doi.org/10.26398/IJAS.0034-003

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

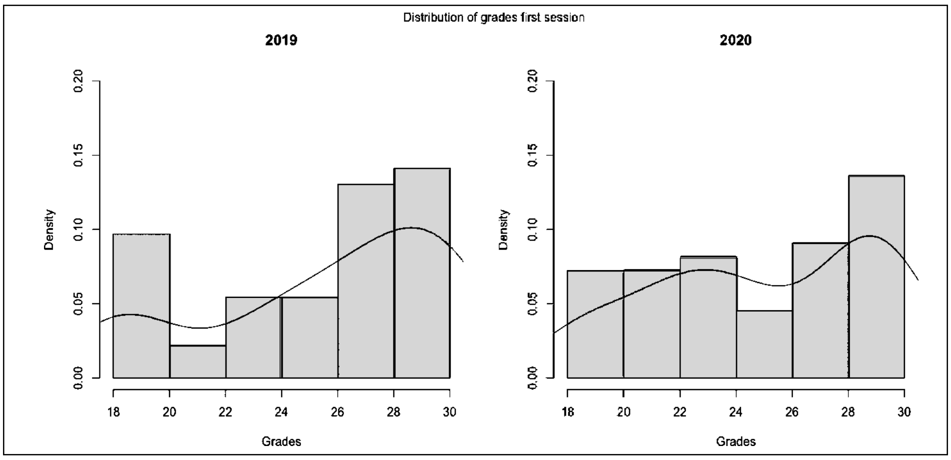

Fig. 1: Histogram and estimated densities of the grades of 2019 and 2020

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 X_i 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 = $(\text{GPA})^2$) 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:

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

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)).

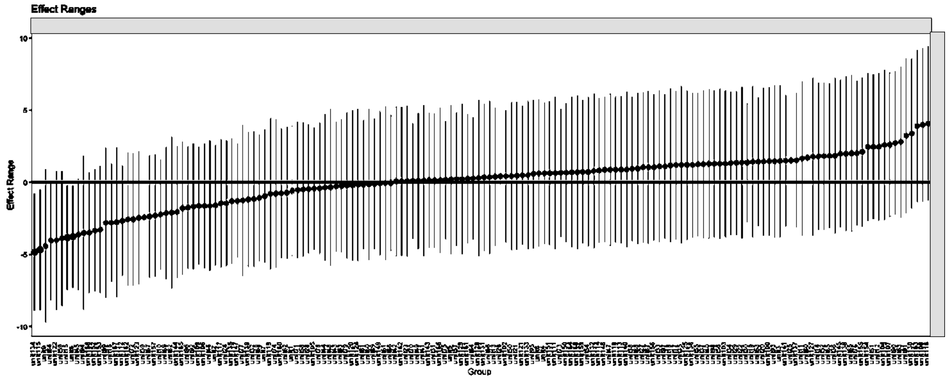

Fig. 2: Simulated random effects from posterior distributions, model (4) of Table 5

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

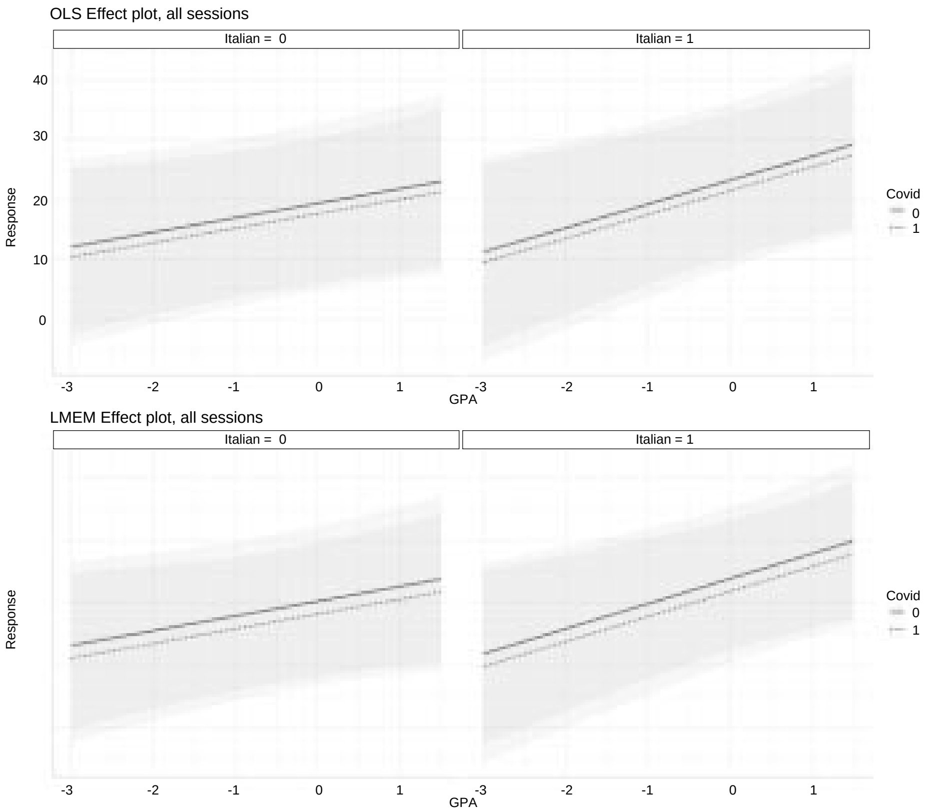

Fig. 3: Estimated marginal means of Grade computed from linear and linear mixed effects model, over GPA, Covid and nationality from models (2) and (4) of Table 5

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, $h = 1,\dots, H$ and, for each one of them, estimate a probit model for the binary dependent variable Grade $\leq y_h$:

$$P(Grade_i < y_h \left| X_i \right.) = \Phi(X_i^T \beta_h),\ h = 1,\dots,H,$$

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

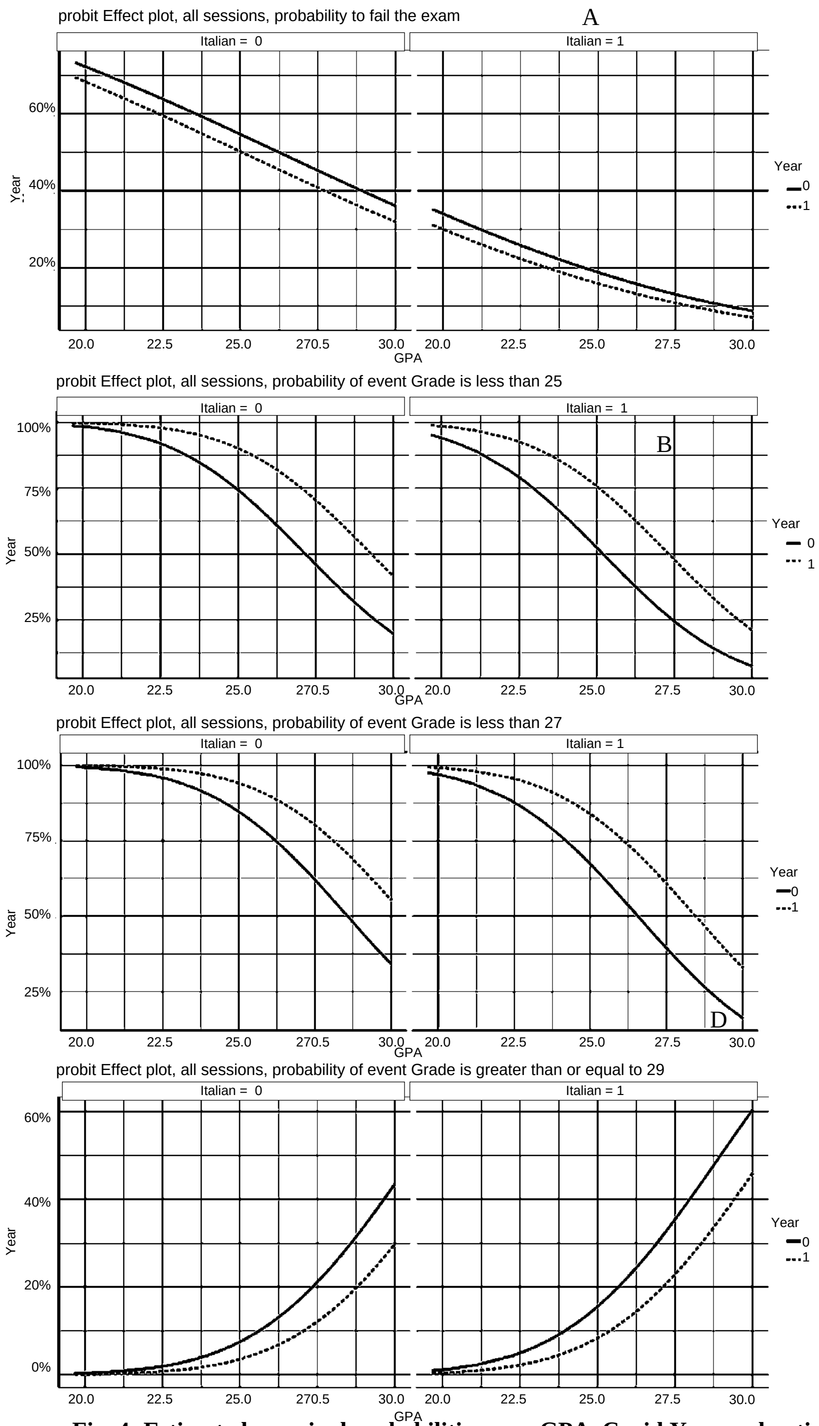

Fig. 4: Estimated marginal probabilities over GPA, Covid Year and nationality

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 ≈ 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.