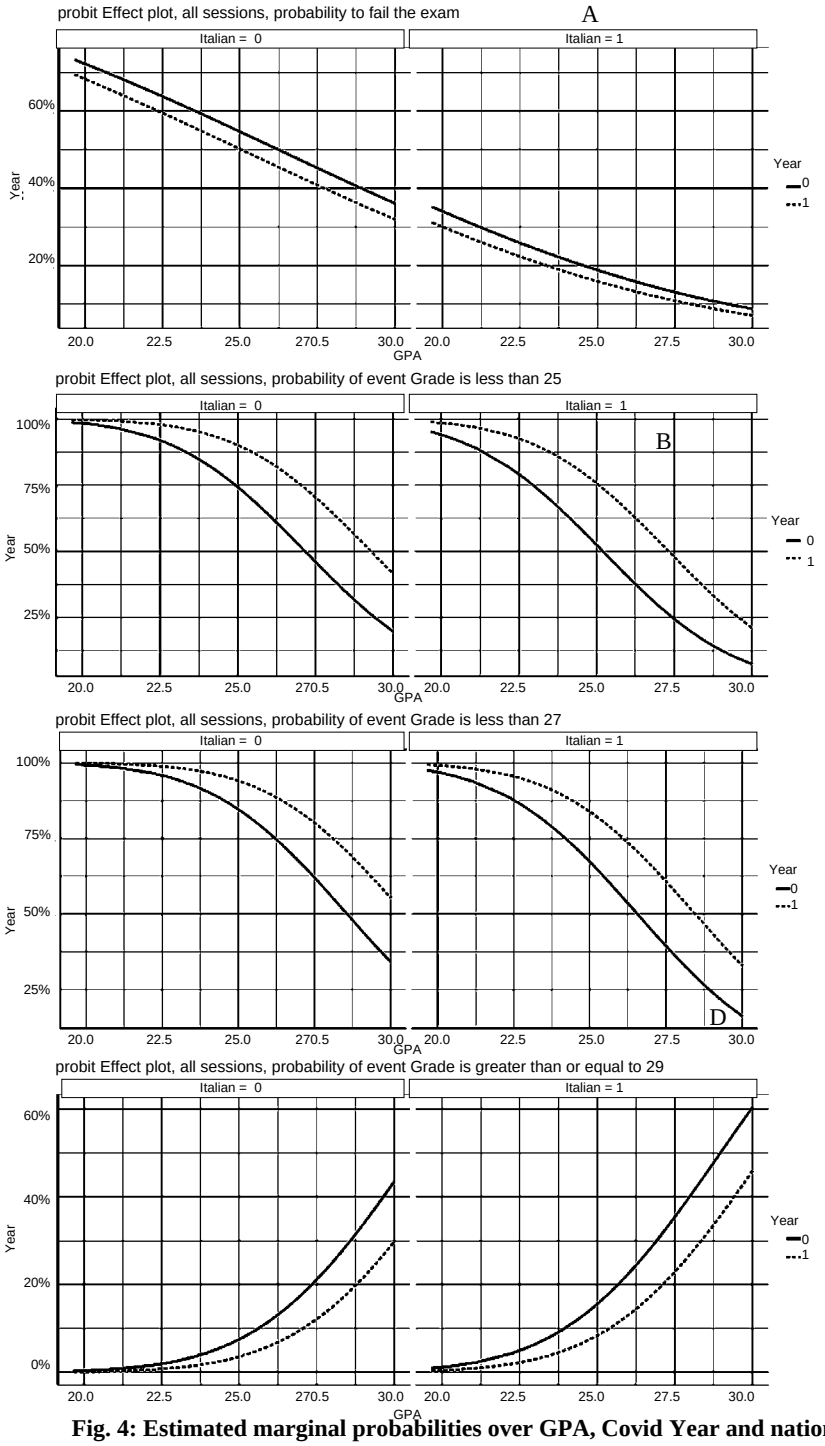


Fig. 4: Estimated marginal probabilities over GPA, Covid Year and nationality

to have suffered a stronger Year effect than their non Italian peers. The average Year effect, that is, the distance between the solid and dashed lines, tends to increase with GPA, except in the probability of Fail (panel A). In the case of Italian students, however, the average Year effect seems to peak when the GPA is approximately 2 points larger than the threshold y_h and after that level the distance between the solid and dashed curves is slowly reduced (see graphs B and C). A similar effect is not visible for non Italian students.

4. CONCLUSIONS

The aim of this paper is to measure the effect of the lockdown imposed by the pandemic and the consequent transition of the universities to teaching in distance and online exams. The variable of interest refers to the year when all this changes occurred and it therefore incorporates all factors falling under the same umbrella of a somewhat general *Covid effect*, such as psychological distress or technological and logistic difficulties (slow or unstable wifi connection, lack of comfortable space for studying or taking tests). The results show in particular a mildly significant negative effect of giving the exam in the year of the pandemic.

From the linear models for the conditional mean, coefficients associated to Covid are significant and suggest that taking the exam after the Covid's breakdown had a negative effect on the performance, especially in the first exam session. According to the specifications (2) and (3) of Table 4, the measured effect corresponds to a reduction of the final grade by 29-32% of standard deviation, while the coefficients in Table 5 range between 0.20 (column (1) OLS specification) and 0.28 (column (4) Mixed-effects model) times the standard deviation (equal to 7.23 throughout the 12 sessions).

Although smaller, these effects are essentially in line with the effect of online courses in Bettinger et al. (2017), who found an expected reduction of one third of the standard deviation.

The inclusion of the variable Iratio, when data from all sessions are available, permits to account for class composition and thus to eliminate part of the effect of the variable Covid that is due to cohort differences. For this reason, the OLS coefficients in Table 5 are significantly smaller, in absolute value, than those from Table 4.

Mixed-effects models are able to account for potential students' random effects that are not captured by the other regressors; this affects the estimates of Covid effect, subject to an increase in absolute terms and in the standard deviation.

It must be pointed out that time is a factor that could possibly mitigate the

Covid effect: students got used with the new online examination procedure and, at the same time, the health situation improved and the quarantine measures were gradually relieved. This is also suggested by the time variable Session, that has a positive and 5% significant effect in the mixed-effects models.

The effect that Covid year had on the conditional distribution of the final grade is also estimated. Four binary variables $D_h = \{\text{Grade} < y_h\}$ are defined at different threshold levels y_h and independent probit models are estimated corresponding to each threshold (D_h, for $h = 1, 2, 3$ and $1 - D_4$ thus allowing for the effects of Covid and of the other covariates on the conditional distribution to vary across the support. The heterogeneity of Covid year effect is highlighted in Figure 4, where the average marginal effects on the cumulative conditional distribution function of Grade, over Covid, nationality and GPA are computed, for the four levels.

The results show no evidence of an effect of the probability to fail the exam, while it seems that the pandemic increased the probability to underperform: for example, students with high GPA (29 or above) experienced a higher probability to get a grade below 27 (panel C), with an increase by about 20 percentage points relative to 2019. This difference is particularly relevant for Italian students, for which the probability $P(\text{Grade} < 27 \mid X)$ almost doubled going from 0.25 to 0.45 (for students with GPA $\approx$ 29).

The main limit of this paper is the sample size used for the analysis, which does not allow to obtain decisive results, especially once the class composition effect is taken into account by the inclusion of the variable Iratio.

It would be of interest to identify the contributions of the different factors entering the broad definition of Covid effect used in this paper, such as psychological or practical issues, but this is unfortunately impossible with the available data. Further investigation on a larger sample of students, from several courses from different faculties would allow a deeper insight of the effects of the disruption caused by the pandemics, on alternative measures of student's achievements (not only grades, but also the number of exams in a semester, delayed graduation time, etc…) and its persistence through time.

REFERENCES

Alpert, W.T., Couch, K. A. and Harmon, O.R. (2016). A randomized assessment of online learning. *American Economic Review*, 106(5): 378–82.

Andrew, A., Cattan, S., Costa Dias, M., Farquharson, C., Kraftman, L., Krutikova, S., Phimister, A. and Sevilla, A. (2020). Inequalities in children's experiences of home learning during the covid-19 lockdown in england. *Fiscal Studies*, 41(3): 653–683.

Aucejo, E.M., French, J., Ugalde Araya, M.P. and Zafar, B. (2020). The impact of covid-19 on student experiences and expectations: Evidence from a survey. *Journal of Public Economics*, 191: 104271.

Bates, D., Mächler, M., Bolker, B. and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1): 1–48.

Bettinger, E.P., Fox, L., Loeb, S. and Taylor, E. S. (2017). Virtual classrooms: how online college courses affect student success. *American Economic Review*, 107(9): 2855–2875.

Bowen, W.G., Chingos, M.M., Lack, K.A. and Nygren, T.I. (2014). Interactive learning online at public universities: Evidence from a six-campus randomized trial. *Journal of Policy Analysis and Management*, 33(1): 94–111.

Cao, W., Fang, Z., Hou, G., Han, M., Xu, X., Dong, J. and Zheng, J. (2020). The psychological impact of the covid-19 epidemic on college students in china. *Psychiatry Research*, 287: 112934.

Chernozhukov, V., Fernández-Val, I. and Melly, B. (2013). Inference on counterfactual distributions. *Econometrica*, 81(6): 2205–2268.

Crawford, J., Butler-Henderson, K., Rudolph, J. and Glowatz, M. (2020). Covid-19: 20 countries' higher education intra-period digital pedagogy responses. *Journal of Applied Learning and Teaching*, 3(1): 1–9.

Dietrich, H., Patzina, A. and Lerche, A. (2021). Social inequality in the homeschooling efforts of german high school students during a school closing period. *European Societies*, 23: S348–S369.

Driscoll, J. and Kraay, A. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *The Review of Economics and Statistics*, 80(4): 549–560.

Figlio, D., Rush, M. and Yin, L. (2013). Is It live or is It internet? Experimental estimates of the effects of online instruction on student learning. *Journal of Labor Economics*, 31(4): 763–784.

Foresi, S. and Peracchi, F. (1995). The conditional distribution of excess returns: An empirical analysis. *Journal of the American Statistical Association*, 90(430): 451–466.

Gonzalez, T., de la Rubia, M.A., Hincz, K.P., Comas-Lopez, M., Subirats, L., Fort, S. and Sacha, G.M. (2020). Influence of covid-19 confinement on students' performance in higher education. *PLOS ONE*, 15(10): 1–23.

Halekoh, U., jsgaard, S.H. and Yan, J. (2006). The r package geepack for generalized estimating equations. *Journal of Statistical Software*, 15/2: 1–11.

Joyce, T., Crockett, S., Jaeger, D.A., Altindag, O. and O'Connell, S. D. (2015). Does classroom time matter? *Economics of Education Review*, 46: 64 – 77.

Kapasia, N., Paul, P., Roy, A., Saha, J., Zaveri, A., Mallick, R., Barman, B., Das, P. and Chouhan, P. (2020). Impact of lockdown on learning status of undergraduate and postgraduate students during covid-19 pandemic in west bengal, india. *Children and Youth Services Review*, 116: 105–194.

Lee, J. (2020). Mental health effects of school closures during covid-19. *The Lancet Child & Adolescent Health*, 4(6): 421.

Maldonado, J. and De Witte, K. (2020). The effect of school closures on standardised student test. *FEB Research Report Department of Economics, KU Leuven, Faculty of Economics and Business*.

Murphy, L., Eduljee, N.B. and Croteau, K. (2020). College student transition to synchronous virtual classes during the covid-19 pandemic in northeastern united states. *Pedagogical Research*, 5(4), em0078.

Odriozola-Gonzàlez, P., Planchuelo-Gomez, A. Irurtia, M.J. and de Luis-Garcà, R. (2020). Psychological effects of the covid-19 outbreak and lockdown among students and workers of a spanish university. *Psychiatry Research*, 290: 113108.

OECD (2020). Learning remotely when schools close: How well are students and schools prepared? insights from pisa. *OECD Policy Responses to Coronavirus (COVID-19). OECD Publishing, Paris, https:// doi.org/ 10.1787/3bfda1f7-en.*

Pietro, G., Biagi, F., Costa, P., Karpinski, Z. and Mazza, J. (2020). The likely impact of covid-19 on education: Reflections based on the existing literature and recent international datasets. JRC Technical Report.

Sahu, P. (2020). Closure of universities due to coronavirus disease 2019 (covid-19): impact on education and mental health of students and academic staff. *Cureus*, 12(4).

Scherer, R., Howard, S. K., Tondeur, J. and Siddiq, F. (2021). Profiling teachers' readiness for online teaching and learning in higher education: Who's ready? *Computers in Human Behavior*, 118: 106675.

Toquero, C. M. (2020). Challenges and opportunities for higher education amid the covid-19 pandemic: The philippine context. *Pedagogical Research*, 5(4).

Yan, J. (2002). geepack: Yet another package for generalized estimating equations. *R-News*, 2/3: 12–14.

Yan, J. and Fine, J. P. (2004). Estimating equations for association structures. *Statistics in Medicine*, 23: 859–880.

Zalite, G.G. and Zvirbule, A. (2020). Digital readiness and competitiveness of the eu higher education institutions: The covid-19 pandemic impact. *Emerging Science Journal*, 4: 297–304.

Zimmermann, M., Bledsoe, C. and Papa, A. (2021). Initial impact of the covid-19 pandemic on college student mental health: A longitudinal examination of risk and protective factors. *Psychiatry Research*, 305: 114254.

SPLIT-PLOT DESIGNS AND MULTI-RESPONSE PROCESS OPTIMIZATION: A COMPARISON BETWEEN TWO APPROACHES

Rossella Berni[1], **Lorenzo Piattoli**
Department of Statistics, Computer Science, Applications "G. Parenti", University of Florence, Italy

Christine M. Anderson-Cook
Los Alamos, New Mexico, USA

Lu Lu
Department of Mathematics and Statistics, University of South Florida, Florida, USA

Abstract: *Nowadays split-plot designs play a crucial role in technological fields, both for their flexibility when applying a robust design approach and in relation to the modelling step, by considering mixed Response Surface models and/or the class of Generalized Linear Mixed Models (GLMMs). In this paper, a split-plot design is studied in a process optimization scenario involving several response variables, a multi-response situation, where two optimization methods are compared. More precisely, by considering a real case study related to the improvement of a measurement process of a Numerical Control machine for measuring dental implants, the optimization is carried out with the Pareto front approach and then compared with an analytical optimization method obtained starting from the definition of a risk function. In the final discussion advantages and disadvantages of application for both methods are evaluated.*

Keywords: *split-plot designs, multi-response process optimization, Pareto front approach, analytical optimization methods*

1. INTRODUCTION

Process optimization is a key issue for statistical quality control, and its relevance has increased since the long and fruitful scientific debate related to Taguchi's two-step procedure for robust design (Nair, 1992). Currently, the robust design approach involves three key-steps: experimental design, modelling, and optimization. For a successful implementation, it is important to

[1] Corresponding author: Rossella Berni, email: rossella.berni@unifi.it

involve noise factors starting from the experimental planning, and then modelling them through a suitable analysis in the subsequent optimization phase. Therefore, the concept of process optimization is extended to robust process optimization, where control and noise variables are jointly studied to attain the best set of control factor levels that simultaneously achieves the target value and minimize the process variability with a robust configuration. In this context, the process optimization strictly depends on the designed experiment and on the class of statistical models applied. Specifically, Response Surface Methodology (RSM) approaches (Myers et al., 2016) may be used, or, alternatively, an experimental design may be planned outside the RSM context, modelling the experimental data with a more flexible class of statistical models, e.g. Generalized Linear Models-GLMs (McCullagh and Nelder, 1989; Nelder and Lee, 1991; Lee and Nelder, 2003) or Generalized Linear Mixed Models-GLMMs (Dror and Steinberg, 2006; Robinson et al., 2006; Berni and Bertocci, 2018).

Undoubtedly, the two choices of experimental design and modelling are related to what is known about the process to be studied and optimized; the same line of reasoning should be applied when deciding on which, how many, and the type of response variables have to be considered. In fact, the multi-response situation should be evaluated when the real scenario shows that several response variables are naturally involved, and they are important to the overall process under study. If this is not the case, then a simpler choice should be used, since collinearity, often present among responses (Box et al., 1973; Chiao and Hamada, 2001), can lead to serious complications both in analysis and optimization.

Moreover, in a multi-response situation, the optimization step is crucial, since it is generally not feasible in practical terms to reach an ideal optimum simultaneously for all the responses. In literature there are several methods for achieving optimal solutions as a compromise among several response variables, including Derringer and Suich (1980), Khuri and Conlon (1981), Del Castillo et al. (1996), Copeland and Nelson (1996).

A further issue concerns the conjunction of a multi-response case and the dual response approach. The simultaneous optimization of several variables, jointly with the consideration of two statistical models, e.g., location and dispersion, complicates the analysis and interpretation of results. Indeed, we should make a distinction when considering the differences between the dual response approach, or, alternatively, when a single response model, opportunely weighted with respect to the estimated dispersion values, is applied in a "true" multi-response case. In the latter case, the application of analytical methods for

optimizing can give fruitful and satisfactory results, particularly by considering recent developments involving noise factors.

To this end, several optimization measures are suggested in the literature (Lin and Tu, 1995; Tang and Xu, 2002). The Pareto front approach is a multi-response optimization method of the analytical-qualitative type, consisting of two sequential steps: a first step based on objective conditions identifies the dominant solutions, and a second step based on examining the identified alternatives and then selecting the best solution conditions that subjectively match the experimenters' priorities (Lu et al., 2011; Chapman et al., 2014a).

In this paper, data from a split-plot experiment (Berni, 2010) are optimized through the Pareto front approach (Chapman et al., 2014b), and following, the results obtained are compared with a proposal of the analytical optimization method. More specifically, both methods are compared and discussed through an empirical example in the orthodontic field, in order to improve the accuracy in the measurements of a Numerical Control (N/C) machine, which provides some automatic control of machining tools.

This paper is structured as follows: in Section 2, the basics of split-plot designs are reviewed and briefly illustrated. Section 3 provides a short description of both optimization methods, and Section 4 presents the case study including optimization results. The paper ends with a general comparison between the two approaches, and final remarks.

2. SPLIT-PLOT DESIGNS FOR STATISTICAL QUALITY CONTROL: A REVIEW

The split-plot design (Cochran and Cox, 1957) has been developed and characterized over the years, proving to be an experimental design widely used in industrial, technological, and environmental fields, because of the need to restrict randomization on expensive or hard-to-change factors in the experiment.

By considering the developments that the fractional factorial and RSM designs, and modelling have had since the 1980s, the split-plot design has experienced a particular renewal (Box and Jones, 1992), expounding its theoretical features, specific usefulness for the statistical quality control and robust design concepts, initially introduced by Genichi Taguchi, (Nair, 1992). In this context, the two seminal papers of Vining and Myers (1990) and Myers et al. (1992) extended the two-step procedure into the dual response approach, and the combined-array is considered as a milestone for recent developments and robust process optimization. Within this methodological framework, the split-plot design plays a central role, starting from the tutorial by Box and Jones

(1992), in which the authors proposed the split-plot design as an efficient alternative in many experiments, for example with hard-to-change factors, to Taguchi's product-array for a robust design approach, and also in a fractional factorial setting (Bisgaard, 2000).

Recently, split-plot designs have great relevance for the latest developments in robust process optimization, extending the initial concept of the robust design approach, with a focus on the design and modelling steps (Kowalski and Potcner, 2003; Kowalski et al., 2007; Jones and Nachtsheim, 2009). More recently, the split-plot design has been revised and included in the class of *crossed bi-randomized* experimental designs (Myers et al., 2016), given the possibility of including environmental/noise factors as Whole-Plot (WP) factors and process factors as Sub-Plot (SP) factors. The standard allocation of the environmental/noise factors as WP allows for the most accurate estimate of the factors of interest, as well as the estimate of the 1st order interactions, e.g., the 1st order interaction between a WP factor (for example a noise factor) with a SP (process) factor, in order to perform a robust design (Berni and Nikiforova, 2022). This structure is common in many applications, given the generally high cost of controlling the noise factors in production.

Nevertheless, a split-plot design in a RSM context generally requires that all the variables included in the experimental plan (irrespective of whether WP or SP factors) must be quantitative in nature. In fact, in case of a qualitative process variable, the optimization step is conditioned to the levels of the categorical variable involved. To this end, the inclusion of a qualitative variable should be limited (Berni, 2010) or restricted to two levels where they can be treated as quantitative in standard models. Moreover, the presence of measurable noise factors, involved as random effects, is possible when the split-plot design is applied through mixed Response Surface (RS) models, or alternatively, through GLMMs.

Additional developments in the literature have contributed significantly to the inclusion of the split-plot design in RSM, showing the equivalence of *Ordinary Least Squares* (OLS) with *Generalized Least Squares* (GLS) for split-plot designs and mixed RS models (Vining et al., 2005); and improving inference issues (Vining and Kowalski, 2008).

2.1. THEORY ABOUT THE SPLIT-PLOT DESIGN

When implementing a split-plot design, it is essential to begin with a primary classification between WP factors and SP factors. It is therefore desirable to carefully evaluate what is possible in the experimental set-up for the specific process (industrial process, laboratory experiment) under study,

considering the exact definition of the response (quantitative) variables. Also, it is necessary to define the role of each variable in the study, in order to plan the split-plot design based on the most efficient arrangement for the specific scenario. This step plays a central role, not only in the attribution of factors as whole-units or sub-units, but also determines the subsequent model estimation, in which each variable plays a specific role, according to its nature (qualitative, discrete quantitative or continuous). It is also a crucial point to clarify the distinction between fixed and random effects.

In short, a Whole-Unit (WU) is defined by runs involving the set of WP factors, and the run order of all the WUs is randomized. Subsequently, the Sub-Units (SUs), defined through the combination of levels (runs) of the SP factors within the WUs, are associated with a particular WU and randomized separately.

In this study, we consider a split-plot design in a RSM context. More precisely, we are interested in applying a standard second-order polynomial model with random effects. The model with random effects for the single experimental observation y_u $(u = 1, \dots, n)$ and J variables $(x_1, \dots, x_j, \dots, x_J)$ is (Khuri, 1996):

$$y_u = \beta_0 + \boldsymbol{f}'(\boldsymbol{x}_u)\boldsymbol{\beta} + \boldsymbol{z}_u'\boldsymbol{\gamma} + \boldsymbol{g}'(\boldsymbol{x}_u)\boldsymbol{\Delta}\boldsymbol{z}_u + \varepsilon_u \qquad (1)$$

where β_0 is the intercept; $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ is the column vector $[p\times 1]$ of the unknown fixed parameters ($p \geq J$); $\boldsymbol{x}_u = (x_{u1}, \dots, x_{uj}, \dots, x_{uJ})'$ is the vector of design settings at the u-th experimental run; $\boldsymbol{f}(\boldsymbol{x}_u)$ is a vector function of dimension $[p\times 1]$ defined for each $\boldsymbol{x}_u$ and related to the p second-order effects for the J variables. Therefore, $\boldsymbol{F}$ is the so-called "extended" matrix of dimension $[n\times p]$, formed by the n rows $\boldsymbol{f}'(\boldsymbol{x}_u)$; ε_u is the residual error. For the random effects, $\boldsymbol{z}_u = (z_{u1}, \dots, z_{ub})'$ is the vector of binary values $(0,1)$ to describe the presence and structure of the block factors; $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_b)'$ is the column vector $[b\times 1]$ of the unknown coefficients relating to the random effects. The matrix $\boldsymbol{\Delta}$ characterizes the 1st order interactions between polynomial effects (fixed) and random effects. The maximum dimension of $\boldsymbol{\Delta}$ is achieved if the interactions of all fixed effects with random effects are included in model (1). Note that this matrix contains the key estimated coefficients for evaluating the robust design as the control-by-noise interactions that can be exploited to achieve robustness.

Starting from the model expression (1), the second-order polynomial model of response surfaces from a split-plot design is outlined, considering quantitative variables only. For further details, see Myers et al. (2016).

Let's consider two sets of factors: $\boldsymbol{z} = (\boldsymbol{z_1}, \ldots \boldsymbol{z_i}, \ldots, \boldsymbol{z_I})$ are WP random factors/variables and $\boldsymbol{x} = (\boldsymbol{x_1}, \ldots \boldsymbol{x_j}, \ldots, \boldsymbol{x_J})$ are SP variables/factors.[2] Furthermore, let y_u be the u-th observation of the k-th block, for the i-th WP factor and the j-th SP factor respectively $(i = 1, \ldots, I; j = 1, \ldots, J; k = 1, \ldots, K)$; therefore, the second-order mixed RS split-plot model for a single replicate $(K = 1)$ and a single observation u is defined as follows:

$$y_u(\boldsymbol{z}, \boldsymbol{x}) = \beta_0 + \boldsymbol{z_i'\gamma} + \boldsymbol{z_i'\Gamma z_i} + \boldsymbol{x_{ij}'\beta} + \boldsymbol{x_{ij}'Bx_{ij}} + \boldsymbol{z_i'\Delta x_{ij}} + \psi_u + \varepsilon_u \qquad (2)$$

where β_0 is the intercept; $\boldsymbol{z_i} = (z_{i1}, \ldots, z_{iu}, \ldots, z_{in})'$ is the vector of the i-th WP factor; $\boldsymbol{\gamma}$ contains the unknown coefficients of linear terms of the WP variables; $\boldsymbol{\Gamma}$ is the array related to the coefficients of the 1st order interaction and quadratic terms of the WP variables; $\boldsymbol{x_j} = (x_{j1}, \ldots, x_{ju}, \ldots, x_{jn})'$ is the vector of the j-th SP factor; $\boldsymbol{\beta}$ contains the unknown coefficients of linear terms of the SP variables; $\boldsymbol{B}$ is the array containing 2nd order coefficients for the fixed effects of SP variables; the matrix $\boldsymbol{\Delta}$, contains coefficients of the 1st order interaction effects between the WP and SP factors. The model terms in the matrix $\boldsymbol{\Delta}$ are of primary interest in the context of robust design evaluation. Regarding the two error components, ψ_u is the WP error component and ε_u is the SP error component, where in general the two error components are assumed to be independent and identically Normally distributed, i.e. $\psi \sim iid\ N(0, \sigma_\psi^2)$ and $\varepsilon \sim iid\ N(0, \sigma_\varepsilon^2)$, respectively. In addition we also assume that $Cov(\psi_u, \varepsilon_u) = 0\ \forall u$. The assumptions about the error variances and covariances are equivalent to assume constant covariance between two observations belonging to the same WU, across all its observations.

In the case study illustrated in Section 4, the multi-response case is related to the optimization involving three split-plot models, one for each response, estimated applying the RSM model above.

3. OPTIMIZATION METHODS

This Section includes a short description of both optimization methods considered. The Pareto front approach (Lu et al., 2011; Chapman et al., 2014a; Chapman et al., 2014b) is also illustrated within the case study, considering the application (Subsections 4.2 and 4.3); a brief introduction of the analytical

[2] Please note that we are referring to a factor/variable considering that the experimental region χ is defined by the factor ranges; a finite number of experimental points, forming the experimental design, is then selected by the experimental region. Following, the model estimation is performed within the whole experimental region, by inferring from a discrete set of points, e.g., the experimental points, to a continuous one.

method is illustrated in Subsection 3.2. For further details see (Berni and Gonnelli, 2006; Berni, 2010; Berni and Burbui, 2014).

3.1. THE PARETO FRONT APPROACH

The Pareto front approach is a multi-response analytical-qualitative optimization method, which allows the search for optimum to take subjective priorities and constraints into account, such as those due to a company's requirements (for example, costs or technical/engineering specifications). It consists of two sequential steps (Chapman et al., 2014a; Myers et al., 2016; Anderson-Cook, 2017), as outlined below.

Suppose that χ defines the entire experimental region; within this region a finite set, possibly a grid, of points, is selected and used to estimate the responses of interest and used to define a Pareto-optimal set. A possible solution is called non-inferior (or Pareto-optimal), if and only if, there is no other combination within the set for which the values of all the responses are at least as good, and the value of at least one response is strictly better; otherwise, the solution is called inferior or dominated. The set of non-inferior (or Pareto-optimal) input combinations is called the Pareto-optimal set, and the corresponding set of vectors for the responses under consideration is known as the Pareto front or frontier. Since the inferior solutions are not rational choices conditional on the choice of responses under consideration, they are not considered further and definitively discarded (Zitzler, 1999; Marler and Arora, 2004; Coello Coello et al., 2007). This leads to a reduced number of alternative solutions to be considered further in later stages of the optimization.

The Pareto front approach can be summarized with the following two steps:

1. An objective step, where the Pareto-optimal set is identified from the initial set of choices, based on the corresponding estimated response values;
2. A subjective step, in which the points belonging to the Pareto-optimal set are examined and then compared. Only points that provide the best combination of responses are considered as a compromise among all the estimated response values (quantitative considerations). This choice is based on evaluation and incorporation of the priorities/preferences of the company.

It must be noted that several optimal points corresponding to input combinations could be considered and compared for selection, by considering the priorities of different teams (decision-makers) involved in the study. Therefore, the best optimal solution combines the quantitative results with the

decision-makers' priorities. Moreover, graphical methods are a useful tool for discussion, comparison and achieving a consensus among all stakeholders (Anderson-Cook and Lu, 2018).

3.2. THE ANALYTICAL METHODS FOR A ROBUST PROCESS OPTIMIZATION

When dealing with several response variables, it is generally not feasible in practical terms to simultaneously achieve the optimum for each of them with a single input combination. To this end, many authors, starting from the methods suggested by Derringer and Suich (1980) and Khuri and Conlon (1981), have proposed methods to synthetize and optimize the responses, such as Ames et al. (1997), Del Castillo et al. (1996), Rajagopal et al. (2005).

In addition, a further issue emerges when considering the multi-response case and the dual response approach. Here, the simultaneous optimization of several variables jointly with the consideration of two statistical models, e.g., location and dispersion models, increases the complexity and dimensionality of the problem.

In order to solve the latter issue, which could imply a notable computational burden, analytical optimization methods can be defined and simplified starting from the dual approach theory and the building of a suitable performance measure (Leon et al., 1987). To this end, we consider a multiplicative relationship between the expected value $\big(E(y) = \mu(\boldsymbol{x})\big)$ and the process variance $\big(Var(y) = \sigma^2(\boldsymbol{z}, \boldsymbol{x})\big)$ defined as the variance of the response variable. Moreover, the expected value of the response could be identified in relative to the target value (e.g., $E(y) = \tau$), according to the Nominal the Best (NTB), Smaller the Better (STB), or Larger the Better (LTB) situations. At the beginning a general risk function is expressed as follows:

$$R(\boldsymbol{z}, \boldsymbol{x}) = (\mu(\boldsymbol{x}) - \tau)^2 + f\big(\mu(\boldsymbol{x})\big)\sigma^2(\boldsymbol{z}, \boldsymbol{x}) \tag{3}$$

Formula (3) explicitly involves two terms: i) $(\mu(\boldsymbol{x}) - \tau)^2$ which expresses the adjustment to the target value, while ii) $f\big(\mu(\boldsymbol{x})\big)\sigma^2(\boldsymbol{z}, \boldsymbol{x})$ is related to the multiplicative relation between location (adjustment) and process variance (dispersion). Therefore, formula (3) allows for defining specific objective functions in a dual response approach perspective, (Berni and Gonnelli, 2006), and particularly, to optimize several response variables without separately estimating the two statistical models for each response. This approach provides a simplification as well as a weighting of the responses according to their relative importance.

More recently, split-plot designs and modelling have been optimized by explicitly involving one model only for each response in a robust process optimization context, in which random effects are also evaluated (Berni and Bertocci, 2018; Berni and Nikiforova, 2022).

Let's start by defining a general response surface model, y_t $(t = 1, \dots, T)$, for each of the T responses. The simultaneous optimization may be performed considering the T estimated surfaces; each estimated model is evaluated as a single function to be included in the objective function to be optimized. Starting by formula (3) and considering the concept of a dual response approach, a general objective function can be defined as the distance between the estimated surface $\hat{y}_t$ and the corresponding desired target value τ_t:

$$S_t(\boldsymbol{z}, \boldsymbol{x}) = (\hat{y}_t(\boldsymbol{z}, \boldsymbol{x}) - \tau_t)^2 \quad \forall t$$

The approach can be easily adapted for responses where the goal is to achieve a maximum or minimum value. Subsequently, the minimization on the coded experimental region χ is performed by minimizing the sum of all the distances, as follows:

$$\min_{\chi} \left\{ \sum_t S_t(\boldsymbol{z}, \boldsymbol{x}) \right\} \tag{4}$$

The objective function (formula (4)) is optimized conditional on the whole experimental region χ defined by the process variable ranges (and potentially any limiting constraints for other technological issues), as well as involving the estimated confidence interval for each random coefficient when random noise factors are present.

In the following section, we compare the two optimization methods, the Pareto front approach (Subsection 3.1), and the objective function of formula (4), where the goal is to improve the accuracy of the measurement process of a N/C machine, used in the orthodontic field; the measurements of which are analyzed for a generic dental implant.

4. THE CASE STUDY: DATA DESCRIPTION AND PROCESS OPTIMIZATION

In this Section, comparison of multi-response optimization methods is made, after a short description of the experimental planning and data. For further details see Berni (2010).

4.1. SPLIT-PLOT DESIGN AND DATA DESCRIPTION

The aim of the study is to improve the accuracy in measurements for a N/C machine, jointly with a reduction of the measuring time. The machine uses a feeler pin with a movable bridge framework to facilitate the positioning of the measured piece (dental implant). The machine needs specific environmental conditions to function properly, all of which were ensured previously (see Berni and Gonnelli, 2006).

In Berni (2010), five response variables, $T = 5$, were optimized simultaneously applying formula (4) and related to the different positionings of the feeler pin on the dental implant during the process measurement steps. In this paper we focus on the optimization comparison by looking at a subset of three response variables.

By considering the dental implant used to set the measurement process, the three responses are (with their respective targets in brackets): maximum circle diameter-$crmax$ $(\tau_1: 3.000)$, minimum circle diameter-$crmin$ $(\tau_2: 2.790)$, and eccentricity-$eccen$ $(\tau_5: 0.000)$. There is no problem with correlation among the three dependent variables, since each type of measurement is carried out as a distinct step; moreover, each response variable is independent from the others during the measurement of the piece. In order to reduce the measuring time, it is possible to intervene on the process phase related to the identification of the cone frustum, identified by three circles, at three different distances. In Figure 1, the frustum of cone is shown by highlighting the three circles used to locate it.

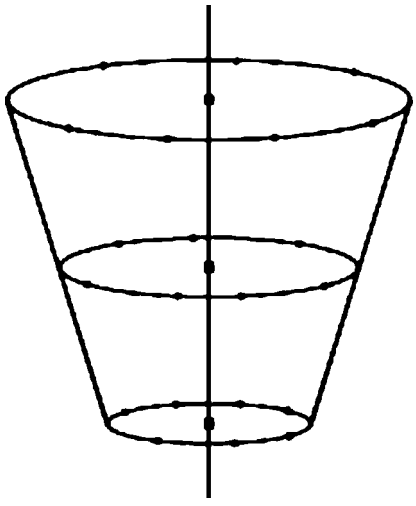


Fig. 1: Location of the frustum of cone by three circles and definition of the "circle-point" factor

In the initial setting, the N/C machine measures 7 points on each circle (7,7,7), each point is denoted with a dot in Figure 1. A categorical input factor "circle-point-cp" is then defined at four levels with each level corresponding to a different number of points measured on each circle: (1) 7,7,7; (2) 7,5,7; (3) 5,7,5; (4) 5,5,5.

Two other variables are involved in the split-plot design: measurement speed-$mspeed$ (mm/sec), and drift speed-$dspeed$ (mm/sec). Therefore, a split-plot design with three factors is planned: two WP process factors, both at two levels (measurement and drift speeds), and only one SP control factor, the cp categorical factor at four levels. The final split-plot has 112 runs with seven replicates.

Standardization of the responses was carried out (Berni, 2010) to compensate for differences in magnitude among responses, even though both responses and WP factors are expressed with the same unit of measurement.

4.2. THE PARETO FRONT APPROACH: OBJECTIVE PHASE

In order to identify the Pareto-optimal set, a series of 1764 combinations of the factor $mspeed$, $dspeed$ and cp levels were identified, from which, the predicted response values $\widehat{crmax}$, $\widehat{crmin}$ and $\widehat{eccen}$ were estimated using the model form described in formula (2). The set of possible input combinations (Figure 2) was formed by constructing a grid of points based on discrete levels of $mspeed$ and $dspeed$ for each level of the factor cp. The fineness of the mesh of each grid is 0.1, since this choice balances the complexity of calculation and valid coverage of the two-dimensional region, (ranges of $mspeed$ and $dspeed$ factors). The combinations of the possible solutions are labelled from 1 to 1764 according to the approach described in Chapman et al. (2014a): i) from the first grid on the upper left to the lower right grid; ii) inside each grid starting from the bottom row and moving from left to right, then starting at the end of each row, from the leftmost point of the next row.

The obtained Pareto-optimal set consists of 61 combinations, highlighted by the solid circles in Figure 2. These all combinations involve $cp = 4$, which requires the smallest number of measured points. Therefore, irrespective of the choice in the subjective phase, the Pareto front results ensure that an improvement in the measurement time is always obtained.

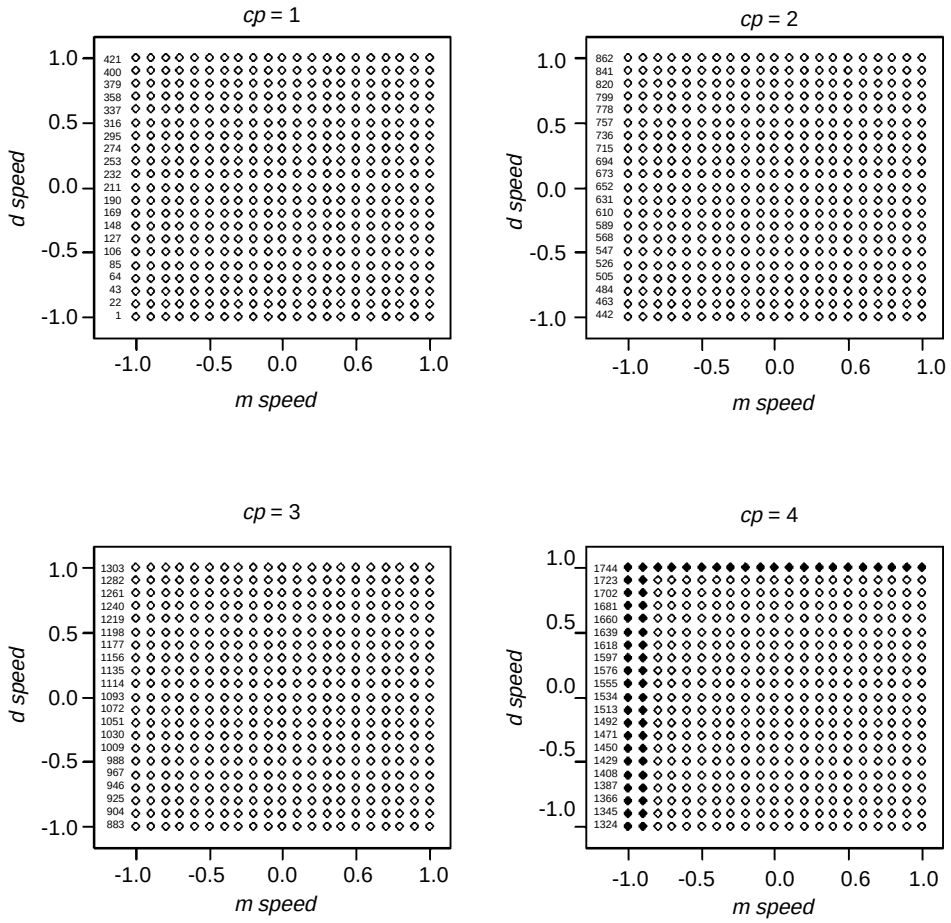


Fig. 2: Input grid plot

Figure 3 shows the pairwise scatterplots of the points belonging to the Pareto front (Chapman et al., 2014a; Anderson-Cook, 2017). The analysis of this set of points shows that there is a strong trade-off between the maximum circle diameter and eccentricity. The trade-offs between the other two pairs of response variables appear less important. Finally, by observing the relatively small ranges of the predicted responses, we note that all 61 combinations of the Pareto-optimal set lead to values of $\widehat{crmax}$, $\widehat{crmin}$ and $\widehat{eccen}$, close to the respective desired targets.

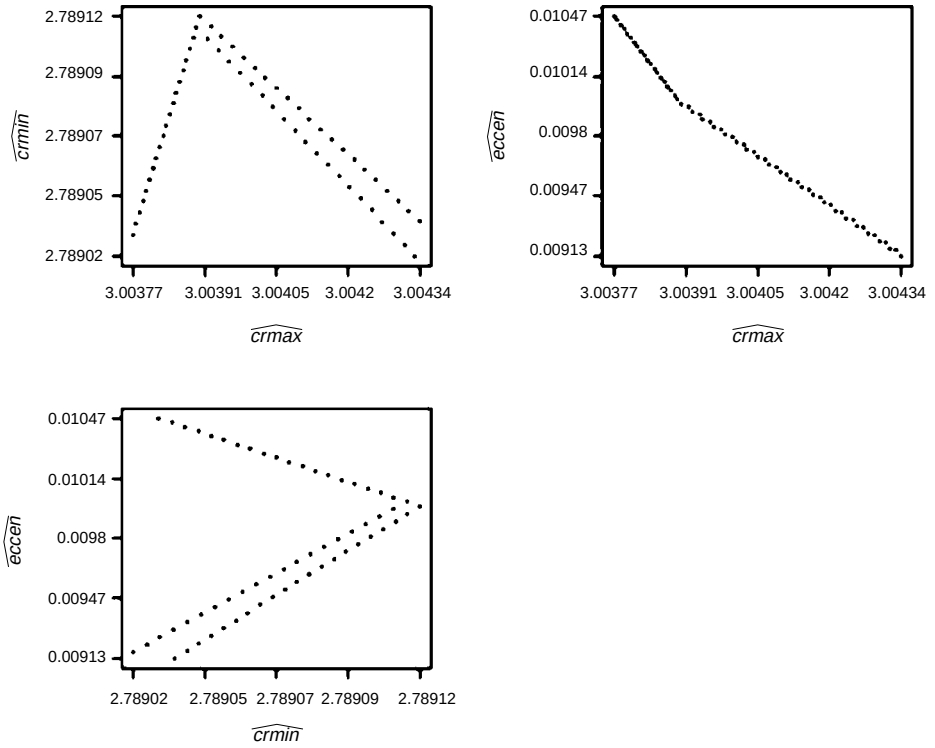


Fig. 3: Pairwise scatterplots of the points belonging to the Pareto front

4.3. THE PARETO FRONT APPROACH: SUBJECTIVE PHASE

In order to compare the 61 combinations of the Pareto-optimal set, the following procedure is carried out (Chapman et al., 2014a; Myers et al., 2016; Anderson-Cook, 2017): i) the Pareto front values of each predicted response are transformed into desirability values, so that the best value obtained (from the set of solutions comprising the front) for each response is scaled to one, while the worst value is scaled to zero; ii) for each combination of the Pareto-optimal set, the respective desirability values are combined in a single global desirability function. Since it was considered appropriate to heavily penalize undesirable predicted response values, we choose the standard multiplicative desirability form, based on the geometric mean expression, as follows:

$$D(\boldsymbol{x}_{\mathcal{P}}, \boldsymbol{w}) = d_{crmax}(\boldsymbol{x}_{\mathcal{P}})^{w_{crmax}} \, x \, d_{crmin}(\boldsymbol{x}_{\mathcal{P}})^{w_{crmin}} \, x \, d_{eccen}(\boldsymbol{x}_{\mathcal{P}})^{w_{eccen}}$$

where $\boldsymbol{x}_{\mathcal{P}}$ is a combination of the Pareto-optimal set; $d_{crmax}(\boldsymbol{x}_{\mathcal{P}})$, $d_{crmin}(\boldsymbol{x}_{\mathcal{P}})$,

$d_{eccen}(\boldsymbol{x}_{\mathcal{P}})$ the single desirability values related to the three predicted responses; $\boldsymbol{w} = (w_{crmax}, w_{crmin}, w_{eccen})'$ a weight vector, with $w_{crmax}, w_{crmin}, w_{eccen} \geq 0$ representing the weights assigned to the three response variables and $w_{crmax} + w_{crmin} + w_{eccen} = 1$. We note here that the small deviations of the response values from their desired targets mean that even small misses from the target are being strongly penalized, because one minimal error in measurements can lead to a serious risk for a patient.

Figure 4 shows the mixture plot, which identifies the best combination (i.e., the optimum point for achieving the highest value of the global desirability function) for each possible weighting of the response variables. Each point of the mixture plot represents a weight vector (e.g., the left bottom vertex represents $\boldsymbol{w} = (1,0,0)'$, the centroid marked with a black cross represents $\boldsymbol{w} = (1/3, 1/3, 1/3)'$, and the bottom edge represents the weight vectors with $w_{crmax}, w_{crmin} > 0$ and $w_{eccen} = 0$). For further details see Cornell (2002).

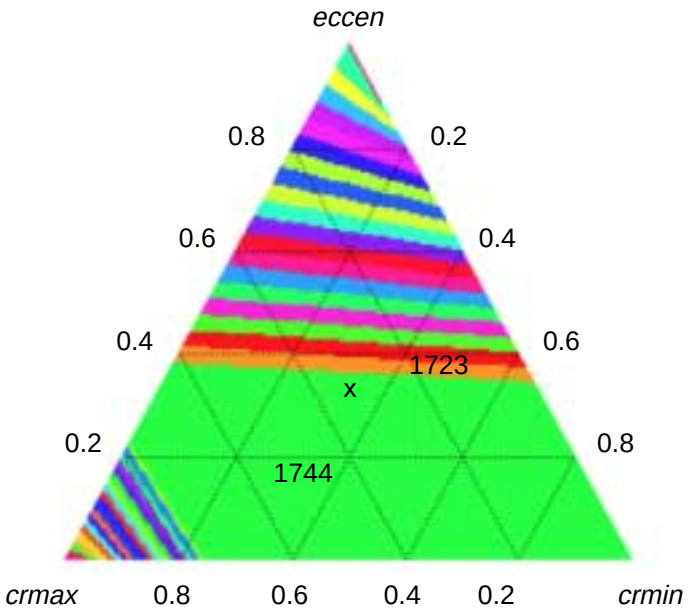


Fig. 4: Mixture plot

In the case study, 41 of the 61 combinations belonging to the Pareto-optimal set appear in the mixture plot (each colored area identifies a different combination), that is, they are best for at least one weight vector. Assuming that the three response variables are thought to be of equal importance, the weights reflecting company priorities/preferences are those around the centroid of the triangle. The two best points for these weight combinations are 1723 and 1744. In particular, 1744 is better for most of these weights, including the one directly at the centroid of the triangle as well. Table 1 shows the detailed results obtained for these two points, (1723 and 1744), differing only in the *dspeed* level value, and providing similar predicted responses.

Tab. 1: Factor levels and predicted response values for the combinations 1723 and 1744

Combination	Factors			Predicted responses		
	mspeed	*dspeed*	*cp*	$\widehat{crmax}$	$\widehat{crmin}$	$\widehat{eccen}$
1723	-1	0.9	4	3.00392	2.78911	0.00994
1744	-1	1	4	3.00390	2.78912	0.00998

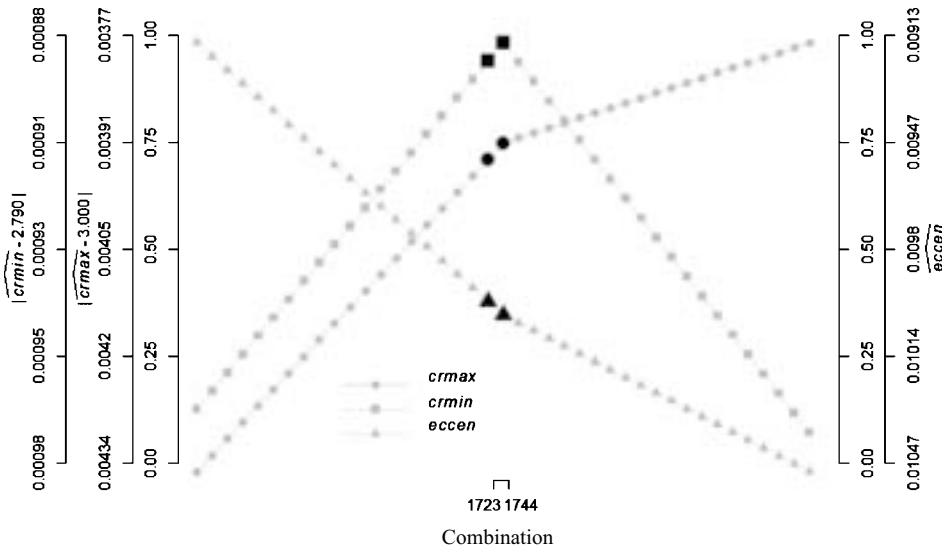


Fig. 5: Trade-off plot of the 41 best point combinations for at least one weight combination

Figure 5 contains the trade-off plot illustrating the desirability values (internal vertical axes) and absolute value differences between the predicted response values and the respective targets (external vertical axes) considering the 41 solutions that are best for at least one weight combination. In Figure 5, the trade-offs between the pairs of responses are similar to those highlighted by the pairwise scatterplots in Figure 3. Moreover, as shown in the mixture plot (Figure 4); the point combinations 1723 and 1744 provide an ideal balance among the three responses when they are all prioritized as being equally important.

In order to better analyze and compare these two combinations of interest, Figure 6 shows the synthesized efficiency plots (Lu and Anderson-Cook, 2012) which allow comparison of the relative efficiency of individual solutions with the best available across all the possible weight vectors.[3] The synthesized efficiency of a point combination (belonging to the Pareto-optimal set) $\boldsymbol{x}_{\mathcal{P}}$, with weight vector $\boldsymbol{w}$, is defined as follows:

$$\frac{D(\boldsymbol{x}_{\mathcal{P}}, \boldsymbol{w})}{max_{\boldsymbol{x}_{\mathcal{P}}}[D(\boldsymbol{x}_{\mathcal{P}}, \boldsymbol{w})]}$$

The shading, from white to black, represents the transition from high to low values of the synthesized efficiency. Each of the 19 shades of grey, starting from the lightest, corresponds to a decrease in the synthesized efficiency of 0.05.

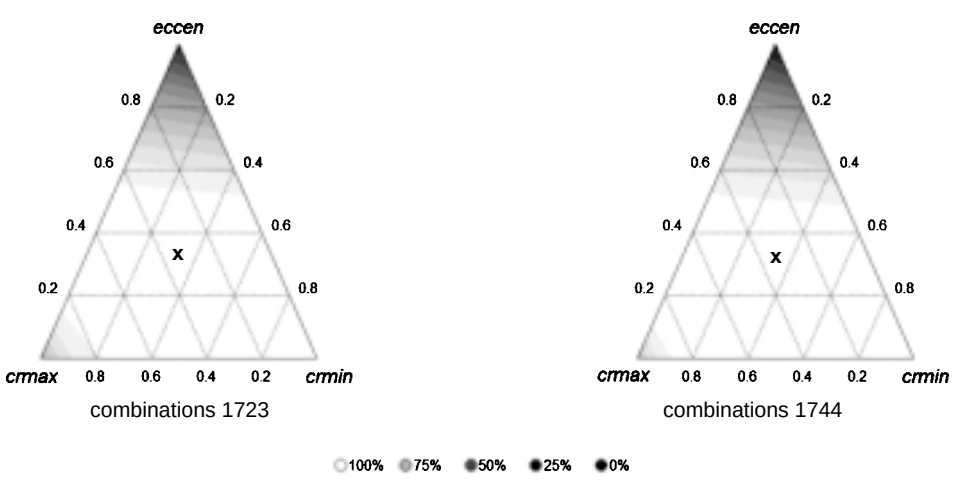


Fig. 6: Synthesized efficiency plot for the point combinations 1723 and 1744

[3] It should be noted that for the construction of the synthesized efficiency plots (Figure 6) and the mixture plot (Figure 4), a set of 20301 weight combinations has been defined, where adjacent weights related to a same response variable are separated by a distance equal to 0.005.

The large white region characterizing the two graphs, represents approximately 75% and 74% of the triangles, respectively, and indicates that both points have a synthesized efficiency of at least 0.95 for a substantial number of weight combinations. In particular, the white region around the centroid of the triangle, indicated in Figure 6 with a black cross, shows that both point combinations give excellent performance at the weighting region, which reflects the company's priorities/preferences. The graphical plots provide detailed information about the relative performance of the different contending solutions and allow the experimenter to understand what alternatives are available.

Moreover, we select the optimal solution as represented by combination 1744, since it is slightly better for a large number of weight combinations, and in particular, for the weighting giving equal importance to the three response variables. However, input combination 1723 provides a similar performance, thus representing a valid competitive alternative.

4.4. RESULT COMPARISON BETWEEN THE TWO APPROACHES

Table 2 shows the results from both multi-response optimization methods: the Pareto front approach and the analytical method applied in Berni (2010). By comparing (Table 2) the optimum point combination 1744 (Piattoli, 2020) and the optimal analytical solution, we can observe how only one process factor, *cp*, shows the same optimal level; nevertheless, it must be noted that the circle-point variable is the main process variable that we are interested in optimizing. Although both combinations provide similar predicted response values, the analytical method allows for obtaining a better value for the response $\widehat{eccen}$. This is an important result in view of the relevance that this response variable has in the actual process. The constructed Pareto front contains similar solutions to those identified by the optimal analytical solution, but corresponding to different weight combinations. Hence, with a more thorough exploration of the solution set identified with the Pareto front, a similar solution could be selected relaxing the assumption that all the responses were of equal importance.

Tab. 2: Optimization results: the comparison

Method	Factors			Predicted responses		
	mspeed	***dspeed***	***cp***	$\widehat{crmax}$	$\widehat{crmin}$	$\widehat{eccen}$
Pareto front	-1	1	4	3.00390	2.78912	0.00998
Analytical	0.710	0.362	4	3.00300	2.78500	0.00100

It is important to note, however, that although a Pareto front can be constructed for any number of responses of interest, the graphical tools considered here only used three response variables, unlike the five considered in Berni (2010). For this reason, it was only possible to make a partial comparison between the results obtained through the two different methods. Nevertheless, the Pareto front approach offers the possibility of using additional graphical tools (Lu et al., 2017), which enable multi-response optimization of more than three response variables.

Moreover, through the analytical approach, the optimization was carried out considering both non-standardized and standardized data, where the latter gave the best optimization results.

5. GENERAL COMPARISON AND FINAL REMARKS

The results obtained through the case study allows us to perform an empirical comparison between the two approaches, where some specific differentiations could be viewed in a theoretical perspective, as outlined in the following scheme (Table 3). Both methods use the same experimental plan, data and analysis, but then differ in how choose to optimize the settings of inputs.

Tab. 3: Theoretical step comparison between the Pareto front approach and the analytical method

Step	***Method***	
	Pareto front	**Analytical**
1	DoE: planning and trials	DoE: planning and trials
2	Statistical modelling	Statistical modelling
3	Optimization: A) objective phase - identification of Pareto-optimal set; B) subjective phase - choice of the optimal solution among the points belonging to the Pareto-optimal set, taking the quantitative results and the decision-makers' priorities into account; C) validation of the results obtained at step (B), intrinsic in the subjective phase.	Optimization: a) definition of objective function (formula (4)); b) minimization (or maximization) of the objective function, and identification of the optimal solution (optimal process variable levels); c) validation of the obtained results at step (b) by: 1. the objective function value, (results also checked through: convergency, gradient estimates, determinant of the Hessian matrix); 2. application of the optimal solution (obtained through step (b)) in the real (actual) production process, by involving the stakeholders (engineers).

Undoubtedly, the Pareto front approach offers the advantage of using graphical tools in a simple and intuitive way, enabling straightforward identification of leading solutions with discussion allowing for consensus of the optimal solution among the various company teams involved. The elimination of non-competitive choices streamlines where to focus further discussion. Moreover, a subjective evaluation (Table 3, step B) can also be performed, with the possible achievement of a unanimous decision among different stakeholders, and considering different weightings of how important the performance on each response is to overall results. It is possible to compare different identified solutions, and see their relative strengths and weaknesses for each of the responses of interest. Indeed, it allows for accurate comparison among several input combinations of interest. A further advantage is the flexibility in response weighting to handle multiple combinations of business priorities and to examine the impact of these choices on the identified results. The transparent nature of the Pareto front presents the experimenter with different alternatives that can be explored and compared. Nevertheless, this is also possible by performing analytical optimization methods (see Lin and Tu, 1995). In addition, response weighting and analytical methods assign relative importance to each response according to the estimated corresponding weight (Berni, 2010). This is particularly helpful for solving technological issues and constraints, that can be measured and evaluated in a wide and general context.

An advantage of the analytical method is the ability to include random effects, within both the modelling and the optimization steps. Therefore, the fixed as well as random effects are wholly involved, and as a result a robust process optimization can be carried out, and the final validation (Table 3, step c.2) specifically verified in the firm, allows for checking the validity in the actual context. It is straightforward to use the same optimization function based on the inclusion of random effects for each of the responses as the basis for constructing the Pareto front.

The aforementioned advantages and disadvantages highlight the significant relevance of both methods, as each has specific strengths and weaknesses that would be relevant for a wide range of empirical situations (real industrial processes, technological contexts) where they can be effectively applied.

REFERENCES

Ames, A.E., Mattucci, N., Macdonald, S., Szonyi, G. and Hawkins, D. M. (1997). "Quality loss functions for optimization across multiple response surfaces". *Journal of Quality Technology*, 29 (3): 339-346.

Anderson-Cook, C.M. (2017). "Optimizing in a complex world: a statistician's role in decision making". *Quality Engineering*, 29 (1): 27-41.

Anderson-Cook, C.M. and Lu,L. (2018). "Graphics to facilitate informative discussion and team decision making". *Applied Stochastic Models in Business and Industry*, 34 (6): 963-980.

Berni, R. (2010). "Split-plot and robust designs: weighting and optimization in the multiple response case". In: Giovagnoli, A., Atkinson, A. C., Torsney, B., and May, C., eds. *mODa 9 – advances in model-oriented design and analysis: proceedings of the 9th international workshop in model-oriented design and analysis held in Bertinoro, Italy, June 14-18, 2010*. Heidelberg: Physica-Verlag, 25-32.

Berni, R. and Bertocci, F. (2018). "Optimization of gas sensors based on advanced nanomaterials through split-plot designs and GLMMs". *Sensors*, 18 (11): 3858.

Berni, R., Bertocci, F., Nikiforova, N.D. and Vining, G. G. (2020). "A tutorial on randomizing versus not randomizing split-plot experiments". *Quality Engineering*, 32 (1): 25-45.

Berni, R. and Burbui, M. (2014). "Process optimization of a superfinishing machine through experimental design and mixed response surface models". *Quality Engineering*, 26 (4): 404-415.

Berni, R. and Gonnelli, C. (2006). "Planning and optimization of a numerical control machine in a multiple response case". *Quality and Reliability Engineering International*, 22 (5): 517-526.

Berni, R. and Nikiforova, N. D. (2022). "A joint multiresponse split-plot design modeling and optimization including fixed and random effects". *Austrian Journal of Statistics*, 51: 83-101. https://doi.org/10.17713/ajs.v51i1.1211

Bisgaard, S. (2000). "The design and analysis of $2^{k-p} \times 2^{q-r}$ split plot experiments". *Journal of Quality Technology*, 32 (1): 39-56.

Box, G.E.P., Hunter, W. G., MacGregor, J. F. and Erjavec, J. (1973). "Some problems associated with the analysis of multiresponse data". *Technometrics*, 15 (1): 33-51.

Box, G.E.P. and Jones, S. (1992). "Split-Plot designs for robust product experimentation". *Journal of Applied Statistics*, 19 (1): 3-26.

Chapman, J.L., Lu,L., and Anderson-Cook, C.M. (2014a). "Process optimization for multiple responses utilizing the Pareto front approach". *Quality Engineering*, 26 (3): 253-268.

Chapman, J. L., Lu, L. and Anderson-Cook, C. M. (2014b). "Incorporating response variability and estimation uncertainty into Pareto front optimization". *Computers & Industrial Engineering*, 76: 253-267.

Chiao, C.-H. and Hamada, M. (2001). "Analyzing experiments with correlated multiple responses". *Journal of Quality Technology*, 33 (4): 451-465.

Cochran, W.G. and Cox, G.M. (1957). *Experimental designs*. 2° ed. New York: Wiley.

Coello Coello, C.A., Lamont, G.B. and Van Veldhuizen, D.A. (2007). *Evolutionary algorithms for solving multi-objective problems*. 2º ed. New York: Springer, 1-13.

Copeland, K.A.F. and Nelson, P.R. (1996). "Dual response optimization via direct function minimization". *Journal of Quality Technology*, 28 (3): 331-336.

Cornell, J.A. (2002). *Experiments with mixtures: designs, models, and the analysis of mixture data*. 3° ed. New York: Wiley.

Del Castillo, E., Montgomery, D.C. and McCarville, D.R. (1996). "Modified desirability functions for multiple response optimization". *Journal of Quality Technology*, 28 (3): 337-345.

Derringer, G. and Suich, R. (1980). "Simultaneous optimization of several response variables". *Journal of Quality Technology*, 12 (4): 214-219.

Dror, H. A. and Steinberg, D.M. (2006). "Robust experimental design for multivariate generalized linear models". *Technometrics*, 48 (4): 520-529.

Jones, B. and Nachtsheim, C.J. (2009). "Split-plot designs: what, why, and how". *Journal of Quality Technology*, 41 (4): 340-361.

Khuri A. I. (1996). "Response surface models with mixed effects". *Journal of Quality Technology*, 28 (2): 177-186.

Khuri, A.I. and Conlon, M. (1981). "Simultaneous optimization of multiple responses represented by polynomial regression functions". *Technometrics*, 23 (4): 363-375.

Kowalski, S.M., Parker, P.A. and Vining, G.G. (2007). "Tutorial: industrial split-plot experiments". *Quality Engineering*, 19 (1): 1-15.

Kowalski, S.M. and Potcner, K.J. (2003). "How to recognize a split-plot experiment". *Quality Progress*, 36 (11): 60-66.

Lee, Y. and Nelder, J.A. (2003). "Robust design via generalized linear models". *Journal of Quality Technology*, 35 (1): 2-12.

León, R.V., Shoemaker, A.C. and Kacker, R. N. (1987). "Performance measures independent of adjustment". *Technometrics*, 29 (3): 253-285.

Lin, D.K.J. and Tu, W. (1995). "Dual response surface optimization". *Journal of Quality Technology*, 27 (1): 34-39.

Lu, L. and Anderson-Cook, C.M. (2012). "Rethinking the optimal response surface design for a first-order model with two-factor interactions, when protecting against curvature". *Quality Engineering*, 24 (3): 404-422.

Lu, L., Anderson-Cook, C.M. and Robinson, T. J. (2011). "Optimization of designed experiments based on multiple criteria utilizing a Pareto frontier". *Technometrics*, 53 (4): 353-365.

Lu, L., Chapman, J.L. and Anderson-Cook, C.M. (2017). "Multiple response optimization for higher dimensions in factors and responses". *Quality and Reliability Engineering International*, 33 (4): 727-744.

Marler, R.T., and Arora, J. S. (2004). "Survey of multi-objective optimization methods for engineering". *Structural and Multidisciplinary Optimization*, 26 (6): 369-395.

McCullagh, P. and Nelder, J.A. (1989). *Generalized linear models*. 2° ed. London: Chapman & Hall.

Myers, R.H., Khuri, A.I. and Vining, G.G. (1992). "Response surface alternatives to the Taguchi robust parameter design approach". *The American Statistician*, 46 (2): 131-139.

Myers, R.H., Montgomery, D.C. and Anderson-Cook, C.M. (2016). *Response surface methodology: process and product optimization using designed experiments*. 4° ed. Hoboken, New Jersey: Wiley.

Nair, V.N. (1992). "Taguchi's parameter design: a panel discussion". *Technometrics*, 34 (2): 127-161.

Nelder, J.A. and Lee, Y. (1991). "Generalized linear models for the analysis of Taguchi-type experiments". *Applied Stochastic Models and Data Analysis*, 7 (1): 107-120.

Piattoli, L. (2020). *Disegno split-plot per l'ottimizzazione multiresponse in ambito tecnologico* [MD Thesis]. Florence: University of Florence, pp. 48-60.

Rajagopal, R., Del Castillo, E. and Peterson, J.J. (2005). "Model and distribution-robust process optimization with noise factors". *Journal of Quality Technology*, 37 (3): 210-222.

Robinson, T.J., Wulff, S.S., Montgomery, D.C. and Khuri, A.I. (2006). "Robust parameter design using generalized linear mixed models". *Journal of Quality Technology*, 38 (1): 65-75.

Tang, L.C. and Xu, K. (2002). "A unified approach for dual response surface optimization". *Journal of Quality Technology*, 34 (4): 437-447.

Vining, G.G. and Kowalski, S.M. (2008). "Exact inference for response surface designs within a split-plot structure". *Journal of Quality Technology*, 40 (4): 394-406.

Vining, G.G., Kowalski, S.M. and Montgomery, D. C. (2005). "Response surface designs within a split-plot structure". *Journal of Quality Technology*, 37 (2): 115-129.

Vining, G.G. and Myers, R.H. (1990). "Combining Taguchi and response surface philosophies: a dual response approach". *Journal of Quality Technology*, 22 (1): 38-45.

Zitzler, E. (1999). *Evolutionary algorithms for multiobjective optimization: methods and applications* [dissertation]. Zurich: Swiss Federal Institute of Technology Zurich, 5-10.

CIRCULAR ECONOMY IN SMALL AND MEDIUM-SIZED ENTERPRISES IN THE EUROPEAN UNION: HETEROGENEITY BETWEEN AND WITHIN COUNTRIES

Francesca Bassi[1], Dario Costa

Department of Statistical Sciences, University of Padova, Italy

***Abstract**. In this paper, we analyze the different behavior of small and medium-sized enterprises (SMEs) with reference to eight specific Circular Economy (CE) actions. Data come from a Flash Eurobarometer survey that investigates efficiency in use of resources. We estimated classification trees (CART) in order to identify homogeneous groups of European countries with regard to the adoption of CE practices by SMEs and multilevel regression models to measure differences among SMEs in adopting sustainability management, considering firms' characteristics. Results of the analyses revealed heterogeneity between and within European countries. Five groups of countries are identified, with SMEs having a similar attitude towards CE. Within each group of factors, however, specific firms' characteristics have a non-negligible effect on firms' decision to adhere to sustainability.*

***Keywords**: Efficiency of resources, Circular Economy, SMEs, CART, multilevel analysis.*

1. INTRODUCTION

In this paper, we analyze the different behavior of small and medium-sized enterprises (SMEs) with reference to eight specific Circular Economy (CE) actions. The transition from a linear economic context to a circular one implies for products and services a change from a production system with the phases of conception, construction, use and disposal, to a system committed to having less waste and to environmental issues. The market is giving way to a circular idea of the value chain, which means that the environmental impact, that the materials that compose it will have, is assessed from the very beginning of the production phase, from its conceptualization and design (Suárez-Eiroa et al., 2019). Therefore, in the last few decades, the major companies begun to consider as a new resource the opportunity of processing the materials released by production and the products themselves, once their use is finished. Reuse, and recycling are among the Rs on which the Circular Economy is based (Vermeulen et al., 2019).

[1] Corresponding author: Francesca Bassi, email: bassi@stat.unipd.it

In order to understand how SMEs are dealing with CE and specifically to evaluate heterogeneity of behavior in European countries, we analyzed the data collected with a Eurobarometer survey, precisely with Flash Eurobarometer 456, conducted from 11th to 26th September 2017 on sample of European firms. These data refer to eight green actions to be implemented in the production process: saving water, saving energy, using renewable energy, saving materials, minimizing waste, recycling, designing products that are easy to maintain or repair, selling scrap material to another company.

In this paper, we will use classification trees (CART, Breiman et al. 1984) in order to identify homogeneous groups of European countries with regard to the adoption of CE practices by SMEs and multilevel regression models (Hox, 2002) to measure differences among SMEs in adopting sustainability management, considering firms' characteristics.

The paper is organized as follows: section 2 reviews the concept of CE and the recent reference literature, with a specific attention to SMEs; section 3 describes the data; section 4 presents the results of the analyses and section 5 concludes.

2. CIRCULAR ECONOMY: ORIGINS, DEVELPOMENT AND IMPLEMENTATIONI IN SEMS

The origin of the modern idea of industry started at the end of the eighteenth century with the Industrial Revolution, this era was characterized by the overproduction of goods and the extensive use of fossil energy sources. Only during the World Wars, companies were forced to reconsider a really different method of using raw materials, due to the enormous costs to support the war front. The dynamics of the production system of the time consists in proceeding from virgin raw materials to transformation, consumption and final conferment in landfill; this production model can be defined as linear. The linear model can be improved and optimized, but still what remains is waste, pollutants and scraps of industrial production and consumption with negative environmental and social implications (Sharma et al., 2021). The introduction of the concept of Circular Economy dates back to the end of the twentieth century, when several publications arouse the attention of scholars (Lieder and Rashid, 2016). Several authors, such as Andersen (2007), Ghisellini et al. (2016), and Su et al. (2013) attributed the introduction of the concept to Pearce and Turner (1989) in their work "Economics of Natural Resources and the Environment". These authors wrote how natural resources sustain the economy by providing inputs for production and consumption, but, at the same time, natural resources produce outputs, which are represented almost

entirely from waste. From these considerations, the idea of Circular Economy was formulated. Pearce and Turner were influenced by the work of Boulding (1966), who described the Earth as a circular and closed system with limited assimilation capacity, deducing that the economy and the environment should coexist in equilibrium. Stahel and Reday (1976) introduced some features to this economic approach, with a particular focus on the industrial sector; they devised a continuous cycle economy to write industrial strategies on waste prevention, job creation, resource efficiency and dematerialization. Later, Stahel (1982) further stressed that granting use, instead of relinquishing ownership of goods, is the most relevant sustainable business model for a closed economy, thus allowing companies to profit from waste without having costs and risks deriving from them. The term Circular Economy was coined in China in 2002, when the government approved the first CE Promotion Law of the People's Republic of China, which became effective in January 2009 (The Standing Committee of the National People's Congress China, 2008). The main goal was to reduce pollution and protect the planet by making important public decisions. After this turning point, institutions from all over the world, including the European Union, had to inevitably adapt to addressing these problems. The first European countries to adopt CE practices were Sweden, the United Kingdom and Spain (Lieder and Rashid, 2016).

The use of the term CE has evolved in the business world in an attempt to find a compromise between economic growth and environmental protection. This idea of economy wants to be in contrast with the linear one. The most renowned definition of CE was given by the Ellen MacArthur Foundation (2015), introducing the Circular Economy as "a restorative industrial economy or regenerative in intention and design", from then, many different ways to describe the process were proposed. Thinking about the eco-environmental industry, the CE can also be defined as a closed flow of materials within a well-structured economic system (Geng and Doberstein, 2008). Kirchherr et al. (2017) analyzed over one hundred definitions of Circular Economy and concluded that all are in line with 3Rs (Reduction, Reuse and Recycling) paradigm, which purpose is to make the flow of materials closed or circular (Yuan et al., 2006). The fact that the present economic development model has taken a dead-end way, it is recognized not only by those who work in the sectors directly concerned, but by the majority of the people. It is evident that it is necessary to rearrange the bases of the production system in order to guarantee adequate well-being to future generations.

The European Union is encouraging the business activities to seriously face ethical problems, which concern the economic cycle in an active way. (European Commission, 2008; Dalhammar 2015). The rules to guide the cycle of a product can

be summarized as it follows: prohibition of dangerous substances, products energy-efficient and correct disposal of materials after their use. The European Union presented in 2015 an executive plan for CE, which includes legislative proposals and measures for the management of the production, consumption and waste (Dalhammar, 2016). The plan is divided into two parts: the first part explains how CE measures can be introduced into the product life cycle and the second one, instead, is devoted to the care and the specific treatment of scraps (European Commission, 2014).

One of the objectives of the recommendations is to inform and address not only the behavior of companies, but also that of consumers. Therefore, the goal is making everyone more aware that the cycle of a product involves the whole of society and that the gain following a correct behavior is for everyone's life, not only for the economic benefit. The action plan seeks to put industries in a sustainable but competitive context, stimulate economic growth and create new types of jobs. From this, it arises the need for companies to have a qualified workforce with new and specific green skills (Bassi and Guidolin, 2021, ⁄beda-García et al. 2021). The professions that fall into this category can be defined in many ways, Burger et al. (2019), for example, did an in-depth exploration of the US market.

This transition, however, has a very high cost. The European Commission is moving in this direction, with measures to encourage the adoption of CE practices and employment growth. The first action plan (European Commission, 2014) was designed for the entire life cycle of a product: from production to consumption, from repair to regeneration, to waste management. The intent was to direct it to all the administrations involved, starting from the member States of the European Union, then passing through the regions, cities, businesses and finally citizens. The European Commission attributes a very important role to the production phase in the chain, encouraging companies to replace harmful chemicals and/or to have innovative technologies for production processes. For this reason, the European Resource Efficiency Excellence Center was created, helping companies to improve their production efficiency. The devised action plan takes into account also the contribution and the responsible choices that consumers will make in the purchasing phase, this in fact is an element that will directly affect the functioning of the CE. An example regards the fact the price of a product that has been conceived and put on the market using sustainable production techniques with a good environmental impact, will be higher. It is important that customers appreciate this effort and are willing to pay the additional cost. The cost of a product with these characteristics will be proportional to the attention it has for the environmental effects, the companies in this will be supported through incentives from the European Union

but the consumer will have a great responsibility in considering environmental protection as a quality during the purchase (European Commission, 2018a). In the European Union the CE regards almost exclusively small and medium-sized enterprises (SMEs), which represent over 99% of all European companies and around two thirds of total employment. SMEs have been defined by the European Commission as companies that have less than 250 employees and whose turnover does not exceed 50 million euros and/or whose total balance sheet does not exceed 43 million (European Commission, 2003). In detail, micro enterprises are those with less than 10 workers and an annual balance sheet lower than 2 million euros; small businesses are those with less than 50 workers and a budget of less than 10 million; businesses are defined medium-sized if they have between 50 and 250 workers and a budget between 10 and 43 million euros. From the annual report of the European Commission on SMEs 2018/2019 (European Commission, 2020b), it emerged that in the European Union about 25.1 million SMEs are operating: 23.3 million micro-enterprises, 1.47 million small ones and about 236 thousand medium-sized enterprises. It is estimated that SMEs create between 60 and 70% of total air pollution (Hoogendoorn et al., 2015). The European Union considers the contribution of SMEs to be fundamental to the CE also because they should be more active and predisposed to changes in the sectors of recycling, repair and product innovation. The distribution of SMEs in the EU is not homogeneous, compared to an average of 92% on all active companies, this figure can vary considerably from state to state. For example, in Germany, SMEs represent 82% of all businesses, unlike countries like Greece, Poland or the Czech Republic where they are over 96%.

The circular chain is a model based on the supply of renewable, recyclable and biodegradable products. With the recovery of resources, at the end of a production process, waste continues to have an intrinsic value and can be used in further transformation processes. The sharing model promotes collaboration between users of goods and services in order to exploit overcapacity and underutilization. To implement the transition to sustainable and Circular Economy models, the European action plan defines 54 measures to close the loop life of products, identifying five priority areas to accelerate the transition along their value chain (plastics, food waste, essential raw materials, construction and demolition, biomass and materials biological, European Commission, 2015). The plan places a strong emphasis on creating a solid foundation on which investment and innovation can thrive. The transition is financially supported by the European Structural and Investment Funds, from Horizon 2020, from the European Fund for Strategic Investments (EFSI) and the LIFE program, founded in 1992 to promote protection strategies of the environment. A recent step taken by the European Union was to

implement the second action plan for the Circular Economy on 11th March 2020 (European Commission, 2020a). This initiative is one of the main measures of the European Green Deal, defined as the roadmap to make the EU economy as sustainable. The new plan describes the way to progress towards a climate-neutral and competitive economy, in which consumers are empowered. The objectives are: making sustainable products as the standard within the Union. The Commission proposed a legislative action on the strategy for sustainable products to ensure that they are designed to last longer, easy to reuse, repair and recycle, and contain as much recycled materials as possible rather than primary raw materials. The measures will also limit single-use products, they will deal with premature obsolescence and ban the destruction of unsold durable goods. Secondly, the empowerment of consumers who will have access to reliable information on issues such as reparability and durability of products so they can make better informed choices. As Najami et al. (2020) noted, sustainability cannot be fulfilled without the collaboration of end consumers.

Finally, it is important to recognized that the Circular Economy will produce net assets in terms of GDP growth and job creation; it is estimated that the application of ambitious Circular Economy measures in Europe will be able to increase GDP by a further 0.5% between now and 2030, creating around 700.000 new jobs (European Commission 2020a).

3. THE DATA

Our data were collected within the Flash Eurobarometer 456 survey, conceived and proposed by the European Commission. The data collection period is included between 11th September 2017 and 26th September 2017; questions were answered by 13,117 SMEs belonging to the 28 countries of the European Union (before the Brexit)[2]. The intent was to understand how many efficiency measures were developed by firms. With reference to the single SME, the following "demographic" information was collected: country, economic activity sector, number of workers, year of foundation, if turnover changed or remained stable in the past two years, turnover in 2016, type of output, whether products or services were sold.

Eight specific CE actions were investigated: paying attention to the waste and reuse of water; minimize energy use while maximizing performance; using mainly renewable energy (including own production through solar panels or other); saving

[2] Quota sampling was used with quotas applied to company size and sector, adjusted according to country's universe. Interviews were conducted by CATI mode.2 For the analyses the software R was used.

of raw materials; waste minimization; the permission to other companies of the use of waste; recycling, reusing materials or waste from the company itself; creation of products that are easier to maintain, repair or reuse. With reference to these eight actions, SMEs had to declare if they were implemented in the preceding two years and/or if there was the intention to consider them for the future two years. Additional information with reference to these actions was collected, regarding costs, percentage of turnover invested, eventual financial support received and from which source, workers employed in green jobs, difficulties and needs related to the implementation of sustainability practices.

We expected that the size of the company had direct effects on the choice to undertake activities related to the CE (Bianchi and Noci, 1998). Larger companies have access to more resources to invest, while smaller ones can suffer from the absence of a strong economic structure that supports them for investments and measures that have a targeted production scheme to recycling (Hogg et al., 2011). From the point of view of sustainability, all companies have the goal of creating a type of environmental economics. It is noted that companies that generate one low portion of waste in relation to what they produce are less motivated to think about recycling methods (Reike et al., 2018). On the other hand, in large companies, ethics plays a central role in their behavior as they are more importantly exposed to criticism and it is therefore a necessity to preserve their reputation (Inigo and Blok, 2019).

The age of an SME has a direct effect on the willingness to undertake Circular Economy practices (Hoogendoorn et al., 2015). The competence and social responsibility of a company can derive precisely from the experience that has been accumulated in the area (Trencansky and Tsaparlidis, 2014). Social responsibility derives from an economic and corporate stability. When business procedures for the environment are put in place although costs might increase, it is in all respects a way to please stakeholders. The same concept also applies to new companies, who, having to set up new working strategies, can gain in considering the idea of the Circular Economy as a new perspective model both for the environment and from the business point of view. Older and newer SMEs have more interest in undertaking a business model that follows the CE, more than companies with an intermediate age.

The sector in which an SME operates influences its willingness to undertake sustainable activities or to follow green economy policies (Bradford and Fraster, 2008). SMEs operating in sectors with production processes are more tangible and producing greater quantities of waste are the keenest to follow CE expedients. The sectors that are most inclined to suggest sustainable activities are manufacturing, construction, agriculture and waste management. In these sectors, the production process disperses a lot of waste and this leads to requesting and having large

quantities of raw materials. Furthermore, strict environmental and corporate parameters have been devised by nations and by institutions to stem waste and give a common direction to SMEs (European Commission, 2018b)). The need for greater quantities of materials for the sectors with the most organic value chain composed of tangible materials is an element that makes the sustainable choice of SMEs a priority. In fact, in the EU action plan, the plastics, food and raw materials sectors, constructions and demolitions, biomasses and products biologicals have priority for the implementation of efficiency measures.

SMEs that participated in the survey employ on average 18 workers: 80.4% of them have between 1 and 9 employees, 15.5% between 10 and 49, 3.0% between 50 and 249, and 1.1%, large companies, have more than 250 employees. 57.4% of SMEs have no employees engaged in a green job. The average age of SMEs is 25.7 years: 76.9% were founded before to 2010, 9.3% between 2010 and 2012, 23.3% between 2013 and 2016, and only 1.5% were founded in 2017. For what regards economic activity sector, 10.1% of SMEs belong to the manufacturing sector, 15.9% to the industrial one, 30.1% are active in retail, and 43.9% in services. Another relevant aspect to consider to assess the propensity of a company to undertake CE actions is its annual turnover. In the two years preceding the questionnaire, the turnover of SMEs grew for 42.5% of them, decreased for 21.2% while it remained almost the same for the other 20.1%. 19.6% of SMEs have a turnover of less than 100,000 euro in the reference year, 23.3% a turnover between 100 and 500 thousand, 22.8% between 500 thousand and 2 million, 18.8% between 2 and 10 million, 10.4% between 10 and 50 million and only 5.1% had a turnover that is greater than 50 million euros.

As already introduced, the survey aimed at measuring the adoption of specific CE practices by European SMEs. The most adopted efficiency action was the minimization of waste, undertaken by 65.5% of SMEs; minimizing energy use by keeping stable or increasing performance was adopted by 63.2% of them, saving materials regards 56.8% of SMEs, and minimization of water waste is adopted by 47.3%. Recycling inside company through reuse and use of waste was undertaken by 41.8% of companies; designing ad hoc of products that are easier to maintain, repair or reuse them is applied by 25.4% of firms. The sale of waste to other companies is done by 21.1% European SMEs, while the least adopted sustainability practice is the choice to use mainly renewable energy (14.0%).

Table 1 shows the relationship between the adoption of sustainability practices and the characteristics of the firms by number of employees, economic activity sector, age, turnover in 2016; all relationships are statistically significant according to the Chi-squared test.

Tab. 1: Percentage of European SMEs adopting CE practices by characteristics.

	Minim. waste	Saving energy	Saving materials	Saving water	Recycling	Design products	Selling scrap	Renewable energy
EU 28	65.5	63.2	56.8	47.3	41.8	25.4	21.1	14.0
Size								
1-9	64.7	62.3	55.4	46.9	40.1	24.3	18.0	12.6
10-49	66.3	64.1	62.3	46.1	45.9	28.2	31.5	17.4
50-249	77.2	75.9	64.7	56.3	58.6	38.1	47.7	30.2
250+	80.3	81.9	62.0	69.6	59.1	26.8	29.7	26.8
Turnover								
-100,000	57.0	58.1	54.2	43.4	38.0	21.2	17.2	11.6
100,00-500,000	66.7	63.3	57.0	48.3	39.0	26.5	18.8	14.6
500,000-2mil	68.8	67.1	59.7	46.4	46.1	28.9	26.6	15.5
2-10mil	71.3	69.5	63.6	46.5	47.4	22.3	30.0	17.7
10-50mil	78.4	77.9	72.4	58.0	56.9	42.9	53.0	43.3
+50mil	84.7	80.9	64.5	68.5	50.5	17.3	23.4	21.6
Sector								
Manufacturing	71.3	64.6	64.2	43.4	41.9	33.2	31.5	12.7
Retail	65.1	66.9	56.9	48.3	44.1	24.3	21.6	11.7
Services	62.7	61.3	54.2	46.4	38.5	23.2	15.1	14.1
Industry	70.4	60.4	59.1	46.5	46.4	28.1	30.0	18.7
Age								
-31 Dec 2010	66.7	64.4	57.2	47.9	41.7	25.3	22.4	14.4
1 Jan 2010-31 Dec 2012	62.2	59.3	55.8	42.0	41.9	24.5	16.3	13.3
1 Jan 2013-31 Dec 2017	61.0	58.5	55.1	46.7	40.2	26.5	17.5	13.0
1 Jan 2017+	66.8	63.8	52.0	47.4	53.1	26.0	21.1	5.6

Entrepreneurial sustainability is by definition linked to the social and economic context. In developed countries, both legal framework and financial resources are very solid, favoring sustainable entrepreneurship. In developing countries, transition to Circular Economy is more challenging, especially in terms of infrastructures and new technologies (Abarca Guerrero et al., 2020). The legal context is one of the factors which varies most from state to state also in the European Union; even if our analyses do not directly focus on this aspect, it is an element that must be kept in consideration for appropriate comparisons. Economic indicators at national level are also important to understand the wealth of the country since the economic dimension of sustainability coincides with the large amount of liquidity that can be used to satisfy the needs and requests of stakeholders. In more developed countries, investment policies are directed mostly to the private sector, since it is more inclined

to support innovation and competitiveness, which favor a sustainable growth (Cadil et al., 2018). The level of innovation goes hand in hand with new technologies, which can also be considered a prerequisite. Although SMEs do not usually make innovation as the most important aspect of their structure, there is significant financial support by the European Union to increase their performance in terms of sustainability. The reference literature shows that entrepreneurial sustainability is also influenced by factors such as gender, age, education, skills, family context and community background. Therefore, differences between countries with regard to CE practices are the result of the complex mixture of all aspects mentioned above (Spangenberg et al., 2002). Table 2 lists the percentage of firms implementing each sustainability action in the 28 European MSs.

A preliminary exploratory cluster analysis on the data reported in Table 2 classifies the 28 EU MSs in four homogeneous groups for the percentage of SMEs operating in the country and adopting CE practices. The greenest firms are located in France, Ireland, Portugal, Spain, Sweden and Great Britain, countries were the percentage of SMEs adopting CE practices is higher than the average for at least seven among the eight considered actions, except for the efficiency practice of using renewable energy. A second group is formed by Austria, Belgium, Croatia, Czech Republic, Denmark, Finland, Germany, Italy, Luxemburg, Poland, Slovenia, The Netherlands; in these countries the percentage of firms adopting CE practices is higher than the average for many practices, these percentages, however, are lower than those observed for SMEs located in the first group of countries. The percentage of firms in these countries that are saving water and energy is lower than that in the average sample. The third group refers to Greece, Hungary, Latvia and Slovakia whose SMEs implement only in small percentages, lower than those observed at European level, the majority of actions; however, in these countries, firms are involved in selling scrap material to other companies. Finally, in the last group of countries, Bulgaria, Cyprus, Estonia, Lithuania, Malta, and Romania only a very small percentage of SMEs adopts green practices. This first result highlights the presence of a discrete level of heterogeneity in adopting efficiency practices in the 28 European countries, this evidence will be further explored in the next section of the paper.

Tab. 2: Percentage of European SMEs adopting CE practices by country, figures in *italics* indicate a value greater than the average.

	Minim. waste	Saving energy	Saving materials	Saving water	Recycling	Design products	Selling scrap	Renewable energy
EU 28	65.5	63.2	56.8	47.3	41.8	25.4	21.1	14.0
Austria AT	60.0	*71.8*	52.3	44.1	*47.7*	*31.3*	*25.6*	*32.3*
Belgium BE	*75.0*	*69.8*	61.6	46.7	40.5	25.3	*25.6*	*19.9*
Bulgaria BG	27.7	36.4	30.6	29.5	16.9	9.8	16.3	4.4
Cyprus CY	28.6	50.0	28.6	28.6	*46.4*	7.1	14.3	7.1
Czech Republic CZ	64.0	60.9	46.2	42.4	34.7	*32.6*	*29.1*	7.2
Germany DE	61.3	*70.2*	57.1	35.8	38.6	24.1	23.0	*32.0*
Denmark DK	50.0	56.3	52.3	41.4	28.9	*26.6*	*26.4*	9.4
Estonia EE	8.6	20.0	14.3	8.6	11.8	5.7	5.7	2.9
Spain ES	65.5	*72.4*	*69.5*	*54.8*	*56.9*	*31.9*	20.0	7.4
Finland FI	55.2	50.7	53.7	26.9	31.6	23.3	18.7	*14.3*
France FR	*83.1*	*71.2*	*59.4*	*67.8*	*42.5*	*32.7*	18.0	5.5
Great Britain GB	*81.9*	*66.8*	*61.8*	*55.7*	*70.2*	*25.8*	*28.8*	*16.3*
Greece GR	36.8	51.4	45.4	32.8	30.1	18.3	*25.9*	12.3
Croatia HR	64.3	*65.5*	*62.4*	*51.2*	29.8	17.9	*28.2*	8.3
Hungary HU	40.1	57.8	45.0	40.1	18.8	16.3	20.9	7.4
Ireland IE	*84.9*	*68.5*	*57.4*	*59.3*	*70.4*	24.5	*25.9*	*18.5*
Italy IT	*73.6*	57.3	52.5	44.4	37.3	23.0	15.2	*15.0*
Lithuania LT	20.5	42.0	33.0	35.2	6.8	6.8	15.9	3.4
Luxemburg LU	60.0	50.0	52.4	33.3	*45.0*	*28.6*	25.0	*15.0*
Latvia LV	34.5	60.0	54.5	43.6	14.5	16.4	10.9	3.6
Malta MT	64.3	*71.4*	35.7	28.6	*50.0*	14.3	21.4	*14.3*
Netherlands NL	65.2	*64.7*	*61.3*	32.1	36.9	20.6	*25.7*	*26.7*
Poland PO	55.1	57.3	*59.9*	*49.5*	23.6	16.7	21.1	4.1
Portugal PT	55.1	*75.6*	*74.8*	*63.1*	*65.9*	*42.2*	*23.7*	9.2
Romania RO	31.0	32.9	29.5	23.4	21.5	6.3	12.0	4.4
Sweden SE	*76.3*	57.9	*66.1*	36.1	*61.1*	*32.4*	*26.3*	*35.3*
Slovenia SI	50.7	47.9	52.1	35.6	32.9	*26.0*	21.9	*16.4*
Slovakia SK	44.1	57.7	43.6	44.9	35.7	14.5	15.9	5.7

4. STATISTICAL ANALYSIS[3]: METHODS AND RESULTS

4.1 CLASSIFICATION TREE

In order to identify homogeneous groups of countries with reference to the number of CE practices, among the eight considered in the survey, adopted by SMEs, we

[3] For the analyses the software R was used.

estimated a classification tree (CART) (Breiman et al., 1984). CART is a statistical method of a-priori segmentation, i.e, segments are identified on the basis of a criterion variable *Y*. In our application, this variable quantifies the adherence of a generic SME to the efficiency measures proposed in the questionnaire with the following three categories: 5 or more measures adopted (class 1), between 1 and 4 CE practices (class 2), no sustainable measures (class 3). From the dataset, it emerged that 36.4% of SMEs in the sample belong to the first class, 52.3% to the second one and 11.3% had not carried out any efficiency action in the reference period. As independent variable for the classification tree, we considered: the countries where the SME operates (28 levels). Considering these values for all SMEs operating in the 28 European countries, the CART algorithm identifies the best partition of countries through an iterative algorithm, which is reported in Table 3. These groups will be the starting point to analyze and quantify the heterogeneity between EU member states (MSs), in terms of adopted resource efficiency variable *Y* for a generic SME in each of the five groups of countries identified by the CART algorithm. The CART procedure can be divided into two distinct parts: tree growing and tree pruning. In the first part of the model, the response variable is repeatedly split, starting with the variable that has the highest association with it. The splitting continues till a predetermined stopping criteria (Zhang, 2016). In the pruning phase of the algorithm, which is performed on another subsample of data in order to avoid overfitting, the best grouping is identified on a basis of a measure of fit; we used total deviance.

Tab. 3: Groups of homogeneous countries: CART best partition.

Group of countries	
A	Austria, Belgium, France, Great Britain, Ireland, Portugal, Sweden Spain
B	Croatia, Germany, Italy, Malta, Slovakia, The Netherlands
C	Czech Republic, Denmark, Luxemburg, Poland
D	Cyprus, Finland, Greece, Hungary, Latvia, Slovenia
E	Bulgaria, Estonia, Lithuania, Romania

Tab. 4: Probabilities (%) of belonging to the three classes of variable *Y* in the five groups of countries

Group of countries	Class 1 5+ actions	Class 2 1-4 actions	Class 3 No actions
A	42.63	51.95	5.42
B	30.98	60.76	8.26
C	28.89	56.45	14.66
D	20.60	57.77	21.63
E	8.68	58.32	33.01

Groups of countries in the Tables 3 and 4 are ordered by decreasing probabilities of adopting CE practices by SMEs operating in their territories. There is a big difference of behavior between firms in groups A and E; for example, the percentage of SMEs that do not implement efficiency actions increases from 5.42% to 33.01%. In group A we find the European countries with the greenest businesses; in group E, the least green ones. Group B includes countries belonging to Western and Southern Europe; the proportion of SMEs not activating sustainability practices is a bit higher than that of group A, 8.26%. Group C is composed of only four countries, where a high percentage of SMEs is adopting at least one CE practice. Countries in group D are the most different for what concerns location in Europe and with one fifth of them not implementing CE practices. Finally, countries in group E are all located in Eastern Europe.

4.2 MULTILEVEL ANALYSIS

The CART analysis identified homogeneous groups of European Union MSs, confirming heterogeneity of behavior between countries. In each group, however, it is important to explore further differences among SMEs, especially in relation to their characteristics that might affect their decision to adopt CE. The identification of homogeneity and heterogeneity among European countries must be followed by an appropriate analysis of heterogeneity within each MS. As written in the introduction, there is a rich literature on the factors affecting companies' decision to comply with sustainability; some of these evidences are confirmed also by the descriptive analyses of our data as reported in section 1. In order to face this problem, it is necessary to use statistical methods for hierarchical data (Hox, 2002). In our specific sample, SMEs are nested into countries and this originates a multilevel dataset. In this section of the paper, we will explore how decisions towards CE actions are related to factors as size, turnover, number of workers, etc., i.e., we aim to quantify how the probability of adopting CE actions is determined by SMEs characteristics. A second goal of the analyses is to understand how economic investment affects environmental measures and if this aspect varies among the groups of countries. And finally, we will look at the eight specific CE actions considered in the survey to evaluate which factors are specifically determinant in their adoption. We now indicate with Y_{ij} the number of resource efficiency actions undertaken by the *i*-th SME belonging to the *j*-th group with *j=1,..,28*, as we are considering the 28 EU MSs; Y_{ij} will assume values from 1 to 8. Our multilevel model is given by equation (1):

$$\ln\left(\mu_{ij}\right) = \beta_{o\,j} + \beta_{1\,j}\, x_{ij} + R_{ij}$$

where Y_{ij} is assumed to follow a Poisson distribution with mean μ_{ij}. In the random

intercept model, β_{0j} are random variables representing differences among groups:

$$\beta_{0j} = \gamma_{00} + U_{0j}$$

and U_{0j} is a random effect, at state level, following a Normal distribution with 0 mean and variance equal to σ_U^2. As independent variables (vector $\boldsymbol{X}$), we consider: SME dimension, the year of foundation, average turnover in the reference year, the sector of economic activity; the number of full-time workers employed in green jobs. Tables 5 and 6 list the result of the estimation of this multilevel random intercept model with our dataset, the estimates refer to the best fitting model assessed with the lowest values of AIC and BIC indexes.

Tab. 5: Estimated parameters: random intercept model

	Estimate	Standard deviation	p-value
Intercept	0.966	0.059	<0.001
# workers			
-9 ref. category			
10-49	0.097	0.019	<0.001
50-249	0.058	0.046	0.209
250+	0.119	0.067	0.041
Age			
-31 Dec 2010 ref. category			
1 Jan 2010-31 Dec 2012	-0.026	0.017	0.128
1 Jan 2013-31 Dec 2017	-0.007	0.015	0.620
1 Jan 2017+	-0.037	0.041	0.042
Turnover			
-100,000 ref- category			
100,00-500,000	0.048	0.013	<0.001
500,000-2 million	0.077	0.015	<0.001
2-10 million	0.086	0.022	<0.001
+10 million	0.265	0.030	<0.001
Sector			
Manufacturing ref. category			
Retail	-0.068	0.017	<0.001
Services	-0.166	0.016	<0.001
Industry	-0.062	0.018	0.002
# workers in green jobs			
0 ref. category			
1-5	0.320	0.010	<0.001
6-9	0.283	0.026	<0.001
10-50	0.512	0.025	<0.001
51-99	0.504	0.084	<0.001
100+	0.713	0.115	0.003
var(μ_j)	0.086	0.021	<0.001

Tab 6. Countries' effects: random intercept model

AT	3.261	EE	1.145	HU	2.398	NL	3.108
BE	3.348	ES	3.416	IE	3.668	PL	2.624
BG	1.730	FI	2.442	IT	2.734	PT	3.656
CY	2.172	FR	3.587	LT	1.718	RO	1.532
CZ	2.875	GB	3.954	LU	2.724	SE	3.350
DE	3.023	GR	2.247	LV	2.109	SI	2.721
DK	2.693	HR	2.770	MT	3.015	SK	2.337

The magnitudes of country intercepts mirror the groups obtained with the regression tree analysis. In general, from Table 5 we see which are the factors significantly affecting the decision to implement resource efficiency practices and how they might act. For example, yearly turnover and the number of workers employed in green jobs have a direct effect on the number of implemented actions, while the manufacturing is the sector where SMEs are more prone to sustainability activities. There is also a non-negligible effect of dimension and age, in the sense that bigger and older firms are more inclined to resource efficiency practices. For what regards dimension, small and large firms adopt a greater number of CE practices, while medium businesses do not.

Starting from the above evidences, we consider important to explore another aspect related to the effect of economic investments on sustainability. The question is "do equal investments for sustainable measures led to a proportional total adhesion to CE in the homogeneous groups of countries obtained with the CART segmentation procedure?". The dataset contains the information on the amount invested on average per year by each SME in order to become more resource efficient. This information gives rise to a categorical variable with the following classes: no investment, less than 1% of yearly turnover, between 1 and 5%, between 6 and 10% and more than 10%. Le Y_i the number of CE practices adopted by SME *i*, that is assumed to follow a Poisson distribution with mean μ_i, we estimate the model in equation (2) separately for the SMEs in the five groups of countries in Table 3.

$$ln(\mu_i) = \eta_i + x_{ik}\beta_k + z_i y \tag{2}$$

where $k=1,..,5$, x_k are the covariates used in the multilevel regression model and z is the categorical variable indicating the amount of yearly turnover invested to improve the sustainability of the business by SME *i*. Table 7 lists the results of the estimation of the generalized linear model (GLM) in equation (2) for the SMEs operating in the five groups of countries described in Table 3; i.e., 5 GLMs are estimated.

Tab. 7: GLMs: estimation results

Group A	Estimate of γ	Standard error	p-value
0% ref. category			
<1%	0.335	0.019	<0.001
1-5%	0.419	0.018	<0.001
6-10%	0.516	0.030	<0.001
>10%	0.610	0.045	0.002
Group B			
0% ref. category			
<1%	0.407	0.025	<0.001
1-5%	0.575	0.024	<0.001
6-10%	0.598	0.033	<0.001
>10%	0.726	0.056	<0.001
Group C			
0% ref. category			
<1%	0.700	0.039	<0.001
1-5%	0.743	0.040	<0.001
6-10%	0.746	0.059	<0.001
>10%	0.410	0.099	<0.001
Group D			
0% ref. category			
<1%	1.030	0.061	<0.001
1-5%	1.019	0.058	<0.001
6-10%	0,976	0.082	<0.001
>10%	1.016	0.099	<0.001
Group E			
0% ref. category			
<1%	1.311	0.095	<0.001
1-5%	1.347	0.105	<0.001
6-10%	1.208	0.153	0.003
>10%	1.506	0.144	<0.001

Values of estimated parameters clearly show that the higher the percentage of turnover invested, the higher the number of CE practice adopted by SMEs in all five groups of countries. However, this relationship has a different magnitude in the five groups, increasing from SMEs operating in countries classified in group A to SME operating in European countries classified in group E. In countries of group A, SMEs implement the highest number of sustainability practices, for this reason the result that in this group investments have the lowest impact deserves attention. In our opinion, this evidence shows that the transition from a linear economic system to a circular one in these countries, once started, does not require high

extra investments to be maintained, we can briefly say that a Circular Economy system, once implemented, continues to increase business sustainability, somehow self-expanding.

As a further analysis, we want to obtain a measure of adhesion for each state to the single actions studied; to answer this question, again we must apply a statistical method that takes into account the hierarchical nature of the data.

For this scope, we define a new variable Y_{ij} that takes value 1 if SME *i*, operating in European country *j*, adopted the specific considered efficiency action, while it takes value 0, otherwise; the following multilevel logit model in equation (3) is estimated for the eight surveyed actions:

$$\log\left[P\left(Y_{ij} = 1 \middle| \mu_j \right)\right] = x_{1ij}\,\beta_1 + x_{2ij}\,\beta_2 + x_{3ij}\,\beta_3 + x_{4ij}\,\beta_4 + x_{5ij}\,\beta_5 + \mu_j \qquad (3)$$

x_1 represents the number of workers, x_2 the age of the SME, x_3 average yearly turnover, x_4 the sector of economic activity and x_5 the number of workers employed in green jobs; μ_j is the random intercept with Normal distribution with mean 0 and variance σ^2_μ. Table 8 lists the values of the random intercepts for the eight models, estimated for the corresponding efficiency actions, obtained for each European country and refer to micro-enterprises with a number of employees between 1 and 9, founded before 1st January 2010, belonging to the sector manufacturing, with an average yearly turnover of less than 100 thousand euros and without any workers in green jobs. The eight actions are: minimize waste of water, minimize energy use, use of renewable energy, attention to raw materials, waste minimization, selling of waste to other companies, recycling of waste or others materials and designing of sustainable products ready for reuse, reuse or with minimal environmental impacts. Figures in the table show that preferences in adopting specific CE actions are different in the 28 MSs; these evidences emerged also from descriptive statistics in Table 2; they are confirmed taken into account the multilevel structure of our data.

Tab. 8: Random intercepts

	Minimizing waste	Saving energy	Saving materials	Saving water	Recycling	Design products	Selling scrap	Renewable energy
AT	0.35	0.63	0.23	0.51	0.56	0.26	0.38	0.25
BE	0.39	0.64	0.14	0.62	0.73	0.26	0.33	0.24
BG	0.26	0.33	0.04	0.34	0.30	0.22	0.14	0.12
CY	0.27	0.47	0.07	0.39	0.63	0.21	0.37	0.14
CZ	0.38	0.58	0.06	0.49	0.65	0.32	0.30	0.33
DE	0.28	0.62	0.23	0.56	0.57	0.22	0.29	0.22
DK	0.36	0.50	0.06	0.52	0.47	0.25	0.22	0.25
EE	0.17	0.29	0.04	0.28	0.17	0.19	0.15	0.08
ES	0.45	0.64	0.05	0.68	0.61	0.21	0.47	0.29
FI	0.21	0.42	0.09	0.51	0.50	0.20	0.24	0.21
FR	0.64	0.67	0.04	0.62	0.83	0.19	0.36	0.33
GB	0.50	0.63	0.13	0.65	0.84	0.34	0.66	0.26
GR	0.25	0.41	0.08	0.43	0.32	0.27	0.22	0.17
HR	0.38	0.53	0.05	0.58	0.57	0.25	0.22	0.15
HU	0.36	0.54	0.07	0.48	0.40	0.24	0.16	0.17
IE	0.47	0.60	0.12	0.57	0.81	0.25	0.60	0.22
IT	0.33	0.45	0.10	0.49	0.69	0.15	0.19	0.27
LT	0.33	0.41	0.04	0.38	0.23	0.21	0.08	0.11
LU	0.28	0.45	0.09	0.51	0.53	0.24	0.34	0.22
LV	0.32	0.48	0.03	0.51	0.31	0.18	0.12	0.15
MT	0.32	0.61	0.10	0.46	0.62	0.23	0.42	0.19
NL	0.26	0.58	0.20	0.62	0.64	0.29	0.30	0.20
PO	0.41	0.49	0.03	0.60	0.53	0.23	0.18	0.16
PT	0.54	0.67	0.06	0.73	0.51	0.27	0.58	0.40
RO	0.19	0.27	0.03	0.30	0.29	0.16	0.17	0.07
SE	0.27	0.47	0.24	0.62	0.73	0.25	0-32	0.28
SI	0.47	0.44	0.11	0.54	0.50	0.18	0.29	0.25
SK	0.36	0.49	0.04	0.41	0.39	0.42	0.27	0.13

The sector of economic activity and the dimension of the firm have a different effect on the different actions. For example, actions as 2 (minimize energy use), 3 (use of renewable energy) and 6 (sale of waste to other companies), which require large investments, are more chosen by larger SMEs.

5. CONCLUDING REMARKS

The scope of this paper is to investigate differences in behavior towards sustainability practices of European SMEs. Heterogeneity emerges both between and within

European countries.

Segmentation analysis identified five groups of European countries, homogeneous for the attitude of SMEs to CE actions. In eight states, Austria, Belgium, Spain, France, Great Britain, Ireland, Portugal and Sweden, (group A), firms show the highest level of innovation in the field of sustainability, only a very small percentage of businesses (5.4%) do not adopt any efficiency measure. In this group of countries, the average number of green actions implemented by each SME is 3.9, out of the eight investigated by the Flash Eurobarometer survey. The estimation of a multilevel regression model shows a similar behavior among the eight nations, i.e., low level of within group heterogeneity. On the opposite side we find SMEs that operate in countries classified in group E, Latvia, Lithuania, Romania and Estonia, where we found the lowest attention to Circular Economy practices; SMEs adopt on average only 1.4 CE actions.

For what concerns firms' characteristics, the yearly turnover and the sector of economic activity proved to be significant in determining an efficient business management; SMEs in the manufacturing sector are the most inclined to perform green actions. The presence or absence of employees involved in green jobs is another important factor. The proportion of yearly turnover invested in sustainability directly affects the number of resource efficiency actions implemented by European SMEs, i.e, as investment increases, more actions are implemented; however, this effect is not the same in its magnitude in the five groups of homogenous countries, it is lower in the group of greenest countries. This result shows that investment in sustainability has decreasing marginal returns on implementation of resource efficiency actions.

As a general consideration, results from our analyses show that, for what regards CE attitudes in European SMEs, there is a lot of between and within country heterogeneity. Policies that aim to increase CE practices adoption must take these differences into account and therefore should be tailored for specific SMEs within each country. For what concerns countries, differences are related to geographical location, SMEs in Western-European countries exhibit more attention to green matters that SMEs in Eastern-European countries. However, there are some exceptions. Moreover, the geographical location is strictly correlated to economic and social conditions in the European MSs. A limitation of this study is that we did not insert in our models covariates collected at country level; this is a topic that deserves further attention. Further attention has to be paid also to the specific CE actions; also in this case there are difference between and within countries. These differences are related to SMEs' characteristics and to costs of implementation. Incentives to favor green economy should consider these elements as well.

REFERENCES

Abarca Guerrero, L., Maas, G. and Hogland, W. (2020). Solid waste management challenges for cities in developing countries. *Waste Management.* 33: 220-232.

Andersen, M.S. (2007). An introductory note on the environmental economies of the circular economy. *Sustainability Science.* 2: 133-140.

Bassi, F. and Dias, J.D. (2019). The use of circular economy practices in SMEs across the EU. *Resources, Conservation & Recycling.* 146: 523-533.

Bassi, F. and Guidolin, M. (2021). Resource Efficiency and Circular Economy in European SMEs: Investigating the Role of Green Jobs and Skills. *Sustainability.* 13: 12136.

Bianchi, R. and Noci, G. (1998). Greening SMEs' Competitiveness, Business Economics. *Small Business Economics.* 11: 269-281.

Boulding, K. (1966). The Economics of the Coming Spaceship Earth. In H. Jarrett, Editor, *Environmental Quality in a Growing Economy, Resources for the Future.* Johns Hopkins University Press, Baltimore: 3-14.

Bradford, J. and Fraser, E.D.G. (2008). Local authorities, climate change, and small and medium enterprises: identifying effective policy instruments to reduce energy use and carbon emissions. *Corporate Social Responsibility and Environmental Management.* 15: 156-172.

Breiman, L., Friedman, J., Stone, C.J. and Olshen, R.A. (1984). *Classification and Regression Trees.* Chapman and Hall, London.

Burger, M. and Douformont, J. (2019). The heterogeneous skill-base of circular economy employment. *Research Policy.* 48: 248-261.

Cadil, J., Mirosnik, K., Petkovova, L. and Mirvald, M. (2018). Public support of private effects on economic sustainability. *Sustainability.* 10: 4612.

Choongo, P., Van Burg, E., Paas, L.J. and Masurel, E. (2016). Factors influencing the identification of sustainable opportunities by SMEs: empirical evidence from Zambia. *Sustainability.* 8: 81.

Dalhammar, C. (2015). The application of "life cycle thinking" n European environmental law: theory and practice. *Journal for European Environmental and Planning Law.* 12: 97-127.

Dalhammar, C. (2016). Industry attitudes towards ecodesign standards for improved resource efficiency. *Journal of Cleaner Production.* 123: 155-166.

De Wit, G., Uhlaner, L., Berent-Braun, M. and Jeurissen, R. (2010). Family ownership, innovation and other context variables as determinants of sustainable entrepreneurship in SMEs: An empirical research study. *Scales Research Reports H201006.* EIM Business and Policy Research.

Ellen MacArthur Foundation (2015). *Towards the Circular Economy: Economic and Business Rationale for an Accelerated Transition.* Isle of Wight.

European Commission, Directorate-General for Enterprise and Industry (2002). *European SMEs and social and environmental responsibility.* Publications Office, Brussels.

European Commission (2003). *Official Journal of the European Union 20/05/2003.*

European Commission (2008). *Sustainable Consumption and Production and Sustainable Industrial Policy Action Plan* (COM(2008) 397). Brussels.

European Commission (2014). *Towards a Circular Economy: a Zero Waste Program for Europe.* B. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52014DC0398R%2801%29. Last access: 05/01/22.

European Commission (2015). *Closing the Loop - an EU Action Plan for the Circular Economy, 614 Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions.* http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52015DC0614. Last access: 05/01/22.

European Commission (2018a). *Behavioral Study on Consumers' Engagement in the Circular Economy.* Brussels.

European Commission, Executive Agency for Small and Medium-sized Enterprises, Doranova, A., Mueller, M., Zhechkov, R., et al. (2018b). *Green action plan for SMEs: addressing resource efficiency challenges and opportunities in Europe for SMEs: implementation report, Publications Office.* https://data.europa.eu/doi/10.2826/031664. Last access: 05/01/22.

European Commission (2019). *Report from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions.* Brussels.

European Commission (2020a). *Circular Economy Action Plan.* Brussels.

European Commission, Executive Agency for Small and Medium-sized Enterprises, Schroder, J., Farrenkopf, J., Pedersen, B., et al. (2020b). *Annual report on European SMEs 2018/2019: research & development and innovation by SMEs: background document, Publications Office,* https://data.europa.eu/doi/10.2826/603707. Last access: 05/01/22.

Geng, Y. and Doberstein, B. (2008). *The Further Benefits of Business Resource Efficiency.* Department for environment, Food and Rural Affairs, London.

Ghisellini, P., Cialani, C. and Ulgiati, S. (2016). A review on circular economy: the expected transition to a balanced interplay of environmental and economic system. *Journal of Cleaner Production.* 114: 11-32.

Hogg, D., Sherrington, C. and Vergunst, T. (2011). *A Comparative Study on Economic Instruments Promoting Waste Prevention.* Eunomia Research and Consulting, Bristol.

Hoogendoorn, B., Guerra, D. and van der Zwan, P. (2015). What drives environmental practices of SMEs? *Small Business Economics.* 44: 759-781.

Hox, J. (2002). *Multilevel Analysis: Techniques and Applications. Quantitative Methodology Series.* Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US.

Inigo, E.A. and Blok, V. (2019). Strengthening the socio-ethical foundations of the circular economy: Lessons from responsible research and innovation. *Journal of Cleaner Production.* 233: 280-291.

Kirchherr, J., Reike, D. and Hekkert, M. (2017). Conceptualizing the circular economy: An analysis of 114 definitions. *Resources, Conservation & Recycling.* 127: 221-232.

Lieder, M. and Rashid, A. (2016). Towards circular economy implementation: a comprehensive review in context of manufacturing industry. Journal of Cleaner Production. *115: 36-51.*

Liu, Z., Adams, M., Cote, R.P., Chen, Q., Wu, R., Wen, Z., Liu, W. and Dong, L. (2018). How does circular economy respond to greenhouse gas emissions reduction: An analysis of Chinese plastic recycling industries. *Journal of Renewable and Sustainable Energy Reviews.* 91: 1162-1169.

Najmi, A., Kanapathy, K. and Aziz, A.A. (2020). Exploring consumer participation in environment management: Findings from two-staged structural equation modelling-artificial neural network approach. *Corporate Social Responsibility and Environmental Management.* https://doi.org/10.1002/csr.2041.

Pearce, D.W. and Turner, R.K. (1989). *Economics of Natural Resources and the Environment.* Johns Hopkins University Press, Baltimore.

Reike, D., Vermeulen, W.J.V. and Witjes, S. (2018). The circular economy: New or Refurbished as CE 3.0? Ñ Exploring Controversies in the Conceptualization of the Circular Economy through a Focus on History and Resource Value Retention Options. *Resources, Conservation & Recycling. 135: 246-264.*

Sharma, N.K., Govindan, K., Lai, K.K, Chen, W.K. and Kumar, V. (2021). The transition from linear economy to circular economy for sustainability among SMEs: A study on prospects, impediments, and prerequisites, *Business Strategy and the Environment.* 30: 1803-1822.

Spangenberg, J.H. (2002). Economic sustainability of the economy: concepts and indicators. *International Journal of Sustainable Development*. 8: 47-64.

Stahel, W.R. (1982). *The Product-Life Factor*. http://www.product-life.org/en/major-publications/the-product-life-factor. Last access: 05/01/22.

Stahel, W.R. and Reday, G. (1976) *The potential for substituting manpower for energy. Report to the Commission of the European Communities.* Brussels.

Su, B., Heshmati, A., Geng, Y. and Yu, X. (2013). A review of the circular economy in China: moving from rhetoric to implementation. *Journal of Cleaner Production*. 42: 215-227.

Suárez-Eiroa, B., Fernández, E., Méndez-Martínez, G. and Soto-Oñatec, D. (2019). Operational principles of circular economy for sustainable development: Linking theory and practice. *Journal of Cleaner Production*. 214: 952-961.

The Standing Committee of the National People's Congress China (2008). *Circular Economy Promotion Law of the People's Republic of China*.

Trencansky, D. and Tsaparlidis, D. (2014). *The effects of company s age, size and type of industry on the level of CSR.* Umeå School of Business and Economics.

/beda-García, M., Claver-Cortés, E., Marco-Lajara, B. and Zaragoza-Sáez, P. (2021). Corporate social responsibility and firm performance in the hotel industry. The mediating role of green human resource management and environmental outcomes. *Journal of Business Research.* 123: 57-69.

Vermeulen, W.J.V., Reike, D. and Witjes, S. (2019). Circular Economy 3.0 - Solving confusion around new conceptions of circularity by synthesising and re-organising the 3R's concept into a 10R hierarchy. *Renewablematter*. 27: 1-15

Yuan, Z., Bi, J. and Moriguichi, Y. (2006). The circular economy: a new development strategy in China. *Journal of Industrial Ecology*. 10: 4-8.

Zhang, Z. (2016). Decision tree modeling using R, *Annals of Translational Medicine*. 4: 275.

JOB SATISFACTION AND TELEWORKING: A STUDY ON PUBLIC ADMINISTRATION WORKERS IN ITALY

Stefania Capecchi[1]

Department of Political Sciences, University of Naples Federico II, Naples, Italy

Giustina Orientale Caputo[2]

Department of Social Sciences, University of Naples Federico II, Naples, Italy

Abstract. This paper presents the results of one of the first surveys carried out in Italy on the living and working conditions of public administration employees who were engaged in compulsory telework during the first stages of the COVID-19 pandemic (March–May 2020). Although this study examines a small sample of public workers in Campania region, interesting results emerge in a modelling implementation. In fact, by means of a heteroskedastic Ordered Probit model, some findings are presented with job satisfaction being the response variable. Considering the workers' need to adjust to a completely novel situation, our results reveal a significant role played by a potential lack of concentration and by the satisfaction of using their own home as a workplace as well as by the differences experienced in work efforts. The presence of children in the household turns out to be slightly significant, whereas childcare duties do seem to exert some impacts on job satisfaction, implying relevant effects on work-life balance. Workers' concern regarding a possible lack of recognition of their job by supervisors or managers is also highlighted.

Keywords: Job Satisfaction, Teleworking, Covid-19, Ordered Probit Model

1. INTRODUCTION

In 2020, the global COVID-19 pandemic crucially altered global economies and people's ordinary social and working lives. To contain the spread of the virus, governments imposed various social-distancing measures and, as a result, private and public companies began experimenting with strategies to reduce the number of people in contact with each other, thus making it possible for employees to work at their homes. Consequently, all over the world an unprecedented number of workers

[1] Corresponding author: Stefania Capecchi, ORCID: 0000-0003-2842-8132, e-mail: stefania.capecchi@unina.it

[2] Giustina Orientale Caputo, ORCID: 3.0000-0002-5897-4786, e-mail: giustina.orientale @unina.it

were asked to or allowed to work from home, after decades of often ineffective attempts to develop the use of telework (Ogbonna and Harris, 2006, Welz and Wolf, 2010; Eurofound, 2010 and 2012; Pyöriä, 2011).

In Italy, teleworking in public administration ("*lavoro agile*" in Italian or "smart working" in everyday language) was introduced by Law No. 124/2015 and is presently regulated by Law No. 81 of 2017. Until the outbreak of the COVID-19 pandemic, such modality of work was not common. In fact, with the arrival of the restrictive measures due to the pandemic, in combination with the *low uptake* of teleworking in the past decades, Italy displayed one of the largest increases in the figures of employees working from home among European Union (EU) member states: an upsurge of 39.9% as compared to an average EU27 growth close to 36.5% (Eurofound, 2021; EU-OSHA, 2021b). According to data from the National Observatory on Smart Working, an agency of the Ministry of Public Administration (Osservatorio Nazionale del Lavoro Agile, 2020), at the beginning of February 2020 only 2.5% of public employees were classified as teleworkers. Within just a few weeks, this figure had risen to almost 65%. In such a short period, about 86% of public administration offices activated forms of "smart working" for at least part of their staff, and this percentage rose to 99% when considering administration offices with more than 50 employees.

With specific reference to such circumstances, this paper presents the results of one of the first studies of the living and working conditions of public administration personnel in Italy who began teleworking during the first lockdown period, from March to May 2020. The survey was conducted with our scientific support on behalf of the *Confederazione Generale Italiana del Lavoro, Funzione Pubblica* (CGIL for Public Sector), a largely representative Italian trade union. More specifically, the employed data stem from an observational study aimed at analysing the conditions of public administration workers in the Campania region and in the province of Naples during the lockdown periods, to investigate the unexpected large-scale application of emergency "smart working" arrangements. The survey was designed to determine, on the one hand, the employees' satisfaction with work conducted under these unprecedented conditions and, on the other hand, the material and environmental conditions under which the telework was carried out. Finally, a further focus of the investigation was the complex balance, sometimes resulting in disruption, between private and working life as compared to the pre-existing circumstances. The fieldwork of the survey was conducted by CGIL for Public Sector, Naples and Campania section, which administered the questionnaire directly through its social media channels.

By means of a heteroskedastic Ordered Probit model, some results are presented with job satisfaction being the response variable. The paper is organised as follows. After a brief review of the recent literature on teleworking conditions in Section 2, data of interest and results of the estimated model are presented in Section 3, while Section 4 discusses the main findings and presents some concluding remarks.

2. LITERATURE REVIEW AND MOTIVATION

During the COVID-19 pandemic, teleworking has been regarded as the most effective and cost-efficient approach to preserve and restore the functioning of the entire economy (among others: ILO, 2021 and 2022). Apart from "native teleworkers", those employed generally as telemarketers and customer care staff, home-based work assisted by information and communication technology (ICT) was regarded with some degree of scepticism by both employers and workers themselves. On the employers' side, teleworking was often considered as likely to lead to low productivity because of the lack of direct control (Harker Martin and MacDonnell, 2012; Hilbrecht et al., 2013; Putnam et al., 2014; Messenger, 2019). On the workers' side, two different attitudes can be distinguished. First, employees' concerns about career dynamics were heightened by the physical distance from the company premises, due to the perceived difficulty of recognition by managers or supervisors of their work performance. Second, blurred feelings of interest and aversion have arisen towards a condition often perceived as a "privilege" experienced especially by public sector employees in the perspective of reducing their difficulties in balancing work and non-work duties, and mostly in the case of working women (Chung and Van der Horst, 2018; Kaduk et al., 2019).

However, despite media interest and both corporate and academic debates on the potentials of teleworking, only a comparatively low number of establishments and organisations had adopted some home-based teleworking practices in the past decades. Until March 2020, primarily because of the inadequacy of legal and welfare frameworks, working from home essentially remained a seductive proposition in highly developed countries (Baruch, 2001). Even though the advances in ICT have undoubtedly triggered improvements at both business and societal levels, home-based teleworking benefits have long been regarded as a preferential treatment (Parry et al., 2021).

Additionally, for some scholars, teleworkers seem to be exposed to greater levels of stress than their office worker counterparts, even presenting additional physical health symptoms. Mann and Holdsworth (2003) underline some practical

benefits of teleworking, such as increased flexibility, less commuting time, and a better work-life balance in general, while clearly highlighting the potentially unfavourable consequences of telework on workers' mental health, such as perceived loneliness, social isolation, blurring of boundaries, and presenteeism (i.e. the lost efficiency which occurs when employees are not fully functioning in the workplace because of an illness, injury, or other condition) (see also: Steidelmüller et al., 2020).

Wide-ranging practices of flexible working settings have appeared only with the outbreak of COVID-19, and their effects are compared using "before" and "after" benchmarks (Parry et al., 2021; Dunatchik et al., 2021). Furthermore, a new digital divide based on teleworkability is now discussed in literature as a possible driver of increased disparities (among others, Fana et al. 2020; Sostero et al., 2020). Teleworkability indicates the degree to which an activity can be performed remotely thanks to ICT devices, thus implying that job tasks requiring physical handling or duties must necessarily be performed on-site, at the employers' premises (ILO, 2021). Such a concept is essential for properly exploring the impacts of the 2020–2021 actions to develop telework practices in the coming years.

As teleworking remained a marginal issue in the past, mostly confined to private companies and adopted in a few countries, its implementations were rarely the subject of detailed statistical surveys in Western developed countries. With reference to EU countries, the scant statistical evidence on teleworking across member states has been obtained by extrapolating relevant information from surveys focusing on other related topics, such as the European Working Conditions Survey, the European Survey of Enterprises on New and Emerging Risks, and the European Labour Force Survey itself (EU-OSHA, 2021a and 2021b). Although not up to date, these official statistics are currently the only ones that can be consulted to obtain representative information.

During the first months of the 2020 pandemic, Eurofound conducted a non-statistically representative survey across European countries. Moreover, in 2021 a special wave of the European Working Conditions Survey was implemented to provide comparable and representative data on working conditions during the pandemic across EU 27 member states and other European countries, but those micro-data are not available yet. For the specific Italian context, the situation is similar, since most of the studies conducted in the past two years are qualitative or do not meet the requirements of reliable sample surveys (Eurofound, 2021). The same can be said for the public sector across Europe and in Italy as well. Consequently, we have chosen to employ information stemming from an observational study developed in collaboration with CGIL Campania, which is long

established[3] in the public sector, to examine the response patterns of a small sample of workers towards job satisfaction during the first lockdown period in the Campania region.

3. DATA AND METHODS

The research targeted the five provinces of Campania, with the aim of assessing organisational aspects and impacts on both working life and perceived job quality during the pandemic, as perceived by public employees.

To set the scene, it should be mentioned that, according to the Ministry of Economy and Finance (2020), in 2019 the total number of public employees[4] in Italy was 3,186,014. In the Campania region, 279,077 people were employed in the public sector, thus representing about 8% of the national aggregate. Actually, in the city of Naples, belonging to a large metropolitan area (province), the absolute value is 45,947. Our respondents come mainly from local functions based in the area of Naples. In particular, most of the interviewees (78.5%) live in the province of Naples, with 7.2% in the province of Salerno, 6.1% in both Avellino and Caserta, and only 2.15% in the province of Benevento.

This disproportion is related to the fact that the responses were collected by a trade union, which traditionally has a stronger presence in more sizeable public offices, therefore helping to explain the larger ratio of respondents from Naples. The fieldwork, in fact, was carried out directly by CGIL Campania, which disseminated 320 questionnaires through its online social media channels between October 2020 and January 2021.

The administered questionnaire focuses on the analysis of workers' conditions following the emergency imposition of remote work practices, with particular reference to: i) assessed job satisfaction with respect to the tasks completed under

---

[3] According to the data provided by ARAN (Agency for the Negotiation Representation of Public Administrations), in the three-year period 2019-2021, when distinguishing public employment by sector, for the local functions the percentage distribution of union membership saw CGIL covering 34.3% of unionised workers, CISL 27.4% and UIL 18%. Hence, the CGIL is considered the most representative trade union in the labour sector investigated (see: https://www.aranagenzia.it/rappresentativita-sindacale-loader/rappresentativita/triennio-2019-2021-provvisorio.html).

[4] The contingent of public employees (excluding those with flexible, temporary, or other non-standard contracts) in Italy is 3,186,014. Of these, 7.1% of the employed people belong to the Central Functions sector, while 38.3% of the employees belong to the Education and Research sector. Three sectors register a number of employees close to each other: about 20% are employed in the public Health sector, while those with public law contracts are about 17.8%; the amount of employees in the local government sector is 15.5% of the total. A marginal quota, out of these sectors, accounts for 1.5% of the total.

novel and unexpected circumstances; ii) physical and environmental conditions in which their work was carried out; and iii) suitability or disruption of the work-life balance in comparison to the previous period.

In addition to the usual demographic variables (gender, marital status, education, composition of the household, type of work), information[5] was collected on the type of public institution to which respondents belonged, as well as on their work organisation, such as workspace at home, working time and procedures. Perceptions and assessments regarding job satisfaction, work-life balance, and relationships with colleagues and supervisors were also considered. Overall, 279 individuals answered the survey, resulting in a response rate of approximately 87%, thus confirming the effective interest of respondents towards the topics of the questionnaire.

The sample, of course, is small; nevertheless, the positive response rate is quite consistent with such types of investigations. The main descriptive statistics to understand the composition of the sample of respondents are presented below.

Our sample of interest is equally distributed by gender; 72.8% of the respondents are married or declare to have a partner, while 10% are single. As far as age is concerned, 50.2% of the sample is in the 35–54 age group, while 43% of the sample is between 55 and 67 years old, and only 7% are younger than 34 years. Because of this age composition, the presence of young children of primary school age is extremely limited, which means that parental care duties may not influence the satisfaction of working from home in our sample. However, 72% of the respondents state that they do have children, and 36.2% of the entire sample report that their partner has a job. When considering only the married respondents, almost 41% affirm that their partner is employed.

Regarding the composition of the household, 8.5% of the sample declare that they live alone (regardless of their marital status), 54.7% of the sample say they live in a household of two or three people, and 36.8% report living in households of more than three people.

With respect to the size of the respondents' employer, as indicated in Table 1, more than 56.6% of the workers belong to an administration office with more than 1,000 employees.

The distribution by workers' reported net income is presented in Table 2 and includes four income classes. Most of the sample report an annual income from employment ranging from 20,000 to 30,000 euros per year.

---

[5] Data are available from Authors on request.

Tab. 1: Respondents by Administration Size

Administration Size	Freq.	Percent	Cum.
< 15	10	3.58	3.58
15–50	13	4.66	8.24
51–100	32	11.47	19.71
101–250	38	13.62	33.33
251–500	18	6.45	39.78
501–1,000	10	3.58	43.37
> 1,000	158	56.63	100.00
Total	279	100.00	

Participants were also asked to respond to questions about various aspects of family and work life, using a Likert-type scale to assess their degree of agreement or disagreement with the statements. The results are depicted in Figure 1.

Our variable of interest is the satisfaction with telework. The original variable is expressed on an 11-point scale from 0 to 10 (with 0=totally disagree and 19=totally agree) and presents a substantial dispersion, with a considerable number of responses assigned to the "balance" modality of the scale (rating=5), and quite high ratings assigned to both the low modalities and the high ones, as seen in Figure 1.

Tab. 2: Respondents by Income Classes

Income classes (Euros)	Freq.	Percent	Cum.
< 20,000	56	20.07	20.07
20,001–30,000	154	55.20	75.27
30,001–40,000	55	19.71	94.98
> 40,000	14	5.02	100.00
Total	279	100.00	

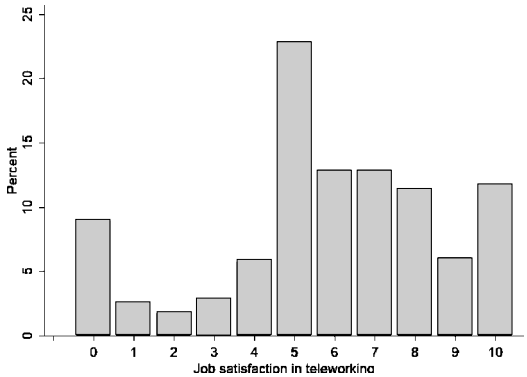


Fig. 1: Job Satisfaction in Teleworking, Original Likert Scale (percentage)

We also explore interviewees' satisfaction with their own "home as a place of work". Furthermore, the perceived "lack of recognition" from supervisors or managers regarding work performed in teleworking mode is investigated, as well as the employees' perception of a "never-ending" working day. All of the considered variables are expressed on the same 11-point scale.

As indicated in Table 3, nearly 6% of the sample declare to be totally dissatisfied (rating=0) with their home as a place of work, while more than 43% reported a high or very high level of satisfaction (ratings between 7 and 9). Additionally, 20.43% affirm that they are more than satisfied (rating=10). Conversely, it is remarkable that the concern that one's work may not be fully acknowledged due to teleworking is diversely distributed. In fact, while more than 16% of respondents do not rate this as a prominent concern, approximately 22% assign a medium-high rating (between 7 and 9) and as many as 17% claim they fear a significant lack of appreciation of their work (rating=10). With respect to the long working days which "do not seem to finish", responses indicate that 17% totally disagree, 19% totally agree, and 10% express an "intermediate" answer, leaving the remaining categories equally distributed.

Tab. 3: Respondents' Self-Assessments (percentages)

Level of proposed scale	Satisfaction towards home as a place of work	Lack of recognition of work done in teleworking	Never-ending working days
0	6.1	16.1	17.6
1	2.2	4.7	3.9
2	2.2	4.3	3.9
3	2.2	4.3	3.6
4	3.9	6.8	2.9
5	8.2	14.7	10.4
6	11.8	10.4	9.3
7	15.4	7.5	7.2
8	19.0	9.0	14.7
9	8.6	5.0	7.5
10	20.4	17.2	19.0
Total	100.0	100.0	100.0

Similar quite diverse response patterns are registered for other self-reported assessments, such as the level of concentration experienced during teleworking. In fact, approximately 58% of respondents assert that their work effort has increased overall while teleworking, whereas 38.8% do not report any change. With respect to housework and other unpaid duties, 61% of respondents state they share child-

rearing tasks with their partners, 23% report caring for their offspring alone, and only 10% said they received some help with parental duties from people outside the family. Finally, approximately 61% of interviewees report incurring some additional expenses for useful ICT devices to enable them to telework from home.

3.1 ESTIMATED MODEL RESULTS

Various modelling approaches can be applied in the case of ordinal responses. Agresti (2010), Tutz (2012), and Piccolo and Simone (2019) are the main references in this field. Taking into account the small sample size and the nature of variables, in our opinion, a simple Ordered Probit model is the most suitable methodology. However, due to the dispersion of responses to the job satisfaction question on the original 11-point scale, the response variable was conveniently recoded on a three-level scale (see Table 4), and the probability of being "dissatisfied", "indifferent", or "satisfied" (*job satisfaction*) is studied using an Ordered Probit model.

Tab. 4: Recoded Dependent Variable

Telework satisfaction (original modalities)	Freq.	Percent
0 dissatisfied (0–4)	61	21.86
1 indifferent (5–7)	136	48.75
2 satisfied (8–10)	82	29.39
Total	279	100.00

Given the available information, the model takes into account a number of basic demographics: *gender* (dummy variable, 1=woman), a dummy variable for presence of children in the household (*children*), and *marital status* (categorical variable, 1=single; 2=married; 3=other). Moreover, *age* classes, dimension of public administration office (*size*), *income*, and *expenses* for buying ICT devices to work at home are inserted. Finally, some ordinal variables are considered: the expressed level of satisfaction with home as workplace (*home as wp*), perceived differences in efforts pursued at work (*work effort change*), lack of concentration (*concentration*), endless working days (*neverending*), perceived lack of recognition of work done (*lack of recog*), and activities devoted to care for children (*carechildren*).

In the Ordered Probit model, the probability of an outcome *j* is given by the probability that the estimated linear function, plus random error, lies within the range of the estimated cut-points for the outcome. Formally, the model is:

$$Pr(Y_i = j) = Pr(k_{j-1} < \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_m x_{mi} + u_i \leq k_{J-1}) \quad (1)$$

where $u_i \sim N(0, \sigma^2)$, coefficients ($\beta_1 \ldots \beta_m$), and cut-points ($k_1 \ldots k_{J-1}$) are the parameters to be estimated, J is the number of possible outcomes; and $i = 1 \ldots n$, k_0

is taken as $-\infty$ and k_J is taken as $+\infty$. When in binary or an ordinal regression model the homoskedastic error hypothesis is incorrectly assumed, the standard errors are wrong, and the parameter estimates are biased (Yatchew and Griliches, 1985). Therefore, the inferential conclusions based on the usual z-test statistics can be misleading. To address the potential heteroskedasticity within the data, Williams (2009 and 2011) proposed the heteroskedastic ordered models, in which the factors affecting the heteroskedasticity are explicitly specified.

In particular, in heteroskedastic ordered models, the log-variances are specified by:

$$\log(\sigma_i^2) = \sum_{j=1,h} z_{ij} y_j \qquad j = 1 \ldots n \tag{2}$$

where z_{ij} is the value assumed by variable Z_j for the i-th observation. The vector $Z = (Z_1, Z_2, \ldots, Z_h)$ may include dummy or continuous variables and define groups with different error variances.

The estimated coefficients for the model are presented in Table 5 as obtained by maximising the likelihood function with the ordinal generalised linear models (OGLM) package in STATA14 (Williams, 2011). Given the results, it is possible to assume that, since the survey refers to the beginning of a strict lockdown period, response patterns appear to be mainly affected by a potential lack of concentration and by the availability of a comfortable home, suitable as a workplace. Considering workers' necessity to adjust to a completely new situation, the impacts of such variables are somehow expected, since they are, of course, strongly interconnected.

The presence of children is significant at 10%, since it should be considered that, as mentioned, children in the households are almost all in their late teens, and, therefore, substantially autonomous. The variable *carechildren* is significant in the variance equation, thus indicating some impact of the overall amount of usual family caregiving duties.

No statistical significance is attached to age, gender, and marital status (80% of the sample were married), consistent with the respondents' status of public employees, therefore presumably sharing similar overall conditions. Income class, instead, is slightly significant.

To consider the different effects of determinants, the model was estimated with the variance taking into account factors such as age, caring for children as a parental duty, and a variable referred to the lack of acknowledgement of the work carried out (*lack of recognition*). The latter driver turns out to be significant only in the variance equation but not as a general explanatory variable in the model: In fact, it was deleted from the model main equation, so as not to overload it with non-significant explanatory variables.

Tab. 5: Heteroskedastic Ordinal Probit for Teleworking Job Satisfaction, Coefficient Estimates, Variance Equation Coefficients, and Cut Points Estimates

TW job satisfaction	Coef.	Std. Err.	Z	P > z	
Gender	0.009	0.136	0.060	0.948	
Children	0.354	0.211	1.680	0.093	*
Marital status	-0.174	0.143	-1.210	0.225	
Age	-0.098	0.074	-1.320	0.185	
Size	0.045	0.036	1.250	0.211	
Income	-0.189	0.111	-1.700	0.090	*
Home as wp	0.157	0.066	2.370	0.018	**
Work effort change	0.351	0.166	2.110	0.035	**
Concentration	0.368	0.141	2.610	0.009	***
Neverending	0.008	0.023	0.340	0.735	
Expenses	0.307	0.168	1.820	0.068	**
Log(sigma) equation					
Children0.350	0.225	1.550	0.120		
Age class	-0.070	0.097	-0.730	0.468	
Lack of recognition	0.041	0.023	1.790	0.073	*
Carechildren	-0.165	0.083	-1.970	0.048	**
Cut point1	1.534	0.794	1.930	0.053	
Cut point2	3.142	1.241	2.530	0.011	
Pseudo R^2= 0.21			LR test $\chi^2(15)$=119.83		

***: significant at 1%; **: significant at 5%; *: significant at 10%

The discussion of the results from the estimated model must take into account that in the Ordinal Probit model the magnitude and sign of the coefficients cannot be interpreted by themselves, so it could be worth examining specific response profiles. In particular, we analyse the probability to be “very satisfied” and “not satisfied” as a function of the level of satisfaction with home as a workplace and the perceived variation in work effort, assigning to all the remaining variables in the model their median value, considering a female individual with children who has encountered some expenses to be able to work from home.

In general, as it may be observed in Figure 2, the probability of being more satisfied with telework clearly grows as the degree of satisfaction with home as a place of work increases and the work efforts varies. This circumstance is more evident in case of greater work effort, meaning that home conditions do play a prominent role. On the contrary, the probability of being less satisfied is undoubtedly higher for the same conditions, when respondents’ satisfaction for home as a

workplace declines (Figure 3), even in case of a lower work effort. It should be noted that since gender is not statistically significant, the corresponding estimated probabilities for a male with the same characteristics are almost identical.

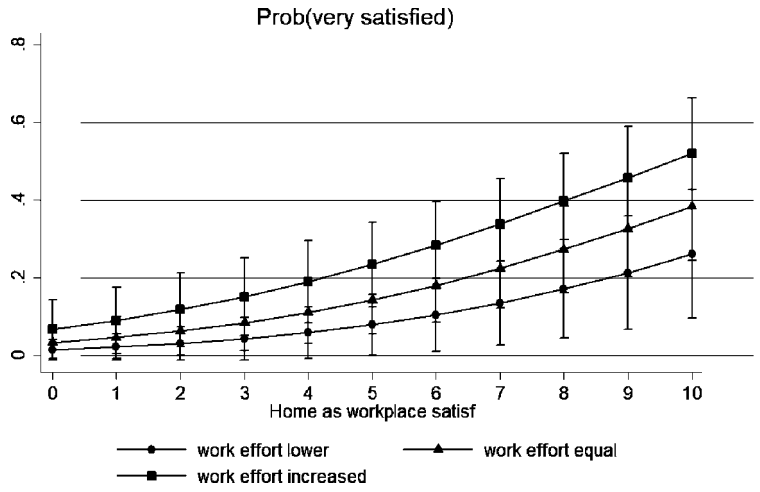


Fig. 2: Marginal effects of being "very satisfied" with telework for a female respondent with children, by varying work effort

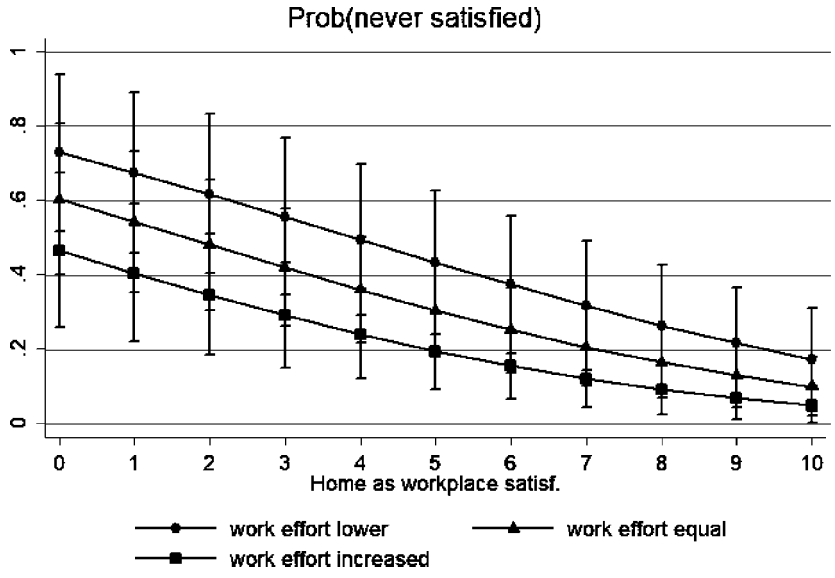


Fig. 3: Marginal effects of being "not satisfied" with telework for a female respondent with children, by varying work effort

4. DISCUSSION AND CONCLUDING REMARKS

The probability of being satisfied with telework, as well as with using their own home as a place of work, is rather high for the surveyed public employees. This first observation might seem to contrast with some findings of several research studies on the level of satisfaction with this specific job arrangement. The widespread consensus found in this study was rarely seen in previous transformation processes of work organisation, therefore representing a resource not to be disregarded as a basis for future improvements (among others: Messenger, 2019).

The respondents clearly perceived the unusual circumstances they experienced in the first phase of the COVID-19 pandemic to be potential advantages, envisaging the benefits in terms of less commuting, greater productivity, and increased self-management of workload. This outcome could not have been predicted, especially in light of the upheavals to daily life caused by the pandemic as well as of the lack of an established culture of telework in Italy.

Beyond a general good reception of compulsory teleworking in the early stages of containment and social distancing measures, some essential aspects and limitations of the study should be underlined. First and foremost, our research refers to a small sample of interviewees with respect to conditions experienced in the period between March and May 2020, a time in which the workers were basically forced to telework. We may assume that the novelty of working from home exerted an initially strong positive effect because of the possibility of staying "safe", not commuting to workplaces, and, at the same time, remaining connected to the world.

Additionally, respondents' positive perceptions may have benefitted from other effects, such as the awareness of being observed and studied as a sub-group of interest. This could also be a manifestation of the well-known Hawthorne effect, which refers to the set of modifications in a phenomenon or a behaviour which occur as a result of the presence of observers, but which is not likely to last over time (for a statistical interpretation of the Hawthorne effect, see Franke and Kaul, 1978).

Negative impacts, as underlined in the literature (Fana et al., 2020; EU-OSHA, 2021a; Eurofound, 2021) seem to take place only at a subsequent stage, when the problems and inconveniences of teleworking have begun to become more palpable in everyday life. In fact, the detrimental effects of such a new working arrangement likely would arise only when employees encountered an increase in their workload, more pressure regarding their work performance, and, in the long run, a failed opportunity to settle work-life balance. The situation was entirely new in the Italian context, since the country's teleworking implementation figures before the pandemic were extremely low in comparison to other similar European economies. Then, the continuation of the pandemic and protracted periods of involuntary work from home began to negatively affect the perceived quality of life

of teleworkers, also in terms of technostress[6]. These occurrences were higher for teleworkers (28% compared to 22% of other employees) and even stronger for women (29% compared to 22% of male colleagues). Moreover, teleworkers have been considerably affected by the reduction or absorption of social contacts and by the blurring of boundaries between working and non-working time (INPS, 2021).

In general, although with some intrinsic limits due to the small sample size, our results reveal that the probability of being satisfied with teleworking increases as satisfaction with home as a place of work rises, even in the case of work effort intensification. These circumstances lead us to surmise that workers in our sample are deeply engaged and can express their own ability to control their work processes.

It seems remarkable that many respondents in middle- and low-income classes have undertaken expenses to enable them to work from home. This observation is apparently at odds only with the fact that our sample consists of workers employed in the public sector: Italian public employees have seen many of their advantages eroded over the past 30 years. To date, figures demonstrate that, in comparison with the rest of Europe, the Italian public sector workers are in the lowest positions[7] with respect to almost all the available indicators.

With respect to a better balance of living and working time, our results – consistent with evidence at a national and international level (among others, Del Boca et al.; 2020; Dunatchik et al., 2021; EU-OSHA, 2021b) – have demonstrated that teleworking may be considered an effective contributor to a better work-life balance only if it is properly managed. In fact, research on the impacts of telework has begun to disclose that telework activities, when not properly managed, could lead to a non-sustainable overlapping of care and professional roles, with more severe consequences for women, especially with regard to their career prospects (Rodríguez-Modroño and López-Igual, 2021).

These results call for further research on larger datasets and deep investigation of the impacts of telework on the quality of working life, aiming to provide valuable information to both policy makers addressing specific regulatory measures and to managers in charge of implementing companies' welfare.

---

6 Some possible negative effects of the so called "technostress" are deterioration of the work-life balance and overworking. Overall, overworking (i.e. dedicating a large amount of time to work and neglecting moments of rest) involved 13% of workers and to a greater extent teleworkers than other workers (17% compared to 9%), women than men (19% compared to 11%) and managers than collaborators (19% compared to 9%) (Osservatori.net, 2021).

7 As an instance, the percentage of public workers out of the total number of workers in Italy in 2017 (13.4%) is lower than it is in France (19.6%), Spain (15.9%), and in the United Kingdom (16%), and higher only as compared to Germany (10.8%) (Eurofound and ILO, 2017).

REFERENCES

Agresti, A. (2010). *Analysis of ordinal categorical data*. Hoboken, N.J.,Wiley.

Baruch, Y. (2001). The Status of Research on Teleworking and an Agenda for Future Research. *International Journal of Management Reviews* 3(2): 113-129.

Chung, H., Van der Horst, M. (2018). Women's employment patterns after childbirth and the perceived access to and use of flexitime and teleworking. *Human relation* 71(1): 47-72.

Del Boca, D., Oggero, N., Profeta, P., Rossi, M. (2020). Women's and men's work, housework and childcare, before and during COVID-19. *Review of Economics of the Household* 18: 1001-1017.

Dunatchik, A., Gerson, K., Glass, J., Jacobs, J.A., Stritzel, H. (2021). Gender, Parenting, and The Rise of Remote Work During the Pandemic: Implications for Domestic Inequality in the United States. *Gender & Society* March 2021, doi:10.1177/08912432211001301

EU-OSHA (2021a). *Teleworking during the COVID-19 pandemic: risks and prevention strategies*. Publications Office of the European Union, Luxembourg.

EU-OSHA (2021b). *Home-based teleworking and preventive occupational safety and health measures in European workplaces: evidence from ESENER-3.*

Publications Office of the European Union, Luxembourg Eurofound (2010). *Telework in the European Union.* Publications Office of the European Union, Luxembourg.

Eurofound (2012). *Organisation of working time: Implications for productivity and working conditions*. Publications Office of the European Union, Luxembourg.

Eurofound (2021). *European Working Conditions Survey 2021*. https://www.eurofound.europa.eu/surveys/2021/european-working-conditions-survey-2021. Last access: 15/02/2022

Eurofound and ILO. (2017). *Working anytime, anywhere: The effects on the world of work*. Luxembourg and Geneva, Switzerland.

Fana, M., Tolan, S., Torrejón, S., Urzi Brancati, C., Fernández-Macías, E. (2020). *The COVID confinement measures and EU labour markets*. EUR 30190 EN, Publications Office of the European Union, Luxembourg.

Franke, R.H., Kaul, J.D. (1978). The Hawthorne Experiments: First Statistical Interpretation. *American Sociological Review* 43(5): 623-643.

Harker Martin, B., MacDonnell, R. (2012). Is telework effective for organizations? A meta-analysis of empirical research on perceptions of telework and organizational outcomes. *Management Research Review* 35(7): 602-616.

Hilbrecht, M., Shaw, S.M., Johnson, L.C., and Andrey, J. (2013). Remixing work, family and leisure: teleworkers' experiences of everyday life. *New Technology, Work and Employment* 28(2): 130-144.

ILO (2021). *Working from home: From invisibility to decent work.* ILO, Geneva, Switzerland.

ILO (2022). *World Employment and Social Outlook, Trends 2022.* Geneva, Switzerland.

INPS (2021). *Rapporto di Ricerca 2/2021. Indagine sullo Smart Working.* Direzione Centrale Studi e Ricerche. Available at: https://www.inps.it/docallegatiNP/Mig/Dati_analisi_bilanci/Attivita_ricerca/Studi_e_analisi/Rapporto_Smart_ Working_Inps2020_n2-2021.pdf. Last access: 10/03/2022.

Kaduk, A., Genadek, K., Kelly, E.L., Moen, P. (2019). Involuntary vs. voluntary flexible work: Insights for scholars and stakeholders. *Community, Work & Family* 22(4): 412-442.

Mann, S., Holdsworth, L. (2003). The Psychological Impact of Teleworking: Stress, Emotions and

Health. *New Technology, Work and Employment* 18(3): 196-211.

Messenger, J. (ed.) (2019). *Telework in the 21st century: an evolutionary perspective*. ILO-Edward Elgar (elgaronline.com)

Ministry of Economy and Finance (2020). *Conto Annuale, Andamento dell'Occupazione*, Ministero dell'Economia e delle Finanze, Ragioneria Generale dello Stato, Roma: https://www.contoannuale.mef.gov.it/ext/Documents/ANDAMENTO%20 DELL'OCCUPAZIONE.pdf Last access: 15/02/2022

Ogbonna, E., Harris, L. (2006). Organisational Culture in the Age of the Internet: An Exploratory Study. *New Technology, Work and Employment* 21(2): 162-175.

Osservatori.net (2021). Rivoluzione Smart Working, Un futuro da costruire adesso. https://www.osservatori.net/it/ricerche/infografiche/rivoluzione-smart-working-futuro-da-costruire-adesso-infografica. Last access: 15/02/2022.

Osservatorio Nazionale del Lavoro Agile (2020). *Monitoraggio sull'attuazione del lavoro agile nelle pubbliche amministrazioni nel periodo gennaio-aprile 2020*. https://www.funzionepubblica.gov.it/sites/funzionepubblica.gov.it/files/documenti/SW_COVID/Monitoraggio/Report_Monit_LavoroAgile_%20gen_apr2020.pdf. Last access: 28/02/2022

Parry, J., Young, Z., Bevan, S., Veliziotis, M., Baruch, Y., Beigi, M., Bajorek, Z., Salter, E., Tochia, C. (2021). *Working from Home under COVID-19 lockdown: Transitions and tensions, Work after Lockdown*. https://www.employment-studies.co.uk/system/files/resources/files/Working%20from%20Home%20under%20Covid-19%20Lockdown%20-Transitions%20and%20Tensions.pdf. Last access: 28/02/2022

Piccolo, D., Simone, R. (2019). The class of CUB models: statistical foundations, inferential issues and empirical evidence. *Statistical Methods and Application* 28, 389-435.

Putnam, L., Myers, K.K., Gailliard, B.M. (2014). Examining the Tensions in Workplace Flexibility and Exploring Options for New Directions. *Human Relations* 67(4): 413-440.

Pyöriä, P. (2011). Managing Telework: Risks, Fears and Rules. *Management Research Review* 34(4): 386-399.

Rodríguez-Modroño, P., López-Igual, P. (2021). Job Quality and Work-Life Balance of Teleworkers. *International Journal of Environmental Research and Public Health* 18(6): 3239. https://doi.org/10.3390/ijerph18063239

Sostero, M., Milasi, S., Hurley, J., Fernández-Macías, E., Bisello, M. (2020). *Teleworkability and the COVID-19 crisis: a new digital divide?* JRC121193, European Commission, Seville.

Steidelmüller, C., Meyer, S.C., Müller, G. (2020). Home-Based Telework and Presenteeism Across Europe. *Journal of Occupational and Environmental Medicine* 62(12): 998-1005.

Tutz, G. (2012). *Regression for categorical data*. Cambridge, Cambridge University Press.

Welz, C., Wolf, F. (2010). *Telework in the European Union*. European Foundation for the Improvement of Living and Working Conditions, Dublin.

Williams, R. (2009). Using heterogenous choice models to compare logit and probit coefficients across groups. *Sociological Methods & Research* 37: 531-559.

Williams, R. (2011). Fitting heterogeneous choice models with oglm. *The Stata Journal* 10(4): 540-567.

Yatchew, A., Griliches, Z. (1985). Specification error in probit models. *Review of Economics and Statistics* 67: 134-139.