INPS
Istituto Nazionale Previdenza Sociale

Ottobre 2022 – numero 56

WorkINPS *Papers*

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

Daniela Sonedda

Paolo Ghinetti

Giorgia Casalone

In copertina: uno storico "Punto cliente" a Tuscania
INPS, Direzione generale, Archivio storico

I WORKINPS PAPER

Le basi dati amministrative dell'*INPS* rappresentano una fonte statistica unica per studiare scientificamente temi cruciali per l'economia italiana, la società e la politica economica: non solo il mercato del lavoro e i sistemi di protezione sociale, ma anche i nodi strutturali che impediscono all'Italia di crescere in modo adeguato. All'interno dell'Istituto, questi temi vengono studiati sia dai funzionari impiegati in attività di ricerca, sia dai *VisitInps Scholars*, ricercatori italiani e stranieri selezionati in base al loro curriculum vitae e al progetto di ricerca presentato.

I **WORKINPS** hanno lo scopo di diffondere i risultati delle ricerche svolte all'interno dell'Istituto a un più ampio numero possibile di ricercatori, studenti e policy markers.

Questi saggi di ricerca rappresentano un prodotto di avanzamento intermedio rispetto alla pubblicazione scientifica finale, un processo che nelle scienze sociali può chiedere anche diversi anni. Il processo di pubblicazione scientifica finale sarà gestito dai singoli autori.

Maurizio Franzini

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

Daniela Sonedda

(Università degli Studi dell'Insubria)

Paolo Ghinetti

(Universit`a degli Studi del Piemonte Orientale)

Giorgia Casalone

(Centre For North South Economic Research (CRENoS), Cagliari, Italy. Dondena Centre for Research on Social Dynamics and Public Policy, Milan, Italy)

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

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

[a] *Università degli Studi dell'Insubria.*
[b] *Università degli Studi del Piemonte Orientale.*
[c] *Centre For North South Economic Research (CRENoS), Cagliari, Italy.*
[d] *Dondena Centre for Research on Social Dynamics and Public Policy, Milan, Italy.*

September 2022

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. Women benefited more than men from the reform. The reform led them to choose a different higher education path and opened up the doors to occupations and earnings they would not have earned absent it.

Keywords: Education Returns; Auditing; High education tracks; Earnings; Gender gap.

JEL Codes: I20; J30; M42.

*Contact information: Daniela Sonedda (corresponding author), daniela.sonedda@uninsubria.it; Paolo Ghinetti, paolo.ghinetti@uniupo.it; Giorgia Casalone, giorgia.casalone@uniupo.it.
We thank Fabio Berton, Massimiliano Bratti, Edoardo Di Porto, Paolo Naticchioni, Lia Pacelli, Matteo Paradisi, Paolo Sestito, Silvia Vannutelli and audiences at AIEL, IAAEU Workshop on Labour Economics (Trier), Turin (EST seminar) and IWAEE workshop 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. The realization of the present article was possible thanks to the sponsorship and the donations in favour of the "VisitINPS Scholars" program. The findings and conclusions expressed are solely those of the author and do not represent the views of INPS.

Il valore dell'auditing del voto di maturità sul percorso di istruzione terziaria e sugli esiti nel mercato del lavoro*

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

[a] *Università degli Studi dell'Insubria.*
[b] *Università degli Studi del Piemonte Orientale.*
[c] *Centre For North South Economic Research (CRENoS), Cagliari, Italy.*
[d] *Dondena Centre for Research on Social Dynamics and Public Policy, Milan, Italy.*

Settembre 2022

Abstract

In questo saggio, analizziamo il valore dell'audit (valutazione esterna) dei voti finali delle scuole superiori sui percorsi di istruzione terziaria e gli esiti nel mercato del lavoro. Utilizziamo la riforma che nel 2007 ha introdotto in Italia la presenza di esaminatori esterni nelle commissioni dell'esame di maturità. Confrontiamo le coorti trattate (post-riforma) e non trattate (pre-riforma) utilizzando un modello ad effetti fissi del tipo *two-way* per mostrare che la riforma ha aumentato i guadagni dei diplomati delle scuole superiori. Adottiamo una strategia simile, combinato con uno a variabili strumentali che utilizza l'approccio degli *special regressors* per mostrare che la riforma ha innalzato gli anni di scolarizzazione dei diplomati superiori. Estendiamo la metodologia ad effetti fissi combinati di (Altonji and Zhong 2021) per mostrare che, tra le coorti trattate, il rendimento di una laurea è di circa sei punti percentuali più alto di quello dei non trattati. Le donne hanno beneficiato della riforma più degli uomini. La riforma le ha portate a scegliere un diverso percorso di istruzione superiore e ha aperto le porte a occupazioni e guadagni migliori, che non avrebbero avuto altrimenti.

Parole chiave: Rendimento dell'istruzione; Auditing; Percorsi di studio di istruzione terziaria; Retribuzioni; Differenziali di genere.

Codici JEL: I20; J30; M42.

*Contatti: Daniela Sonedda (autore di riferimento), daniela.sonedda@uninsubria.it; Paolo Ghinetti, paolo.ghinetti@uniupo.it; Giorgia Casalone, giorgia.casalone@uniupo.it.
Ringraziamo Fabio Berton, Massimiliano Bratti, Edoardo Di Porto, Paolo Naticchioni, Lia Pacelli, Matteo Paradisi, Paolo Sestito, Silvia Vannutelli e i partecipanti alla conferenza AIEL 2019, IAAEU Workshop on Labour Economics (Trieste), Torino (EST seminar) e IWAEE Workshop per i suggerimenti e gli utili commenti ricevuti. Questo studio utilizza dati anonimizzati forniti dall'Istituto Nazionale di Previdenza Sociale (INPS), il cui accesso è stato possibile grazie al programma VISITINPS Scholars. Per questo siamo grati a Daniele Checchi, Monia Monachini, Alfredo Arpinelli e Barbara Ceremigna per il loro aiuto nel rendere possibile questa ricerca, e a Massimo Ascione e Elio Bellucci per il loro supporto coi dati. Ringraziamo anche il Rettore e gli uffici dell'Amministrazione Centrale dell'università che ci ha fornito i suoi dati. La realizzazione del presente articolo 'e stata possibile grazie alle sponsorizzazioni e le erogazioni liberali a favore del programma VisitINPS Scholars. Le opinioni espresse in questo articolo appartengono esclusivamente agli autori e non riflettono necessariamente la posizione dell'INPS.

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 market's 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

[1]For instance record-breaking increases in grades were awarded in the UK in the summer of 2021. Nearly 45% of A-level were recorded in England, a 6.2 percentage point rise on 2020's results and a 19.1 increase compared with the pre-pandemic exam outcomes of 2019.

and Pistaferri 2010).

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, understands 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 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. We assume that these cohorts are statistical identical but the treatment. In this part of the analysis, we retain those whose maximum level of completed education is high school. To do that, we link the INPS archives with data from the Mandatory Communications, maintained by the Ministry of Labour and Social Policies. This hypothesis rules out the existence of heterogeneous treatment effects across groups.

For us, the cohorts identify the groups. Individuals are heterogeneous within the cohorts, but heterogeneity is statistically identical across cohorts. Under this assumption, we do not control for individual fixed effects. This assumption factors in treated and untreated cohorts sharing abilities and preference distributions. However, we allow for gender differences. For each age, we consider a group/time average effect. As we have more than one birth cohort for each treatment group and monthly outcomes, we can separate birth cohort from time effects.

We rule out negative weights in this context due to heterogeneous treatment effects across groups. The group weights come as a proxy of the sample share of each treatment group at each age and cannot be negative (de Chaisemartin and D'Haultfoeuille 2020, Goodman-Bacon 2018, Callaway and Sant'Anna 2020, Sun and Abraham 2020). Yet, when (month/age) workers start their working career and with which job is part of the treatment effects. Hence, we allow for heterogeneous treatment effects over time. These treatment effects vary with job experience, and from this point of view, the treatment is staggered even though the reform is pre-determined to the labour market entry.

We then trace these two groups' earnings (income) profiles from 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. We estimate the fitted values at each age. Yet, these fitted values are not the treatment effects. We calculate these effects as the difference across the two groups of the difference between the fitted value at each age compared with the baseline at age 19. This way of proceeding shares some features of the methodology proposed by Borusyak, Jaravel and Spiess (2022). They propose an estimator that is robust to heterogeneous treatment effects over time as the difference between each treated unit outcome and the fitted values of the counterfactual outcome of the untreated group.

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 than the latter one. For instance, six years is the minimum length of Medicine. These courses are always longer than whichever degree in the two-tier track, as the two-tier track is five years long.

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. We checked that 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 and it 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.[2] We use the special regressor method Dong and Lewbel (2015) to estimate the impact on the enrolment probability of being enrolled at the one-tier track. We measure these effects on the outcome after one 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.[3] 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 (i.e. high-school achievers) 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.

[2] With the exceptions of Medicine degree that requires six years.

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

Second, high school degree achievers would have shared a common experience profile without the reform.

We extend this setting by comparing cohorts treated and untreated by the 2007 reform. These two groups overlap in terms of ability and preferences but differ because of the treatment. They would have shared the same labour market experience without the reform. The reform didn't affect the probability to observe a university graduate worker at any age between 19 and 30. Yet, it impacted the choices people made on the academic track and, through this channel, it changed the returns to a university degree.

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 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. This policy benefited those who did not enrol at university, providing a higher reward for their high-school grades. Moreover, it changed the pupils' choices. Through this different educational path, the labour market perspectives of treated workers improved. 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. With the attendance of university courses that they would not have attended absent the reform, women reap the returns of this choice in the labour market. However, as we show in our analysis of the first remark, the gender income gap in levels is far from being filled.

Finally, we follow Hendren and Sprung-Keyser (2020) to provide a welfare analysis of this reform. After seven years from high school completion, we show that it paid for itself. 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. Section 7 reports the welfare analysis. We conclude in Section 8.

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.[4] 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, issued 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,[5] the department office for education

[4]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.

[5]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.

chooses the subjects of the external examiners. Three written assessments and an interview make the exam, and the board votes the grade assessment in a range from 60 to 100. When there is no agreement, the majority of the votes sets the final grade. 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, understands 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 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. When this issuance comes from internal teachers, there is no guarantee of an unbiased judgement. 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. The issuance of a final high-school grade provides for accountability on the worker to be. External examiners, the auditors, increase the labour market's confidence in the established grades. The labour market is 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 cannot monitor this evaluation since they do not participate in the high-schooling system. 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. 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 may lead 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 (equal or above high school). 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 (in particular firms and students) 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.[6] We further impose the following

[6] 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.

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.

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

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 adapt the presentation developed in Card (1999) in which an individual chooses S to maximise a utility function $U(S, y(S, \xi_S))$:

$$U(S, y(S, \xi_S)) = log(y(S, \xi_S)) - h(S) \quad (3)$$

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

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

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' (and students') updated beliefs on worker productivity. These 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.

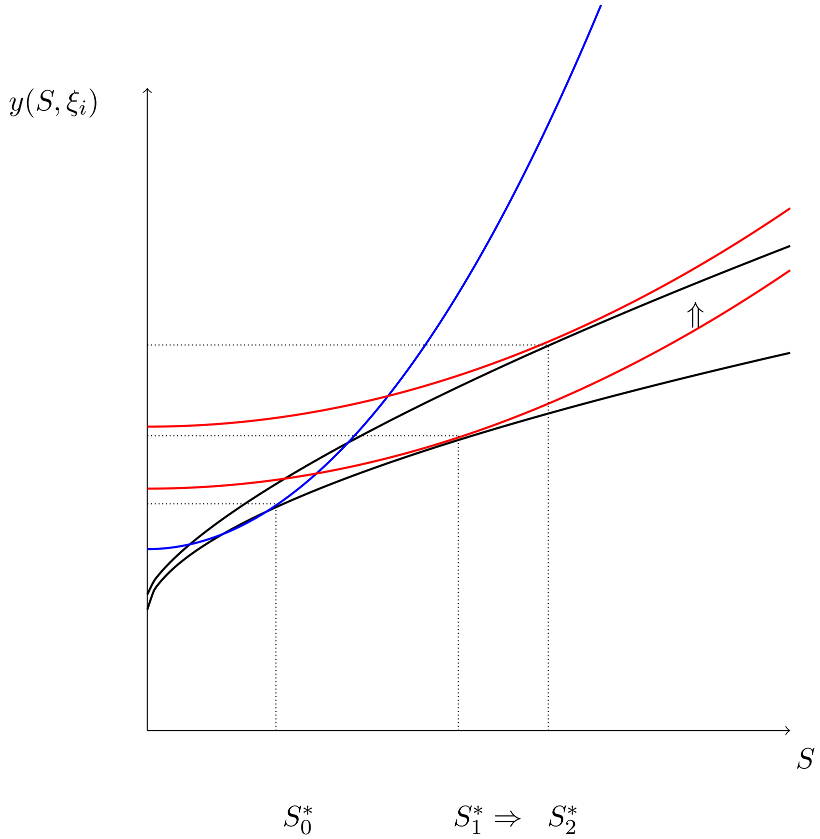


Figure 1: Optimal choice of years of schooling

The big picture that emerges is that there are three consequences of this exogenous shift. Assume that lower ability individuals will always choose low education S_0^*. They may be those students who would stop at high school both with and without the reform. 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. The second is an archive of careers of university students enrolled in a medium-size university located in the North of Italy. This is used to test the second prediction. About INPS data, we use two representative samples to address our research questions.

The first is pulled from the archive on social security contributions and covers all sectors. Data are drawn from four birth dates for each month and year. With social security records we construct two different measures of earnings. The first includes all earnings from dependent work, while the second is the sum of all sources of income from private, public and self-employment. To distinguish the two, we name the first earnings and the second monthly income. We compute both at the monthly level. Like a bank account, social security accounts record events (e.g. being employed for a certain period) that are relevant for social security purposes (e.g. movement in the account when the employer pays pension contributions). We observe the total earnings/income received from each job during its duration spell. We know when each spell started and ended, within each calendar year.[7]. We use this information to compute the average monthly earnings/income from each job, and attribute it to the months when that job was active. This is a proxy of actual monthly job payments since they do not vary across the months of duration of each spell (within the year). Workers may have overlapping spells and may be observed in more than one job in a given month. In that case, monthly earnings and income are the sum of payments received from various sources in that month.

The second is an archive on working careers for the whole population of private sector workers. Here, our two outcomes are monthly and daily earnings. We cannot conduct our analyses on wages because INPS data do not record the hours of work. The advantage here is that we can compute both monthly and daily earnings (based on the number of days worked each month), and that both are actual and not reconstructed measures. The limitation is that, 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 university end up in professional occupations that cannot be found in a private sector archive as a doctor for the National Health System or a lawyer.

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. Enlarging the sample to other cohorts would require to deal with two opposite reforms of the compulsory schooling age. Compulsory schooling age was first

[7]If a worker is employed from day 1 to 365 in a given job, we know the total pay received that year. If he/she was employed only for the first 35 days, we had the total payment relative to that spell, and so on.

raised to 15 in 1999 and then was lowered to 14 in 2003.[8] These changes could have impacted the age profiles of the earnings of the involved cohorts.[9]

The information on the education level is missing in the two INPS archives. We merge them with the archive of mandatory communications (*Comunicazioni OBbligatorie*, COB hereafter) on job creation or destruction to retrieve it. This archive records starting and ending date of any job (as well as any job transformation, e.g. from temporary to permanent) as well as some information on the worker employed in that job.[10] We do not observe the high school graduation year. Yet, we consider it a minor issue if the retention rate in high school is the same across treatment groups.[11] For each matched individual, we know the highest education degree possessed at the time of the communication. If an individual appears in more than one mandatory communication in the observation period, we can observe the changes over time in the education level (e.g. from high school to university degrees). In general, we know the education degree but not when this degree was achieved. We assume a hypothetical age of high school graduation of 19 years old for all individuals, 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. Since we define the treatment status by the birth year only (i.e. being 19 years old from 2007 onwards or not), we use two treated and two untreated cohorts of high-school graduates.

Our time window covers from 0 to 11 years after age 19 for each birth cohort. Considering all cohorts, our data cover the 2005-2019 period. We have 6,497,948 monthly/year observations and 109,126 individuals for the sample that covers all sectors.[12] For each individual, we construct a monthly panel of his/her working history to document the exogenous shift in expected earnings/income after the reform.[13] The unbalanced nature of the panel reflects the different timing of individuals labour market entry.

Our second data source comes from administrative records of a public university based in

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

[9]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.

[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, for reasons explained below) provide all the relevant information on the education level of any of them.

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

[12]When we employ the data for all workers in the private sector, we have 38,829,228 observations.

[13]For empirical convenience, we take advantage of our large sample and operate a trimming and exclude observations - not individuals - in the upper and bottom 5% of earnings/income distributions. This allows us to discard extreme values of little interest - in particular disproportionally low monthly earnings/income (close to zero or of few dozens of euros) or high ones (hundreds of thousands euros, perhaps by professional football players) - and to focus on the part of the distribution more relevant for policy analysis.

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 9,150 individuals make our working sample. They 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.

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. To this purpose we retain from the matched INPS-COB dataset only those whose higher education degree is high school. Using the Social Security Contributions archive, we end up with 4,015,519 observations.[14]

In our application, we have two groups: untreated and treated birth cohorts. We rule out heterogeneous treatment effects over groups by assuming that these cohorts share abilities, preferences and common labour market experience at age 19. This assumption is trivially satisfied as we select workers with a high school degree. It is pretty unlikely that they could have worked while in high school. We allow for heterogeneous treatment effects over time. We

[14]With the INPS archive on workers in the private sector, after applying the same selection rules we have 25,438,187 observations. This part of the analysis is reported in Appendix A1.1.

can substantiate this claim by bringing in the economic modelling. A well-recognised fact is that earnings vary with experience. In our context, this fact might have the following implication. With employers more able to trust in the high value and rigorous assessment of the high school degree, workers' job experience could differ from what would have been without it. For instance, as high-school grades provide a more accurate description of the workers' skills, more stable jobs could have emerged.[15] If this argument is valid, treatment effects vary over time, as summarised in Table A1.

Table 1: Two groups, Eleven-period

$E[Y_{it}]$	(i=T)	(i=U)
t=0	α_T	α_U
t=1	$\alpha_T + \beta_1 + \tau_1$	$\alpha_U + \beta_1$
t=2	$\alpha_T + \beta_2 + \tau_2$	$\alpha_U + \beta_2$
t=3	$\alpha_T + \beta_3 + \tau_3$	$\alpha_U + \beta_3$
t=4	$\alpha_T + \beta_4 + \tau_4$	$\alpha_U + \beta_4$
t=5	$\alpha_T + \beta_5 + \tau_5$	$\alpha_U + \beta_5$
t=6	$\alpha_T + \beta_6 + \tau_6$	$\alpha_U + \beta_6$
t=7	$\alpha_T + \beta_7 + \tau_7$	$\alpha_U + \beta_7$
t=8	$\alpha_T + \beta_8 + \tau_8$	$\alpha_U + \beta_8$
t=9	$\alpha_T + \beta_9 + \tau_9$	$\alpha_U + \beta_9$
t=10	$\alpha_T + \beta_{10} + \tau_{10}$	$\alpha_U + \beta_1 0$
t=11	$\alpha_T + \beta_{11} + \tau_{11}$	$\alpha_U + \beta_1 1$

We assume a location effect specific to each group at age 19 (parameters α_T and α_U), as soon as pupils earn their degree. Under such a case, it is too early for a treatment effect that varies with experience to materialise. We then consider the potential outcome at this age as the one in the pre-treatment period. The parameter β captures the common trend. The treated group would have had the same age profile as the untreated one without the reform. Borusyak et al. (2022) thoroughly discuss the short-run bias of the TWFE OLS static model. In Online Appendix A1, we follow them to show how it applies to our case. We report there its estimation results while here, we focus on our primary model specification. In Table A1, we abstract from gender-specific age profiles, but in Equation 5, we allow them. We run Equation 5 for each treatment group: the untreated cohorts gather workers born in 1986 and 1987, and treated cohorts are made of those born in 1988 and 1989.

$$y_{ijt} = \sum_{a=0}^{11} \alpha_{1aj} AGE_{it}^{a} + \sum_{a=0}^{11} \alpha_{2aj} AGE_{it}^{a} * F_i + \sum_{r=2005}^{2019} \theta_{jr} + \sum_{m=2}^{12} \mu_{jm} + \gamma_j F_i + u_{ijt} \quad (5)$$

[15] We plan to provide more evidence on this issue as soon as we access the data.

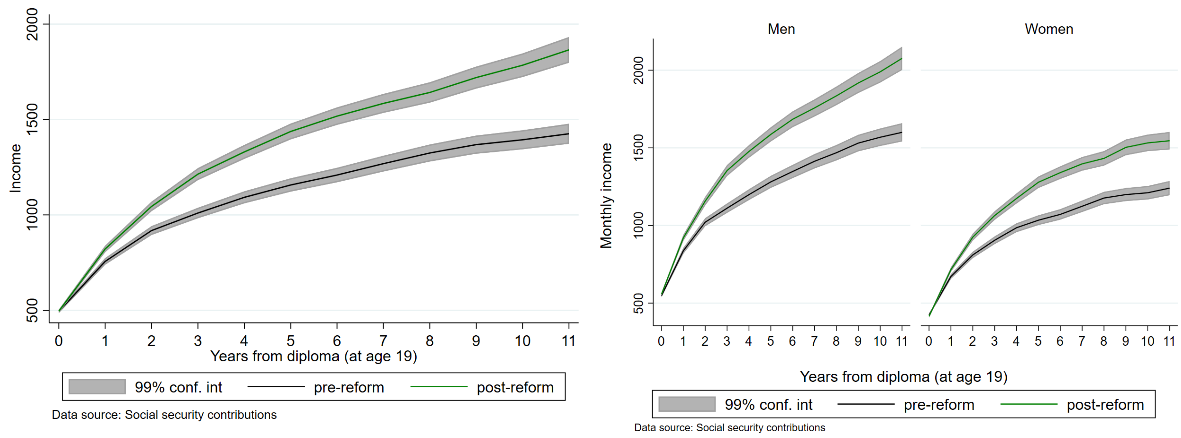
The outcome y (log earnings or income) of individual i at time t (which we use as a compact notation of year r and month m) of relative age (years from 19) $age_{it} = a$ who belongs to the treatment group j is regressed on year θ_r and month μ_m dummies, the gender-specific age profile summarized by a set of indicators AGE^a_{it} that are 1 when $age_{it} = a$; female F_i and treatment $T_i = j, j = 0, 1$ dummies.

We estimate our treatment effects τ, age by age. This procedure amounts to calculating for each age a from 1 to 11:

$$(E[Y_{Ta}] - E[Y_{T0}]) - (E[Y_{Ua}] - E[Y_{U0}]) = \tau_a \tag{6}$$

We proceed as follows. First, we estimate the fitted values of our regression model for each treatment group. We work them out from exponentiated predictions to report them in Figure 2 in levels. Confidence intervals are obtained from the *nlcom* command in Stata. Our outcome is the monthly income as measured by the INPS social security archive.

Figure 2: Age profiles of income by treatment group



(a) Income: all workers

(b) Income: by gender

In panel (a), for each treatment group, these fitted values are averaged across gender for each age; in panel (b), we do not average across gender. We include in the model treatment group fixed effects rather than workers' fixed effects. This choice follows the assumption of birth cohorts sharing ability and preference distributions. Yet, combined with group/age effects, it eliminates the risk of negative weights suggested by de Chaisemartin and D'Haultfoeuille (2018), Goodman-Bacon (2018), Callaway and Sant'Anna (2020), Sun and Abraham (2020), de Chaisemartin and D'Haultfoeuille (2020). As we have more than one birth cohort, we can separate birth cohort effects from time (year/month) effects. The 2007 reform is pre-determined to the workers' labour market participation. Yet, when (i.e. year/month) and how (i.e. with which job) workers started their first job is part of the treatment effect. From this point of view, treatment effects are staggered, and standard TWFE (with workers fixed effects) suffers from negative weight issues. However, averaging for groups (each of them made of two birth cohorts) and age (across each month and year for each worker of that age) is meant to approximate the sample share of each treatment group at each age. These approximations cannot be negative.

The fitted values of income levels before and after the reform can be appreciated on the vertical axes of Figure 2. This picture shows how they vary over age as the labour market experience increases. Yet, these are not treatment effects. To calculate them, we apply Equation 6, where we substitute to outcomes their fitted values.

Table 2: Treatment effects

	Men		**Women**	
	Euros	(%)	**Euros**	(%)
τ_1	74	8	54	8
τ_2	126	11	124	14
τ_3	233	17	169	16
τ_4	268	18	198	17
τ_5	296	19	254	20
τ_6	327	19	279	21
τ_7	334	19	283	20
τ_8	356	19	265	19
τ_9	377	20	314	21
τ_{10}	412	21	331	22
τ_{11}	467	23	315	20

This procedure shares some features of the one proposed by Borusyak et al. (2022). Yet, their imputation methodology estimates the difference between the outcome of each treated unit and the estimated potential untreated outcomes [$\hat{\tau}_{it} = Y_{it} - \hat{Y}_{it}(0)$]. Here, we are calculating the difference between fitted values at each age compared with the baseline at age 19. The gender gap in the fitted values is stark and amplifies with age. Women are paid less and have a flatter profile in the ages where the profile might (and should) grow steeper. The reform does not appear to advantage men more than women in percentage terms (see Table 3). However, as the fitted values of women's monthly income plateaued at age 29, the treatment effect dropped at age 30. We do not find this pattern for men whose profile keeps growing. With women delaying childbirth in the late twenties, the peak of their working career has passed at that age. Yet, as we stop there, we do not have the data to discuss this issue thoroughly. Nevertheless, we can underline that this pattern is observed before and after the reform.

We document the robustness of these findings in the Online Appendix A1.1 from a twofold perspective. In Figure A1, we use the same INPS data employed here but monthly 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 amount of education 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 only 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.[16] 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 being enrolled at the one-tier ($y_{it} = 1$) or two-tier path ($y_{it} = 0$) at time t. This regression model is our first stage that is specified as follows:

$$y_{it} = \sum_{a=0}^{6} \alpha_a AGE_{it}^{a} + \sum_{c=1985}^{1990} \theta_c + \beta T_i + \gamma X_i + u_{it} \tag{7}$$

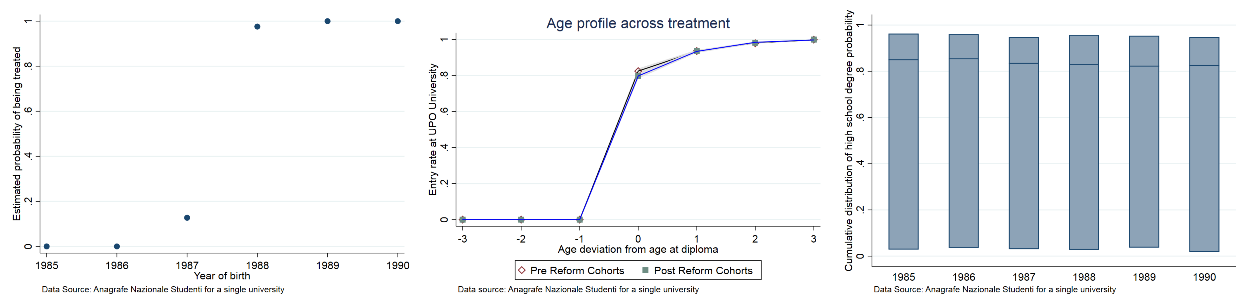
Here, θ_c controls for the birth year effect. All pupils born in the same year belong to a specific θ pool. Here the cohort of birth plays the role of time. We introduce in the model as covariates

[16] 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.

X: indicator functions for gender, public school, the type 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 7 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.[17] In Online Appendix B1, we report an additional test on the common trend hypothesis.

[17] More multiple-comparison tests of these distributions are available upon request from the authors. We always

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 on average more, switching from a two-tier to a longer one-tier track. Table 3 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 3: 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

find their statistical equality.

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 4: 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.[18] Our age windows are from one up to six years after high school graduation. 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)):

$$\begin{aligned} Y &= \mathbb{1}(X^e\beta_e + X^o\beta_o + V + \epsilon \geq 0) \\ X^e &= Z\alpha + X^o\beta_o + e \end{aligned} \tag{8}$$

Y denotes our discrete outcomes: the enrolment probability observed up to six years since high school completion.[19] 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.

[18] In Online Appendix B1, we report results on the graduation probability.

[19] In Online Appendix B1, we report results on the graduation probability.

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

	Outcome: Enrolment probability		
	Probit	IVProbit	SR
Years from high school graduation	(1)	(2)	(3)
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^{***}
	0.004	0.400	0.033
4	0.476^{***}	-0.404	0.300^{***}
	0.004	0.608	0.103
5	0.342^{***}	2.944^{***}	0.209^{***}
	0.004	0.073	0.062
6	0.210^{***}	1.974^{***}	0.165^{***}
	0.004	0.528	0.050

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

Table B14 reports our results. Column 1 reads the probit estimator; column 2 the ivprobit one, and column 3 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 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 B14 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. The reform proved valuable in two ways: it raised the years of schooling and changed the nature (i.e. different degree) 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. In particular, we aim at comparing earnings before and after the university degree for those treated and untreated by the 2007 reform. In this analysis, we use all selected individuals: everyone has a high school diploma, some of them eventually get a university degree. Since we use the combined INPS-COB archive, a university degree can reflect both a one-tier or the first degree of the two-tier track.

Here we have also data on university graduates to estimate the returns to a university (vs high school) degree, and in particular how they changed following the 2007 reform. Let's consider a standard 'mincerian' model augmented to account for the effect of the reform on average earnings (intercept effect) and on the college premium (slope effect):

$$y_{it} = (\sum_{a=0}^{11} \alpha AGE^{a}_{it}) * H_i + \beta T_i + \gamma G_{it} + \delta T_i * G_{it} + bX_{it} + u_{it} \tag{9}$$

where, to focus on main issues, X_{it} collects gender, year, month effects, as well as gender-specific age profiles.

G_{it} is a dummy variable that takes the value of one for workers with a university degree at time t. $T_i * G_{it}$ captures the differential effect of the reform on university returns.

Here, the set of age dummies measured as deviation from 19 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. To emphasise this issue, we multiply this term to H_i, a dummy variable equal to 1 for high-school achievers. As all workers in our sample are at least high-school achievers, this dummy is trivially 1 for all.

Reported estimates in Online Appendix C1, where we run separate equations by treatment status, are robust to this simple model specification. Hence, some restrictions we are putting in are harmless.

Different hypotheses on u_{it} reflect different regression models. An OLS model estimates Equation 10 as it stands; a fixed-effects model writes $u_{it} = e_i + \epsilon_{it}$ and treats e_i as a time-invariant person specific component in the estimation. Altonji and Zhong (2021) decompose $u_{it} = \delta HG_i + \nu_i + \epsilon_{it}$ with ν_i treated as random. HG_i takes the value of one for high-school achievers we also observe as university graduates by the last time we saw them in the data and it is time-invariant.

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 graduation (hg) fixed effects estimator (FE_{hg}) makes full use of the data. In this context, once we add the interaction HG_i as an additional control, we can identify the effect by including the number of individuals observed only before and the many observed only after their university degree. We also retain those who will never graduate to adjust the counterfactual experience profile without university graduation. Hence, our primary empirical model is:

$$y_{it} = (\sum_{a=0}^{11} \alpha AGE_{it}^{a}) * H_i + \beta T_i + \gamma G_{it} + \delta T_i * G_{it} + bX_{it} + \delta HG_i + \nu_i + \epsilon_{it} \qquad (10)$$

There are two conditions under which FE_{hg} 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, it assumes a common experience profile for high school achievers absent the reform.[20]

This choice feeds in the assumption that treated and untreated cohorts share abilities, preferences and labour market experience up to age 19. 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 $3,405,373$ observations for untreated individuals and

[20] In Equation 10, we assume that the treated and untreated cohorts share the age profile conditional on high school graduation and combined fixed effects, while separate regressions allow it to differ across treatment. We report here this restricted model to show the estimated standard errors. In the separate regressions, standard errors might be bootstrapped. We plan to conduct these estimates once we access the data. Nevertheless, comparing the last column on the right of Table 6 and the difference between the last column on the right of Table C5 and the third column on the left in Online Appendix, we see that the estimated coefficient is the same.

$3,092,575$ for treated ones. We use this sample to provide OLS, FE of Equation 9, and FE_{hg} estimates as in Equation 10. Then, we restrict our sample to those who have obtained a university degree by the last time we observe them. Here we have $1,329,215$ untreated observations and $1,153,214$ treated ones. We can apply the OLS and FE estimators for this restricted sample.

Our outcomes y_{it} are, as before, individuals' log earnings and income from INPS-COB records. In addition, we follow Altonji and Zhong (2021) and also consider as outcomes the occupational earnings (income) premiums. They are computed from OLS estimates of on an earnings/income log equation using the dataset with high school achievers and those with a higher education degree. In these equations, we control for occupational dummies,[21] 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. The estimated premiums are the estimated coefficients of the occupational dummies. We can identify 34 occupational categories in our data, based on the three digits International Standard Classification of Occupations (ISCO) codes. Table C1 reports the summary statistics of occupational outcomes by treatment groups. These premiums are then attributed to each individual according to the occupation possessed at time t. This additional analysis allows to understand how much of the returns to a university degree are due to job occupations.

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.[22] 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 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

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

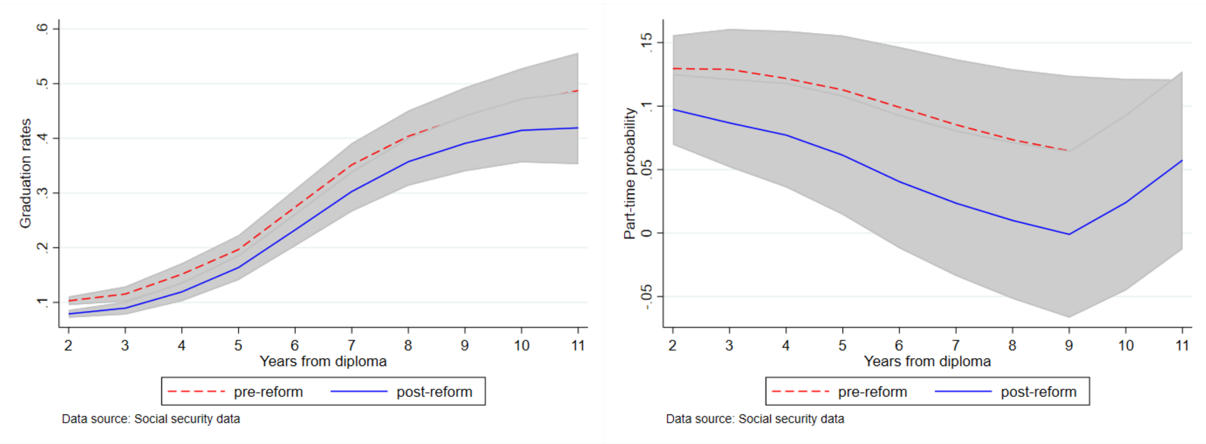
[22] Controlling for age and time dummies is equivalent to controlling for birth year dummies.

4 highlights that the FE and FE_{hg} estimators are likely similar.

Panel (a) of Figure 4 shows that treated and untreated cohorts share the same age profile for the propensity of being observed with a university degree. Here, the graduation indicator takes the value of 1 for all higher education degrees. Hence, those who get their degree after the first tier of the two-tier track are assigned the value of 1. Panel (a) of Figure 4 indicates that the reform did not increase the number of higher educated workers. Yet, it changed the higher education degree, leading people to switch from the two-tier to the one-tier track. This latter track raises the years of schooling and offers degrees better rewarded in the labour market (such as Medicine and Engineering). Our estimated impact on the return of graduation feeds into these different degrees. We also check if our results could be due to a different use of part-time jobs across the two groups. Panel (b) of Figure 4 reports this is not the case.

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

Table 6 reports our results. We estimate Equation 10 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

Table 6: Returns to a university degree

	Baseline			Differential impact of the 2007 reform		
Outcomes:	OLS	FE	FE_{hg}	OLS	FE	FE_{hg}
Monthly 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
Monthly 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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. As this feature holds before and after the reform, those who graduate earlier after it are more likely to graduate earlier in the one-tier rather than in the two-tier track. We find that the 2007 reform raised the earnings (income) returns to university degrees by 6 percentage points.

The baseline effect on occupational earnings/income premiums is small and about one tenth of that on actual data. This captures the part of the college return that is occupation-specific and occurs within occupations. This suggests that most of the college premium is across occupations, i.e. that the reason why university graduates earn higher wages than secondary school graduates is because they (on average) have better paid occupation, and not because they earn more in the same job. The last column shows that the reform contributes to raise the within occupation component of the college premium, but by the same amount than in the baseline. Hence, the reform contributed to increases the college premium mostly by reallocating college graduates across (better paying) occupations.

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 10 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.

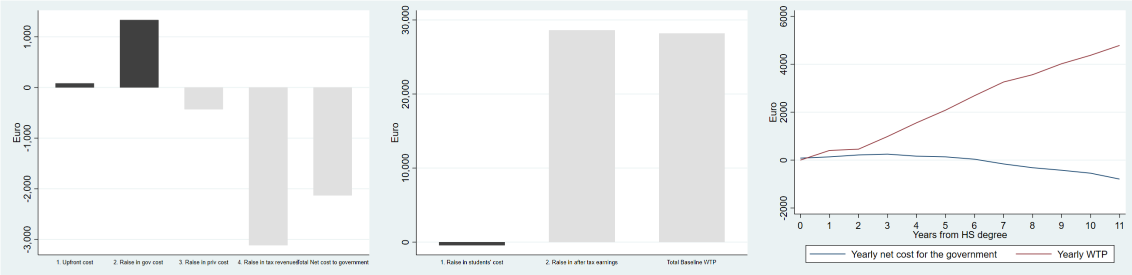
Our results send a clear message: auditing the high-school grade assessment matters. It matters because it likely impacts the choices pupils make and the skills and knowledge they develop. It matters because the labour market values and rewards it.

7 Welfare Analysis

We follow Hendren and Sprung-Keyser (2020) to conduct a welfare analysis of the policy that changed the high school grades assessment rules, and we calculate the Marginal Value of Public Funds (MVFP). The ratio between the policy benefits and its net costs defines this MVFP. Two ingredients make this calculation. The first is the sum of the initial spending associated with the policy and the long-run impact on the government budget, which accounts for the costs. The second is the recipients' willingness to pay for this policy that measures the benefits. The MVFP reckons the delivered welfare per euro of government spending on the policy. We will use our estimates to gauge the reform's impact on the earnings of workers aged from 19 to 30. We show that this short time is enough for the policy to pay for itself.

Net Costs: Figure 5 panel (a) shows how we calculate the net cost of the reform. We

Figure 5: WTP and Cost Components for auditing the grade assessment



(a) Net Government Costs (b) Willingness to Pay (c) Yearly

start from the upfront monetary costs of an external board to high school exams. We recover these upfront costs from the accompanying report to the reform.[23] The overall spending for the examining boards of the high school final exam in the year before the reform (i.e. the schooling year 2005/2006) amounted (in euros) to 103,24 million, while the updated amount for the first year of application of the reform was 143 million, with an increase of only 39.76 million. As high school graduates were around 480.000 in 2007, the per capita cost of the reform was about €83 per student.

We then add to these upfront costs of the reform the net government costs of each additional year of tertiary education. We compare these costs between treated and untreated cohorts observed in the 19-30 age range. OECD (2013) provides the annual expenditure per student in tertiary education amounting to €7.226 in 2010.[24] Yet, 32.4% of this spending (of which households pay 24.4%) are covered by individuals, businesses and other private sources. Hence, we subtract 2.341 euros to calculate the costs net of these private expenditures.[25] We can estimate these net government costs for each year between 0 and 11 from high school graduation because of one feature of the tertiary education system: the long time-to-degree. This feature affects the government net costs and people's foregone and earned incomes. It comes as a side effect of the flexibility to be enrolled without time limits and sit exams how many times a student wishes. As the number of examination sessions is limited, students have several sessions in an academic year at their disposal. They can choose the exam to sit in, reject the grade and re-take the exam to grade up. With these rules put in place, few students comply with the legal duration of the degrees.

We use the same INPS data of Section 6 to calculate the graduate share and the share of those observed as graduates in each year. In our data, 38% of workers are recorded as graduates.

[23]The Italian Constitution mandates a report to account for any additional public spending.
[24]See OECD (2013) Chart B1.2 page 165 (figures in dollars). We apply the 2010 average exchange rate of 1.3257 dollars per euro.
[25]See OECD (2013) Table B3.2b page 207.

This proportion reflects the drop-out of some people who enrol at university but never complete their course degrees. We keep this percentage constant. Yet, we multiply it by the proportion of workers whose education is a university degree. These proportions vary over time, as workers get their degrees and are positive from the third year. We calculate them in the subsamples of treated and untreated workers as it takes more time to graduate after the reform. These calculations return our measure of the yearly net public costs of one more year of tertiary education. We then add up all these yearly costs to obtain the total net public costs after 11 years from high school completion by treatment groups.[26]

The yearly and total difference in these net costs between treated (who study longer) and untreated workers measures the change in these costs caused by the high school reform. After 11 years from high school completion, the government costs (in euros) for treated cohorts are 1.332 as high as for the untreated ones, while the estimated increase in private costs is 431. Hence, the net costs amount to 901 euros. The other costs for the government stem from revenues changes induced by the reform. They come as revenues losses due to foregone earnings when students stay at university for more time. Yet, the increased earnings of treated workers compared with the untreated ones raise revenues. We calculate tax revenues by applying the Italian Personal Income Tax schedule (IRPEF) to our earnings profiles as estimated in Sections 2 and 3.[27]

We use the birth year to map age profiles in tax years. We can separate these figures by gender, education and treatment group for each age between 19 and 30. When treated cohorts are 30, they gained 17,533 euros more than untreated ones, raising tax revenues by 2,046 euros. As these tax revenues overcome the upfront costs of the reform (83 euros) and the government net costs (901 euros), the reform has paid for itself. In the Online Appendix, we show that the reform paid for itself after six years from high school completion.

Willingness to pay: We have shown that the reform led to net savings to the government. Hence, the policy has an infinite MVPF, provided that the willingness to pay (WTP) is greater than zero. We calculate this WTP as in Hendren and Sprung-Keyser (2020). The estimation of the WTP follows the logic of the envelope theorem and revealed preference. Under this logic, the reform increased income is due to higher returns to human capital rather than to a higher level of effort. We can then estimate the WTP through the policy's impact on net income. Panel (b) of Figure 5 shows how we calculate it. Treated workers paid 325 euros more than the

[26] All costs, taxes and earnings changes are scaled back to the time of the implementation of the reform using a 3% discount rate.

[27] The Italian Personal Income Tax (IRPEF) was reformed in 2008. Therefore, we apply to incomes earned in the first three years (2005, 2006 and 2007) the "old" IRPEF and the "new" IRPEF structure in the following years. Changes concerned tax brackets (from 4 to 5) and the substitution of tax deductions with tax credits. Tax rates remained in a range comprised between 0 (no tax area) and 43%.

untreated ones to cover the additional spending in education.[28] However, they earned 15,486 euros more compared with the untreated ones. As the benefits are higher than the costs, the WTP is positive and equal to 15,162 euros.

8 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 records social security contributions for a random sample of the universe of Italian workers in all sectors, including self-employment periods. 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. Yet, we do not find relevant differences in treatment effects across gender.

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 by about 21 percentage points.

[28] We calculate this 325 euros as 24.4% of the yearly expenditures for tertiary education. As the OECD (2013) reports, this percentage is the share paid by the households.

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 would have experienced the same age profile without the policy. 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 (about nine percentage points for women versus two percentage points for men). 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. On top of that, our welfare analysis proves that this reform paid for itself just after six years from high school completion.

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.

Borusyak, K., Jaravel, X. and Spiess, J. (2022). Revisiting event study designs: Robust and efficient estimation, *Discussion Paper DP17247*, CEPR.

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.

Hendren, N. and Sprung-Keyser, B. (2020). A unified welfare analysis of government policies, *Quarterly Journal of Economics* **135**(3): 1209–1318.

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).

Appendix

We organise this appendix as follows. Section A complements our analysis on the first remark; Section B1 refers to our second remark where we also present as a robustness check a replication of the analysis using ISTAT national data (Section B2). Section C1 relates to our third one. Additional materials can be found in Section D1.

A Additional analysis on our first remark

A.1 Static Two-way fixed effects model

We here review why the static OLS estimator is short-run bias if treatment effects vary over time, as assumed in Table A1.

Table A1: Two groups, Twelve-period

$E[Y_{it}]$	(i=T)	(i=U)
t=0	α_T	α_U
t=1	$\alpha_T + \beta_1 + \tau_1$	$\alpha_U + \beta_1$
t=2	$\alpha_T + \beta_2 + \tau_2$	$\alpha_U + \beta_2$
t=3	$\alpha_T + \beta_3 + \tau_3$	$\alpha_U + \beta_3$
t=4	$\alpha_T + \beta_4 + \tau_4$	$\alpha_U + \beta_4$
t=5	$\alpha_T + \beta_5 + \tau_5$	$\alpha_U + \beta_5$
t=6	$\alpha_T + \beta_6 + \tau_6$	$\alpha_U + \beta_6$
t=7	$\alpha_T + \beta_7 + \tau_7$	$\alpha_U + \beta_7$
t=8	$\alpha_T + \beta_8 + \tau_8$	$\alpha_U + \beta_8$
t=9	$\alpha_T + \beta_9 + \tau_9$	$\alpha_U + \beta_9$
t=10	$\alpha_T + \beta_{10} + \tau_{10}$	$\alpha_U + \beta_1 0$
t=11	$\alpha_T + \beta_{11} + \tau_{11}$	$\alpha_U + \beta_1 1$

As a reference point, let's start from the standard difference-in-differences estimator in the two groups and two time periods case.

Table A2: Two groups and two time periods

$E[Y_{it}]$	(i=T)	(i=U)
t=0	α_T	α_U
t=1	$\alpha_T + \beta + \tau_1$	$\alpha_U + \beta$

$$\beta_{did} = E(Y_{T1} - Y_{T0}) - E(Y_{U1} - Y_{U0}) = \tau_1 \tag{A1}$$

This result also applies when treatment effects are homogeneous.

However, if effects are heterogeneous, as in Table A1, the static OLS estimator is equal to:

$\beta_{static} = E(Y_{T1} - Y_{U1}) - \frac{1}{11}E(Y_{T0} - Y_{U0}) - \frac{1}{11}E(Y_{T2} - Y_{U2}) - \frac{1}{11}E(Y_{T3} - Y_{U3}) - \frac{1}{11}E(Y_{T4} - Y_{U4}) - \frac{1}{11}E(Y_{T5} - Y_{U5}) - \frac{1}{11}E(Y_{T6} - Y_{U6}) - \frac{1}{11}E(Y_{T7} - Y_{U7}) - \frac{1}{11}E(Y_{T8} - Y_{U8}) - \frac{1}{11}E(Y_{T9} - Y_{U9}) - \frac{1}{11}E(Y_{T10} - Y_{U10}) - \frac{1}{11}E(Y_{T11} - Y_{U11})$

$\beta_{static} = \tau_1 - \frac{1}{11}\tau_2 - \frac{1}{11}\tau_3 - \frac{1}{11}\tau_4 - \frac{1}{11}\tau_5 - \frac{1}{11}\tau_6 - \frac{1}{11}\tau_7 - \frac{1}{11}\tau_8 - \frac{1}{11}\tau_9 - \frac{1}{11}\tau_{10} - \frac{1}{11}\tau_{11}$

It then suffers from a short-run bias as it places a negative weight on medium-run effects.

A.1.1 Estimates

In each of the three rows of Table A3, 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 INPS archive that covers the private sector only. We display the treatment effects for monthly earnings (available from both archives), for monthly income (income from self-employment included, Columns 1, 3 and 5) and daily earnings (remaining columns).

Table A3: Treatment effects: sample of high school graduates

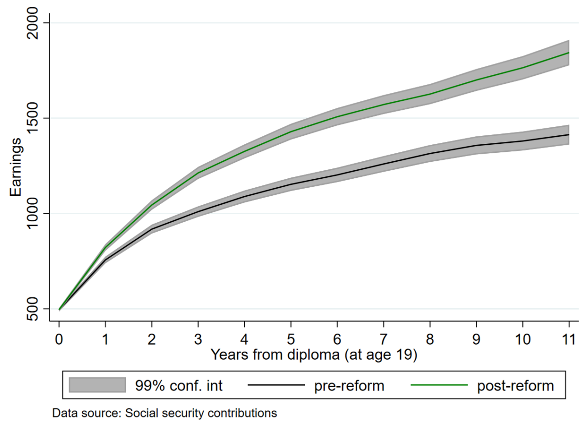
	All		**Males**		**Females**	
	All sectors	Private	All sectors	Private	All sectors	Private
Outcomes:	(1)	(2)	(3)	(4)	(5)	(6)
Monthly Earnings	-0.001	0.001	0.007^{***}	0.012^{***}	-0.008^{***}	-0.011^{***}
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
Monthly Income	-0.001	-	0.006^{***}	-	-0.014^{***}	-
	(0.002)	-	(0.002)	-	(0.002)	-
Daily Earnings	-	0.000	-	0.007^{***}	-	-0.007^{***}
	-	(0.000)	-	(0.000)	-	(0.000)

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

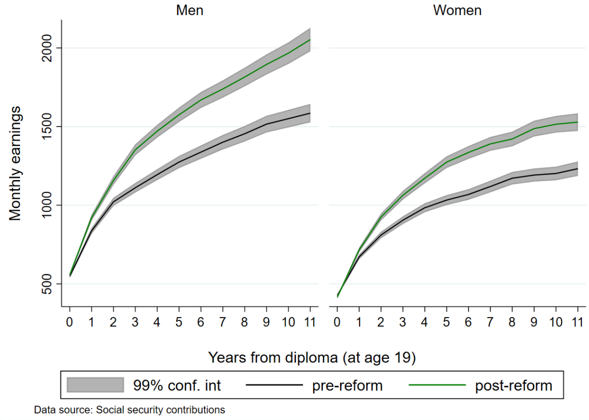
A.2 Robustness using different earning measure and INPS archive for the entirety of the private sector

Figure A1 uses data from INPS social security contributions, thus considering all earnings from dependent work. Figures A2 and A3 use INPS data on the universe of workers in the private sector only.

Figure A1: Age profiles of monthly earnings by treatment group

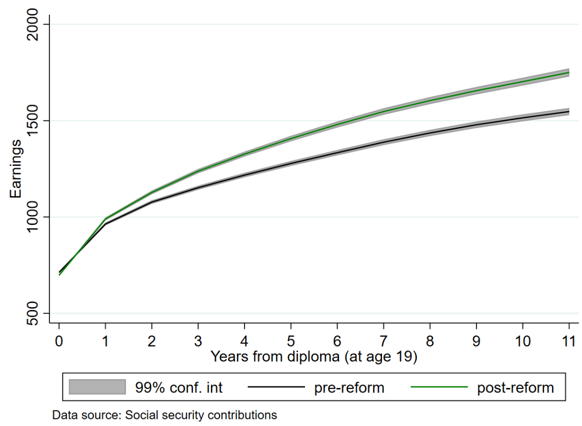


(a) Monthly Earnings: all workers

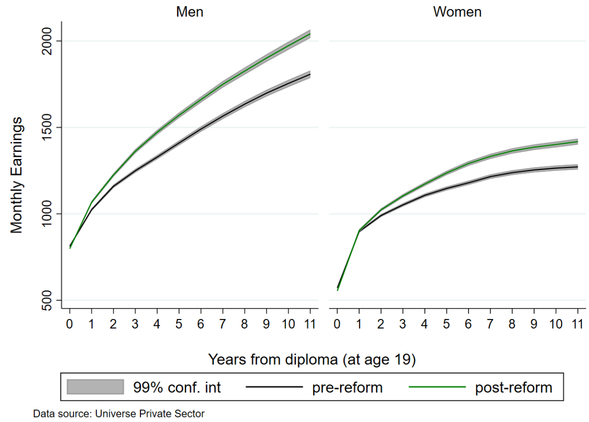


(b) Monthly Earnings: by gender

Figure A2: Age profiles of monthly earnings by treatment group

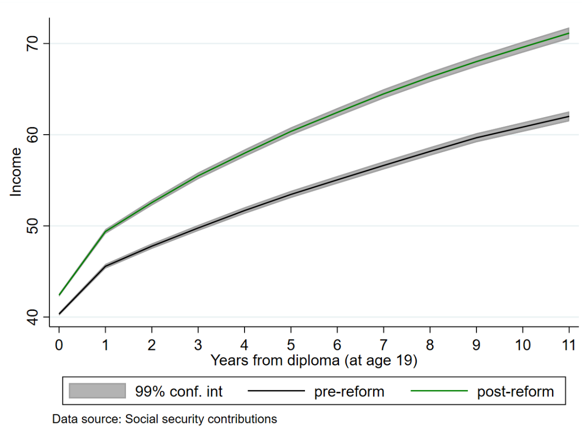


(a) Monthly Earnings: all workers

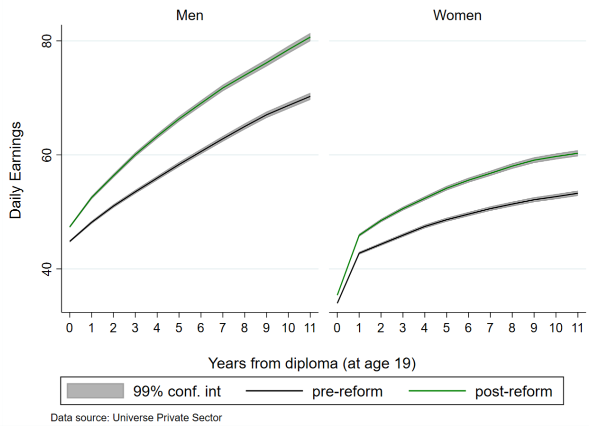


(b) Monthly Earnings: by gender

Figure A3: Age profiles of daily earnings by treatment group



(a) Daily Earnings: all workers



(b) Daily Earnings: 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	70,770	80,880

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

Table B14: Marginal Effects of enrolment at the one-tier track.

	Outcome: Graduation Probability		
	Probit	IVProbit	SR
Years from high school graduation	(1)	(2)	(3)
3	-0.312^{***}	-3.496^{***}	-0.237
	0.002	0.081	0.159
4	-0.332^{***}	-3.146^{***}	-0.087
	0.002	0.009	0.150
5	-0.241^{***}	-2.908^{***}	-0.141
	0.003	0.046	0.143
6	-0.136^{***}	-2.664^{***}	-0.133
	0.004	0.119	0.194

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

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 B15: 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

*N*otes: 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	11	10	11	11
90th quantile	11	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

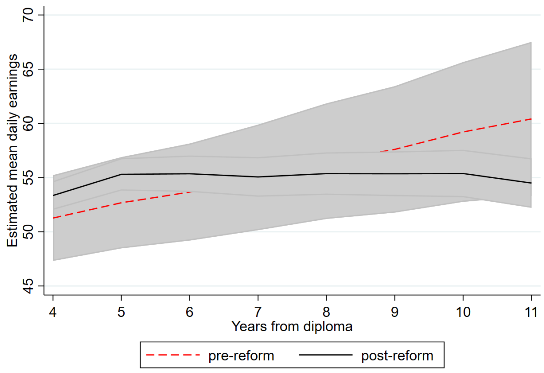
	Baseline			Differential impact		
Outcomes:	OLS	FE	FE_{hg}	OLS	FE	FE_{hg}
Earnings	0.136^{***}	0.149^{***}	0.331^{***}	0.068^{***}	0.091^{***}	0.069^{***}
	0.002	0.002	0.003	0.002	0.004	0.003
Daily Earnings	0.150^{***}	0.045^{***}	0.194^{***}	0.048^{***}	0.040^{***}	0.048^{***}
	0.001	0.001	0.002	0.002	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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

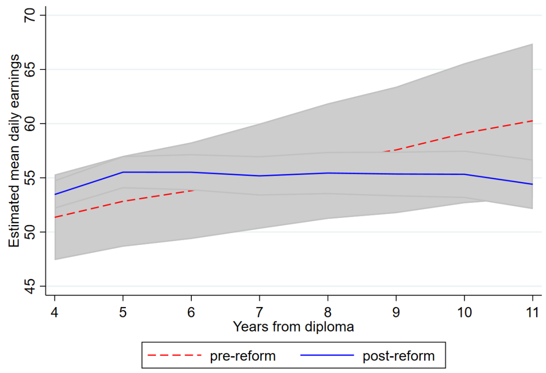
C1.1 Estimating earnings profile using University-INPS data

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.

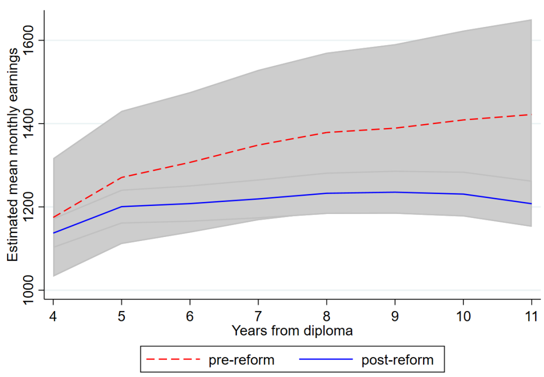
Figure C1: Earnings profiles: private sector



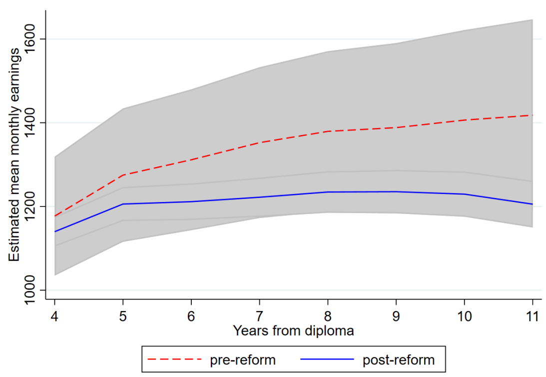
(a) Daily earnings: unbalanced panel



(b) Daily earnings: balanced panel

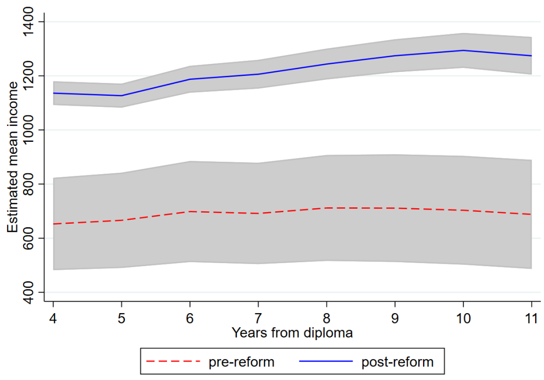


(c) Monthly earnings: unbalanced panel

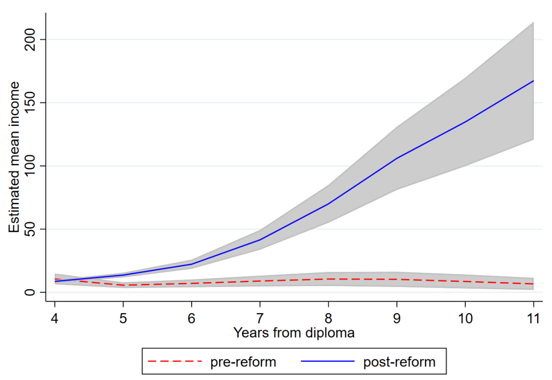


(d) Monthly earnings: balanced panel

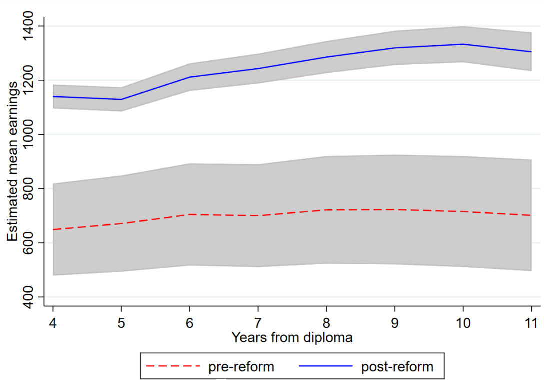
Figure C2: Earnings profiles: all sectors and self-employment (source: Social Security Contributions)



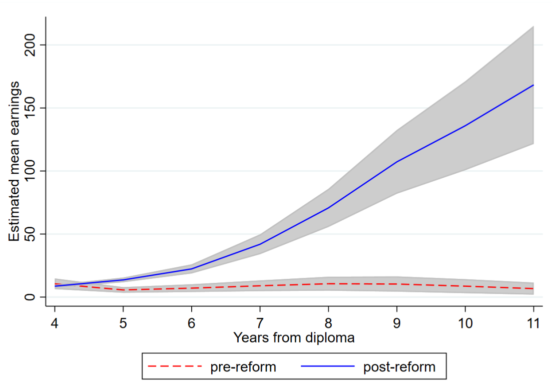
(a) Daily earnings: unbalanced panel



(b) Daily earnings: balanced panel

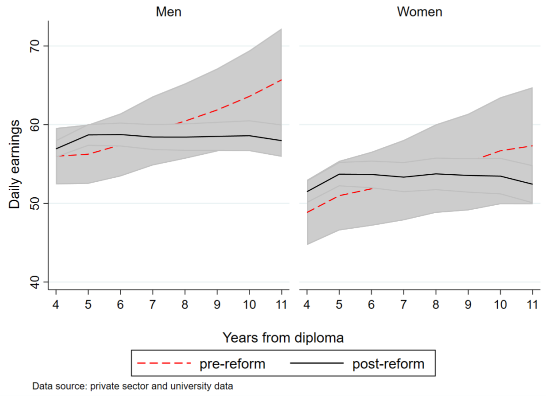


(c) Monthly earnings: unbalanced panel

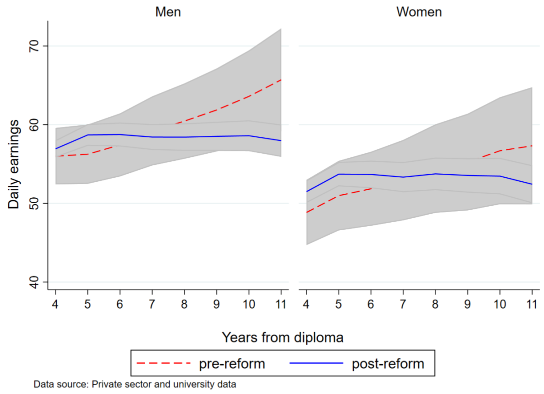


(d) Monthly earnings: balanced panel

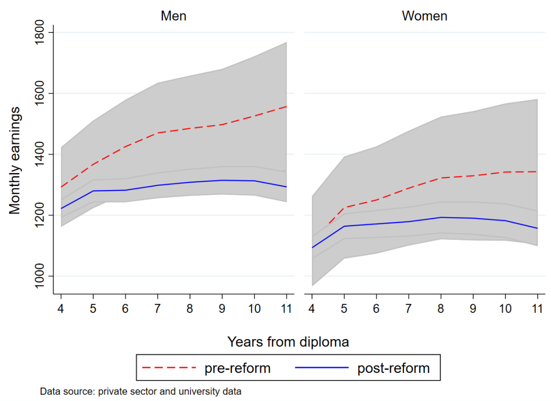
Figure C3: Earnings profiles: private sector



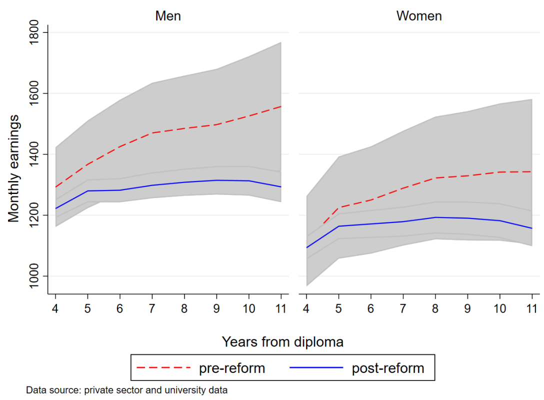
(a) Daily earnings: unbalanced panel



(b) Daily earnings: balanced panel

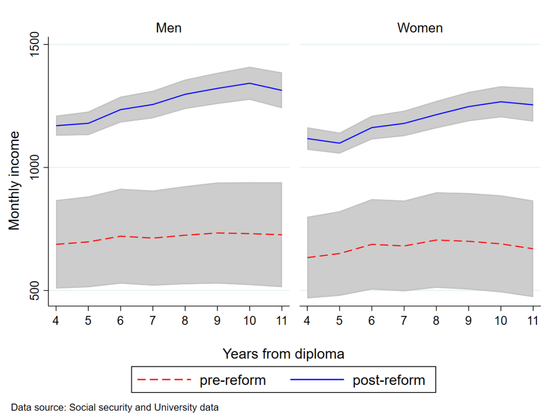


(c) Monthly earnings: unbalanced panel

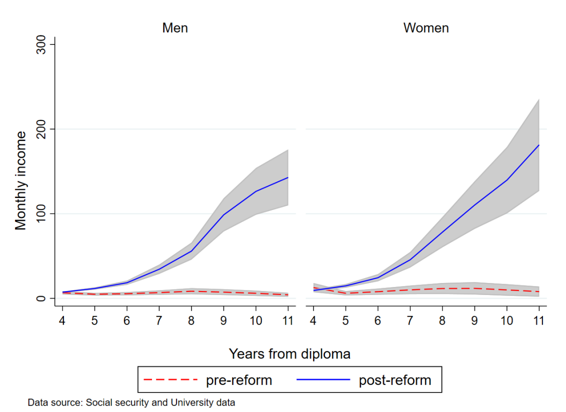


(d) Monthly earnings: balanced panel

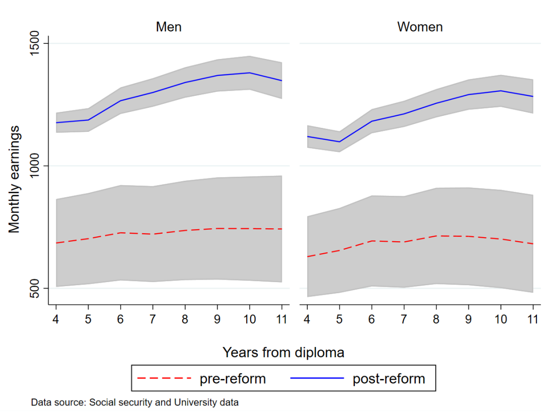
Figure C4: Earnings profiles: all sectors and self-employment (source: Social Security Contributions



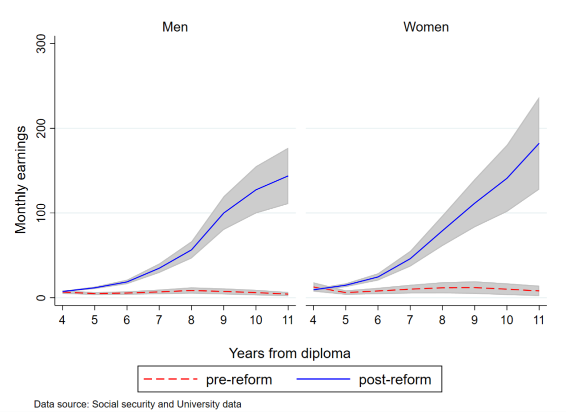
(a) Daily earnings: unbalanced panel



(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

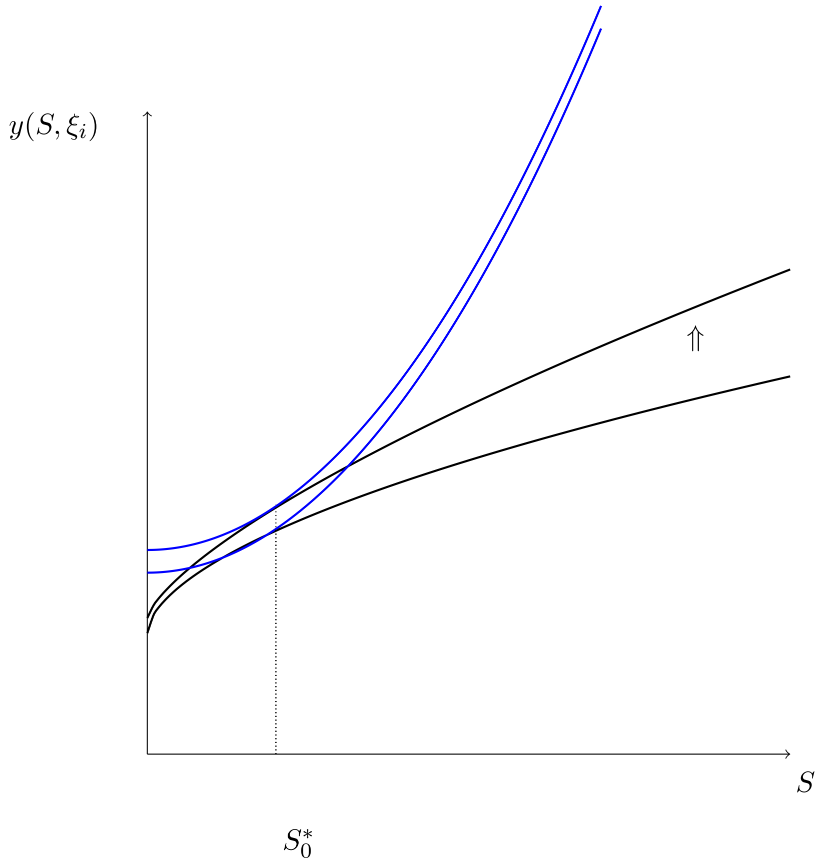


Figure D1: Optimal choice of years of schooling