It's All About What You Learn

Isolating the Human Capital Component in the Returns to Higher Education

Sonja Kovacevic*

This version: December 4, 2023

Job Market Paper

Click for most recent version

Abstract

Returns to higher education can reflect either an increase in human capital or job market signalling. In this paper, I isolate the human capital resulting from higher education, capitalizing on a distinctive feature of Norway's implementation of the Bologna reform. Students who enrolled in 1999 inadvertently acquired more human capital compared to those who enrolled in 2000, despite earning identical degrees. This anomaly, two sets of graduates with identical degrees, of which one has more education, allows me to answer a key question concerning investment in higher education. How much does an increase in human capital increase an individuals' earnings, in the absence of self-selection and signalling? I use rich longitudinal register data to estimate returns to university-acquired human capital. 15-20 years after first enrolling, receiving reform-induced human capital leads to gaining 1.5 percentile ranks in the earnings distribution. While I can rule out negative effects, the precision of the estimates does not allow me to distinguish between moderate and large effects. Using a novel 3-step cross-fit test-and-select estimator exploiting first-stage heterogeneity reduces standard errors by 16%. Human capital increases long-run earnings potential even when employers cannot observe it directly, but signalling matters when first entering the labor market.

Keywords: Education, human capital, signalling

JEL classification: I23, I25, I26, J24, J31.

*PhD Candidate, Institute of Economics, University of Oslo.

I am deeply indebted to my supervisors Jo Thori Lind and Andreas Kotsadam for their thoughtful guidance and invaluable support. I am grateful for key suggestions by Edwin Leuven and Yagan Hazard. Helpful comments and enlightening discussions with Adam Altmejd, Martin Andresen, Kai Barron, Bernt Bratsberg, Eleonora Brandimarti, Manudeep Bhuller, Ruslana Datsenko, Dirk Engelmann, Jon Fiva, Erlend Fleisje, Tilman Fries, Giulia Gupponi, Karl Harmenberg, Nathan Hancart, Ingrid Huitfeldt, Steffen Huck, Dorothea Kübler, Lars Kirkebøen, Maxwell Kellogg, Karl Ove Moene, Espen Rasmus Moen, Maria Nareklishvili, Adam Reimero, Ragnhild Schneider, Ingrid Semb, Henrik Sigstad and Fernando Stipanicic moved this project further. I thank participants at the EEA conference, UiO Seminar, BI Seminar, Frisch Centre Seminar, the UiO applied group, the WZB work in progress group and the BI-UiO PhD Initiative for helpful questions and comments. This project was initially conceived together with Karl Ingar Kittelsen Røberg. Though his professional obligations led him to concentrate on different projects, his ideas and experience with register data were invaluable inputs for this paper. All remaining errors are my own.

1 Introduction

Higher education pays handsomely in the labor market (Oreopoulos and Petronijevic, 2013). The higher education premium may reflect self-selection into higher education, returns to signalling and returns to human capital. Whether human capital or signalling drives the returns to higher education matters for education policy. Governments have more reasons to subsidize higher education when human capital plays an important role. This paper answers the question how much human capital contributes to the returns to higher education.

Separately identifying the human capital or signaling component in the returns to higher education is a long-standing empirical challenge in the field. The two most prominent theories about returns to higher education – human capital theory and signaling theory – predict higher wages for highly-educated individuals. Human capital theory, as introduced by Becker (1962) and Mincer et al. (1974), argues that higher education increases productivity by improving human abilities. Signaling theory, as formulated by Spence (1973) and Arrow et al. (1973) challenges this: high-ability individuals may signal their type through degrees in higher education, and employers pay higher wages accordingly. When job market signals and acquired skills co-vary, neither the returns to human capital nor the returns to signals can be identified. University degrees, which certify higher education, entangle the variation in students' acquired human capital with the signalling effect of their degrees in the job market. Studies that leverage exogenous variation in human capital remove students' signaling intentions. Yet, employers observing the degrees they earn as a result of induced human capital may still consider the degree as a signal. To isolate human capital effects, additional human capital should be exogenously induced while not resulting in higher-level degrees. The Bologna reform in Norway presents such a case.

The 2003 Bologna reform was a Europe-wide harmonization of higher education systems. Transition cohorts experienced as-good-as-random encouragement of different length of education while earning the same distribution of degrees. In Norway, the 4-year undergraduate degree was shortened to a 3-year bachelor. In 2003, accordingly, students who had enrolled in 1999 were about to earn the "old" undergraduate degree; they had accumulated 4 years' worth of study credits. But the Bologna reform was introduced with retroactive effect. The 1999 cohort was thus given the new bachelor's degree. Students enrolling in 2000 were likewise planning to earn the old undergraduate degree. By 2003, they had accumulated only 3 years' worth of study credits - which, with the reform, also sufficed for a bachelor's degree.

A comparison between the 1999 and 2000 enrollment cohort identifies variation in human capital. The 1999 cohort spent up to one year longer at the university and accumulated more study credits; its level of human capital is higher. Neither cohort knew about the future Bologna reform, which allows me to rule out self-selection into cohort. Both cohort had wanted, and eventually received, the same degrees. Neither their choices nor their final bachelor's or master's degree differ. If there are differences between these two cohorts' later earnings, they would, plausibly, derive not from self-selection nor signalling, but from returns to human capital. Comparing cohorts' earnings in calendar time captures the double advantage of the 1999 cohort; not only do they receive more education, they also have an additional year to build a career. In any calendar year, 1999 cohort students are on average one year older and had one additional year to complete their studies or gather experience.

In the main specification, I account for the advantage of earlier enrollment using the 2001-2002 cohort pair. Unlike the 1999-2000 cohort pair, students in 2001-2002 cohorts do not differ in how long they took to complete their studies. All four cohorts obtained the same degree types – the new Bologna reform's bachelor's and master's degrees. Consequently, I can employ the 2001-2002 cohort as a placebo group and estimate a quasi-difference-in-difference design. This involves measuring earnings in calendar time for each cohort pair but estimating differences relative to the enrollment year of the older cohort within each pair. In an alternative specification, I estimate a 2SLS in an event study fashion. Instrumenting study duration with enrollment cohort, I measure outcomes of the 1999 and 2000 cohort in time since enrollment, rather than in calendar time.

During the first 10 years after enrollment, earnings are not significantly affected by the amount of university-acquired human capital. Pre-enrollment, not-yet-acquired human capital does not predict earnings, underscoring that selection is not a concern. While studying, earnings remain comparable, suggesting that partially acquired human capital does not significantly impact part-time employment income, as pursued by students while studying. Around the typical graduation period, returns on university-acquired human capital tend to be marginally negative. This reflects the opportunity costs associated with prolonged university education compared to entering the job market sooner. Approximately a decade after enrollment, when about 80% of the sample has completed their studies, returns on university-acquired human capital become positive.

In 15 to 20 years post-enrollment, being part of a cohort that receives, on average, an additional 3.5 months of education results in a 1.5 percentile-rank increase in earnings. The confidence intervals' width suggests potential gains ranging from as low as 0.25 to as high as 2.7 ranks. While this allows me to rule out negative effects of university-acquired human capital on earnings, the upper bound lacks precision. The same qualitative pattern of earnings over time emerges using the event study 2SLS specification, with long-run gains of 2.6 ranks per additional year of education, albeit non-significant at conventional levels. The main specification estimates opportunity-cost adjusted returns to acquiring extra human capital, where I do not control for experience (less experience due to longer study time is part of the causal effect). However, controlling for experience is fairly common in the literature (Arteaga, 2018; Carneiro et al., 2011) and identifies the partial effect of more education under very strong assumptions. Estimating this benchmark increases estimates by 10-20% relative to the baseline specification, and returns to human capital become significant earlier in a students' or graduates' career.

Overall, these findings suggest that university-acquired human capital pays off even when candidates do not get a degree that shows their extra studies. While signalling may still play a part, employers appear to learn about and eventually compensate for non-certified skills. This is all the more remarkable because the majority of students in the sample earn a master's degree in the end, masking potential differences in their course of study for their bachelor's. In contrast to studies where students receive certified extra education, it takes a lot longer for earnings benefits to materialize, in line with the findings of Arcidiacono et al. (2010).

This paper contributes to four literatures. Firstly, several studies attempt to separately estimate either the signaling or the human capital component in returns to education. Arteaga (2018) estimates the human capital component in returns to the economics and business degree at an elite university in Colombia. I expand on this analysis by incorporating a large number different degrees from different universities, exploring how returns to university-acquired education evolve over time and offering an empirical rather than theoretical approach to self-selection. Adapting my specification to Arteaga's strategy, estimating partial returns and using log earnings as the outcome, I obtain comparable albeit somewhat lower estimates. Targeting the signaling component, Clark and Martorell (2014) find little evidence for high school diplomas acting as a signal, while Tyler (2003) finds some signaling returns to the GED. Both studies were designed to eliminate the variation in human capital to isolate the signaling effect of educational credentials. My research serves as a complement to these studies, isolating instead the returns to human capital. Bedard (2001) makes a theoretical argument that limited access to universities increases pooling at the high school level in a signaling model and measures educational choices in the data. By focusing on higher education rather than schooling, I illuminate further the relation of human capital and signaling in an educational area with more individual choice.

Secondly, there is an immense body of literature estimating the returns to higher education. I contribute to this literature by investigating what drives these returns, focusing on the human capital mechanism. Many studies in this literature estimate the returns to a specific certified unit of education. Using distance to college (Carneiro et al., 2011; Brandimarti, 2023), admission thresholds (Kirkeboen et al., 2016; Hastings et al., 2013; Bleemer and Mehta, 2022), or lotteries (Ketel et al., 2016) as an instrument or at a discontinuity, these studies estimate the returns to attaining specific degrees, levels of education, or completing different fields of study. The estimated returns may reflect university-acquired human capital, employers' perceptions about the costs of completing of a particular course (signal), monopoly or licensing rents for some occupations and the match between a candidate and their program of choices. While taking care of selection, these studies identify joint variation in human capital and acquired degrees. Isolating the university-acquired human capital component in the returns to higher education enriches the findings in this literature by shedding light on the mechanisms that reward higher education.

Thirdly, my works contributes to the literature on employer learning. Farber and Gibbons (1996) and Altonji and Pierret (2001) argue that employers use schooling to form expectations about candidates productivity. If employers learn the candidates true productivity over time, the signaling contribution to the returns to schooling should diminish. Lange (2007) uses their framework, observing that returns to signaling depend on the speed of employer learning, possibly asymmetrically (Schönberg, 2007). This literature makes strong assumptions on workers discount factor and employers speed of learning and estimated bounds on the return to signalling depend on these parameters. Recent work by Aryal et al. (2022) relaxes this assumption, deriving learning speed from the difference in IV estimates to the returns to schooling when the employer have information about whether schooling was chosen or induced. This literature shows that employers learn quickly about candidate ability, leading to limited returns to signaling. In this study, there are no returns to signaling because degrees do not differ. However, I find that employers learn slowly about non-certified human capital. This result, when considered in the context of the employer learning literature, suggests that signaling returns might decrease faster than currently estimated. This is because the returns on human capital are not immediate but unfold over time, while the employer learning literature assumes that employers observe human capital from schooling instantaneously.

Finally, there is a growing literature on dealing with low compliance, and the resulting problems with weak instruments and the variance of the 2SLS estimator (Abadie et al., 2023; Hazard and Löwe, 2022; Coussens and Spiess, 2021). Students in this study progress at different rates, making some too fast or too slow to be affected by the Bologna reform. This leads to moderate overall adherence to the reform. While simple testing on the strength of the first stage is practiced in applied work (Stephens and Yang, 2014), this introduces bias in the 2SLS estimates and invalidates inference. Interacting the instrument with the covariates that predict compliance gives consistent estimates of a weighted average of LATEs but puts disproportionately large weights on subgroups with high compliance (Huntington-Klein, 2020; Abadie et al., 2023). Hazard and Löwe (2022), propose a novel 2-step cross-fit test-and-select estimator, where first-stage testing and estimation are performed on different partitions of the data, each grouped by covariates. I use a modified version of this approach with a 3-step cross-fit procedure, suited for cases where the covariates determining compliance are not grouped in categories. In my case, students' study speed is continuous, with theoretical reasons to expect low compliance in both tails of the distribution. I optimize cut-offs with respect to different measures of study speed on a separate partition of the data before applying the test-and-select estimator. Using this procedure reduces standard errors by 16%.

The paper, below, is organized as follows. Section 2 describes the Bologna reform and how it affected transition cohorts' length of education and degree types. Section 3 constructs the baseline sample, targeting students that may experience the Bologna reform. I show balancing of the treated cohorts and argue why the timing of enrollment and reform announcement made it impossible for students to select into treatment. Section 4 lays out the empirical strategy; this includes defining the parameter of interest, introducing the placebo quasi difference-in-difference main specification, and applying a cross-fit test-and-select method to improve efficiency for 2SLS estimation. Section 5 reports estimates for the short- and long-run returns to university-acquired human capital. Section 6 extends to partial returns, addresses some students receiving the old degree types through estimating their counterfactual earnings using non-parametric bounds and discusses monotonicity concerns. Section 7 consists of my conclusion.

2 Background

In 1999, 29 European nations initiated the Bologna Process, a comprehensive plan to standardize and unify higher education across Europe. They aimed to align university curricula, credit systems, and degree structures to ensure comparability. The goal was to enable students to transfer seamlessly among European universities and to ensure that graduates' degrees are recognized across Europe's job market without question. All participating countries committed to introduce bachelor's and master's degrees as the principal academic qualifications.

When Norway adopted the bachelor-master system, this changed the duration of higher education cycles in Norway. Before the reform, students in Norway could enroll in a *cand.mag.*[1], the undergraduate degree and then continue with a *hovedfag*[2], the postgraduate degree. The *cand.mag.* was expected to last for 4 years in the social sciences and humanities, and 3.5 years in the natural sciences. When adopting the Bologna reform, the bachelor replaced the *cand.mag.* degree. This reduced the expected duration of an undergraduate degree from 4 - or 3.5 years in the case of natural sciences - to 3 years. Similarly, the master replaced the *hovedfag*, the postgraduate degree. Both the previous and the new postgraduate degree programs were standardized to a 2-year duration of study, resulting in no change to the duration of postgraduate studies. To sum up, Norway replaced a 4+2 higher education system with a 3+2 higher education system.

Despite the Bologna reform, several study programs remained unaffected. To begin with, Norwegian degrees in fields like nursing and engineering had already adopted a 3+2 structure prior to the Bologna reform. Additionally, professional degrees, such as medicine and psychology, retained their original study structure. In total, approximately 50 academic degrees were replaced due to the reform. Affected degrees were taught at one of the four large universities, while degrees at the community colleges[3] remained largely unaffected. For this reason, I restrict the sample to university students.

Norway implemented the Bologna reform rigorously and wholeheartedly. While some countries delayed introducing the new degrees and granted numerous exceptions, Norway followed the Bologna declaration to the letter. Starting with the implementation in the fall semester of 2003, students faced a completely overhauled curriculum, with changes both in didactic set-up, course structure, and exam frequency. Even students that enrolled before the reform in 2003 were subject to the reform. No matter when students had initially enrolled and which old degree they might have qualified for, from July 2003 all Norwegian

[1]Short for candidatus magisterii/candidata magisterii. Available since 1959, initially directed towards teachers, this became the necessary undergraduate degree before beginning with a higher level degree. This degree could be earned in the humanities, social sciences, natural sciences and even across faculties.

[2]Umbrella term for postgraduate degrees, including cand.philol. for languages, cand.polit., cand.sociol., og cand.oecon. in the social sciences and cand.scient. for natural sciences

[3]The Norwegian higher education system consists of universities and "høykoler". These could be translated as regional or community colleges. They typically offer degrees in a particular direction, technical or social, and are more applied and less research-oriented than universities. The bulk of courses taught at community colleges are undergraduate degrees but they also offer some graduate degrees. Pre-Bologna, many degrees offered at the community colleges were standardized to two or three years of study. Although some few degrees were reformed in the same way as university degree, they constitute a negligibly small share.

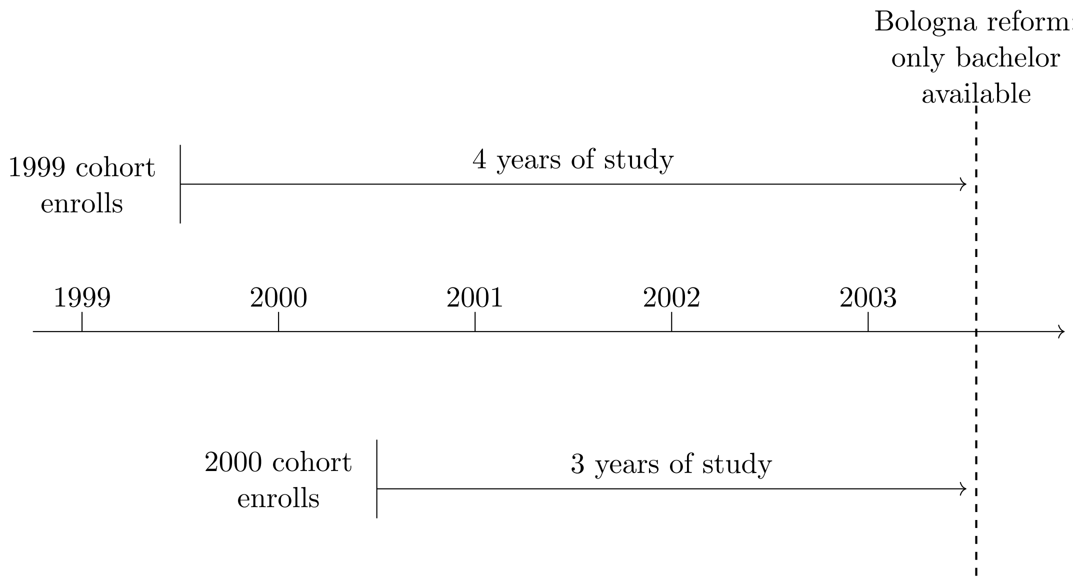


Figure 1: Stylized effect of the Bologna reform on study time across cohorts

universities awarded the new degrees as a general rule. I concentrate exclusively on the retroactive implementation of the Bologna reform; this consitutes the centerpiece of my identification strategy. Examining pre-Bologna enrollment cohorts allows me to isolate the returns to induced study duration, without interfering effects from curriculum changes and study practices that took effect for cohorts enrolling after the Bologna reform.

When Norwegian universities enacted the Bologna reform retroactively, they generated exogenous variation in the amount of human capital for already enrolled students. Students who enrolled in the years before the Bologna reform planned to earn a *cand.mag.* degree, possibly followed by a *hovedfag* degree. I first describe the stylized case where students progress at the rate educational programs were normed (60 study credits/year), see Figure 1, before turning to factual progression rates and their consequences for exposure to the reform. The *cand.mag.* degree was normed to 4 years of study. Students enrolling in 1999 were about to earn their *cand.mag.* in 2003 and had accumulated 4 years' worth of study credits. Because of the retroactive implementation of the Bologna reform, they did not receive their *cand.mag.* but were instead awarded a bachelor degree if they had collected sufficient credits from July 2003 and onwards. Students enrolling in 2000 were likewise planning to earn a *cand.mag.* degree. By July 2003, they had accumulated 3 years' worth of study credits. Since a bachelor requires three years' of study credits, they had no reason to collect additional study credits.

Comparing students from the 1999 enrollment cohort and 2000 enrollment cohort identify exogenous variation in human capital. The 1999 cohort spends up to one year longer at the university and accumulates more study credits, acquiring a higher level of human capital. With enrollment into the same programs long before the Bologna reform and the same degrees (signals) available at graduation, neither their choices nor final degree contain any information that could serve as a signal to employers. If employers reward this extra education in the labor market, it must be because the 1999 students are more productive due to the extra courses they took and not because they signal ability through their choices.

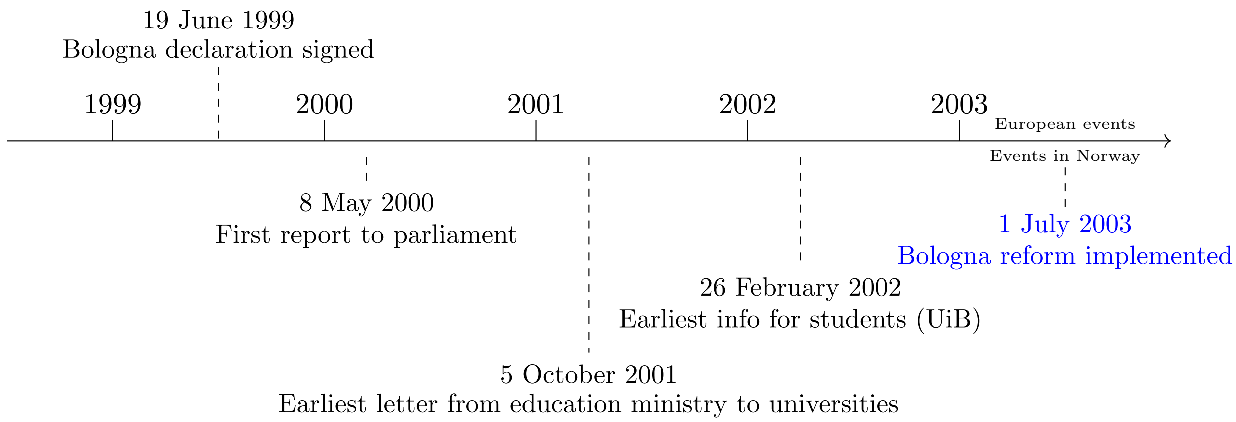


Figure 2: Timeline of the Bologna reform and it's implementation

Until February 2002, students were unaware of the Bologna process, see Figure 2 above, and even then, information was incomplete. This absence of information rules out selection into the 1999 and 2000 cohort and minimizes opportunities for students to change course choices in response to the reform. European states signed the Bologna declaration in 1999. At this stage, Norwegian authorities had not decided how to transform the general intentions stated in the Bologna declaration into national law. In October 2001, the Ministry of Education notified the universities that a reform of the higher education system was to come, but important details such as the introduction date were missing, see Quotes2 and 2 below. As late as 2002 – when the University of Bergen took up the issue with their students – the reform date was not decided. The following exchange between a student and the university administration is the earliest official communication about the reform between any university and their students. It shows considerable uncertainty around the introduction date of the reform, even for students that enrolled after 1999 and 2000:

> Question: *What will happen to students who were admitted to the University of Bergen in the fall semester 2001, and who decided to enroll in exphil/exfac. [shared introductory course covering philosophy and research ethics] in the first semester? What degree*

> *will they receive? Are they cand.mag. students or are they bachelor students? Can they choose their degree?*
>
> Answer: *The letter on the 05.10.01 from KUF [the Ministry of Education and Research] states: "All those who satisfied the requirements for the cand.mag. degree up to the time of introduction shall have this degree. Everyone who satisfies the requirements for the bachelor's degree from the time of introduction will become a bachelor. There won't be the opportunity to receive both a cand.mag. and a bachelor degree. Students who have completed three years' worth of studies without fulfilling a bachelor degree will be allowed to fulfil a cand.mag. in a transition time of up to two years." The essential question becomes: When is the introduction date? For the time being, students who enrolled [in the fall 2001] should be regarded as cand.mag. students, but the moment the introduction date is set, the students must relate to the above.*
>
> Åsheim (2002). Published on the website of the University of Bergen on 26 February 2002. Own translation.

The uncertainty reflected in the quote is not limited to university-student communication; rather, the introduction date was genuinely undecided at this point. As late as March 2002, communication between the different government branches and the parliament suggest that fall 2004 was considered the most likely introduction date. The introduction in fall 2003 came earlier than anticipated and as a change of schedule. Even if students were following government-parliament communication, they would expect fall 2004 rather than fall 2003 as the introduction date.

> *The studies recommended here for a reduced duration should adopt a new degree structure [...] by the fall of 2004 at the latest [...]. This timeline has been established to ensure that the institutions have ample time to implement the necessary adjustments associated with the reduction in study time.*
>
> Ministry of Edcuation (2002). Published on the website of the Ministry of Education on 18 March 2002. Own translation.

Students who enrolled in the 1999 and 2000 cohorts – they enrolled earlier than the 2001 cohort discussed above – could not have based their enrollment date on the Bologna reform, since the date of this reform was not even decided until three (two) years after they enrolled. At most, 1999 students could have decided to slow down their studies early in 2003, when they learned that they may not need four years' worth of studies to graduate. This type of behaviour would reduce treatment intensity, weakening the effect of enrollment cohort on accumulated human capital. 1999 students would receive less extra education vis-à-vis the 2000 cohort if they slowed down compared to the case where they continued at previous study speed - but would not cause selection bias.

Students' exposure to the reform depends on how quickly they progress in their studies; I use study credits to assess study progress. Pre-Bologna, 80 Norwegian study credits were required to earn a *cand.mag.* in the social sciences and humanities. Implementing the Bologna reform, Norway converted the Norwegian credits to European Credit Transfer and Accumulation System (ECTS) credits at a rate of 1 to 3. Expressed in ECTS credits, a *cand.mag.* was worth 240 credits for social sciences and humanities and 210 credits for the natural sciences. The new undergraduate degree that replaced the *cand.mag.* was the bachelor; to earn it a student had to collect 180 study credits. I define study speed as the number of (ECTS) study credits earned per year of study. 60 credits per year correspond to the government-set normed study used by universities to plan study programes, both pre- and post-Bologna. A student in the social sciences or humanities studying at this normed speed would collect 240 credits in four years and earn 60 credits more than required for a bachelor at the time of the reform introduction. These 60 extra credits represent the maximum expected treatment for a student from the social sciences or humanities. Quicker students will not be treated more, because 240 credits suffice for the old degree, self-limiting the maximum extra education. Students from the natural sciences have both a lower likelihood to be treated and a lower treatment intensity if they are treated. A student studying at normed speed in the natural sciences will fulfil the requirements for the *cand.mag.* already in January 2003, earning a *cand.mag.* before the reform is implemented. A student studying slightly slower may not finish before the reform, but will at most collect 30 credits too much. If they had collected even more, they could have graduated before the reform. Students in the natural sciences should be less affected by the reform, and will control for being a science student in all specifications.

Students progress in their studies at different rates; this distribution of study speed determines which students are exposed to the reform. I proxy study speed by calculating the average number of credits per year during the 5-semester pre-screening period. This measure is imperfect, because students may speed up or slow down later, and because there is under-reporting in the number of study credits. Using this low-biased measure as a proxy for study speed, 85% of students study slower than normed. By 2003, all 1999 students with an average actual study speed between 46 and 60 credits per year should have accumulated extra credits. Students in the lower end of that bracket will earn fewer extra credits than students in the upper end. Corresponding 2000 students will have earned less than 180 credits. They simply continue their studies until they reach 180 credits and earn a bachelor without any extra credits.[4] Using the average study speed in the first five semesters of study as a proxy for study speed – computed before sufficiently accurate information about the reform emerges – 30.1% percent of students in both cohorts fall into 46 to 60 study credit bracket that should collect extra credits. Taking under-reporting of study credits into account, this group is likely larger. These students represent the expected complier group.

Note that this is only the group of expected-treated students. Initially quicker students slowing down or initially slower students speeding up may also end up being treated, while students initially in this bracket may change their speed and end up non-treated.

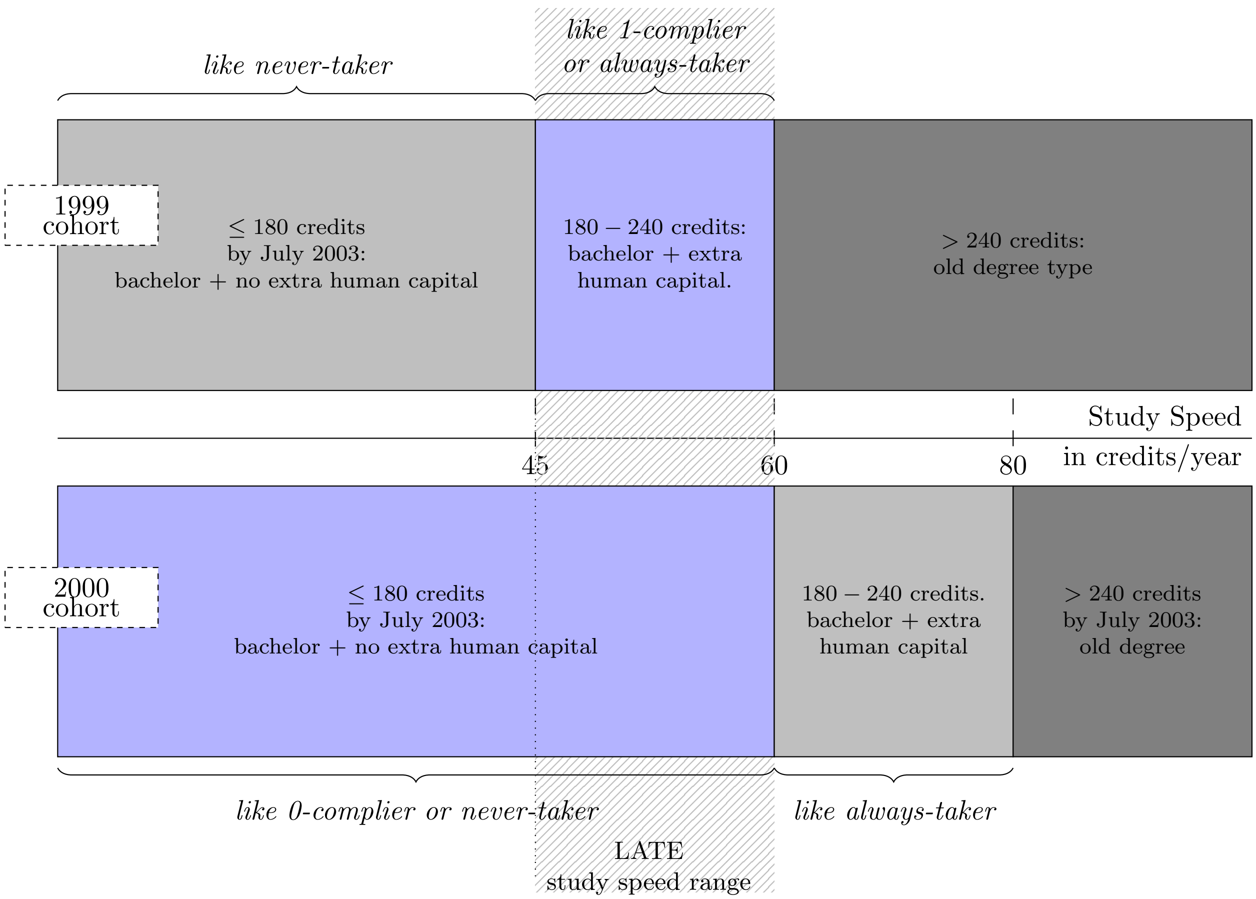


Figure 3: Expected reform exposure by study speed

1999 students who keep progressing at less than 45 credits per year will not collect any overshooting credits by 2003 which makes them expected never-takers.[5] Students in the 2000 cohort who keep progressing at more than 60 credits per year are expected always-takers.

[4] In principle, students can always choose to take more subjects. 2000 cohort students collecting extra credits in their bachelor are like always-takers.

[5] Note that the classical definition of never-takers, always-takers and compliers does not apply since the

See Figure 3 below for an overview of the expected reform exposure by study speed bracket. I use cohort membership - a binary variable for either enrolling in the 1999 and 2000 cohort - as an instrument for human capital, measured as total study length (after screening) in the main specification. See section 4.4 for first stage estimates and a detailed discussion of the induced variation in human capital across cohorts. In section 4.5 I build on the relation between study speed and reform exposure to improve efficiency, using a 3-cross-fit test-and-select procedure.

Some students that study quicker than norm study speed may obtain the old degree types; this is a threat to my design. 1999 students that progress at a rate of 60+ credits per year will have earned a *cand.mag.* by the time the reform is introduced. 2000 students will need to study at 80+ credits per year to receive the old degree before the reform is announced. For this reason, the sample is imbalanced with respect to students receiving the old degree types across cohorts. Students in the speed bracket from 60 to 79 credits per year are expected to earn a *cand.mag.* degree if in the 1999 cohort, but a bachelor degree (with overshooting credits) if in the 2000 cohort. Employers may consider *cand.mag.* degrees as a signal, even if students in the two enrollment cohorts were identical in planning to earn a *cand.mag.* degree initially. The practical importance of this problem is limited, since only 2% in the 1999 cohort and 0.3% in the 2000 cohort earn *cand.mag.* degrees. Theoretically, the problem amounts to a violation of the exclusion restriction, since cohorts differ both in human capital and degree composition, with degrees likely having a direct effect on earnings. I address imbalance in the degree composition by estimating non-parametric bounds on the counterfactual earnings of *cand.mag.* degree holders in section 6.2.

3 Data and baseline sample

3.1 Data

I use several registry databases collected by Statistics Norway. The course registry, containing information on all courses taken by all students registered at any education institution, identifies when, where, and how long students study and which degree they eventually earn. Checking which courses students take reveals whether a student pursues a program subject to the Bologna reform. I match this information with demographic data from the population registry which lets me match students to their families, allowing to construct de-

amount of human capital - measured as study length or study credits - is a continuous variable. Students can collect as many more credits in the future as they wish in the future, so that even students that are not pushed into reform-induced overshooting credits in 2003 may collect extra credits. However, 1999 students in the 45 to 59 study credit bracket, if maintained during the undergraduate degree, will collect *too many* credits.

mographic background variables such as parental education. Additionally, I use information from the tax register; this contains employment status, wages, earnings, post-tax earnings and hours worked for students from 5 years prior to enrollment until 2021, up to 21 (22) years after enrollment.

3.2 Baseline sample

I focus on the 1999 and 2000 enrollment cohorts and choose students that study in Bologna-relevant subjects. Initially, there is no reason to expect differences in ability or other earnings-relevant traits for students in the same broad fields on a year-by-year basis. Although some student characteristics such as parental education or number of siblings change somewhat over time, this happens over an extended time span, so that differences from one cohort to the next are negligible. Potentially, enrollment cohorts could also differ because the time lag from finishing high school to enrolling changes. The end of compulsory military service as well as the expansion of exam-free pre-college (Folkehøyskole) could change cohort composition, but neither event falls into the relevant enrollment years. Therefore, Bologna-subject enrollment cohorts constitute a base sample with comparable characteristics before the reform.[6]

Individuals from the 1975-1981 birth cohorts were most likely to end up in the pre-Bologna enrollment cohorts 1999 and 2000.[7] Between 50% (1975) and 52% (1981) percent of people in these birth cohorts enroll in higher education at some point after finishing school.

Within each enrollment cohort, I select students that enroll in a course leading to a Bologna-affected degree within their first five semesters of study. An enrollment cohort follows the academic year starting in August in one year and ending in July the next year; so a student in the 1999 enrollment cohort can enroll anytime between August 1999 and July 2000. In practice, most students - 81 percent - enroll in the start of the fall semester. The remaining 19 percent enroll in the spring semester. Given that these students enroll slightly later, they would need to study more quickly in order to accumulate more than three years' worth of study by the time the Bologna reform is implemented. With most students on the

[6] Targeting birth cohorts may seem like a natural choice since birth cohorts are even less likely to suffer from selection bias. Using full birth cohorts would include students in programs not affected by the Bologna reform and even people not enrolling in higher education at all, making the birth cohort instrument too weak. In addition, Norway had compulsory military service at that time, so that men enroll systematically later than women relative to their birth cohort. Using birth cohort as an instrument for more higher education would thus pick up sex differences in enrollment age as well. Focusing on enrollment cohorts avoids this effect.

[7] Most students are 19 (42%) or 20 years (23%) when they first enrol, with 7.5% starting at age 25 or older and 4.9% starting at age 30 or older. Students that enroll later in life will be from much earlier birth cohorts, while no student in the sample is born later than 1986.

slow side, spring semester students should have a lower treatment intensity and represent a different complier group. Interacting the main cohort instrument with whether a student starts their studies in spring allows these different groups to differ in their first stages.

The raw 1999 and 2000 enrollment cohorts have 44 410 and 44 823 students respectively. First, I exclude the roughly 76 percent of students enrolled at community colleges. These colleges offered 3-year undergraduate degrees such as nursing and were thus unaffected by the Bologna reform. Some of the undergraduate degrees at the community colleges were even extended in length, making it paramount to exclude community college students to avoid defiers. Removing students enrolling in community colleges leaves university enrollment cohorts of 10 593 and 10 701 students. However, almost 40 % of these later switch to community colleges. With pre-treatment sample construction, these remain in the sample. Second, Norway exempted several professional degrees such as medicine, dentistry and theology from the Bologna reform. I exclude the 31% of students enrolled in professional degrees from the university enrollment cohorts. Finally, I impose conditions that help to identify whether students start studying (explained in detail below). The resulting sample consists of 3598 students in the 1999 cohort and 3521 students in the 2000 cohort.

Pre-screening for Bologna students lasts for five semesters, see Figure 4. For the later-enrolling 2000 cohort, pre-screening ends one semester before the Bologna reform is implemented. Pre-Bologna, students did not enroll into specified study programs but rather chose which subjects to attend; subjects within broad academic fields could be combined into one degree fairly liberally, and even degrees cutting across academic fields were possible. An example for a subject leading to a Bologna degree would be "Introduction to scientific methods for the social sciences", with a workload of half a semester at the Social Sciences Faculty. An example for a course in a professional degree not leading to a Bologna degree would be "Pediatric diseases" at the Medicine Faculty.

In order to exclude all university students planning to earn a professional degrees, I exclude a student even if they take a single course belong to a professional degree study program. Completing a Bologna-related course provides less information about the student's intended program compared to taking a course in a professional degree. Many of the professional degrees were heavily oversubscribed and some students take courses in other subjects while waiting for admission in their desired program. For instance, if a student completes an introductory sociology course, it could indicate either waiting for a spot in medicine or a genuine interest in sociology. On the other hand, observing a student completing a medicine course is a strong indicator that they are on track for a medical degree, as students usually

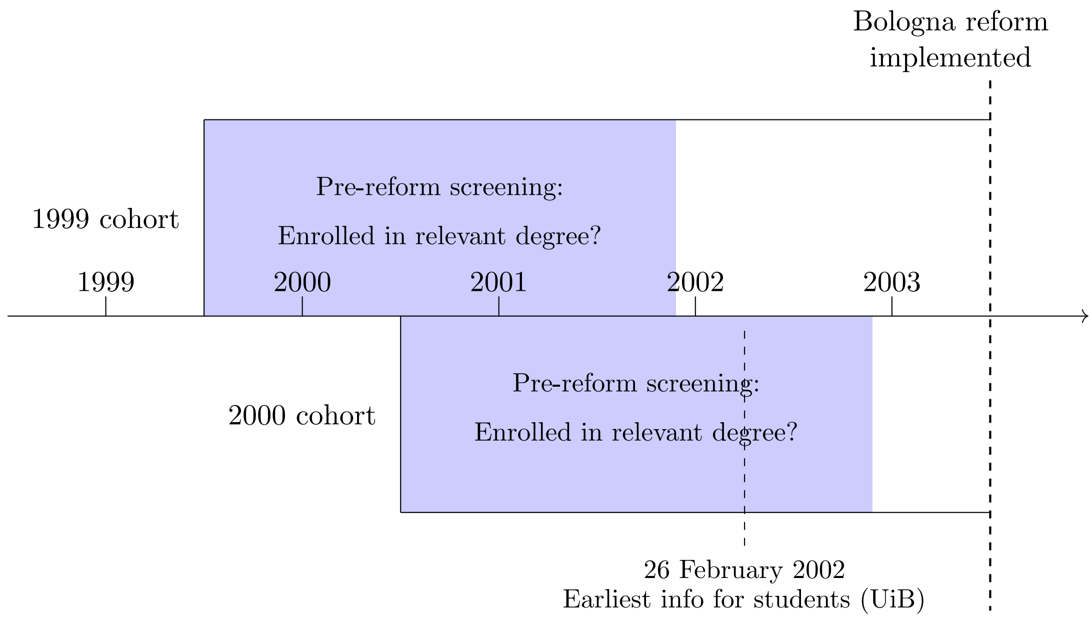


Figure 4: Pre-screening period for sample selection

follow through with professional degrees once enrolled.

Some people that enroll do not study. Students in Norway receive considerable discount for public transport and other services, giving an incentive to enrol without intending to study. Students that never finish a degree should not be excluded, since some students may try to study without succeeding and could still be affected by the reform in that process. But I can run a minimal check on whether people attempt to study in the pre-screening period. I consider students that do not collect a single study credit at a university during the first five semesters as a non-student, this is the case for 20% of students that enroll. 78% of those never complete a degree at a university, but a large fraction later completes a degree at a community college. Similarly, I exclude students that stop being enrolled during the screening period. Since they quit long before the reform is implemented, dropping out should not be a consequence of the reform. Some of these students are early dropouts that never obtain any higher education degree. Another group are students that finish a higher level degree in Norway very quickly, after having obtained an undergraduate degree abroad.

These selection rules produce a sample with of a total of 7119 individuals, with 3521 students in the 1999 cohort and 3598 students in the 2000 cohort. 45% of these students eventually graduate from a Bologna-affected degree, while 7% graduate from a non-Bologna degree at the universities, 40% graduate from a regional college and 6% never earn any degree.[8] The single most common degree type in the sample is the master's degree, with roughly 1000 students in each cohort earning a master's compared to roughly 300 in each cohort that earn a bachelor's. Master students are genuine Bologna students in the sense that they (almost always) first earn a bachelor and experience the Bologna reform while earning their bachelor. Concerns that Bologna-induced study credits on bachelor's transcripts may act as a minor signal to employers are lessened when individuals pursue higher-level degrees. For students with a master degree, the master likely serves as the main signal, reducing attention to course level signals at the bachelor level.

Table 1 below shows that cohorts are well balanced on observables, indicating that excluding community college students, screening out professional degrees, and excluding non-students does not select systematically different portions of the full enrollment cohort. The timing of enrollment within an enrollment cohort matters for how students are affected by the reform. Spring semester students enroll on average half a year later than fall semester students in their respective enrollment year. For this reason, they have less time to accumulate study credits before the reform is introduced, so that relatively quicker spring students will be affected by the reform, relative to the affected group among the fall students. Since the average student is on the slow side of studying, this also means that the first stage for spring semester starters is lower. When estimating the first stage, I interact cohort with spring semester start in the first stage to capture that the reform affects different sub-groups of students in the spring semester compared to the fall semester. In both enrollment cohorts, approximately the same proportion of students enrolls in the spring semester. While fall and spring semester students differ in their characteristics, they differ in the same way across cohorts. Spring semester students in the 1999 and 2000 enrollment cohort are just as indistinguishable on observables as the full enrollment cohorts, see Table A.1 in the appendix.

[8] This captures the last degree earned. Some students from the community/regional colleges may still be affected if they first earned a bachelor at a university, got delayed due to the reform and then completed a masters at a community college. I discuss the occasional presence of old degree types in detail in section 6.2 with regard to signaling and the distribution of degree levels in 4.3 with regard to monotonicity.

Variable	Mean 2000	SD 2000	Mean 1999	SD 1999	p-value
Mothers' education	4.6	1.72	4.55	1.72	0.29
Fathers' education	5.0	4.96	5.0	4.96	0.25
Female	0.64	0.48	0.63	0.48	0.49
Nr of siblings	2.14	0.88	2.15	0.87	0.68
Age at enrollment	20.81	4.33	20.69	4.12	0.22
Natural science student	0.14	0.35	0.13	0.33	0.15
Enrolled in spring semester	0.19	0.39	0.18	0.39	0.86
N	3598		3521		

Table 1: Balancing of observables across enrollment cohorts

4 Empirical strategy

4.1 Defining the parameter of interest

I am interested in the returns to acquiring more human capital at university. In order to identify the returns to university-acquired human capital from the data, other factors that typically affect earnings jointly with human capital must be ruled out. Usually, earnings reflect the returns to human capital, the returns to a degree (signalling), the returns to unobserved determinants of productivity such as ability and motivation that correlate with the choice to enroll in higher education (selection) and, in a given calendar year, how much time has passed since enrolling first in higher education (earnings trajectory and experience). That is, $Y(HC, D, U, T)$. Estimating OLS returns identifies joint variation in human capital, ability and degrees $Y(HC, D, U)$, and I report that as a benchmark in Section 5.1.

I want to estimate

$$E[Y(HC^{+}|D = D, U = U, T = T)] - E[Y(HC|D = D, U = U, T = T)]$$

where HC^{+} represents the human capital a student accumulates when affected by the Bologna reform while HC represents counterfactual human capital.

Using variation in human capital induced by the Bologna reform – where the 1999 cohort accumulates extra human capital while the 2000 cohort does not – takes care of omitted variable bias. Students with more induced human capital should not differ on unobservables U such as ability or motivation, see Section 4.3 for details. While students should not differ on unobservables, enrolling in the different cohorts creates differences in human capital (HC), degrees (D) and time since enrollment in a given year (T). Both the direct effect of enrollment cohort on degrees and on time since enrollment violate the exclusion restriction in an IV framework.

The distribution of degrees differs little across cohorts. The retroactive implementation of the Bologna reform should have ensured that students in the 1999 and 2000 cohorts all earn the new degree types, bachelor's and master's. Avoiding the direct effect of degrees to isolate returns to human capital is the goal of this paper. However, some quick students in the 1999 cohort and extraordinarily quick students in the 2000 cohort earned the new degrees by graduating before the reform was implemented. This direct effect of enrollment cohort on degrees is very small – 2% in the 1999 cohort and 0.3% in the 2000 cohort earn *cand.mag.* degrees. I address direct degree effects in section 6.2 by constructing non-parametric bounds for potential outcomes of old degree holders had they earned the new degrees instead. Under mild assumptions on monotone selection into old degrees and the valuation of old degree types in the labor market, bounds on the returns to human capital are tight. This indicates that the (minimal) degree differences across cohorts are inconsequential for earnings differences and I am going to abstract from them here.

The direct effect of enrollment cohort on time since enrollment in a given calendar T year makes a big difference. In a given calendar year, the 1999 cohort mechanically has had an extra year, $\tau + 1$ years since enrollment compared to the 2000 cohort that has had τ years since enrollment to study and/or gather experience. Measured in calendar time, the reduced form effect of enrolling in a different cohort becomes:

$$E[Y(HC^+, T)|Z_1 = 1] - E[Y(HC, T)|Z_1 = 0]$$

This identifies the joint effect of accumulating extra human capital and having more time since enrollment, where year T for $Z_1 = 1$ is $\tau + 1$ and T for $Z_1 = 0$ is τ. This constitutes an upper bound on the human capital effect, and I present estimates for this upper bound in section 4.2.

In order to identify the returns to human capital, I use a placebo cohort pair which only differs in time since enrollment in each given calendar year, that is, a pair of cohorts where one enrolls one year earlier, but without any differences in human capital. The reduced form effect for the placebo cohorts with the placebo instrument Z_2 captures the effect of enrolling earlier on its own:

$$E[Y(HC, T)|Z_2 = 1] - E[Y(HC, T = T)|Z_2 = 0]$$

If wages only depend jointly on human capital and time since enrollment, the placebo cohort pair's earnings account for the direct effect of enrolling one year earlier. This requires that earnings only depend on calendar year through time since enrollment and possibly a constant calendar year effect, but that the calendar year effect (other than through τ) does not interact with human capital. If that is the case, both reduced forms can be rewritten as:

$$E[Y(HC^+)|T = \tau + 1, Z_1 = 1] - E[Y(HC)|T = \tau, Z_1 = 0]$$

and

$$E[Y(HC)|T = \tau + 1, Z_2 = 1] - E[Y(HC)|T = \tau, Z_2 = 0]$$

As long as neither instrument affects earnings directly other than through human capital and $E[Y(HC)|T=\tau, Z_1=0] - E[Y(HC)|T=\tau, Z_2=0] = c$, the difference between treated and placebo cohorts earnings pair identifies the return to human capital, net of time effects.

Alternatively, earnings can be measured at a given time since enrollment. At any given time since enrollment, enrollment cohort does not determine how many years have passed, so that:

$$E[Y(HC^+)|T=\tau, Z_1=1] - E[Y(HC)|T=\tau, Z_1=0]$$

Both ways of isolating human capital identify returns to human capital relative to what a student would have earned had they not been induced to stay longer at university, for the group of students that are affected by the reform. The counterfactual implies that students who do not study longer can use their time on something else. Time advantages are either accounted for via a placebo cohort pair, or eliminated by measuring in relative time. Both measures incorporate the opportunity cost of studying longer.

This human capital return parameter captures this opportunity-cost adjusted return to Bologna-induced human capital. This is a special case of a policy-relevant treatment effect (PRTE) because policy-makers truly did influence students' study time at this margin (Mogstad and Torgovitsky, 2018).[9] Policy makers can only influence how people allocate their time, but they cannot change the total time people have available. In terms of education policy, policy makers can incentivize people to take more education, but that will translate into having less (potential) experience at a given age. In order to estimate whether extra higher education increases a persons' earnings, the counterfactual for receiving extra education should be spending the time working (or unemployed) instead and comparing outcomes after a given amount of time. I rely on the earnings of 2000 cohort students to estimate the counterfactual earnings of 1999 cohort students. Due to the Bologna reform, 1999 cohort students study on average 3.5 months longer. 2000 cohort students graduate 3.5 months earlier and the typical students enters the labor market after graduating. At any given time since enrollment, the 2000 cohort will have slightly more experience. Having less experience is a direct consequence of receiving more education for the 1999 cohort. 2000-cohort students' experience-boosted earnings constitute the most plausible estimate of counterfactual earnings for the 1999 students. This is what the 1999 students would have

[9]Characterizing the group of affected students in 1999, this PRTE is for students in the upper-middle of the study speed distribution. Estimating the effect for other enrollment cohort could identify the returns for slower students (earlier cohorts) or quicker students (later cohorts). This could give an indication whether the PRTE estimated in this study would differ immensely from PRTEs at a different margin. However, earlier cohorts have students that graduate with hugely different final degrees, identifying a mixture of human capital and degree effects; for later cohorts, the pre-screening period does not end before the Bologna reform.

earned, had they not been encouraged by the Bologna transition to study longer. Returns to staying longer in education are thus directly contrasted with a realistic other way to use one's time; graduating earlier, working and collecting experience. In principle, students could also be unemployed or outside the labor force after graduating. In this sample of Norwegian graduates, the by far most common alternative to studying longer is working, with over 85% being in full-time employment one year after graduating.

Since the comparison group will have to spend their time on something, any policy-relevant treatment parameter should incorporate that alternative use of time. In this application, this means that one should not attempt to control for experience; extra experience for the comparison group that does not acquire more human capital is part of the variation that a policy-maker could conceivably induce. Policy-makers cannot induce students to receive more education without reducing the potential experience at a given time since enrollment. The PRTE captures the opportunity-cost corrected returns to acquiring extra human capital in this setting. That is, it captures how much more a student earns compared to the counterfactual of acquiring (potential) experience.

In the returns to education literature, it is very common to estimate returns using a Mincerian equation. This equation specifies separate parameters for returns to education and experience (Heckman et al., 2006). Estimating a version of this equation is common in applied work (Arteaga, 2018; Carneiro et al., 2011; Bhuller et al., 2017). When measuring outcomes at a given point in time, study time and experience are almost always negatively correlated. When study time is arguably exogenous, (less) experience results directly from (more) study time. One may argue that controlling for experience isolates the direct controlled effect of education (Pearl, 2009). This requires very strong assumptions on potential outcomes for students with more or less experience in different cohorts. In fact, in most cases, experience will be a bad control (Angrist and Pischke, 2009). Since this "partial-effect" parameter is an important benchmark in the literature, I discuss when it can be estimated and present some assumption-heavy estimates of this benchmark parameter in section 6.

4.2 Specifications

To estimate returns to higher education, I use percentile ranks (1-100) of earned incomes as the outcome, where earned income is defined as the sum of wages and earnings from self-employment.[10] Earnings ranks allow to compare earnings that are systematically measured

[10] One could argue for defining income as all income, including state transfers and capital income. University-acquired human capital may well improve the capacity to figure out which benefits one is entitled to and to apply for them successfully. However, earned income offers a more comparable earnings measure relative to other settings where there are not as many state transfers available as in Norway.

in different earnings years and most specifications below rely on measuring earnings in some variation of time since enrollment. Inflation changes nominal incomes even in the absence of any fluctuation in real earnings and most years see some changes even to real earnings, because of economic growth or recessions. Measuring earnings as percentile ranks of earned income can partly, but not fully adjust for that. I report log earnings as a benchmark where feasible.

OLS benchmark

OLS estimates of the endogenous return to study time as a benchmark and capture $Y(HC, D, U)$. Study time varies endogenously across students because some students study slowly and other study quickly for a given number of credits (negative selection), and because they choose different amounts of study credits (positive selection), also resulting in different final degrees. OLS estimates capture joint positive and negative selection, returns to acquired human capital and returns to signalling. Students studying longer for a given degree are negatively selected. While high ability students may choose to study quickly or slowly, low ability students do not have the capacity to finish a degree quickly, so that more low ability students should be among the slow students. Students studying longer because they earn a higher degree level are most likely positively selected, if ability and the cost of earning a degree are negatively correlated. In addition to positive selection, these students also acquire more human capital and will receive signalling returns to the higher degree they earned. These counteracting effects make it difficult to predict whether returns to study time should be positive or negative. Previous work estimating endogenous returns to education on Norwegian data, such as Bhuller et al. (2017), find large positive effects returns to endogenous education length. However, they look at returns to length of education across all levels, while this paper focuses on the length of university education.

Equation 1 estimates returns to endogenous study length, named HC for human capital because I later instrument study length as a measure of human capital. γ_t reflects returns to study time in a given calendar year and I estimate Equation 1 separately for each year. X contains the same set of time-invariant control variables as in the main specification to increase efficiency. These are mothers' and fathers' education, age at enrollment, broad study field and gender, all included as factors. Additionally, I control for enrollment cohort to eliminate the exogenous portion of variation in human capital.

$$Y_{i,t} = \alpha_t + \gamma_t HC_i + \delta_t Cohort_i + X_i\beta_t + U_{i,t} \quad (1)$$

For the transitioning cohorts of the Bologna reform, study length varies because of the reform timing relative to enrollment date, but the induced extra time does not lead to higher degrees. In contrast, endogenous study time co-varies with obtained degrees. Equation 1 above jointly estimates the returns to endogenous study time and endogenously chosen degrees. This moves a step closer to the kind of variation the Bologna reform induces – changes in study length without changes in degrees.

$$Y_{i,t} = \alpha_t + \gamma_t HC_i + \theta_t Degree + \delta_t Cohort_i + X_i\beta_t + U_{i,t} \tag{2}$$

In Equation 2 I control for the level of a degree, using dropout, undergraduate university, postgraduate university, community college and PhD as levels. This removes returns to signals, but also positive selection into higher degree types. Longer study time driven by students that study more slowly for a given degree yield a downward-biased estimate of the return to study time due to negative selection. Equation 2 serves as a lower benchmark for the returns to study time.

I estimate both equations on the baseline sample (as in the main specification), with both enrollment cohorts pooled. γ_t in years prior to enrollment reflect selection effects since neither human capital nor signals are realized at this point. Since these are earnings where the majority of the sample still attends school and works mostly part-time, insignificant γ_t are a very mild test on selection effects. γ_t in the early 2000s reflect opportunity costs of staying longer in education. γ_t in the 2010s converge to the long-term earnings differences for people spending more time at university. In the appendix Figure A.3 I present results where with human capital HC measured study credits as opposed to study time.

Upper bound benchmark: calendar time reduced form

As a first approximation, I estimate the reduced form effect of being in the 1999 rather than the 2000 enrollment cohort on earnings. This specification captures the effect of enrolling in 1999 rather than in 2000, which serves as an upper bound on the human-capital returns to higher education. The exclusion restriction is violated, preventing 2SLS estimation. When measuring outcomes in calendar time, enrollment cohort predicts both how much human capital a student acquires and that students from the 1999 cohort have one extra year between the time they enroll and the time their earnings are measured. Having extra time in itself should make 1999 students earn higher wages. In any given calendar year, students from the 1999 cohort are on average one year older, receive on average 3.5 extra months of education (because of the reform) and almost 9 extra months of experience (since few people in the sample are unemployed).

$$Y_{i,t} = \alpha_t + \gamma_t Cohort_i + \theta_t Cohort_i * S_i + \delta_t S_i + X_i \beta_t + U_{i,t} \tag{3}$$

γ_t reflect returns to enrolling one year earlier. γ_t before enrollment reflects the pure effect of being one year older. After graduation, γ_t represent the combined effect of being one year older and receiving Bologna-induced extra education. I estimate Equation 3 separately for each year from 1997 to 2020. Earlier earnings data only capture earnings in the right tail of the age at enrollment distribution. The median 2000 cohort student is 17 years in 1997. At that age, a sizable fraction already works part-time. Going further back in time, a large fraction of the enrollment cohort will not even be legally permitted to work.

Main specification: Placebo-corrected returns to human capital

The main specification takes the calendar-time reduced-form model as a starting point. I account for the benefit of being one year older by making use of a post-Bologna cohort pair selected with the same inclusion rules as the Bologna cohort. Earnings differences between the cohorts in the placebo pair should only reflect the advantage of studying one year longer since neither cohort in the pair was affected by the reform. This design is close to a difference-in-difference set-up, but the baseline effects represent already a difference. The earnings advantage of the 1999 treatment cohort is measured relative to their adjacent cohort 2000 and the placebo earnings advantage of the 2001 cohort is measured relative to their adjacent cohort 2002. Besides, the main difference is measured in event time. Effectively, this compares how well both cohorts do relative to the enrollment date of the treated and placebo-treated cohort.

For the placebo, students in Bologna subjects at a non-reform time are more suited than non-Bologna students at the same time. The professional degrees that were not subject to the Bologna reform map into highly regulated occupations. Students from Bologna do not have access to these occupations and will end up in different labor market segments. This makes it unlikely that they would be on the same earnings trajectory in the absence of the Bologna reform, which is the identifying assumption for a diff-in-diff design.[11]

Focusing on students in Bologna subjects makes it more likely that the placebo cohorts have the same time-since-enrollment-earnings profile as the treated cohorts. Directly adjacent cohort pairs are unsuitable, since students in different speed brackets experience the reform for different enrollment years. The amount of credits in July 2003 determines whether

[11] Rationing of study slots in the profession as well as restrictions to enter professions without professional education generally translates into very high starting wages. Additionally, the state is a crucial employer for the largest of these occupations, meaning that state budgets and laws may have an independent impact on the wages of this group.

a student will get too many study credits and thus experiences the Bologna reform. Students from other cohorts than the 1999 cohort will be treated according to their study speed, with slower students receiving extra credits when enrolling in earlier years and quicker students receiving extra credits when enrolling in later years. Table A.2 in the appendix shows the expected exposure to the Bologna reform conditional on study speed for different enrollment cohorts.

The pre-Bologna cohorts will receive a bachelor with overshooting human capital at 30-39 credits per year for 1997 and 36-47 credits per year for 1998, while the 2000 cohort will be treated at 61-79 credits per year and the 2001 cohort at 90-119 credits per year. While many students study more slowly, less than one percent of students complete more than 90 study credits per year. Earlier cohorts should therefore have a first stage as well, which makes them unsuitable to serve as placebo cohorts. Testing cohort pairs for first stages in the data gives significant first stages for all pre-Bologna cohorts as well as for the 2000-2001 comparison. In addition, the directly adjacent pairs, that is 1998-1999 and 2002-2001 cannot be used, since some students would then be part of both the placebo and the control group.

The 2001-2002 cohort pair is more promising, because even for the 2001 cohort, only students from 90 credits and upwards should be treated. Verifying this in the data, it turns out that there is no significant first stage for the 2001-2002 cohort pair. The enrollment-cohort coefficient for the 2001 cohort is insignificant and very close to zero with 0.039 extra years for the 2001 cohort with a p-value of .58, see Table A.4.

Choosing much earlier cohorts would be an alternative but comes with two different challenges. Firstly, study credits were only collected starting in 1998 which makes it difficult to determine a sample with exactly the same selection rules for, say, 1995-1996. Secondly, earlier cohorts are inherently worse placebo candidates because most students in these cohorts receive the old degrees, making them less similar to the treatment cohorts 1999 and 2000 where the overwhelming majority receives the new degree types. Thirdly, since more students study slowly than quickly, the earliest post-Bologna placebo is the 2001-2002 cohort, while the earliest pre-Bologna placebos would be much further apart in time from the 1999-2000 cohort.

I estimate a quasi difference-in-difference placebo model of the following form

$$Y_{i,t-r} = \alpha_{t-r} + \theta_{t-r} Treat_i + \delta_{t-r} Advant_i + \gamma_{t-r} Treat_i * Advant_i + \beta_{t-r} X_i + U_{i,t-r} \tag{4}$$

γ_{t-r} is the coefficient of interest and captures the difference in the older cohorts earnings' advantage between the treatment cohorts (1999 and 2000) and the placebo cohorts (2001 and 2002). $Treat_i$ is an indicator for being in the treatment cohort pair, that is, enrolling either in 1999 or 2000, while $Advant_i$ is an indicator for being in the older cohort in any of the cohort pairs, that is, enrolling in 1999 or 2001. $t - r$ is time relative to the enrollment of the older cohort for each cohort pair. This means that within each cohort comparison, earnings differences are estimated in calendar time, while the earnings differences between the cohort pairs are estimated in time since enrollment of the older cohorts. X contains the fixed set of time-invariant controls.

The identifying assumption for this design is that earnings differences for the treated cohort pair 1999/2000 would have followed the same trend in the absence of the Bologna reform as the earnings differences for the 2001/2002 cohort pair. Although I cannot verify this after the onset of treatment, that is, once students have accumulated different amounts of study time at university, I can test for pre-trends. Compared to a classical difference-in-difference design, pre-trends should be checked in time since enrollment of the older cohort in each pair, since cohort pairs are offset, but income differences within cohort pairs are measured in the same year.

I use a placebo difference-in-difference design on a parameter that already reflects a difference since I want to account for the year advantage of the 1999 enrollment cohort relative to the 2000 cohort. For this be valid, earnings differences in the placebo cohort must only reflect the year advantage. While there is no first stage for the 2001-2002 cohort, it could be that the enrollment cohorts differ from each other at enrollment. Table ?? below reports balancing on the same observable characteristics that used to assess the 1999/2000 cohort.

	2002 cohort		2001 cohort		
Variable	Mean	SD	Mean	SD	p.value
Female	0.64	0.48	0.63	0.48	0.21
Mothers' education	4.58	1.73	4.63	1.73	0.25
Fathers' education	4.89	4.95	4.89	4.95	0.11
Age at enrollment	21.16	4.66	21.27	5.2	0.31
Nr of siblings	2.2	0.9	2.15	0.87	0.01
Natural science student	0.11	0.32	0.13	0.34	0.01
Spring semester	0.18	0.38	0.19	0.39	0.28
N	4281		4004		

Table 2: Balancing on observable for the placebo cohort pair

Most notably, both enrollment cohorts are larger. I select the 2001 and 2002 cohorts using the same selection rules: Enrolling at a university in that year, being enrolled in both pre-screening years, not taking a professional degree in the first five semesters and collecting at least one point. For the 2001 and 2002 cohort part of their screening time is close to and after the Bologna reform. The Bologna reform imposed stricter rules on reporting for faculties. As the data quality improves, data on where students actually enroll and whether they take any study credits have fewer missing values. For this reason, there are generally *more* students in the placebo cohorts. Although most observable characteristics are balanced across cohorts and some p-values may be significant without any underlying differences when testing many variables, balancing overall looks slightly worse. One reason for concern may be that a slightly larger fraction of students in the 2001 cohort (13%) studies in the natural sciences compared to the 2002 cohort (11%). Since I measure the year advantage of enrolling one year earlier and the 2001 cohort has more natural science students in addition to enrolling earlier, the earnings advantage of the 2001 cohort may be overestimated. This makes the estimate more conservative in terms of returns to university-acquired human capital. Regressing cohort on all balancing variables gives an F statistic of 1.23.

I estimate Equation 4 separately for each $t-r$, where r is the relative enrollment year for reform and placebo cohort pairs, 1999 and 2001 respectively. γ_{t-r} before enrollment checks whether both cohort pairs follow the same pre-trends in earnings differences. If assuming that students' human capital matters for jobs after graduation, rather than for part-time jobs that students may pursue while studying, the first university years where most students still study represent an additional test on pre-trends. In the most intense graduation time around $t - r = 8$ the placebo cohort should earn slightly better, if anything, since they graduate slightly earlier on average. These years capture the opportunity cost to staying longer in education. In the long run, the 1999 cohort should earn better relative to the 2000 cohort than the 2001 cohort relative to the 2002 cohort if there is a return to university-acquired human capital. To improve efficiency and to obtain a point estimate, I pool several years and estimate a single coefficient for the returns to human capital. I estimate one coefficient for the interaction of being treated (enrolling in 1999 or 2000) and having a year advantage (starting in 1999 or 2001) as well as the baseline effect of being treated and having an advantage, but keep the α_{t-r} from Equation 4 as relative time fixed effects as follows:

$$Y_{i,t-r} = \alpha_{t-r} + \theta Treat_i + \delta Advant_i + \gamma Treat_i * Advant_i + \beta X_i + U_{i,t-r} \tag{5}$$

In this model, I cannot interact the $Treat_i * Advant_i$ (capturing acquired-human capital) term further with the spring semester indicator, because the two two-way interactions together with the baseline terms would be a linear combination of the three-way interaction. However, estimating this model only on fall semester students (80% of the sample) makes sure that the γ coefficient from the event diff-in-diff specification (Equation 5) coincides perfectly with the difference of year advantages between placebo and treated cohort pairs.

Pooling the last 5 years represents the closest approximation for the effect of university-acquired human capital on earnings potential. This is because earnings converge towards the long-run earnings potential with increasing time since graduation. Individuals' earnings tend to stabilize fairly late in life, especially for college graduates. Previous studies find the 30s to late 40s most predictive of lifetime earnings in the US (Haider and Solon, 2006), including both college and non-college workers. Baker and Solon (2003) find the lowest variance around age 45 for Canadian men while Björklund (1993) conclude that earnings are not informative about lifetime earnings before age 30 in Sweden. In Norwegian data, Bhuller et al. (2017) document that earnings profiles do not even start to flatten before age 40 for college-educated men and reach their peak at age 47. They restrict their analysis to men because women's' incomes tend to take even longer to become predictive of lifetime earnings. Nilsen et al. (2012) find that mens' earnings take longer time to converge to their earnings potential for each subsequent cohort, in line with educational expansion. Markussen and Røed (2020) use joint parents' income at ages 52-57 to approximate families' earnings potential.

In 2020, the last year available in the data, students are on average 40.5 years. The birth cohorts mapping into the 1999 and 2000 enrollment cohorts are much more recent than the ones used in the literature above. Additionally, university enrollment cohorts consist by definition of individuals enrolling in higher education, who take longer to reach their earnings potential. Therefore, individuals in this particular sample should take as long or longer to reach their long-run earnings potential compared to the studies discussed above. Having both men and women in the sample likely delays convergence to lifetime earnings even further, especially because the sample period ends before the fertile window ends. For this reason, the last earnings years available are most informative about earnings potential.

Event study 2SLS

Our main specification allows to estimate the earnings differences both between the two treatment and the two placebo cohorts in their calendar time. This automatically discounts economic shocks to the labor market and takes care of wage growth and inflation. However, the differences between cohorts must be meaningfully comparable across income years, since I measure each difference in time since enrollment of the older cohort. These are two years for the 1999 vs. 2001 cohort. In addition, I need to assume that the 2001-2002 earnings differences would have followed the same trajectory as the 1999-2000 earnings differences if the Bologna reform had not been enacted.

To check the robustness of the results with respect to differences being comparable across years and the placebo cohort following the same trend, I estimate the same parameter of interest – returns to substituting experience with education due to the Bologna reform – with a different specification. Rather than using the placebo cohorts to difference out the earnings advantage of enrolling one year earlier, I measure estimate returns to human capital in an event study fashion. $t - \tau$ indexes time since enrollment, where t is calendar time and τ the enrollment year. I estimate a 2SLS event study of the following form:

$$Y_{i,t-\tau} = \alpha_{t-\tau} + \gamma_{t-\tau} HC_i + X_i \beta_{t-\tau} + U_{i,t-\tau} \tag{6}$$

I estimate Equation 6 separately for each $t - \tau$ so that $\alpha_{t-\tau}$ effectively becomes a constant. $\gamma_{t-\tau}$ estimates returns to university-acquired human capital at a given time since enrollment. Human capital HC is measured as study time in years. All controls in X_i are fully saturated so that this 2SLS specification has a LATE interpretation (Blandhol et al., 2022; Angrist and Pischke, 2009). I include students earliest three years prior to enrollment because earlier earnings data only represent earnings of students in the upper tail of the distribution of age at enrollment that already work. My data series ends in 2020, so that I can follow them up until 20 years after enrollment (2019 and 2020 respectively). I estimate Equation 6 separately for each $t - \tau$. In principle, one may want to interact only study time and the instruments with each relative time period and estimate a vector of constant β coefficients for the time-invariant controls. This specification with 2*2*23=92 instruments suffers from a many IV bias. Therefore, I estimate the model separately for each time period.

As in the main specification, returns prior to enrollment $\gamma_{t-\tau} < 0$ serve as an implicit test on pre-trends. Measuring in relative time forces a direct trade-off between returns to education and returns to experience. In the long run, this direct trade-off is desirable because it incorporates the opportunity costs of education and more education would only be worthwhile if the returns to education exceed that of experience. However, upon graduating, incomes grow quickly with experience for the first years in the labour market. Even with a sizable return to an extra year of education, comparing a new graduate to an employee with one year of experience may yield a negative net return. This is part of the opportunity costs of staying longer in education. Once the experience curve flattens after some years in the labour market, returns to human capital for the long-run earnings potential should become visible. Both study speed and earned degrees vary considerably in the sample, leading to large variation in total study time and graduation dates. Thus, for most time points in the sample, some students will be recent graduates with experience effects dominating. The latest observations in the sample should start to converge to differences in earnings potential due to studying one year longer. Estimating Equation 6 separately for each year recovers how the earnings premium moves over time.

Measuring income in earnings percentile ranks makes incomes across earnings years as comparable as possible. Measuring the 1999 cohort systematically one year earlier, inflation, real wage growth, and wage shocks due to business cycles distort nominal or log earnings substantially. Ranks of earned income faithfully measures the status in the earnings distribution across years. However, some events such as sudden recession may influence people differentially by their level of experience. For example, a newly hired graduate may have a higher risk of losing their job in a recession compared to a person with somewhat longer tenure. The income year in which a recession happens will show up in different relative time measures across cohorts and the impact most likely differs by experience. Since students graduate over a broad time window, some students in both cohorts experience a recession (2008, 2020, less growth in 2016-2017) shortly after graduating and since this is measured at different time since enrollment, this introduces more noise in the estimates. Additionally, the 2000 cohort has more experience at any given time since enrollment, so that their incomes should be less susceptible to economic shocks for a given level of human capital. This direct benefit for the 2000 cohort is part of the opportunity cost for the 1999 cohort taking more education.

4.3 Instrument validity

The Bologna reform provides exogenous variation in the length of study and in the amount of study credits for the 1999 and 2000 transition cohorts. Enrollment cohort will serve as an instrument for university-acquired human capital, both to estimate reduced form effects and 2SLS. To obtain a local average treatment effect while respecting that treatment effects may be heterogeneous the four assumptions laid out in Imbens and Angrist (1994) need to be satisfied.

The independence assumption requires that the instrument is as good as randomly assigned, implying that students from the two enrollment cohorts should neither differ in their expected acquired human capital nor in their earnings potential if the enrollment cohort instrument were not to affect how cohorts experience the Bologna reform. A testable implication of that is whether students enrolling in 1999 or 2000 differ on observables. The balancing exercises in section 3.1 do not indicate reason for concern. Additionally, I demonstrate in the event study specifications that cohort membership does not predict labor earnings before people enroll and the treatment realizes. From a theoretical perspective, the fact that the reform happens several years after enrollment mitigates concerns about students self-selecting into different enrollment cohorts because they anticipate the reform. Therefore, I do not expect students to differ on unobservables that may matter for earnings or duration of education, such as motivation.

This design faces two possible challenges for the exclusion restriction. Firstly, enrollment cohort reflects enrollment time in addition to predicting study duration. When measuring outcomes in calendar time, this leads to a direct effect of the instrument on experience and thus on earnings. In any given calendar year, 1999 cohort students are one year older and had an extra year since enrollment to build their career through education or experience. Without further adjustment, this set-up violates the exclusion restriction. In the main specification, I address this challenge through a placebo design, estimating the pure year-advantage from the earnings profiles of an enrollment-cohort pair not affected by the Bologna reform. I then difference out the year-advantage in a quasi difference-in-difference design, recovering the effect of university-acquired human capital as the additional earnings' advantage of the 1999 cohort over the 2000 cohort relative to the pure year advantage measured for the placebo cohorts. In the alternative specification, I address the year advantage by estimating an event study 2SLS, so that outcomes are measured at a given time since enrollment. Both approaches remove the instruments' direct effect on earnings through the year advantage, mitigating concerns about the exclusion restriction.

Secondly, differences in degrees earned by graduates in the different cohorts could violate the exclusion restriction. In observational data, additional human capital (study time or study credits) is associated with higher level degrees on average; higher earnings of more educated individuals reflect both signalling and human capital in addition to selection. Finding a valid instrument for taking more higher education typically identifies returns to education; this includes both returns to acquired human capital and returns to obtaining higher level degrees. My design rests on an instrument that should only induce variation in university-acquired human capital, without simultaneously changing degrees. In principle, Bologna transition cohorts unexpectedly only had new degree types available in 2003. With some students from the 1999 cohort having progressed further into their studies than required for a bachelor degree but less than necessary to obtain the old degree before 2003, enrollment cohort should predict changes in human capital only.

In practice, some students still earn the old degree types. Students who were successfully induced towards earning more study credits by the reform – too successfully in a way – will end up with the old degree types. The additional time that 1999 cohort students have at their disposal that causes overshooting point for students that progress as planned allows very quick students to earn the old degree types. Any student earning more than 240 credits before or by June 2003 will earn a *cand.mag.* before the reform is introduced. In addition, students that had studied three years' worth of study by 2003, but selected courses that did not map into the new bachelor degrees were given a two year grace period to complete a *cand.mag.* degree. Students in the 2000 cohort need to study quicker to end up in either situation than students from the 1999 cohort. For this reason, more students from the 1999 cohort earn the old degree types. This violates the exclusion restriction because being in the 1999 enrollment cohort increases both human capital and induces a higher fraction of students to earn the old degree types and both human capital and degree types matter for earnings.

I cannot simply control for degree types, because the quality of old degree holders likely varies across cohorts. Students in the 2000 cohort need to study exceptionally quickly to still earn the old degrees, so that these old degree holders are likely positively selected. 1999 cohort students are positively selected as well, but to a lesser degree. From a practical perspective, the impact of this problem is limited because very few people receive the old degrees. 85 (2.4%) students in 1999 earn a *cand.mag.* relative to 14 (0.3%) in 2000 and 95 (2.6%) students in 1999 earn a *hovedfag* relative to 12 (0.3%) in 2000. In section 6.2, I estimate the counterfactual earnings of old degree holders for the scenario where they had instead received the new degree types, using non-parametric bounds. I re-estimate all specifications separately using the upper and lower bound on earnings for old-degree holders, thereby bounding the returns to university-acquired human capital.

Theoretically, monotonicity can be a challenge in the setting of the Bologna reform, since the discreteness of university degrees offers a plausible motivation for students to defy. If students ideal length of education is continuous, but universities can only certify specified amounts of education, degrees, students may want to adjust their choices if the length of offered degrees changes. This concern would arise both in a human capital and a signalling framework, as long as employers have trouble detecting and interpreting time spent at university that is not related to a degree.

In our sample, students do not choose to earn higher level degrees at different rates across cohorts, see Table 3 below. In section 6.3 I discuss why students may choose master degrees

Degree level	N 2000	N 1999	N	% 2000	% 1999	p_value
Dropout	246	241	487	0.51	0.49	0.75
Low Degree	326	334	660	0.49	0.51	0.66
High Degree	1033	1031	2064	0.50	0.50	0.95
PhD	194	199	393	0.49	0.51	0.72
Non-Bologna	341	403	744	0.46	0.54	0.01
Community College	1458	1455	2913	0.50	0.50	0.94

Table 3: Evidence on monotonicity: Differential degree choices

at different rates when the length of the undergraduate degree is reduced and why this is not the case for the 1999 cohort compared to the 2000 cohort.

4.4 Instrument relevance

I verify instrument relevance through the first stage and estimate a first stage of the following form:

$$HC_i = \alpha + \theta Cohort_i + \gamma Cohort_i * S_i + \delta S_i + X_i\beta + V_i \tag{7}$$

HC stands for human capital. In the main specification, I measure human capital as years spend in higher education after the pre-screening period. In alternative specifications I also measure human capital as study credits accumulated after the pre-screening period or as the number of chosen subjects after the pre-screening period. Cohort is a binary variable indicating whether a student enrolled in 1999 or in 2000. An enrollment year starts with the academic year. That is, students enrolling between August 1999 and July 2000 are in the 1999 enrollment cohort and students enrolling between August 2000 and July 2001 are in the 2000 enrollment cohort.

I interact the cohort instrument with a binary indicator for whether a student started in the spring semester. In both enrollment cohorts, spring semester students enroll half a year later than fall semester students. Among spring semester students, students need to study quicker to be affected, leading to a different complier group among spring students. Students in the study speed bracket from 51-68 study credits per year should be affected. Since fewer students fall into this bracket compared to the 45-60 study credit bracket for the fall students, the first stage among spring semester students should be weaker. The interaction between cohort membership and spring semester accounts for this. The matrix X contains the set of flexible time-invariant controls, mother's and father's education as factors for each of the 9 education levels, gender, natural science and age at enrollment. Age at enrollment consists

Table 4: First stage estimates

	Human capital measure:		Study time	Study credits	Subjects
	(1)	(2)	(3)	(4)	(5)
1999 Cohort	0.252***	0.301***	0.288***	6.635*	0.782***
	(0.075)	(0.082)	(0.081)	(2.877)	(0.195)
Spring student		−0.857***	−0.715***	−14.714**	−0.274
		(0.134)	(0.137)	(4.809)	(0.325)
1999 Cohort:Spring		−0.257	−0.212	−0.568	0.095
		(0.191)	(0.189)	(6.662)	(0.451)
Controls	No	No	Yes	Yes	Yes
R^2	0.002	0.016	0.060	0.085	0.051
Adj. R^2	0.001	0.016	0.055	0.080	0.045
Num. obs.	7230	7230	7028	7058	7058

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$
Models 1-3 report the main first stage with study time after pre-screening in years as the dependent variable. Model 4 and 5 report for alternative measures of human capital, study credits after pre-screening and number of chosen subjects after pre-screening. Flexible controls for age at enrollment, gender, natural sciences and mothers' and father's education level.

of one dummy for each age under 30, five-year brackets for age groups between 30 and 50 and one dummy for being older than 50.[12]

I measure university-acquired human capital fixed within person as the number of semesters a student was enrolled. One could argue that human capital accumulates over time, but this variation cannot be used. Firstly, human capital accumulated during the pre-screening period to predict outcomes. This is mostly relevant for the study credit and subject robustness checks of the first stage, since study time human capital is constant during pre-screening for all students. Secondly, having human capital at it's final value turns pre-enrollment 2SLS estimates into a check for pre-trends; the later-revealed human capital is not realized and should thus not affect earnings. Recording zero university-acquired human capital for everyone before enrollment would make it impossible to estimate this.

The empirical cdf, Figure 5 below, illustrates the timing of the first stage in time relative to enrollment. The figure displays the fraction of students who completed their studies after a given number of semesters for each cohort. Few students progress at normed study speed. Therefore, there are no distinct graduation jumps for the bachelor after 3 years (4 for the 1999 cohort) or for the master after 5 years (6 for the 1999 cohort). Rather, students

[12]Since I do not restrict age at enrollment, some students are up to 68 years old when first enrolling in higher education. Using 5-year-brackets for older students preserves degrees of freedom.

complete their studies at many different times.

Around 8 years after enrollment, the two cohorts differ most in how many students still study. For this time span, income differences between the two cohorts should be driven by differences in (full-time) labor force participation. Over 80 percent finish their studies within 10 years and over 90 percent within 12 years. At this point, earnings differences start to reflect wage differences among working individuals and not a mechanical extensive margin effect. While the difference in the graduation rates pick up the opportunity costs for staying longer in education, they are not a good measure of long-run earnings potential. Similarly, experience is likely to have a disproportionate effect in the first few years after graduating.

Students in this sample study longer compared to the full enrollment cohort. This is a side-effect of excluding people that do not collect a single study credit during the pre-screening cohort. Many of these students drop out and never earn any degree. For this reason, they tend to leave university earlier than the students that actually study. See Figure A.1 for the same figure, but including people without any credits.

In most specifications I estimate the first stage repeatedly and separately for each income year t or each time period since enrollment $t-\tau$, because I trace in the second stage how the effect of university-acquired human capital evolves over time. The first stage then becomes:

$$HC_i = \alpha_t + \theta_t Cohort_i + \gamma_t Cohort_i * S_i + \delta_t S_i + X_i \beta_t + V_{i,t} \tag{8}$$

$$HC_{i,t-\tau} = \alpha_{t-\tau} + \theta_{t-\tau} Cohort_i + \gamma_{t-\tau} Cohort_i * S_i + \delta_{t-\tau} S_i + X_i \beta_{t-\tau} + V_{i,t-\tau} \tag{9}$$

All first-stage variables are time-invariant, but first stage estimates will still differ minimally for different time periods. This is because some individuals have missing earnings data in some years. While at most 28 income-years can be observed per person, the average number of observed income years is 27.4, indicating that this is a minor deviation from a balanced panel. In 2000, most individuals' earnings are observed, with 7231 while in 2019 fewest are observed, with 7024 individuals; in total, 7263 individuals have full information on time-invariant characteristics. For the main specification, the cohort coefficient θ_{t_τ} varies between 0.279 and 0.304 when estimated on slightly different samples at different times since enrollment $t-\tau$, with the mean first stage coefficient of 0.288 coinciding with the cross-section first stage coefficient.

The F statistics for the cohort-only first stage is 11.72. This is on the margins for enrollment cohort being a weak instrument for human capital measured as study time. When predicting study credits rather than study time, the F statistics is as low 7.58. While there is

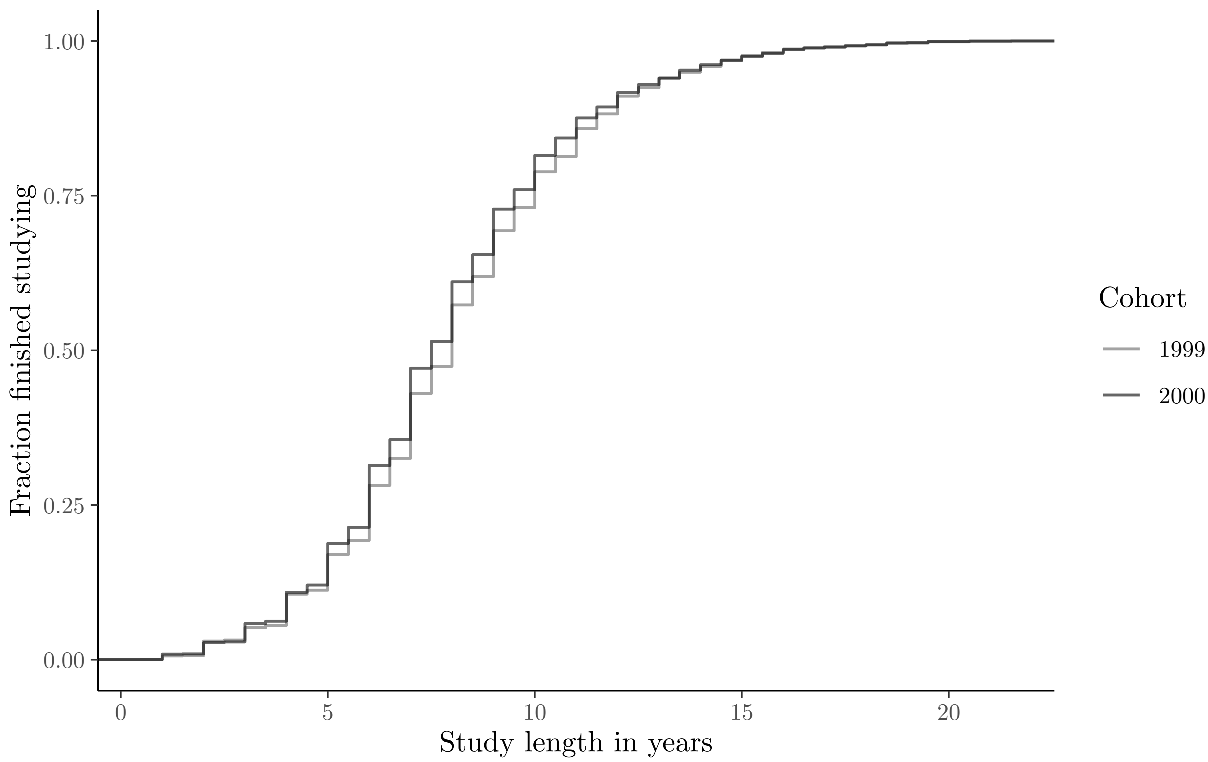


Figure 5: Timing of extra study time: empirical cdf across cohorts

Study length as net study time, that is, the total number of years a student spends in higher education, irrespective of gaps. I define ended study time as not being enrolled anymore and not enrolling again until 2020, the last year in the data. I do not condition on whether students graduate or drop out when they stop being enrolled.

a stronger first stage with F 25.7 for the attempted number of subjects, this variable is worst at capturing human capital. Collecting study credits and spending longer at university should reflect how much human capital a student accumulated; this is less obvious for attempted number of subjects, since this also counts subjects where students either failed the exam or withdrew from the exam. The reason for the fairly weak first stage in the baseline sample is two-fold. Firstly, registering which courses people take does not infer precisely whether they actually study in a Bologna-relevant degree and end up with a large number of students that finish professional degrees or graduate from untreated community colleges and therefore, unsurprisingly, do not react to the cohort instrument. Secondly, even for students that actually take up a Bologna subject, only those in a study speed range of 45-60 study credits per semester should be affected by the reform. In section 4.5 I use a sample-splitting and cross-fitting procedure to leverage information on how study speed affects reform exposure to construct an alternative sample with higher reform exposure.

4.5 Screening for expected reform exposure

As discussed in section 4.4, both non-Bologna students and students outside the reform-relevant study-speed window of 45-60 credits per year will not experience the reform. For this reason, there is a low concentration of compliers in the baseline sample. Having a low share of compliers in the sample leads to a high variance of the 2SLS estimator. With a small share of compliers π, the enumerator becomes quickly very small since π enters squared, leading to a large variance. If one could increase the share of compliers in the sample, by removing non-compliers, the 2SLS estimates should become much more efficient (Hazard and Löwe, 2022).

$$Var[LATE_{2SLS}] = \frac{1}{N}\frac{1}{\pi^2}\frac{\sigma_\epsilon^2}{p(1-p)}$$

While I cannot use students' final degrees to remove non-compliers (since degree is an outcome), study speed during the pre-screening period is a powerful predictor of whether a student will be exposed to the reform. Firstly, students in the 1999 cohort with less than 45 credits per year are unlikely to acquire any extra study credits by the time the reform is introduced, and will have no reason to acquire more credits relative to their 2000 counterparts after the reform. Among slow students there should be few compliers and many never-takers. Secondly, students in the 1999 cohorts with more than 60 credits per year do collect extra credits, but so do their 2000 counterparts. Among quick students, there should be few compliers and many always-takers. Since these subpopulations of very slow and very quick students do not react to the enrollment cohort instrument, they do not contribute to identifying the LATE.

As a first indication of treatment propensity in the sample, the histogram below illustrates the variability in study speed in the student population. I use the average number of study credits per year in the first five semesters as a proxy for latent study speed.

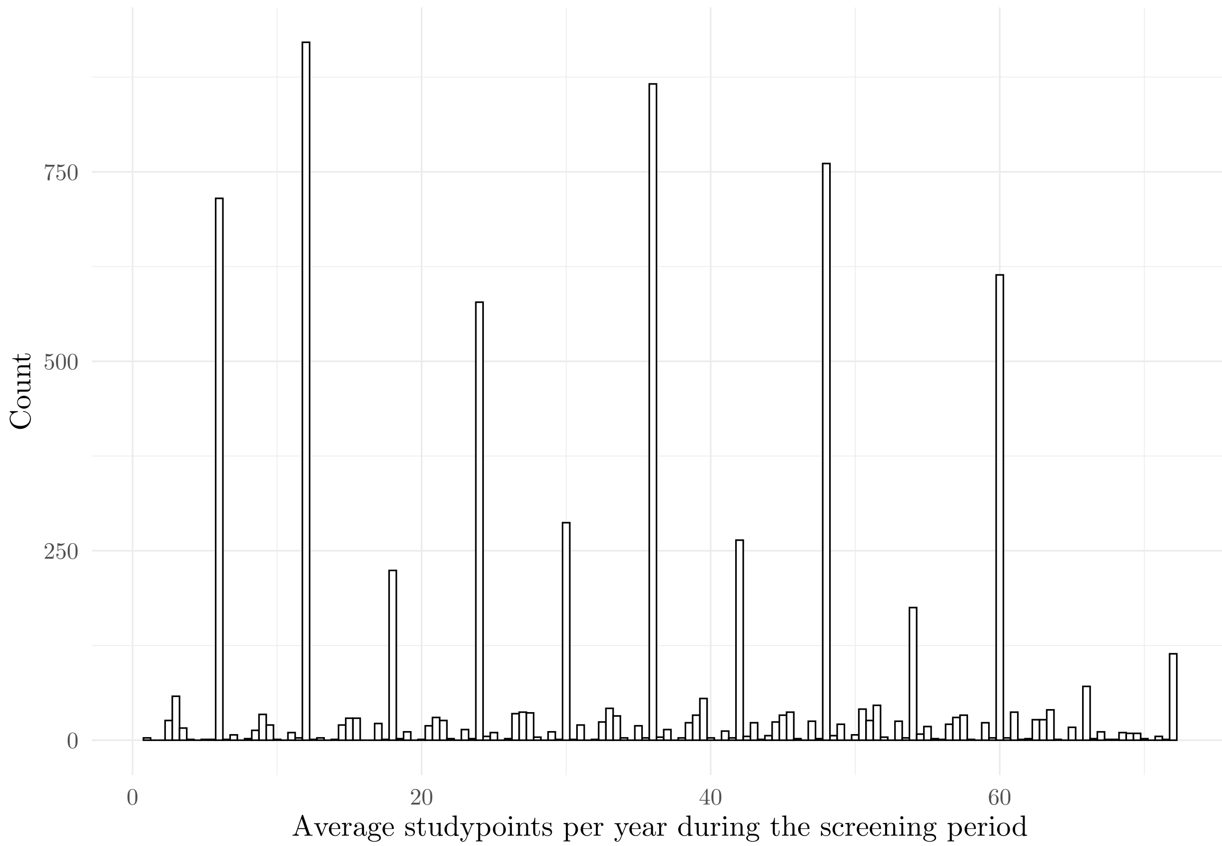


Figure 6: Average number of credits earned per year in the pre-screening period.

Study speed is calculated by summing all study credits accumulated at a university during the first five semesters and dividing by 2.5. Study-speed is top-censored at the 99% percentile to improve readability in this graph, but not in the data.

The dispersion of study speed suggests that a large fraction of the sample will likely not be exposed to the reform, with less than a third of students nominally in the expected-exposure bracket of 45-60 study credits. However, the theoretical expectation of treatment in the 45-60 credit bracket does not translate well into measured study speed, because there is measurement error in study credits and because study speed may vary inherently over time.

Before the Bologna-reform, faculties were under-reporting study credits earned by students. I trace under-reporting in the data by comparing the total amount of credits a student supposedly accumulated with the minimum number of study credits necessary for the highest degree they achieved. Note that summing all credits also account for credits from language courses or music projects or whichever other courses students may follow in their leisure that are unrelated to the discipline in which they eventually earn a degree. Likewise, if students begin studying in one discipline and then switch they may not be able to use all their credits towards this degree. Almost 50% of the students in the data do not finish the degree they initially embarked on. Assuming that all study credits a student ever accumulated count towards the degree they finally earn identifies a lower bound of under-reporting. 20.8% of students in our sample have fewer credits reported than necessary for the highest degree they earned, and accounting for the fact that some of these credits may not even be related to their highest degree, this indicates that under-reporting was fairly common. Some of this discrepancy may stem from study credits accumulated outside of Norway, but even in the group of students born in Norway and without immigrant parents, 19.1% still have fewer credits registered than necessary for their respective highest degree. Looking at all credits students ever collect in the Norwegian higher education system – this requires that credits earned at community colleges count fully and directly for all degrees earned at university – 11% still do not earn the minimum number of credits necessary for the highest degree they obtained. See appendix, Figure A.2, for the distribution of the difference between required and obtained study credits. This indicates that study credits tend to understate the number of credits a student achieves, but it is difficult to determine how much.

In addition, study speed varies a lot, both for accounting reasons and because of genuine study speed changes. Firstly, many pre-Bologna subjects[13] span a whole year where students only receive their study credits at the end, so that it looks as if they did not take any courses in the first of the two semesters in that subject. For measuring study speed, this means that students generally look slower than they are, even if the total amount of study credits were reported correctly, since they receive their credits in chunks rather than on a rolling basis. Secondly, students may genuinely speed up or slow down due to changes in motivation or personal reasons. After the pre-screening period, students may also speed up or slow down because they anticipate the effects of the reform. 1999 students on the quicker end of the spectrum may speed up in spring 2003 to earn a *cand.mag.* before the reform makes it impossible to earn that degree. 1999 students in the normal study speed range may slow

[13]Especially the so-called "grunnfag", an introduction to one discipline, where up to two of these basic introductions could be credited towards a *cand.mag.* degree.

down in spring 2003 when they learn that they already have sufficient study credits for the bachelor they will be forced to earn. The groups set in the theoretical cut-offs depend on correctly measured constant study speed. Both reporting issues and study speed variation indicate that study speed measured as average study credits per year in the first five semester will not match the theoretical study speed treatment bracket.

As a further complication, I cannot observe treatment status at the individual level. I measure human capital primarily as study duration, and secondarily as number of completed subjects. Both variables enter continuous in the estimation. While one could argue that study time could be measured as discrete ordered – for example in study months – study time does not lend itself to a binary definition of treatment status. Students can choose among several levels of degrees, and which level one chooses, undergraduate or postgraduate, may be affected by the reform. Within each level, students study at different speeds, leading to a fairly flat distribution of study duration. A binary definition of treated vs. untreated study length would could across degree categories and study speed brackets rather than capturing the variation in the length of the undergraduate degree the reform induces. Note also, that the expected treatment is at most one year of study, and thus always a small fraction of the time spent in higher education.

While I cannot tell for a given individual whether they are treated or not, I can estimate cohort-differences in study length. Estimating the first stage separately by study speed quintiles bears out the expected u-shaped relation between study speed and the propensity to be treated. This indicates that the slow and quick subpopulations do not react to the instrument and do not contribute to estimating the LATE. See below for an estimate of the first stage by study speed quintiles, along with the distribution of study speed. The average first stage in the baseline sample conceals a strong first stage for students with intermediate study speed and no first stage for slow and quick students. First stage heterogeneity leads to inefficient estimators (Abadie et al., 2023). One could obtain a stronger first stage by selecting the sample based on the strength of the instrument in different subgroups. This is fairly common in the literature. Stephens and Yang (2014), for example, exclude African Americans from the analysis, noting that they are little reactive to compulsory schooling instruments, while Johnson (2020) excludes regions with a weak first stage. However, simply selecting parts of the sample with a high first stage leads to first-order bias in LATE estimates as well as too small second stage standard errors.[14] When a different but comparable sample

[14]Testing for the first stage and selecting in the sample sample leads to overestimating the first stage. If no first stage was present in truth, some subgroups may still look as if they have a first stage due to random variation. Testing and selecting in the sample would include those groups without a first stage in the analysis

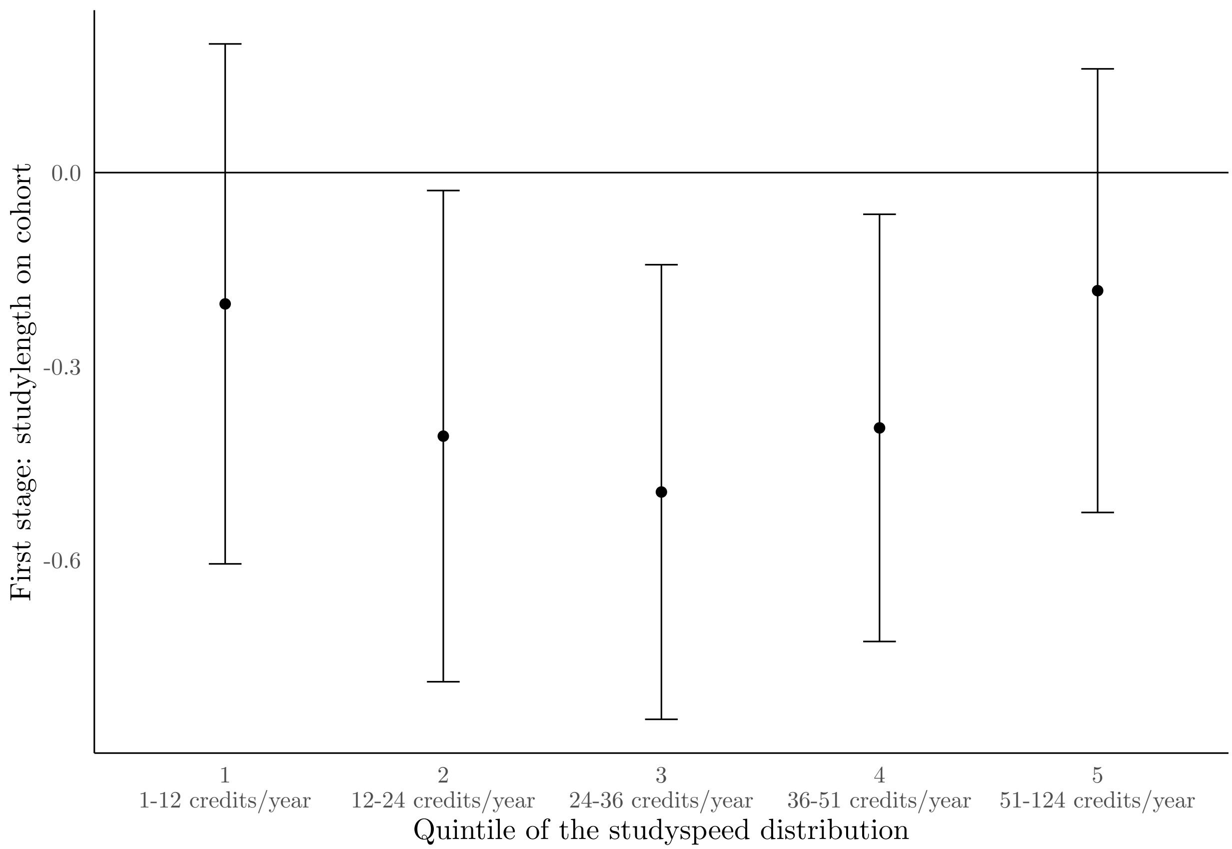


Figure 7: Strength of the first stage by study speed group

is available to test what predicts treatment, like in Altmejd et al. (2021) for the US sample, this problem can be circumvented. An expanding literature investigates how to increase efficiency in the presence of first-stage heterogeneity without introducing bias and how to recover appropriate standard errors.

A commonly used alternative to selecting subgroups of the sample based on instrument strength, is to interact the instrument with the variables along which the first stage differs, see for example, Jackson et al. (2016), Løken et al. (2012) and Carneiro et al. (2011). A practical problem with this approach is that interacting can generate many (weak) instruments. In addition, the estimates represent a convex weighted average of LATEs if there is second stage heterogeneity, giving more weight to groups with high compliance rates. Abadie et al. (2023) address many-IV bias by selecting groups with strong first stages, then using the interacted-IV 2SLS in a different sample. This procedure still identifies a convex weighted average of LATEs. Coussens and Spiess (2021), building on Huntington-Klein (2020), propose weighting groups of the sample by their estimated share of compliers. The estimator converges to weighing each LATE by the square of the share of compliers in each group, deviating from the normal LATE procedure which weights each LATE by the share of compliers. If there is treatment effect heterogeneity across groups, this introduces bias.

Hazard and Löwe (2022) suggest a sample-splitting cross-fit test-and-select estimator where testing for compliance is conducted in the first half of the sample $\mathcal{I}1$, and estimation with the selected groups in the other second of the sample $\mathcal{I}1$. After using one half of the sample to select subgroups and estimating on the other half, the roles of the two subsamples are reversed, so that then the second sample is used for testing and the first sample for estimation. This approach yields smaller variance gains compared to Coussens and Spiess (2021) and Abadie et al. (2023), but still recovers the LATE as long as treatment effect heterogeneity does not exceed the sampling variation.

I adopt a modified a variation of the procedure suggested by Hazard and Löwe (2022) because both splitting by theoretical expected study speed brackets and by quantile works has distinct disadvantages. Testing group-wise first stage coefficients requires to define groups G along which to split each sample. In many applications, G will be intuitively defined, for instance through demographic groups. In this project, there is a clear theoretical reason to

(while testing and selecting on different partitions would not). In those group with a "random noise" first stage, the LATE is effectively estimated on the difference between always-takers and never-takers, which could take any direction. Concerning inference, note that selecting on the first stage induces correlation between the first-stage error term and the instrument, violating the exclusion restriction. See (Hazard and Löwe, 2022), and Abadie et al. (2023), for more details.

believe that some groups should be more exposed to the treatment: Students with a study speed between 45 and 60 credits may be moved by the instrument, while students with less than 45 credits, or more than 60 credits a year should not. As explained above, under-reporting of total study credits, delay in study credit reporting due to multiple-semester courses and genuine variation in study speed makes it difficult to take this prediction to the data. Proxied study speed in the data performs poorly in delineating expected "too slow","exposed" and "too quick" groups of students in the theoretical 45-60 study credit bracket. Applying Hazard and Löwe's method directly, that is, splitting the study speed distribution into quantiles and testing for each subgroup is sub optimal as well. Quantiles may cut across groups of students with and without a first stage. Bunching of students around the multiples of some common subject credit values (see Figure 6) makes it impossible to define finely grained quantiles to mitigate the issue. Splitting into quintiles is feasible, but leads almost mechanically to loosing at least 40% of the sample (assuming that the slowest and quickest quintile's first stage is non-significant because of the large presence of non-compliers there). Within the upper and the lower quintile, there may still be many students affected by the reform that will not be included because they are grouped together with non-affected students.

For this reason, I modify Hazard and Löwe's approach by conducting a three-step sample-split and cross-fitting procedure. I start by splitting the baseline sample into three subsamples of the same size, stratified by study speed. Then, I use a predict-test-select procedure to estimate on relevant portions of the sample. First, I determine the optimal cut-off value for study speed groups in sample $\mathcal{I}1$ to form three groups, "too slow", "exposed" and "too quick" students. Then, I use cut-offs for the optimal groups from sample $\mathcal{I}1$ to estimate the first stage in each group in sample $\mathcal{I}2$. Finally, I select groups with a t-statistic significant at the 5% level in sample $\mathcal{I}2$ for sample $\mathcal{I}3$, using the cut-off values for group membership from sample $\mathcal{I}1$. This ensures that each step is conducted on a new partition of the data. Afterwards, I cross-fit, so that the whole procedure is repeated twice, but each subsample acting in a different role.

I optimize the cut-offs for study speed based on two variables, study credits and subjects chosen. For both, I record how many credits (and subjects) a student earned in the five semester pre-screening period. Recording for five semesters or computing the average per year or semester is equivalent and only changes the scale. Using both study credits and chosen subjects improves predictions because they suffer from different types of measurement error. Intuitively, number of subjects is a noisy measure of study speed, since some subjects are worth more study credits than others. At the same time, study credits suffer from under-reporting.[15] Combining the information in these two variables helps to pin-point students actual progression in their degrees, and thus their propensity to be exposed to the reform. A student with a moderate amount of study credits may be too slow or of sufficient speed and have some unreported courses. Taking into account the number of subjects they took can help to counterbalance that, while a correspondingly low number of subjects may support that the student was indeed to slow to be exposed to the reform.

I estimate the first stage for each combination of lower and upper cut-offs of study credits and lower and upper cut-offs of number of subjects and record the p-value. Then I pick the combination of cut-offs that correspond to the lowest p-value. The first stage contains the same control variables used in the main specification of this paper, see Equation 7. In line with the evidence on a u-shaped relation between study speed and the first stage for study credits by quintiles, I find that p-values drop most sharply when increasing the lower cut-off from the lowest values and the upper cut-off from the highest values. P-values proceed smoothly to the minimum region. For high lower cut-offs and low higher cut-offs p-values increase again due to loss in sample size when tightening cut-offs further in the core sample with high first stage coefficients. The plots below illustrates how the p-value of the first stage changes with different study speed cut-offs. The procedure is equivalent when testing for subjects and study credit cut-offs simultaneously.

In the validation exercise, optimal cut-offs in one sample are used to test the significance of the first stage by group; optimal cut-offs in the initial sample become sub-optimal cut-offs in the next sample. This captures actual heterogeneity in how well study speed predicts the strength of the first stage. Below, I show the first stage estimates per group in sample $\mathcal{I}_{N+1}$ where the cut-offs were optimized in sample $\mathcal{I}_N$. The optimal cut-offs from the training sample do not pick up the first stage perfectly in the sense that some cut-offs optimal in one are too wide or narrow in the other, so that some first stage coefficient remains in another group. In my data, the first stage coefficient estimated in sample $\mathcal{I}3$ in the too slow group with cutoffs from sample $\mathcal{I}1$ is significant. Therefore, in $\mathcal{I}2$, both the exposed and the the too slow group (with cut-offs from sample $\mathcal{I}1$) will be selected for estimation.

Using Hazard and Löwe's procedure requires to run the analysis with the selected group separately for each sample and then average the coefficients in a final step. Pooling the sample after selection gives the same coefficients. Regarding inference, their paper proves

[15] Subjects are not affected by under-reporting in the same way. There are several subjects in the data for which zero study credits are recorded, while no subject in Norway at that time was worth less than 5 credits.

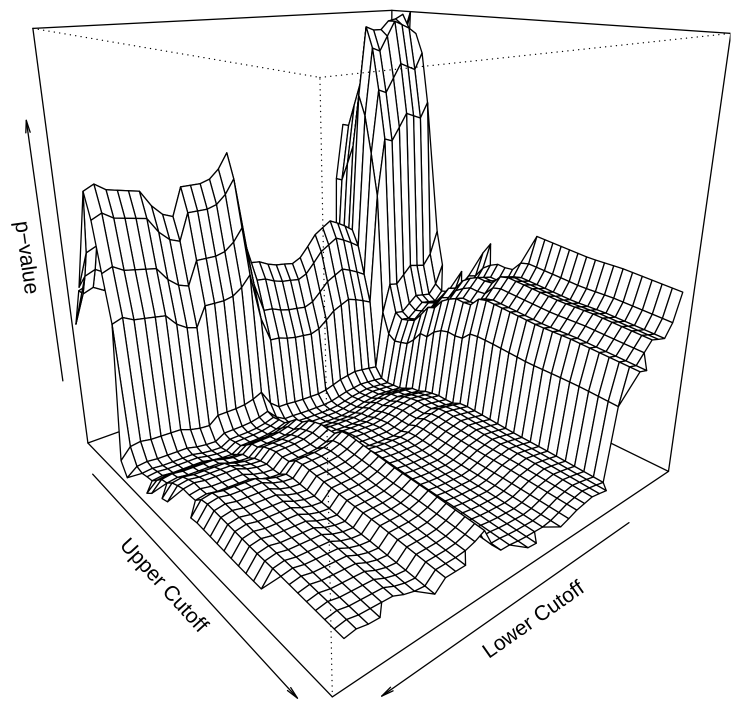


Figure 8: First stage p-value optimization in one of the three sub-samples

In the two variable optimization with number and subjects and an upper and lower cut-off for each, I pick the lowest p-value in a grid search. This illustration shows changes in the p-value of the first stage in one of the sub-samples as the upper and lower cut-off for the amount of credits in the first five semesters are changed, holding the cut-offs for number of subjects in the first five semesters constant at their optimal value.

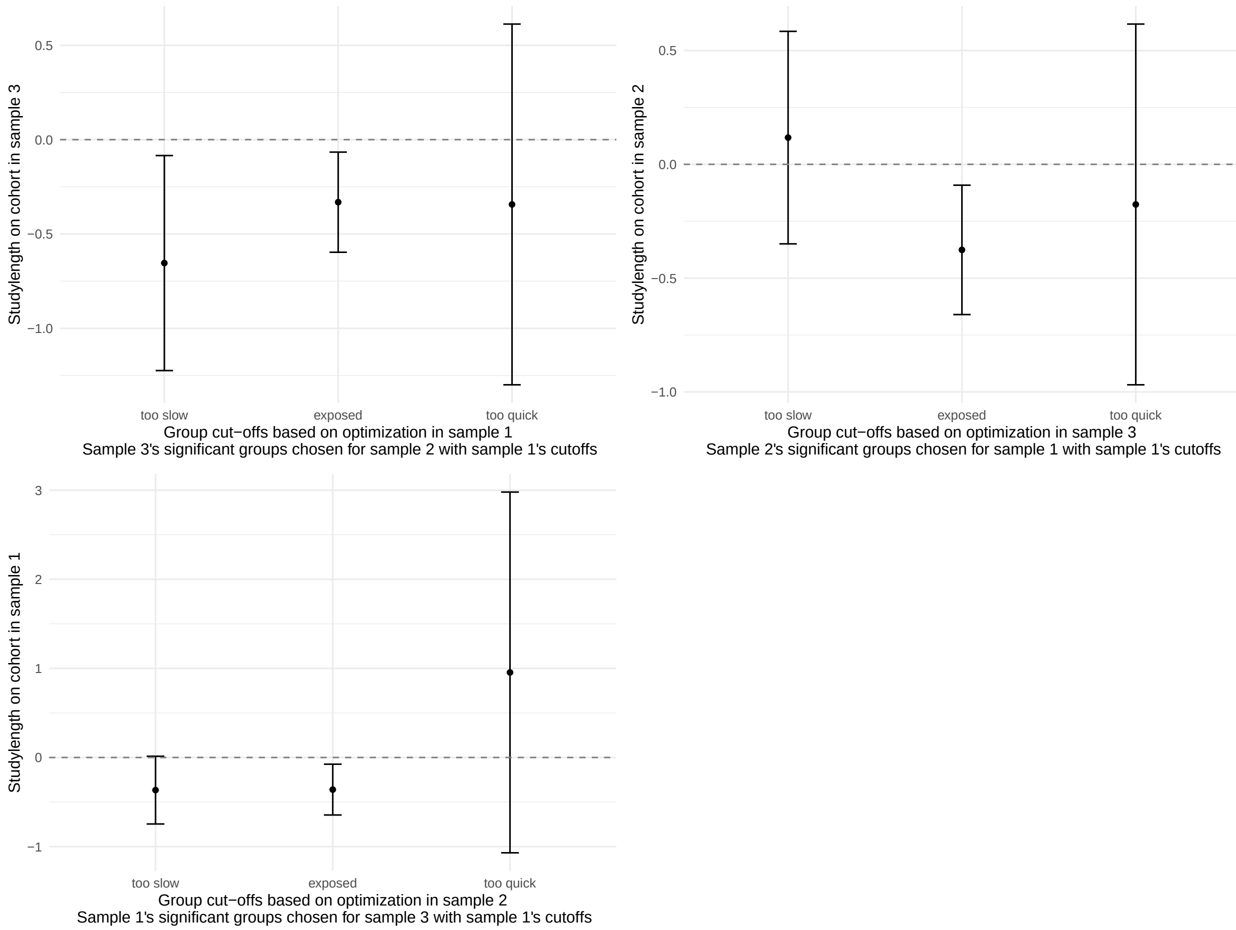


Figure 9: Group selection in each subsample based on the previous samples' optimal cut-offs.

that standard errors have the correct coverage for the sample-by-sample estimation and averaging procedure, but not separately for the pooled case. I use this procedure for every result presented in this paper, but present descriptive statistics on the pooled sample to verify that the selection procedure did not change balancing of covariates by cohort.

In Table 5 below, I compare samples based on different selection rules. Applying the 3-stage cross-fitting method explained above reduces the pooled sample about 25%. Compliance is higher, with a first stage coefficient of 0.34. In the observations excluded during the cross-fitting procedure, there is no significant first stage, although the sign remains the same. Balancing holds in the cross-fitted sample as shown below.

Method	N	FS Coef	St error	F
Benchmark samples				
Fully unrestricted sample	8720	0.13	0.069	3.8
Baseline sample	7230	0.25	0.075	11.3
Bologna degree holders	2906	0.43	0.106	15.4
Cross-validated sample				
3-cross-fit, 2 cut-off vars	5319	0.34	0.08	16.8
Excluded from cross-fit sample	1911	0.065	0.15	0.16

Table 5: Cross validated sample properties vs. benchmarks

[a] P-values are optimized for the combination of lower and upper cut-off(s) of the study credit (and number of subjects) variable(s).

Variable	Mean 2000	SD 2000	Mean 1999	SD 1999	p-value
Mothers' education	4.58	1.72	4.54	1.7	0.32
Fathers' education	4.97	4.96	4.97	4.96	0.84
Female	0.64	0.48	0.64	0.48	0.93
Nr of siblings	2.16	0.87	2.16	0.86	0.99
Age at enrollment	20.72	4.01	20.65	4.03	0.53
Natural science student	0.14	0.35	0.13	0.33	0.08
Enrolled in spring semester	0.18	0.38	0.18	0.38	0.73
N	2751		2574		

Table 6: Balancing of observables across enrollment cohorts

5 Results

5.1 OLS benchmark

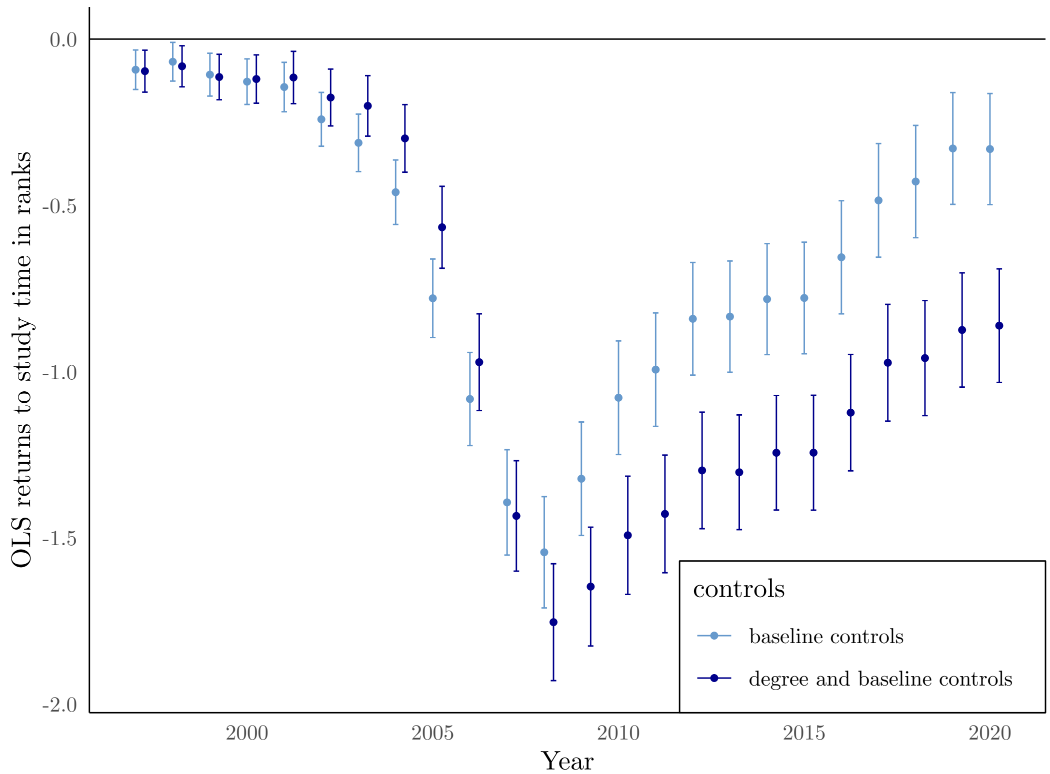
Study length is not associated with higher earnings when students choose themselves how long to study. The light blue coefficients in Figure 10 show the time-varying human capital coefficients γ_t from Equation 1. Chosen study length in this case is both due to students studying slowly and students studying a lot, since degrees vary freely. The negative selection from being slow exceeds the positive selection into higher level degrees and degree returns as well as human capital returns, at least for the available time frame.

The short run costs of staying longer are highly visible. In the years where most students graduate, between 2006 and 2010, the students that study a bit longer are still in education, leading to lower wages because of the extensive margin difference. Already graduated students enter the labor market and earn full-time wages compared to working few or no hours while studying. The high penalty on studying longer at this time reflects these opportunity costs.

The small but statistically significant coefficients for study time before enrolling in higher education takes place reflects negative selection. Students who will end up studying longer tend to work less already before they enter higher education. This may be lower capacity to work while being in school, but it could also reflect that students who have less economic pressure to work can afford to study long and do not see the need to take up a side-job while still in school. The dark blue coefficients are the γ_t estimated in Equation 2 which includes a control for the level of a degree students eventually obtain. This removes signalling returns and positive selection from the study time returns. Notably, estimates differ mainly in the long run. The penalty for endogenous study time without controlling for degrees decreases over time, with the positive trend not ruling out that study time may start to pay off eventually. Study time when controlling for degree seems to settle at a penalty of around one rank per year of study. This is clearly expected, since the remaining variation in study time only reflects being slow or indulging in consuming courses that do not fit.

From an employers perspective, study time - and especially longer study time for a given degree - may be considered as a bad sign, potentially revealing lower ability and/or motivation. This is important to keep in mind when interpreting returns to exogenously-induced, degree-invariant study time. Even when a student is not negatively selected as in the genuinely “slow” student group, employers may still pay lower wages because of their knowledge of overall selection into slowness. Candidates ability to disclose the reason for their delay may determine whether employers discount their expected ability.

Figure 10: OLS benchmark: returns to chosen study time



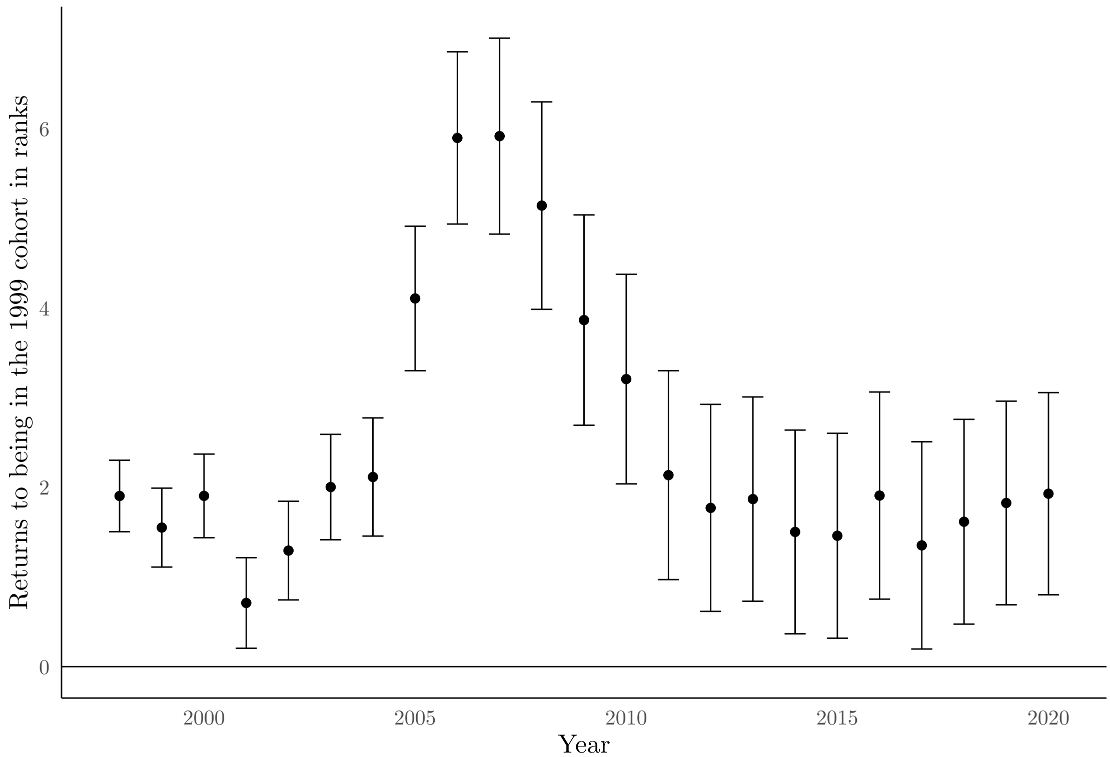
Notes: Coefficients in light blue show γ_t coefficients from estimating Equation 1, and coefficients in dark blue show the corresponding γ_t from Equation 2. Years t range from 1997 to 2020. Equation 2 includes flexible controls for the level of the obtained degree. Study time is measured in years. Both regressions include controls for parental education, gender, natural science, enrolling in spring and age at enrollment. Enrollment cohorts are pooled and a dummy for enrollment cohort is included. Everything is estimated in the baseline sample.

Examining endogenously chosen study credits rather than study time presents a different perspective, see Figure A.3. Pre-trends reveal no earnings disparity between individuals accumulating more or fewer study credits. In the medium term, the opportunity-cost earnings penalty remains consistent, as accruing credits requires time (the correlation of study time and study credits is 0.67). Nevertheless, in the long term, endogenously chosen study credits yield a return of 0.02 ranks per credit, equating to a 1.2 earnings rank increase for a full year's 60 normed credits. When accounting for degrees, this advantage vanishes, suggesting employers disregard degree-unrelated study credits.

5.2 Calendar time reduced form

Estimating the returns to enrolling one year earlier clearly establishes the year advantage of the 1999 cohort. Before enrolling and being one year older on average, 1999 students are

Figure 11: Reduced form effects in calendar time.



Notes: The plot shows the set of estimated γ_t coefficients from Equation 3. This captures for each year t the earnings differential for being in the 1999 cohorts in ranks and enrolling in the fall semester. All regressions include controls for parental education, gender, natural science, enrolling in spring and age at enrollment, as well as the interaction of enrolling in spring and being in the 1999 cohort. Estimation on the baseline sample.

already placed two percentile ranks higher in the earnings distribution. Differences diminish while both cohorts are at university and nobody has graduated yet. Since the 1999 cohort enrolls one year earlier, they also graduate 9 months earlier despite staying on average 3 months longer in education. As the empirical cdf of the first stage indicated, see Figure 5, the difference in students who already graduated across cohorts peaks around 8 years after enrollment. In those years, a larger fraction of the 1999 cohort has graduated. This extensive margin effect explains the large earnings differences in the years 2007-2011. In the long run, when virtually all students have graduated in both cohorts, the earnings advantage stabilizes around 2 earnings ranks. This specification does not allow to infer whether these are returns to university-acquired human capital or whether the year advantage of enrolling one year earlier simply persists. Since both effects are positive, this estimate represent an upper bound for the return to university-acquired human capital.

5.3 Main specification: Placebo Study

In the main specification, I separate between university-acquired human capital and the year advantage by estimating the year advantage separately. The 2001 and 2002 enrollment cohorts did not receive extra human capital due to the Bologna reform because they learned about the reform so early in their studies that they did not acquire any extra study credits. Tracing how earnings differences for this placebo cohort pair evolves differences out the human capital portion in the earnings advantage of the 1999 cohort.

The coefficient plot below shows γ_t estimates from Equation 3 separately for the 1999/2000 sample and the 2001/2002 sample, where $Cohort_i$ is 1999 and 2001 respectively. Since Equation 3 includes the Cohort*Spring interaction, this effectively displays the effect for fall students. t is set as the enrollment date of the older cohort in each cohort pair. The difference between those two coefficients at each relative time corresponds to the γ_{t-r} coefficients from Equation 4, but without the interaction with spring semester. The relative year in the plot corresponds to the values of $t - r$.

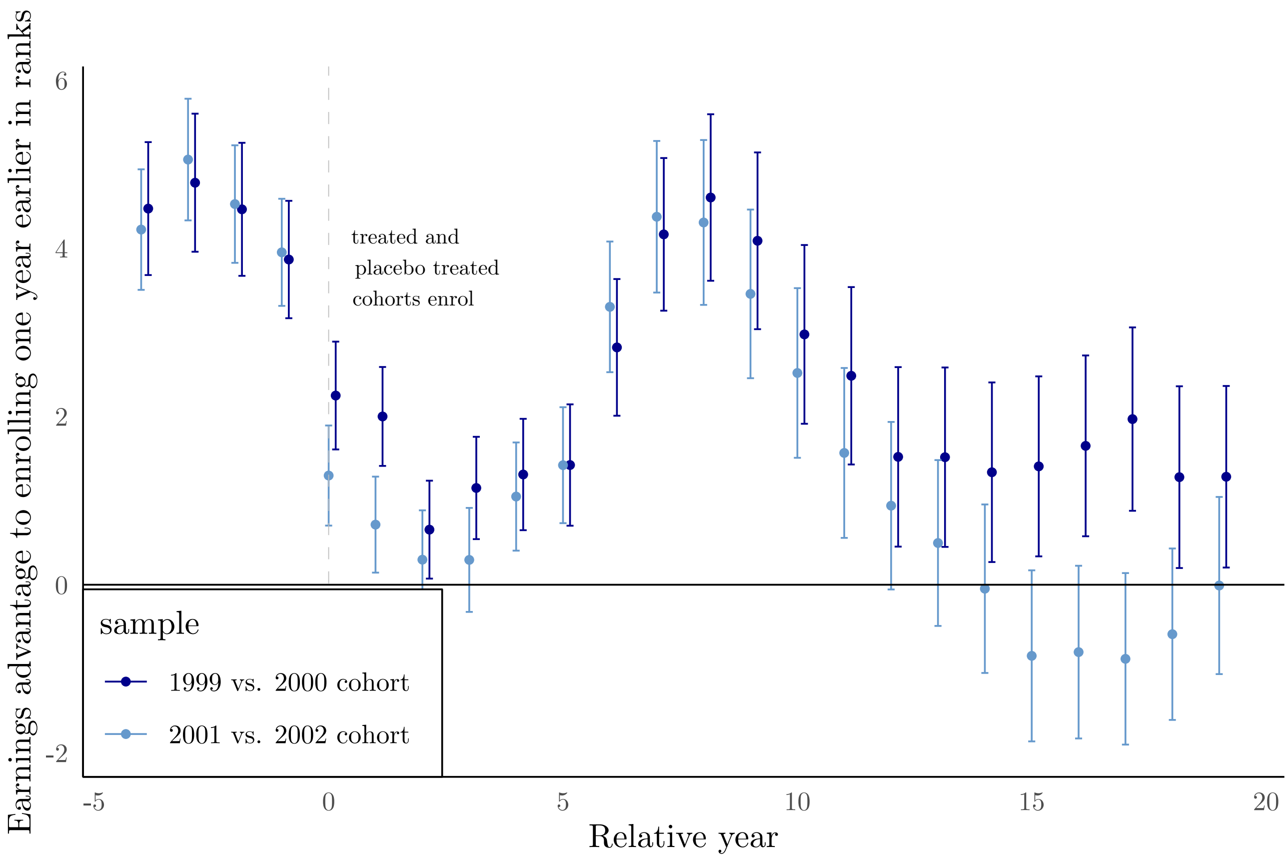
The four pre-enrollment years, $t - r < 0$, indicate that the two cohort pairs share the same pre-enrollment earnings differentials. In both cohort pairs, the older cohort earns more than the younger cohort, with exactly the same pattern. During higher education, both cohorts pairs' earnings differences follow each other closely and both cohort pairs share the earnings boost for graduating earlier. Since the 1999 cohort stays slightly longer in higher education than the 2001 cohort relative to their counterparts, one could have expected the 1999 cohort to do slightly worse around 6-10 years after enrollment. There is little evidence for that.

As late as around 15 years after the older cohort in each pair enrolls, earnings diverge. The numerical difference between the differences is large, with a university-acquired human capital return of 2 to 2.5 ranks for substituting on average three month of education and experience.[16] At the same time, estimates for both cohort pairs are quite noisy.

The cohort coefficient baseline terms in repeated relative-time cross sections where cohort is interacted with spring — the ones displayed in the coefficient plot — are equivalent to estimating the effect on fall students only. Fall students are more exposed to the reform than spring students because they start earlier than spring students. They have more time to collect overshooting credits by 2003, while spring students would need to be extraordinarily

[16]This specification has to use reduced form estimates, since there is no first stage to estimate 2SLS on for the 2001/2002 cohort. 2SLS for the placebo cohorts results in extremely wide confidence intervals with around 100 ranks because of the extremely weak/non-existing instrument for that cohort pair.

Figure 12: Human capital and placebo returns



Notes: The plot shows the set of estimated γ_t coefficients from Equation 3 separately for the treatment sample, consisting of the 1999 and 2000 cohort, and the placebo sample, consisting of the 2001 and 2002 cohort. All regressions include flexible controls for parental education, gender, natural science, enrolling in spring and age at enrollment, as well as the interaction of enrolling in spring and being in the 1999 cohort.

quick to be treated.[17] Table 7 below presents pooled diff-in-diff estimates for the pre-trend period as well as for the last 10 and 5 years of the data (Equation 5). I show results both for the full sample and for fall semester students only, where the γ coefficient for fall semester students coincides perfectly with the difference of the estimates in the coefficient plot (over several years).

I do this separately for different time windows. Pre-enrollment, $t-r<0$, both the treated and placebo cohort pairs year advantage should follow the same earnings trajectory, so the γ coefficient for the $Treat_i * Advant_i$ term should be zero. Aiming to estimate the long-run

[17] For the equivalent coefficient plot of the difference of year advantages between cohort pairs, but for all students rather than fall students, see Table A.4.

returns to acquiring more human capital, the last years should be most informative. More and more students graduate as time goes by, reducing the graduation advantage of both older cohorts so that earnings in the labor market rather than graduation timing start to dominate. As explained above, individuals earnings tend to stabilize and converge towards their earnings potential relatively late in life. The main estimates pool the last five available years, balancing long-run earnings potential with reaping efficiency gains from pooling years. At the same time, more than 80% of students finish 10 years after enrollment. To trace when the human capital advantage starts to become relevant, I present alternative estimates from the 10th year after the older cohorts enrolls; these are the last ten years in the sample.

	Baseline sample			Fall students only		
	(1)	(2)	(3)	(4)	(5)	(6)
Year advantage * Reform cohorts	0.052	1.045	1.522*	−0.037	1.317*	2.018**
	(0.375)	(0.598)	(0.656)	(0.396)	(0.638)	(0.696)
Year advantage	4.243***	0.390	−0.470	4.427***	0.458	−0.529
	(0.259)	(0.424)	(0.466)	(0.266)	(0.441)	(0.481)
Reform cohorts	−1.196***	−1.037*	−1.359**	−1.324***	−0.959*	−1.525**
	(0.269)	(0.431)	(0.475)	(0.277)	(0.448)	(0.491)
Spring student	1.375***	0.776	0.977*			
	(0.296)	(0.458)	(0.497)			
Baseline controls included	Yes	Yes	Yes	Yes	Yes	Yes
Time period FEs	Yes	Yes	Yes	Yes	Yes	Yes
Included time periods	<0	10-19	15-19	<0	10-19	15-19
R^2	0.296	0.037	0.053	0.256	0.054	0.049
Adj. R^2	0.295	0.037	0.052	0.255	0.054	0.049
Num. obs.	56602	138607	69208	48359	118396	59127

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7: Placebo diff-in-diff multiple period estimates

Year advantage is an indicator for enrolling earlier in each cohort pair (1999 or 2001), Reform cohorts are the enrollment cohorts 1999 and 2000. The interaction term reflects the additional earnings advantage the 1999 cohort enjoys relative to the 2001 cohort. Models 1-3 show estimates for the full baseline sample, models 4-6 show estimates for the sample of fall students. All regressions include flexible controls for age at enrollment, gender, natural sciences and mothers' and father's education level as well as fixed effects on the relative time period. Standard errors are clustered at the individual level.

Table 7, shows estimates of the returns to university-acquired human capital (γ from Equation 5) for the pre-enrollment period, 10 years since enrollment of the older cohort and onward, and the last five available years. The first three columns are for all students, the second three columns for the sample of fall students. For both samples, there is no significant interaction effect, mirroring the matching pre-trends from the coefficient graph. Looking at mostly-post graduation (10-19 year columns) and fully post-graduation (15-19 year columns) effects, 1999 students do better than 2001 students, where both are evaluated relative to their non-year-advantage counterpart 2000 and 2002. The benefits of university-acquired human capital materialize fairly late. 10 years after the enrollment of the older cohort in each pair, while some students still study, effects are borderline significant. In the latest available earning years, the effect stabilizes. As expected, effects are more pronounced for the sample of fall students relative to the full sample, since spring students were less likely to be exposed to the reform.

Although pooling years relative to the year-by-year specification shown in the coefficient plot increases efficiency, estimates are still noisy. With a standard error of 0.65, earnings' effects of substituting 3.5 months of experience with 3.5 months of education lie between 0.25 and 2.7 (full sample, last five years). Estimates in the upper end of that interval are larger than what seems reasonable for 3.5 months more of higher education without a degree to show for it and with the experience-substitution accounted for. Naively scaling up to a full year gives an interval of 0.75 and 9.2 for an induced extra year of higher education. While effects in the lower end may seem plausible, a simple comparison with the raw earnings difference for being female — -5.6 ranks — and enrolling in the natural sciences — 6.7 ranks — indicate that the confidence interval is too wide to give new information compared to a very broad set of possible priors one may hold about the importance of university-acquired human capital for wages. Although uninformative for the upper bound, negative effects of substituting education with experience can be ruled out.

While I estimate each cohort pairs earnings' in calendar time, differences between cohorts are measured in time since the older cohort enrols. If those differences can be compared across labor markets, one could estimate Equation 5 also for log earnings as an outcome variable. While the same pattern of findings persists, estimates are more noisy in logs, see Table A.3 in the appendix.

Inspecting the coefficient plots for each cohort-year-advantage separately indicates that part of the difference effect comes from point estimates for the 2001/2002 cohorts being (non-significantly) lower than zero. One may imagine that the year advantage becomes less important over time so that the earnings of the 2001 and 2002 cohorts converge, but there is no reason to believe that the one-year-older 2001 cohort should be worse off around the age 35-40. For that reason, I contrast findings with the 2SLS event study 2SLS, which does not rely on placebo cohort pairs.

5.4 Event study 2SLS

Estimating the effect of university-acquired human capital in time since enrollment removes the year advantage directly (rather than addressing it through placebo cohorts' year advantage as in the main specification). This specification can be estimated using the cross-fitted test-and-select samples.[18] However, comparing incomes directly and systematically across different income years is problematic. The coefficient plot 13 below shows $\gamma_{t-\tau}$ estimate from Equation 6.

Pre-trends in earnings measured in relative time are negative. The effect of a not-yet-realized year of higher education should be zero if percentile ranks were a sufficient adjustment to compare the cohorts systematically across different income years. The negative pre-trend indicates that students' earnings in adjacent years at a given time before enrollment and measured in ranks are not the same. One potential explanation is that certain pre-enrollment calendar years exhibited particularly advantageous or disadvantageous conditions for young workers. Measuring in enrollment time, these good or bad years would show up in different cohorts' earnings at different times. Overall, the fact that earnings measured before enrollment differ although they capture earnings at the same ages, suggests that measuring earnings' in enrollment time may not reliable.

Other than that, the earnings effects show the expected time dynamics. Point estimates tend to be negative, albeit not significant so, around the time where 2000 cohort students are more likely to be in the labor market since they on average graduate 3.5 months earlier. In the long run, returns to an extra year of higher education are estimated to be around 2 ranks, but not significant. This is around 50% lower than in the main specification (when scaling the placebo estimates to one year). Estimating on the test-and-select cross-fitted samples reduces the standard error by 16%, but efficiency does not improve sufficiently to reject the null hypothesis.

Moving to pooling years in the same groups as for the main specification - pre-enrollment, last 10 years and last 5 years - improves efficiency. Pre-trends remain significantly negative and point estimates for the time periods are estimated with a tighter overall confidence interval. But even the most efficiently estimated model - the last 5 years for the test-and-select model - has a p-value of 0.12 so that I cannot reject the null. Comparing the placebo specification and the relative time specification, both indicate qualitatively comparable returns to university-acquired higher education, but the placebo specification tends to be more efficiently estimated.

[18] In principle, one could use the cut-offs and included groups determined in the 1999/2000 sample for the 2001/2002 sample and obtain a cross-fitted version of this sample as well. However, the measurement of study credits improved vastly in 2003 and since the screening period for those two cohorts include 2003, all cut-offs will identify a systematically slower group of students. In addition, one cannot estimate the placebo model in 2SLS because there is no first stage, and the efficiency gains from the test-and-select procedure only apply for 2SLS, not the reduced form.

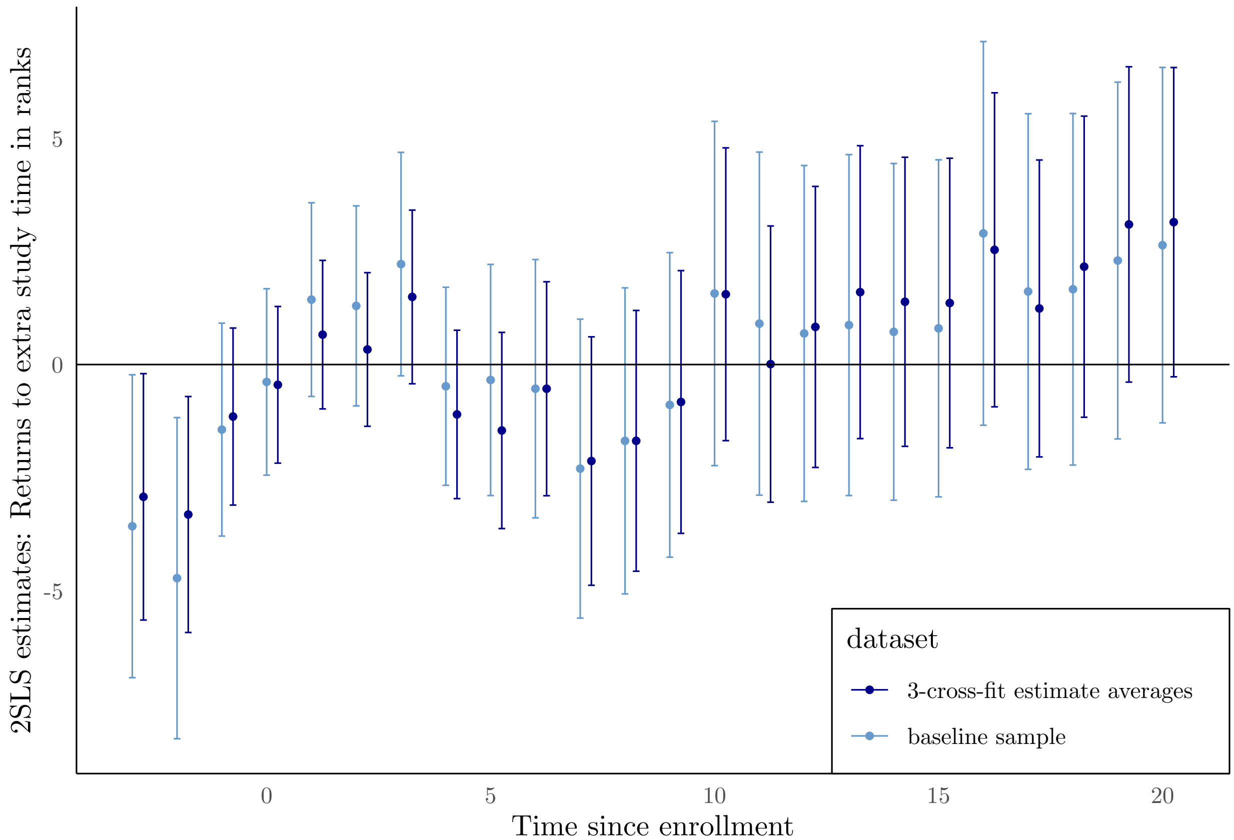


Figure 13: 2SLS in enrollment time

Each coefficient represents a $\gamma_{t-\tau}$ estimate from Equation 6 for the effect of an extra year of (instrumented) study time on earnings' ranks at the indicated time since enrollment. Coefficients in light blue are obtained from estimation on the baseline sample, and coefficients in dark blue from estimation in the three test-and-select samples. All regressions include flexible controls for age at enrollment, gender, natural sciences and mothers' and father's education level.

	Baseline sample			test-and-select 3-cross fit samples		
	(1)	(2)	(3)	(4)	(5)	(6)
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Study time	-3.376^{*}	1.347	2.006	-2.520^{*}	1.750	2.518
	(1.366)	(1.734)	(1.900)	(1.035)	(1.467)	(1.639)
Baseline controls included	Yes	Yes	Yes	Yes	Yes	Yes
Time period FEs	Yes	Yes	Yes	Yes	Yes	Yes
Included time periods	<0	16-20	11-20	<0	11-20	16-20
R^2	0.16	0.019	0.011	0.185	0.013	0.007
Adj. R^2	0.163	0.018	0.010	0.182	0.012	0.004
Num. obs.	20963	68478	34170	15433	50392	25140

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 8: Event study 2SLS multiple period estimates

Study time is instrumented with cohort (binary for 1999/2000), interacted with being a spring student. Models 1-3 show estimates for the full baseline sample, models 4-6 show estimates for the test-and-select samples. All regressions include flexible controls for age at enrollment, gender, natural sciences and mothers' and father's education level as well as time since enrollment fixed effects. Standard errors are clustered at the individual level.

6 Robustness and Extensions

6.1 Partial returns to education

In the literature on returns to education, it is common to control for experience in a setting where education is arguably exogenous (Arteaga, 2018; Carneiro et al., 2011; Bhuller et al., 2017). Returns after controlling for experience can be referred to as the partial returns to education. Partial returns are supposed to reflect how much better a student does if receiving more education, holding everything else constant. Experience is a powerful predictor of earnings after graduation. In the logic of this literature, controlling for experience means moving a step closer to estimating the partial returns to education.

Experience is a a post-treatment variable that is directly affected by the treatment. Staying longer in education decreases experience at any given time point, both in calendar and enrollment time. Earnings of the compliers in the 2000 cohort – not receiving any extra education in line with their instrument value – are used as an estimate of counterfactual earnings for the compliers in the 1999 cohort. While the compliers in the 1999 cohort study longer, the compliers in the 2000 cohort leave university and gather (potential) experience. In other words, having less experience is a direct consequence of the treatment; this makes experience a bad control (Angrist and Pischke, 2009).

One may want to argue that when controlling for experience, one removes a side-effect of higher education that by definition should not be included in the partial effect. In that sense, one could consider removing part of the effect as desirable, since the total effect of more higher education (for this study: more human capital) is not of interest. One could call this partial effect of higher education the returns to higher education net of opportunity costs. In that logic, potential experience captures the netting out of other effects more broadly than measured experience.

According to Angrist and Pischke (2009), estimating this partial effect does not have a causal interpretation. Assume that the cohort instrument is valid in the sense that it is independent of potential outcomes and independent of potential experience. Then cohort-induced earnings differences in earnings at a certain level of experience (imagine experience as a binary) can be written as:

$$E[Y_{1999}|Exp_{1999} = 1] - E[Y_{2000}|Exp_{2000} = 1]$$

By the independence of the cohort of potential earnings and potential experience, this is equivalent to

$$E[Y_{1999} - Y_{2000}|Exp_{1999} = 1] + E[Y_{2000}|Exp_{1999} = 1] - E[Y_{2000}|Exp_{2000} = 1]$$

The first term corresponds to the direct causal effect of earnings differences between the cohorts, that is, returns to cohort-induced human capital (assuming exclusion holds), for a certain level of experience. This is the partial effect that one may aim for, the pure returns to human capital, disregarding the opportunity costs due to less experience. The second term represents selection bias, that is, people with the same experience levels from different enrollment cohort may not be comparable even for the potential outcome where no education is induced at the cohort level. To estimate partial returns to human capital, one would need to argue why the selection bias should equal zero.

Following Pearl (2009) and Cinelli et al. (2022), the first term above corresponds to the controlled direct effect (partial returns), which can be estimated without bias if there is no confounder affecting both experience and earnings (and assuming there is a remaining direct effect of cohort on earnings) This is equivalent to saying that the selection bias is zero. In this design, factors determining non-exogenous study length (the remaining variation after regressing study length on cohort) are relevant confounders. Experience and non-exogenous study length are highly negatively correlated and non-exogenous study length is likely to affect outcomes. At the same time, endogenous study length likely depends on earnings-relevant factors as well. A potential candidate would be ability, governing both how quickly a student can study (and in turn experience at a given time) and earnings potential. Or in Angrist and Pischke (2009)'s terms above, students with the same experience level from different cohorts would differ in their earnings even if they had not received different amounts of human capital due to their cohort membership.

The problem can also be formulated as a violation of instrument independence. Conditional on experience, potential outcomes may not be independent of the instrument anymore. Without experience controls, observable are balanced across cohort, which is evidence in favor of instrument independence, see Table 1. Regressing cohort on the balance variables for each discrete value of experience separately (rounding experience to full years) gives F values ranging from 0.3 to 1.4, with jointly significant imbalances most common for years with untypical experience values (very slow or quick students) while for the most common years, balancing typically holds. Parental education tends to be most likely to predict cohort membership in experience cells, which is troubling since parental education also predicts earnings. Keeping in mind that selection bias is likely present, I still compute the experience-controlled returns to human capital, since this is an important benchmark specification in studies relevant to this paper.

If assuming for a moment that the selection bias discussed above is zero, controlling for experience should remove opportunity costs from the return to higher education and give higher returns compared to the total effect model. Arteaga (2018), for instance, estimates a difference-in-difference model for the effect of a curriculum reduction at the economics and business departments at the Universidad de Los Andes, where students from non-affected economics departments at other universities serve as the control group. The difference-in-difference interaction term captures how much less students earn after the reform if they earn a shorter economics degree relative to other students after the reform whose degree was not affected. Controlling for experience in this setting discounts earnings of the treatment post-group, since they graduate earlier than all other groups. In any given income year, they will have more experience, although they enrolled at the same time as their control-group counterparts. Arteaga (2018) reports 12-15% higher earnings per year of experience across varying specification. Students treated with a shorter curriculum gather mechanically half a year extra of experience, and are estimated to receive 16-22% lower earnings due to the reform. This suggests controlling for experience for those gathering more experience when denied the longer study program may be influential.

The most comparable "partial effect" or direct controlled effect estimate in my setting is the main placebo specification, pooling years. Estimation is feasible in contrast to the 2SLS event specification.[19] While cohort predicts study length, this only accounts for a tiny bit of the variation. The 1999 cohort studies 3.5 months longer on average, but (post-screening) study length has an average of 5.56 years with a standard deviation of 3.17 years. The correlation between experience and cohort length in the cross-section is -.11, and reduces further to -.05 when pooling income years. The exogenous part in the variation in study time cannot be used to estimate the coefficient on experience; adding experience as a control amounts to estimating the earnings-experience profile using endogenous variation in study time as well as within-person earnings variation as experience increases.

To estimate partial returns to university-acquired human capital and compare estimates to the main specification, I vary Equation 4 by estimating one time-independent parameter for all variables except for time fixed effects α_{t-r}, and add experience controls. The issue with selection bias at different levels of experience is argued for binary experience and extends to experience factors. I still report on estimating with linear and quadratic experience terms, since this makes it easier to benchmark against the literature. Adding experience factors instead makes little difference.

$$Y_{i,t-r} = \alpha_{t-r} + \theta Treat_i + \delta Advant_i + \gamma Treat_i * Advant_i + \beta X_i + \kappa E_{t-r} + \lambda E^2_{t-r} + U_{i,t-r} \quad (10)$$

The γ coefficient on the placebo diff-in-diff term reflects the partial returns to education since experience-related opportunity costs are controlled for; if people with the same experience in different cohorts were comparable in the absence of receiving more education through their cohort status, which, as discussed above is unlikely to hold.

Pre-enrollment, nobody has obtained any experience when defining experience as potential experience measured as years since graduation. Therefore, I do not report the experience-controlled specification for the pre-enrollment period. Table 9 shows estimates for the partial returns to university-acquired human capital. Compared to the opportunity-cost corrected PRTE estimates, point estimates increase between 10 and 20% across all specifications, while standard errors decrease minimally since the experience controls account for some variation in earnings ranks. Differences tend to be larger for the estimation using the last ten years,

[19] Estimating partial returns to university-acquired human capital by controlling for experience cannot be estimated using 2SLS, since it is not identified. In any given cross-section, potential experience and study time are almost perfectly negatively correlated (-.98) because studying longer means graduating later. The correlation in the cross-section is not exactly -1 because I measure study time as total length of education, removing gap semesters where student do not study, but the potential experience counter starts when students graduate (or drop out) for good. Experience has to be added to the first stage when wanting to control for it in the second stage, effectively adding the first stage dependent variable again on the right hand side.

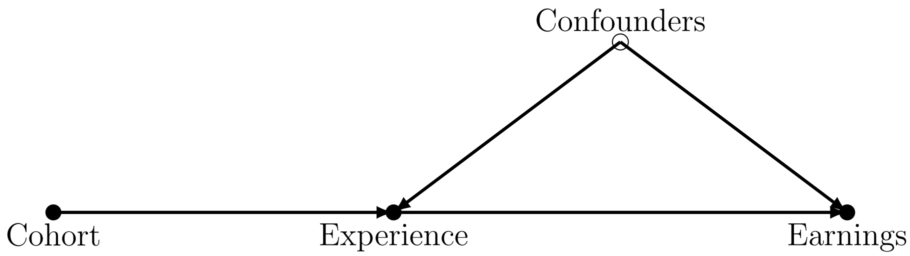


Figure 14: Conditions for bias and unbiased controlled direct effects

	Baseline sample		Fall students only	
	(1)	(2)	(3)	(4)
Year advantage * Reform cohorts	1.251*	1.615*	1.429*	2.129**
	(0.591)	(0.651)	(0.631)	(0.691)
Year advantage	−0.400	−0.732	0.103	−0.846
	(0.419)	(0.464)	(0.439)	(0.479)
Reform cohorts	−0.873*	−1.267**	−0.842	−1.406**
	(0.426)	(0.471)	(0.443)	(0.487)
Experience	1.663***	2.063***	1.342***	2.027***
	(0.053)	(0.139)	(0.069)	(0.151)
Experience^2	−0.073***	−0.110***	−0.088***	−0.107***
	(0.004)	(0.009)	(0.006)	(0.010)
Spring student	0.805	1.013*		
	(0.452)	(0.493)		
Baseline controls included	Yes	Yes	Yes	Yes
Time period FEs	Yes	Yes	Yes	Yes
Included time periods	10-19	15-19	10-19	15-19
R^2	0.079	0.066	0.070	0.064
Adj. R^2	0.078	0.065	0.070	0.063
Num. obs.	138207	69008	118326	59092

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 9: Placebo diff-in-diff estimates of the partial returns to education

Variation of the main specification, controlling for experience and experience squared to estimate partial returns to extra human capital. This has only a causal interpretation if people with the same experience are not differently selected across cohorts. Year advantage is an indicator for enrolling earlier in each cohort pair (1999 or 2001), Reform cohorts are the enrollment cohorts 1999 and 2000. The interaction term reflects the additional earnings advantage the 1999 cohort enjoys relative to the 2001 cohort. Models 1-2 show estimates for the full baseline sample, models 3-4 show estimates for the sample of fall students. All regressions include flexible controls for age at enrollment, gender, natural sciences and mothers' and father's education level as well as fixed effects on the relative time period. Standard errors are clustered at the individual level.

since there are many recent graduates in the sample where experience profiles are still very steep. Point estimates are significant at least at the 5% level when controlling for experience. Overall, this indicates that the price graduates pay for staying longer in education matters for their net returns to university-acquired human capital.

6.2 Bounds for degree effects

Although the Bologna reform in 2003 was supposed to ensure that students can only graduate with the new degree types – bachelor's and master's – some students receive the old degrees[20]. Students may study so quickly that they graduate before the reform is implemented, or they may collect a combination of at least 180 credits that does not map into a bachelor degree and be allowed to finish a *cand.mag.* within two years.

Students that earn old degrees may earn different salaries compared to students with new degrees because employers may trust the old degrees more or because employers are not aware that students' enrollment date governs which degree they earn and interpret the degree as a signal. I cannot simply control for degree types, because the quality of old degree holders likely varies across cohorts. Students in the 2000 cohort need to study exceptionally quickly to still earn the old degrees. Old degree holders from both cohorts are positively selected, and students in the 2000 cohort particularly so.

To remove degree effects from the estimates, I treat the earnings of old degree holders as missing and construct non-parametric bounds on the earnings they may have had had they instead earned the new degrees (Manski, 2009; De Haan, 2017). I construct upper and lower bounds for the counterfactual earnings of old degree holders and re-estimate the main specification with the imputed data for old degree holders for the lower and upper bound in turn. If I make the following assumption (Equation 6.2) about *new* degree holders' earnings, non-parametric bounds on old degree holders' counterfactual earnings recover the effect of university-acquired human capital on earnings.

$$E[Y^0|D = N, Z = 1] - E[Y^0|D = N, Z = 0]$$

For new degree holders (D=N), being in the 1999 or 2000 cohort must not affect wages other than through the human capital effect. In other words, the potential outcome for students from both cohorts, had the 1999 cohort not received human capital, should be the

[20] I define *cand.mag.* and *hovedfag* as old degrees and all other degrees as new degrees. Bachelor's and master's are clearly new degrees. The classification is arbitrary for degrees unrelated to Bologna. These are degrees at community colleges as well as Bologna-unrelated professional degrees at universities, and I classify them as new degrees.

same. This assumption is not innocuous. Slightly fewer students in the 1999 cohort earn new degrees relative to the to the 2000 enrollment cohort, because a larger fraction (2.4%) earn a *cand.mag.* compared to the 1999 cohort (0.3%).

In order to remove any direct effect of (old) degrees from the estimation, I construct non-parametric bounds on the counterfactual earnings of old degree holders. If students that did earn old degree types (O) had received new degree types instead (N), their earnings might have been different. With no assumptions, counterfactual earnings of old degree holders must be between the lowest and highest possible earnings. Since I measure earnings in percentile ranks, no assumption bounds on old degree holders' counterfactual earnings are:

$$1 \leq E[Y^N|D = O] \leq 100$$

Old degree holders would have earnings between the 1st and the 100th earnings' rank had they instead earned the new degrees. In the data, I substitute factual earnings of old degree holders with rank 1 to obtain a lower bounds sample and with rank 100 to obtain an upper bound sample. Students that earned a *cand.mag.* (2.4% and 0.3% respectively) or a *hovedfag* (2.3% and 0.3% respectively) classify as old degree holders.[21] Then I re-estimate the difference-in-difference specification separately for the lower bound and upper bound sample (Equation 5), with the last five relative years pooled as in the main specification. These bounds are uninformative. As Figure 15 shows, assuming the highest possible and lowest possible earnings introduces a lot of variation, leading to bounds on the returns to education between -1.3 and 2.4 earnings ranks for 3.5 months of reform-induced human capital.

In order to tighten these no-assumption bounds, I need to make an assumption about the degrees old degree holders would have earned otherwise. In the sample, students earn degrees of many different *levels*, undergraduate degrees, postgraduate degrees, professional degrees and doctoral degrees. I will assume that old degree holders would have earned a new degree at the same level had they not earned the old degrees. That is, I assume that *cand.mag.* students had earned a bachelor instead and *hovedfag* students had earned a master instead.

[21] It may seem puzzling why roughly the same fraction earns a *hovedfag* as a *cand.mag.* as their last degree, although one can only earn a *hovedfag* if finishing a *cand.mag.* and then enrolling in a *hovedfag* degree before the reform. A common reason for this to happen is that students have credits from outside Norway when they first enroll. Roughly as many students earn a *hovedfag* as a *cand.mag.* as their last degree, although a *hovedfag* requires a much higher speed because many more students earn a cand.mag. first and then proceed to a master so that the cand.mag. is not visible as their last degree.

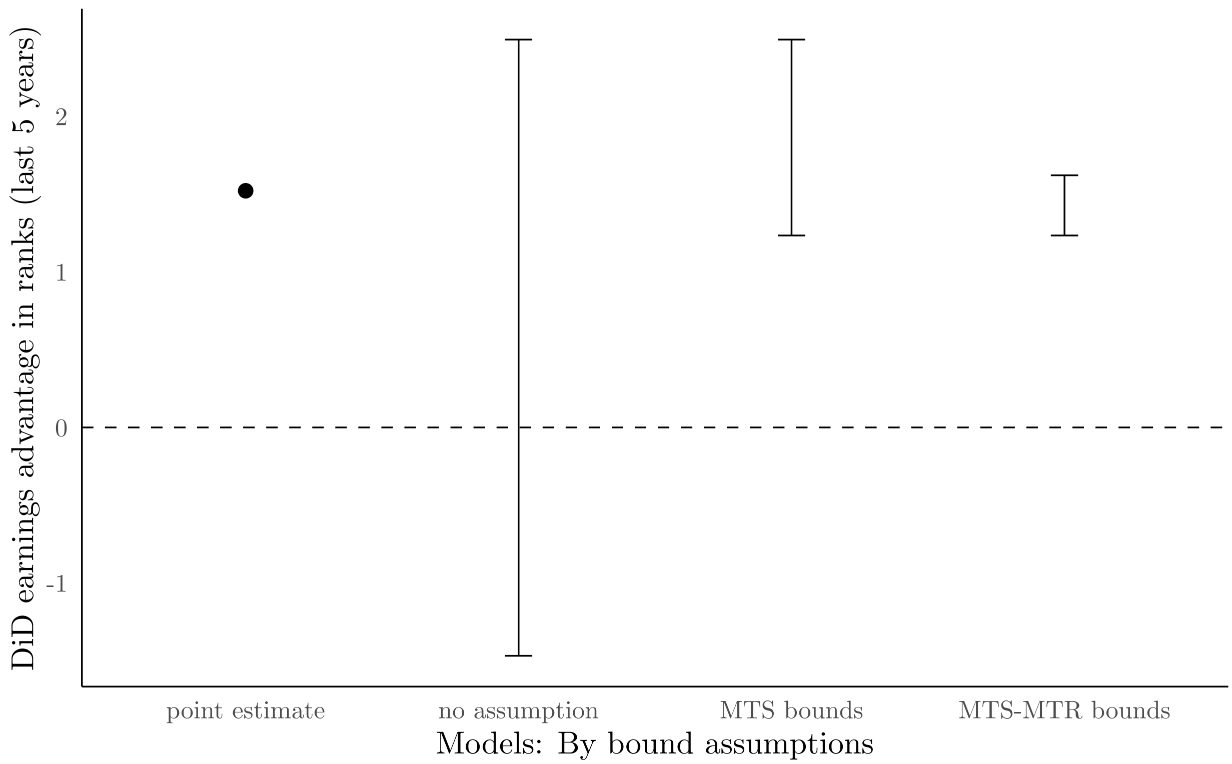


Figure 15: Bounds on returns to human capital, accounting for direct degree effects

The first columns shows the point estimate for returns to Bologna-induced human capital in ranks, this corresponds to the estimate in Table 7, column 3, "Reform * Year Advantage" that isolates returns to human capital. The second column shows no-assumption bounds for the same estimate where earnings for old degree holders (students that earned at cand.mag. or hovedfag degree as their last degree) are set to rank 1 and rank 100 respectively. The third column shows how the lower bound tightens when taking into account positive selection into old degrees (Monotone Treatment Selection (MTS)). The last column shows how the upper bound tightens assuming a Monotone Treatment Response (MTR). Since the old degrees are longer than the short degrees, candidates of equal quality should not earn more if they instead had earned the shorter degree. Counterfactual earnings are computed as the respective degree groups' average earnings in ranks within each income year.

Students that earned the old degrees had to progress in their studies faster than planned for the program. Less able students may not be able to study extraordinarily quickly, so that there should be positive selection into old degrees. Acknowledging positive selection, one can expect that students that did earn the old degrees would not have done worse that students that earned the new degrees had they themselves also earned the new degrees. Even if the old degree holders had earned the new degree they would still be positively selected and should earn at least as much as students that actually earned the new degrees. This gives the following condition on Monotone Treatment Selection (MTS) bounds:

$$E[Y^N|D = N] \leq E[Y^N|D = O] \leq 100$$

The first term corresponds to the observed earnings for new degree holders, while the second term corresponds to potential outcome for old degree holders had they earned the new degree. Re-estimating the upper and lower bound on returns to human capital with this assumption (Figure 15, model 2) leads to a significantly tighter lower bound, with a bound on returns between 1.23 and 2.5 ranks.

Invoking the monotone treatment response (MTR) assumption allows to tighten the upper bound as well. Old degree holders have two advantages in the labor market. They are positively selected and they hold degrees that were designed to last one year longer. [22] If employers were to treat old and new degree holders differently, they may consider the old degrees as a positive signal (if they are not well enough aware of the reform to know that the transition cohorts did not have choice) or they may know that students were positively selected and by had to acquire weakly more human capital to earn these degrees (if they were well informed about the reform). In both cases, they should pay old degree holders higher wages, if anything. I assume that old degree holders if they had instead earned the new degrees, would not have earned more than they actually did when they earned the old degree. Taking the monotone treatment selection (MTS) and the monotone treatment response (MTR) assumption together gives:

$$E[Y^N|D = N] \leq E[Y^N|D = O] \leq E[Y^N|D = O]$$

Using this assumption in the data pushes the upper bound down to the unadjusted point estimate for the returns to human capital. This is as expected; accounting for the fact that some 1999 students earned the (slightly longer) old degree types should not lead to higher

[22] 4 instead of 3 years for the undergraduate degrees, and 6 instead of 5 years for the postgraduate degrees.

estimates. Taking both assumptions together gives a bound on the returns to human capital on earnings ranging from 1.2 to 1.6 ranks. This indicates that returns to degree types does not drive the estimates for the returns to human capital. If one is willing to assume that old degree holders are positively selected and that the longer old degrees were not worse than the new degrees, bounds on the pure human capital effect are tight and very close to the unadjusted estimates.

6.3 Monotonicity concerns

If students in the 2000 enrollment cohort respond to the shortening of the undergraduate degree by enrolling in postgraduate degrees at a higher rate than than students from the 1999 cohort, monotonicity would be violated. This may happen when students have an ideal investment in higher education (or signals) that they would like to pursue, but choose discrete levels of education if employers can only observe those discrete levels, degrees, directly.

Imagine the case where time spend at university increases directly increases human capital and, in turn, productivity. If employers can observe human capital, there would be no reason for universities to award degrees. Students would invest optimally, based on their cost of education and the wage offered for different marginal products and employers would hire and pay according to observed marginal products. If employers cannot observe human capital directly, they may instead pay students for the degrees they obtained.[23] In this framework, students would pick the degree that maximizes the difference between wages and education costs. In a pure signalling framework, imagine instead that studying does not increase productivity, but that the taste for education is correlated with innate ability. As in the human capital case, students will choose the degree that leads to the largest difference between wages education costs. Employers are still paying higher wages for students with higher degrees since higher degree holders have higher innate ability.

In both cases, moving from a system with two relatively longer degrees – *cand.mag.* and *hovedfag* – to two relatively shorter degrees – *bachelor* and *master* – implies that some of the people that marginally preferred the shorter over the longer degree in the pre-Bologna system will choose the longer degree in the post-Bologna system. Some 2000 cohort students that would have been content with a *cand.mag.* in the old system will earn a master in the new system. More precisely, students that would have chosen between four and five years of education if education were continuously observable should switch in their degree level if the undergraduate degree is shortened. Students with an ideal point below four or above

[23] This may happen without any signalling taking place if factors that vary the cost of education are uncorrelated with ability. This would be the case with an intrinsic taste for education that modifies the net cost of education.

five years should maintain their degree choice.

If students in the 1999 cohort behave like pre-Bologna students, 2000 cohort reform-induced master degree holders would be defiers. Imagine two students from the 1999 and 2000 cohort with an ideal point of 4.2 years of education. The 2000 student that would have chosen a *cand.mag.* in the old system (4 years) decides for a masters when forced to choose between a master (5 years) and a bachelor (3 years), while their 1999 counterpart with also 4.2 years in the old system would keep choosing the *cand.mag.*. In that case, the cohort instrument moves students in the 4-5 year ideal point bracket in the wrong direction while students with ideal points above 5 and below 4 years would comply. In the direct pre/post Bologna comparison, changing the degree structure should push some students to comply and some to defy.

However, the 1999 transition cohort does not face the old systems' degree structure. The 1999 counterparts of the 2000 cohorts' would-be-*cand.mags.* pay the cost for (up to) a *cand.mag.*, but still earn bachelor degrees. Their choices will depend on what they expect to earn for a too-long bachelor. If employers only observe degrees, 1999 cohort too-long bachelor holders would earn the same wage as all bachelors. On earning their bachelor, they will have to make a new choice of whether to study for a masters degree; whether they upgrade to a master's just as much as their 2000 cohort counterparts depends on whether education costs escalate. They face the same wage increase for the same increase in the duration of higher education.

If 1999 students can use their excess bachelor's credits towards a master's, their returns to studying one year extra becomes larger than for the 2000 cohort, implying that *more* 1999 students earn a master's degree compared to the 2000 cohort. Bologna transition rules specify that excess credits from a bachelor can be used towards a master's, but only if these are credits from master level courses. This is only the case for some few students; most students do not carry over a significant amount of credits. But if 1999 students were to earn master's degrees at a higher rate, this would not be an issue with monotonicity (but possibly with degree signalling).

To sum up, in a pure pre-Bologna post-Bologna degree structure comparison, defiers would be expected. Since the 1999 transition cohort does not have the old degrees available, one can either expect to have no defiers and both cohorts upgrading to masters at the same rate vis-à-vis the old degree structure or having more 1999 students choosing master, which would increase the first stage but not violate monotonicity. Table 3 shows no evidence of students upgrading degree choices; about the same proportion of students obtain lower level (bachelor and *cand.mag.*) and higher level degrees (master and *hovedfag*) across cohorts.

7 Conclusion

Societies and individual students invest substantial resources in higher education. While the returns to higher education are well documented and consistently found to be significant and substantial across countries, institution types, and field of study, there is less evidence on the mechanism through which higher education is rewarded. Conventional estimates of the returns to higher education capture the joint returns to university-acquired human capital, signalling and potentially monopoly and licensing rents. This paper separates the returns to university-acquired human capital from other sources of higher education returns.

I capitalize on a distinctive feature of Norway's implementation of the Bologna reform. Students who enrolled in 1999 inadvertently acquired more human capital while obtaining identical degrees compared to their counterparts who enrolled in 2000. Since both the 1999 and 2000 enrollment cohorts began their studies before the reform was ratified, let alone communicated, differential self-selection into enrollment cohorts can be ruled out. The 1999 cohort had already studied longer for their undergraduate degree than necessary for the new bachelor degree when the reform was announced, resulting in this cohort accruing more more human capital. Still, both cohorts were awarded the new degree formats - bachelor's and master's - so that students from both cohorts do not differ in the distribution of degree types.

I estimate the economic returns of university-acquired human capital by examining the earnings differential between the 1999 student cohort and the 2000 cohort, and then comparing this with the differential between the 2001 cohort and the 2002 cohort. The 2001 and 2002 cohort learned early in their studies about the Bologna reform so that neither cohort received unplanned extra study credits like the 1999 cohort. Using a quasi difference-in-difference design, I differentiate the returns attributed to extended study specific to the 1999 cohort as opposed to the 2000 cohort from the benefits of earlier enrollment, which are common to both the 1999 and 2001 cohorts when compared to their subsequent peers.

Spending more time in higher education without receiving a degree results in at least as high earnings as graduating earlier and collecting more experience. In observational data, spending longer at university for a given degree is associated with a substantial, significant, and lasting wage penalty. When employers observe students that studied longer because of the reform without earning higher-level degrees, they do not penalize this, but if anything, these students earn more. In the medium run, earnings for students exposed to the reform and staying longer at university are not significantly different from students that were not exposed. In the long run, students with additional education climb 1.5 percentile ranks in the earnings distribution, indicating substantial returns. I show that the time dynamics and effect sizes are comparable across different specifications.

My baseline estimates lack the precision to distinguish between moderate and large effects of university-acquired human capital on earnings. Using a novel 3-step cross-fit test-and-select estimator to leverage first-stage heterogeneity improves compliance and reduces standard errors by 16%. This still leaves the upper bound for an additional year of reform-induced education at 4.6 ranks. Comparing this effect size to observational differences between sexes and for natural science vs. social science and humanities students which stand at 5.6 and 6.7 ranks respectively, illustrates the issue.

References

Abadie, Alberto, Jiaying Gu, and Shu Shen (2023). "Instrumental variable estimation with first-stage heterogeneity." In: *Journal of Econometrics*.

Altmejd, Adam, Andrés Barrios-Fernández, Marin Drlje, Joshua Goodman, Michael Hurwitz, Dejan Kovac, Christine Mulhern, Christopher Neilson, and Jonathan Smith (2021). "O brother, where start thou? Sibling spillovers on college and major choice in four countries." In: *The Quarterly Journal of Economics* 136.3, pp. 1831–1886.

Altonji, Joseph G and Charles R Pierret (2001). "Employer learning and statistical discrimination." In: *The quarterly journal of economics* 116.1, pp. 313–350.

Angrist, Joshua D and Jörn-Steffen Pischke (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Arcidiacono, Peter, Patrick Bayer, and Aurel Hizmo (2010). "Beyond signaling and human capital: Education and the revelation of ability." In: *American Economic Journal: Applied Economics* 2.4, pp. 76–104.

Arrow, Kenneth J et al. (1973). "Higher education as a filter." In: *Journal of Public Economics* 2.3, pp. 193–216.

Arteaga, Carolina (2018). "The effect of human capital on earnings: Evidence from a reform at Colombia's top university." In: *Journal of Public Economics* 157, pp. 212–225.

Aryal, Gaurab, Manudeep Bhuller, and Fabian Lange (2022). "Signaling and employer learning with instruments." In: *American Economic Review* 112.5, pp. 1669–1702.

Åsheim, Oddveig (2002). *"Kvalitetsreformen" ved Universitetet i Bergen. Spørsmål og svar.nbsp;* URL: https://web.archive.org/web/20020226063202/http://www.uib.no/fa/kfsk/st-meld-27/sporsmal-og-svar.htm,%20accessed%2020.10.2023.

Baker, Michael and Gary Solon (2003). "Earnings dynamics and inequality among Canadian men, 1976–1992: Evidence from longitudinal income tax records." In: *Journal of Labor Economics* 21.2, pp. 289–321.

Becker, Gary S (1962). "Investment in human capital: A theoretical analysis." In: *Journal of Political Economy* 70.5, Part 2, pp. 9–49.

Bedard, Kelly (2001). "Human capital versus signaling models: university access and high school dropouts." In: *Journal of Political Economy* 109.4, pp. 749–775.

Bhuller, Manudeep, Magne Mogstad, and Kjell G Salvanes (2017). "Life-cycle earnings, education premiums, and internal rates of return." In: *Journal of Labor Economics* 35.4, pp. 993–1030.

Björklund, Anders (1993). “A comparison between actual distributions of annual and lifetime income: Sweden 1951–89.” In: *Review of Income and Wealth* 39.4, pp. 377–386.

Blandhol, Christine, John Bonney, Magne Mogstad, and Alexander Torgovitsky (2022). *When is TSLS actually late?* National Bureau of Economic Research.

Bleemer, Zachary and Aashish Mehta (2022). “Will studying economics make you rich? A regression discontinuity analysis of the returns to college major.” In: *American Economic Journal: Applied Economics* 14.2, pp. 1–22.

Brandimarti, Eleonora (2023). *Self-Selection, University Courses and Returns to Advanced Degrees*. Working paper.

Carneiro, Pedro, James J Heckman, and Edward J Vytlacil (2011). “Estimating marginal returns to education.” In: *American Economic Review* 101.6, pp. 2754–2781.

Cinelli, Carlos, Andrew Forney, and Judea Pearl (2022). “A crash course in good and bad controls.” In: *Sociological Methods & Research*, p. 00491241221099552.

Clark, Damon and Paco Martorell (2014). “The signaling value of a high school diploma.” In: *Journal of Political Economy* 122.2, pp. 282–318.

Coussens, Stephen and Jann Spiess (2021). “Improving inference from simple instruments through compliance estimation.” In: *arXiv preprint arXiv:2108.03726*.

De Haan, Monique (2017). “The Effect of Additional Funds for Low-ability Pupils: A Non-parametric Bounds Analysis.” In: *The Economic Journal* 127.599, pp. 177–198.

Farber, Henry S and Robert Gibbons (1996). “Learning and wage dynamics.” In: *The Quarterly Journal of Economics* 111.4, pp. 1007–1047.

Haider, Steven and Gary Solon (2006). “Life-cycle variation in the association between current and lifetime earnings.” In: *American Economic Review* 96.4, pp. 1308–1320.

Hastings, Justine S, Christopher A Neilson, and Seth D Zimmerman (2013). *Are some degrees worth more than others? Evidence from college admission cutoffs in Chile*. National Bureau of Economic Research.

Hazard, Yagan and Simon Löwe (2022). *Improving LATE estimation in experiments with imperfect compliance*. Working paper.

Heckman, James J, Lance J Lochner, and Petra E Todd (2006). “Earnings functions, rates of return and treatment effects: The Mincer equation and beyond.” In: *Handbook of the Economics of Education* 1, pp. 307–458.

Huntington-Klein, Nick (2020). “Instruments with heterogeneous effects: Bias, monotonicity, and localness.” In: *Journal of Causal Inference* 8.1, pp. 182–208.

Imbens, Guido W and Joshua D Angrist (1994). "Identification and Estimation of Local Average Treatment Effects." In: *Econometrica: Journal of the Econometric Society*, pp. 467–475.

Jackson, C Kirabo, Rucker C Johnson, and Claudia Persico (2016). "The effects of school spending on educational and economic outcomes: Evidence from school finance reforms." In: *Quarterly Journal of Economics* 131.1, pp. 157–218.

Johnson, Matthew S (2020). "Regulation by shaming: Deterrence effects of publicizing violations of workplace safety and health laws." In: *American economic review* 110.6, pp. 1866–1904.

Ketel, Nadine, Edwin Leuven, Hessel Oosterbeek, and Bas van der Klaauw (2016). "The returns to medical school: Evidence from admission lotteries." In: *American Economic Journal: Applied Economics* 8.2, pp. 225–254.

Kirkeboen, Lars J, Edwin Leuven, and Magne Mogstad (2016). "Field of study, earnings, and self-selection." In: *The Quarterly Journal of Economics* 131.3, pp. 1057–1111.

Lange, Fabian (2007). "The speed of employer learning." In: *Journal of Labor Economics* 25.1, pp. 1–35.

Løken, Katrine V, Magne Mogstad, and Matthew Wiswall (2012). "What linear estimators miss: The effects of family income on child outcomes." In: *American Economic Journal: Applied Economics* 4.2, pp. 1–35.

Manski, Charles F (2009). *Identification for prediction and decision*. Harvard University Press.

Markussen, Simen and Knut Røed (2020). "Economic mobility under pressure." In: *Journal of the European Economic Association* 18.4, pp. 1844–1885.

Mincer, Jacob A et al. (1974). *Schooling, Experience, and Earnings*. National Bureau of Economic Research.

Ministry of Edcuation (2002). *Kvalitetsreformen. Om vurdering av enkelte unntak fra ny gradsstruktur i høyere utdanning. [The Quality reform. About the evaluation and exemptions from the new degree structure in higher education]*. URL: https://www.regjeringen.no/no/dokumenter/stmeld-nr-11-2001-2002-/id195363/?ch=1,accessed%2020.10.2023.

Mogstad, Magne and Alexander Torgovitsky (2018). "Identification and extrapolation of causal effects with instrumental variables." In: *Annual Review of Economics* 10, pp. 577–613.

Nilsen, Øivind Anti, Kjell Vaage, Arild Aakvik, and Karl Jacobsen (2012). "Intergenerational earnings mobility revisited: Estimates based on lifetime earnings." In: *The Scandinavian Journal of Economics* 114.1, pp. 1–23.

Oreopoulos, Philip and Uros Petronijevic (2013). *Making college worth it: A review of research on the returns to higher education*. National Bureau of Economic Research.

Pearl, Judea (2009). *Causality*. Cambridge university press.

Schönberg, Uta (2007). "Testing for asymmetric employer learning." In: *Journal of Labor Economics* 25.4, pp. 651–691.

Spence, Michael (1973). "Job Market Signaling." In: *The Quarterly Journal of Economics*, pp. 355–374.

Stephens, Melvin and Dou-Yan Yang (2014). "Compulsory education and the benefits of schooling." In: *American Economic Review* 104.6, pp. 1777–1792.

Tyler, John H (2003). "Economic benefits of the GED: Lessons from recent research." In: *Review of Educational Research* 73.3, pp. 369–403.

A Appendix

	Spring 1999		Spring 2000		
Variable	Mean	SD	Mean	SD	p.value
Female	1.46	0.5	1.5	0.5	0.15
Mothers' education	4.57	1.75	4.59	1.71	0.87
Fathers' education	4.94	4.98	4.94	4.98	0.66
Natural science student	0.04	0.19	0.03	0.18	0.58
Age at enrollment	21.71	6.37	21.86	6.44	0.68
Nr of siblings	2.15	0.92	2.19	0.93	0.49
N	647		667		

Table A.1: Balancing of observables for students starting in the spring semester

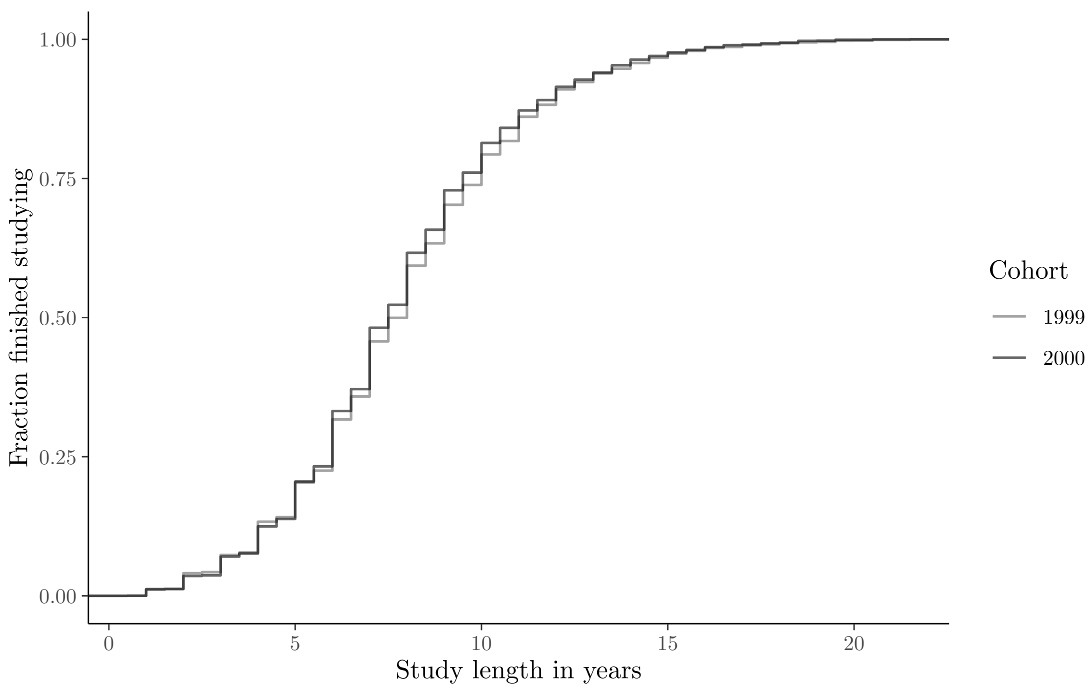


Figure A.1: Timing of extra study time: empirical cdf across cohorts

Notes: This figure includes individuals that did not collect any study credits at all during the pre-screening period. Study length is measured as net study time, that is, the total number of years a student spends in higher education, irrespective of gaps. I define ended study time as not being enrolled anymore and not enrolling again until 2020 when the data series ends. I do not condition on whether students graduate or drop out when they stop being enrolled.

		Treatment	Control
	Old degree & 240+ study credits	New degree & 180-240 study credtis	Control: New degree & 180 study credits
1997	$>$ 40 credits/year	30-39 credits/year	$\leq$ 30 credits/year
1998	$>$ 48 credits/year	36-47 credits/year	$\leq$ 36 credits/year
1999	$>$ 60 credits/year	45-59 credits/year	$\leq$ 45 credits/year
2000	$>$ 80 credits/year	60-79 credits/year	$\leq$ 60 credits/year
2001	$>$ 120 credits/year	90-119 credits/year	$\leq$ 90 credits/year

Table A.2: Link between study credits in July 2003 and treatment status

Table A.3: Placebo diff-in-diff robustness check: Log earnings as outcome

	Baseline sample			Fall students only		
	(1)	(2)	(3)	(4)	(5)	(6)
Year advantage	1.258***	0.045	−0.021	1.296***	0.050	−0.018
	(0.069)	(0.053)	(0.060)	(0.071)	(0.054)	(0.061)
Reform cohorts	−0.570***	−0.090 ·	−0.098	−0.586***	−0.102 ·	−0.118 ·
	(0.071)	(0.054)	(0.061)	(0.074)	(0.056)	(0.063)
Year advantage * Reform cohorts	−0.050	0.046	0.109	−0.080	0.073	0.151 ·
	(0.098)	(0.073)	(0.082)	(0.105)	(0.077)	(0.087)
Baseline controls included	Yes	Yes	Yes	Yes	Yes	Yes
Time period FEs	Yes	Yes	Yes	Yes	Yes	Yes
Included time periods	<0	10-19	15-19	<0	10-19	15-19
R^2	0.303	0.031	0.041	0.298	0.031	0.038
Adj. R^2	0.302	0.031	0.040	0.297	0.030	0.037
Num. obs.	56602	138607	69208	48359	118396	59127

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\cdot}p < 0.1$

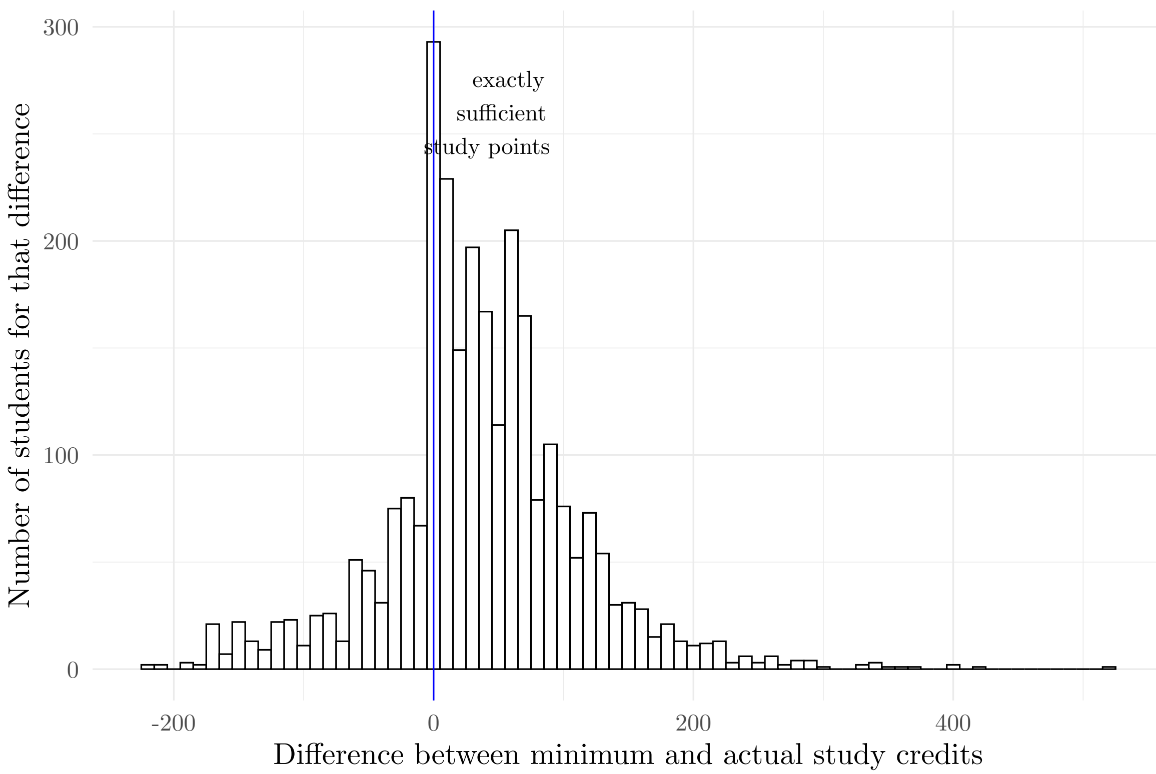
I estimate the effects on log earnings rather than ranks. Year advantage is an indicator for enrolling earlier in each cohort pair (1999 or 2001), Reform cohorts are the enrollment cohorts 1999 and 2000. The interaction term reflects the additional earnings advantage the 1999 cohort enjoys relative to the 2001 cohort. Models 1-3 show estimates for the full baseline sample, models 4-6 show estimates for the sample of fall students. All regressions include flexible controls for age at enrollment, gender, natural sciences and mothers' and father's education level as well as fixed effects on the relative time period. Standard errors are clustered at the individual level.

Table A.4: Placebo First stage

	Human capital measure: Study time			Study credits	Subjects
	(1)	(2)	(3)	(4)	(5)
2001 cohort	0.039	0.063	0.034	0.256	0.095
	(0.067)	(0.074)	(0.074)	(2.653)	(0.176)
Spring semester		-0.942^{***}	-0.664^{***}	-12.208^{**}	5.658^{***}
		(0.122)	(0.125)	(4.472)	(0.297)
2001 cohort*Spring semester		-0.009	-0.091	1.764	-5.435^{***}
		(0.171)	(0.173)	(6.164)	(0.410)
Controls	No	No	Yes	Yes	Yes
R^2	0.002	0.016	0.060	0.085	0.051
Adj. R^2	0.001	0.016	0.055	0.080	0.045
Num. obs.	7230	7230	7028	7058	7058

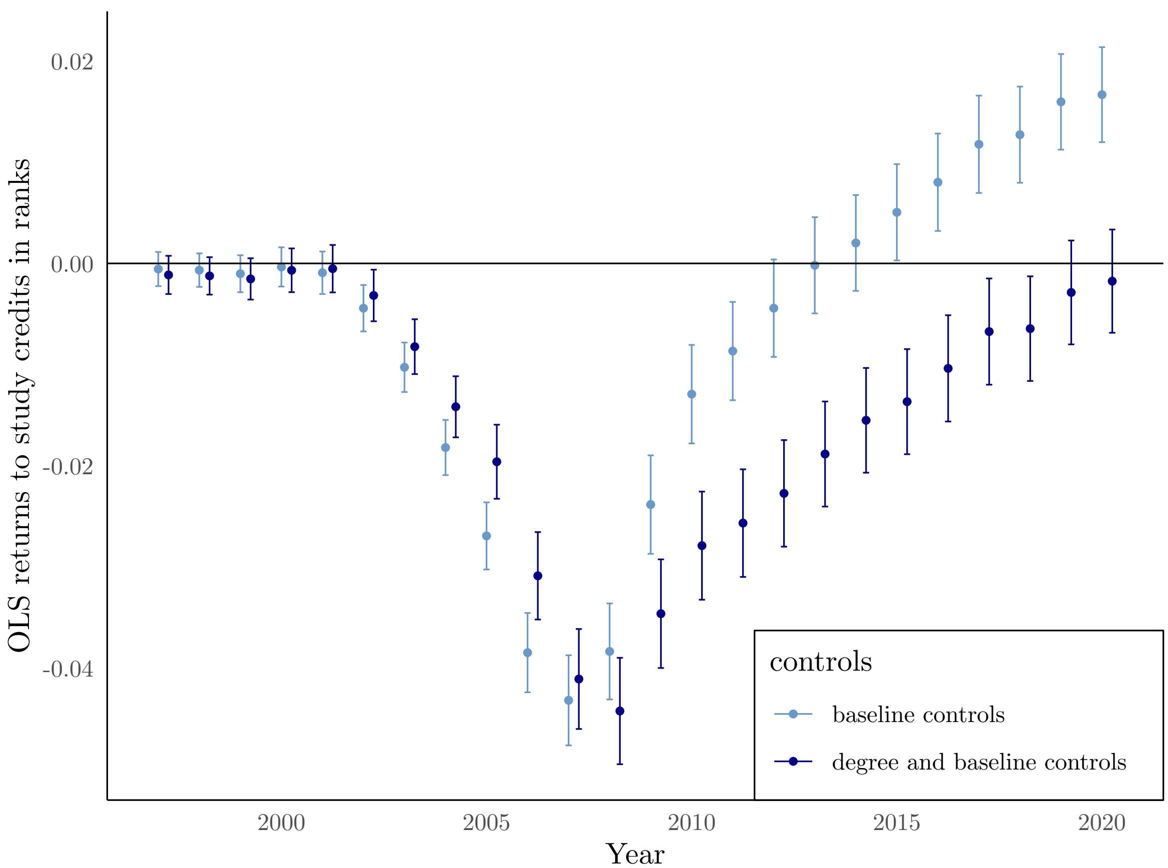
$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$
All models are estimated on the placebo sample, the 2001 and 2002 enrollment cohort. 2001 cohort is an indicator for enrolling in the 2001 cohort. Models 1-3 report the main first stage with study time after pre-screening in years as the dependent variable. Model 4 and 5 report for alternative measures of human capital, study credits after pre-screening and number of chosen subjects after pre-screening. Flexible controls for age at enrollment, gender, natural sciences and mothers' and father's education level included where indicated.

Figure A.2: Evidence for measurement error in study credits



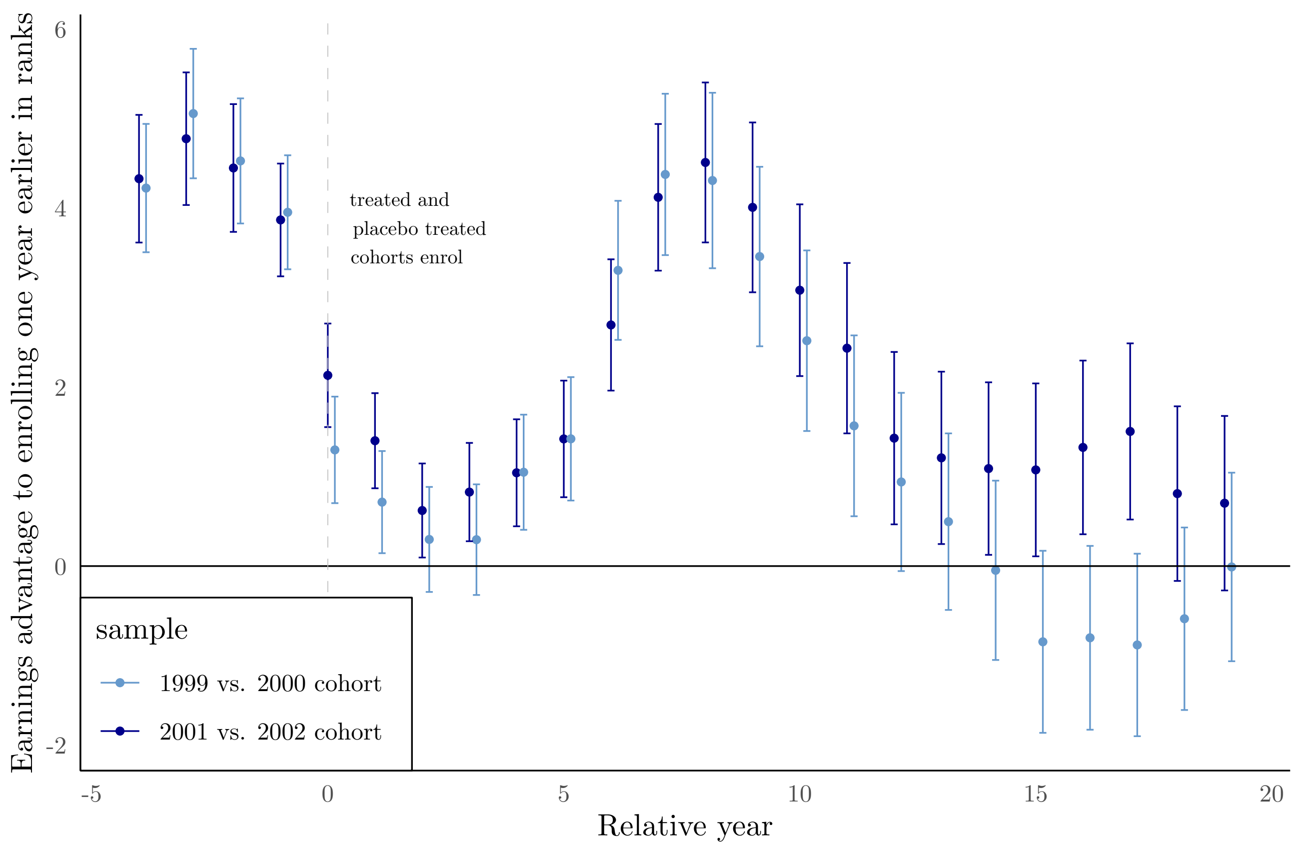
Notes: This figure displays the difference between the minimum number of study credits necessary for obtaining the highest degree a person earns and the total recorded number of study credits. Fewer credits than necessary indicate under-reporting in the number of study credits. If credits collected abroad can be counted towards a degree, they are supposed to registered as study credits as well. Students not residing in a municipality in Norway in some semesters (who likely are on exchange) are only weakly more likely to be registered as having fewer credits than necessary. More credits result when students obtain degrees that do not match into a prescribed educational career. A student with a bachelors in chemistry and art history and a master in chemistry would show up as having 180 credits too many as the chemistry master will be recorded as needing 300 credits (120 for the master and 180 for the prerequisite chemistry bachelor). Courses that are taken out of interest, i.e. language classes, would lead to positive numbers as well. Finally, all study credits of students dropping out are recorded as positive, since their non-acquired degree counts as 0 study credits.

Figure A.3: OLS benchmark: returns to chosen study credits



Notes: The upper panel shows γ_t coefficients from estimating Equation 1, and the lower panel shows corresponding γ_t coefficients from Equation 2. Years t range from 1997 to 2020. In Equation 2 I control flexibly for the level of the obtained degree. Results are return to each obtained study credits, where a normed speed student obtains 60 study credits per year. Both regressions include controls for parental education, gender, natural science, enrolling in spring and age at enrollment. Enrollment cohorts are pooled and a dummy for enrollment cohort is included.

Figure A.4: Full sample robustness: Human capital and placebo returns



Notes: I omit the interaction of enrollment cohort and spring semester in Equation 3, so that the estimated cohort advantages are for all students rather than being estimated only for fall students as in the main specification. The plot shows the set of estimated γ_t coefficients from Equation 3 separately for the treatment sample, consisting of the 1999 and 2000 cohort, and the placebo sample, consisting of the 2001 and 2002 cohort. All regressions include flexible controls for parental education, gender, natural science, enrolling in spring and age at enrollment.