To whom it may concern: the value of auditing grade assessment on educational and labour market outcomes.*

Giorgia Casalone[a], Paolo Ghinetti[a], and Daniela Sonedda[a,b]

[a] *University of Eastern Piedmont.*
[b] *Centre For North South Economic Research (CRENoS), Cagliari, Italy.*

October 2021

Abstract

In this paper, we assess the value of auditing the final high school grades on educational and labour market outcomes. We leverage a 2007 reform in Italy that introduced the presence of external examiners on the board. We compare treated and untreated cohorts in a two-way fixed effects model to show that the reform increased the earnings of high school graduates. We carry out two-way fixed effects and special regressor methods to prove that the reform raised the pupils' years of schooling. We extend the combined fixed effects approach (Altonji and Zhong 2021) to attest that treated cohorts' returns to graduation are about six percentage points as high as the untreated ones.

Keywords: Returns to Education; Auditing; Earnings gender gap

JEL Codes: I20; J30; M42

*We thank Edoardo Di Porto, Paolo Naticchioni, Matteo Paradisi and Silvia Vannutelli for helpful comments and suggestions. This study uses anonymous data from the Italian Institute of Social Security (INPS), and data access was provided as part of the VISITINPS Programme. We are very grateful to Daniele Checchi, Monia Monachini, Alfredo Arpinelli and Barbara Ceremigna for their invaluable help in making this project possible and Massimo Ascione and Elio Bellucci for their support with the data. We also thank the Chancellor and the Administrative Offices of the University that provided us with its data.
Corresponding author: Daniela Sonedda, Via Perrone 18, 28100 Novara, Italy;+39 0321 375325; daniela.sonedda@uniupo.it.

1 Introduction

Record numbers of top high school grades were awarded to pupils in developed countries in 2020 and 2021 compared to figures from 2019, the last year before the Covid pandemic.[1] These grades were primarily teacher-assessed. Heated debates followed suit on how fair these grades were. The primary argument is simple. An external examiner, like an auditor, gives the guarantee of their fairness. In her absence, one can cast doubts on them even when they are well deserved. The economic costs of these doubts are earnings losses. These earnings losses come from three sources: inaccurate information on pupils' knowledge and skills, wrong choices of the educational path, and lower paid jobs in worse occupations.

In this paper, we explore all these three issues in a novel way. Inaccurate information on pupils' knowledge and skills impacts employers' expectations on workers productivity. Hence, firms reward the value of the final high school grade less and offer lower earnings. A final grade backed by external examiners is more likely to make the pupils feel that it is legitimately earned. Therefore, their choice about which track to go through may differ for those who enrol at university. As a final result, the returns to graduation rise for two reasons. First, the labour market rewards them more. Second, the different educational paths lead workers to find better jobs with higher earnings. To the best of our knowledge, this is the first paper that accounts for all these issues.

A well-established fact is that education increases earnings but bears some costs. Becker (1967) shows that people reach an optimal schooling decision by weighing the benefits of higher schooling (which are picked over the lifecycle) against the costs. However, recognising individuals' different aptitudes and tastes for schooling required a more general framework to account for them. These different skills and preferences lead to optimal choices that vary. Card (2001) provides a theoretical and empirical model that allows the returns to education to differ across the population. Individual heterogeneity in the optimal schooling choice can arise from differences in the marginal return or marginal costs to schooling.

New insights about the connection between education and earnings came out with asymmetric information and uncertainty taken into consideration. For instance, one may think of education as a multi-period investment with uncertain returns but certain costs (Levhari and Weiss 1974). This argument hails an insurance device against such uncertainty. For instance, borrowing, saving and labour supply adjustments could play that part (Low, Meghir and Pistaferri 2010).

[1]See, for instance figures for the UK.

One can stress the importance of the process through which individuals and firms form expectations to relate education, earnings and productivity. How clear these expectations are is essential. Grades contribute to this process, and those who assess them bear a huge responsibility and are implicitly accountable for that. For instance, school and teachers' reputations are at stake. This context is equivalent to the reason why it is worthy of making external auditors assess a firm's balance sheet.

The relevance of the external audit is now a well-recognised fact. For instance, Demski and Swieringa (1974) highlighted that the external audit is essential for the auditee's financial reporting system. However, in 1980, the nature of the auditee's benefits of the audit service was at least in part still unclear. Simunic (1980) suggested that the benefits come from liability avoidance. External auditing takes the form of a guarantee provided to shareholders and creditors. This guarantee reduces the legal liability of an auditee.

In the first part of the paper, we present a simple theoretical model of endogenous schooling with external auditing of high school final grades. This model is then used to motivate the second part of the paper, where we test its three predictions. We introduce some elements of the auditing service into the Card (1999) model. We assume the existence of a loss function of an unfair assessment of the high school grades. We define an assessment as unfair when the grades do not reflect what a student knows, understand and can do. When the assessment is perceived as unfair, earning losses follow suit that can be reduced by external auditing. In this context, three model predictions arise. First, external auditing raises the earnings of high-school degree owners. Second, some marginal pupils would make a different choice about their schooling and increase it. Third, university courses will determine what people will earn. Hence, choosing a longer and more challenging course because of the external auditing will lead to higher returns to graduation. In the second part of the paper, we test these three remarks.

We leverage a 2007 reform in Italy that changed the composition of the final high school exam board. After this exam, pupils become high school degree owners with a grade from 60 to 100, summarising their knowledge and competencies. Before the reform, the board was made of internal teachers only. The reform ruled that external examiners comprised half of the board whose head was also external to the school. Hence, the majority of the votes was not under the control of the school teachers.

We employ a common strategy to test our three model's predictions: we compare treated and untreated cohorts. However, we carry out a specific method for each of them. We start using the data of the Italian Institute of Social Security (INPS) to conduct our analysis on high school degree owners. We use two archives to provide robust evidence. The first archive builds on social security contributions paid in all sectors. The second one is the universe of Italian workers in the private sector. We select two untreated cohorts born in 1986 and 1987 and two treated cohorts born in 1988 and 1989. In this part of the analysis, we retain those with a high school degree. To do that, we link the INPS archives with data from the Mandatory Communications, maintained by the Ministry of Labour and Social Policies. We assume that these cohorts are statistical identical but the treatment. This hypothesis rules out the existence of heterogeneous treatment effects across groups. However, we allow for gender differences. We then trace the earnings (income) profiles of these two groups from age 19 to 30 under a common age profile assumption. The untreated group's profile sets the benchmark of what the treated group would have achieved without the reform. Our treatment is pre-determined to the labour market participation[2], and therefore constant over time. This circumstance dismisses heterogeneous treatment effects over time. Hence, we meet the two conditions under which a two-way fixed effects model is valid.[3] We can, then, compare the age profiles of treated and untreated groups to show whether the reform raised earnings (income) of high school degree achievers.

To test our second remark, we use administrative data from an Italian University. We divide this part of the analysis into two stages. In the first stage, we establish whether treated cohorts increased their schooling by exploiting a feature of the Italian university system. Since the 2001 reform, there have been two university tracks: the one-tier and the two-tier track. The former is longer and more challenging than the latter one. Hence, we study whether, as a consequence of the 2007 reform, enrolment rates switched from the two-tier to the one-tier. To do that, we employ a two-way fixed effects estimator that is valid under the conditions described above, to which we add another one borne out by our figures. Treated and untreated cohorts share the same enrolment propensity at this University. What this first stage does is providing evidence of a switch in the track. The second stage proves helpful to argue that the number of years of schooling increased without consequences on the graduation rates. This second stage is essential to our primary argument. Those enrolled at the two-tier track can register at the second tier after completing the first. If this circumstance occurs regularly, there could not be differences in the years of schooling in the two tracks.[4] We use the special regressor method (Dong and Lewbel 2015) to estimate the impact on the enrolment and graduation probability of being

[2]It is unlikely that high school pupils work while at school.

[3]For issues on the two-way fixed effects and difference-in-differences estimators when treatment effects are heterogeneous, see for instance, de Chaisemartin and D'Haultfoeuille (2018), Goodman-Bacon (2018), Callaway and Sant'Anna (2020), Sun and Abraham (2020), de Chaisemartin and D'Haultfoeuille (2020).

[4]An exception is the Medicine degree that requires six years.

enrolled at the one-tier track. We measure these effects on the former (latter) outcome after one (three) to six years from the first enrolment.

We extend the Altonji and Zhong (2021) method to test our third remark and establish whether the 2007 reform increased the returns to graduation. We employ the same datasets used for the first remark but enlarged to those with a university degree. We identify our combined fixed effects as a pair of high school and university degrees.[5] In this context, the age profile of pre-university earnings of individuals who later obtain a graduate degree approximate what they would have earned had they not gone to university. This approach generalises the fixed-effects case that identifies the returns to a university degree using only people with earnings observations before and after university graduation. Instead, the combined fixed-effects model exploits the earnings observations of all workers. Those observed only before or only after contribute to adjust for the counterfactual experience profile without a degree. Two conditions are required to validate this method. First, information about ability or preferences does not change between labour market entry and the decision to enrol at university. Second, high school degree achievers share a common experience profile. We extend this setting by comparing cohorts treated and untreated by the 2007 reform. These two groups overlap in terms of ability, preferences and labour market experience distributions but differ because of the treatment. Hence, we could identify the reform's effect even if some assumptions on the ability, preferences and labour market experience distributions were not satisfied within each cohort. In such a case, we would need that treated and untreated cohorts share a common bias due to these factors, and this bias is washed by making the cohorts comparison.

Our paper contributes in a novel way to the vast literature on the returns to schooling. We extend the Card (1999) model to estimate the effect of auditing the final high school grade on educational and labour market outcomes. Over the past years, the literature has moved in the direction to estimate the returns to college majors (see Altonji, Blom and Meghir (2012), Altonji, Arcidiacono and Maurel (2016), and Altonji and Zhong (2021)). It has studied the determinants of college major choice (Kinsler and Pavan 2015), also highlighting the importance of information on skills (Wiswall and Zafar 2015). Other works focus on gender differences in this choice (see, for instance, Zafar (2013)). In the context of the 1994 educational reform in Norway, Bertrand, Mogstad and Mountjoy (2021) show divergent consequences by gender. The reform worsened the gender gap in adult earnings while proving helpful to disadvantaged men.

In dealing with testing our three model's predictions, we address the issue of the returns to

[5] Altonji and Zhong (2021) instead perform combined fixed effects for pair of field-specific undergraduate and graduate degrees.

the university track. Enrolment at the longer and more challenging one-tier track rather than the two-tier one helps shape the labour market prospects. We show that pupils have got to choose carefully what they study at university because the track determines what they go on to earn. We find that the 2007 reform increased the enrolment in the one-tier track by about four percentage points and reduced that in the two-tier track by the same amount. Five years after completion of high school, treated cohorts had the enrolment probability of 21 percentage points as high as untreated ones. We do not detect a worsened graduation probability. Hence, the 2007 reform changed pupils' choices and the labour market outcomes of those who did not change them. On average, the reform raised the earnings of high school achievers by about one percent but only for men. When we adjust for the experience profile, the gender differential impact is still there. The beneficial effect of the reform increases with labour market experience for men but less so for women. If we consider the occupational earnings premium as an outcome, on average, both women and men benefited from the reform, with women slightly less than men. However, the gender differences in the experience-adjusted impacts of the reform on income are striking: the men profiles are steep and become steeper after the reform, the women ones keep being flatter. To complete our reform insights, we find that returns to a university degree of treated cohorts were six percentage points as high as those of untreated. Women are the ones who benefited most from it. To the best of our knowledge, this is the first paper that provides a comprehensive view of the value of auditing the final high school grades.

The paper proceeds as follows. In Section 2, we illustrate the institutional framework, and we present the model. In Section 3, we describe the data. In Section 4, we first discuss the econometric specification and present the estimates of the reform impact on earnings and income of high school achievers. In Section 5, we account for the reform effect on pupils education choices and outcomes. In Section 6, after discussing the empirical model, we provide evidence on the reform effect on the return to graduate degrees. We conclude in Section 7.

2 The setting

2.1 Institutional Setting

At the beginning of the 2000s, Italy's university system needed an overhaul to accomplish the Bologna process. A reform in 2001 kicked in to substitute the one-tier with a two-tier path. A three-year bachelor degree course makes the first of the two tiers. A two-year master degree course can be added to complete the graduation programme. This master degree is not compulsory, and those who do not go on for it after completing the first tier have a higher education degree. Yet, some degree courses such as Medicine and Law maintain their one-tier structure.[6] At the onset of the reform of the high-school examination board, pupils could head to the one or the two-tier path depending on the degree programme.

This reform took place in 2007 and was barely foreseen. Law no.1/2007, issue in January, established a 50% quota of external examiners on the board. This rule replaced the grade assessment with a 100% internal teachers board. Since the final high-school exam is in June and July of each year, there was little scope for adjustment. Maximum six people plus the Head can sit on the board. The Head cannot be an internal teacher allotting to the external quota the majority of the votes. Each year for each type of school,[7] the department office for education chooses the subjects of the external examiners. Three written assessments and an interview make the exam, and the board votes the grade assessment. When there is no agreement, the majority of the votes sets the final grade in a range from 60 to 100. The board can assess a maximum of seventy-five points, forty-five for the written part of the exam and thirty for the interview. The twenty-five points left reflect the pupil's high-school career.

The exam admission had to be set by the internal board of teachers, and it has not come to pass for some pupils. This measure was part of the increased fairness in the grade assessments. The debate that followed the law's introduction emphasised the need for an accountable grade system to replace one that had lost the faith of many. The law intended to entrust the external board the task to restore this lost credibility. Our model explains why.

2.2 A simple model

High-school final grades are regarded as a fair and accurate reflection of the ability of pupils receiving them. While grades and abilities are undeniably related, they do not perfectly match. A less able pupil could achieve the same grade as a more able peer putting in more effort. Grades are the best way of assessing what a student knows, understand and can do. Exams can be a real catalyst for pulling together learning and for synthesising learning. Hence, each grade reveals a certain degree of knowledge, cognitive and non-cognitive skills that convert into a certain level of productivity in the labour market. The precision of this conversion lies in how accurate and fair is the grade assessment. Internal teacher assessment could give rise to controversy over possible

[6] Law degree courses followed the two-tier path in the phase-in of the reform. After a couple of years in 2006, they stepped back to the one-tier programme.

[7] In Italy, high schools differ on subjects and vocational/academic tracks. Several forms of lyceums, depending on the subjects, head to the academic track. For instance, Scientific Lyceum focuses more on maths and physics while ancient Greek and Latin are taught in the Classic Lyceum. Other high schools are more vocational.

grade inflation or biased judgements even when the grades are fair. On the one hand, one could argue that these assessments could reflect experiences from years of personal relationships with the pupils. On the other hand, it could be said that the teachers could fear being the ones who are judged. Oversight carried out by an external exam board limits these two concerns because it is equivalent to auditing a firm.

Auditing high-school final grades can serve different purposes linked by a common thread: information. Audited high-school grades add credibility of observed grades is as fair as possible. External examiners, the auditors, increase the firm's confidence in the established grades. The firms are believed to gain from the increased credibility. These benefits are deemed as improved quality of hiring decisions that are based on more reliable information. For their part, firms demand accountability from the school examination board, but this demand could fall in the void in the absence of a direct link between the firm hirings and how the grades are determined. The issuance of a final high-school grade provides for accountability on the worker to be. When this issuance comes from internal teachers, there is no guarantee of an unbiased judgement. Firms cannot monitor this evaluation. In such a case, increased reliability of this information is in the interests of all the third parties as well as the high-school management. The high-school reputation is at stake. However, firms fail to appoint external examiners because they do not participate in the high-schooling system. The urged government intervention sorts this problem out. These auditors should strive to meet the firms, parents, and students expectations of fairness and unbiasedness in their judgements. The increased reliability of the high-school grades helps firms to determine market values of student productivity. Firms value more each grade and are ready to offer a higher earning. This expected firms behaviour leads students to modify their higher-education investment decisions.

We sketch the following simple model to draw this big picture on the value of auditing the high-school grading system. Consider a framework in which students are the auditees, and internal and external teachers are the auditors. The potential losses to auditees and auditors caused by a biased and unfair high-school grade drive the design of an external reporting system. Reduced losses are in the nature of higher expected earnings at each level of education. Students are liable to future penalisations in the labour market for losses attributable to defects in the final examination. Reputation costs and the added stress of knowing the consequences of failing to deliver a fair grade burden on teachers. In this setting, the benefits from external auditing derive from reductions of these losses.

Let the random variable, L, denote the expected present value of possible future losses which may arise from the high-school grade if it does not reveal the correct information on students' knowledge and skills. We assume:

$$E(L(\xi_S)) = f(a(\xi_S), q(\xi_S), \xi_S) \tag{1}$$

We define a as the quantity of information produced by the auditee in the examination procedure and q as the quantity of information produced by internal teachers. We measure the credibility of the observed grades of the high school final exam through the parameter ξ_S. We assume ξ_S exogenous and different across educational levels S. In other words, third parties can update the value of the high-school grade after higher-education completion (if any). It is beyond the scope of this model to show the students (and teachers) optimal choices of a and q. The auditors and auditees will provide information to the point in which the marginal reduction in expected losses is equal to the marginal costs.[8] We further impose the following reasonable assumptions: $\frac{\partial E(L)}{\partial a} < 0$; $\frac{\partial^2 E(L)}{\partial a^2} > 0$; $\frac{\partial^2 E(L)}{\partial a \partial q} > 0$;$\frac{\partial E(L)}{\partial q} < 0$; $\frac{\partial^2 E(L)}{\partial q^2} > 0$. The interpretation is the following: the more the information provided on the pupil's knowledge and skills, the lower the expected losses. What we do is very simple and is a direct application of the envelope theorem. We show how the expected loss function changes in response to an exogenous variation of the parameter ξ_S. This exogenous shift reflects the increased credibility of the observed grades oversight by an exam board with external teachers.

$$\frac{dE(L(\xi_S))}{d\xi_S} = \underset{-}{\frac{\partial f(a(\xi_S), q(\xi_S), \xi_S)}{\partial a}} \underset{+}{\frac{\partial a}{\partial \xi_S}} + \underset{-}{\frac{\partial f(a(\xi_S), q(\xi_S), \xi_S)}{\partial q}} \underset{+}{\frac{\partial q}{\partial \xi_S}} + \underset{-}{\frac{\partial f(a(\xi_S), q(\xi_S), \xi_S)}{\partial \xi_S}} < 0 \tag{2}$$

Equation 2 indicates that increased credibility of the high-school grades reduces losses attributable to failures in the examination process. The first and second terms drop to zero when evaluated at the optimal choice.

This loss function is a relevant part of the expected level of earnings $y(S, \xi_S)$ (per year) if an individual acquires schooling level S equal or higher than high school. The lower the losses are, the higher the expected earnings will be. We can frame this statement in a context in which audited high-school grades impact the students' higher-education choice and the labour market's returns. We follow the presentation developed in Card (1999) in which an individual chooses S to maximise a utility function $U(S, y(S, \xi_S))$:

[8]For instance, the student's choice comes out from the minimisation of the expected total costs: $E(TC) = va + E(L(\xi_i)) = f(a(\xi_S), q(\xi_S), \xi_S)$ where v measures the unit cost of a that includes the opportunity costs.

$$U(S, y(S, \xi_S)) = log(y(S, \xi_S)) - h(S) \tag{3}$$

and $h(S)$ denotes some increasing convex function. A well-known result is that the optimal schooling choice satisfies the first-order condition:

$$h'(S) = \frac{y'(S, \xi_S)}{y(S, \xi_S)} \tag{4}$$

Figure (1) describes what is new. The increased credibility of high-school grades shifts the production function upwards. This upward shift raises log earnings for all levels of schooling but differently. This result comes out from the firms' updated beliefs on worker productivity. These firms' updates reflect how the knowledge and skills developed at the high school complement the knowledge and skills developed at university. The higher the knowledge and skills, the more important their fair assessment becomes. To clarify, we make an example. Think about a low qualified job as a waiter. The restaurant owner does not need a huge amount of information on her skills. In this job, labour is pretty homogeneous. An employer who posts a vacancy for a managerial occupation faces a different challenge for filling it: she has to find the right person, and to do so, more precise information is demanded.

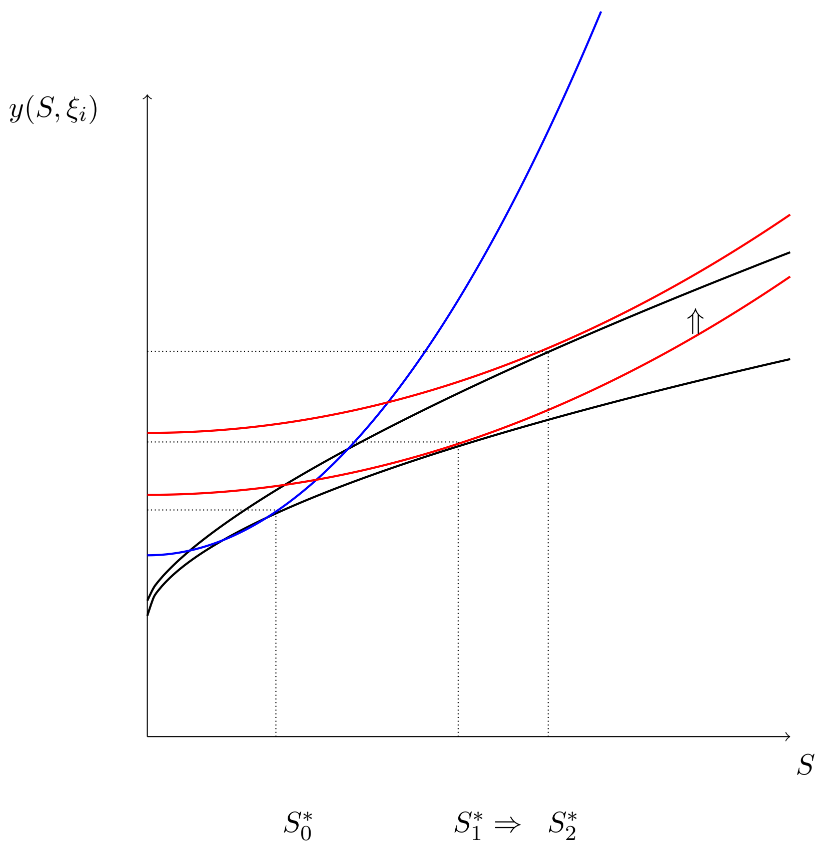


Figure 1: Optimal choice of years of schooling

The big picture that emerges is that there are three consequences of this exogenous shift. Assume that lower ability individuals always assume low education S_0^*. Hence, they will not acquire further education to S_0^* but will benefit from higher earnings. In Online Appendix B, we replicate this graph under the magnifying glass to make it clear. Second, higher ability individuals will increase their optimal amount of schooling from S_1^* to S_2^*. Third, increased earnings rise with schooling. The distance between the two production functions amplifies when schooling increases. Hence, for a given schooling level, the increased earnings are higher when schooling is higher.

We submit these three remarks to test.

3 Data

To test our three model's implications, we use two administrative data sources. Archives from the Italian Social Security Institute (INPS) are our first source. We employ these data to conduct our analysis on the first and third implications of our model. We gather data from Social Security Records for all sectors and an archive on working careers for the private one. We have monthly employment, earnings and income for the whole working population for the private sector. For all sectors, we have a representative sample. In our definition of income, we consider all sources of income from private, public and self-employment, while earnings relate to jobs in the private sector. We link these data with an archive on demographics to select our working sample of all people born in 1986-1989. These four cohorts are enough to satisfy our requirement to compare similar cohorts but the treatment because of our data. As discussed above, we have full coverage of the private sector and broad coverage of all sectors. Enlarging the sample to other cohorts would require to deal with two opposite reforms of the compulsory schooling age. Compulsory schooling age was first raised to 15 in 1999 and then was lowered to 14 in 2003.[9] These changes could have impacted the age profiles of the earnings of the involved cohorts. We expect this issue to be more severe for the oldest cohorts because those who did not keep studying at 15 would have stopped at 14 without achieving a high school degree. Those who did not keep studying at 14 could have taken the chance to complete high school if forced to be at school longer. The information on the education level is missing in the INPS archives. We merge the INPS data with the archive of mandatory communications on job creation or destruction to retrieve it. This archive records starting and ending date of any job (as well as

[9] Cohorts between 1985 and 1988 are in the same regime. Our sample selection limits this issue to one cohort only, 1989.

any job transformation, e.g. from temporary to permanent).[10] We do not observe the high school graduation year. We assume a hypothetical age of graduation of 19 years old for all individuals, such that we define the treatment status by the birth year only (i.e. being 19 years old from 2007 onwards or not). Hence, we can compare two treated and two untreated cohorts of high-school graduates. We construct a panel at the individual level of working histories to document the exogenous shift in expected earnings after the reform. Our time window covers from zero to eleven years after age nineteen for each birth cohort (i.e. years 2005-2019). Finally, to test our third remark, we add to these data on high school degree owners, people with a higher education degree. This largest dataset is made of 6497948 monthly/year observations and 109126 individuals for the sample that covers all sectors.[11] The unbalanced nature of the panel reflects the different timing of individuals labour market entry.

Our second source comes from administrative records of a public university based in the Northern part of Italy. This university offers a wide range of bachelor degrees, master programmes and one-tier degrees. We use these data to test our second model's implication. Hence, we characterise and model these students' choices when they enrol for the first time at university. We focus on students who graduated high school between 18 and 20 years and enrolled at the university within three years. We restrict the analysis to the cohorts of those born between 1985 and 1990 to compare three treated and untreated cohorts. With these data, we enlarge the cohort window for two reasons. First, we can test whether the cohorts are similar in their propensity to achieve a high school degree. Second, we have much fewer observations here than in the other archives. Yet, in these data, we know the high school graduation year. We define our treatment status as graduating before or after the reform. Hence, the treatment depends on both the birth year and the age at diploma. We keep individuals from zero to six years from high-school graduation in 2005-2016. Hence, 694,594 monthly/year observations and 9150 individuals make our working sample. These 9150 individuals are observed at each distance from graduation between zero and six.

We merge these two administrative datasets to provide some robustness checks. These linked data constitute the INPS subsamples of individuals recorded in the university archive. We use these data to confirm what we find when using the entire INPS archives. This robustness check proves the consistency of the students' outcomes pattern and the population one to support our model.

[10] Data on mandatory communications regard job flows. Yet, this is not a big issue for our analysis. We focus on high-school graduates who enter the labour market for the first time. We verify that turnover data for a specific time window (in our case, 2005-2019) provide all the relevant information on the education level of any of them.

[11] When we employ the data for all workers in the private sector, we have 38829228 observations.

4 Remark 1: lower ability individuals will not go on education, but they earn more from their high school degree.

The first remark faces questions over increased earnings caused by the reform for high school graduates. We use a representative sample pulled from the INPS archive on social security contributions paid in all sectors to address these questions. These data are drawn from four birth dates for each month and year. We start by selecting the 1986-1989 cohorts to have two of them in each treated and untreated group. We link these data to the archive from Mandatory Communications to retain those with a high school degree. We observe the earnings of these cohorts from age 19, the typical age to achieve the high school degree in Italy, to 30, ending up with 4015519 observations. We do not know the exact high school graduation age. Yet, we consider it a minor issue if the retention rate in high school is the same across treatment groups.[12] We construct two different measures of earnings with these data. The first includes all earnings from dependent work, while the second also adds self-employment income. To distinguish the two, we name the first earnings and the second income. We report results on income in the main text and those on earnings in the Online Appendix. We employ another INPS archive on the earnings of all workers in the private sector. After applying the same selection rules, we have 25438187 observations. With these data, our two measures of earnings are monthly and daily earnings. We cannot conduct our analyses on wages because INPS data do not record the hours of work. This part of the study is meant as a robustness check and can be found in the Appendix.

We add two other outcomes to understand whether the treated group benefited from occupational earnings (income) premiums. Following Altonji and Zhong (2021), we estimate these premiums using an OLS estimator on a earnings/income equation using the dataset with high school achievers and those with a higher education degree. In all the other parts of this analysis, we will use the subsample of high school degree owners. Yet, here, we do not disregard how occupational premiums relate to the education degree. In these equations, we control for occupational dummies,[13] a gender indicator, year and month dummies, age as deviation from 19 and its interaction with the gender dummy, a dummy for whether the person has a university degree, and an indicator for working part-time. We then merge the estimated premiums to our data using the three digits International Standard Classification of Occupations (ISCO) codes.

We perform a two-way fixed effects model under the assumption of the absence of

[12] While we cannot prove this assertion with the INPS data, we provide evidence of it in the next Session 5.

[13] Chief executives, senior officials and legislators are the reference category.

heterogeneous treatment effects. We rule out heterogeneous treatment effects over groups by assuming that our cohorts share abilities, preferences, and labour market experience distributions. Once we control for age distance from 19 and time dummies, nothing is left. In Section 6 we provide some descriptive evidence that backs this assumption. Heterogeneous treatment effects over time is not a concern in our case because our treatment is constant and pre-determined to the labour market outcomes. We assume a common age profile between treatment and control groups. The treated group would have had the same age profile as the untreated one without the reform. However, we allow for gender-specific age profiles. We start by estimating the following regression model:

$$y_{ijt} = \sum_{a=0}^{11} a\alpha_a + \sum_{a=0}^{11} b\alpha_a * f_i + \sum_{r=2005}^{2019} c\theta_r + \sum_{m=2}^{12} e\mu_m + \beta d_j + \gamma f_i + u_{ijt} \quad (5)$$

The outcome y of individual i who belongs to the treatment group j at time t (year/month) is regressed on year θ_r and month μ_m dummies, the gender-specific age profile, female f_i and treatment d_j indicator functions. The treatment indicator takes the value of one (zero) for the two (un)treated cohorts (1986 and 1987 vs 1988 and 1989). In Equation 5, we restrict the treatment effect to be equal across gender and over the age profile. We then allow for a gender-specific impact:

$$y_{ijt} = \sum_{a=0}^{11} a\alpha_a + \sum_{a=0}^{11} b\alpha_a * f_i + \sum_{r=2005}^{2019} c\theta_r + \sum_{m=2}^{12} e\mu_m + \beta d_j + \gamma f_i + \delta d_j * f_i + u_{ijt} \quad (6)$$

We consider a more flexible model as the last one. It is a gender-specific age profile with year/month dummies estimated by each treatment group:

$$y_{ijt} = \sum_{a=0}^{11} a_j\alpha_a + \sum_{a=0}^{11} b_j\alpha_a * f_i + \sum_{r=2005}^{2019} c_j\theta_r + \sum_{m=2}^{12} e_j\mu_m + \gamma_j f_i + u_{ijt} \quad (7)$$

In each of the four rows of Table 1, our different measures of earnings are reported. In odd columns, we use data for all sectors from the INPS archive on social security contributions. In even columns, we employ the other INPS archive that covers the private sector only. Hence, in rows two and four, the earnings definition differs because not the same data are available to us. We display the treatment effects on monthly income (income from self-employment included) in columns 1, 3 and 5. The outcome is the daily earnings in the remaining columns. In the last two rows, we have our measure of the earnings/income occupational premium.

Table 1: Treatment effects: sample of high school graduates

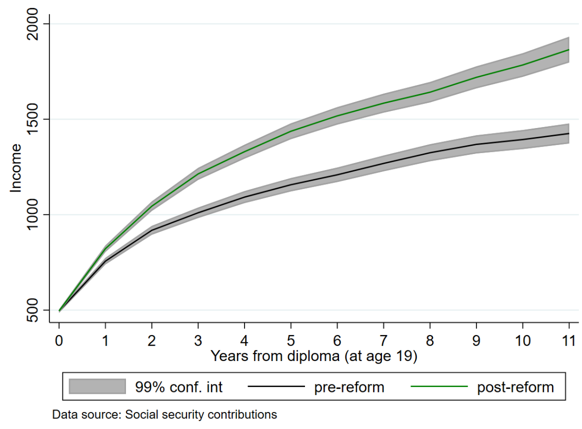
	All		Males		Females	
	All sectors	Private	All sectors	Private	All sectors	Private
Outcomes:	(1)	(2)	(3)	(4)	(5)	(6)
Monthly Earnings	-0.001	0.001	0.007^{***}	0.012^{***}	-0.008^{***}	-0.011^{***}
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
Monthly Income	-0.001	-	0.006^{***}	-	-0.014^{***}	-
	(0.002)	-	(0.002)	-	(0.002)	-
Daily Earnings	-	0.000	-	0.007^{***}	-	-0.007^{***}
	-	(0.000)	-	(0.000)	-	(0.000)
Monthly Earnings Occup. Premium	0.086^{***}	0.090^{***}	0.086^{***}	0.092^{***}	0.085^{***}	0.087^{***}
	(0.004)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Monthly Income Occup. Premium	0.096^{***}	-	0.096^{***}	-	0.096^{***}	-
	(0.004)	-	(0.000)	-	(0.000)	-
Daily Earnings Occup. Premium	-	0.100^{***}	-	0.102^{***}	-	0.098^{***}
	-	(0.000)	-	(0.000)	-	(0.000)

Notes: The model includes age distance from 19, year and month dummies, a gender indicator, the interaction of the gender indicator with the age distance dummies. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

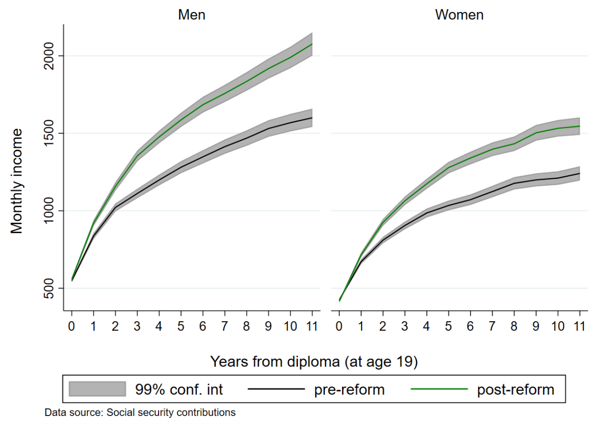
Table 1 summarises results for regression models 5 and 6. The first two columns show the homogeneous treatment effects that are not statistically different from zero for two out of four earnings measures. The 2007 reform raised the occupational earnings/income premiums, but it did not appear to impact the earnings and the income. However, the latter finding masks two relevant sources of heterogeneity. The first of them is reported in Columns 3 to 6 and is a stark gender differential. The reform benefited males but harmed females. The earnings increase (decrease) for the former (latter) group is slightly less than one percent. The gender gap in the treatment effect is minor in terms of occupational earnings premiums. The treatment group for both genders raised their occupational premiums by about nine percent. Auditing the final high school grade provides a guarantee to the labour market. Yet, the value of this guarantee is not the same for all when we look at the big picture of average earnings. Discrimination against women might seem to appear. To better understand how this discrimination takes root, we study the age profile of the reform impact in Figure 2. We report here results using the income measure from the INPS social security archive.

Figure 2 presents the age profile from 19 (0) to 30 (11) of the expected income or the expected income occupational premiums, estimated by Equation 7. We measure age as the distance from 19. In panels (a) and (c), these age profiles are averaged across gender; in panels (b) and (d), we relax this hypothesis. The expected outcomes are averaged within treatment status, and the value of their levels before and after the reform can be appreciated on the vertical axes. For each age from 19 to 30, we can calculate the treatment effect as the distance between the two

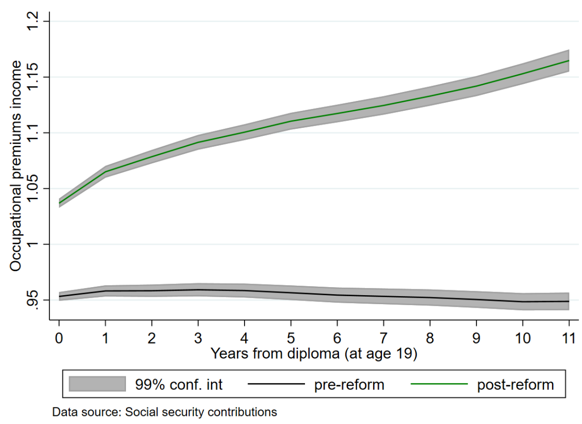
Figure 2: Age profiles of income by treatment group



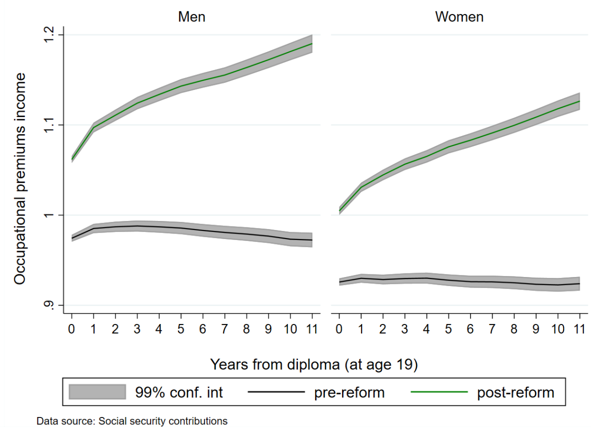
(a) Income: all workers



(b) Income: by gender



(c) Occupational income premiums: all workers



(d) Occupational income premiums: by gender

expected outcome functions. The gender gap in the expected outcomes is stark and amplifies with age. Women are paid less and have a flatter profile in the ages where the profile might grow steeper. Panel (b) of Figure 2 strikingly shows that men, but not so women, benefited from the reform at an increasing rate with age. For males, the treatment effect goes from slightly more than one per cent at age 19 to about 22 per cent at age 22, about 25 per cent at age 25 and 28, and about 30 per cent at age 30. For females, the impact is flatter and negative at age 19. It is about -2 per cent at age 19, +18 at age 22, 25 per cent at age 25 and 28, and then it decreases at age 30 and equal to about 24 per cent. When we look at the occupational income premiums as an outcome, both women and men gained from the reform. Moreover, the women's profile of the treatment effect is on par with the men's one. For men and women, the impact ranges from about nine per cent at age 19, 14 per cent at age 22, 17 per cent at age 25, 20 per cent at age 28 and about 22 per cent at age 30.

We document the robustness of these patterns in the Online Appendix A1.1 from a twofold perspective. In Figure A1, we use the same INPS data employed here but a different measure of earnings (i.e. all earnings from dependent work). In Figures A2 and A3, we use recorded monthly and daily earnings from the INPS archive on all jobs in the private sector.

5 Remark 2: higher ability individuals will increase their optimal amount of schooling

5.1 Empirical Model: first stage

Our second remark states the increased schooling of some more able individuals when the high school grade assessment is more credible and reliable. We leverage the 2007 reform as a game-changer to shift earnings. We use the university administrative data to test this assertion. In these data, we can observe someone if she enrols at this university. We cannot study the reform impact on university enrolment for two reasons. First, we do not observe those who graduated from high school and didn't enrol at university. Second, we do not have the data to argue how enrolment at this university could differ from the national average. In the Online Appendix B, we address this issue using national survey data to show robust results.

We exploit the entry choice between one and two-tier paths to ascertain the remark's validity. Those who register at the one-tier path are expected to have higher years of schooling than those who follow the two-tier track and stop as soon as they complete the first tier.[14] We follow a two stages approach. In the first stage, we estimate whether the reform altered the one-tier versus the two-tier choice. Here, we are testing whether some individuals would have enrolled at the two-tier rather than the one tier track hadn't the reform occurred. In the second stage, we study how enrolment at the one-tier path impacts educational outcomes. We use two-way fixed effects regressions to estimate the effect of the 2007 reform on the probability of enrolment at the one-tier or two-tier path. This regression model is our first stage that is specified as follows:

$$y_{ij} = \sum_{a=0}^{6} a\alpha_a + \sum_{j=1985}^{1990} c\theta_j + \beta d_j + \gamma X_{ij} + u_{ij} \tag{8}$$

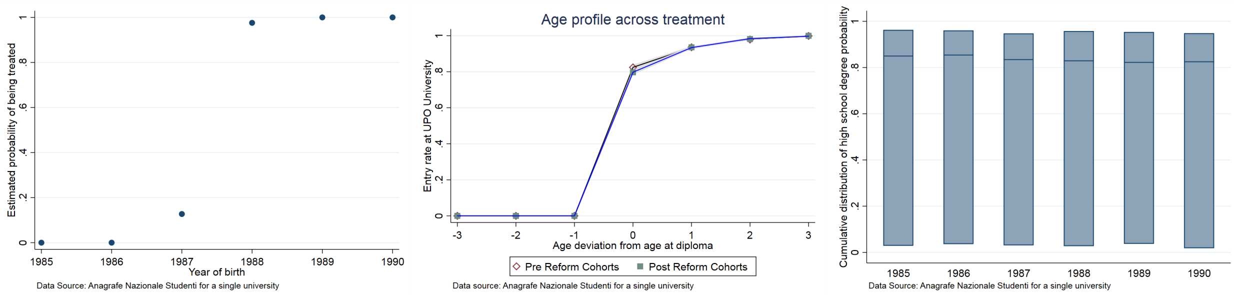
α control for the group effect and θ for the birth year effect. Each group α is made of individuals with the same year distance from high school graduation. All pupils born in the same year belong to a specific θ pool. Here the cohort of birth plays the role of time. We introduce in the model as covariates X: indicator functions for gender, public school, the type

[14]In the raw data, we observe that roughly 17% of those who chose the two-tier track enrolled to the second one within six years on the high school graduation.

of high school (lyceums, vocational, polytechnic); the high-school grade; the average high-school grade in the same type of school; the average high school grade in the same local labour market and the high school class size.

The exposure to treatment lies in being born in a cohort that turned the age at high school graduation after introducing the reform. In our selected sample, individuals graduate when aged between 18 and 20. Hence, some individuals are treated within the 1987 and 1988 cohorts at the same age, and others are not. This slight difference in the intensity of treatment is borne out by panel (a) of Figure 3.

Figure 3: Model validity



(a) Intensity of treatment

(b) Entry rate by treatment

(c) Cumulative probability distributions

Equation 8 estimates the effect under the standard common trends assumption when the treatment effect is constant across groups and over birth years. We discuss these hypotheses in turn. First, we can rule out heterogeneous treatment effects. Effects over birth years are likely constant. We compare contiguous and similar cohorts. Once we include in the model birth years dummies, no other form of heterogeneity might be left. These cohorts share similar observable characteristics described by Table B1 (See Online Appendix B1). Panel (b) of Figure 3 bears out the same entry rate at this university for both treatment groups. Second, effects over groups are constant over time because the first registration at the university is made once for all. Under these circumstances, our two-way fixed effects model is valid when the common trends assumption holds (de Chaisemartin and D'Haultfoeuille 2020). Panels (b) and (c) of Figure 3 stand for it, showing a clear common pre-trend in high school graduation. Cohorts could differ for the timing in this graduation, with some cohorts having older pupils. Panel (c) rules out this circumstance supporting the model's soundness. The cumulative distributions of high school graduation probability are on par over birth cohorts.[15] In Online Appendix B1, we report

[15]More multiple-comparison tests of these distributions are available upon request from the authors. We always find their statistical equality.

an additional test on the common trend hypothesis. We exploit its different timing to regress the probability of high school graduation on a trend, the treatment indicator, their interaction, dummies for age at graduation and birth year. The interaction term is not statistically different from zero supporting the common trend assumption. We complete this preliminary analysis proving the treatment exogeneity, and we report it in Online Appendix B1.

Auditing high school grade assessment induces some pupils to study more, switching from a two-tier to a longer one-tier track. Table 2 unfolds how the 2007 reform increased the probability to enrol at this longer track. It came as the enrolment rate at the two-tier track dropped by the same proportion (see Online Appendix B1). We read from columns 1 to 4 that this picture is consistent across all the sample periods. We estimate a four percentage points increase.

Table 2: First Stage: Impact on the enrolment probability at the one-tier track

Covariates:	(1)	(2)	(3)	(4)
Treatment	0.041***	0.039***	0.041***	0.055***
	(0.01)	(0.01)	(0.01)	(0.01)
Female	-0.001	-0.002	-0.018***	-0.026***
	(0.01)	(0.01)	(0.00)	(0.00)
High school final grade	0.005***	0.005***	0.005***	0.005***
	(0.00)	(0.00)	(0.00)	(0.00)
Average grade in the same type of high school	-0.003***	-0.003***	-0.003***	-0.003***
	(0.00)	(0.00)	(0.00)	(0.00)
Average grade in the same local labour market	0.005***	0.004***	0.005**	0.010***
	(0.00)	(0.00)	(0.00)	(0.00)
Public school	-0.029***	-0.028***	-0.034**	-0.041
	(0.01)	(0.01)	(0.01)	(0.02)
Polytechnic high school	0.009	0.011	0.006	-0.003
	(0.01)	(0.01)	(0.01)	(0.01)
Lyceums	0.128***	0.130***	0.114***	0.107***
	(0.01)	(0.01)	(0.01)	(0.01)
Vocational high school	0.011	0.011	-0.003	0.007
	(0.02)	(0.02)	(0.02)	(0.01)
Class size	0.005***	0.005***	0.006***	0.007***
	(0.00)	(0.00)	(0.00)	(0.00)
Sample period	2003-2010	2004-2009	2005-2008	2006-2007

Notes: The model includes age distance from high school graduation, birth year, region birth dummies. Standard errors are clustered at birth year, age distance from high school graduation level, age at high school graduation. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This figure comes out from the whole sample observed up to six years after high school graduation. But the enrolment choice is made once and for all up to two years after graduation. Tables in Online Appendix B1 show that our results do not vary when we restrict the age distance window to two years. They are also insensitive to how we cluster standard errors.

We conduct two analyses based on placebo reforms taking place in 2004 and 2002. Here, we report what we find under the first of the two. Results concerning the latter are in the Online Appendix B.

Table 3: Placebo First Stage: Impact on the enrolment probability at the one-tier track

Covariates:	(1)	(2)	(3)
Treatment	-0.030***	-0.028***	-0.033***
	(0.01)	(0.01)	(0.01)
Female	-0.014**	-0.015***	-0.016**
	(0.01)	(0.01)	(0.01)
High school final grade	0.004***	0.004***	0.004***
	(0.00)	(0.00)	(0.00)
Average grade in the same type of high school	-0.002***	-0.002***	-0.002***
	(0.00)	(0.00)	(0.00)
Average grade in the same local labour market	0.003*	0.003	0.005**
	(0.00)	(0.00)	(0.00)
Public school	0.009	0.008	0.015*
	(0.01)	(0.01)	(0.01)
Polytechnic high school	0.010*	0.010*	0.006
	(0.01)	(0.01)	(0.01)
Lyceums	0.136***	0.137***	0.129***
	(0.01)	(0.01)	(0.01)
Vocational high school	0.045***	0.046***	0.059***
	(0.01)	(0.01)	(0.02)
Class size	0.003**	0.003**	0.002
	(0.00)	(0.00)	(0.00)
Sample period	2001-2006	2002-2005	2003-2004

Notes: The model includes age distance from high school graduation, birth year, region birth dummies. Standard errors are clustered at birth year, age distance from high school graduation level, age at high school graduation. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The fake reform seems to reduce access to the one-tier track. This figure goes with a lower probability to enrol at the other track (see Table B9). These figures come from differences in the composition of treated and untreated groups rather than due to a fake reform. Tables B7 and B8 bear out this conclusion. Table B7 shows that covariates are less balanced, although differences are still slight. Table B8 reports the F-test statistics that reject the null hypothesis of the reform exogeneity. We draw the same conclusion when we consider the other fake reform. In the absence of other confounding policies, one expects contiguous birth cohorts to be similar at the national level once fixed effects are controlled. This property may fail when, as in our case, one moves from the national to a narrower level. The strength of our results lies in a switch from the two- to the one-tier track of the same proportion. This proportionality proves powerful to discharge concerns about biases due to cohorts specific selection into this university. Without this proportionality, it would be harder to argue that treated and untreated cohorts are similar. For instance, one could say that the reform induced more able (less) able people to enrol at this university, casting doubts on what the estimated effect is picking up. It could pick up the reform' s impact but also people's skills or both. People's abilities are not observable, but one way to detect this problem could be to look at the correlation between the reform and people's observable characteristics. The exogeneity test on the 2007 reform has come to pass, with the F statistics failing to reject the null.

5.2 Empirical Model: second stage

Our second stage is meant to estimate the impact of enrolment at the one-tier track on the probability of enrolment in the following years and graduation probability. Our age windows are from one up to six years after high school graduation for the former and three up to six years for the latter outcome. In such a setting, we have discrete outcomes with an endogenous binary regressor. We deal with it following Dong and Lewbel (2015) and employ their special regressor estimator. The reason lies in the inconsistency of the control function (i.e. ivprobit stata command) and maximum likelihood (i.e. biprobit stata command) estimators in a context such as ours.

Dong and Lewbel (2015) assume that the model includes a special regressor with four primary properties. First, it is exogenous and, second, it must be conditionally independent of the error term. Third, it enters as an additive term, and forth, it is continuously distributed with large support. A special regressor with greater kurtosis will prove more powerful. Its thick-tailed distribution is not strictly necessary, but it helps.

The binary choice special regressor has the threshold crossing form (see also Lewbel (2000)):

$$\begin{aligned} Y &= \mathbb{1}(X^e\beta_e + X^o\beta_o + V + \epsilon \geq 0) \\ X^e &= Z\alpha + X^o\beta_o + e \end{aligned} \tag{9}$$

Y denotes our discrete outcomes: the enrolment and graduation probability observed up to six years since high school completion. X^e is our discrete endogenous regressor, the probability of enrolment at the one-tier track; and Z is our instrument, a 2007 reform's treatment indicator. V is the special regressor; in our context, is age measured in days from birth. The month in which one was born matters for the educational and working career. On average, younger children perform worse than their peers because they sit the same exam earlier than older cohort members. This disadvantage stays on beyond compulsory schooling. For instance, Crawford, Dearden and Meghir (2010) find that with age 19/20, schooling participation declines monotonically with the month of birth. Hence, we expect the enrolment probability to increase with age measured in days for each age distance from high school graduation.

Table 4: Marginal Effects of enrolment at the one-tier track.

	Outcome: Enrolment probability			**Outcome: Graduation Probability**		
	Probit	IVProbit	SR	Probit	IVProbit	SR
Years from high school graduation	(1)	(2)	(3)	(4)	(5)	(6)
1	-0.005	2.380^{***}	0.090	-	-	-
	0.004	0.184	0.069	-	-	-
2	-0.019^{***}	0.644	0.029	-	-	-
	0.004	0.666	0.077	-	-	-
3	0.348^{***}	2.213^{***}	0.217^{***}	-0.312^{***}	-3.496^{***}	-0.237
	0.004	0.400	0.033	0.002	0.081	0.159
4	0.476^{***}	-0.404	0.300^{***}	-0.332^{***}	-3.146^{***}	-0.087
	0.004	0.608	0.103	0.002	0.009	0.150
5	0.342^{***}	2.944^{***}	0.209^{***}	-0.241^{***}	-2.908^{***}	-0.141
	0.004	0.073	0.062	0.003	0.046	0.143
6	0.210^{***}	1.974^{***}	0.165^{***}	-0.136^{***}	-2.664^{***}	-0.133
	0.004	0.528	0.050	0.004	0.119	0.194

Notes: For the special regressor model, we report bootstrapped standard errors (30 replications)

Table 4 reports our results. Columns 1 and 4 read the probit estimator; columns 2 and 5 the ivprobit one, and columns 3 and 6 the special regressor one. The latter appears more reliable than the former two. The probit estimator presents the wrong sign; the ivprobit is inconsistent and too large. Up to two years since high school graduation, enrolment at the one-tier track does not improve the chance to be enrolled at university compared to the two-tier one. In part, this result could be due to a low precision in the estimates. The number of replications in our bootstrap procedure of the special regressor model is limited to 30. But, it could also be the case that the drop-out probability is the same between the two tiers. The 2007 reform is our instrument. Hence, some individuals would have enrolled at the two-tier rather than the one-tier track without the reform. Our results prove that these individuals can make it even in the more challenging and longer one-tier track. Three years on the high school graduation and enrolment at the one-tier track leads to more years of schooling. Column 3 of Table 4 unveils this effect to persist after four, five and six years. This effect is not mechanical because the one-tier track is longer than the two-tier one. Our data includes those enrolled on the two-tier track and running across the second track after completing the first. To them, we assign a value of one to the indicator for being enrolled. The choice to enrol at the one-tier track brings about more education. Five years on the high school graduation and the increased enrolment probability is about 21 percentage points. This choice is not detrimental to university graduation, as borne out by Column 6. Those who switched the tier track have a probability of completing it as high as they would have it without switching. The reform proved valuable in two ways: it raised the years of schooling and changed the quality of these years of schooling. These claims are precisely what the second remark of our model tells.

6 Remark 3: increased earnings rise with schooling

Our third remark asserts that the earnings increase, which follows a fairer and more reliable high school grade assessment, rises with the schooling level. We estimate the labour market returns to a university degree to test it. We employ the universe of those born between 1986 and 1989 in the INPS archives. Two measures of income define our earnings functions. The first measure includes all earnings from dependent employment for which social security contributions are paid. The second one adds to the former income from self-employment activities to embrace all types of jobs. In our context, we would miss part of the story if we limited our attention to the private sector. Many of those who enrol at the one-tier track end up in occupations that cannot be found in a private sector archive as a doctor for the National Health System or a lawyer. The archive of mandatory communications provides the workers' education. We know the education degree but not when this degree was achieved. Hence, we focus on an age interval of 11 years since 19, the typical age to high-school graduation in Italy. We restrict our sample to those who have a high-school degree by the first time we observe them. We follow Altonji and Zhong (2021) and use a fixed-effects combination of high-school and university degrees to compare earnings before and after the university degree for those treated and untreated by the 2007 reform. In our analysis, a university degree can reflect both a one-tier or the first degree of the two-tier track. Our empirical model is the following:

$$y_{ijt} = a_1 + (\alpha_0 + \sum_{1}^{11} \alpha_{age})\mathbb{1}H_{ij} + \gamma_g \mathbb{1}G_{ijt} + \delta d_j + X_{ijt}\beta + u_{ijt} \tag{10}$$

H is an indicator function for having a high-school degree, equal to one for all our selected workers. G denotes a dummy variable that takes the value of one for workers with a university degree. The indicator function d stands for being treated by the 2007 reform.

Different hypotheses on u_{ijt} reflect different regression models. An OLS model estimates Equation 10 as it stands; a fixed-effects model writes $u_{ijt} = e_i + \epsilon_{ijt}$ and treats e_i as a time-invariant person specific component in the estimation. We follow Altonji and Zhong (2021) and decompose $u_{ijt} = \beta_{hg}\mathbb{1}HG_{ijt} + \nu_i + \epsilon_{ijt}$ with ν_i treated as random. HG_{ij} takes the value of one for those with the pair of high-school and university degrees observed the last time we saw them in the data. $\sum_1^{11} \alpha_{age}$ are age dummies measured as deviation from 19. These dummies adjust for experience the pre-graduate earnings of individuals who later obtain a university degree to provide the counterfactual earnings they would have earned had they not enrolled at university.

A fixed-effects estimator (FE) identifies the returns to a university degree limited to people observed with positive labour income before and after university graduation. In our sample, about 39% of people changed education status and would be used in such estimation. A combined high-school-university fixed effects estimator (FE_{hg}) makes full use of the data. In this context, we identify the effect by including the number of individuals observed only before and the many observed only after the university degree. We also retain those who will never graduate to adjust the counterfactual experience profile without university graduation. Yet, their occupations differ from the counterfactual ones for those who graduate. For this reason, we cannot simply compare earnings with a higher education degree to those without it.

Our primary sample is made of 3405373 observations for untreated individuals and 3092575 for treated ones. We use this sample to provide OLS, FE, and FE_{hg} estimates. Then, we restrict our sample to those who have obtained a university degree by the last time we observe them. Here we have 1329215 untreated observations and 1153214 treated ones. We can apply the OLS and FE estimators for this sample.

Differently from Altonji and Zhong (2021), we compare the returns to a university degree before and after the 2007 reform. Our regression model is the following:

$$y_{ijt} = a_1 + (\alpha_0 + \sum_{1}^{11} \alpha_{age}) \mathbb{1}H_{ij} + \gamma_g \mathbb{1}G_{ijt} + \gamma_{gd} \mathbb{1}G_{ijt} d_j + \delta d_j + X_{ijt}\beta + \beta_{hg} \mathbb{1}HG_{ij} + \nu_i + \epsilon_{ijt} \quad (11)$$

Reported estimates in Online Appendix C1, where we run Equation 10 separately for treatment status, are robust to this simple model specification. Hence, some restrictions we are putting in are harmless.

The strength of our data lies in having a highly representative sample of the entirety of the working population of the 1986-1989 birth cohorts. These cohorts are likely similar. Hence, we expect their distribution of abilities, preferences, and labour market experience profile to be the same once we control for time-fixed effects.[16] If true, treated and untreated cohorts have similar distributions of time gaps between educational experience and earnings observations. We prove this claim in online Appendix C1 and Figure 4. Column 1 of Table C2 reports the 10th, 25th, 50th, 75th and 90th quantiles of the number of years from age 19 for earnings observed before a university degree. The 10th, 50th and 90th quantiles are 1, 11 and 11 (the maximum) for the

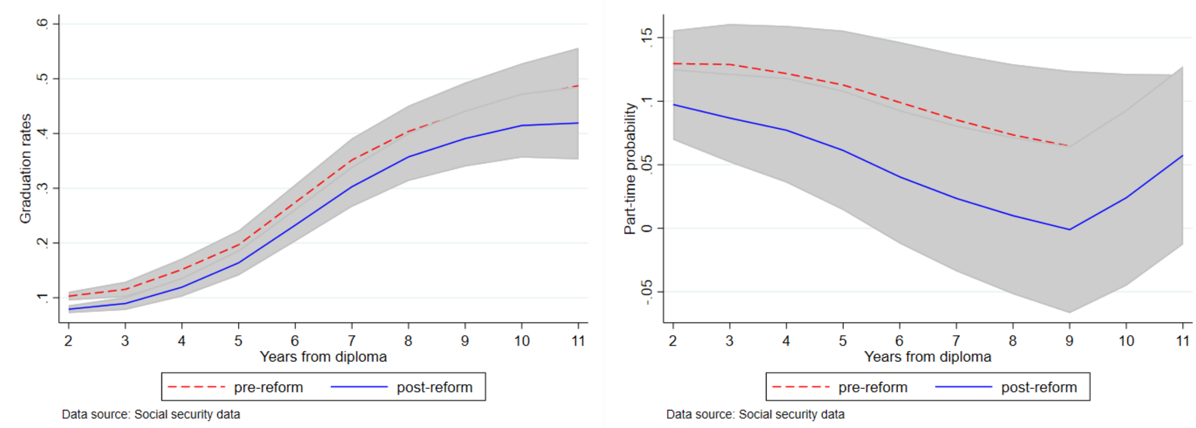
[16]Controlling for age and time dummies is equivalent to controlling for birth year dummies.

untreated group. The treated group differs for the 10th quantile, which is equal to 3. About 50% of pre-university degree earnings observations occur between 0 and 3 (4) years from age 19 for the untreated (treated) group (Column 2). Column 3 bears out the 10th, 50th, and 90th quantiles of years from 19 to observing post-university earnings equal 10 (9), 11 and 11 for the untreated (treated) group. The corresponding values are the same for individuals with earnings observed before and after the university degree (Column 4). The equivalence of Columns 3 and 4 highlights that the FE and FE_{hg} estimators are likely similar.

Figure 4 shows that treated and untreated cohorts share the same age profile for the propensity of being observed with a university degree or in a part-time job. What makes the difference between these two groups cannot be attributed to these two factors. It is worth stressing that the same age profile in university graduation does not contradict our model predictions. Our second remark calls on an increase in the years of schooling. Switching from the two-tier to the one-tier track raise the years of schooling of those who graduate.

Finally, Table C3 presents the age distribution of our sample by treatment groups. The 10th, 50th and 90th quantiles are 21, 26 and 30 for the untreated one. These quantiles are almost identical for the treated one, differing the 90th (29). These are also the age distributions for those with a high school degree. The median age is slightly smaller (25 and 24) before graduation and higher (28) after it.

Figure 4: Similarity of Treated and Untreated Cohorts



(a) Probability of being observed with a university degree

(b) Probability of working part-time

Hence, we meet the two conditions under which the combined high-school-university graduation fixed effects is unbiased. First, new information on abilities and preferences impacts earnings in the same way as in the counterfactual case in which the person would not have enrolled at university. The 2007 reform does not harm this assumption because it updates information before the observed pre-university degree earnings. Our sample covers individuals aged 19 or more with a high-school degree. It is an improbable event that they could have worked while at the high school. Second, we assume a common experience profile conditional on high school graduation.

Another challenge is to understand how much of the returns to a university degree are due to job occupations. We consider the occupational earnings (income) premiums as outcomes to address this issue. Table C1 reports the summary statistics of our outcomes by treatment groups.

Table 5: Returns to a university degree

	Baseline			**Differential impact of the 2007 reform**		
Outcomes:	OLS	FE	FE_{hg}	OLS	FE	FE_{hg}
Earnings	0.030^{***}	0.299^{***}	0.423^{***}	0.049^{***}	0.102^{***}	0.061^{***}
	0.005	0.009	0.008	0.007	0.013	0.007
Income	0.041^{***}	0.301^{***}	0.433^{***}	0.045^{***}	0.102^{***}	0.057^{***}
	0.005	0.009	0.008	0.007	0.013	0.007
Earnings Occupational Premium	-0.010^{***}	-	0.035^{***}	0.004^{*}	-	0.006^{**}
	0.002	-	0.002	0.002	-	0.002
Income Occupational Premium	-0.006^{***}	-	0.038^{***}	0.095^{***}	-	0.005^{**}
	0.002	-	0.002	0.003	-	0.002

Notes: The model includes age distance from 19, year and month dummies, a gender indicator, the interaction of the gender indicator with the age distance dummies. Standard errors are clustered at individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5 reports our results. We estimate Equation 11 for the entire sample. The first three columns read the OLS, FE and FE_{hg} baseline impact that applies to the untreated group. The last three columns measure the differential impact for the treated group compared to the untreated one. Each of the four rows refers to a different outcome. Looking at the first three columns, we observe a theme consistent with Altonji and Zhong (2021). The FE_{hg} is large relative to the OLS one because OLS is small. The gap between these two coefficients negatively correlates with the average for the graduate degree of the high school's degree premium. Hence, OLS tends to understate returns to graduation of students from low-paying jobs. These students accept these jobs as a temporary measure in waiting for a better job (or job conditions) after graduation.

The last three columns show the estimated impact of the 2007 reform. They highlight how much the treated group gained in earnings and income and the extent to which these gains come from occupational premiums. OLS estimate is still downward bias, but not much. Comparing similar cohorts with a common experience profile proves effective in reducing the bias. FE estimate is larger than FE_{hg}. People who are more likely to be observed before and after graduation periods are those who graduate earlier and possibly more able. We find that the

2007 reform raised the earnings (income) returns to university degrees by six percentage points. Occupation accounts for about 0.6 percent of the return.

In Online Appendix C1 (see Tables C4-C6), we replicate our analysis for the subsample of education changers and estimating it separately by treatment groups. Results are robust, and it is worth stressing two points. First, the differential impact of the 2007 reform estimated by FE for the subsample is slightly smaller (about three percentage points). We attribute it to having a few years of observations after graduation. Second, the impact of the reform as estimated by Equation 11 is the same as the difference of the treatment group-specific regression model.

Our figures could be lower bound estimates of the returns over the workers' life-cycle because we observe just a few years after graduation. For the same reason, we cannot directly compare all our estimates presented in Section 4 and those reported here. Yet, looking at the average effects for men, our evidence supports the model predictions. The reform impact on the returns to graduation is twice as high as the returns to high school degrees (about two versus one percentage point). Female high school achievers were penalised, but going on to university after the reform paid off. Women who did it increased their income by about nine percentage points (see Table C8).

In Section 4, we show that the reform gains increased substantially with the labour experience of high school achievers. Yet, we do not have the data to test this hypothesis for graduates. If we had, it would also be more challenging to disentangle the reform impact. As predicted by our model, the labour market likely updates its beliefs on people's skills and knowledge once they graduate.

Our results send a clear message: auditing the high-school grade assessment matters. It matters because it impacts the choices pupils make and the skills and knowledge they develop. It matters because the labour market values and rewards it.

7 Conclusions

Due to the COVID 19 pandemic's disruption of pupils' studies, high-school exams have been scrapped worldwide and primarily replaced with a teacher-assessed grading system. The need to make sure pupils' grades were as fair as possible calls for a debate on how to provide this information optimally. A natural experiment to study this issue can hardly be found.

A 2007 reform in Italy provides us with a unique setting to study this issue. This reform enforced the presence of external examiners and changed the composition of the board. Hence, we can ascertain the impact of auditing the high-school final grades on educational and labour market outcomes. We sketch a simple model to make clear what we expect from this reform. Three are the model's predictions that we test.

First, auditing the final grades raises the earnings of high school graduates. We find evidence of this remark using two different INPS archives in a two-way fixed effects model. The first archive contains the entirety of Italian employees in the private sector. To support our claim, we restrict the sample to two untreated and two treated cohorts to compare their earnings profile. The second archive is a random sample of the universe of Italian workers in all sectors. This second source is based on the social security contributions paid, and we can include an individual's self-employment periods. We select the same cohorts, and we show that the 2007 reform on average increased male but not female earnings and income. We find evidence that the value of auditing differs across genders. We also document stark gender gaps in the age profile of earnings and income. Moreover, women and men benefited from the reform when the labour market experience increased but women less.

Our second remark states that auditing the high school final grades impacts the choice made by pupils and the skills and knowledge they develop. We use administrative data of an Italian university to underpin this claim. We compare three treated and untreated cohorts in a two-way fixed effects model to show that enrolment at the one-tier track increased by four percentage points. This shift is due to students choosing the one-tier rather than the two-tier track. We prove that this behaviour raised their years of schooling. For instance, five years on high school completion, having enrolled at the one-tier track increases the probability of being enrolled still by about 21 percentage points. Yet, we do not estimate any harmful effect on the probability of graduation.

Our third model prediction asserts that the increase in earnings rises with the years of schooling. We employ a randomised sample of the entirety of workers in all sectors to prove it. We carry out a combined fixed effects regression model as suggested by Altonji and Zhong (2021). We extend their model to a context where we compare cohorts who find themselves in similar circumstances but before or after the reform treatment. About six percentage points is the estimated advantage in the earnings returns to graduation led by the reform. Going on to university after the reform paid off, mainly for women. The different choices they made on the tier track could have led them to find better jobs.

Many institutions have hailed to tackle the education crisis triggered by the Covid-19 pandemic. On the one hand, the pandemic widened fissures in education and likely derailed the social and economic progress of many from disadvantaged families. One could also add to the list the costs in the well-being due to the difficult circumstances. On the other hand, we show that it could have brought other costs in terms of earnings losses due to the disruptions in the high school final grade. We prove that these costs are not negligible. Putting all these costs together, these cohorts of youth are hit hard by the pandemic. More generally, like auditing a firm gives a guarantee to the stakeholders, auditing the high school final grades gives a guarantee valued by the labour market.

References

Altonji, J., Arcidiacono, P. and Maurel, A. (2016). The analysis of field choice in college and graduate school:determinants and wage effects., Vol. 5, Elsevier, chapter Handbook of the Economics of Education, Chapter 7, pp. 305–396.

Altonji, J., Blom, E. and Meghir, C. (2012). Heterogeneity in human capital investments: High school curriculum, college major, and careers, *Annual Review of Economics* **4**(1): 185–223.

Altonji, J. G. and Zhong, L. (2021). The labor market returns to advanced degrees, *Journal of Labor Economics* **39**(2): 303 – 360.

Becker, G. (1967). *Human Capital and the Personal Distribution of Income*, 1 edn, Ann Arbor Michigan: University of Michigan Press.

Bertrand, M., Mogstad, M. and Mountjoy, J. (2021). Improving educational pathways to social mobility: Evidence from norway's "reform 94", *Journal of Labor Economics* **39**(4): 965–1010.

Callaway, B. and Sant'Anna, P. (2020). Difference-in-differences with multiple time periods, *Papers*, arXiv.org.

Card, D. (1999). Chapter 30 the causal effect of education on earnings, *in* O. C. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, Vol. 3, Part A of *Handbook of Labor Economics*, Elsevier, pp. 1801 – 1863.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems, *Econometrica* **69**(5): 1127–60.

Crawford, C., Dearden, L. and Meghir, C. (2010). When you are born matters: the impact of date of birth on educational outcomes in england, *IFS Working Papers W10/06*, Institute for Fiscal Studies.

de Chaisemartin, C. and D'Haultfoeuille, X. (2018). Fuzzy differences-in-differences, *Review of Economic Studies* **85**(2): 999–1028.

de Chaisemartin, C. and D'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects, *American Economic Review* **110**(9): 2964–96.

Demski, J. and Swieringa, R. (1974). A cooperative formulation of the audit choice problem, *The Accounting Review* pp. 506–513.

Dong, Y. and Lewbel, A. (2015). A simple estimator for binary choice models with endogenous regressors, *Econometric Reviews* **34**(1-2): 82–105.

Goodman-Bacon, A. (2018). Difference-in-differences with variation in treatment timing, *NBER Working Papers 25018*, National Bureau of Economic Research, Inc.

Kinsler, J. and Pavan, R. (2015). The specificity of general human capital: Evidence from college major choice, *Journal of Labor Economics* **33**(4): 933 – 972.

Levhari, D. and Weiss, Y. (1974). The effect of risk on the investment in human capital, *American Economic Review* **64**(6): 950–963.

Lewbel, A. (2000). Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables, *Journal of Econometrics* **97**(1): 145–177.

Low, H., Meghir, C. and Pistaferri, L. (2010). Wage risk and employment risk over the life cycle, *American Economic Review* **100**(4): 1432–1467.

Simunic, D. (1980). The pricing of audit services - theory and evidence, *Journal of Accounting Research* **18**(1): 161–190.

Sun, L. and Abraham, S. (2020). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects, *Papers*, arXiv.org.

Wiswall, M. and Zafar, B. (2015). Determinants of college major choice: Identification using an information experiment, *Review of Economic Studies* **82**(2): 791–824.

Zafar, B. (2013). College major choice and the gender gap, *Journal of Human Resources* **48**(3).

To whom it may concern: the value of auditing grade assessment on educational and labour market outcomes.

October 2021

Content

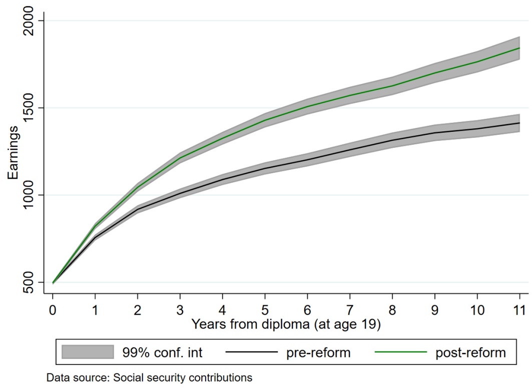
We organise this appendix as follows. Section A1 complements our analysis on the first remark; Section B1 refers to our second remark and Section C1 to our third one. Additional materials can be found in Section D1.

A1 Additional empirical analysis on our first remark

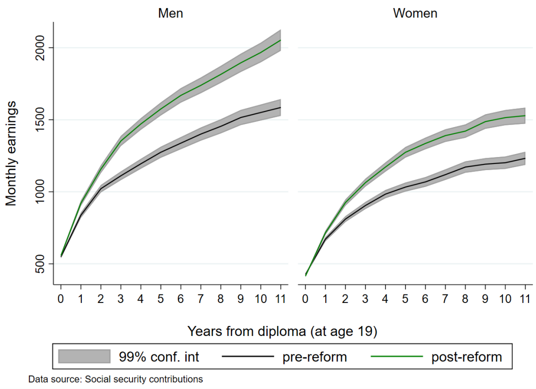
A1.1 Robustness using different earning measure and INPS archive for the entirety of the private sector

A1.1.1 Source INPS social security contributions: all earnings from dependent work

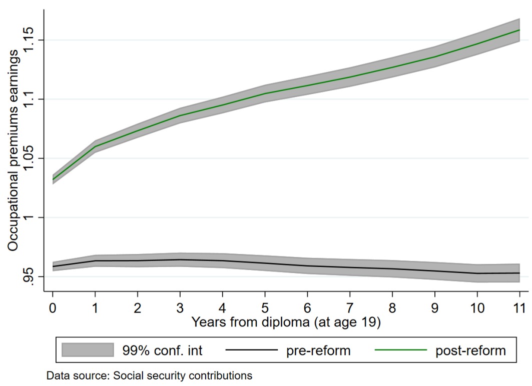
Figure A1: Age profiles of monthly earnings by treatment group



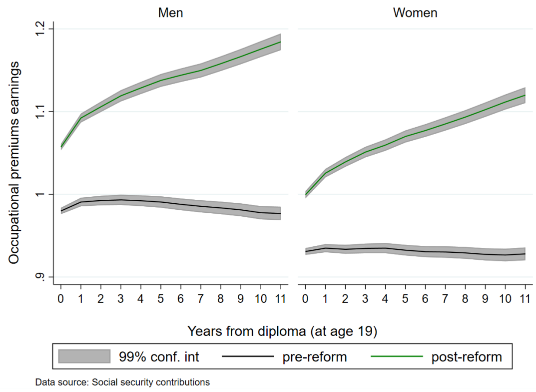
(a) Monthly Earnings: all workers



(b) Monthly Earnings: by gender



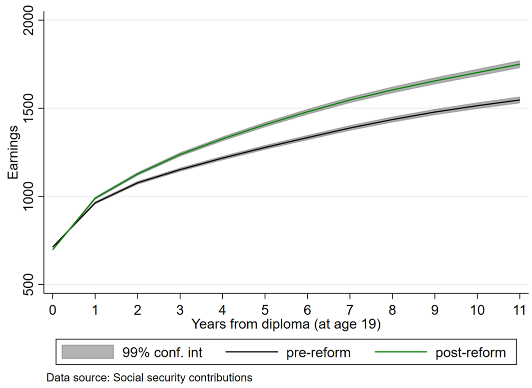
(c) Occupational earnings premiums: all workers



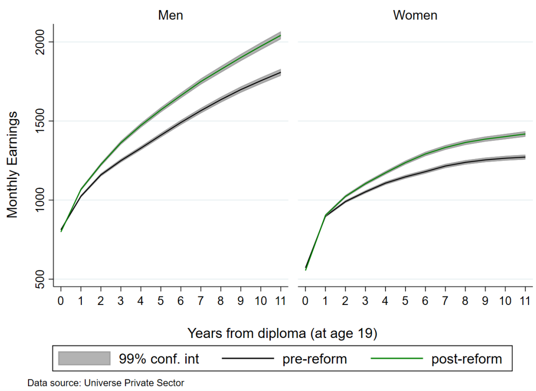
(d) Occupational earnings premiums: by gender

A1.1.2 Source INPS, universe of workers in the private sector

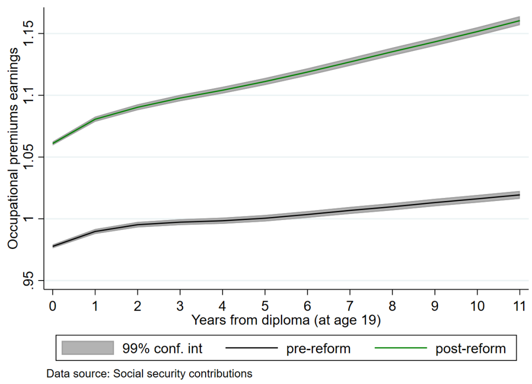
Figure A2: Age profiles of monthly earnings by treatment group



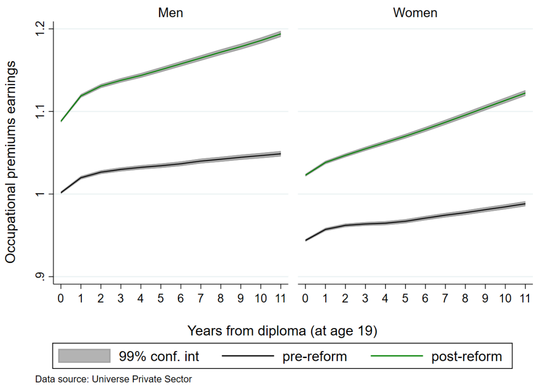
(a) Monthly Earnings: all workers



(b) Monthly Earnings: by gender

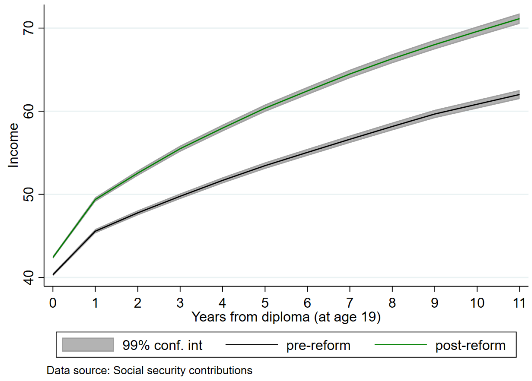


(c) Occupational monthly earnings premiums: all workers

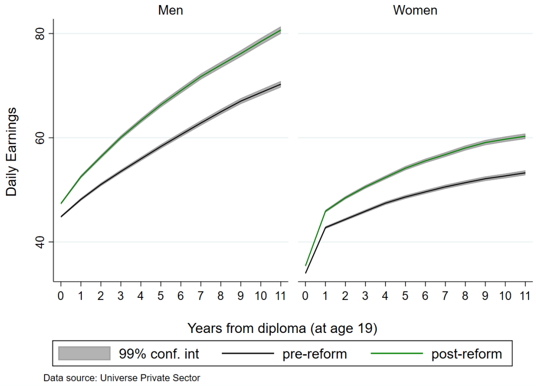


(d) Occupational monthly earnings premiums: by gender

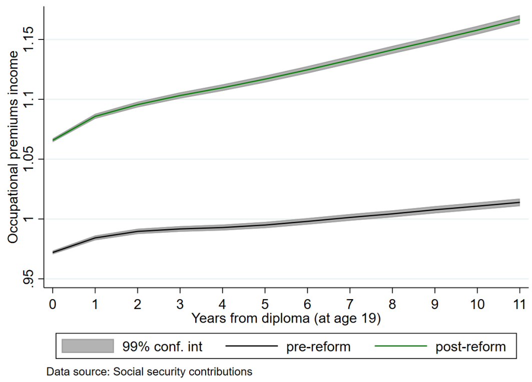
Figure A3: Age profiles of daily earnings by treatment group



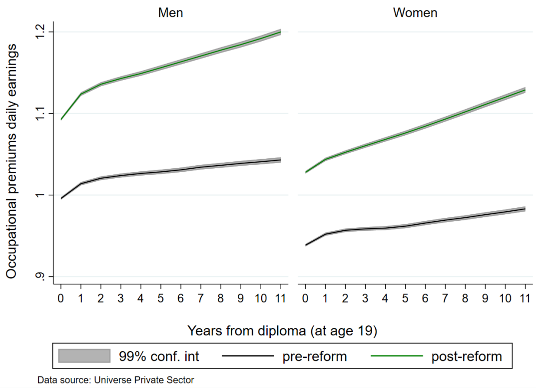
(a) Daily Earnings: all workers



(b) Daily Earnings: by gender



(c) Occupational daily earnings premiums: all workers



(d) Occupational daily earnings premiums: by gender

B1 Additional empirical analysis on our second remark

Table B1: Summary statistics on observable characteristics

	Controls		**Treated**	
	Mean	Std. Dev.	Mean	Std. Dev.
Female	0.64	0.48	0.63	0.48
Age at high school graduation	19.05	0.33	19.14	0.39
Age	20.46	0.65	20.55	0.69
High school graduation year	2004.99	0.82	2008.08	0.90
High school final grade	80.30	12.98	77.09	11.94
High school graduation region	2.30	4.17	2.88	5.07
Public school	0.93	0.26	0.93	0.25
Polytechnic high school	0.35	0.48	0.28	0.45
Lyceums	0.51	0.50	0.60	0.49
Vocational high school	0.10	0.31	0.10	0.29
Class size	20.72	2.72	21.12	2.45
Average grade in the same type of high school	79.34	4.46	76.39	4.02
Average grade in the same local labour market	79.18	2.14	76.30	1.88

Table B2: Testing common trends on the high school graduation probability

	First difference	Level
Trend	0.0018	0.1890***
	(0.0024)	(0.0010)
Trend*Treatment	0.0038	-0.0008
	(0.0046)	(0.0017)
Treatment	-0.0188	0.0038
	(0.0228)	(0.0078)
Number of observations	70770	80880

Table B3: Treatment exogeneity

	Treatment
Female	0.00**
	(0.00)
High school final grade	-0.00*
	(0.00)
Average grade in the same type of high school	-0.00*
	(0.00)
Average grade in the same local labour market	-0.01**
	(0.01)
Public school	0.01**
	(0.00)
Polytechnic high school	0.02*
	(0.01)
Lyceums	0.02
	(0.01)
Vocational high school	0.02**
	(0.01)
Class size	-0.00
	(0.00)
F test individual characteristics	1.654
F test school characteristics	1.237
F test all covariates	0.846
Age, birth year, region birth dummies	YES

Table B4: First Stage: Impact on the enrolment probability at the two-tier track

Covariates:	(1)	(2)	(3)	(4)
Treatment	-0.037***	-0.035***	-0.037***	-0.050***
	(0.01)	(0.01)	(0.01)	(0.01)
Female	0.000	0.001	0.017***	0.025***
	(0.01)	(0.01)	(0.00)	(0.00)
High school final grade	-0.005***	-0.005***	-0.005***	-0.005***
	(0.00)	(0.00)	(0.00)	(0.00)
Average grade in the same type of high school	0.003***	0.003***	0.003***	0.003***
	(0.00)	(0.00)	(0.00)	(0.00)
Average grade in the same local labour market	-0.005***	-0.004***	-0.005**	-0.009***
	(0.00)	(0.00)	(0.00)	(0.00)
Public school	0.027***	0.027**	0.032**	0.043*
	(0.01)	(0.01)	(0.01)	(0.02)
Polytechnic high school	-0.008	-0.010	-0.005	0.003
	(0.01)	(0.01)	(0.01)	(0.01)
Lyceums	-0.127***	-0.129***	-0.114***	-0.106***
	(0.01)	(0.01)	(0.01)	(0.01)
Vocational high school	-0.009	-0.011	0.005	-0.004
	(0.02)	(0.02)	(0.02)	(0.01)
Class size	-0.006***	-0.005***	-0.006***	-0.007***
	(0.00)	(0.00)	(0.00)	(0.00)
Sample period	2003-2010	2004-2009	2005-2008	2006-2007

Notes: The model includes age distance from high school graduation, birth year, region birth dummies. Standard errors are clustered at birth year, age distance from high school graduation level, age at high school graduation. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B1.1 Robustness: Changing the age distance window

Table B5: First Stage: Impact on the enrolment probability at the one-tier track

Covariates:	(1)	(2)	(3)	(4)	(5)
Treatment	0.041**	0.040***	0.041***	0.041***	0.041**
	(0.01)	(0.01)	(0.01)	(0.00)	(0.02)
Female	-0.019**	-0.018***	-0.018***	-0.018***	-0.018***
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
High school final grade	0.005***	0.005***	0.005***	0.005***	0.005***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Average grade in the same type of high school	-0.003**	-0.003***	-0.003***	-0.003***	-0.003***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Average grade in the same local labour market	0.005	0.005*	0.005**	0.005***	0.005***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Public school	-0.034	-0.034*	-0.034**	-0.034***	-0.034***
	(0.02)	(0.02)	(0.01)	(0.00)	(0.01)
Polytechnic high school	0.005	0.005	0.006	0.006	0.006
	(0.02)	(0.02)	(0.01)	(0.00)	(0.01)
Lyceums	0.114***	0.114***	0.114***	0.114***	0.114***
	(0.02)	(0.02)	(0.01)	(0.00)	(0.01)
Vocational high school	-0.004	-0.004	-0.003	-0.003	-0.003
	(0.04)	(0.03)	(0.02)	(0.00)	(0.02)
Class size	0.006***	0.006***	0.006***	0.006***	0.006***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age distance window from graduation	0	0-1	0-2	0-2	0-2
Standard Errors	AYA	AYA	AYA	ROB	AYAC

Notes: The model includes age distance from high school graduation, birth year, region birth dummies. Standard errors are clustered at birth year, age distance from high school graduation level, age at high school graduation (AYA); birth year, age distance from high school graduation level, age at high school graduation and birth region (AYAC); robust (ROB). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B6: First Stage: Impact on the enrolment probability at the two-tier track

Covariates:	(1)	(2)	(3)	(4)	(5)
Treatment	-0.037**	-0.036***	-0.037***	-0.037***	-0.037**
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)
Female	0.018**	0.017***	0.017***	0.017***	0.017***
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
High school final grade	-0.005***	-0.005***	-0.005***	-0.005***	-0.005***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Average grade in the same type of high school	0.003**	0.003***	0.003***	0.003***	0.003***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Average grade in the same local labour market	-0.005	-0.005**	-0.005**	-0.005***	-0.005**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Public school	0.032	0.032*	0.032**	0.032***	0.032**
	(0.02)	(0.02)	(0.01)	(0.00)	(0.01)
Polytechnic high school	-0.005	-0.005	-0.005	-0.005	-0.005
	(0.02)	(0.02)	(0.01)	(0.00)	(0.01)
Lyceums	-0.113***	-0.113***	-0.114***	-0.114***	-0.114***
	(0.02)	(0.02)	(0.01)	(0.00)	(0.01)
Vocational high school 0.005	0.005	0.005	0.005	0.005	
	(0.04)	(0.03)	(0.02)	(0.00)	(0.02)
Class size	-0.006***	-0.006***	-0.006***	-0.006***	-0.006***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age distance window from graduation	0	0-1	0-2	0-2	0-2
Standard Errors	AYA	AYA	AYA	ROB	AYAC

Notes: The model includes age distance from high school graduation, birth year, region birth dummies. Standard errors are clustered at birth year, age distance from high school graduation level, age at high school graduation (AYA); birth year, age distance from high school graduation level, age at high school graduation and birth region (AYAC); robust (ROB). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B1.2 Placebo Reform in 2004

Table B7: Summary statistics on observable characteristics

	Controls		**Treated**	
	Mean	Std. Dev.	Mean	Std. Dev.
Female	0.65	0.48	0.62	0.48
Age at high school graduation	19.10	0.37	19.15	0.40
Age	21.62	1.56	21.63	1.58
High school graduation year	2003.00	0.00	2004.54	0.58
High school final grade	80.05	12.69	79.27	12.89
High school graduation region	1.95	3.53	2.23	4.08
Public school	0.92	0.27	0.93	0.26
Polytechnic high school	0.34	0.47	0.36	0.48
Lyceums	0.51	0.50	0.50	0.50
Vocational high school	0.11	0.31	0.11	0.31
Class size	20.40	3.03	20.65	2.80
Average grade in the same type of high school	79.91	4.89	78.92	4.80
Average grade in the same local labour market	79.87	2.41	78.72	2.29

Table B8: Treatment exogeneity

	Treatment
Female	-0.02*
	(0.01)
High school final grade	-0.00*
	(0.00)
Average grade in the same type of high school	-0.00*
	(0.00)
Average grade in the same local labour market	-0.01**
	(0.01)
Public school	-0.01
	(0.01)
Polytechnic high school	0.04**
	(0.02)
Lyceums	0.01
	(0.01)
Vocational high school	0.06**
	(0.03)
Class size	-0.00**
	(0.00)
F test individual characteristics	2.172
F test school characteristics	3.140
F test all covariates	4.083
Age, birth year, region birth dummies	YES

Table B9: Placebo First Stage: Impact on the enrolment probability at the two-tier track

Covariates:	(1)	(2)	(3)
Treatment	-0.055***	-0.058***	-0.051***
	(0.02)	(0.02)	(0.02)
Female	0.056***	0.054***	0.074***
	(0.01)	(0.01)	(0.01)
High school final grade	0.000	-0.000	0.000
	(0.00)	(0.00)	(0.00)
Average grade in the same type of high school	0.002*	0.002*	0.003**
	(0.00)	(0.00)	(0.00)
Average grade in the same local labour market	-0.006***	-0.006***	-0.010***
	(0.00)	(0.00)	(0.00)
Public school	0.110***	0.113***	0.098***
	(0.02)	(0.02)	(0.02)
Polytechnic high school	-0.003	-0.005	0.009
	(0.02)	(0.02)	(0.02)
Lyceums	-0.048**	-0.053**	-0.033*
	(0.02)	(0.02)	(0.02)
Vocational high school	-0.097***	-0.095***	-0.099***
	(0.02)	(0.03)	(0.03)
Class size	-0.001	-0.001	-0.002
	(0.00)	(0.00)	(0.00)
Sample period	2001-2006	2002-2005	2003-2004

Notes: The model includes age distance from high school graduation, birth year, region birth dummies. Standard errors are clustered at birth year, age distance from high school graduation level, age at high school graduation. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B1.3 Placebo Reform in 2002

Table B10: Summary statistics on observable characteristics

	Controls		Treated	
	Mean	Std. Dev.	Mean	Std. Dev.
Female	0.66	0.47	0.66	0.47
Age at high school graduation	19.00	0.00	19.00	0.00
Age	21.48	1.46	21.52	1.51
High school graduation year	2001.00	0.00	2002.00	0.00
High school final grade	79.86	12.08	80.48	12.51
High school graduation region	1.65	2.83	1.49	2.08
Public school	0.93	0.26	0.93	0.26
Polytechnic high school	0.33	0.47	0.34	0.47
Lyceums	0.50	0.50	0.52	0.50
Vocational high school	0.07	0.25	0.09	0.28
Average grade in the same type of high school	79.83	5.14	78.46	4.93
Average grade in the same local labour market	80.03	2.47	78.20	2.21

Table B11: Treatment exogeneity

	Treatment
Female	0.01
	(0.01)
High school final grade	0.00
	(0.00)
Average grade in the same type of high school	-0.01
	(0.00)
Average grade in the same local labour market	-0.08*
	(0.01)
Public school	0.03
	(0.03)
Polytechnic high school	0.13*
	(0.01)
Lyceums	0.15*
	(0.01)
Vocational high school	0.18**
	(0.01)
F test individual characteristics	1.763
F test school characteristics	0.607
F test all covariates	0.607
Age, birth year, region birth dummies	YES

Table B12: Placebo First Stage: Impact on the enrolment probability at the one-tier track

Covariates:	(1)
Treatment	0.010**
	(0.00)
Female	0.004
	(0.00)
High school final grade	0.003***
	(0.00)
Average grade in the same type of high school	0.000
	(0.00)
Average grade in the same local labour market	0.002
	(0.00)
Public school	0.023***
	(0.00)
Polytechnic high school	-0.022
	(0.01)
Lyceums	0.116***
	(0.02)
Vocational high school	-0.016*
	(0.01)
Sample period	2001-2002

Notes: The model includes age distance from high school graduation, birth year, region birth dummies. Standard errors are clustered at birth year, age distance from high school graduation level, age at high school graduation. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B13: Placebo First Stage: Impact on the enrolment probability at the two-tier track

Covariates:	(1)
Treatment	-0.010
	(0.02)
Female	0.019
	(0.01)
High school final grade	0.000
	(0.00)
Average grade in the same type of high school	-0.001
	(0.00)
Average grade in the same local labour market	0.002
	(0.00)
Public school	0.098***
	(0.02)
Polytechnic high school	0.067***
	(0.01)
Lyceums	0.000
	(0.02)
Vocational high school	-0.003
	(0.01)
Sample period	2001-2002

Notes: The model includes age distance from high school graduation, birth year, region birth dummies. Standard errors are clustered at birth year, age distance from high school graduation level, age at high school graduation. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B2 External validity

In order to provide external validity to our main findings, we analyse the choices of a nationally representative sample of high school graduates before and after the 2007 reform that introduced audited grades. We aim at proving whether the Italian graduates' behaviour changed consistently with the theoretical model and the empirical findings on our university administrative data. We employ two cross-sections of the Italian National Institute of Statistics (ISTAT) survey *Percorsi di Studio e di Lavoro dei Diplomati.* This survey collects information about high school graduates a few years after their graduation. The first cross-section covers the 2004 graduates' cohort, examined by all internal teachers; the second is on the 2007 cohort, the first to be assessed by all external members except one. The 2004 cohort was interviewed three years after graduation (i.e. in 2007), while the 2007 cohort four years after (i.e. in 2011). This does not represent an issue in our case, as we are interested in retrospective information, namely in the choices made right after their high school graduation.

We estimate simple linear probability models for the following outcomes. First, the probability to further study after the high school diploma in whatever type of post-secondary institution. Second, the probabilities to enrol at a two-tier (three-years plus eventually two years) degree programme, at a one tier (five or six years) degree programme, or at any post high school non-tertiary degree program for those who decided to continue their studies. We control for the students' gender, age at diploma, area of residence, type of high school attended (public or private), and type of high school degree achieved (academic, technical or professional). We also exploit information about students' ability by adding their final grades both at lower secondary school (i.e. at the compulsory schooling level) and at the high school. The variable of interest is a dummy *Post* which is 0 if the student graduated in 2004, 1 if she graduated in 2007.

We find that the probability to continue the studies decreased by 2% in 2007, even if it is barely significant (at 10%). Findings at national level confirm that post-reform graduates are more likely to enrol in long single cycle degree programs (+ 4.3%), and less in bachelor degrees (-4%).

Table B14: High School graduates

	(1) Yes_univ	(2) Yes_bac	(3) Yes_cu	(4) Yes_other
Post	-0.0214*	-0.0405***	0.0430***	-0.00119
	(-1.82)	(-2.69)	(3.42)	(-0.11)
Female	0.00924	0.00242	0.0163	-0.0201**
	(0.81)	(0.15)	(1.21)	(-1.96)
Age at high school diploma	-0.0811***	0.00155	-0.0203	0.0187
	(-5.63)	(0.08)	(-1.62)	(1.08)
NorthEast	-0.0264	0.00892	-0.0144	0.000246
	(-1.55)	(0.41)	(-0.83)	(0.02)
Centre	-0.0268	-0.0396	0.00735	0.0272*
	(-1.44)	(-1.58)	(0.34)	(1.71)
South	-0.0520***	-0.0377*	0.0185	0.0139
	(-3.21)	(-1.70)	(0.99)	(0.91)
Islands	-0.0669***	-0.0730***	0.0139	0.0535***
	(-3.02)	(-2.73)	(0.68)	(2.58)
Public school	0.0470*	0.0450	0.00670	-0.0497**
	(1.78)	(1.43)	(0.28)	(-2.16)
HighS_acad	0.556***	0.106***	0.144***	-0.247***
	(37.78)	(4.63)	(9.58)	(-12.22)
HighS_tech	0.225***	0.0936***	0.0206*	-0.114***
	(13.99)	(4.30)	(1.83)	(-5.35)
HighS_others	0.368***	0.149***	0.0110	-0.159***
	(24.85)	(6.97)	(0.85)	(-7.94)
LowS_grade2	0.0734***	0.0761***	-0.0109	-0.0661***
	(3.74)	(2.87)	(-0.82)	(-2.59)
LowS_grade3	0.116***	0.0874***	0.00675	-0.0929***
	(5.14)	(3.11)	(0.40)	(-3.58)
LowS_grade4	0.112***	0.0715**	0.0204	-0.0930***
	(5.08)	(2.49)	(1.08)	(-3.76)
HighS_grade2	0.107***	0.00959	0.0661***	-0.0759***
	(5.74)	(0.42)	(3.98)	(-4.09)
HighS_grade3	0.192***	0.0357	0.0566***	-0.0875***
	(11.71)	(1.47)	(3.22)	(-4.57)
HighS_grade4	0.267***	0.0583***	0.0902***	-0.147***
	(17.20)	(2.61)	(5.73)	(-8.37)
Constant	1.677***	0.479	0.339	0.183
	(5.74)	(1.21)	(1.35)	(0.52)
Observations	34264	21681	21681	21681

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C1 Additional empirical analysis on our third remark

Table C1: Summary statistics of the outcomes

	Untreated Cohorts		**Treated Cohorts**	
Outcomes:	Mean	Standard dev.	Mean	Standard dev.
Log earnings	7.095	0.896	7.132	0.869
Log income	7.104	0.897	7.137	0.870
Earnings Occupational Premium	−0.052	0.182	0.034	0.178
Income Occupational Premium	−0.055	0.182	0.040	0.178

Table C2: Distribution of time gaps between educational experience and earnings observation

		Time from 19:		
Untreated cohorts:	to pre-graduation	when first observed	to graduation	to graduation (education changers)
10th quantile	1	0	10	10
25th quantile	10	0	11	11
Median	11	3	11	11
75th quantile	11	10	11	11
90th quantile	11	10	11	11
count	992118	679501	230835	230835
Treated cohorts:				
10th quantile	3	0	9	10
25th quantile	9	2	10	11
Median	11	4	11	11
75th quantile	8	10	11	11
90th quantile	10	10	11	11
count	1012192	589077	220654	220654

Table C3: Age distribution of the earnings observation

Untreated cohorts	Full sample	Individuals with high school degree	Individuals observed before graduation	Individuals observed after graduation
10th quantile	21	21	20	24
25th quantile	24	23	22	26
Mean	25.901	25.393	24.902	27.269
Median	26	26	25	28
75th quantile	29	28	28	29
90th quantile	30	30	30	30
count	3405373	2074139	323918	1007316
Treated cohorts				
10th quantile	21	21	20	24
25th quantile	23	23	22	26
Mean	25.707	25.166	24.147	27.307
Median	26	25	24	28
75th quantile	28	28	27	29
90th quantile	29	29	28	30
count	3092575	1937069	253193	902313

Table C4: Returns to a university degree: subsample of education changers

	Baseline		**Differential impact**	
Outcomes:	OLS	FE	OLS	FE
Earnings	0.288^{***}	0.128^{***}	0.000	0.031^{***}
	0.010	0.008	0.014	0.011
Income	0.298^{***}	0.127^{***}	-0.002	0.034^{***}
	0.010	0.008	0.014	0.011
Earnings Occupational Premium	0.037^{***}	-	-0.001	-
	0.003	-	0.004	-
Income Occupational Premium	0.040^{***}	-	-0.001	-
	0.003	-	0.004	-

Notes: The model includes age distance from 19, year and month dummies, a gender indicator, the interaction of the gender indicator with the age distance dummies. Standard errors are clustered at individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C5: Returns to a university degree: all sample; separated regressions by treatment

	Untreated Cohorts			**Treated Cohorts**		
Outcomes:	OLS	FE	FE_{hg}	OLS	FE	FE_{hg}
Earnings	0.029^{***}	0.297^{***}	0.420^{***}	0.079^{***}	0.403^{***}	0.488^{***}
	0.005	0.009	0.010	0.005	0.010	0.011
Income	0.041^{***}	0.300^{***}	0.431^{***}	0.087^{***}	0.406^{***}	0.493^{***}
	0.005	0.009	0.010	0.005	0.010	0.011
Earnings Occupational Premium	-0.010^{***}	-	0.036^{***}	-0.006^{***}	-	0.038^{***}
	0.002	-	0.002	0.002	-	0.002
Income Occupational Premium	-0.005^{***}	-	0.039^{***}	-0.002	-	0.041^{***}
	0.002	-	0.002	0.002	-	0.002

Notes: The model includes age distance from 19, year and month dummies, a gender indicator, the interaction of the gender indicator with the age distance dummies. Standard errors are clustered at individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C6: Returns to a university degree: subsample of education changers; separated regressions by treatment

	Untreated Cohorts		**Treated Cohorts**	
Outcomes:	OLS	FE	OLS	FE
Earnings	0.290^{***}	0.126^{***}	0.286^{***}	0.161^{***}
	0.011	0.008	0.011	0.008
Income	0.300^{***}	0.125^{***}	0.293^{***}	0.164^{***}
	0.011	0.008	0.011	0.008
Earnings Occupational Premium	0.037^{***}	-	0.034^{***}	-
	0.003	-	0.003	-
Income Occupational Premium	0.041^{***}	-	0.038^{***}	-
	0.003	-	0.003	-

Notes: The model includes age distance from 19, year and month dummies, a gender indicator, the interaction of the gender indicator with the age distance dummies. Standard errors are clustered at individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C7: Returns to a university degree: men

	Baseline			Differential impact		
Outcomes:	OLS	FE	FE_{hg}	OLS	FE	FE_{hg}
Earnings	0.030^{***}	0.299^{***}	0.422^{***}	-0.002	0.110^{***}	0.024^{***}
	0.005	0.009	0.008	0.005	0.018	0.009
Income	0.041^{***}	0.301^{***}	0.432^{***}	-0.007^{***}	0.111^{***}	0.019^{***}
	0.005	0.009	0.008	0.009	0.018	0.009
Earnings Occupational Premium	-0.010^{***}	-	0.034^{***}	-0.005^{*}	-	-0.002
	0.002	-	0.002	0.003	-	0.003
Income Occupational Premium	-0.006^{***}	-	0.038^{***}	-0.006^{***}	-	-0.003
	0.002	-	0.013	0.003	-	0.003

Notes: The model includes age distance from 19, year and month dummies, a gender indicator, the interaction of the gender indicator with the age distance dummies. Standard errors are clustered at individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C8: Returns to a university degree: women

	Baseline			Differential impact		
Outcomes:	OLS	FE	FE_{hg}	OLS	FE	FE_{hg}
Earnings	0.030^{***}	0.299^{***}	0.422^{***}	0.091^{***}	0.097^{***}	0.092^{***}
	0.005	0.009	0.008	0.009	0.015	0.009
Income	0.041^{***}	0.301^{***}	0.432^{***}	0.088^{***}	0.096^{***}	0.090^{***}
	0.005	0.009	0.008	0.009	0.015	0.009
Earnings Occupational Premium	-0.010^{***}	-	0.034^{***}	0.013^{***}	-	0.013^{***}
	0.002	-	0.002	0.003	-	0.003
Income Occupational Premium	-0.006^{***}	-	0.038^{***}	0.012^{***}	-	0.013^{***}
	0.002	-	0.013	0.003	-	0.003

Notes: The model includes age distance from 19, year and month dummies, a gender indicator, the interaction of the gender indicator with the age distance dummies. Standard errors are clustered at individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C9: Returns to a university degree: private sector

	Baseline			Differential impact		
Outcomes:	OLS	FE	FE_{hg}	OLS	FE	FE_{hg}
Earnings	0.137^{***}	0.149^{***}	0.332^{***}	0.043^{***}	0.080^{***}	0.048^{***}
	0.002	0.002	0.003	0.002	0.004	0.003
Daily Earnings	0.150^{***}	0.045^{***}	0.195^{***}	0.027^{***}	0.036^{***}	0.028^{***}
	0.001	0.001	0.002	0.002	0.002	0.002
Earnings Occupational Premium	-0.007^{***}	-	0.033^{***}	0.008^{*}	-	0.009^{***}
	0.001	-	0.001	0.001	-	0.001
Daily Earnings Occupational Premium	-0.004^{***}	-	0.035^{***}	0.008^{***}	-	0.009^{***}
	0.001	-	0.001	0.001	-	0.001

Notes: The model includes age distance from 19, year and month dummies, a gender indicator, the interaction of the gender indicator with the age distance dummies. Standard errors are clustered at individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C10: Returns to a university degree: men private sector

	Baseline			Differential impact		
Outcomes:	OLS	FE	FE_{hg}	OLS	FE	FE_{hg}
Earnings	0.136^{***}	0.149^{***}	0.331^{***}	0.015^{***}	0.068^{***}	0.025^{***}
	0.002	0.002	0.003	0.003	0.005	0.003
Daily Earnings	0.150^{***}	0.045^{***}	0.194^{***}	0.002	0.040^{***}	0.004^{***}
	0.001	0.001	0.002	0.002	0.003	0.002
Earnings Occupational Premium	-0.008^{***}	-	0.033^{***}	-0.002^{*}	-	0.000
	0.001	-	0.001	0.001	-	0.001
Daily Earnings Occupational Premium	-0.004^{***}	-	0.035^{***}	-0.003^{**}	-	-0.001
	0.006	-	0.001	0.001	-	0.001

Notes: The model includes age distance from 19, year and month dummies, a gender indicator, the interaction of the gender indicator with the age distance dummies. Standard errors are clustered at individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C11: Returns to a university degree: women private sector

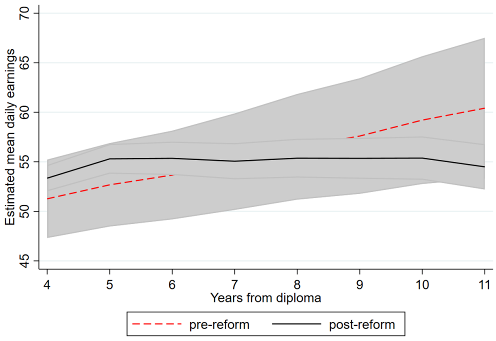
	Baseline			Differential impact		
Outcomes:	OLS	FE	FE_{hg}	OLS	FE	FE_{hg}
Earnings	0.136^{***}	0.149^{***}	0.331^{***}	0.068^{***}	0.091^{***}	0.069^{***}
	0.002	0.002	0.003	0.002	0.004	0.003
Daily Earnings	0.150^{***}	0.045^{***}	0.194^{***}	0.048^{***}	0.040^{***}	0.048^{***}
	0.001	0.001	0.002	0.002	0.003	0.002
Earnings Occupational Premium	-0.008^{***}	-	0.033^{***}	0.017^{***}	-	0.017^{***}
	0.001	-	0.001	0.001	-	0.001
Daily Earnings Occupational Premium	-0.004^{***}	-	0.035^{***}	0.016^{***}	-	0.017^{***}
	0.006	-	0.001	0.001	-	0.001

Notes: The model includes age distance from 19, year and month dummies, a gender indicator, the interaction of the gender indicator with the age distance dummies. Standard errors are clustered at individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

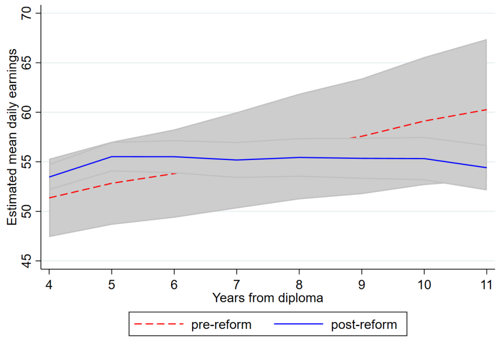
C1.1 Estimating earnings profile using University-INPS data

We merge our University administrative data with the INPS archives. We use this selected sample to provide further evidence on the relevance of the 2007 reform in increasing earnings. These figures bear out three facts. First, our second remark predicts that the pupils switched from the two-tier to the one-tier track. The impact of the reform shows up when we use data covering all sectors and self-employment. Graduation in the one-tier track raises the likelihood of being (self)employed in sectors other than the private one. Second, cohorts differences are larger when we partly control the unemployment periods by assigning a zero value to earnings. Third, differences enlarge with students' age as they graduate. This last fact supports our third remark. We cannot disentangle these three effects with these data. Nevertheless, we can argue that the 2007 reform raised workers' earnings.

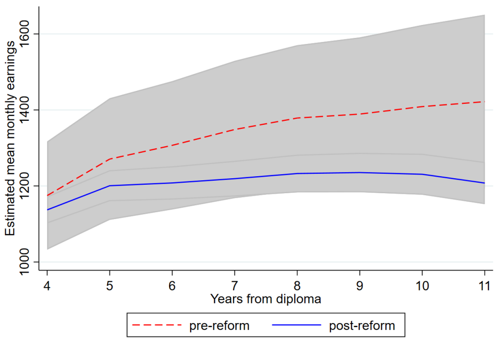
Figure C1: Earnings profiles: private sector



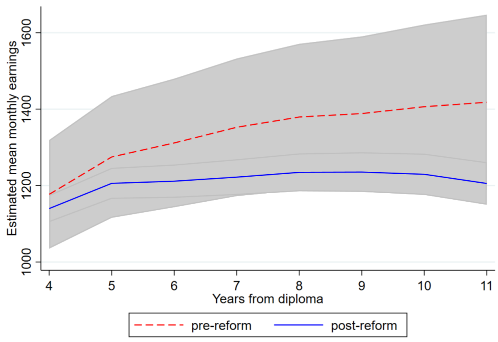
(a) Daily earnings: unbalanced panel



(b) Daily earnings: balanced panel

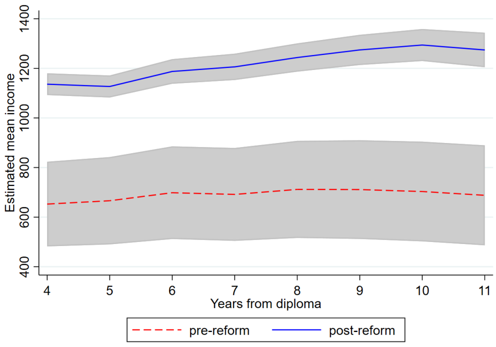


(c) Monthly earnings: unbalanced panel

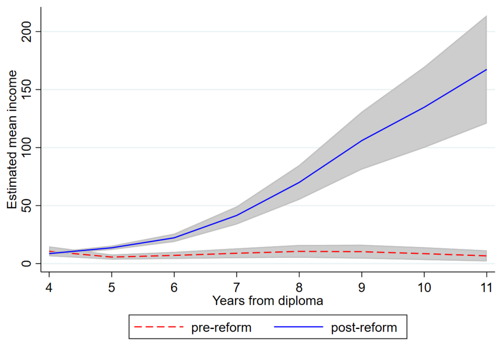


(d) Monthly earnings: balanced panel

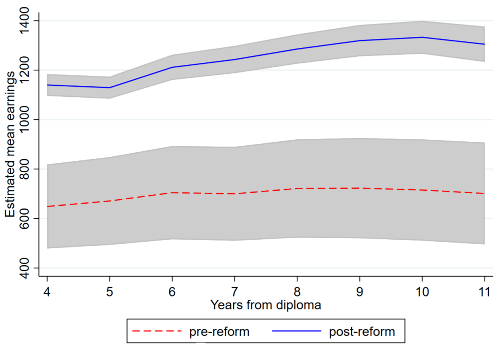
Figure C2: Earnings profiles: all sectors and self-employment (source: Social Security Contributions)



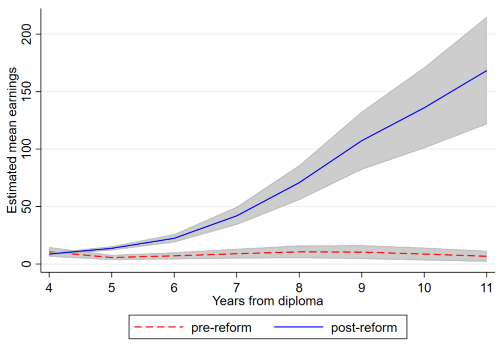
(a) Daily earnings: unbalanced panel



(b) Daily earnings: balanced panel

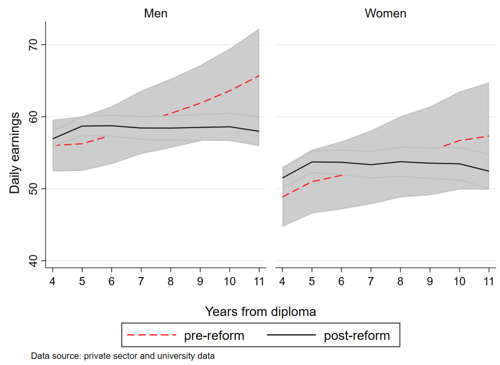


(c) Monthly earnings: unbalanced panel

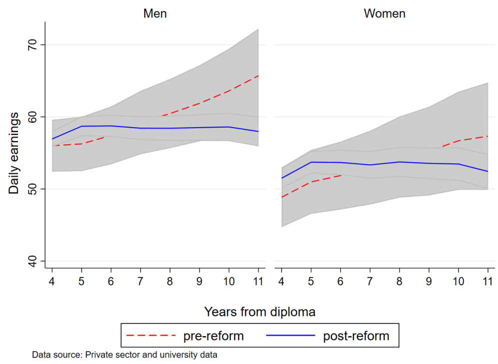


(d) Monthly earnings: balanced panel

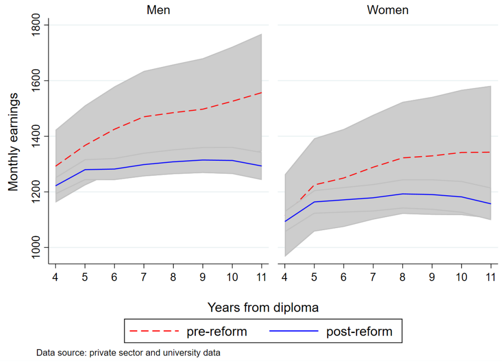
Figure C3: Earnings profiles: private sector



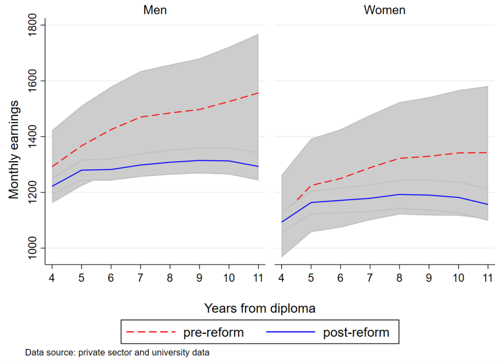
(a) Daily earnings: unbalanced panel



(b) Daily earnings: balanced panel

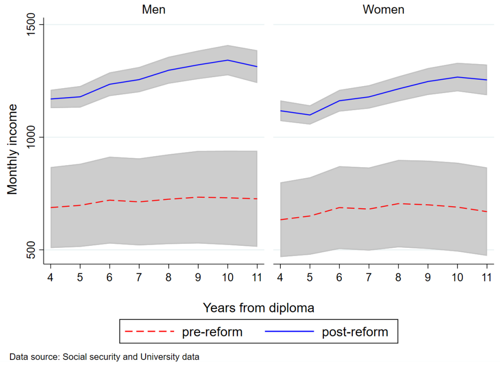


(c) Monthly earnings: unbalanced panel

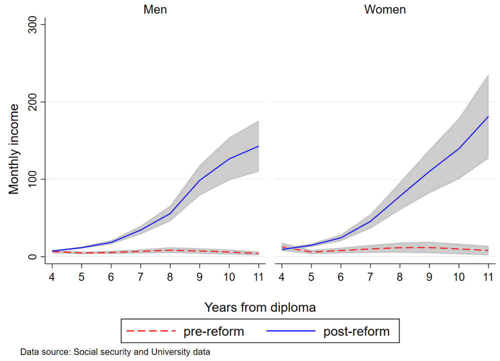


(d) Monthly earnings: balanced panel

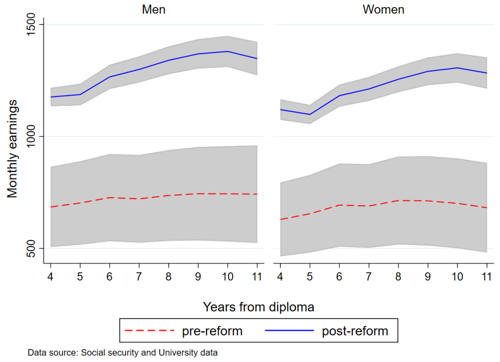
Figure C4: Earnings profiles: all sectors and self-employment (source: Social Security Contributions



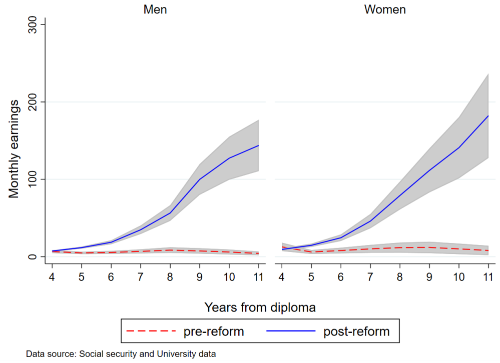
(a) Daily earnings: unbalanced panel



(b) Daily earnings: balanced panel



(c) Monthly earnings: unbalanced panel



(d) Monthly earnings: balanced panel

D1 Additional Materials

D1.1 Graph on the optimal choice under the magnifying glass

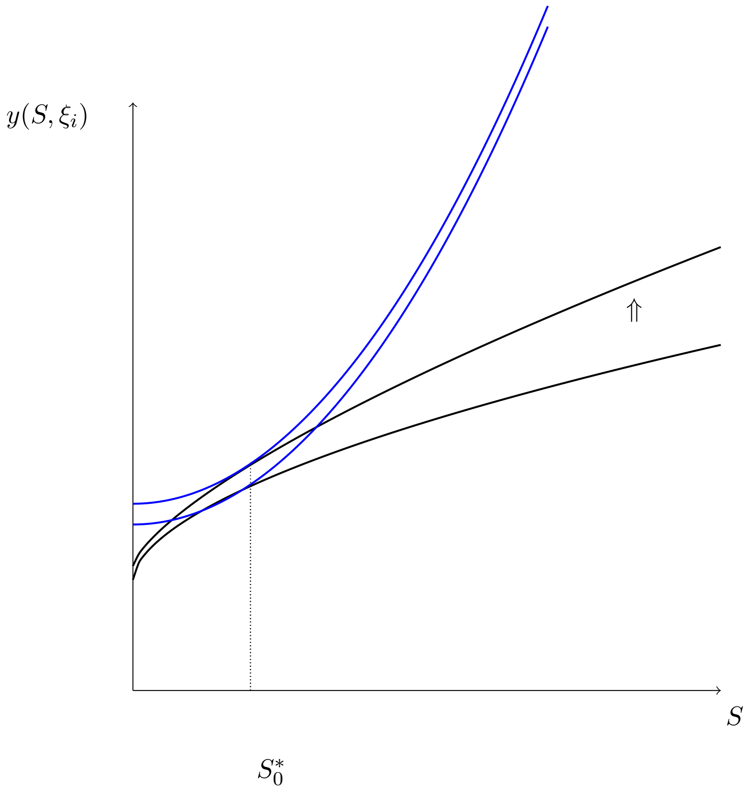


Figure D1: Optimal choice of years of schooling