CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

The Impact of COVID-19 on Students' Academic Performance: The Case of the Faculty of Social Sciences, Charles University

Bachelor's Thesis

Author of the Thesis: Adam Bruzl

Study Programme: Ekonomické teorie

Supervisor: Mgr. Barbara Pertold-Gebicka M.A., Ph.D.

Year of the Defense: 2022

Declaration

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

In Prague on the 3rd of May 2022 Adam Bruzl

Abstract

In order to determine the effect of the COVID-19 crisis on academic performance of students of the Faculty of Social Sciences at Charles University, unbalanced panel of over 5000 students ranging from Winter 2016 to Summer 2020 is analyzed using panel data estimation methods and binary dependent variable models. The estimated effect of COVID-19 on GPA is negative, i. e. improved GPA, for both bachelor's and master's students. On average, bachelor's students seem to be affected more severely than master's students. The magnitude of the effect varies substantially across institutes of the faculty, especially when considering master's students. The probability of 3rd year bachelor's or 2nd year master's graduating in the academic year 2020/21 is significantly lower than when compared to the same-grade students of 2018/19. When considering students of 2019/20, no significant difference in probability of graduating was found.

JEL Classification

A22, C23, C25, I20

Keywords

COVID-19, academic performance, GPA, higher education, distance learning, online learning

Title

The Impact of COVID-19 on Students' Academic Performance: The Case of the Faculty of Social Sciences, Charles University

Abstrakt

Za účelem zjištění vlivu krize COVID-19 na studijní výsledky studentů Fakulty sociálních věd Univerzity Karlovy jsou analyzována nevyvážená panelová data skládající se z více než 5000 studentů v rozmezí od zimy 2016 do léta 2020 za použití modelů individuálních vlivů a modelů binární vysvětlované proměnné. Odhadovaný vliv krize COVID-19 na vážený průměr je negativní, což implikuje zlepšení váženého průměru, jak pro studenty bakalářského, tak i magisterského studia. V průměrném případě jsou studenti bakalářského studia postiženi vážněji než studenti magisterského studia. Velikost efektu se mezi instituty fakulty podstatně liší, zejména pokud jde o studenty magisterského studia. Pravděpodobnost absolvování 3. ročníku bakalářského nebo 2. ročníku magisterského studia v akademickém roce 2020/21 je výrazně nižší než ve srovnání se studenty stejného ročníku 2018/19. Při posuzování studentů ročníku 2019/20 nebyl zjištěn žádný významný rozdíl v pravděpodobnosti absolvování.

Klasifikace JEL

A22, C23, C25, I20

Klíčová slova

COVID-19, akademické výsledky, průměr známek, vysoké školy, distanční výuka, online výuka

Název práce

Dopad pandemie COVID-19 na akademické výsledky studentů: případ Fakulty sociálních věd Univerzity Karlovy

Acknowledgement

I would like to express my gratitude to the supervisor of my thesis, Mgr. Barbara Pertold-Gebicka M.A., Ph.D., for her patience and valuable suggestions during the process of writing this thesis.

Table of Contents

1 Introduction

As of 1st March 2020, the Czech Republic had recorded its first confirmed cases of COVID-19 infections, followed by a period of a rapid increase in the number of infected (Česká televize, 2020). By 11th March 2020, nation-wide school closures had been implemented as per governmental restrictions (Czech Republic Ministry of Health, 2020), resulting in abandoning face-to-face classes and switching to online education completely.

In light of these events, this thesis' purpose is to answer, whether the COVID-19 crisis and its consecutive switch from face-to-face classes to online education had any effect on the academic performance of students of the Faculty of Social Sciences at Charles University.

As we observe individual students across several time periods, i. e. panel data structure, the conceptual approach towards answering the research question is to estimate the treatment effect attributed to a dummy variable(s) attaining one for semesters affected by the pandemic, and thus identify the difference between the pre-COVID-19 and during-COVID-19 academic performance. Academic performance in the context of this thesis is defined by two distinct measures – GPA and a binary variable denoting whether a given student graduated or not.

The existing literature regarding the effect of COVID-19 on academic performance usually suggests that grades improved as proposed by Rodríguez-Planas (2022a) for example. When considering graduation, Aucejo et al. (2020) reports that 13% of those that delayed graduation did so on the account of the pandemic.

The thesis is structured in the following way: Chapter 2 outlines the current state of knowledge, including pre-COVID-19 comparison of traditional and online education. In Chapter 3, expected results are stated. Chapters 4 and 5 present data and methodology utilized in estimating the results, which are presented in chapter 6 and discussed in chapter 7. Chapter 8 offers concluding remarks.

2 Literature Review

As the COVID-19 crisis has affected the whole world in an unprecedented manner, a vast number of studies and research emerged to determine the crisis' consequences, one of the investigated consequences being the impact on academic performance of students in higher education. The aim of this chapter is to introduce and summarize relevant literature regarding the crisis' effect on academic performance of students in higher education, i. e. results of various analyses and methodologies implemented shall be discussed as they play an important role in explaining the results of my own analysis.

The first subchapter summarizes research made on comparison of face-to-face and online education with respect to GPA in the pre-COVID-19 era. The purpose of the following subchapters is to outline how and possibly why did the COVID-19 crisis affect academic performance, including a discussion on pandemic-induced delayed graduation.

2.1 Pre-COVID-19 Comparison of Traditional and Online Education

Even before the breakout of the COVID-19 crisis and massive use of online tutoring, numerous studies were investigating whether online teaching differs from traditional teaching in terms of students' exam performance. The usual findings imply that students of online programs perform worse than students in traditional programmes.

When comparing the pre-COVID-19 online and traditional education, one needs to take into account that choosing traditional or online education, unlike the pandemic-induced switch to online learning mandatory for everyone, is a deliberate act. Thus, there might be differences among students based on which mode of education they chose. For example, employed students might choose online courses as these students generally have less time for studying and might have worse grades even if attending traditional classes. Therefore, the estimated effect might not be the effect of online education but the effect of selection. The following cited studies employ various estimation methods to mitigate the selection bias, such as instrumental

variables, controlled experiments, fixed effects estimation or propensity score matching.

Bettinger et al. (2017), using data of more than 200 000 students attending undisclosed American for-profit university, analyzed the effect of taking online course on student outcome for a given student, course and semester. The authors tackled the selection bias by utilizing an instrumental variable approach, where the instrument is an interaction of two variables. Namely, the distance from a student's residence to her local campus and a dummy variable equal to one if a given course is offered in face-to-face setting at the student's local campus. The analysis' results imply that taking a course online has negative effect on grades not only in the currently taken course, but also on grades in future courses. Additionally, taking a course online is reported to decrease the probability of a student remaining enrolled in the following semesters.

Cacault et al. (2021) conducted a randomized experiment on students attending bachelor programme in economics and management at University of Geneva. The experiment consisted of live-streaming lectures and allowing random part of students to access the streamed lectures while also having the opportunity to attend in-person lecture. Then the information whether a student was allowed to access the streamed lectures was used as an instrumental variable. The authors report that attending online lectures decreases grades for low-performing students while increasing them for high-performing students.

A longitudinal analysis, carried out by Xu and Jaggars (2014) at an American university, employs the within transformation to eliminate the selection bias. The estimated effect of a course being taken online on grades is negative and the most affected groups are males, younger and low-performing students.

To eliminate the selection bias and the endogeneity of choosing online learning, Coates et al. (2004) utilize 2SLS. Uncovering that students in online sections performed worse on the Test of Understanding College Economics than students in face-to-face classes did. Additionally, the authors found that not accounting for the selection of students into the respective modes of education biases the estimated difference of the two modes of education towards zero.

Heissel (2016) analyzes the effect of choosing virtual Algebra I course instead of its in-person counterpart on American middle school students. To deal with the selection bias, authors implement propensity score matching where prior math achievement serves as covariate that accounts for the selection process. The author reports that students that chose Algebra I in online mode underperformed when compared to those students that attended Algebra I in-person.

In view of the presented research, online teaching usually results in students' academic performance to diminish. However, Cacault et al. (2021) report that online courses are beneficial to high-performing students. The authors propose such explanation that high-performing students are less likely to attend in-person class, particularly when the opportunity cost of going to class is high, as they are usually more autonomous than low-performing students and thus are able to manage learning without any supplemental explanation from a lecturer. Therefore, introducing online classes might induce high-performing students to attend as the opportunity costs of attending decrease. That is, high-performing students attend online lectures instead of self-studying, which results in improved learning and thus improved grades.

2.2 Effect of COVID-19 on GPA

Since the breakout of the COVID-19 crisis numerous studies evaluating its effects on students have been written. They investigate how the crisis-related school closures and move towards online teaching has affected students' well-being, satisfaction, or academic performance. In this review I focus on the latter stream of research.

It should be stressed that the COVID-19 crisis affected not only the mode of education, but also other factors that might affect students' performance. Firstly, Savage et al. (2020) suggest that the COVID-19 pandemic has increased stress and has had negative effect on the overall mental wellbeing of students. Increased stress levels are believed to negatively impact learning capacities and academic performance as argued by Pascoe et al. (2020). Secondly, the comparison of pre-COVID-19 traditional and online education suggests that students tend to perform worse when partaking in online education. Therefore, under the ceteris paribus assumption, the change of mode

of education is expected to have negative effect on students' performance. On the other hand, students had more time as most free-time activities were banned due to the ongoing lockdown which could positively affect students' performance assuming that more time was devoted to studying-related pursuits.

The list of possible factors that might affect students' performance presented in the preceding paragraph is not exhaustive, nevertheless, sufficient for the purpose of demonstrating that the expected effect or even the direction of the expected effect of COVID-19 on students' performance is uncertain and unpredictable as there are various factors that could have affected students' performance both positively and negatively.

Rodríguez-Planas (2022a) conducted a longitudinal analysis at City University of New York to assess the effect of moving to online education during Spring 2020 on academic performance with respect to students' social conditions. By utilizing panel regression models and event study approach, along with controlling for socioeconomic conditions of students, the analysis uncovered that both higher- and lower-income students performed better in terms of GPA than they would, had not the COVID-19 crisis occurred. The treatment effect was estimated as the coefficient by the dummy variable equal to one if the observation belonged to the summer semester of the academic year 2019/20, that is the focus of this analysis is purely on the short-term effect of the COVID-19 crisis.

The positive effect of COVID-19 on students' performance is confirmed by research of Gonzalez et al. (2020), carried out at Universidad Autónoma de Madrid, which utilized the comparison of means of academic results in the academic year 2019/20 and two preceding academic years. When comparing the means of 2017/18 and 2018/19, no difference was found, whereas when comparing the year 2019/20 to each of the two preceding, a significant increase in the mean of 2019/20 was found. The study also suggests that this improvement in grades is related to a real improvement in students' learning strategy. However, this suggestion is based on the fact that the students of 2019/20 performed better in exams that were conducted in online mode even before the COVID-19 crisis occurred, i. e. the assessment process

is assumed constant for these exams. This claim seems implausible since the assessment process is not the only factor that needs to be taken into consideration.

Collecting data from five Turkish universities, utilizing ANCOVA and controlling for class size, class level, educator's academic degree, individual students' university entrance scores and field, Karadag (2021) finds that students of the summer semester 2019/20 performed better than students of the summer semester 2018/19. As the analysis compares grades from time periods set only one year apart, the contents of the given courses are assumed to be identical for both pre-pandemic and pandemic settings. Moreover, the author also argues that online education requires substantially more preparation and planning than face-to-face tutoring. Thus, implementing distance learning on such a short notice could not have resulted in adequate substitution of traditional education. Therefore, it is suggested that improvement in grades is a result of grade inflation, i. e. the sudden change of mode of education forced the lecturers to be more lenient to mitigate the negative conditions caused by the COVID-19 crisis without any real improvement in students' academic performance.

A study by Iglesias-Pradas et al. (2021) who test differences in academic performance across the academic years 2017/18, 2018/19 and 2019/20 at the School of Telecommunication Engineering (Spain), by utilizing the ANOVA technique, also reports improved academic performance. However, it should be noted that the authors explicitly state that they consider the result to be counterintuitive. The framework on which the analysis was built is rooted in the idea that teaching offered as a result of forced school closures is unlike any kind of online education approach known until this very moment. The authors, using findings of Hodges et al. (2020), define the term emergency remote teaching in the following way:

"The main difference between online learning and emergency remote teaching lies in that online learning results from careful instructional design and planning, requiring an investment in a whole ecosystem of learner supports that takes time to build, whereas emergency remote teaching emerges as a response to a crisis and entails a temporary shift of instructional delivery to an alternate delivery mode that involves the use of fully remote solutions for instruction that would otherwise be delivered using face-to-face, blended or hybrid courses" (Iglesias-Pradas et al., 2021, p. 2).

Under the assumption that emergency remote teaching is truly inferior to the pre-COVID-19 online teaching, and considering the comparison of pre-COVID-19 traditional and online education and its suggestion that students attending online courses usually perform worse than those attending face-to-face classes, it may be reasonable to believe that the positive effect of COVID-19 on academic performance is indeed counterintuitive as we would expect it to be negative. Hence, if any positive effect is observed, then it is probably grade inflation.

Rodríguez-Planas (2022a) explicitly states that academic performance was expected to drop and proposes a number of reasons for which it increased. Firstly, the author suggests that the improvement in grades may be related to a real improvement in students' learning strategy and since the ongoing lockdown made it almost impossible to participate in other activities, the opportunity cost of studying decreased. Secondly, due to the uncertain conditions brought about by the crisis, the course instructors were forced to conduct exams in an unusual way, i. e. remote examination, which could have potentially resulted in easier exams or more lenient grading. Thirdly, online exams are by their nature much more difficult to manage, especially in terms of cheating prevention. Some students may have taken advantage of this situation to improve their academic performance.

A more in-depth analysis of student cheating practices during COVID-19 crisis by Balderas et al. (2020) also supports the claim that some students may have cheated during exams. For instance, the authors' findings suggest that when conducting an asynchronous exam, i. e. students do not have to take the exam at the same time, students organize themselves to collaborate. Noting that the students that started the exam at the first possible moment required at least twice as much time to finish than certain students that delayed the start.

Contradictory findings to those mentioned above are reported by Aucejo et al. (2020). The authors of this study conducted a survey administered during the last week of instruction of the summer semester of 2019/20 among Arizona State University students asking them about their actual performance and about their expected performance if there was no pandemic. By taking the difference between an individual's actual GPA and expected GPA, had there been no pandemic, the average

treatment effect of COVID-19 is -0.17, which could be interpreted as "GPA decreased with COVID-19 by 0.17 on average". Bear in mind that this specific study was conducted at an American university and that decrease in GPA implies worsened academic performance since the American grading scale ranges from zero to four where four is the best possible grade. Had it been a Czech institution, the scale would have been reversed.

Additionally, the authors also delve into the heterogeneity in the effect of COVID-19. They report that the magnitude of the effect was far greater for men, first-generation students, or lower-income students. When considering the results of this analysis, critical judgment is advised given that this study's results are highly dependent on the students' expectations and may not reflect the true state of the world had there been no COVID-19 crisis. For example, the magnitude of the estimated effect for men could have been greater due to the fact that men tend to overestimate their performance (Bench et al., 2015).

It is very likely that the effect is heterogeneous not only with respect to gender as suggested by Aucejo et al. (2020) or socioeconomic conditions (Rodríguez-Planas, 2022a) but also with respect to other numerous factors such as the given institution's characteristics, its equipment, the way the institution communicates with students or the overall preparedness of lecturers. Meaning that the various results of the above-mentioned distinct analyses may not be easily comparable since each of those analyses is limited to its own research setting and for the results to be comparable, we would need to control for the given institution.

In view of the presented literature, students' academic performance is usually found to have improved with the COVID-19. Nevertheless, these results are found to be in contradiction of the researchers' expectations as indicated by Karadag (2021), Iglesias-Pradas et al. (2021) and Rodríguez-Planas (2022a). The latter author explores possible explanations of the effect as per discussion above, however, she only presents suggestive evidence without any statistical testing.

The results and approach proposed by Rodríguez-Planas (2022a) are considered to be of most relevance for this thesis not only due to similar methods,

i. e. panel data estimation methods, being utilized, but also due to the fact that the author disposes of much larger and diverse sample than the remaining authors do.

As some of the studies presented analyze academic performance of students attending American institutions, conducting a similar study on a university in continental Europe might produce not only interesting but also different results since there are major differences between European and American higher education systems (Gapinski, 2010).

Additionally, the presented literature lacks analysis by field or level of study – the studies either focus on one specific degree or university as a whole. Therefore, heterogeneity analysis of the effect of COVID-19 on GPA with respect to institutes of the Faculty of Social Sciences, i. e. field-related subgroups of the faculty, shall be conducted for both bachelor's and master's students, respectively, to determine whether the effect of COVID-19 differs across these given subgroups.

2.3 Effect of COVID-19 on Graduation

As GPA is not the only performance measure this thesis takes interest in, literature regarding the effect of COVID-19 pandemic on students delaying their graduation is presented.

It should be noted that pre-COVID-19 research by Bettinger et al. (2017), provides evidence that online teaching affects not only grades, but also the probability of finishing studies. It is proposed that taking a course online decreases the probability of a student remaining enrolled in the following semesters. Therefore, we might expect the effect of COVID-19 on the probability of graduating to be negative.

In the online survey administered at the end of the summer semester 2019/20 at the Arizona State University, Aucejo et al. (2020) also inquire into the possibility that the COVID-19 crisis may have affected whether students delayed graduation or not. The approach consists of comparing the actual state of the world and the students' expectations had there been no pandemic and investigating what share of students expect delaying graduation. The findings propose that 13% of students planned

delaying graduation in the summer semester of 2019/20 on the account of the COVID-19 crisis.

Rodríguez-Planas (2022b) implements a similar approach, that is measuring the differences between the pandemic-affected world and students' expectations obtained from a survey conducted during the Summer of 2019/20 at City University of New York. For the results to be valid, the author assumes that students' beliefs about the alternative state of the world are accurate. The findings imply that the share of students that planned to delay their graduation as a consequence of the COVID-19 crisis is approximately 13%.

As proposed by Saw et al. (2020), whose findings are based on an online survey of STEM students across numerous institutions in the US, as of June 2020, 18% of master's and 7.6% of bachelor's students reported delaying their graduation due to the COVID-19 pandemic. The survey also offers possible explanations of the delayed graduation. Among those that delayed graduation, almost 62% of students, reported limited access to academic facilities and resources, slightly more than 45% describe declined mental or physical health and nearly 41% report set-back in degree-related tasks.

Given that the presented literature regarding the effect of COVID-19 crisis on whether students delayed graduation or not is based solely on data obtained from online surveys, there are some issues that need to be addressed. Foremost, Rodríguez-Planas (2022b) argues that it is possible that certain part of the student population might have been more likely to respond due to feeling more affected by the pandemic. If this were true, then the samples obtained from the surveys are not representative which could potentially result in biased results. Additionally, the question of whether students' expectations are a trustworthy approximation of the state of the world under the assumption that there was no COVID-19 crisis is also relevant given that individuals, especially young individuals, tend to give more weight to recent experiences (Malmendier & Nagel, 2016; Kuchler & Zafar, 2019).

Due to the reasons discussed above, the results of the presented literature are open to disputes. This thesis' approach of analyzing such phenomenon consists of

comparing the third-year bachelor's students or second-year master's students and whether they graduated or not across numerous academic years. The author of this thesis believes that approach of this sort may deliver more trustworthy results as the data at hand are based on reality and not on students' self-perceived subjective truths.

Moreover, there is one more important advantage of our research over the presented ones: they usually collected data in 2020, while we also have data for the academic year of 2020/2021, three semesters into the pandemic.

3 Expected Results

The purpose of this thesis' analysis is to determine the effect of COVID-19 on the academic performance of students of the Faculty of Social Sciences, the effect on GPA and a binary variable denoting whether a given student graduated or not, to be more specific. Effect on both variables shall be studied in greater detail as we explore possible heterogeneity across level of studies, genders, or institutes of the faculty (field-related subgroups of students). Additionally, as the data available cover longer period of the pandemic-induced online education, i. e. 3 semesters, we will focus not only on the aggregate effect on the given student outcome, but also on the effect across each of the semesters.

Given that majority of the presented studies, especially Rodríguez-Planas (2022a) who uses similar approach to the approach of our own, report improved academic performance under the pandemic, it is expected that our analysis will yield similar results. That is, we expect GPA to have improved.

However, each of the cited studies collected data only on the first pandemic semester. As argued by Karadag (2021) and Iglesias-Pradas et al. (2021), implementing distance learning on such a short notice could not have resulted in adequate substitution of traditional equation. Therefore, not only teaching but also the way of conducting exams during the academic year 2020/21 could have been substantially different than that of the summer semester of 2019/20 due to there being more time to prepare. Thus, the effect for the remaining semesters does not necessarily have to be what we would expect.

Regarding the effect on the probability of graduating, the literature offers a very limited overview. Nevertheless, as per findings of Bettinger et al. (2017) who report that taking an online course decreases the probability of a student staying enrolled in the upcoming semesters, the probability of graduating is expected to have decreased. Probably not immediately, i. e. not in the summer semester of 2019/20 but in the following academic year.

4 Data

This chapter introduces the data that were used for the analysis of the effect of the COVID-19 crisis on students' academic performance at the Faculty of Social Sciences of Charles University. To begin with, the data source is stated. Followed by preliminary data manipulation description and brief characterization of the data set and its variables. Following subchapters offer detailed view of the dependent variables.

The data were obtained from the Charles University's information system via the Study Department, which also included anonymization of the students' records to avoid any privacy-related issues. However, the data cannot be provided as we consider them to be sensitive information of the Faculty of Social Sciences.

The data set is two-dimensional, each record corresponds to a given student at a certain point in time. The time dimension ranges from the academic year 2016/17 to the academic year 2020/21. The frequency at which we observe the sample is half-yearly, that is each time unit represents a given semester which means that there are 10 time periods in total.

We observe such students that attend full-time study programmes and started their studies in the academic years 2016/17, 2017/18, 2018/19 or 2019/20. Therefore, the data set contains both students that have already finished (either successfully or not) their studies and those that have yet to do so. Since the standard length of studies for both bachelor and master students is less than the number of time periods observed, we do not observe every individual student in each time period, i. e. the data set is an unbalanced panel.

The cross-sectional index is defined by two variables – a unique student specific identification number and an identification number corresponding to a given study programme attended, i. e. a student attending two different programmes is regarded as two non-identical cross-sectional units.

Table 1: Number of students with respect to the number of study programmes attended

	Bachelor's	Master's
Number of students attending 1 study programme	2762	2335
Number of students attending 2 study programmes	106	91
Number of students attending 3 study programmes	3	3
Total number of students	2871	2429

For further analysis, each of the students attending more than one study programme has been reduced to a single cross-sectional unit by eliminating the additional study programmes. The elimination process is based on the assumption that the programme for which there is the highest number of records is of most importance for the given student and thus reflects her or his abilities and effort the best.

The available variables are defined with respect to each semester, for example, GPA is not cumulative GPA over the whole duration of a student's studies, it is solely the result of the exams in the given semester. The list of the given or calculated variables is as follows:

- Current academic year
- Current semester
- Unique student identification number
- Year of admission
- Birth year
- Sex
- Field
- Institute
- GPA
- Status
- Age = Current academic year – Birth year
- Year of study = Current academic year – Year of admission + 1

It should be noted that the Institute variable has been created by looking up the study programme code at the Faculty of Social Sciences website (2022) and mapping the programme code to its respective institute. The variables GPA and Status will be discussed in greater detail in the following sections as they play a pivotal role in the analysis.

Table 2 summarizes the number of students with respect to available grouping criterions which are an important aspect of the further analysis as each of the regression models will be run for the faculty as a whole and then with respect to each of the criterions. More detailed view can be found in the appendix in Table A.1 and Table A.2 which offer number of students grouped by field.

Table 2: Number of students with respect to various grouping criterions

		Number of students	
		Bachelor's	Master's
Total		2871	2429
Sex	Female	1443	1384
	Male	1428	1045
Institute	IES	605	346
	ICSJ	637	570
	IIS	568	722
	IPS	604	560
	ISS	457	231
Year of starting the study	2016	768	596
	2017	664	570
	2018	671	578
	2019	768	685

4.1 GPA

Due to the fact that GPA is one of the dependent variables in the following analysis, a more detailed description, summary statistics and figures depicting average GPA with respect to several grouping criterions are provided.

Let X be a student's final assessment for a given class expressed as a percentage, i. e. the number of points received divided by the total number of points that could have been received. The Faculty of Social Sciences' A-F classification grading system can be defined in the following way:

- A: $90\% < X$
- B: $80\% < X \leq 90\%$
- C: $70\% < X \leq 90\%$
- D: $60\% < X \leq 90\%$
- E: $50\% < X \leq 90\%$
- F: $X \leq 50\%$.

For the purpose of calculating semestral GPA, the Study Department, which has given us the access to the data, converted the A-F classification system to numerical grading scale 1-4, where 1 represents A, 1.5 B, 2 C, 2.5 D, 3 E and 4 stands for F. Finally, to determine the semestral GPA for each of the student the average of a given student's exam results weighted by the number of ECTS credits was computed.

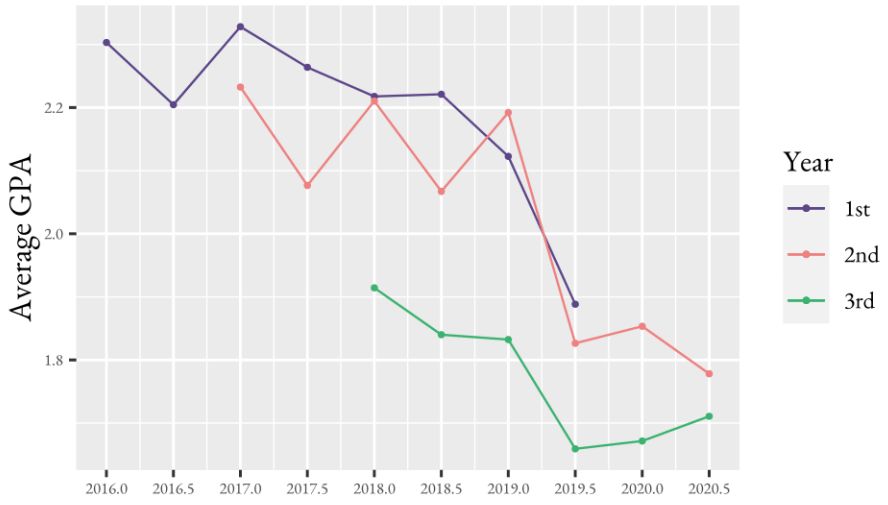
Table 3: Five-number summary and mean of the GPA variable

	Min.	1st Qu.	Median	3rd Qu.	Max.	Mean
Bachelor's	1	1.5	2	2.5	4	2.038
Master's	1	1.25	1.5	2	4	1.681

Table 3 presents simple summary statistics for the given GPA variable. It is apparent that the variable attains values in the closed interval [1, 4] and NAs in some cases. These cases occur either when a student attends such classes that lead to only a pass or fail result or does not register for any of the graded exam dates.

Given that the data for bachelor's students are generated in such way that the observations in the academic year 2016/17 consist of first-year students only, observations in the academic year 2017/2018 contain first- and second-year students and observations in the academic year 2020/21 entail second- and third-year students, presenting aggregated average GPA across time would have created somewhat biased view (such view can be found in the appendix, see Table A.3 and Table A.4). Therefore, the following figure depicts average GPA across time with respect to the year of study.

Figure 1: Average GPA by semester and year of study (bachelor's)

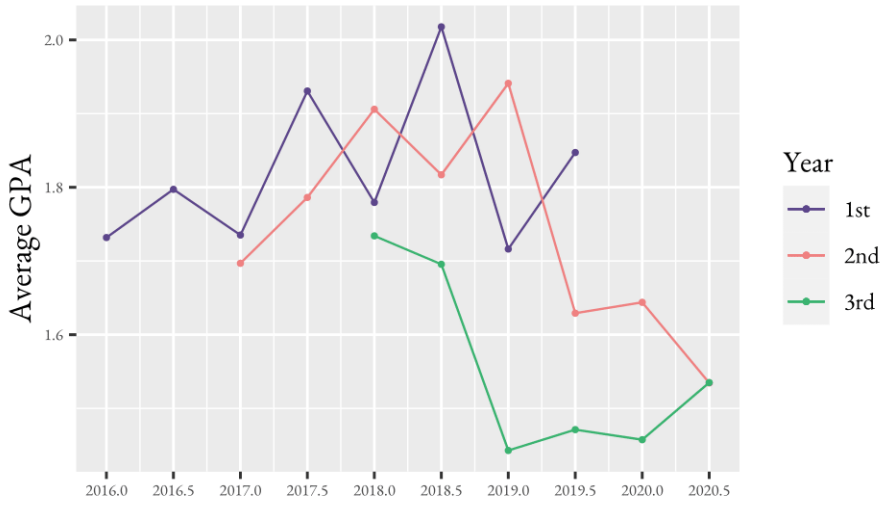


As illustrated in Figure 1, it is quite apparent that there is a non-negligible drop in average GPA of first-, second- and third-year bachelor's students as the COVID-19 crisis began, that is the summer semester of the academic year 2019/20.

Figure 2 and Figure 3 show average GPA by semester and year of study for IES bachelor's and ICSJ bachelor's, respectively. These two institutes have been arbitrarily chosen to demonstrate that the magnitude or even the direction of the change seems to be heterogeneous across institutes when comparing the summer semester of the academic year 2019/20 to the preceding periods. The remaining institutes' average GPA plots can be found in the appendix in Figure A.1, Figure A.2 and Figure A.3.

Figure 2: Average GPA by semester and year of study (IES bachelor's)

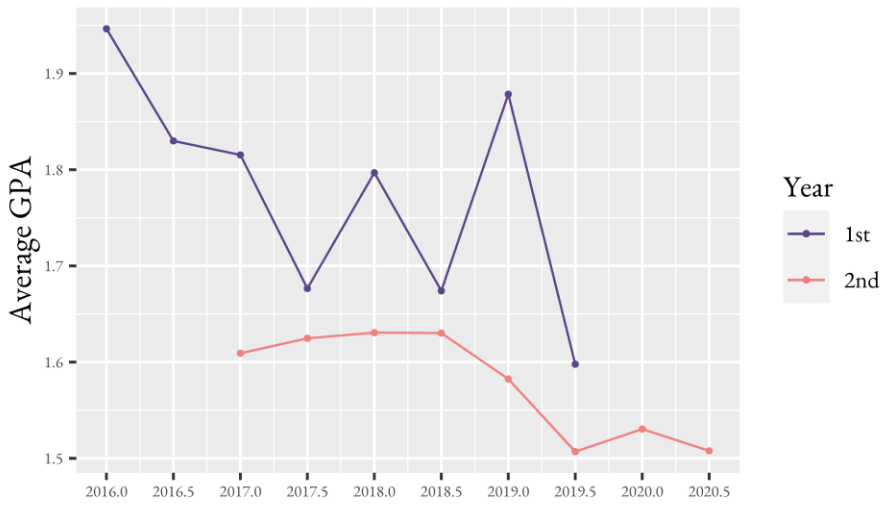
Figure 3: Average GPA by semester and year of study (ICSJ bachelor's)



As per Figure 2, average GPA of IES bachelor's students has plummeted significantly in the summer semester of 2019/20 and attained its all-time minimum value for each of the three groups. Whereas ICSJ bachelor's students, as shown in Figure 3, evinced less severe change when compared to the preceding periods.

For master's students, analogous problem arises due to the observations in the academic year 2016/17 containing only first-year students, as well as the observations in the academic year 2020/21 containing second-year students only. Thus, Figure 4 presents master's students' average GPA across time with respect to the year of study.

Figure 4: Average GPA by semester and year of study (master's)



As demonstrated in Figure 4, both first- and second-year master's students recorded an unprecedented decrease in average GPA in the summer semester of 2019/20. Nevertheless, this decline is not as severe as the decline of bachelor's students' average GPA.

Figure 5 and Figure 6 depict average GPA of master's students for two arbitrarily chosen institutes for the purpose of discussing possible heterogeneity of the change across institutes. For analogous plots of the remaining institutes' average GPA see Figure A.4, Figure A.5 and Figure A.6 in the appendix.

Figure 5: Average GPA by semester and year of study (IES master's)

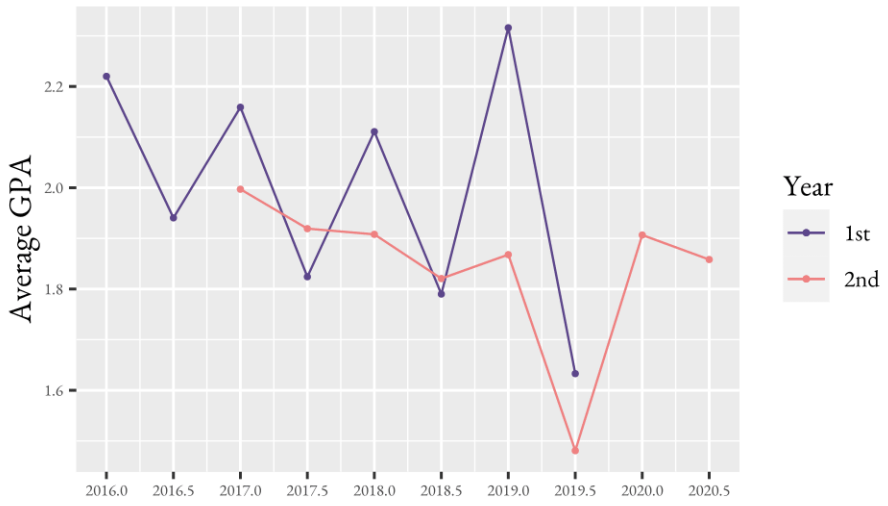
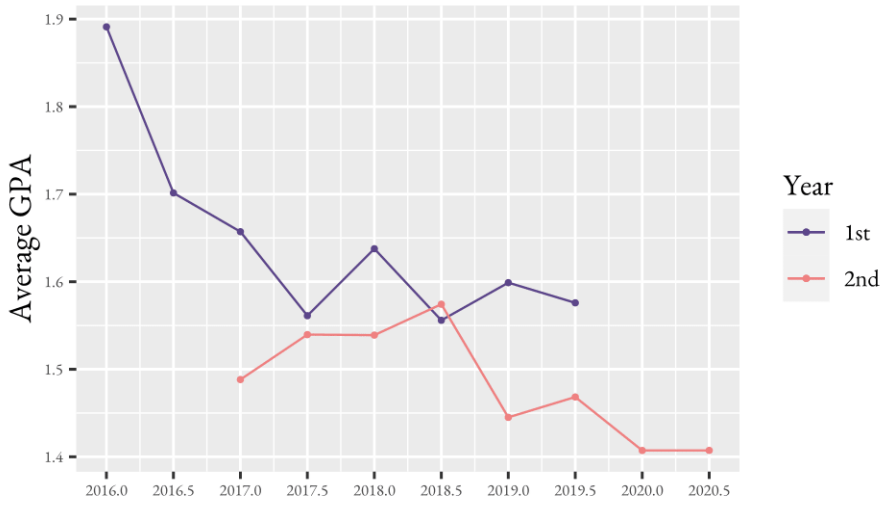


Figure 6: Average GPA by semester and year of study (IIS master's)



Based on Figure 5, both first- and second-year IES master's students' average GPA attained its minimum value in the summer semester of 2019/20 and then stabilized for the second-year students. First-year IIS master's students' average GPA even slightly increased when comparing the summer semester of 2019/20 and 2018/19 as per Figure 6.

4.2 Graduated – a Dummy Variable

This section is dedicated to describing the process behind encoding the Graduated variable, which is used in the follow-up analysis as a dependent variable. Firstly, the Status variable, which is present in the original data set, is described as it is the source of information for encoding the Graduated variable, then we proceed to outline the process of encoding the said variable.

The Status variable represents a given student's state with respect to her or his study programme as of November 2021, which is the moment the source data set was built. Thus, the variable is time-invariant within the scope of the panel data set. The Status variable is a nominal variable, as it attains four values which have no intrinsic ordering. The values are as follows:

- S: a student is studying
- A: a student has graduated
- P: a student has interrupted their studies
- Z: complement to the previous possible outcomes (i. e. quitting or finishing unsuccessfully).

Table 4: Summary statistic of the Status variable

	S	A	P	Z
Bachelor's	815	1005	56	995
Master's	427	1336	91	575

Table 4 reports the number of students with respect to the value of the Status variable for both bachelor's and master's students.

Table 5: Average number of semesters before graduating by institute

	IES	ICSJ	IIS	IPS	ISS
Bachelor's	6.28	6.04	6.09	6.30	6.19
Master's	4.43	4.39	4.06	3.91	4.48

Bachelor's students are expected to graduate in 3 years (6 semesters) and master's students in 2 years (4 semesters). Table 5 shows average number of semesters

before graduating with respect to institute. It is evident that an average student would finish as expected. Table A.5 and Table A.6 offer greater detail as they present average number of semesters before graduating by field.

For the purposes of modeling the effect of COVID-19 on the probability of a given student successfully finishing their studies in standard time, an additional data frame has been built to accommodate the needs of modeling a discrete dependent variable as the methods used will assume cross-sectional analysis framework.

Firstly, a set of such students (the cross-sectional units) that have ever been in their 6th semester, which is considered to be the standard finishing period for bachelor's students, has been created. Then, we say that i-th student, belonging to the above-defined set, graduated in their 3rd year, if the Status variable for the i-th student equals "A" and their 6th semester is their last semester. Such students that do not comply with the stated conditions have not graduated in their 3rd year.

The information in the previous paragraph has been encoded into a dummy variable, i. e. the Graduated variable, carrying the information whether i-th student graduated in their 3rd year or not. Hence, we are working with binary dependent variable. The cross-sectional data frame thus contains the just defined dummy variable, the information about which of the academic years was the 3rd and control variables for a given student.

The process of creating the data frame for master's students was analogous, the only difference is that we assume that the standard finishing period is the 4th semester, that is the 2nd year.

Table 6: Overview of 3rd year bachelor's students graduating across time

	Academic year		
	2018/19	2019/20	2020/21
Number of 3rd year students	427	414	431
Number of those that graduated	281	257	230
Ratio of the 2nd and the 1st row	0.66	0.62	0.53

Table 6 offers aggregated view on the key variables in the above-defined data frame for bachelor's students. There has been a non-negligible decrease in the ratio of 3rd years that graduated to all 3rd years when comparing the academic year 2020/21 to the previous ones.

Table 7: Overview of 2nd year master's students graduating across time

	Academic year			
	2017/18	2018/19	2019/20	2020/21
Number of 2nd year students	382	431	456	530
Number of those that graduated	206	240	273	220
Ratio of the 2nd and the 1st row	0.54	0.56	0.60	0.42

Based on Table 7, the ratio of 2nd year students that graduated to all 2nd year students slightly increased in 2019/20 and then decreased dramatically in 2020/21. However, the number of all 2nd year students appears to have increased exceptionally when compared to the previous academic years which might have been caused by numerous master's programmes being opened in the academic year 2019/20.

5 Methodology

This chapter presents statistical methods that are utilized in the analysis of the data set described in the preceding chapter. Specifically, panel data estimation methods and binary dependent variable models shall be discussed as they are of service in answering the research question, i. e. determining the effect of COVID-19 crisis on the academic performance of students at the Faculty of Social Sciences.

General approach undertaken to estimate the effect of COVID-19 and related switch towards online teaching on students' performance is captured by equation (1) and (2):

$$y_{it} = \beta_0 + \beta_1 COVID_t + Z_{it}\gamma + e_{it}, \tag{1}$$

where y_{it} is a performance measure of i-th student in semester t, specifically GPA or a dummy capturing the graduation status. $COVID_t$ is a dummy variable equal to 1 for semesters affected by the COVID-19 pandemic, Z_{it} is a vector of student characteristic (available control variables are sex, age, field of study), and e_{it} is the unobserved disturbance. Coefficient β_1 corresponds to the effect of the COVID-19 crisis on student's performance. Analogously, we define:

$$y_{it} = \alpha_0 + \alpha_1 sem_t^{2019/20S} + \alpha_2 sem_t^{2019/20W} + \alpha_3 sem_t^{2020/21S} + Z_{it}\delta + d_{it}, \tag{2}$$

where sem_t^j is a dummy variable equal to 1 if $j = t$ and d_{it} is the disturbance term. Thus, the coefficients α_1, α_2 and α_3 correspond to the effect of the COVID-19 crisis on student outcome in a given pandemic-affected semester.

5.1 Panel Data Estimation Methods

The approach of estimating the effect of the COVID-19 crisis on GPA of students of the Faculty of Social Sciences is discussed in this section. The following discussion will focus on equation (1), discussion for equation (2) is analogous.

As we observe GPA of each student for each semester since the winter semester of 2018/19, the data have panel structure. Therefore, time-invariant

unobservable individual characteristics can be accounted for by decomposing the error term e_{it} into a_i and u_{it}. The baseline equation takes the following form:

$$GPA_{it} = \beta_0 + \beta_1 COVID_t + Z_{it}\gamma + a_i + u_{it}, \quad (3)$$

where all variables are as described above, a_i is the time-invariant, individual-specific unobserved effect, u_{it} is the idiosyncratic error and β_1 is the desired effect.

The fixed effect a_i represents the unobserved characteristics of a given student, such as student's inherent abilities, intelligence, academic motivation, family, or financial background. If we were to include each of the available control variables into the model, it is possible that some of them would be correlated with a_i. For example, field of study is very likely to be correlated with a given student's abilities and innate skills assuming that students choose their field of study with regards to their capabilities. If the time-invariant, individual specific unobserved effect a_i is correlated with any of the explanatory variables, OLS estimates are biased and inconsistent (Wooldridge, 2013). Such bias is called the heterogeneity bias.

To deal with the heterogeneity bias, assuming there is any, several approaches can be employed. The most straightforward ones are the first-difference and fixed effects estimators as they are built to deal with heterogeneity bias by eliminating time-invariant effects, which eliminates not only the unobserved effect a_i, but also the variables controlling for sex and field of study as they are constant over time. Such approach may lead to unbiased results, however, we might lose precision in the estimates since GPA is highly heterogeneous across institutes, i. e. field-related subgroups of the faculty, as shown in the preceding chapter.

Another way to deal with heterogeneity bias is omitting the explanatory variable that is correlated with a_i. Assuming we have successfully identified the endogenous variable, omitting such variable solves the issue at hand, however, may introduce another omitted variable bias. For example, suppose that the field of study is correlated with unobserved student's ability. Omitting it from the regression equation would solve that problem but introduce a new one if field of study is correlated with sex.

Nevertheless, under the assumption that sex and field of study are uncorrelated, omitting the field variable leads to successful elimination of the bias. Given that there is no heterogeneity bias present, random effects estimator can also be utilized. Therefore, only sex and age will be controlled for when using random effects estimator.

Panel data methods, such as first-differencing, fixed effects, and random effects, are estimation approaches relying on transforming the baseline model, i. e. transforming the data, followed by applying pooled OLS on the transformed models to obtain the estimates.

It should be noted that for simpler and more transparent description of the mentioned transformations, the explanatory variables and their respective slope coefficients shall be represented by $X_{it}\beta$, where X_{it} is a row vector of the explanatory variables and β is a column vector of their respective slope coefficients.

5.1.1 First-Difference Estimator

The first-difference (FD) model is obtained by subtracting the lagged model equation from the baseline model equation.

$$GPA_{it} = \beta_0 + X_{it}\beta + a_i + u_{it}$$

$$GPA_{it-1} = \beta_0 + X_{it-1}\beta + a_i + u_{it-1}$$

Subtracting the second equation from the first one yields the FD model:

$$GPA_{it} - GPA_{it-1} = \underbrace{\beta_0 - \beta_0}_{=0} + (X_{it} - X_{it-1})\beta + \underbrace{a_i - a_i}_{=0} + u_{it} - u_{it-1}$$

Thus, the final model, which will be estimated by pooled OLS, takes the following form:

$$\Delta GPA_{it} = \Delta X_{it}\beta + \Delta u_{it},$$

where Δ denotes the change from period $t-1$ to period t. Notice that this approach eliminates the intercept, the unobserved effect a_i and any variable in X_{it} that is constant over time within all cross-sectional units. They have been "differenced away"

since they are time-invariant, which is exactly the solution needed for dealing with heterogeneity bias.

The FD transformation allowed us to remove the heterogeneity bias problem. However, eliminating heterogeneity bias is not sufficient for the estimates to be unbiased and consistent. For the FD estimator to be unbiased and consistent, additional assumptions need to be satisfied.

The first assumption states that the model follows the baseline equation, where the slope coefficients are the parameters to estimate and a_i is the unobserved effect. Secondly, the same random sample of cross-sectional units is observed. Thirdly, the rank condition, i. e. no explanatory variable is constant over time and there are no perfect linear relationships among the explanatory variables.

The fourth assumption, namely the strict exogeneity assumption, states that the expected value of the error term in any given time period conditional on the explanatory variables in all time periods and the unobserved effect equals zero:

$$E(u_{it}|X_i a_i) = 0.$$

Note: X_i contains all explanatory variables for all time periods for the i-th observation.

Be that as it may, the strict exogeneity assumption is stronger than necessary as we do not have to take the unobserved effect a_i into consideration since it has been differenced away. The following implication of the excessively strong assumption is sufficient for our purposes:

$$E(\Delta u_{it}|X_i) = 0, t = 2, \dots, T.$$

Given that the above-described assumptions hold, the FD estimator is unbiased. Under the same set of assumptions, FD estimator is also consistent for fixed number of time periods and large number of cross-sectional units.

For the inference to be at least asymptotically valid, i. e. for the standard errors and the test statistics to be calculated correctly, we need an additional set of assumptions. To begin with, the variance of the differenced error term conditional on the independent variables is assumed to be constant, that is the homoskedasticity assumption:

$$Var(\Delta u_{it}|X_i) = \sigma^2, t = 2, \dots, T.$$

And finally, conditional on the explanatory variables in all time periods, the idiosyncratic errors in any two different time periods are assumed to be uncorrelated, the so-called no serial correlation assumption:

$$Cov(\Delta u_{it}, \Delta u_{is}|X_i) = 0, t \neq s.$$

5.1.2 Fixed Effects Estimator

The fixed effects (FE) model eliminates the heterogeneity bias by utilizing the within transformation, which is obtained by averaging the baseline model over all time periods for each i and subtracting the averaged equation from the baseline model.

$$GPA_{it} = \beta_0 + X_{it}\beta + a_i + u_{it}$$

$$\overline{GPA}_i = \beta_0 + \bar{X}_i\beta + a_i + \bar{u}_i,$$

where $\overline{GPA}_i = T^{-1}\sum_{t=1}^{T} GPA_{it}$, analogously for $\bar{X}_i$ and $\bar{u}_i$. Subtracting the second equation from the first gives the FE equation:

$$GPA_{it} - \overline{GPA}_i = \underbrace{\beta_0 - \beta_0}_{=0} + (X_{it} - \bar{X}_i)\beta + \underbrace{a_i - a_i}_{=0} + u_{it} - \bar{u}_i$$

$$G\ddot{P}A_{it} = \ddot{X}_{it}\beta + \ddot{u}_{it}$$

The assumptions under which the FE estimator is unbiased and consistent are identical to those of the FD estimator. Strictly speaking, random sampling assumption, no perfect collinearity among the explanatory variables, the strict exogeneity assumption and linearity in parameters assumption.

Similarly, under the assumptions of homoskedasticity and no serial correlation of u_{it}, the validity of statistical inference relies on asymptotic approximations. For these approximations to be trustworthy, large number of cross-sectional units and small number of time periods is essential.

5.1.3 Random Effects Estimator

Random effects (RE) estimator is designed to efficiently estimate a regression model with panel data when none of the components of the composite error e_{it} is correlated with explanatory variables.

Due to the presence of a_i, the composite errors $e_{it} = a_i + u_{it}$ are by definition serially correlated. Therefore, RE removes serial correlation by quasi-demeaning the model, i. e. by removing a fraction of the time average from the model. Let

$$\lambda = 1 - \left(\frac{\sigma_u^2}{\sigma_u^2 + T\sigma_a^2}\right)^{\frac{1}{2}},$$

where σ_u^2 represents the variance of the idiosyncratic error, σ_a^2 denotes the variance of the unobserved individual effect a_i and T is the number of time periods. Followingly, by applying the RE transformation, we obtain the quasi-demeaned model:

$$GPA_{it} - \lambda\overline{GPA}_i = \beta_0(1-\lambda) + (X_{it} - \lambda\bar{X}_i)\beta + a_i - \lambda a_i + u_{it} - \lambda\bar{u}_i.$$

For the estimates to be consistent, we assume linearity in parameters, random sampling, no perfect collinearity among the explanatory variables, the expected value of the idiosyncratic error conditional on the explanatory variables and the fixed effect being equal 0, and constant expected value of a_i conditional on the explanatory variables.

Additionally, we assume homoskedasticity for both the idiosyncratic error and the unobserved individual effects, along with no serial correlation for the idiosyncratic error for the inference to be valid.

Given that we estimate both FE and RE models, the Hausman test for consistency can be utilized. The null hypothesis for the Hausman test states that both FE and RE models are consistent, whereas the alternative states that only FE is consistent. If the null is not rejected, then the RE model is the preferred one due to being more efficient than the FE model (Wooldridge, 2013).

5.1.4 Discussion of the Assumptions

This section is dedicated to the discussion of the underlying assumptions needed for the FD, FE and RE models to be valid.

The regression model specified by equation (3) is linear in parameters. We observe the full population of students affected by the change of modes of education in the summer semester of 2019/20. No perfect collinearity assumption is also met as for the FD and FE models, only the differenced or time-demeaned $COVID_t$ variable remains, for the RE model, we can safely assume that $COVID_t$, Sex_i and Age_{it} are not linearly dependent.

The zero conditional mean assumption is also met given that the COVID-19 crisis and the sudden lockdown in the summer semester of the academic year 2019/20 was exogenous.

The remaining two assumptions needed for the inference to be valid, i. e. homoskedasticity and no serial correlation assumptions, unlike the strict exogeneity assumption, can be tested. To test for presence of heteroskedasticity, the Breusch-Pagan test shall be used (Breusch & Pagan, 1979). The null hypothesis of the Breusch-Pagan test is "the variances of the error term are all equal". To test for presence of serial correlation, Breusch-Godfrey test will be applied (Breusch & Godfrey, 1978), where the null hypothesis assumes no serial correlation.

If the null hypothesis for either of the tests is rejected, then heteroskedasticity or autocorrelation is present and needs to be controlled for. To tackle this problem, heteroskedasticity- and or autocorrelation-robust standard errors shall be implemented (Arellano, 1987).

5.2 Binary Dependent Variable Models

In order to analyze the effect of COVID-19 crisis on the probability of graduating on time, linear probability models (LPM) and logistic regression will be used. This is because the dependent variable - an indicator of graduation at the end of the 3rd year of study (for bachelor's students) or at the end of the 2nd year of study (for master's students) - is binary. Moreover, given that each student is observed at the end of their 3rd (bachelor's) or 2nd (master's) year of study only once, the data used in this analysis is of cross-sectional nature.

For the purpose of illustration, the econometric models will be presented only for the case of bachelor's students, as the procedure for master's students is analogous. The baseline equations take the following form:

$$Graduated_i = \beta_0 + \beta_1 COVID_i + Z_i\gamma + e_i, \tag{4}$$

where $Graduated_i$ is a dummy variable equal to 1 if i-th student graduated at the end of their 3rd year, 0 otherwise and $COVID_i$ is a dummy variable equal to 1 if i-th student attended her 3rd year summer semester in the academic year 2019/20 or 2020/21. There are three groups of 3rd year students included in the data set – 3rd year students of 2018/19, 2019/20 and 2020/21. The 3rd year students of 2018/19 constitute the comparison group. Consequently, the coefficient β_1 can be interpreted as the difference in graduation probability between the COVID-19 years and the pre-COVID-19 year.

Similarly, model capturing the semestral detail of the effect (or in this case, yearly detail given that students can graduate standardly only in the summer semester, i. e. once per academic year):

$$Graduated_i = \alpha_0 + \alpha_1 sem_i^{2019/20S} + \alpha_3 sem_i^{2020/21S} + Z_i\delta + d_i, \tag{5}$$

where sem_i^t is also a dummy variable, which is equal to 1 if semester t equals i-th student's 3rd year summer semester, 0 otherwise. The coefficients α_1 and α_3 can be interpreted as the difference in graduation probability between the respective COVID-19 year and the pre-COVID-19 year, that is the academic year 2018/19.

The explanatory variables and their respective coefficients shall be referred to as $X_i\beta$, where X_i is a row vector of the explanatory variables and a one, and β is a column vector of the slope coefficients and the intercept.

5.2.1 Linear Probability Model

By applying OLS estimation method on the baseline equation where the dependent variable attains only 0 or 1, the linear probability model is obtained. Under the assumption of zero conditional mean of the disturbance, the following holds:

$$E(Graduated_i|X_i) = X_i\beta.$$

Due to the fact that $Graduated$ is a binary variable, the expected value on the left-hand side of the equation above can be expressed as:

$$E(Graduated_i|X_i) = 1 * P(Graduated_i = 1|X_i) + 0 * P(Graduated_i = 0|X_i).$$

Thus, the equation $E(Graduated_i|X_i) = P(Graduated_i = 1|X_i)$ states that the probability of success, that is, the probability of i-th student graduating in their 3rd year, is a linear function of the explanatory variables. The underlying assumptions necessary for OLS to produce unbiased and consistent estimates of regression coefficients are linearity in parameters, random sampling, no perfect collinearity and the zero conditional mean assumption.

Linearity in parameters is satisfied as per model specification by equations (4) and (5). The full population of 3rd year students for each of the respective academic years is observed. No perfect collinearity is assumed to hold since the dummy variables representing each of the academic years are by definition linearly independent. The zero conditional mean assumption is assumed to hold given that the dummy variables representing a given student's 3rd year are exogenous by nature.

Variance of a binary dependent variable conditional on the explanatory variables is a function of the explanatory variables which implies the presence of heteroskedasticity (Wooldridge, 2013). Thus, for the inference to be valid, heteroskedasticity-robust standard errors need to be calculated.

Due to the inbuilt heteroskedasticity and the fact that linear probability model may predict values that do not intersect with the range between 0 and 1, the logistic regression shall also be used.

5.2.2 Binary Logit Model

As the linear probability model may incorrectly predict values outside of the range between 0 and 1, we turn to the logistic regression. To model the probability of a 3rd year bachelor's student graduating, a non-linear function G assuming values in the 0 and 1 range is utilized in the following way (indices omitted):

$$P(Graduated = 1|X) = G(X\beta).$$

Where the function $G: \mathbb{R} \to (0, 1)$, also called the logistic function, is defined as follows:

$$G(z) = \frac{exp(z)}{1 + exp(z)}.$$

The logistic function is strictly increasing over all real numbers, and it is centrally symmetric with respect to the point $\left[0, \frac{1}{2}\right]$.

Derivation of the logit model can be done via the latent variable model:

$$Graduated^* = X\beta + u.$$

Where $Graduated^*$ is an unobserved variable linearly influenced by the explanatory variables X. This so-called latent variable and the binary dependent variable $Graduated$ are related in the following manner:

$$Graduated = \begin{cases} 1 \text{ if } Graduated^* > 0, \\ 0 \text{ if } Graduated^* \leq 0. \end{cases}$$

Assuming that the error term u is independent of the explanatory variables and that u has the standard logistic distribution:

$$P(Graduated = 1|X) = P(Graduated^* > 0|X) = P(u > -X\beta|X) = 1 - G(-X\beta)$$
$$\underbrace{=}_{\substack{u \text{ is symmetrically} \\ \text{distributed arround } 0}} G(X\beta).$$

As a consequence of the probability of success being non-linearly dependent on the explanatory variables, maximum likelihood estimator shall be utilized to estimate the binary response model. Let n denote the number of observations in our sample and let the following function be the density of $Graduated$ conditional on the explanatory variables:

$$f(Graduated|X_i;\beta) = [G(X_i\beta)]^{Graduated}[1 - G(X_i\beta)]^{Graduated-1}.$$

To obtain the log-likelihood function, we take the natural logarithm of the density function:

$$\ell_i(\beta) = Graduated_i * log[G(X_i\beta)] + (1 - Graduated_i) * log[1 - G(X_i\beta)].$$

The maximum likelihood estimator, i. e. the logit estimator, of β is found by maximizing the following sum of $\ell_i(\beta)$ across all observations with respect to β:

$$\mathcal{L}_i(\beta) = \sum_{i=1}^{n} Graduated_i * log[G(X_i\beta)] + (1 - Graduated_i) * log[1 - G(X_i\beta)].$$

Under very general assumptions, the maximum likelihood estimator is consistent, asymptotically normal, and asymptotically efficient (Wooldridge, 2013).

As the relationship between the dependent variable and the explanatory variables is non-linear, the interpretation of results is somewhat more complicated than in the case of linear regression (Wooldridge, 2013). To determine the partial effect of a continuous explanatory variable on the response probability, we differentiate with respect to the given variable:

$$\frac{\partial P(Graduated = 1|X)}{\partial x_j} = g(X\beta)\beta_j, \quad \text{where } g(z) = \frac{\partial G(z)}{\partial z}(z).$$

For a binary explanatory variable x_j, the partial effect is defined in the following way:

$$G(\beta_0 + \beta_1 x_1 + \cdots + \beta_{j-1}x_{j-1} + \beta_j) - G(\beta_0 + \beta_1 x_1 + \cdots + \beta_{j-1}x_{j-1}).$$

When considering the beta coefficients alone, the only inference to be made is that of the direction of the partial effect. As the logistic function is strictly increasing

over all real numbers, its derivative will always be positive, therefore the $g(X\beta)$ term is an adjustment factor affecting the magnitude, but not the sign of the partial effect. To obtain a similar measure of the partial effect as when running a linear regression, the average partial effect (APE) shall be calculated by taking the average of the individual effects across the sample[1].

The APE for a continuous variable x_j:

$$APE_j = \frac{1}{n}\sum_{i=1}^{n} g(\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_j x_{ji}\,)\,\hat{\beta}_j.$$

The APE for a discrete variable x_k, i. e. a change in x_k from 0 to 1:

$$APE_k = \frac{1}{n}\sum_{i=1}^{n}\big[G\big(\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_{k-1} x_{k-1,i} + \hat{\beta}_k\big) - G\big(\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_{k-1} x_{k-1,i}\big)\big].$$

[1] Alternatively, one could construct the partial effect at average.

6 Results

The purpose of this section is to present empirical results of the analysis described in the preceding chapters. The regression model estimates of the effect of COVID-19 on GPA and the probability of graduating are presented.

6.1 GPA

To estimate the effect of COVID-19 on GPA of students of the Faculty of Social Sciences, panel data estimation methods were employed. Various models were run using data of all bachelor's and master's students and their subgroups, respectively, to examine heterogeneity of the effect across genders and institutes of the faculty.

Table 8 presents the effect of COVID-19 crisis on GPA estimated on different samples of bachelor's students – all bachelor's students, female bachelor's, male bachelor's, and bachelor's with respect to attended institute.

Table 8: Effect of COVID-19 on GPA (bachelor's)

	Dependent variable:		
		GPA	
All bachelor's students	(1)	(2)	(3)
COVID	**-0.189 *****	-0.219 ***	-0.309 ***
	(0.018)	(0.014)	(0.012)
R^2	**0.014**	0.138	0.247
Adjusted R^2	**0.014**	-0.106	0.247
Number of Observations	**9332**	11984	11984
Female			
COVID	**-0.169 *****	-0.184 ***	-0.290 ***
	(0.025)	(0.02)	(0.018)
R^2	**0.011**	0.136	0.222
Adjusted R^2	**0.011**	-0.105	0.222
Number of Observations	**4811**	6154	6154
Male			
COVID	**-0.210 *****	-0.256 ***	-0.329 ***
	(0.025)	(0.02)	(0.017)
R^2	**0.020**	0.142	0.268
Adjusted R^2	**0.020**	-0.107	0.268
Number of Observations	**4521**	5830	5830

Note: Standard errors in parentheses. *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. Column (1) reports FD estimates, column (2) reports FE estimates, and column (3) reports RE estimates. The reported coefficients correspond to the estimate of β_1 in equation (1). There are no control variables in FD and FE regression models. In RE model, we control for sex and age. Preferred model in bold.

Table 8 (continued)

	Dependent variable:		
	GPA		
IES	(1)	(2)	(3)
COVID	-0.223 ***	-0.223 ***	**-0.256 *****
	(0.044)	(0.037)	**(0.033)**
R^2	0.029	0.076	**0.224**
Adjusted R^2	0.028	-0.223	**0.223**
Number of Observations	1799	2380	**2380**
ICSJ			
COVID	**-0.086 *****	-0.142 ***	-0.198 ***
	(0.032)	(0.022)	(0.019)
R^2	**0.040**	0.116	0.134
Adjusted R^2	**0.039**	-0.108	0.133
Number of Observations	**2425**	3040	3040
IIS			
COVID	**-0.185 *****	-0.148 ***	-0.284 ***
	(0.039)	(0.034)	(0.032)
R^2	**0.010**	0.110	0.188
Adjusted R^2	**0.009**	-0.133	0.187
Number of Observations	**1866**	2375	2375
IPS			
COVID	**-0.298 *****	-0.406 ***	-0.479 ***
	(0.036)	(0.029)	(0.025)
R^2	**0.138**	0.206	0.306
Adjusted R^2	**0.137**	-0.014	0.305
Number of Observations	**1992**	2542	2542
ISS			
COVID	**-0.163 *****	-0.145 ***	-0.393 ***
	(0.048)	(0.04)	(0.042)
R^2	**0.036**	0.295	0.404
Adjusted R^2	**0.034**	0.070	0.403
Number of Observations	**1250**	1647	1647

Note: Standard errors in parentheses. *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. Column (1) reports FD estimates, column (2) reports FE estimates, and column (3) reports RE estimates. The reported coefficients correspond to the estimate of β_1 in equation (1). There are no control variables in FD and FE regression models. In RE model, we control for sex and age. Preferred model in bold.

Hausman test for consistency of FE and RE models was employed, resulting in the null hypothesis being rejected for each of the groups except IES bachelor's students. Each of the models was tested for autocorrelation – no serial correlation assumption was rejected for all of the models – as well as for heteroskedasticity which resulted in the homoskedasticity assumption being rejected and thus, we calculated the heteroskedasticity- and autocorrelation-robust standard errors to assure that inference is valid.

Due to the presence of serially correlated errors, the FD models are a more suitable option since FE models are more efficient when the errors are well behaved which is not the case. Therefore, the results of the FD models are considered to be the most reliable – except for the case of IES bachelor's where RE is preferred as per the result of Hausman test.

The effect of COVID-19 on GPA is negative and highly significant when considering all bachelor's and their respective groupings by sex, and institute, which suggests that GPA improved under the pandemic. The results will be further discussed in the Discussion section.

Analogously to Table 8, Table 9 presents the estimates for master's students.

Table 9: Effect of COVID-19 on GPA (master's)

	Dependent variable:		
		GPA	
All master's students	(1)	(2)	(3)
COVID	**-0.095 *****	-0.115 ***	-0.197 ***
	(0.02)	(0.016)	(0.013)
R^2	**0.004**	0.048	0.124
Adjusted R^2	**0.004**	-0.308	0.124
Number of Observations	**6157**	8456	8456
Female			
COVID	**-0.104 *****	-0.112 ***	-0.187 ***
	(0.026)	(0.02)	(0.017)
R^2	**0.006**	0.053	0.114
Adjusted R^2	**0.005**	-0.295	0.114
Number of Observations	**3584**	4898	4898
Male			
COVID	**-0.080 ***	-0.120 ***	-0.206 ***
	(0.032)	(0.025)	(0.02)
R^2	**0.003**	0.042	0.131
Adjusted R^2	**0.002**	-0.325	0.131
Number of Observations	**2573**	3558	3558

Note: Standard errors in parentheses. *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. Column (1) reports FD estimates, column (2) reports FE estimates, and column (3) reports RE estimates. The reported coefficients correspond to the estimate of β_1 in equation (1). There are no control variables in FD and FE regression models. In RE model, we control for sex and age. Preferred model in bold.

Hausman test proved RE to be consistent only for ISS students Both tests for heteroskedasticity and serial correlation rejected the null hypothesis for each of the models, thus heteroskedasticity- and autocorrelation-robust standard errors were calculated. Similarly, FD is the preferred model since FE does not perform as well due

to the presence of serially correlated errors. Thus, the results of the FD model are assumed to be more reliable except for the case of ISS students, where RE is preferred.

Table 9 (continued)

	Dependent variable:		
	GPA		
IES	(1)	(2)	(3)
COVID	**-0.393 *****	-0.444 ***	-0.426 ***
	(0.058)	(0.05)	(0.039)
R^2	**0.139**	0.117	0.193
Adjusted R^2	**0.137**	-0.258	0.191
Number of Observations	**722**	1027	1027
ICSJ			
COVID	**-0.099 ****	-0.128 ***	-0.225 ***
	(0.031)	(0.026)	(0.024)
R^2	**0.043**	0.144	0.119
Adjusted R^2	**0.041**	-0.137	0.118
Number of Observations	**1645**	2184	2184
IIS			
COVID	**0.036**	0.007	-0.106 ***
	(0.035)	(0.026)	(0.021)
R^2	**0.004**	0.027	0.076
Adjusted R^2	**0.003**	-0.334	0.075
Number of Observations	**1902**	2604	2604
IPS			
COVID	**-0.107 ***	-0.146 ***	-0.200 ***
	(0.045)	(0.036)	(0.03)
R^2	**0.023**	0.036	0.131
Adjusted R^2	**0.022**	-0.364	0.129
Number of Observations	**1298**	1835	1835
ISS			
COVID	-0.206 **	-0.018	**-0.079**
	(0.072)	(0.059)	**(0.048)**
R^2	0.019	0.003	**0.114**
Adjusted R^2	0.016	-0.365	**0.111**
Number of Observations	590	806	**806**

Note: Standard errors in parentheses. *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. Column (1) reports FD estimates, column (2) reports FE estimates, and column (3) reports RE estimates. The reported coefficients correspond to the estimate of β_1 in equation (1). There are no control variables in FD and FE regression models. In RE model, we control for sex and age. Preferred model in bold.

The effect of COVID-19 on GPA is statistically significant when considering master's students except for IIS and ISS students. The results imply improved GPA, will be further discussed in the succeeding section.

The following tables present analogous estimates to those in Table 8 and Table 9 for models specified by equation (2).

Table 10: Semestral effect of COVID-19 on GPA (bachelor's)

	Dependent variable:		
	GPA		
All bachelor's students	(1)	(2)	(3)
2019/20 S	**-0.186 *****	-0.356 ***	-0.356 ***
	(0.018)	(0.016)	(0.015)
2020/21 W	**-0.176 *****	-0.364 ***	-0.384 ***
	(0.024)	(0.016)	(0.016)
2020/21 S	**-0.174 *****	-0.418 ***	-0.423 ***
	(0.029)	(0.019)	(0.018)
R^2	**0.014**	0.105	0.236
Adjusted R^2	**0.014**	-0.150	0.236
Number of Observations	**9332**	11984	11984
Female			
2019/20 S	**-0.174 *****	-0.344 ***	-0.351 ***
	(0.026)	(0.022)	(0.021)
2020/21 W	**-0.146 *****	-0.337 ***	-0.357 ***
	(0.034)	(0.023)	(0.022)
2020/21 S	**-0.173 *****	-0.420 ***	-0.430 ***
	(0.04)	(0.025)	(0.024)
R2	**0.011**	0.096	0.212
Adjusted R2	**0.010**	-0.157	0.211
Number of Observations	**4811**	6154	6154
Male			
2019/20 S	**-0.198 *****	-0.368 ***	-0.361 ***
	(0.026)	(0.023)	(0.022)
2020/21 W	**-0.208 *****	-0.392 ***	-0.411 ***
	(0.034)	(0.023)	(0.023)
2020/21 S	**-0.175 *****	-0.416 ***	-0.414 ***
	(0.042)	(0.029)	(0.028)
R2	**0.020**	0.115	0.261
Adjusted R2	**0.020**	-0.141	0.260
Number of Observations	**4521**	5830	5830
IES			
2019/20 S	**-0.208 *****	-0.385 ***	-0.373 ***
	(0.045)	(0.041)	(0.038)
2020/21 W	**0.056**	-0.226 ***	-0.230 ***
	(0.057)	(0.043)	(0.042)
2020/21 S	**0.116**	-0.201 ***	-0.190 ***
	(0.066)	(0.05)	(0.048)
R2	**0.045**	0.078	0.228
Adjusted R2	**0.043**	-0.222	0.227
Number of Observations	**1799**	2380	2380

Note: Standard errors in parentheses. *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. Column (1) reports FD estimates, column (2) reports FE estimates, and column (3) reports RE estimates. The reported coefficients correspond to the estimates of α_1, α_2 and α_3 in equation (2). There are no control variables in FD and FE regression models. In RE model, we control for sex and age. Preferred model in bold.

Table 10 (continued)

	Dependent variable:		
	GPA		
ICSJ	(1)	(2)	(3)
2019/20 S	**-0.125 *****	-0.273 ***	-0.238 ***
	(0.033)	(0.026)	(0.026)
2020/21 W	**-0.195 *****	-0.283 ***	-0.253 ***
	(0.042)	(0.023)	(0.023)
2020/21 S	**-0.329 *****	-0.463 ***	-0.419 ***
	(0.052)	(0.031)	(0.03)
R2	**0.045**	0.127	0.146
Adjusted R2	**0.043**	-0.096	0.145
Number of Observations	**2425**	3040	3040
IIS			
2019/20 S	**-0.220 *****	-0.399 ***	-0.404 ***
	(0.041)	(0.035)	(0.035)
2020/21 W	**-0.208 *****	-0.254 ***	-0.289 ***
	(0.062)	(0.045)	(0.044)
2020/21 S	**-0.395 *****	-0.509 ***	-0.533 ***
	(0.073)	(0.049)	(0.046)
R2	**0.015**	0.082	0.182
Adjusted R2	**0.013**	-0.171	0.181
Number of Observations	**1866**	2375	2375
IPS			
2019/20 S	**-0.248 *****	-0.346 ***	-0.354 ***
	(0.037)	(0.032)	(0.031)
2020/21 W	**-0.361 *****	-0.503 ***	-0.527 ***
	(0.046)	(0.031)	(0.03)
2020/21 S	**-0.208 *****	-0.379 ***	-0.394 ***
	(0.058)	(0.04)	(0.038)
R2	**0.147**	0.229	0.328
Adjusted R2	**0.145**	0.015	0.327
Number of Observations	**1992**	2542	2542
ISS			
2019/20 S	**-0.127 ***	-0.434 ***	-0.450 ***
	(0.049)	(0.046)	(0.044)
2020/21 W	**-0.086**	-0.637 ***	-0.692 ***
	(0.061)	(0.041)	(0.038)
2020/21 S	**0.138**	-0.513 ***	-0.532 ***
	(0.072)	(0.051)	(0.05)
R2	**0.053**	0.164	0.378
Adjusted R2	**0.050**	-0.105	0.377
Number of Observations	**1250**	1647	1647

Note: Standard errors in parentheses. *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. Column (1) reports FD estimates, column (2) reports FE estimates, and column (3) reports RE estimates. The reported coefficients correspond to the estimates of α_1, α_2 and α_3 in equation (2). There are no control variables in FD and FE regression models. In RE model, we control for sex and age. Preferred model in bold.

Table 11: Semestral effect of COVID-19 on GPA (master's)

	Dependent variable:		
	GPA		
All master's students	(1)	(2)	(3)
2019/20 S	**-0.086 *****	-0.145 ***	-0.141 ***
	(0.02)	(0.018)	(0.017)
2020/21 W	**-0.033**	-0.223 ***	-0.236 ***
	(0.028)	(0.02)	(0.019)
2020/21 S	**0.016**	-0.209 ***	-0.207 ***
	(0.035)	(0.026)	(0.024)
R2	**0.005**	0.033	0.124
Adjusted R2	**0.005**	-0.329	0.123
Number of Observations	**6157**	8456	8456
Female			
2019/20 S	**-0.100 *****	-0.167 ***	-0.158 ***
	(0.026)	(0.023)	(0.021)
2020/21 W	**-0.028**	-0.210 ***	-0.219 ***
	(0.034)	(0.027)	(0.024)
2020/21 S	**0.023**	-0.203 ***	-0.203 ***
	(0.042)	(0.034)	(0.031)
R2	**0.008**	0.032	0.114
Adjusted R2	**0.007**	-0.324	0.113
Number of Observations	**3584**	4898	4898
Male			
2019/20 S	**-0.067 ***	-0.114 ***	-0.120 ***
	(0.032)	(0.029)	(0.026)
2020/21 W	**-0.040**	-0.238 ***	-0.258 ***
	(0.048)	(0.03)	(0.029)
2020/21 S	**0.009**	-0.210 ***	-0.211 ***
	(0.059)	(0.041)	(0.039)
R2	**0.004**	0.035	0.135
Adjusted R2	**0.003**	-0.336	0.134
Number of Observations	**2573**	3558	3558
IES			
2019/20 S	-0.378 ***	-0.372 ***	**-0.346 *****
	(0.058)	(0.057)	**(0.049)**
2020/21 W	-0.169 *	-0.337 ***	**-0.311 *****
	(0.072)	(0.056)	**(0.053)**
2020/21 S	-0.107	-0.249 **	**-0.184 ***
	(0.093)	(0.086)	**(0.082)**
R2	0.153	0.162	**0.225**
Adjusted R2	0.148	-0.198	**0.222**
Number of Observations	722	1027	**1027**

Note: Standard errors in parentheses. *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. Column (1) reports FD estimates, column (2) reports FE estimates, and column (3) reports RE estimates. The reported coefficients correspond to the estimates of α_1, α_2 and α_3 in equation (2). There are no control variables in FD and FE regression models. In RE model, we control for sex and age. Preferred model in bold.

Table 11 (Continued)

	Dependent variable:		
	GPA		
ICSJ	(1)	(2)	(3)
2019/20 S	**-0.118 *****	-0.235 ***	-0.179 ***
	(0.032)	(0.029)	(0.029)
2020/21 W	**-0.074**	-0.354 ***	-0.305 ***
	(0.051)	(0.037)	(0.034)
2020/21 S	**-0.146 ****	-0.496 ***	-0.416 ***
	(0.054)	(0.037)	(0.034)
R2	**0.046**	0.108	0.131
Adjusted R2	**0.044**	-0.186	0.130
Number of Observations	**1645**	2184	2184
IIS			
2019/20 S	0.036	-0.033	**-0.048**
	(0.035)	(0.029)	**(0.028)**
2020/21 W	0.005	-0.120 ***	**-0.158 *****
	(0.056)	(0.034)	**(0.03)**
2020/21 S	-0.004	-0.157 ***	**-0.189 *****
	(0.064)	(0.038)	**(0.033)**
R2	0.003	0.012	**0.082**
Adjusted R2	0.001	-0.355	**0.081**
Number of Observations	1902	2604	**2604**
IPS			
2019/20 S	**-0.047**	-0.104 *	-0.105 **
	(0.044)	(0.046)	(0.04)
2020/21 W	**-0.030**	-0.235 ***	-0.259 ***
	(0.056)	(0.043)	(0.042)
2020/21 S	**0.288 *****	0.048	0.048
	(0.085)	(0.073)	(0.073)
R2	**0.045**	0.059	0.150
Adjusted R2	**0.042**	-0.334	0.148
Number of Observations	**1298**	1835	1835
ISS			
2019/20 S	**-0.211 ****	-0.149 *	-0.157 *
	(0.073)	(0.067)	(0.064)
2020/21 W	**0.028**	-0.002	-0.051
	(0.097)	(0.074)	(0.07)
2020/21 S	**0.052**	-0.029	-0.089
	(0.136)	(0.119)	(0.107)
R2	**0.034**	0.014	0.119
Adjusted R2	**0.027**	-0.354	0.114
Number of Observations	**590**	806	806

Note: Standard errors in parentheses. *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. Column (1) reports FD estimates, column (2) reports FE estimates, and column (3) reports RE estimates. The reported coefficients correspond to the estimates of α_1, α_2 and α_3 in equation (2). There are no control variables in FD and FE regression models. In RE model, we control for sex and age. Preferred model in bold.

The decision-making process behind choosing the preferred model in Table 10 and Table 11 is similar to the process used at the beginning of this section. That is, RE is preferred when Hausman proves RE to be consistent, otherwise FD is the preferred model due to the residuals being autocorrelated in each of the presented models.

Moreover, in each of the models, the disturbances are not only autocorrelated but also heteroskedastic, therefore, heteroskedasticity- and autocorrelation-robust standard errors are provided.

The results are somewhat heterogeneous. The effect of the summer semester 2019/20 is negative and highly significant for most of the subgroups the models were run on. When considering bachelor's students, each of the effects is negative and significant for all subgroups except IES and ISS students. For master's, significance and even the direction of the effect differs. Further discussion on results will be provided in the following section.

6.2 Probability of Graduating

For the purpose of estimating the effect of COVID-19 on the probability of 3rd year bachelor's or 2nd year master's graduating standardly, linear probability model and logistic regression were utilized.

Table 10 reports the regression coefficients estimated on all bachelor's students, female bachelor's, and male bachelor's. Grouping by institute is not included as there are too few observations to divide the sample of 3rd year students by this criterion.

Table 12: Effect of COVID-19 on the probability of graduating after 3rd year (bachelor's)

	Dependent variable:	
	Graduated	
All bachelor's students	(1)	(2)
COVID	-0.080 **	-0.080 **
	(0.028)	(0.028)
R^2	0.016	-
Adjusted R^2	0.013	-
Percent Correctly Predicted	61.08%	61.16%
Number of Observations	1272	1272
Female		
COVID	-0.069	-0.069
	(0.038)	(0.038)
R^2	0.031	-
Adjusted R^2	0.022	-
Percent Correctly Predicted	63.12%	63.12%
Number of Observations	667	667
Male		
COVID	-0.090 *	-0.090 *
	(0.042)	(0.042)
R^2	0.027	-
Adjusted R^2	0.018	-
Percent Correctly Predicted	59.34%	59.34%
Number of Observations	605	605

Note: Standard errors in parentheses. *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. Column (1) reports OLS estimates, and column (2) reports average partial effects based on logit estimates. The reported coefficients correspond to the estimate of β_1 in equation (4). We control for sex and age.

The percent correctly predicted at arbitrarily chosen threshold 0.5 was calculated to decide which of the models performs better. Given that percent correctly

predicted measures do not seem to differ, we cannot use it as means to decide which of the two performs better, however, due to the estimated partial effects being nearly identical when comparing OLS estimates to logit APE, we do not have to choose.

Table 11 presents corresponding results to those in Table 10 for master's students.

Table 13: Effect of COVID-19 on the probability of graduating after 2nd year (master's)

	Dependent variable:	
	Graduated	
All master's students	(1)	(2)
COVID	-0.049 *	-0.049 *
	(0.024)	(0.024)
R^2	0.003	-
Adjusted R^2	0.001	-
Percent Correctly Predicted	52.08%	52.58%
Number of Observations	1799	1799
Female		
COVID	-0.044	-0.044
	(0.03)	(0.03)
R^2	0.053	-
Adjusted R^2	0.047	-
Percent Correctly Predicted	59.41%	59.51%
Number of Observations	1057	1057
Male		
COVID	-0.106 **	-0.105 **
	(0.037)	(0.036)
R^2	0.058	-
Adjusted R^2	0.050	-
Percent Correctly Predicted	61.32%	61.32%
Number of Observations	742	742

Note: Standard errors in parentheses. *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. Column (1) reports OLS estimates, and column (2) reports average partial effects based on logit estimates. The reported coefficients correspond to the estimate of β_1 in equation (4). We control for sex and age.

Similarly, given that the estimates are nearly identical, we do not have to choose whether LPM, or logit performs better. The percent correctly predicted was calculated at arbitrarily chosen threshold 0.5.

The following tables present analogous estimates to those in Table 12 and Table 13 for models specified by equation (5).

Table 14: Semestral effect of COVID-19 on the probability of graduating after 3rd year (bachelor's)

	Dependent variable:	
	Graduated	
All bachelor's students	(1)	(2)
2019/20 S	-0.038	-0.039
	(0.033)	(0.034)
2020/21 S	-0.121 ***	-0.122 ***
	(0.033)	(0.034)
R^2	0.035	-
Adjusted R^2	0.029	-
Percent Correctly Predicted	61.40%	61.40%
Number of Observations	1272	1272
Female		
2019/20 S	-0.019	-0.019
	(0.044)	(0.045)
2020/21 S	-0.121 **	-0.122 **
	(0.045)	(0.046)
R^2	0.038	-
Adjusted R^2	0.028	-
Percent Correctly Predicted	63.87%	64.32%
Number of Observations	667	667
Male		
2019/20 S	-0.058	-0.059
	(0.05)	(0.05)
2020/21 S	-0.119 *	-0.119 *
	(0.048)	(0.049)
R^2	0.030	-
Adjusted R^2	0.018	-
Percent Correctly Predicted	61.65%	61.65%
Number of Observations	605	605

Note: Standard errors in parentheses. *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. Column (1) reports OLS estimates, and column (2) reports average partial effects based on logit estimates. The reported coefficients correspond to the estimates of α_1 and α_3 in equation (5). We control for sex and age.

Again, both models seem to perform similarly. The effect on probability of graduating is statistically insignificant for the academic year 2019/20 and negative for the year 2020/21. There appears to be no major difference among genders.

Table 15: Semestral effect of COVID-19 on the probability of graduating after 3rd year (master's)

	Dependent variable:	
	Graduated	
All master's students	(1)	(2)
2019/20 S	0.032	0.032
	(0.029)	(0.028)
2020/21 S	-0.155 ***	-0.155 ***
	(0.027)	(0.027)
R^2	0.071	-
Adjusted R^2	0.067	-
Percent Correctly Predicted	61.65%	61.65%
Number of Observations	1799	1799
Female		
2019/20 S	0.078 *	0.079 *
	(0.037)	(0.037)
2020/21 S	-0.157 ***	-0.157 ***
	(0.035)	(0.036)
R^2	0.082	-
Adjusted R^2	0.076	-
Percent Correctly Predicted	61.02%	61.02%
Number of Observations	1057	1057
Male		
2019/20 S	-0.037	-0.037
	(0.046)	(0.045)
2020/21 S	-0.158 ***	-0.157 ***
	(0.041)	(0.041)
R^2	0.066	-
Adjusted R^2	0.057	-
Percent Correctly Predicted	60.38%	60.24%
Number of Observations	742	742

Note: Standard errors in parentheses. *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. Column (1) reports OLS estimates, and column (2) reports average partial effects based on logit estimates. The reported coefficients correspond to the estimates of α_1 and α_3 in equation (5). We control for sex and age.

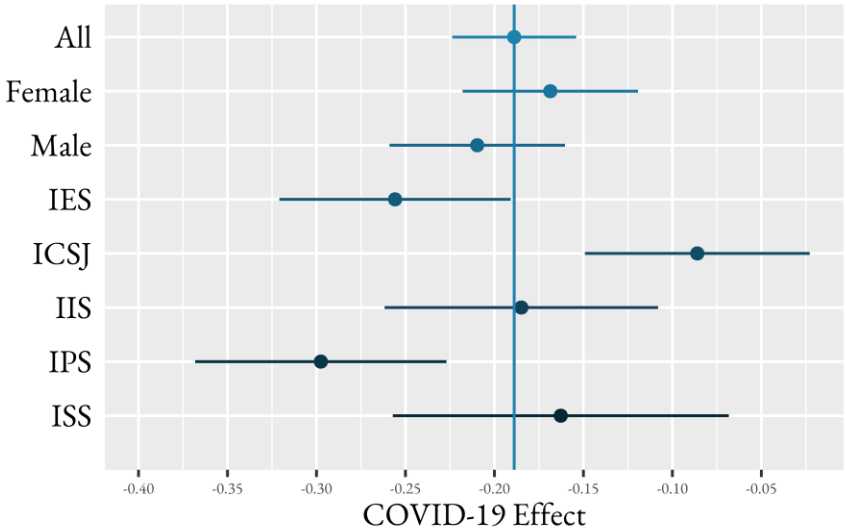
Likewise, logit and OLS do not seem to differ. Results are similar to those of bachelor's students except that the magnitude of the effect in the academic year 2020/21 is greater.

7 Discussion

The aim of this section is to interpret empirical results presented in the preceding chapter, discuss them in context of existing literature covered in literature review, and finally answer the research question.

Figure 7 and Figure 8 depict the estimated effect of COVID-19 on GPA from the preferred models (those specified by equation (1)) across defined subgroups of students for bachelor's and master's students, respectively.

Figure 7: Estimate plot: aggregate effect of COVID-19 on GPA (bachelor's)



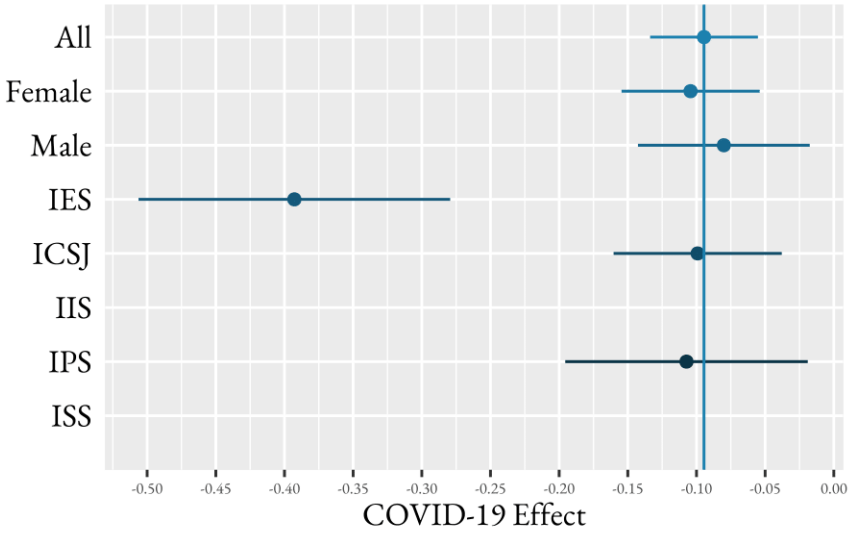
Note: Bars denote 95% confidence interval.

The aggregate effect on bachelor's students could be interpreted as "the closure of universities and classes being conducted online mode caused a decrease in GPA of bachelor's students by 0.189 on average".

The effect is negative for each of the subgroups, that is each group had improved their GPA. The magnitude of the effect seems to be greater for male than female students and students attending IES and IPS when compared to the remaining institutes.

Semestral decomposition of the effect uncovered that the effect on all bachelor's, females, males and all institutes except IES and ISS is negative and highly significant for each of the 3 semesters. For IES and ISS students, the effect is negative for the summer semester of 2019/20 and insignificant for the remaining semesters.

Figure 8: Estimate plot: aggregate effect of COVID-19 on GPA (master's)



Note: Bars denote 95% confidence interval. Missing values indicate insignificant estimate.

The aggregate effect on master's students could be interpreted as "the closure of universities and classes being conducted online mode caused a decrease in GPA of bachelor's students by 0.095 on average".

The effect is negative for each of the subgroups but IIS and ISS students, where we cannot conclude anything due to the insignificance of the estimate. The magnitude of the effect seems to be greater for female than male students. Interestingly, IES master's appear to be the most severely affected group with estimated effect of -0.393, which was probably driven by substantial decline in GPA of second-year IES master's students as per Figure 5. Nevertheless, the root cause of the effect cannot be inferred as the available data do not carry such information.

Regarding the semestral decomposition for master's students, the results vary in significance, magnitude and even direction. When considering all master's, females and males, the effect is negative for the summer semester of 2019/20 and insignificant for the rest. Effect on IES master's is negative and significant for each of the semesters. Interestingly, the effect on IPS master's is insignificant for the first two semesters and positive for the last, i. e. the summer semester of 2020/21.

Considering the pre-COVID-19 comparison of traditional and online education, which generally suggests that students partaking in online education

perform worse than their traditional counterpart, it is very likely that we are facing a case of grade inflation, i. e. students' being given better grades than they deserve.

Unfortunately, the scope of this thesis and the data available do not allow us to draw conclusions on the cause of the improvement in GPA. We can only speculate about the driving factors of the effect based on the research regarding the effect of COVID-19 on academic performance without any supporting evidence.

The direction of the effect, that is grades being improved under the COVID-19 crisis, concurs not only with findings of Rodríguez-Planas (2022a), who utilized similar methods, specifically panel data estimation methods, but also with findings of Gonzalez et al. (2020), Karadag (2021) and Iglesias-Pradas et al. (2021) whose approach consisted of testing differences in academic performance of same-grade students across numerous academic years.

When comparing the estimated effect for master's and bachelor's students, it seems that master's weren't affected as much as bachelor's, at least on the aggregate level since the magnitude of the effect on all bachelor's is almost twice as large as the magnitude of the effect on all master's. The fact that bachelor's students seem to have improved their grades with COVID-19 more than master's students could be considered counter-intuitive due to several reasons.

Firstly, one would expect that more mature students, i. e. master's students, might suffer less from the negative circumstances brought about by the pandemic as they are more accustomed to attending higher education facility. Secondly, bachelor's courses are usually conducted in larger groups of people, for that reason master's students are more likely to have their individual needs met as the lecturer's attention is not spread across greater number of students. Therefore, under the just discussed expectations, master's students gain should outweigh that of bachelor's students.

When considering the effect on different institutes of the faculty, it indeed varies in magnitude for both master's and bachelor's students, for example, IES master's seem to be heavily affected whereas no significant effect was found for IIS and ISS master's. This institutional heterogeneity may be caused by factors attributed to each of the respective institutes which is something we do not observe and thus

cannot come to any conclusions. The only inference to be made is that there is a certain effect on such group of students.

The aggregate effect of the pandemic on the probability of graduating is negative for both bachelor's and master's. However, the semestral (or in this case, yearly) decomposition shows that the effect is insignificant for the academic year 2019/20 and highly significant and negative for 2020/21 for each of the subgroups the models were run on. Effect differs negligibly across genders.

The probability of 3rd year bachelor's graduating in 2020/21 decreased by more than 12% points when compared to the reference group, i. e. 3rd year bachelor's students of 2018/19.

The regression models for master's yield similar results to those of bachelor's students. The probability of a 2nd year master's student graduating in the year 2020/21 has decreased by more than 15% points when compared to the reference group.

When considering the effect of COVID-19 on the probability of graduation in the context of existing literature, our findings confirm the claim of Bettinger et al. (2017), that taking a course online decreases probability of a student remaining enrolled in the following semesters. It is apparent that the effect is not immediate as it is insignificant for the first pandemic-affected semester. Students attending online classes may therefore dropout or not finish their studies. Unfortunately, the underlying mechanism behind the effect is unknown.

Saw et al. (2020) based on an online survey proposes that students may have delayed graduation due to limited access to academic facilities and resources, declined mental or physical health or set-back in degree-related tasks. Given that these are the students that haven't been present in school since March 2020 and were under strict lockdown for most of the academic year 2020/21, their mental and physical health, along with their study motivation, may have been impaired. As a result, a certain part of the students delayed their graduation.

Additionally, given that non-negligible part of students was not able to graduate standardly in the academic year 2020/21, the idea that the improvement in

GPA is grade inflation rather than real improvement in students abilities seems plausible since students with higher ability would be much more likely to finish standardly.

Finally, to answer the research question, which asks whether COVID-19 and its consecutive switch to online education had any effect on GPA, and the probability of graduation of students of the Faculty of Social Sciences at Charles University. The effect on GPA is negative for majority of the analyzed groups, in other words, GPA improved with COVID-19. Regarding the probability of graduating, the probability of 3rd year bachelor's and 2nd year master's graduating in 2020/21 significantly decreased when compared to the reference group, i. e. the students of 2018/19.

8 Conclusion

The purpose of this thesis was to determine whether COVID-19 and its consecutive switch to online education had any effect on GPA and the probability of graduating of students of the Faculty of Social Sciences at Charles University.

Initial investigation of the data consisted of providing summary statistics and graphical representation of available variables. Additionally, the cross-sectional units were discussed in greater detail as their grouping by sex, institute, and level of study plays an important role in the further analysis.

To analyze the effect of COVID-19 crisis on GPA, panel data estimation methods were utilized. The estimated results mostly imply that the effect on GPA is negative, i. e. improvement in GPA, with varying magnitude across given subgroups of students.

In answering if there is any effect on the probability of graduating, binary dependent variable models were employed. According to the reported estimates, there was no effect on students graduating in the academic year 2019/20, however, the probability decreased for students graduating in the following academic year, that is the academic year 2020/21 when compared to the reference group, i. e. the students of the academic year 2018/19.

As the effect of COVID-19 on academic performance is assumed to be highly heterogeneous with respect to a given institution's characteristics, including the country of residence, area of focus, or the overall preparedness of lecturers, the existing research regarding this topic is inherently bound to its own research setting, such as a specific degree or university. Therefore, conducting such research intrinsically contributes as there are no similar analyses for the case of higher education in the Czech Republic.

Furthermore, the existing research lacks analysis by field or level of study, which this thesis attempts to, at least partially, address by inspecting the effect on bachelor's and master's students respectively, along with considering each of the institutes of the Faculty of Social Sciences separately. Additionally, the sample at our

disposal covers 3 pandemic-affected semesters whereas each of the cited studies is based on data from the summer semester of 2019/20 only.

The most obvious issue of this analysis is that the driving mechanism behind the effect cannot be determined as the data at hand do not make this possible. Thus, the possible explanations of the effect are merely suggestions without any supporting evidence.

Therefore, possible extension of the analysis is to study the driving mechanism behind the effect, perhaps by conducting a survey asking both students and academic staff of how they perceived distance learning under the pandemic.

Moreover, as this analysis is limited to the Faculty of Social Sciences which offers various courses, the analyzed group is still very homogenous. Assuming we were given access to data from other faculties of Charles University, the analysis could have been much more thorough as the differences among fields of study at the Faculty of Social Sciences are not as notable as when considering other faculties, such as the Faculty of Science or the Faculty of Mathematics and Physics.

References

Arellano, M. (1987). "Computing Robust Standard Errors for Within-Groups Estimators," Oxford Bulletin of Economics and Statistics, Department of Economics, University of Oxford, vol. 49(4), pages 431-434, November.

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

Balderas, A., & Caballero-Hernández, J. A. (2020, October). Analysis of learning records to detect student cheating on online exams: Case study during COVID-19 pandemic. In Eighth international conference on technological ecosystems for enhancing multiculturality (pp. 752-757).

Bench, S. W., Lench, H. C., Liew, J., Miner, K., & Flores, S. A. (2015). Gender gaps in overestimation of math performance. Sex roles, 72(11), 536-546.

Bettinger, E. P., Fox, L., Loeb, S., & Taylor, E. S. (2017). Virtual classrooms: How online college courses affect student success. American Economic Review, 107(9), 2855-75.

Breusch, T. S. (1978). Testing for autocorrelation in dynamic linear models. Australian Economic Papers, 17(31), 334-355.

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. Econometrica: Journal of the Econometric Society, 1287-1294.

Cacault, M. P., Hildebrand, C., Laurent-Lucchetti, J., & Pellizzari, M. (2021). Distance learning in higher education: evidence from a randomized experiment. Journal of the European Economic Association, 19(4), 2322-2372.

Česká televize. (2020). V Česku jsou tři lidé nakažení koronavirem. Předtím byli v Itálii. ČT24 - Nejdůvěryhodnější zpravodajský web v ČR - Česká televize. Retrieved April 22, 2022, https://ct24.ceskatelevize.cz/domaci/3056228-v-cesku-jsou-tri-lide-nakazeni-koronavirem.

Coates, D., Humphreys, B. R., Kane, J., & Vachris, M. A. (2004). "No significant distance" between face-to-face and online instruction: Evidence from principles of economics. Economics of Education Review, 23(5), 533-546.

Czech Republic Ministry of Health. (2020) Extraordinary measure. Act No. 258/2000 Coll. (MZDR 20584/2020- 1/MIN/KAN).

Econometrics, I. (2013). A Modern Approach, Jeffrey M. Wooldridge, South-Western College.

Gapinski, A. J. (2010, June). Higher education: Europe vs. USA. In Proceedings of LACCEI Conf.

Godfrey, L. G. (1978). Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. Econometrica: Journal of the Econometric Society, 1293-1301.

Gonzalez, T., De La Rubia, M. A., Hincz, K. P., Comas-Lopez, M., Subirats, L., Fort, S., & Sacha, G. M. (2020). Influence of COVID-19 confinement on students' performance in higher education. PloS one, 15(10), e0239490.

Heissel, J. (2016). The relative benefits of live versus online delivery: Evidence from virtual algebra I in North Carolina. Economics of Education Review, 53, 99-115.

Hodges, C. B., Moore, S., Lockee, B. B., Trust, T., & Bond, M. A. (2020). The difference between emergency remote teaching and online learning.

Iglesias-Pradas, S., Hernández-García, Á., Chaparro-Peláez, J., & Prieto, J. L. (2021). Emergency remote teaching and students' academic performance in higher education during the COVID-19 pandemic: A case study. Computers in Human Behavior, 119, 106713.

Karadag, E. (2021). Effect of COVID-19 pandemic on grade inflation in higher education in Turkey. Plos one, 16(8), e0256688.

Kuchler, T., & Zafar, B. (2019). Personal experiences and expectations about aggregate outcomes. The Journal of Finance, 74(5), 2491-2542.

Malmendier, U., & Nagel, S. (2016). Learning from inflation experiences. The Quarterly Journal of Economics, 131(1), 53-87.

Pascoe, M. C., Hetrick, S. E., & Parker, A. G. (2020). The impact of stress on students in secondary school and higher education. International Journal of Adolescence and Youth, 25(1), 104-112.

Rodríguez-Planas, N. (2022a). COVID-19, College Academic Performance, and the Flexible Grading Policy: A Longitudinal Analysis. Journal of Public Economics, 104606.

Rodríguez-Planas, N. (2022b). Hitting where it hurts most: COVID-19 and low-income urban college students. Economics of education review, 87, 102233.

Savage, M. J., James, R., Magistro, D., Donaldson, J., Healy, L. C., Nevill, M., & Hennis, P. J. (2020). Mental health and movement behaviour during the COVID-19 pandemic in UK university students: Prospective cohort study. Mental Health and Physical Activity, 19, 100357.

Saw, G. K., Chang, C. N., & Lomelí, U. (2020). Fall Enrollment and Delayed Graduation Among STEM Students during the COVID-19 Pandemic. The Network for Research and Evaluation in Education (NREED) Data Brief, (1).

The Faculty of Social Sciences. (2022). Karolinka FSV. https://karolinka.fsv.cuni.cz/.

Xu, D., & Jaggars, S. S. (2014). Performance gaps between online and face-to-face courses: Differences across types of students and academic subject areas. The Journal of Higher Education, 85(5), 633-659.

List of Tables

List of Figures

Appendix

Table A.1: Number of bachelor's students by field

Field	Number of students
BEF	47
BP_CNS	15
BP_KSMKP	69
BP_KSMS	67
BP_KSZN	69
BP_PMV	105
BP_PPE	41
BP_PVP	38
BP_SOCSA	19
BP_SOCSS	40
BP_SOSP	58
BP_TSTS	116
CNS	60
EF	558
MKPR	236
MTS	377
PMV	344
PVP	76
SOSA	130
SOSP	210
ZBC	196

Table A.2: Number of master's students by field

Field	Number of students	Field	Number of students
BECES	9	NP_IEPS	28
BS	151	NP_IMESSEB	14
CECS	2	NP_IMESSPE	2
CSF	11	NP_IMESSPS	3
EPS	36	NP_MAS	1
ESA	2	NP_NSS	6
GPS	31	NP_POL	25
IEPS	52	NP_SCM	7
IMESS	21	NP_TSBES	12
ISSA	153	NP_TSES	9
MAIN	57	NP_TSNRS	10
MAS	6	NP_TSSAS	12
MEF	53	NP_TSZES	7
MSP	300	NSS	19
MV	141	P	96
N_IMSISS	156	PASP	17
NEF	282	SCO	112
NMTS	132	SEC	19
NP_BECESSB	3	SKS	15
NP_BECESSR	1	VERASP	83
NP_EPS	46	ZN	263
NP_GPS	24		

Table A.3: Average GPA by semester and study programme (bachelor's)

	2016	2016.5	2017	2017.5	2018	2018.5	2019	2019.5	2020	2020.5
BEF	2.96	2.17	2.57	2.65	2.24	2.60	2.61	2.07	2.17	2.37
BP_CNS	-	-	-	-	-	-	1.68	1.61	1.44	1.32
BP_KSMKP	-	-	-	-	-	-	1.77	1.71	1.56	1.36
BP_KSMS	-	-	-	-	-	-	1.84	2.05	1.95	1.79
BP_KSZN	-	-	-	-	-	-	1.55	1.77	1.43	1.46
BP_PMV	-	-	-	-	-	-	2.46	1.71	1.95	1.70
BP_PPE	-	-	-	-	-	-	2.11	1.77	1.67	1.76
BP_PVP	-	-	-	-	-	-	2.66	2.16	1.98	1.90
BP_SOCSA	-	-	-	-	-	-	2.66	2.08	2.02	2.06
BP_SOCSS	-	-	-	-	-	-	2.55	1.99	1.72	1.95
BP_SOSP	-	-	-	-	-	-	2.59	1.90	1.69	1.98
BP_TSTS	-	-	-	-	-	-	1.88	1.97	2.00	1.84
CNS	2.17	2.21	1.71	1.79	1.63	1.75	1.53	1.45	1.33	1.55
EF	2.28	2.19	2.34	2.18	2.26	2.11	2.21	1.87	2.10	2.06
MKPR	1.54	1.52	1.57	1.73	1.75	1.75	1.70	1.53	1.42	1.47
MTS	2.23	2.34	2.46	2.44	2.23	2.20	2.15	1.87	1.82	1.62
PMV	2.48	2.27	2.54	2.07	2.28	2.03	2.22	1.74	1.71	1.66
PVP	2.58	2.81	2.48	2.15	2.38	2.43	2.25	2.03	2.13	2.24
SOSA	2.69	2.54	2.44	2.43	2.20	2.13	1.95	1.77	1.39	1.64
SOSP	3.04	2.53	2.75	2.56	2.22	2.20	1.89	1.98	1.74	1.82
ZBC	2.05	2.26	1.92	2.04	1.89	1.98	1.69	1.61	1.52	1.62

Note: Missing values indicate non-existence of a study programme at the given time. Keep in mind that the academic years 2016/17, 2017/18 and 2020/21 do not carry the whole population, thus, the average GPA for these academic years in Table A.3 may be biased.

Table A.4: Average GPA by semester and study programme (master's)

	2016	2016.5	2017	2017.5	2018	2018.5	2019	2019.5	2020	2020.5
BECES	2.19	1.33	1.27	1.50	1.75	1.82	1.08	1.00	-	-
BS	1.91	1.66	1.73	1.56	1.77	1.52	1.80	1.60	1.47	1.64
CECS	1.20	1.00	1.18	1.07	1.17	1.10	-	-	-	-
CSF	2.65	2.41	2.83	2.06	2.55	1.65	1.80	1.21	1.50	1.50
EPS	-	-	1.62	1.68	1.63	1.61	1.51	1.73	-	-
ESA	1.57	1.29	1.50	-	-	-	-	-	-	-
GPS	1.73	1.76	1.75	1.72	1.83	2.04	1.79	1.39	2.30	1.29
IEPS	2.21	1.65	1.98	1.75	1.89	1.48	1.67	1.27	1.00	1.00
IMESS	1.53	2.00	1.49	1.54	1.73	1.46	1.46	1.35	1.00	-
ISSA	1.93	1.52	1.56	1.51	1.70	1.65	1.66	1.53	1.51	2.16
MAIN	2.46	2.20	1.81	2.26	1.86	2.00	1.52	1.56	1.66	1.40
MAS	-	-	1.42	1.50	1.57	2.11	1.86	1.50	1.50	-
MEF	2.48	2.28	2.11	1.79	2.37	1.94	2.57	1.75	2.20	2.53
MSP	1.66	1.90	1.70	1.63	1.76	1.83	1.78	1.67	1.53	1.54
MV	1.74	1.72	1.63	1.59	1.77	1.65	1.71	1.64	1.69	1.50
N_IMSISS	-	-	1.45	1.30	1.33	1.42	1.25	1.46	1.27	1.31
NEF	2.17	1.86	2.05	1.88	1.96	1.82	2.09	1.56	1.84	1.74
NMTS	1.73	1.51	1.64	1.47	1.63	1.62	1.50	1.50	1.63	1.47
NP_BECESSB	-	-	-	-	-	-	1.57	1.38	1.44	1.17
NP_BECESSR	-	-	-	-	-	-	1.33	1.25	1.50	0.00
NP_EPS	-	-	-	-	-	-	1.27	1.52	1.40	1.42
NP_GPS	-	-	-	-	-	-	2.03	1.50	1.47	1.65

Note: Missing values indicate non-existence of a study programme at the given time. Keep in mind that the academic years 2016/17 and 2020/21 do not carry the whole population, thus, the average GPA for these academic years in Table A.4 may be biased.

Table A.4 (continued)

	2016	2016.5	2017	2017.5	2018	2018.5	2019	2019.5	2020	2020.5
NP_IEPS	-	-	-	-	-	-	2.24	1.64	1.79	1.73
NP_IMESSEB	-	-	-	-	-	-	2.00	2.14	1.70	1.38
NP_IMESSPE	-	-	-	-	-	-	1.90	2.34	1.90	1.00
NP_IMESSPS	-	-	-	-	-	-	2.06	1.67	1.11	1.11
NP_MAS	-	-	-	-	-	-	1.25	1.50	1.67	1.50
NP_NSS	-	-	-	-	-	-	1.19	1.21	1.09	1.25
NP_POL	-	-	-	-	-	-	2.42	1.60	2.00	1.87
NP_SCM	-	-	-	-	-	-	1.59	1.50	1.29	1.50
NP_TSBES	-	-	-	-	-	-	1.80	1.31	1.39	1.90
NP_TSES	-	-	-	-	-	-	1.60	1.39	1.33	1.29
NP_TSNRS	-	-	-	-	-	-	1.56	1.45	1.45	1.43
NP_TSSAS	-	-	-	-	-	-	1.67	1.48	1.18	1.49
NP_TSZES	-	-	-	-	-	-	1.77	1.23	1.20	1.32
NSS	1.85	1.32	1.46	1.51	1.47	1.67	1.06	1.43	1.08	1.00
P	2.53	2.15	1.95	1.81	2.11	1.89	1.83	1.99	0.00	0.00
PASP	1.58	1.95	1.70	1.57	1.64	1.64	1.71	2.44	2.10	2.11
SCO	1.65	1.73	1.47	1.65	1.59	1.34	1.57	1.48	1.40	1.62
SEC	1.63	1.86	1.35	1.64	1.41	1.69	1.80	1.84	1.96	1.50
SKS	2.75	1.00	1.42	1.46	1.84	1.27	1.54	1.42	1.00	1.50
VERASP	1.81	2.19	2.05	1.99	1.79	2.14	1.88	1.71	1.78	1.54
ZN	1.68	1.74	1.58	1.63	1.61	1.56	1.59	1.48	1.35	1.21

Note: Missing values indicate non-existence of a study programme at the given time. Keep in mind that the academic years 2016/17 and 2020/21 do not carry the whole population, thus, the average GPA for these academic years in Table A.4 may be biased.

Table A.5: Average number of semesters before graduating by field (bachelor's)

Field	
BEF	6.25
BP_PPE	4.00
CNS	4.39
EF	6.28
MKPR	5.99
MTS	6.40
PMV	6.32
PVP	6.29
SOSA	6.16
SOSP	6.21
ZBC	6.10
BP_CNS	-
BP_KSMKP	-
BP_KSMS	-
BP_KSZN	-
BP_PMV	-
BP_PVP	-
BP_SOCSA	-
BP_SOCSS	-
BP_SOSP	-
BP_TSTS	-

Note: Missing values indicate that no student has graduated at the given programme.

Table A.6: Average number of semesters before graduating by field (master's)

Field		Field	
BECES	4.00	NP_IMESSPE	4.00
BS	4.53	NP_IMESSPS	4.00
CECS	4.00	NP_NSS	3.67
CSF	4.20	NP_POL	4.00
EPS	4.04	NP_SCM	3.50
ESA	3.00	NP_TSBES	4.00
GPS	3.35	NP_TSNRS	4.50
IEPS	3.82	NP_TSSAS	4.00
IMESS	4.00	NP_TSZES	4.00
ISSA	2.97	NSS	4.38
MAIN	3.65	P	4.42
MAS	4.33	PASP	4.00
MEF	4.18	SCO	4.54
MSP	4.39	SEC	3.50
MV	4.29	SKS	4.33
N_IMSISS	3.68	VERASP	4.54
NEF	4.47	ZN	4.39
NMTS	4.60	NP_BECESSB	-
NP_EPS	4.00	NP_BECESSR	-
NP_GPS	3.83	NP_MAS	-
NP_IEPS	4.00	NP_TSES	-
NP_IMESSEB	4.00		

Note: Missing values indicate that no student has graduated at the given programme.

Figure A.1: Average GPA by semester and year of study (IIS bachelor's)

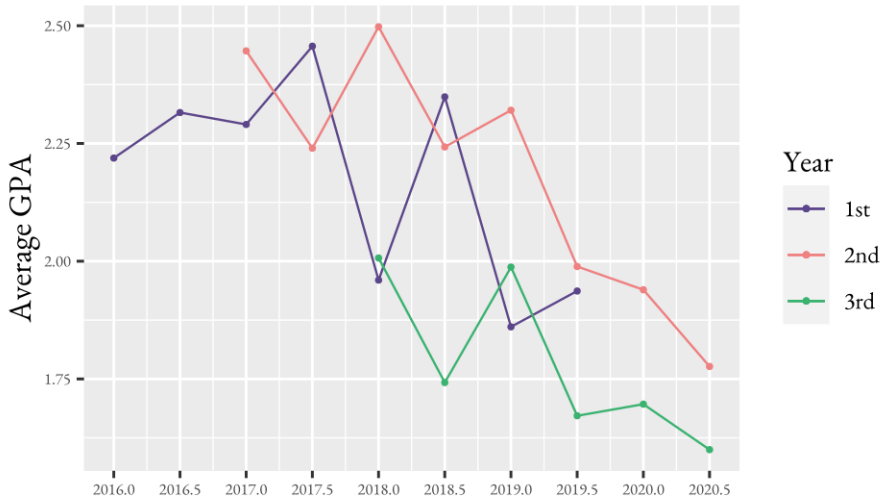


Figure A.2: Average GPA by semester and year of study (IPS bachelor's)

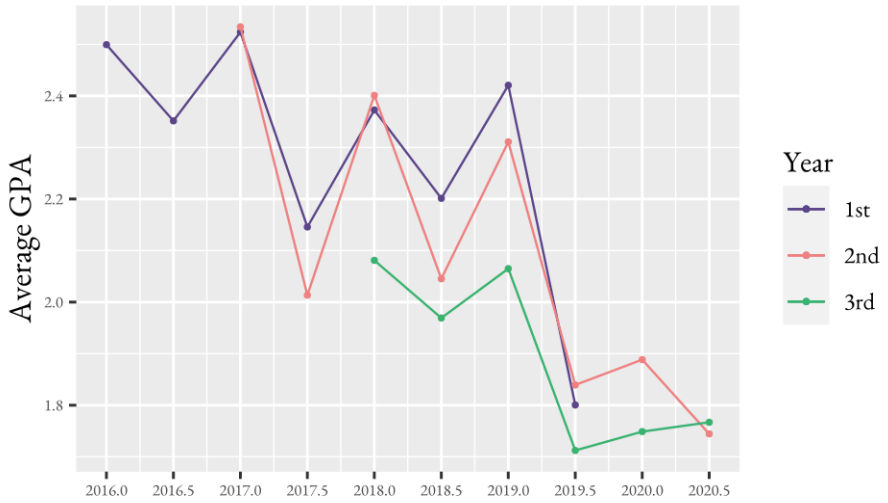


Figure A.3: Average GPA by semester and year of study (ISS bachelor's)

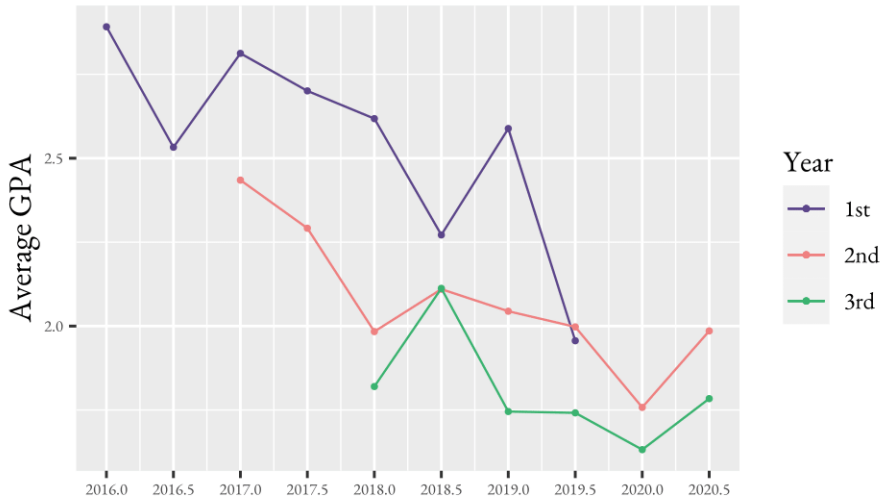


Figure A.4: Average GPA by semester and year of study (ICSJ master's)

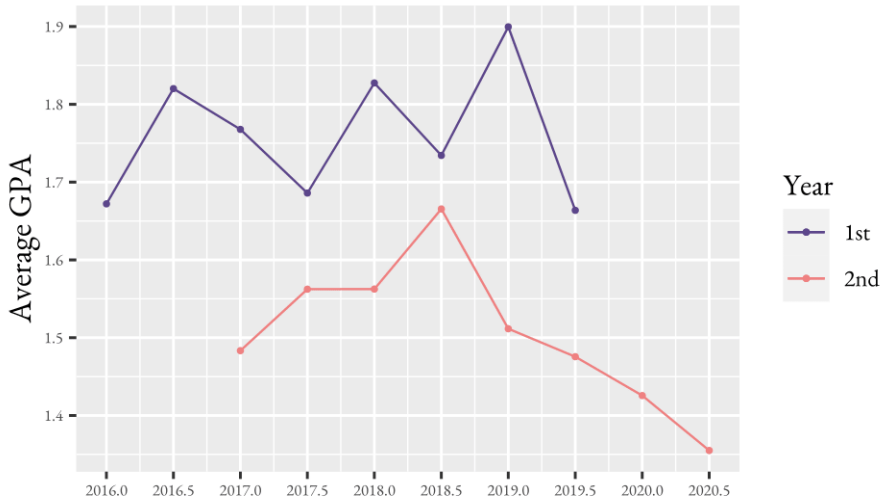


Figure A.5: Average GPA by semester and year of study (IPS master's)

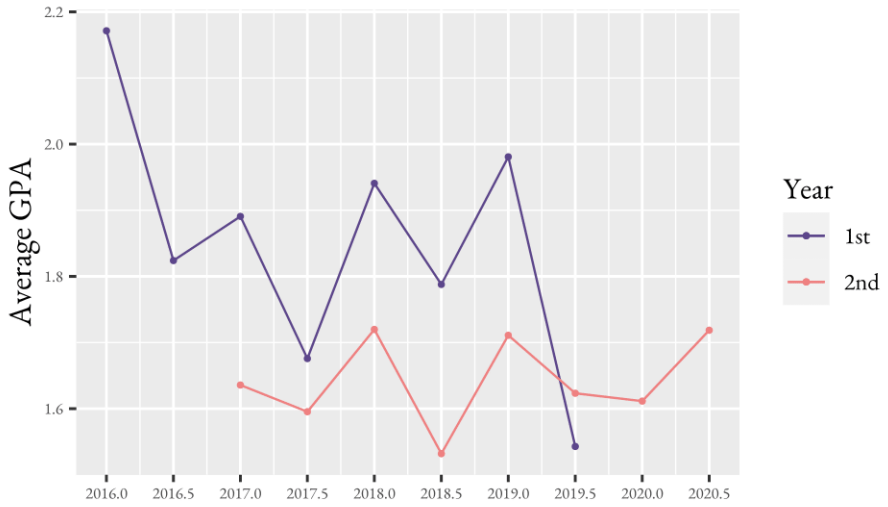


Figure A.6: Average GPA by semester and year of study (ISS master's)