



Marcio Gold Firmo

**Fake Diplomas and Signaling in Labor Markets:
Evidence from Brazil**

Tese de Doutorado

Thesis presented to the Programa de Pós-graduação em Economia of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Economia.

Advisor: Prof. Gustavo Gonzaga

Rio de Janeiro
April 2021



Marcio Gold Firmo

**Fake Diplomas and Signaling in Labor Markets:
Evidence from Brazil**

Thesis presented to the Programa de Pós-graduação em Economia of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Economia. Approved by the Examination Committee.

Prof. Gustavo Gonzaga

Advisor

Departamento de Economia – PUC-Rio

Prof. Juliano Junqueira Assunção

Departamento de Economia – PUC-Rio

Prof. Rodrigo Reis Soares

– Insper Instituto de Ensino e Pesquisa

Prof. Naercio Aquino Menezes Filho

– Insper Instituto de Ensino e Pesquisa

Prof. Breno Gomide Braga

– Urban Institute

Rio de Janeiro, April 27th, 2021

All rights reserved.

Marcio Gold Firmo

Completed his Bachelor of Arts degree in Economics from PUC-Rio in 2005 and obtained his Master of Science degree in Economics from PUC-Rio in 2008. Now holds a PhD degree in Economics from PUC-Rio.

Bibliographic data

Firmo, Marcio

Fake Diplomas and Signaling in Labor Markets: Evidence from Brazil / Marcio Gold Firmo; advisor: Gustavo Gonzaga. – Rio de Janeiro: PUC-Rio , Departamento de Economia, 2021.

v., 66 f: il. color. ; 30 cm

Tese (doutorado)- Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Economia.

Inclui bibliografia

1. Economia - Teses – Teses. 2. Educação;. 3. Mercado de Trabalho;. 4. Sinalização;. 5. Capital Humano..
I. Gonzaga, Gustavo. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Economia. III. Título.

CDD: 004

Aos meus pais, Miriam e Sérgio.

Acknowledgments

First of all, I would like to thank my parents, Miriam Gold and Sergio Firmo, for all the unconditional love received during my whole life, which shaped me and ultimately this thesis. This work is entirely yours: I am because of you.

I am deeply thankful to my advisor Gustavo Gonzaga. Without his patience, empathy and guidance this thesis would not be possible. Each time I got lost in the long marathon of research and writing needed for the thesis, he would pull me back in the track and provide me important academic feedback.

I am grateful to all the members of my committee: Breno Braga, Juliano Assunção, Rodrigo Soares and Naércio Menezes. Their suggestions helped me take this research to another level of quality. It was a great pleasure to have in the committee professors that had accompanied me in the early years of my academic career, either by advising me (Rodrigo during my masters, and Juliano during undergrad) or as a colleague (Breno).

I am indebted to Vitor Pereira, the most formidable education economist I know, for this thesis. From the very first start of the idea, to the development of the research question, collection of data and exploration of the results, he has been an enthusiast of the project and acted like a coadvisor, always there for me. Thank you, my friend.

I am also indebted to the Education Department of Rio de Janeiro State, and specially, to Alessandro Sathler, the head of inspection of schools during these years. He believed in the importance of the project, in our responsibility in treating confidential data, and gave us access to invaluable records that made the research possible.

The ambiance of the Department of Economics in PUC-Rio has always been a great asset and helped flourish collaboration and friendships. During the first year of my PhD studies, while I was commuting from downtown (where my employer, BNDES, is located) and Gavea, Conrado Garcia has helped me a lot. In the following years, my angels (and friends) Pedro Américo and Flavio Carneiro helped me many times in a true altruistic manner, for which I am deeply thankful. I am also thankful to Bianca Belotti, who many times clarified me about procedures, always with patience and kindness.

I am blessed to have so many incredible people and true friends in my life. I do not want to separate between personal and professional experiences when I thank all these people around me, for all the help and companionship in different moments and aspects of my years: Romero Rocha, Pedro James, Bernardo Scotti, Felipe da Cás, Gabriel Buchmann, Carol Salcedo (plus Pedro

and Manu, two little angels that teach me so much), Flavio Moraes, Gabriela Podcameni, Mariana Lopes, Marcelo Conde, Barbara Dias, Luis Cesar, Marcelo Kropf and Roberto Hsu.

I also thank BNDES for supporting my PhD project, and the colleagues directly involved, specially, Leonardo de Oliveira Santos and Pedro Iooty. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Abstract

Firmo, Marcio; Gonzaga, Gustavo (Advisor). **Fake Diplomas and Signaling in Labor Markets: Evidence from Brazil**. Rio de Janeiro, 2021. 66p. Tese de Doutorado – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

This thesis studies different aspects of the interaction between education and labor market outcomes using a still largely unexplored phenomenon: the selling of thousands of high school and undergraduate diplomas in Brazil. The first chapter describes this new database, comprising thousands of Adult High School (AHS, or EJA in Brazil) diplomas sold for individuals that never really completed their degrees. Our analysis of the performance of these individuals in the labor market shows consistent wage gains of buying a diploma, with earnings 5% to 8% higher in comparison with individuals without a high school diploma during our period of analysis. Back-of-the-envelope calculations considering the pecuniary cost of buying a diploma indicate a rate of return over 100%, or present value corresponding to two to six months of wages. We interpret this as potential evidence of alternative mechanisms: (i) signaling effects in a context of labor markets with incomplete information, (ii) asymmetric employer learning or (iii) learning-by-doing. The second chapter explores the same database to investigate human capital accumulation in real AHS. Initial investigation shows that individuals that completed AHS regularly enjoy a wage premium. However, compared to workers that merely bought diplomas, those who studied show no apparent advantage. Taken together, the results from the first two chapters indicate that the benefits from AHS are related to a pure diploma effect, and not to human capital accumulation. The third and last chapter explores a new database of fake college graduates between 2011 and 2016 who were exposed and had their degrees canceled following a large police investigation. We present preliminary evidence about wage gains and establish additional research questions that can be addressed with this database. As a whole, the thesis reinforces the importance of diplomas as signaling devices for the labor market.

Keywords

Education; Labor Market; Signaling; Human Capital.

Resumo

Firmo, Marcio; Gonzaga, Gustavo. **Diplomas Falsos e Sinalização em Mercados de Trabalho: Evidências do Brasil**. Rio de Janeiro, 2021. 66p. Tese de Doutorado – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

Esta tese estuda a interação entre escolaridade e mercado de trabalho a partir de um fenômeno ainda pouco explorado na literatura: a venda de milhares de diplomas de ensino médio e de graduação no Brasil. O primeiro capítulo descreve a nova base de dados construída, composta por milhares de diplomas de Ensino Médio (EM) da modalidade de Educação de Jovens e Adultos (EJA), expedidos para pessoas que nunca concluíram seus estudos. A análise mostra que a compra do diploma resulta em ganhos salariais consistentes, com rendimentos 5% a 8% maiores em comparação com os indivíduos sem diploma de ensino médio durante o período de análise. Considerando apenas o seu custo pecuniário, estima-se que a compra de um diploma tenha uma taxa de retorno superior a 100%, ou um valor presente entre dois e seis meses de salário. Esses resultados são consistentes com: (i) uma possível sinalização de habilidade em mercados com informações imperfeitas, (ii) aprendizagem assimétrica entre o empregador e outras firmas do mercado, ou ainda (iii) fenômenos de *learning-by-doing*. O segundo capítulo explora o mesmo banco de dados para investigar a acumulação de capital humano no EM-EJA. A análise mostra que os indivíduos que concluíram de forma regular o EM-EJA desfrutaram de certo prêmio salarial em relação aos que não completaram o EM, mas não em relação àqueles que compraram diploma. Tomados em conjunto, os resultados dos dois primeiros capítulos indicam que os benefícios do EM-EJA estão relacionados a um efeito-diploma e não à acumulação de capital humano. O último capítulo explora um novo banco de dados de diplomas universitários expedidos entre 2011 e 2016 que foram considerados irregulares e cancelados após investigação judicial. Apresentamos evidências preliminares sobre ganhos salariais e estabelecemos questões de pesquisa adicionais que podem ser estudadas com este banco de dados. De maneira geral, a tese reforça a importância dos diplomas como dispositivos de sinalização para o mercado de trabalho.

Palavras-chave

Educação; Mercado de Trabalho; Sinalização; Capital Humano.

Table of contents

1	Fake Diplomas and Signaling	12
1.1	Introduction	12
1.2	Context	14
1.3	Data	16
1.4	Empirical Strategy	20
1.5	Main Results	23
1.6	Discussion	28
1.7	Conclusion	30
2	Adult High School in Brazil: investigating human capital accumulation	32
2.1	Introduction	32
2.2	Institutional Environment	33
2.3	Data	35
2.4	The effects of Adult High School	37
2.5	Disentangling human capital from diploma effects	42
2.6	Conclusion	48
3	University Diplomas	49
3.1	Introduction	49
3.2	Context and Data	50
3.3	Empirical Investigation	51
3.4	Final remarks	56
4	References	57
A	Appendix to Chapter 1	59
A.1	Evidence gathered	59
A.2	Probability of treatment	61
A.3	Results	61
B	Appendix to Chapter 2	64
B.1	Treatment probabilities	64
C	Appendix to Chapter 3	65

List of figures

Figure 1.1	Distribution of Diploma Acquisitions	19
Figure 1.2	Evolution of the share of workers with high school	19
Figure 1.3	Evolution of wages	20
Figure 1.4	Effects on probability of high school job	24
Figure 1.5	Without controlling for wages	24
Figure 1.6	Controlling for wages	25
Figure 1.7	Individual FE	27
Figure 1.8	Individual FE including individuals with wage 23-25 yrs missing	27
Figure 1.9	Effects on wages	28
Figure 1.10	Individual FE, with missing wagers, only 30+ years old	29
Figure 1.11	Low and high previous experience	29
Figure 2.1	Schooling evolution by cohorts	34
Figure 2.2	Evolution of schooling and wages around AHS graduation	38
Figure 2.3	Evolution of wages: aggregate and cohorts	39
Figure 2.4	Effects of graduating in AHS on schooling	40
Figure 2.5	With control of previous wages	40
Figure 2.6	Individual FE	41
Figure 2.7	Evolution of schooling and wages around AHS graduation	44
Figure 2.8	Effects on Schooling	45
Figure 2.9	Effects on wages	46
Figure 2.10	Effects on wages - fixed effects model	46
Figure 2.11	Effects on wages - fixed effects model	47
Figure 3.1	Distribution across Brazilian states	52
Figure 3.2	Sectoral Distribution	53
Figure 3.3	Distribution of contract types	53
Figure 3.4	Evolution of secondary education and wages	53
Figure 3.5	Diploma Effects	55
Figure 3.6	Diploma Effects - public versus private sectors	55
Figure A.1	Email exchanged with proposal	59
Figure A.2	Two diplomas with the exact same grades	60
Figure A.3	Example of a payment for a diploma, with the same value as proposed in the email received	60
Figure A.4	Kernel	61
Figure A.5	FE, with missing, all ages, low pscores	63
Figure A.6	FE, with missing, all ages, high pscores	63
Figure B.1	Pscores estimated: real AHS versus diploma buyers	64

List of tables

Table 1.1	States	19
Table 1.2	Descriptive Statistics	23
Table 1.3	Diploma Effects on reported schooling and wages	26
Table 2.1	Descriptive Statistics: Adult High School graduates and random sample	39
Table 2.2	AHS effects on schooling and wages	41
Table 2.3	Descriptive Statistics: Diploma buyers versus real AHS graduates	44
Table 2.4	Effects on Schooling and Wages	47
Table 3.1	Regression results	56
Table A.1	Diploma Effects on Wages - other samples	62

1

Fake Diplomas and Signaling

1.1

Introduction

Whether education increases the productivity of individuals or simply signals workers productivity to the firms in a scenario of incomplete information has been an active debate since the formulation of the theories of human capital (Mincer, 1958; Schultz, 1960; e Becker, 1964) and the signaling theory (Spence, 1973). While the human capital theory has been extensively used to explain a wide array of real-world phenomena, as the distribution of wages (Becker, 1967), the life cycle wage profile (Mincer, 1974), productivity differences between cities (Moretti, 2010) and across countries (Lucas, 1988) and has justified the formulation of policies to increase the educational attainment of the whole population, the signaling theory would posit quite different policy prescriptions.

However, both theories are difficult to differentiate empirically, as both imply a positive relationship between education and wages, as more years of schooling or better school quality might improve productivity but also might act as signals of higher innate productivity at the same time. The recent literature has tried to disentangle both effects by comparing individuals with arguably the same levels of human capital, with and without a diploma (Clark and Martorell, 2014, Mazrekaj and Cabus, 2019).

In this chapter, we contribute to this literature by looking at the relationship between human capital acquisition and the signaling value of diplomas through a different perspective, by investigating a still large unexplored phenomenon in developing countries: schools that issue diplomas for individuals who actually did not study.

As a mere piece of paper, diplomas should not increase the productivity of any worker if the sole role of education is to increase the human capital of individuals. By the contrary, if diplomas are used as a steppingstone to get into better jobs or employers believe that diplomas are a signal of workers productivity, workers that buy a diploma would see an increase in wages after obtaining the degree. The wage increase, however, could be only momentary if employers learn that the worker lacks basic skills that should have been learned at school, or if they eventually discover that the diploma is false. Alternatively, if individuals succeed in signaling innate ability with a diploma, wage increases could persist over time.

In order to test these predictions, we use a matched employer-employee data on formal firms in Brazil and detailed data linking students who bought fake diplomas in some selected schools. These hundreds of thousands of diplomas were scrutinized by administrative and criminal investigations after reports of school fraud. Most of the diplomas were issued for individuals over 19 years old (the limit for getting a high school diploma in Brazil) and corresponded to Adult High School (AHS, or EJA in Brazil) diplomas. A key

feature of our data is that we are able to observe individuals' wages before and after the acquisition of the diploma as we have the full history of formal employment since 2006. We apply two strategies that essentially use wages before the acquisition of the diploma to control for individuals' unobserved productivity: first, we consider only individuals that buy diplomas at ages over 25 years old and control for the wages these individuals earned from 23 to 25 years, when the labor market has probably already observed some of their ability. In the second strategy we use the panel structure of our data to account for individual fixed effects.

We estimate wage gains for the diploma buyers that persist over a 6-year horizon, the window for which our data is most representative. These gains increase in the initial years after the diploma is bought and show apparent stabilization in 5% to 8% depending on the specification. These results are consistent with the signaling theory, where higher ability workers signal their ability through the diplomas, and therefore these effects persist over time. Our additional investigation also show evidence of essential heterogeneity between the group of workers that buy diplomas and the control group. Alternatively, the results are also consistent with the hypothesis of asymmetric learning of worker productivity between the employer and other firms. In fact, Kahn (2013) estimates that outside firms reduce the average expectation error over worker ability by only a third of the reduction made by incumbent firms. In a context of high worker turnover like in Brazilian labor markets, this could imply diploma effects that decrease very slowly.

This work mostly contributes to two strands of the literature. First, we provide evidence on the signaling value of diplomas, for which recent research using quasi-experimental evidence has show inconclusive results. In particular, two papers with similar setting show opposite results. Using a discontinuity approach to compare students who barely passed or failed high school exit exam in Texas, Clark and Martorelli (2014) find no evidence of diploma effects on earnings over a 11-year span period. Using a similar regression discontinuity strategy, Mazrekaj and Cabus (2019) find that students that barely passed an standardized high school exit exams in Netherlands earn 4% more than their counterparts that barely failed. As every study, both have their limitations. Clark and Martorelli (2014) use only students who have failed all previous exams and had their final chance to get a high school diploma in the 12th grade, resulting in a sample representing a small proportion (4.83%) of total students (In Texas, the students take exit exams for the first time in 10th grade and can retake them several times before 12th grade). Therefore, their study is focused on the tail of low-ability individuals. Mazrekaj and Cabus (2019), on their side, can only calculate short-run effects (up to two and three years after high school graduation) and are unable to include vocational students, which they highlight are more prone to be high school droupouts. Our paper studies diploma effects on high school dropouts in a developing country in a context that over one-third students drop out of school before having a diploma.

Our work also relates to the literature on employer learning and statistical discrimination (EL-SD), which flourished after Altonji and Perret (2001). These studies consider that employers use easily observable characteristic, such as years of education or credentials and diplomas, to infer workers' unobserv-

able productivity and, as more observations related to workers' productivity become available, employers increasingly distinguish workers based on their real productivity and not on these credentials. Therefore, the weight on these observables drops as workers become more experienced. This is precisely what Altonji and Perret (2001) find: as years pass during an individual's career, wages are less correlated to education, which is interpreted as evidence of employer learning. Revisiting their study, Lange (2007) adopts some additional hypothesis to calculate the speed of employer learning, concluding that employers learn quickly: after three and five years, initial errors made by the employers decline 49% and 64%, respectively. Using a more robust RD approach and data from Chilean university graduates, Bordon and Braga (2017) estimate that employers statistically discriminate workers by their university prestige, rewarding a 19% wage premium, but these effects drop 3% per year, consistent with the EL-SD mechanism. However, as cited above, the speed of learning could be much higher for the employer itself than for outside firms, as Kahn (2013) has suggested. Given high turnout rates in the context of our analysis, our results therefore could be interpreted as evidence of very slow learning by the market that the fake diploma was not a signal of higher productivity.

This chapter is organized as follows. After this introduction, we explain the institutional environment and diploma selling scheme. The next section describes data collection, matching, sample construction and descriptive statistics. The fourth section explains our empirical strategy, and the two following chapters present the results and discuss possible mechanisms. The final section concludes.

1.2 Context

1.2.1 Institutional Environment

In Brazil, primary and high school education is an universal public policy. Since 1988's Constitution, the country has increasingly put resources to include every children and adolescent in school. However, the relative success in expanding access has not been followed by an increase in education quality. This helps explain the persistence of high levels of dropout rates: in 2015, 40% of 17-year-old were out of school. In this context, the number of labor force participants who do not own a high school diploma is very large not only because of the historical debt related to periods before universalization of access but also because of the inflow of dropouts. Therefore, Adult High School is very large: in 2020, the country had 3 million students enrolled in AHS (against 1.9 million students enrolled in the last year of regular high school).

The provision of high school education is responsibility of states, and the regulation balances some standards defined by federal legislation and some degree of state variation. In order to be authorized, private high schools have to comply with some requirements, such as to have a legally authorized

school coordinator, some teachers, a physical structure and a few pedagogical and teaching plans. In the case of remote education, there is a requirement that at least 20% of classes to be in the classroom. Once authorized by the state department, the school is formally able to issue diplomas and, by legal requirement, the relation of high school graduates for each school is published in the public official press, called the Diários Oficiais.

Recent scandals have unveiled large schemes of AHS that irregularly emit diplomas for students without complying with the necessary academic requirements. In practice, these schools are basically selling diplomas for people who did not really study in their establishments. Next section describes how this works.

1.2.2

Diplomas selling scheme

In 2018, authorities unveiled a large scandal, allegedly consisting of 350 thousand diplomas sold during five years¹. The investigation, conducted by Rio de Janeiro Civil Police Department in collaboration with the state's education department, was named Operação Nota Zero (Operation Grade F) and collected information on eleven suspected schools. With the collaboration of the Education Department of Rio de Janeiro State (to which we are extremely grateful), we have been able to understand how the market for fake diplomas works.

On the demand side, there were large number of high school dropouts, as we have seen above. On the supply side, many schools authorized to issue diplomas were in fact poorly inspected and had low accountability on whether the number of diplomas issued were coherent to their structure. In the middle, there were "market markers" that matched demand and supply.

A quick google search for the term "como comprar diploma de ensino médio" (how to buy a high school diploma) shows a number of these "market makers" or sellers of diplomas. In one of these site, as in many others,² the message is explicit:

*"If you don't have time and did not finish your studies, you can easily and quickly get the documentation with us. We have been working for years long in the business, issuing original documents certified by the Ministry of Education"*³.

The process to buy a diploma is allegedly simple: individuals contact these merchants, send some personal information, and make a payment. This personal information is then sent to one school that participates in the scheme, and in a some days, the diploma is issued. We have sent an email to one of these sites. In response, we have received the confirmation of the offer and some additional details: the high school diploma cost R\$600 (around 120 USD) and

¹<https://oglobo.globo.com/rio/operacao-enquadra-escolas-que-emitiram-350-mil-diplomas-falsos-movimentaram-700-milhoes-em-5-anos-23096704>

²<https://www.historicoescolarquente.com.br>

³Free translation from the text: "Para você que não dispõe de tempo para estudar e não conclui os estudos, adquira de forma rápida e fácil sua documentação de procedência conosco. Trabalhamos a anos no ramo somente com documentação original, todos os documentos são reconhecidos pelo MEC (...)"

would be issued in five days. The appendix shows the email exchanged and the offer made.

As we have already seen, the diploma issuers were legally authorized private schools. As per Rio de Janeiro's education department, there were typically two cases. Some schools were created specifically to participate in the diploma selling schemes. Other schools were already regular functioning schools and at, some point, they accepted entering the scheme. In both cases, schools are paid for each diploma, for which the marginal cost of issuing is virtually zero.

Although many of these schemes were in place in the country, Rio de Janeiro state was a hub that concentrated a large part, and the participating schools used to issue diplomas for individuals out of the state. As per RJ's education department, the eleven schools listed in Operação Nota Zero had their market of diploma buyers focused on four states: Rio de Janeiro, São Paulo, Paraná and Mato Grosso do Sul.

Also, by legal requirement, the relation of students the had an issued diploma is publicized in the public official press, called the Diários Oficiais. This will be our main data source, as explained in the next section.

1.3 Data

In this work, we use two data sources: a list of high school graduates and the *Relação Anual de Informações Sociais*, RAIS, which contains every contract registered in the Brazilian formal labor market.

1.3.1 Students lists

In order to obtain the lists of students with diplomas issued by the schools participating in the scheme, we have built a program to make a webscraping in one of the websites that compile all the publications of the DOERJ - the *Diário Oficial* of the Rio de Janeiro state official press ⁴. It is important to notice that there is an official template for the text that a school uses for the publication of their students lists. Nevertheless, the publications show a number of small variations around this template. Therefore, we have inspected all pages containing key expressions commonly used in the publications, from 2008 to 2015, and extracted all the publications with some variations around the official template.

The resulting list consists of 575,774 high school graduates (in schools registered in Rio de Janeiro state), the year of graduation, and information on the school of graduation ⁵. Of these, 82,734 studied in schools classified by SEEDUC-RJ as participating in the scheme and exposed in Operação Nota Zero.

⁴www.escavador.com/doerj

⁵In many cases, the extraction mixed in the same cell the name of the school and some other information, and some extra cleaning was necessary

1.3.2 RAIS

Relação Anual de Informações Sociais (RAIS) is a restricted-access matched employee-firm administrative dataset, with information that the Brazilian Ministry of Labor and Employment requests to firms annually. RAIS has information of all formally employed workers in Brazil, and includes detailed information on the employee (social security number, gender, education), on the employer (tax number, sector of activity, establishment size, geographical location), and on the employment relationship (wage, tenure, hiring date, layoff date and reason, etc). This study uses data from RAIS for the period from 2006 - 2017, considering employment contracts from private firms and workers in the CLT rules (not temporary workers). The sample does not consider employment contracts from the public sector, from agriculture and livestock sector and workers with an average wage equal to zero or missing.

1.3.3 Sample construction

We construct our sample in two steps: (i) identifying high school graduates in RAIS and (ii) building our control sample. Ideally, we would like to match by an individual unique identifier such as the CPF number; however, we lack this data in our students list. Therefore, we match our list of students by name with the yearly registers in RAIS. Our largest concern here is to minimize 'type I' error or a false positive, that is, identifying an individual as a diploma buyer when he was actually an homonym of the individual on our list. With that in mind, we apply the following steps. First we filter for: individuals in the four states cited above as the main market for these schools and aged from 17 to 39 years old⁶⁷. Then, we drop all names that had more than one match in RAIS, keeping only unique matches. Therefore, a common name in our students list will most likely have more than one match in RAIS and will be excluded. Even keeping only unique matches, given the large informal labor market, there is still the concern that the match is a false positive with a non negligible probability. Therefore, as a final step, we use an index of rarity of names built from RAIS and keep only the 50% rarest names⁸.

After both procedures our list of diploma buyers amounts to 10,008 individuals. After dropping outliers on wages ⁹, we have 9,676 individuals that we use in our analysis. Next subsection shows some characteristics of these individuals.

In our control sample, we would like individuals with potentially the same education of our treated sample, that is, no high school completed. Analysing the education level of each individual in RAIS, we find that individual education level trajectories sometimes show inconsistent variations,

⁶We have anecdotal evidence that the vast majority of these high school diplomas were issued for people under this age, and descriptive statistics also suggest that, as it will be show in the following section.

⁷we have matched without this geographical restriction, final dataset is similar and results are insensitive.

⁸Description of the rarity index in the appendix

⁹1% lower and upper tail of hire wages and real wages

for example, a given individual might be classified as primary school for some years, then high school for a few years, and then back to primary school for the following years ¹⁰. Therefore, we calculate the proportion of years that an individual is classified as high school and include only those that have less than a certain threshold ¹¹. In the end, we take a 1% random sample and use it as a control group.

After these procedures, we have a treated sample of individuals that bought diplomas and a control sample in order to perform our analysis. Bellow we show some descriptive statistics to make more tangible the characteristics of the individuals that bought diplomas.

1.3.4 Descriptive statistics of diploma buyers

After these procedures, we have a treated sample of individuals that bought diplomas and a control sample in order to perform our analysis. Bellow we show some descriptive statistics to make more tangible the characteristics of the individuals that bought diplomas and make que preliminary comparison with the random sample of workers in RAIS.

From Table 1.1 we see that 52% of the diploma buyers have worked in Rio de Janeiro in any year of the sample, followed by São Paulo with 45% ¹². Figure 1.1 shows that our sample has a concentration of fake diplomas issued between the years 2008 and 2012, and there is a high concentration of diplomas bought when individuals were aged from 20 to 30 with a steady following this age. Figure 1.2(a) shows the temporal trend in the share of diploma buyers that were classified as finished high school for the diploma buyers and the random sample in RAIS. As we can see, in the first years of our sample the diploma buyers have around 10 percentage points lower share, and in the end of the sample around 20 p.p. higher share than the random sample. Indeed, Figure 1.2(b) shows the evolution of this share around the treatment: in rough numbers, from 50% to 80% in a few years before and after the diploma¹³. Specifically, this share takes a leap from 57% one year before diploma acquisition to 75% one year after diploma acquisition, showing that firms discontinuously raise the educational classification of these individuals in RAIS right after diploma acquisition. The next pairs of figures shows the same analysis for the wages. In Figure 1.3(a), we see that the evolution of wages throughout the years in higher for the treated individuals, and apparently these wages accelerate after the diploma acquisition.

¹⁰In RAIS, the education level of each individual is informed by the employer in each yearly register.

¹¹Even though we take this precaution, results are little sensitive to changes in this filter.

¹²Our tabulation counts each time an individual is observed in a different state, regardless of the year. This is why the column sums more than the number of treated individuals.

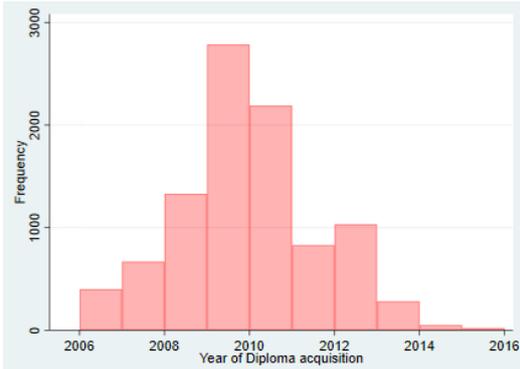
¹³These numbers also show the imprecise characteristic of the education variable in RAIS, as more than 40% of the individuals who will eventually buy a diploma were already classified as high school level.

Table 1.1: States

RJ	5007	52.23
SP	4330	45.17
PR	1071	11.17
MS	118	1.23
Observations	9,587	

Figure 1.1: Distribution of Diploma Acquisitions

(a) By years



By ages

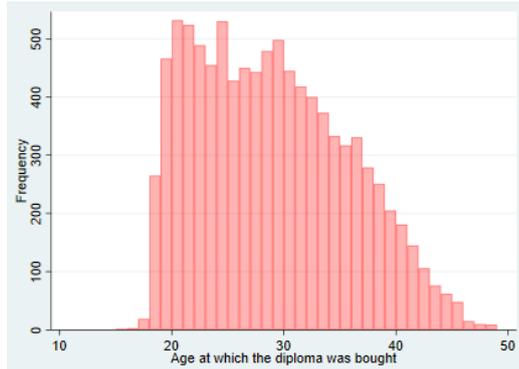
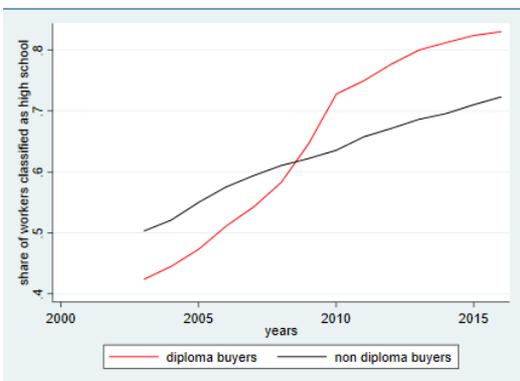


Figure 1.2: Evolution of the share of workers with high school

(a) By years



Around treatment

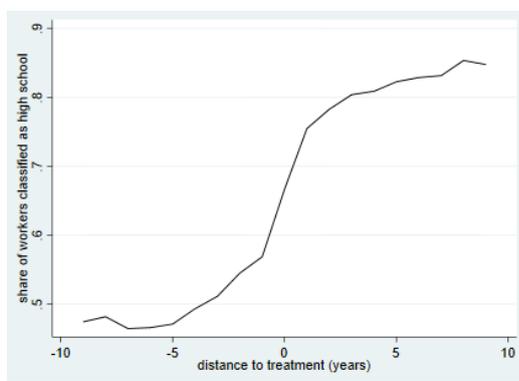
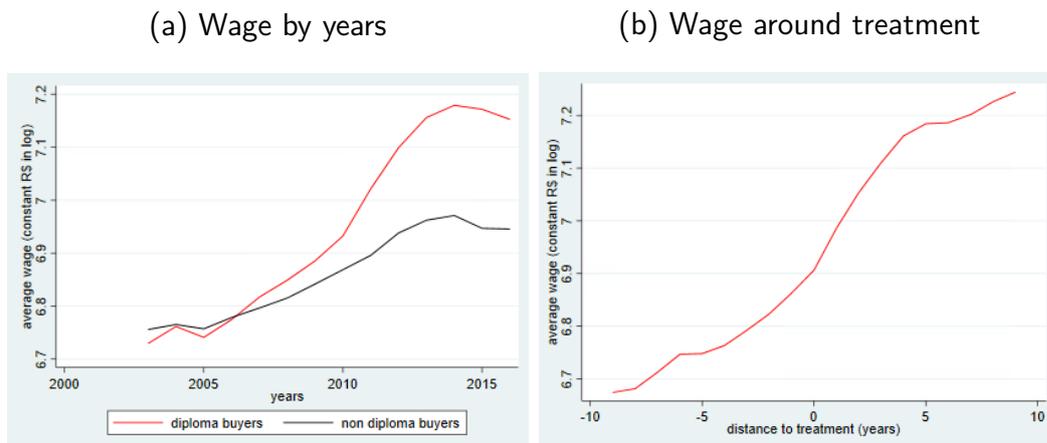


Figure 1.3: Evolution of wages



Source: RAIS

Earlier versions of this work used data received directly from SEEDUC-RJ, consisting of more than 40 thousand students whose diplomas were considered illegal and proposed to be canceled by the technical authorities. These registries were heterogeneous and showed a large variation on the information available, including details on characteristics of schools, courses and student, some of which might certainly be valuable for future extensions of this work. However, many of these records were disorganized and lacked easy-to-access information on the year that the diploma was issued, which is crucial to our identification strategy. Therefore, we proceeded this work using the data constructed above.

1.4 Empirical Strategy

1.4.1 Rationale and underlying model

Our empirical strategy builds upon some assumptions about how the labor market works and uses the data we have collected to design an strategy that can consistently estimate diploma effects.

We consider a model of incomplete information between workers and firms in the labor market. Worker ability, unobservable to the firm, is an innate characteristic that translates into both an individual fixed productivity for a given level of schooling and also into a lower cost of acquiring education. Schooling therefore potentially serves both for raising worker productivity and also for signaling worker innate ability.

At each period, firms observe an imperfect signal of worker productivity and Bayesian update their beliefs, and offer each worker wages equal to their expected productivity.

In this setting, when a firm meets a worker with a diploma, it infers that she might have a higher productivity than an individual without a diploma for two reasons: first, as schooling raises productivity, the expected productivity of

workers with diploma is higher; second, as the cost of schooling is lower for high ability individuals, they use it as a signaling device (as in Spence (1973)) for their higher innate productivity. Therefore, in equilibrium, having a diploma could be associated to higher earnings not only because of accumulation of human capital, but also because firms statistically discriminate workers with diploma, inferring they have larger innate ability than workers without diploma.

In this paper, we are interested in isolating the signaling effect of education. To measure the existence and extent of these signaling effects, we would like to compare two groups that differ only on having a diploma; or (1) workers without high school that bought a diploma – our “treated” group – versus (2) similar workers without a high school that didn’t buy a diploma. Our next subsection explains how we perform this comparison.

1.4.2

Estimation strategy

The obvious challenge to estimate diploma effects is that the decision of buying a diploma is naturally correlated to other characteristics, both observed and unobserved in the data we have. As these characteristics are themselves correlated to individual’s productivity and wages, a naive regression of wages on diplomas would result in biased coefficients and would not tell us real diploma effects. We address this concern with two strategies, both in an event study setting design, allowing for dynamic treatment effects.

In order to control for unobserved ability, we take advantage that we are able to observe individuals’ labor market performance for some years before they buy their diplomas. In fact, most of these individuals buy diplomas in their late 20’s, but have been the labor market since early 20’s or younger.

Our first strategy, then, uses only individuals aged more than 25 years, when market has already observed them some years - and potentially inferred their ability - , and run the following regression:

$$w_{i,t} = \alpha w_{bt} + \lambda_t + \sum_{\substack{l=-6 \\ l \neq -1}}^{l=6} \mu_l T_{i,t}^l + \beta X_{i,t} + v_{i,t}, \quad (1-1)$$

where $w(i, t)$ is our outcome of interest, namely, the log real wages for individual i in year t .

The measure $w(b, t)$ is the mean of the wages each individual received from 23 to 25 years old, always before the acquisition of a diploma. As middle- and high-school individuals typically enter the labor market before 20 years old, by these ages market has already been able to observe some signals of individual performance. Therefore, we expect this variable to capture some of the unobserved ability of each worker. Under the identifying assumption of unconfoundness given this variable, we would could consistently estimate diploma effects. Since we include this control, we restrict our sample to individuals that bought diplomas after they were aged 25 years old and include only observations when they are aged from 25 to 40 years old.

The variables $T(i, t)$ are indicators that equal 1 if treatment occurred in period $t + l$, representing our event study design. Diploma effects are represented by the μ , which are our coefficients of interest. The inequality $l \neq -1$ means that we are considering the baseline period as one year before treatment. We calculate these effects for 6 years before (called *leads*, in the event study literature) and 6 years after the acquisition of the diploma (called *lags*).

Estimated lead coefficients different from zero would raise concerns about the validity of the empirical strategy, as they would indicate that, even under our estimation strategy, control and treatment groups have distinct trajectories pre-treatment. On the other hand, if the lead coefficients show no clear distinction from zero or trend, we could expect that, in the hypothetical absence of the treatment, both groups would continue to have similar trajectories post-treatment. Therefore, the estimated lag coefficients would indicate the causal effects of acquiring a diploma on received wages.

In principle, we do not have priors over the size of these coefficients over time, and allow for complete discretion of each year dummy. If the labor market infers an accumulation of human capital and learns that the individual did not actually acquire this asset, we could expect a declining curve. If, instead, market only sees diplomas as signals of innate ability, then we don't necessarily expect a declining of the estimated coefficients. Moreover, we could have rising coefficients over time, if diplomas are steps for better-performing firms, or if there is learning-by-doing in the jobs these individuals work. That would, indeed, represent a curious phenomena: by fooling the employer about having a diploma, an individual could actually acquire human capital in the new job.

Annual dummies that capture year fixed effects are represented by the λ_t , and $X(i, t)$ are a set of controls that include gender, race, year and age dummies.

As we have different working trajectories, the number of signals emitted by the worker at 23 to 25 years varies substantially. We also have a lot of variation in the age at which diplomas are bought, which translates into variation in the time span from our control measure of wages at 23-25 years and the diploma acquisition ages. Our second strategy takes advantage of the panel feature of our data and estimate individual fixed-effects model is the form of equation bellow. Under this specification, we are able to control for unobservables fixed on individual level over all of the observed working history.

$$w_{i,t} = \alpha_i + \lambda_t + \sum_{\substack{l=-6 \\ l \neq -1}}^{l=6} \mu_l T_{i,t}^l + \beta X_{i,t} + v_{i,t} \quad (1-2)$$

Here we allow for individual fixed-effects, captured by the term $\alpha(i)$ and we omit wages from 23 to 25 years old, since these would be collinear to individual fixed effects. The other terms are exactly those already explained in equation above. Finally, we estimate probability scores of being treated using a probit model in which the probability that individual i acquired a diploma, depends on year, state and age. We then use inverse probability weighting

to estimate the Average Treatment on the Treated (ATT) effect. Indeed, our main parameter of interest is not the effect that diplomas randomly distributed to all the population would have (which would be the Average Treatment Effect - ATE). Instead, we are looking for the effect of diplomas exactly on the population that bought them.

1.5 Main Results

This section presents the main results of the diploma effects estimated in the basic specifications described in the last section. First, Table 1.2 shows the descriptive statistics of our sample, from which we can see that relative to our random sample of workers diploma buyers have a larger share of male and black individuals. Also, the table shows that after making the aforementioned restrictions to build the sample used in the basic regressions, we are left with around 10% of the original sample.

Table 1.2: Descriptive Statistics

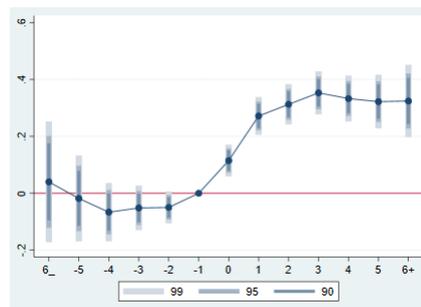
Variable	Full Sample			Basic regression sample		
	Diploma buyers (1)	Other workers (2)	t-test Difference (2)-(1)	Diploma buyers (1)	Other workers (2)	t-test Difference (2)-(1)
Average wage (R\$)	1197.822 [2.905]	1094.807 [0.852]	-103.015***	1338.865 [9.425]	1271.891 [3.297]	-66.974***
male	0.730 [0.002]	0.563 [0.001]	-0.167***	0.796 [0.005]	0.640 [0.002]	-0.156***
black	0.316 [0.002]	0.246 [0.001]	-0.070***	0.366 [0.006]	0.255 [0.002]	-0.111***
age	28.261 [0.023]	27.534 [0.008]	-0.727***	30.128 [0.039]	29.077 [0.014]	-1.051***
schooling (years)	6.593 [0.003]	6.629 [0.001]	0.037***	6.444 [0.011]	6.417 [0.005]	-0.027*
% classified as high school before diploma	0.538 [0.002]	0.645 [0.001]	0.107***	0.216 [0.003]	0.269 [0.001]	0.053***
Obs.	58254	543014		5939	43442	

The value displayed for t-tests are the differences in the means across the groups.
***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

1.5.1 Diploma effects on Schooling

We start by formally evaluating the effects suggested by Figure 1.2: does acquiring a diploma raises the probability that an individual is classified as a high school worker in RAIS? For that, we estimate the fixed-effects model in equation (1.2) placing in the left-side of the equation a dummy variable that equals one if the worker is classified as high school. Figure 1.4 shows a clear impact of finishing school this probability, which jumps 27 percentage points in the first year after the degree and reaches 32 p.p. after five years. This indicates that these workers are indeed increasingly being placed at jobs equivalent to a high school level. The results of the estimation are also shown in the first column of results in Table 1.3.

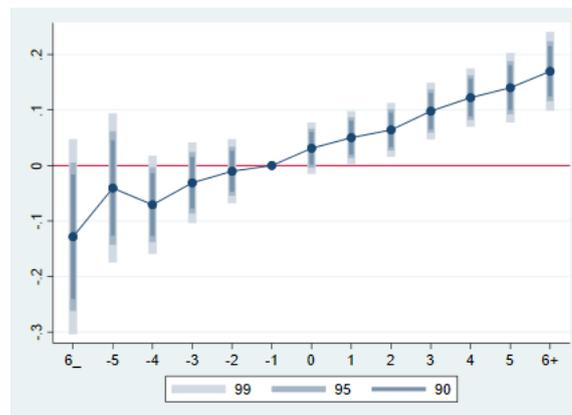
Figure 1.4: Effects on probability of high school job



1.5.2 Diploma Effects on Wages

Once found effects of acquiring a diploma for schooling, in this section we look for evidence of diploma effects on wages. In order to investigate the relevance of our empirical strategy, Figure 1.5 shows diploma effects estimated by equation 1.1 if we omit the control for wages between 23 and 25 years old. Two things call our attention: first, in the years following the acquisition of the diploma, we can see that the estimated coefficients are positive and significant; second, previous to the treatment we also have an apparent upward trend. This reinforces the relevance of the following empirical strategies, which further control for differences between the diploma buyers and the individuals in the control group.

Figure 1.5: Without controlling for wages

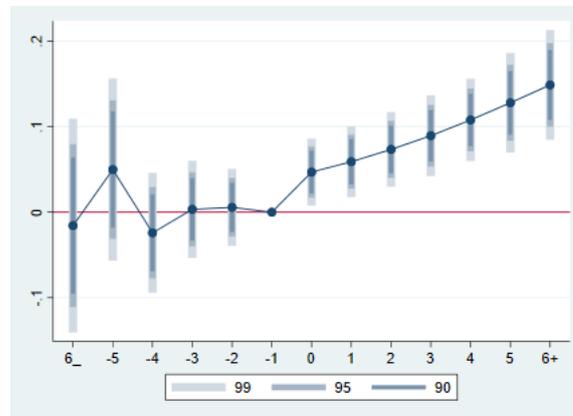


1.5.2.1 Previous wages as controls

Figure 1.6 shows the diploma effects estimated by equation 1.1. As we can see, the coefficients from lead indicators in this specification are all statistically indistinguishable from zero, showing that as we control for previous wages we can see parallel wage trends between individuals that will buy diplomas and those that will not. This raises our confidence that the estimated lag coefficients

can indeed be interpreted as diploma effects. Estimated lag coefficients show an interesting evolution. In the year that the diploma is bought, point estimates indicate a 4.7% wage raise significant at 99% confidence level. As years pass, this wage gap increases to 14.8% six years after diploma acquisition.

Figure 1.6: Controlling for wages



We interpret these results as diploma effects. If not, we would have to consider that there is some sort of unobservable individual characteristic that is relevant to the market and translates into higher wages, but this characteristic only is rewarded by the market after the diploma is bought, and also non-related to the diploma itself. This alternative sounds improbable to us.

Table 1.3: Diploma Effects on reported schooling and wages

	High School	Wages		
		w/o previous wage control	with previous wage control	Fixed-effects model
Mean wages 23-25yrs (log)			0.588 (0.030)***	
<-6	0.040 (0.082)	-0.128 (0.068)*	-0.016 (0.048)	0.032 (0.055)
-5	-0.018 (0.059)	-0.041 (0.052)	0.050 (0.041)	0.067 (0.049)
-4	-0.066 (0.040)*	-0.071 (0.034)**	-0.024 (0.027)	0.012 (0.027)
-3	-0.052 (0.030)*	-0.031 (0.028)	0.003 (0.022)	-0.004 (0.020)
-2	-0.050 (0.022)**	-0.010 (0.022)	0.006 (0.017)	-0.002 (0.014)
baseline	-	-	-	-
0	0.115 (0.022)***	0.031 (0.018)*	0.047 (0.015)***	0.016 (0.013)
1	0.272 (0.026)***	0.050 (0.019)***	0.059 (0.016)***	0.024 (0.015)
2	0.313 (0.028)***	0.064 (0.019)***	0.073 (0.017)***	0.042 (0.018)**
3	0.353 (0.029)***	0.098 (0.020)***	0.089 (0.018)***	0.050 (0.022)**
4	0.333 (0.031)***	0.122 (0.021)***	0.108 (0.019)***	0.079 (0.024)***
5	0.322 (0.037)***	0.140 (0.024)***	0.128 (0.023)***	0.065 (0.030)**
>6	0.324 (0.049)***	0.170 (0.028)***	0.149 (0.025)***	0.110 (0.040)***
male	0.005 (0.063)	0.214 (0.015)***	0.138 (0.013)***	0.066 (0.034)*
black race/color	-0.018 (0.018)	-0.044 (0.013)***	-0.009 (0.011)	0.015 (0.014)
Constant	0.171 (0.087)**	6.613 (0.038)***	2.723 (0.195)***	6.478 (0.068)***
Previous wage as control	No	No	Yes	No
Individual FE	Yes	No	No	Yes
Year FE	Yes	Yes	Yes	Yes
Age dummies	Yes	Yes	Yes	Yes
Observations	38,202	49,381	49,381	38,168
R-squared	0.167	0.123	0.385	0.305
Number of cpf	10,620			10,594

Robust standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

1.5.2.2

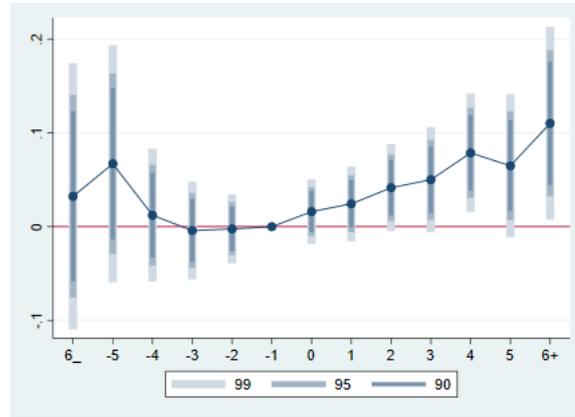
Fixed effects estimation

Figure 1.7 shows the results for the individual fixed-effects estimation. In this specification, we omit previous wages in the list of explanatory variables, as these would be collinear with individual fixed effects. Considering the concerns of the growing literature in event study designs (Borusyak, 2017), we also restrict our sample, dropping years after 2014, in order to minimize concerns about our identification¹⁴. Results show parallel pre-trends before diploma acquisition, comforting us about the validity of the assumptions of

¹⁴We have shown that we have few individuals that bought diplomas after 2014. Although not necessary, a sufficient condition for applying our strategy is to have at least one individual never treated. We also run the regressions including all the years, with similar results.

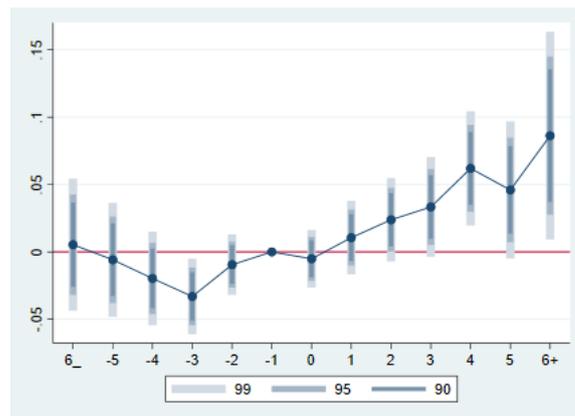
this empirical model. Estimated coefficients after diploma acquisition suggest a positive effect that starts in the year of the diploma acquisition and become significant at the 95% confidence level from year two on. The effects keep increasing up to six years after treatment, achieving 11% in the last year for which we have estimates.

Figure 1.7: Individual FE



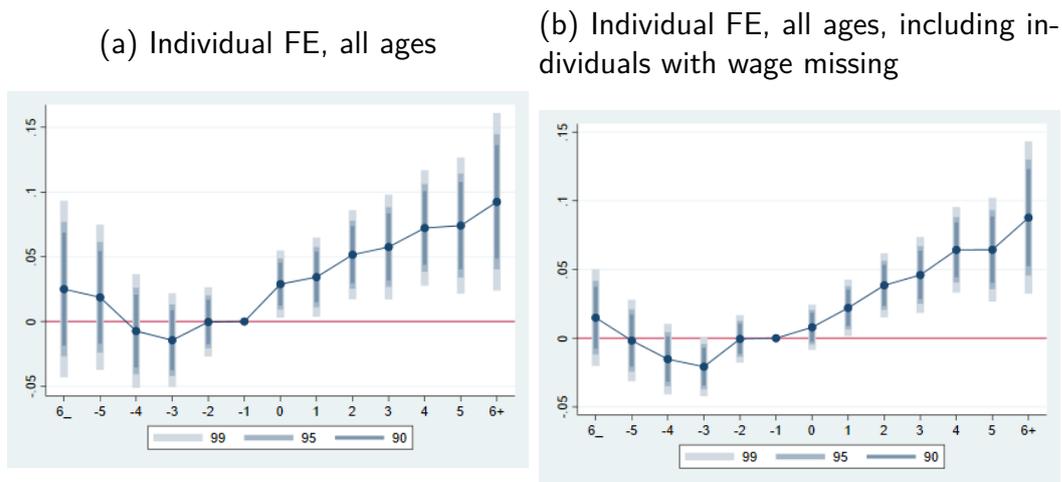
One relative advantage of this strategy is to allow for including individuals that did not participate in the formal labor market from 23 to 25 years old. Figure 1.8 shows results including these observations. Lag coefficients suggest an increasing and significant positive effect, similar to other results using this strategy.

Figure 1.8: Individual FE including individuals with wage 23-25 yrs missing



Finally, this strategy also allow us to drop the restriction of ages. We then run the regressions including all observations of individuals that bought diplomas in any age from 20 to 40 years old. Figures 1.9(a) and 1.9(b) bellow show us the results of these regressions including all ages, respectively keeping and dropping the restriction on missing wages. In both samples, results hold (Table in the Appendix shows the full results).

Figure 1.9: Effects on wages



1.5.3 Is it worth?

The hundreds of thousands of diploma buyers in Brazil suggest that this behavior is not irrational or random. In this section we ask the simple question: at least for these individuals, does it pay off to buy a fake diploma? We answer with simple back of the envelope calculations using the pecuniary cost and our estimated effects.

To calculate the present value and internal rate of return of diploma acquisition, we consider the pecuniary cost of R\$600, a discount rate of 5% per year, mean wage previous to diploma acquisition equal to R\$ 1200 and diploma effects of 8% persisting for 25 years. In this scenario, the present value of a diploma is more than R\$ 10,000, and the internal rate of return of this acquisition is more than 90%.

1.6 Discussion

As we have seen, results show a persistent effect of the diploma on wages, at least for the period we are able to include in the analysis. In this section we discuss and explore possible interpretations of these results. We highlight three possibilities: (i) signaling of a true individual ability previously not rewarded by the market, (ii) pretending to have a different productivity than the real one and succeeding in fooling the market due to asymmetric employer learning and high turnout rates, and (iii) diplomas opening doors to better jobs and leading to learning-by-doing by the worker, that actually raises their productivity in the new job.

The first case would be possible if workers that buy diplomas are potentially more productive than the control group (the group of workers that neither finished high school nor bought diplomas), but only a part of their ability is captured by either the previous wages variable or by the individual fixed effects. Then, these high ability individuals might signal their ability with diplomas and wage effects could persist over time. We develop some preliminary

investigation of this explanations. If signaling a (previously) hidden ability is the relevant mechanism, we might expect effects being lower for those who have already been many times at labor market. We test this prediction with two strategies. The first strategy is to look at older ages, for which we could expect a lower signaling effect of the diploma. We run regression for a sample including only individuals who bought their diplomas after 30 years old. Interestingly, results in Figure 1.10 still show a positive effect, although lower in the initial years. We then construct two samples, one with individuals who were observed more than 3 years in the labor market and the other with individuals who were observed 3 or less years, and run separate regressions. Figure 1.10 bellow shows that the point estimates have similar trajectories, but for the sample with higher previous experience in the labor market, we are unable to reject null effect given higher standard deviations.

Figure 1.10: Individual FE, with missing wagers, only 30+ years old

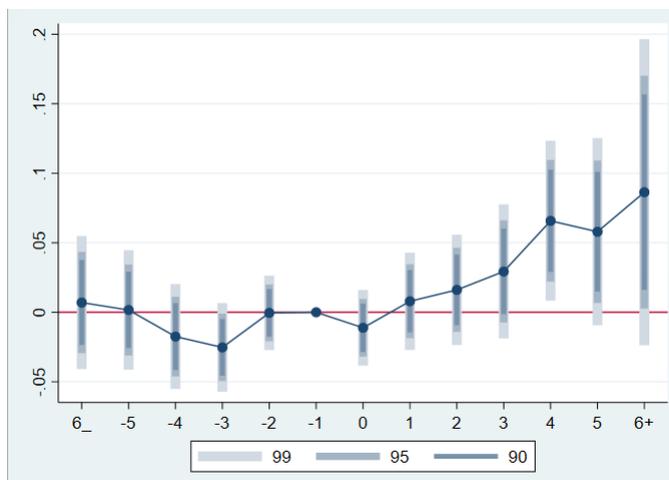
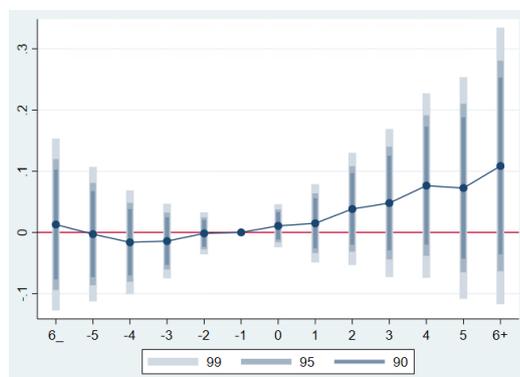
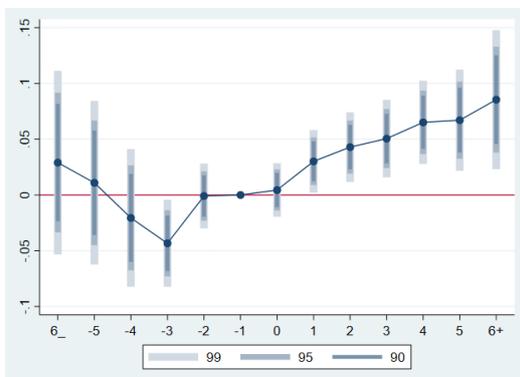


Figure 1.11: Low and high previous experience

(a) FE, with missing, lower previous experience

(b) Individual FE, all ages, higher previous experience



One might wonder why low ability individuals could not also buy diplomas and profit from this effect - something we could call fake signaling. Indeed, if low individuals do buy diplomas and under the assumption of

employer learning through time, we should see a positive effect on earnings of these individuals during some years gradually fading away as time passes. In these cases, even though the value of buying a diploma is higher for a high ability individual, we might still see low ability individuals buying diplomas. In fact, we have seen that in some specifications we have some evidence pointing to diploma effects smaller in magnitude and larger in variance after 6 or more years have passed. This would be consistent with most diploma buyers being high-ability individuals, but some are low ability that fool the market for some time. In the Appendix, we show the results of a simple test that investigates the possibility of essential heterogeneity, which is exactly the case when the treated individuals self-select into the treatment based (at least partly) in their private knowledge of their idiosyncratic gain which is essentially different from those non-treated. We split the sample between individuals with high pscores and low pscores, and separately estimate the fixed-effect model. This results in estimated positive diploma effects concentrated in individuals with high pscores, and, conversely, individuals with low pscores show no apparent gain, as would be expected in the case of selection into treatment.

The second and third alternatives explanation to the signaling of a hidden ability are asymmetric employer learning and learning by doing. In both cases, workers that buy diplomas are no different from workers that don't, but firms demand a diploma for hiring in better positions. If learning the true productivity of the worker is easier for the employer than for the other firms in the market, a worker might be able to be paid more than their productivity for many periods, specially through changing jobs. In order to test this, we plan to analyze how job changes interact with the wage trajectories of the workers. The last alternative would be the case if, after the worker is hired, they accumulate human capital, effectively raising their productivity. In further developments of the research, we will be able to test this by analyzing how the wage trajectories depend on the gap between worker's hire wage and mean firm wages, as for the larger this gap, the higher the space for learning-by-doing.

1.7

Conclusion

Since the formulation of the human capital theory and the signaling theory, the literature has struggled to differentiate empirically whether education increases the productivity of individuals or simply signals workers productivity to the firms in a scenario of incomplete information, as both theories imply a positive relationship between education and wages. The recent literature has tried to evaluate signaling effects by comparing individuals with arguably the same levels of human capital, with and without a high school diploma (Clark and Martorell, 2014, Mazrekaj and Cabus, 2019), with inconclusive results.

In this paper, we contribute to this literature by investigating a still largely unexplored phenomenon in Brazil: schools that issue diplomas for individuals who actually did not study. As a mere piece of paper, diplomas should not increase the productivity of any worker if the sole role of education is to increase the human capital of individuals. However, if diplomas are used as a steppingstone to get into better jobs or employers believe that diplomas are a signal of workers productivity, workers that buy a diploma would see an

increase in wages after obtaining the degree. Moreover, the dynamics of this increase over time could inform us about the learning by the employer of the true productivity of these workers.

In order to test these predictions, we build a novel dataset with tens of thousands of students who bought fake diplomas in schools scrutinized by administrative and criminal investigations after reports of school fraud. Most of the diplomas were issued for individuals over 19 years old and corresponded to Adult High School (AHS, or EJA in Brazil) diplomas. We then identify these individuals in a matched employer-employee data on formal firms in Brazil. Using wages before the acquisition of the diploma and individual fixed-effects as to control for individuals' unobserved productivity, we estimate wage gains for the diploma buyers that persist over a 6-year horizon, the window for which our data is most representative. These gains increase in the initial years after the diploma is bought and show apparent stabilization in 5% to 8% depending on the specification. These results are consistent with predictions from the signaling theory, where employers are only partly able to observe workers' ability. In this context, workers might use diplomas to signal ability and have wage gains that persistent over time. As our work estimates long term diploma effects on wages of school dropouts that represent over one-third of students population, we see our results as complementary to the literature that focused on the very low tail of ability distribution (Clark and Martorell, 2014) or estimated only short term diploma effects (Mazrekaj and Cabus, 2019).

However, signaling of a previously unobserved ability is not the only possible mechanism driving our results. The Employer Learning and Statistical Discrimination (EL-SD) literature has shown us that, the slower the learning process of the market about the true productivity of workers is, the longer diplomas could be used used to fool the market and enjoy wage gains. Specially, if the learning is asymmetric between incumbent and outside firms, the worker might be able fool the market for a long time. Therefore, our results could be seen as evidence of asymmetric employer learning, as Brazil has high turnout rates. Finally, at this point we cannot rule out alternative explanations such as learning-by-doing. Further research could clarify which of these mechanisms are most probably driving these diploma effects.

2

Adult High School in Brazil: investigating human capital accumulation

2.1

Introduction

Although education is one of the most important determinants of wages at the individual level, there is far from consensus about the relative importance of the two main channels that might cause the positive correlation between individual productivity, schooling and wages. While the human capital theory argues that education increases productivity of individuals (Becker, 1962 and Mincer, 1974), the signaling theory (Spence, 1973) posits that higher wages reflect the correlation between education and unobserved ability. Lately, there has been a growing body of evidence that both mechanisms play a role: schooling provides human capital accumulation that translates into increasing wages (Arteaga, 2018), but diploma effects (Mazkeraj and Cabus, 2019) are evidence that individuals also signal their productivity through education.

In developing countries such as Brazil, large annual inflows of high school dropouts into the labor force implies that, in light of the human capital theory, it is crucial to offer these individuals the opportunity to finish their studies. As effect, in 2020, around 1.5 million people were enrolled in Adult High School in Brazil (which takes 3 semesters), versus 1.9 million enrolled in the last year of regular high school. The challenge for estimating labor market effects driven by human capital accumulation is enhanced, however, by the evidence of diploma effects, since the variation in education measures is commonly accompanied by variation in degrees.

This chapter explores the dataset constructed with individuals that bought diplomas in the Brazilian black market, as described in the last chapter, to circumvent this challenge. We investigate two distinct groups of individuals, one composed of people that studied in regular Adult High Schools and the other with people that merely bought diplomas, to search for evidence of human capital effects of AHS in Brazil. Bearing in mind the diploma effects found in the previous chapter, we compare the labor market performance of individuals that finished high school (and got their regular diploma) to those who bought a high school diploma, thus testing if the human capital acquired at high school would confer any labor market advantage on the job, controlling for the diploma status. We also take advantage of the panel structure of the matched employer-employee data of RAIS to account for unobserved productivity before the diploma acquisition. Our results show no evidence of wage gains of studying in an AHS versus simply buying a diploma, which is consistent with no accumulation of productive human capital. We also compare real AHS graduates and individuals that did not finish high school. Results suggest that the benefit of graduating in a regular AHS is similar to that of simply buying a diploma.

Our contribution is twofold. First, we contribute to the literature of the

relative effects of human capital versus signaling, revealing a case where most, if not all, of the wage effects of completing a cycle of study comes from the effects of obtaining the degree itself, and not from human capital accumulation. A number of papers have investigated either human capital effects or diploma effects in primary and secondary education settings and obtained mixed results. Exploring the introduction of one extra year in primary school in China in 1980, Eble and Hu (2016) find a 2% increase in wages. As just a small fraction of the total return to schooling, the authors infer an important role for signaling in primary education. In the context of tertiary education, Arteaga (2018) exploits the reduction of classes for obtaining a bachelor's degree in Economics and in Business in Colombia as to isolate the human capital from diploma effects, finding an estimate of 12%. Our estimates point to no human capital effects on earnings for the case of AHS in Brazil.

Second, we also contribute to the small literature on the relevant subject of human capital accumulation attained by students in Adult High School, which reaches magnitudes of one-third of regular high school. Blundell (2020) finds that workers finishing AHS in Norway have higher earnings. We also find higher earnings for Brazil, but our results suggest that these are driven by diploma effects. To the best of our knowledge, this is the sole work applying the approach of disentangling human capital from signaling effects for Adult High Schools.

This chapter is organized as follows. After this introduction, we briefly recall the institutional environment and the diploma selling scheme, already described in Chapter one. The third section describes the data, and the next section performs a simple test on the impacts of finishing an AHS on the labor market performance of individuals. In the fifth section we develop our main analysis, investigating the evidence of human capital effects by comparing the performance of real AHS graduates with individuals that bought high school diplomas and did not really finished their high school studies. Our conclusion is that the effects of finishing an AHS on earnings apparently come from a diploma effect, and not from human capital accumulation. The final section concludes.

2.2

Institutional Environment

2.2.1

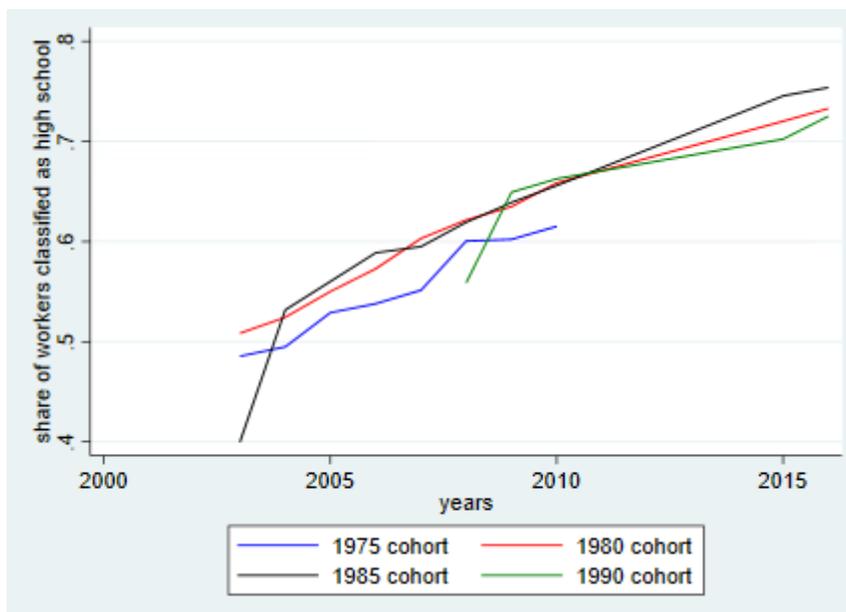
Adult High Schools in Brazil

In the last decades, Brazil has successfully expanded access to primary school, but has struggled to raise education quality and to keep students on track: in 2020, almost 40% of students 18 and 19-year-old had not finished high school. As in many developing countries, the option of many to obtaining a high school diploma is to study in an Adult High School, and the the large size of AHS is partly responsible for the growth of the proportion of the labor force in the country that holds a secondary education degree, which raised from half in 2004 to 70% in 2016¹. In fact, the number students enrolled in AHS has

¹Reis, 2017.

grown steadily in the 2000's, amounting to 1.5 million people in 2020, against 1.9 million students enrolled in the last year of regular high school². For the sake of comparison, in the United States, the number of high school dropouts each year is about 1.2 million and the number of high school graduates is 3.7 million (meaning that around 25% of high school students fail to graduate from high school on time). Of those 1.2 million, around 0.7 million take the GED test, and 73% received the GED certification³. As such, the proportion of young adults without a high school diploma is five times larger in Brazil. The Figure 2.1 below shows the shares of formal workers classified as high school by birth cohorts, from which we can see that workers keep raising their schooling during their working life.

Figure 2.1: Schooling evolution by cohorts



Source: RAIS

In Brazil, the provision of high school education is responsibility of states, and the regulation balances some standards defined by federal legislation and some degree of state variation. In order to be authorized, private high schools have to comply with some requirements⁴. As explained in the last chapter, in the context of thriving demand for high school diplomas by adults and loose supervision of regulatory standards, a black market for fake diplomas has flourished in the last years. In practice, some schools basically sold diplomas for people who did not really study in their establishments. Next section describes how this works.

²Source: tabulations of Censo Escolar, available at www.qedu.com.br.

³Jenpsen, 2016.

⁴Such as to have a legally authorized school coordinator, some teachers, a physical structure and a few pedagogical and teaching plans. In the case of remote education, there is a requirement that at least 20% of classes to be in the classroom.

2.2.2

Diplomas Selling Scheme

In this section we briefly recall the scheme described in the first chapter of this thesis. In 2018, authorities unveiled a large scandal, allegedly consisting of 350 thousand diplomas sold during five years⁵. The schools that issued fake diplomas issuers were legally authorized private schools. The process to buy a diploma is allegedly simple: individuals contact these schools or their sales representative, send some personal information, make a payment, and in a few days the school issues a fake diploma, at the cost of R\$600 (around 120 USD). Although many of these schemes were in place in the country, Rio de Janeiro state was a hub that concentrated a large part, issuing diplomas mostly for individuals living in four states: Rio de Janeiro, São Paulo, Paraná and Mato Grosso do Sul. Finally, the schools also published the lists of names of students that had their high school diploma issued in the Official Journal of Rio de Janeiro State, which we will use as a primary source of data.

2.3

Data

In this work, we use two data sources: a list of high school graduates (both those who studied in regular AHS and those that bought fake diplomas) and the Relação Anual de Informações Sociais, RAIS, which contains every contract registered in the brazilian formal labor market.

2.3.1

Students lists

Our first data source is a list of students graduated from AHS schools in Rio de Janeiro state, which we have built in the same way described in the first chapter of this thesis. Our webscrapping program has inspected all pages containing key expressions commonly used in the publications in the Official Journal of RJ state, from 2008 to 2015. The resulting list consists of 575,774 high school graduates (in schools registered in Rio de Janeiro state), the year of graduation, and the school of graduation. Of these, 82,734 studied in schools classified by SEEDUC-RJ as participating in the scheme and exposed in Operação Nota Zero⁶.

⁵<https://oglobo.globo.com/rio/operacao-enquadra-escolas-que-emitiram-350-mil-diplomas-falsos-movimentaram-700-milhoes-em-5-anos-23096704>

⁶We have written a program to make a webscrapping in one of the websites that compile all the publications of the DOERJ - the Diario Oficial of the Rio de Janeiro state official press - www.escavador.com/doerj. It is important to notice that there is an official template for the text that a school uses for the publication of their students lists. Nevertheless, the publications show a number of small variations around this template. Therefore, we have inspected all pages containing key expressions commonly used in the publications, from 2008 to 2015, and extracted all the publications with some variations around the official template.

2.3.2 RAIS

Relação Anual de Informações Sociais (RAIS) is a restricted-access matched employee-firm administrative dataset, with information that the Brazilian Ministry of Labor and Employment requests to firms annually. RAIS has information of all formally employed workers in Brazil, and includes detailed information on the employee (social security number, gender, education), on the employer (tax number, sector of activity, establishment size, geographical location), and on the employment relationship (wage, tenure and hiring date). This study uses data from RAIS for the period from 2006 - 2017, considering employment contracts from private firms and workers in the CLT rules (not temporary workers). The sample does not consider employment contracts from the public sector, from agriculture and livestock sector and workers with an average wage equal to zero or missing.

2.3.3 Sample construction

In this chapter, we perform two comparisons: (i) real AHS students versus diploma buyers (our human capital test) and (ii) real AHS graduates vs individuals without high school. Therefore, we build our main sample including the three categories (real AHS students, Diploma buyers and those without high school) and for each test we will make the appropriate segregation.

As in Chapter one, as we lack an individual unique identifier, we match our list of students by name with the yearly registers in RAIS. In order to minimize 'type I' error or a false positive (identifying an individual as one present in our list when he was actually an homonym), we apply the same steps already described. First we filter for: individuals aged from 17 to 39 years old and in the four states cited above as the main market for these schools⁷. Then, we drop all names that had more than one match in RAIS, keeping only unique matches. Therefore, a common Brazilian name in our students list will most likely have more than one match in RAIS and will be excluded. Even keeping only unique matches, given the large informal labor market, there is still the concern that the match is a false positive with a non negligible probability. Therefore, as a final step, we use an index of rarity of names built from RAIS and keep only the 50% rarest names⁸. This will give us the labor market observations of both our groups of students, either real AHS graduates or diploma buyers.

For the third category, we would like individuals with no high school completed. As in the first chapter, we will include only individuals classified as non high school graduates in RAIS and take a 1% random sample⁹.

⁷Indeed, a name in our list is either a real AHS student from Rio de Janeiro state or a diploma buyer from (most probably) one of these states

⁸Description of the rarity index in the appendix

⁹Analysing the education level of each individual in RAIS, we find that individual education level trajectories sometimes show inconsistent variations, for example, a given individual might be classified as primary school for some years, then high school for a few years, and then back to primary school for the following years. In RAIS, the education level of each individual is informed by the employer in each yearly register. Therefore, we calculate the proportion of years that

After these procedures, we have our three categories of interest: (i) individuals that studied in real AHS schools, summing 61,048 workers, (ii) individuals that bought diplomas, summing 12,036 people, and (iii) a sample of workers classified as non high school graduates, totalling 130,279 ¹⁰.

2.4

The effects of Adult High School

2.4.1

Empirical Strategy

This section evaluates the effects of AHS graduation, comparing real AHS graduates and workers that never finished high school. We perform an investigation similar to the first chapter of this thesis, with two sets of econometric specifications, both in an event study setting design, in which we allow for dynamic treatment effects. Again, in order to control for unobserved ability, we take advantage that we are able to observe individuals' labor market performance for years before they get their diplomas. Our first strategy uses only individuals aged more than 25 years, and run the following regression.

$$w_{i,t} = \alpha w_{bt} + \lambda_t + \sum_{\substack{l=8 \\ l=-6 \\ l \neq -1}} \gamma_l G_{i,t}^l + \beta X_{i,t} + v_{i,t} \quad (2-1)$$

Where $w(i, t)$ is our outcome of interest, namely, the wage for individual i in year t . The measure $w(b, t)$ is the mean of the wages each individual received from 23 to 25 years old. We restrict our treatment sample for individuals that got their diplomas when they were aged more than 25 years old, so this variable measures mean wages before the graduation in an AHS.

The variables $G_{i,t}^l$ are dummy indicators that equal 1 if the individual got their degree in a (regular) AHS in period $t + l$. The inequality $l \neq -1$ means that we are considering baseline period one year before treatment. We calculate these effects for 6 years before (also called *leads*) and 8 years after the acquisition of the diploma (lags), period for which we have reliable data on. In this empirical model, γ_l represent the (total) effect of finishing studies and getting a diploma in a regular AHS. Finally, annual dummies capturing year fixed effects are represented by the λ_t , and $X(i, t)$ are a set of controls that include gender, race and age dummies. In order for our coefficients to correctly identify our parameters of interest, we would like to observe estimated lead coefficients not to show any noticeable trend, since we need the assumption of parallel trends. Under the validity of our identifying assumptions, the estimated lag coefficients would indicate the causal effects of graduation on wages. Our second strategy estimates individual fixed-effects model in the form of:

an individual is classified as high school and include only those that have less than a certain threshold. Even though we take this precaution, results are little sensitive to changes in this filter.

¹⁰As in Chapter 1, we drop wages outliers, the 1% lower and upper tail of hire wages and real wages

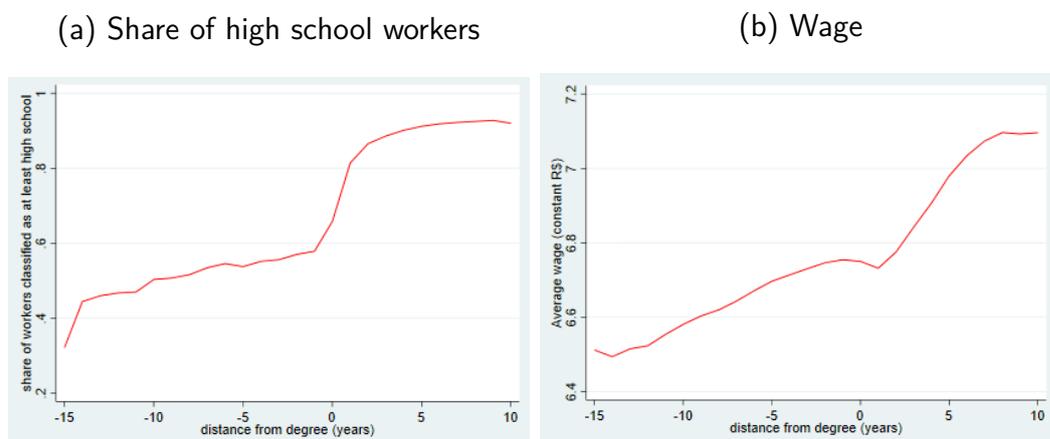
$$w_{i,t} = \alpha_i + \lambda_t + \sum_{\substack{l=-6 \\ l \neq -1}}^{l=8} \gamma_l G_{i,t}^l + \beta X_{it} + v_{i,t} \quad (2-2)$$

Here, we omit wages from 23 to 25 years old (in order to avoid collinearity) and allow for individual fixed-effects, captured by the term $\alpha(i)$, while the other terms are the same as in equation 2.1 above. Finally, we estimate probability scores of being treated using a probit model in which the probability that the individual graduated in an AHS depends on year, state and age. In all regressions, we use this probability to estimate the Average Treatment on the Treated (ATT) effect.

2.4.2 Descriptive Analysis

Before our regression analysis, it is useful to compare the evolution of our treated and control groups. As expected, Figure 2.2(a) shows that right after graduation workers raise their probability to be classified as high school by around 30 percentage points - only 10% are not classified as such after a few years. Coherently, their wage curve accelerates in the years after getting their degree (Figure 2.2(b)s). Comparing the evolution through the years in the left graph of Figure 2.3 shows that our treated group starts in disadvantage in terms of wages but catch up in the last years, when most have already finished the adult high school. In fact, the right graph shows that, for each cohort of AHS graduates, mean wages raise after their respective year of graduation, each cohort contributing to the aggregate increase in the left graph. The Table 2.1 shoes some descriptive statistics for each group, both for the full original sample and for the sample after the restrictions made for applying our regression analysis.

Figure 2.2: Evolution of schooling and wages around AHS graduation

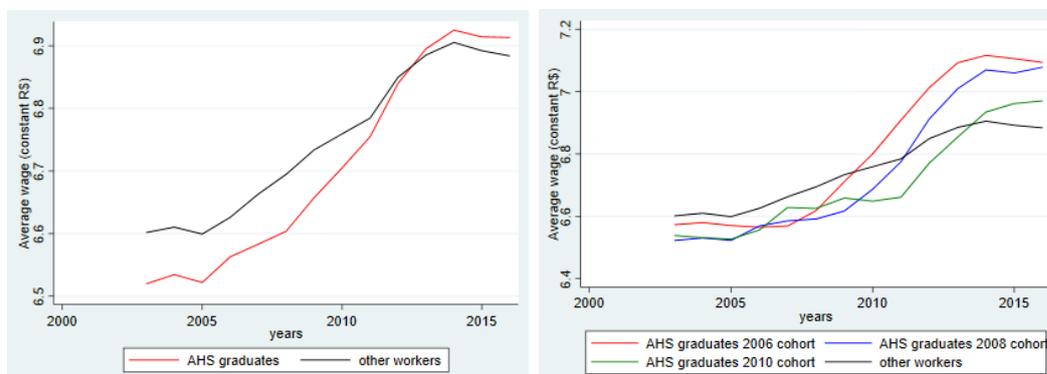


Source: RAIS

Figure 2.3: Evolution of wages: aggregate and cohorts

(a) AHS, by years

(b) AHS, by years - cohorts



Source: RAIS

Table 2.1: Descriptive Statistics: Adult High School graduates and random sample

Variable	Full Sample			Basic regression sample		
	Other workers (1)	AHS grad- uates (2)	t-test Difference (1)-(2)	Other workers (1)	AHS grad- uates (2)	t-test Dif- ference (1)-(2)
Average wage (R\$)	1087.36 [0.828]	1032.77 [1.198]	54.588***	1262.83 [3.246]	1147.84 [5.179]	94.735***
male	0.548 [0.001]	0.577 [0.001]	-0.028***	0.627 [0.002]	0.684 [0.004]	0.050***
black	0.245 [0.001]	0.372 [0.001]	-0.127***	0.255 [0.002]	0.43 [0.004]	-0.047***
age	27.279 [0.008]	25.183 [0.011]	2.096***	29.068 [0.014]	29.805 [0.025]	0.001
schooling (years)	6.644 [0.001]	6.754 [0.002]	-0.110***	6.384 [0.005]	6.248 [0.007]	-0.069***
% classified as high school before diploma	0.641 [0.001]	0.536 [0.001]	0.105***	0.265 [0.001]	0.206 [0.002]	0.153***
Obs.	556734	245953		44162	14238	

The value displayed for t-tests are the differences in the means across the groups.

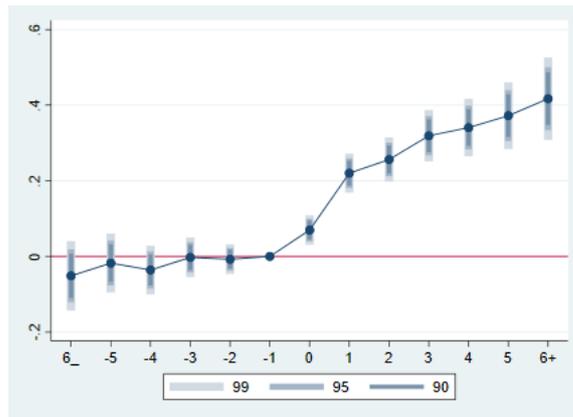
***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

2.4.3 Results

2.4.3.1 AHS Effects on Schooling

We start by formally evaluating the effects suggested by Figure 2.2: does finishing AHS raises the probability that an individual is classified as a high school worker in RAIS? For that, we estimate the fixed-effects model in equation (2.2) placing in the left-side of the equation a dummy variable that equals one if the worker is classified as high school. Figure 2.4 shows a clear impact of finishing school on this probability, which jumps 20 percentage points in the first year after the degree and reaches a 37 p.p. increase after five years. This reassures us that these workers are indeed increasingly being placed at jobs equivalent to their newly acquired high school level. The results of the estimation are in the first column of Table 2.2.

Figure 2.4: Effects of graduating in AHS on schooling



2.4.3.2 AHS Effects on Wages

Next, we evaluate the impacts of a AHS degree on earnings, estimating equation (2.1) with wages in the dependent variable. Figures 2.5 shows that we are unable to observe parallel pre-trends, which suggests that wages at 23 to 25 years are unable to capture all the evolution of the wage difference between these two groups before the graduation in high school. As these differences are upward-sloped, it seems that individuals who will eventually graduate from real AHS have higher wage gains even previous to the AHS graduation¹¹.

Figure 2.5: With control of previous wages

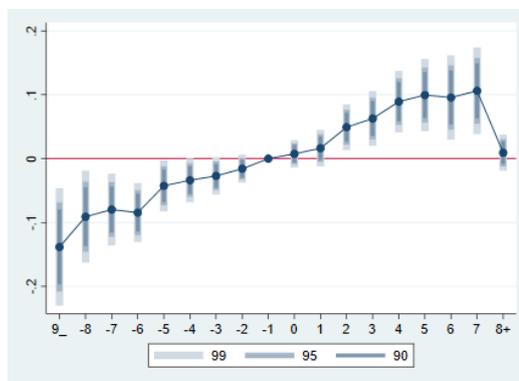


Figure 2.6 then shows results from the individual fixed-effects model. Under this specification, we observe our lead coefficients around zero, consistently with parallel pre-trends between treated and control groups once we account for all the individual wage history before graduating in an AHS. In that case, we interpret the estimated lag coefficients as real treatment effects. The estimates show increasingly positive effects, all of which statistically different from zero at a 95% confidence level after 2 years of graduation, suggesting a 6% wage gain after 4 years of AHS graduation and 13% after 8 years. These

¹¹One possible interpretation is that they are still revealing their ability, and increasingly rewarded, in the years before graduation. Another possibility is that these individuals are somehow acquiring human capital even before going to adult high schools. In any case, further research is planned to address these possibilities.

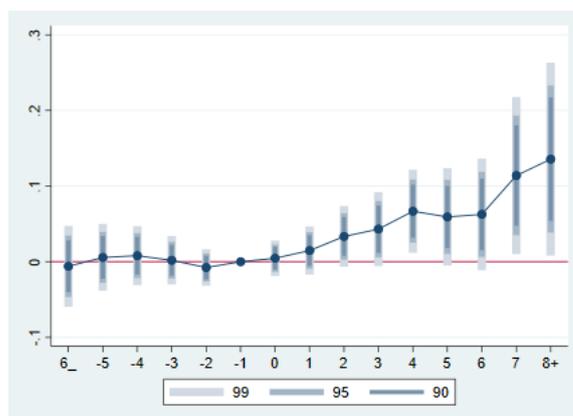
Table 2.2: AHS effects on schooling and wages

	High School	Wages		
		with previous wage control	Fixed-effects model (restricted years)	Fixed-effects model (all years)
Mean wages 23-25yrs (log)		0.461 (0.022)***		
-6	-0.051 (0.036)	-0.084 (0.018)***	-0.006 (0.021)	-0.013 (0.018)
-5	-0.018 (0.030)	-0.043 (0.016)***	0.006 (0.017)	0.001 (0.015)
-4	-0.036 (0.025)	-0.034 (0.013)**	0.008 (0.015)	0.005 (0.013)
-3	-0.003 (0.020)	-0.027 (0.011)**	0.002 (0.012)	-0.006 (0.011)
-2	-0.007 (0.015)	-0.016 (0.009)*	-0.008 (0.009)	-0.005 (0.008)
0	0.070 (0.015)***	0.007 (0.008)	0.005 (0.009)	-0.001 (0.008)
1	0.220 (0.020)***	0.016 (0.011)	0.015 (0.012)	0.006 (0.010)
2	0.256 (0.023)***	0.049 (0.014)***	0.033 (0.016)**	0.026 (0.013)**
3	0.319 (0.026)***	0.063 (0.017)***	0.043 (0.019)**	0.026 (0.016)*
4	0.341 (0.029)***	0.089 (0.019)***	0.067 (0.021)***	0.046 (0.017)***
5	0.372 (0.034)***	0.100 (0.022)***	0.059 (0.025)**	0.048 (0.020)**
6	0.417 (0.042)***	0.096 (0.026)***	0.063 (0.029)**	0.052 (0.023)**
7		0.106 (0.026)***	0.114 (0.040)***	0.048 (0.026)*
8		0.009 (0.011)	0.136 (0.050)***	0.070 (0.032)**
male	0.043 (0.045)	0.206 (0.012)***	-0.009 (0.034)	-0.005 (0.028)
black race/color	0.006 (0.018)	-0.057 (0.010)***	0.011 (0.009)	0.008 (0.008)
Constant		3.599 (0.355)***	6.465 (0.075)***	6.439 (0.053)***
Previous wage as control	No	Yes	No	No
Individual FE	Yes	No	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Age dummies	Yes	Yes	Yes	Yes
Observations	44,527	58,400	44,527	58,400
R-squared	0.097	0.346	0.304	0.283
Number of cpf	12,245		12,245	14,274

Robust standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

results indicate a persistent positive impact of graduating in an AHS on the wage trajectories of these workers.

Figure 2.6: Individual FE



A naive interpretation of the results above section could consider the effects to be human capital effects. However, the literature has already found evidence of diploma effects on earnings, and our chapter 1 has reinforced this possibility. Therefore, the effect finishing an AHS on earnings cannot be credibly attributed to the accumulation of human capital without further analysis. Next section develops a test to isolate human capital effects.

2.5 Disentangling human capital from diploma effects

We start by recalling that, in a context of incomplete information, having a diploma could be associated to higher earnings as for two channels: (i) schooling might raise productivity because of direct human capital accumulation and/or (ii) the diploma itself could have persistent effects. As such, the usual comparison of workers with and without high school would yield estimates that combine two parameters: human capital effects plus diploma effects.

In order to measure pure human capital effects, we would like to compare two groups that differ only on human capital, but with similar diploma achievement status. Luckily we are able to replicate that in our setting, comparing two groups: (1) workers that legitimately got a high school diploma – our treated group, that is, the group with a human capital treatment – versus (2) similar workers that bought a high school diploma - which will be our control group. Our next subsection explains how we do that.

2.5.1 Empirical Strategy

Our strategy to isolate human capital effects from diploma effects relies upon the basic idea used to estimate the effects of AHS graduation, but now not only the treatment group has achieves a diploma status in our panel, but also the control group - which is the workers that bought the diploma. As we have already seen, it might be the case that these diploma buyers enjoy wage increases. The empirical question is, then, whether the wage increase of the workers who graduate in the real AHS is larger than those who just buy diplomas.

Following this strategy, we run two sets of econometric specifications, both in an event study setting design and, again, in order to control for unobserved ability, we take advantage that we are able to observe individuals' labor market performance for years before they get their diplomas - either legally or illegally. Our first strategy uses only individuals aged more than 25 years and run the following regression.

$$w_{i,t} = \alpha w_{bt} + \lambda_t + \sum_{\substack{l=-6 \\ l \neq -1}}^{l=6} \gamma_l G_{i,t}^l + \sum_{\substack{l=-6 \\ l \neq -1}}^{l=6} \mu_l T_{i,t}^l + \beta X_{i,t} + v_{i,t} \quad (2-3)$$

Where $w(i, t)$ is the wage for individual i in year t and the measure $w(b, t)$ is the mean of the wages each individual received from 23 to 25 years old. As

before, we also restrict for individuals that got their diplomas when they were aged more than 25 years old. .

The variables $G_{i,t}^l$ are dummy indicators that equal 1 if the individual graduated from (any) AHS in period $t+l$, and the variables $T_{i,t}^l$ are indicators that equal 1 only if an individual got their degree in a regular AHS in period $t+l$. The inequality $l \neq -1$ means that we are considering baseline period one year before treatment. We calculate these effects for 6 years before (*leads*) and 6 years after the acquisition of the diploma (*lags*). In this empirical model, represent human capital effects, already controlled for diploma effects.

We expect these estimated coefficients to be nonnegative, as it is improbable that spending time in an regular AHS makes an individual less productive. If we further believe these individuals accumulate human capital, then we would expect the estimated coefficients to be positive. The evolution of these coefficients, however, is uncertain. For example, if firms overestimate the human capital accumulated, these coefficients might decline over time; on the contrary, if firms underestimate it, these coefficients could increase as time passes.

Our second strategy takes advantage on the panel feature of our data and estimate individual fixed-effects model in the form of:

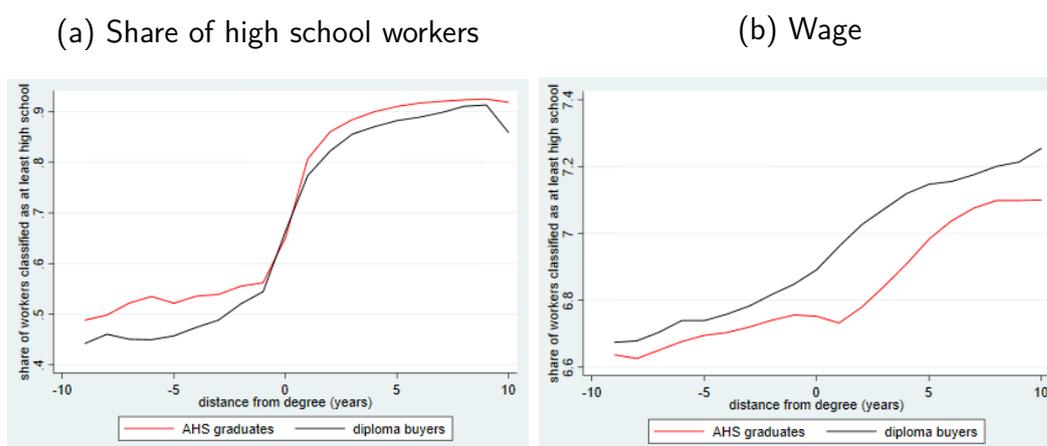
$$w_{i,t} = \alpha_i + \lambda_t + \sum_{\substack{l=-6 \\ l \neq -1}}^{l=6} \gamma_l G_{i,t}^l + \sum_{\substack{l=-6 \\ l \neq -1}}^{l=6} \mu_l T_{i,t}^l + \beta X_{it} + v_{i,t} \quad (2-4)$$

As aforementioned, in this setting we control for the full trajectories of worker wages before the diploma. Once again, we use the probabilities of treatment (finishing studies in a real AHS) to recover the Average Treatment on the Treat (ATT) parameter.

2.5.2 Descriptive Analysis

Figure 2.7 (a) and (b) show that for both treated and control groups, we observe a similar leap in the probability of being classified as high school after getting a diploma - either real or fake. As per wages, the diploma buyers also enjoy an acceleration that looks at least as intense as the real AHS graduates. Table 2.3 show the descriptive statistics, both for the full sample of individuals that we were able to identify in RAIS and for the sample used in the basic regressions, after applying the necessary filters explained previously.

Figure 2.7: Evolution of schooling and wages around AHS graduation



Source: RAIS

Table 2.3: Descriptive Statistics: Diploma buyers versus real AHS graduates

Variable	Full Sample			Basic regression sample		
	Diploma buyers (1)	AHS graduates (2)	t-test Difference (1)-(2)	Diploma buyers (1)	AHS graduates (2)	t-test Difference (1)-(2)
Average wage (R\$)	1183.21 [2.620]	1064.46 [1.357]	118.745***	1296.34 [10.355]	1201.61 [6.387]	94.735***
male	0.728 [0.002]	0.588 [0.001]	0.140***	0.766 [0.006]	0.716 [0.004]	0.050***
black	0.333 [0.002]	0.356 [0.001]	-0.023***	0.357 [0.007]	0.404 [0.005]	-0.047***
age	28.081 [0.021]	25.292 [0.012]	2.789***	30.046 [0.044]	30.045 [0.030]	0.001
schooling (years)	6.576 [0.003]	6.801 [0.002]	-0.225***	6.228 [0.013]	6.297 [0.009]	-0.069***
% classified as high school before diploma	0.66 [0.001]	0.532 [0.001]	0.127***	0.341 [0.003]	0.188 [0.002]	0.153***
Obs.	68110	208674		4879	10621	

The value displayed for t-tests are the differences in the means across the groups.

***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

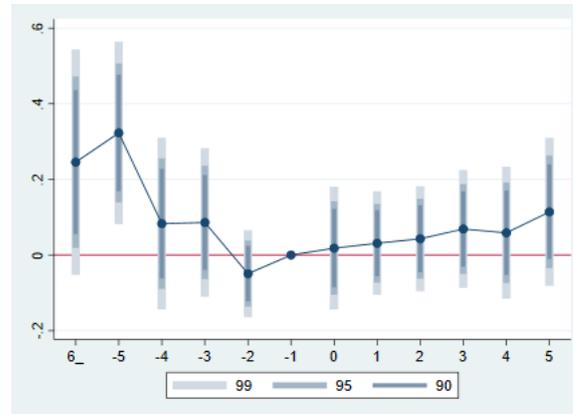
2.5.3 Results

2.5.3.1 Effects on Schooling

As before, the first question we ask is whether finishing the studies in a real AHS has additional impact on the probability of being classified as high school in RAIS beyond any diploma effect. In principle, even if diploma buyers (our control group) enjoy an equal increase in this probability in the short run, as time passes, having really finished the studies in the high school level could be associated with a larger chance of holding a position in a high school level. Figure 2.8 shows the results of the estimation of equation (2.4) replacing the dependent variable for a dummy indicating if the worker is classified as high school. Though the identification is not clear as we fail to observe parallel trends before the treatment, the lag coefficients do not show

any statistical significance, meaning we cannot rule out the case where both diploma buyers and real AHS graduates enjoy the same increase in the chance of being persistently classified as high school by their employers.

Figure 2.8: Effects on Schooling



2.5.3.2 Effects on Wages

We now present the main results of the human capital effects estimated in the basic specifications described above¹². Figure 2.9(a) shows the result of the estimation of equation 2.3 without controlling for wages between 23 and 25 years old. We note that, differently from Chapter 1, we can already observe parallel pre-trends before graduation. This means that even unconditional to wages before graduation, individuals that graduated from regular AHS and individuals that bought diplomas seem to have similar wage trajectories. Comparing to previous chapter's results, this suggests that real AHS students and diploma buyers are more comparable, in terms of wage trajectories before diploma, than diploma buyers and individuals without diplomas. Figure 2.9(b) shows the human capital effects estimated by equation 2.3. As expected, the coefficients from lead indicators are all statistically indistinguishable from zero. Our coefficients of interest, the lag coefficients of the specification, fluctuate around -1% and -8%, but only 3 three of the six coefficients are significant at 10% confidence level, and non at the 95% level. This first set of results suggest that, in comparison to those individuals that bought a diploma, real AHS graduates have no additional wage gains. Also, they leave room for investigation of whether (and why) real AHS perform worse than diploma buyers.

¹²Appendix shows the histogram of probability scores of treated (real AHS students) and nontreated (fake diploma buyers) individuals.

Figure 2.9: Effects on wages

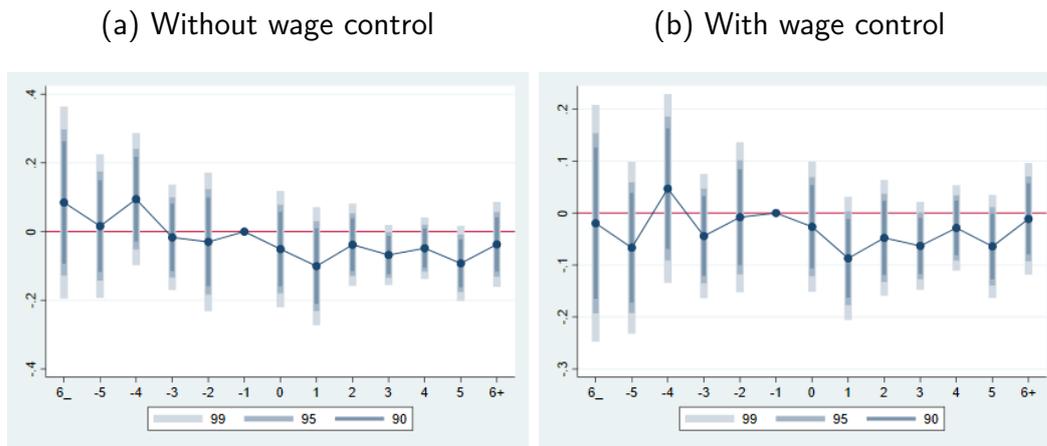


Figure 2.10 shows the results for the individual fixed-effects estimation, restricting the sample for the years before 2014 and also for all the years we have in our sample. As the figure show, the estimated coefficients after graduation are all indistinguishable from zero.

Figure 2.10: Effects on wages - fixed effects model

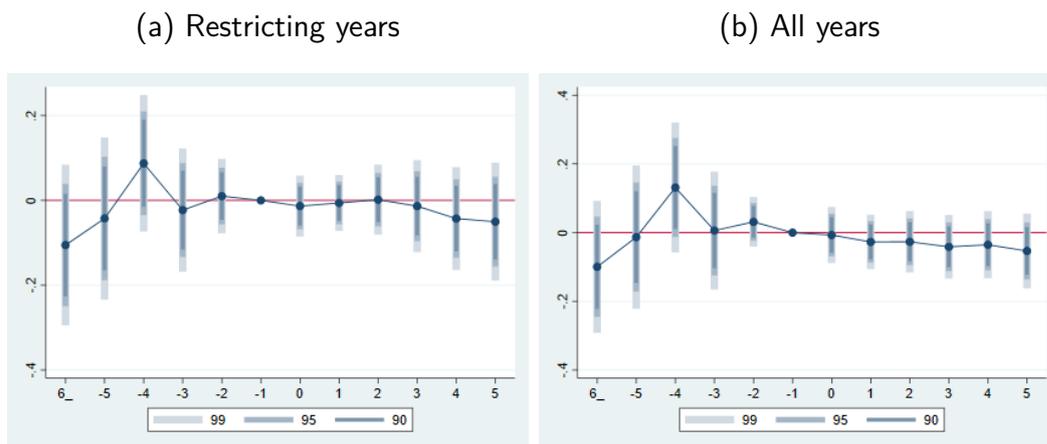


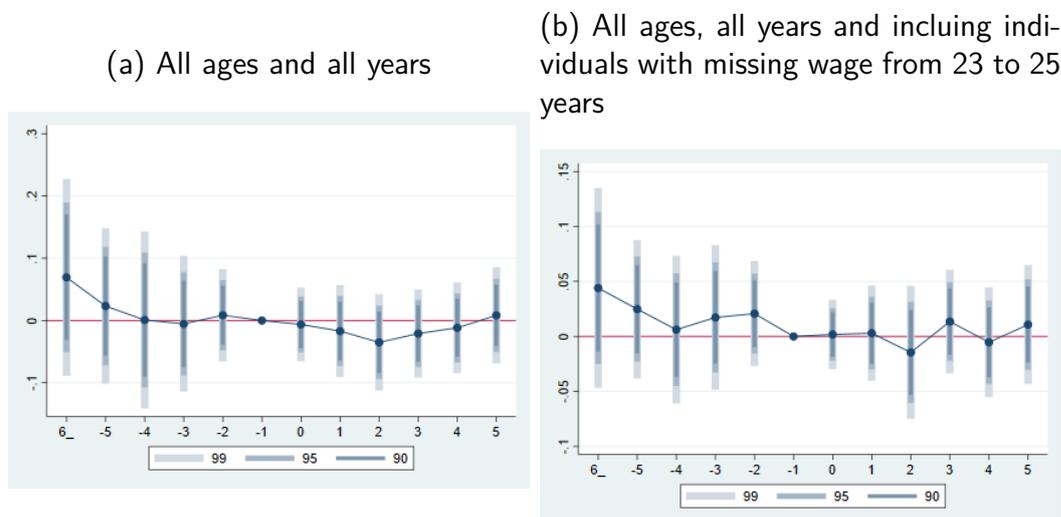
Table 2.4: Effects on Schooling and Wages

	High School	Wages					
		without previous wage control	with previous wage control	Fixed-effects model (restricted years)	Fixed-effects model (all years)	Fixed-effects model (all years and all ages)	Fixed-effects model (all years, all ages and including workers w/o wages 23-25 yrs)
Mean wages 23-25yrs (log)			0.586 (0.051)***				
-6	0.246 (0.116)**	0.085 (0.108)	-0.020 (0.088)	-0.105 (0.074)	-0.100 (0.074)	0.069 (0.061)	0.044 (0.035)
-5	0.323 (0.094)***	0.016 (0.081)	-0.067 (0.064)	-0.043 (0.074)	-0.013 (0.081)	0.023 (0.048)	0.025 (0.024)
-4	0.083 (0.088)	0.094 (0.075)	0.047 (0.070)	0.087 (0.062)	0.131 (0.073)*	0.001 (0.055)	0.006 (0.026)
-3	0.086 (0.076)	-0.017 (0.060)	-0.044 (0.046)	-0.023 (0.056)	0.006 (0.067)	-0.005 (0.042)	0.017 (0.026)
-2	-0.049 (0.045)	-0.030 (0.078)	-0.008 (0.056)	0.010 (0.034)	0.031 (0.028)	0.008 (0.029)	0.021 (0.019)
0	0.018 (0.063)	-0.051 (0.066)	-0.026 (0.049)	-0.013 (0.028)	-0.007 (0.032)	-0.006 (0.023)	0.002 (0.012)
1	0.031 (0.053)	-0.101 (0.067)	-0.087 (0.046)*	-0.006 (0.026)	-0.027 (0.031)	-0.017 (0.029)	0.003 (0.017)
2	0.043 (0.054)	-0.038 (0.047)	-0.048 (0.043)	0.002 (0.032)	-0.027 (0.035)	-0.035 (0.030)	-0.015 (0.023)
3	0.069 (0.061)	-0.068 (0.034)**	-0.063 (0.033)*	-0.014 (0.042)	-0.041 (0.036)	-0.021 (0.027)	0.014 (0.018)
4	0.059 (0.068)	-0.048 (0.035)	-0.029 (0.032)	-0.043 (0.047)	-0.035 (0.038)	-0.012 (0.028)	-0.005 (0.019)
5	0.114 (0.076)	-0.093 (0.042)**	-0.064 (0.038)*	-0.050 (0.054)	-0.053 (0.042)	0.008 (0.030)	0.011 (0.021)
6	0.158 (0.110)	-0.037 (0.048)	-0.011 (0.042)	-0.073 (0.054)	-0.007 (0.046)	-0.002 (0.032)	0.002 (0.025)
male	0.112 (0.097)	0.267 (0.047)***	0.173 (0.037)***	0.122 (0.071)*	0.118 (0.067)*	0.063 (0.031)**	0.039 (0.028)
black race/color	0.033 (0.041)	-0.023 (0.037)	0.008 (0.025)	0.025 (0.018)	0.015 (0.018)	0.013 (0.011)	0.016 (0.009)*
Constant	0.344 (0.162)**	6.532 (0.072)***	2.703 (0.317)***	6.615 (0.102)***	6.651 (0.105)***	6.118 (0.108)***	6.179 (0.059)***
Previous wage as control	No	No	Yes	No	No	No	No
Individual FE	Yes	No	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12,693	15,500	15,500	12,693	15,500	39,814	86,629
R-squared	0.190	0.183	0.403	0.325	0.309	0.363	0.405
Number of cpf	2,635			2,635	2,695	7,640	14,294

Robust standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

Finally, since under this specification we rely on the whole history of wages and do not control specifically for wages from 23 to 25 years old, we can drop the restriction on ages over 25 years old and we also can include individuals without observations of those wages. Figure 2.13 bellow presents the results of the models including both groups, showing that the results aforementioned are not sensitive to the inclusion of these groups in the analysis.

Figure 2.11: Effects on wages - fixed effects model



Taken as whole, our interpretation of the results above is that there is no evidence of human capital accumulation impacting wages in for adult high schools in the context of our analysis, once we control for the impacts of having a diploma.

2.6

Conclusion

The relative importance of the human capital versus the signaling theory in explaining the positive relation between education and earnings has long been object of research. Lately, there has been a growing body of evidence that both mechanisms play a role: schooling provides human capital accumulation that translates into increasing wages (Arteaga, 2018), but sheepskin effects and diploma effects (Mazkeraj and Cabus, 2019) are evidence of mechanisms for individuals to signal their productivity. The understanding that both mechanisms play role provokes a new approach to the agenda: it might not be whether any of these effects might be true for a certain scenario, context, policy or intervention, but more about, for each object of analysis, what is the relative role of each mechanism ¹³.

We study this in the context of a relatively under-researched policy given its size in terms of number of participants and resources invested: Adult High Schools in Brazil. This policy is present in many developing countries as a way of raising schooling of adults and specially for the high school dropouts to finish their studies and get their degrees. Using a novel dataset constructed with individuals that bought diplomas in the Brazilian black market, we search for evidence of human capital effects of Adult High School (AHS) in Brazil. Bearing in mind the diploma effects found in the previous chapter, we compare the labor market performance of individuals that finished regular AHS (and got their regular diploma) to those who bought a high school diploma, thus testing if the human capital acquired at high school would confer any labor market advantage on the job, apart from those granted by the diploma. We take advantage of the panel structure of the matched employer-employee data of RAIS to account for unobserved productivity before the diploma acquisition. We find no evidence of wage gains of studying in a regular AHS versus simply buying a diploma, which is consistent with no accumulation of productive human capital in adult high schools. We also compare real AHS graduates and individuals that did not finish high school. The results suggest that the benefit of completing AHS are similar to that of simply buying a diploma.

Taken together, the results from the first two chapters of this thesis point to the relevant role of the signaling theory in explaining the adult high school premium in Brazil and cast doubts on the importance of the labor market returns to knowledge acquired in adult high schools. Moreover, our analysis in this chapter highlights the salience of considering diploma effects and other implications of the signaling theory in the context of incomplete information, such as employer learning and statistical discrimination, when evaluating effects of schooling on individual earnings.

¹³In a sense, it also poses new challenges for the external validity of estimates of each study.

3 University Diplomas

3.1 Introduction

The first two chapters of this thesis use a database of fake high school diplomas to investigate signaling versus human capital mechanisms behind the premium enjoyed by adult high school graduates in Brazil. In this last chapter we explore a novel database made of fake college graduates, who were exposed and had their degrees canceled following a large police investigation. This allows us to design the expansion of the analysis in at least three dimensions.

First, we are able to investigate the wage evolution of workers acquiring university diplomas. The literature on university diplomas has shown sizeable signaling effects of university prestige, but also significant employer learning as labor market experience increases (Bórdon, 2020). Here, we present preliminary evidence of the premium enjoyed by these workers. Due to data limitations at this point, we are not able to precisely estimate the evolution of this premium, but we show evidence of a magnitude of around 4% premium during some years after the issuance of the diplomas.

Second, the large media propagation of this scandal allows us to design a test for whether employers react to the news that their employee has shown a diploma that was considered of no value. At this moment we cannot run the test also due to data limitations that we expect to briefly overcome. This informational shock might either be interpreted as a revelation of ability or of characteristics such as work ethics. In that field, Hussey (2011) has analyzed the relationship of ethics and labor market performance in the context of MBA degrees and argues for strong gender heterogeneity. The author finds that self-reported measures of ethics are negatively correlated with future performance in the case of men, but positively correlated for women. In another strand, Ferraz (2008) showed that politicians exposed to corruption scandals are punished by the voters with a lower chance of being reelected. We have no knowledge of studies evaluating the effects of this kind of information shock about the workers.

Third, this data also opens the avenue for analyzing differences between public sector versus private sector responses to the diploma issuance and posterior scandal and dispute over its cancellation. Though the public sector corresponds to more than one fourth of the total employment in many countries, its dynamics are known to be very different from the private sector and thus conclusions about private employer responses to signals (either of ability or of work ethics) are not necessarily valid for the public sector.

This chapter is structured in four sections. After this introduction, we briefly describe the context, and present the data. The following section shows the preliminary empirical investigation. The last section concludes and proposes future research.

3.2

Context and Data

Following whistleblowing reports denouncing irregularities in the issuance of undergraduate diplomas, a large investigation conducted by the Legislative Assembly of Pernambuco State condemned thousands of diplomas issued by the University of Nova Iguaçu (UNIG) in Rio de Janeiro State. Basically, this large private university used its legal right to issue diplomas to allow for other, smaller, institutions to use UNIG's seal to issue diplomas in their names, under no supervision¹.

As a result of the investigation, at least 25 smaller secondary education institutions were disqualified by the Ministry of Education. UNIG, also under the threat of disqualification, signed a commitment protocol with the the country's Ministry of Education and representatives of the Federal Prosecutor (*Ministério Público Federal*) indicating the obligation of the university to check the regularity of the documentation of all the diplomas issued under UNIG's seal by other institutions ("external registries"). Following that, UNIG announced the cancellation of more than 65 thousand diplomas, from a total of almost 95 thousand external registries from the period 2011 to 2016 issued with the university's seal². Since then, a dispute has been taking place in the judicial field over the cancellation of these diplomas³.

3.2.1

Data

In this work, we use two main datasets: (i) a novel dataset built with university diplomas that were exposed in the investigation and cancelled and (ii) a matched employer-employee data to analyze the labor outcomes of these individuals.

3.2.1.1

Diplomas cancelled

The list of diplomas cancelled issued by UNIG in the name of the other faculties was made publicly available by order of Justice and could be consulted in the university's website (www.unig.br), from which we extracted the document. Following a court decision that provisionally reestablished the validity of these diplomas, the list was withdraw from the website and replaced for a form to be submitted asking for the data on the individual diploma to be consulted, and the following message: *Temporarily the Consultation of Records of External Diplomas is being made by the form below. Please fill in all the requested fields, send us and wait for our return.*⁴

¹<https://g1.globo.com/rj/rio-de-janeiro/noticia/2019/04/03/unig-cancela-mais-de-65-mil-registros-de-diplomas-no-pais.ghtml>

²Full list of courses cancelled in the Appendix

³<http://www.mpgp.br/portal/noticia/mp-obtem-liminar-para-suspender-cancelamento-de-registro-de-diplomas-universitarios-em-rubiataba#.YG4obuhKjIU>

⁴Free translation from the Portuguese: *Temporariamente a Consulta de Registros de Diplomas Externos está sendo feita pelo formulário abaixo. Por favor preencha todos os campos solicitados, nos envie e aguarde nosso retorno.*

The list we extracted contains 57,774 registries, each one containing the name of the person, the university, the course, and six from the eleven-digit individual taxpayer registration number (CPF - *Cadastro de Pessoas Físicas*).

3.2.1.2

RAIS

Relação Anual de Informações Sociais (RAIS) is a restricted-access matched employee-firm administrative dataset, with information that the Brazilian Ministry of Labor and Employment requests to firms annually. RAIS has information of all formally employed workers in Brazil, and includes detailed information on the employee (social security number, gender, education), on the employer (tax number, sector of activity, establishment size, geographical location), and on the employment relationship (wage, tenure and hiring date). This study uses data from RAIS for the period from 2006 - 2017, the last year we have available. For the next versions, we will soon include the years 2018 and 2019, which will allow us to properly analyze scandal effects.

3.2.1.3

Sample construction

In order to analyze the labor market outcomes of the diploma upholders, we merge the list with the RAIS yearly registries, using as key the combination of the six-digit identifier and the first ten characters of the name. As result, we find 28,983 matches, from our initial list of 57,774 names. In the matching process we do not restrict age or geographical location⁵.

For computational reasons, we build our control sample using a 1% random sample of the individuals in RAIS, resulting in a sample of 888,711 individuals.

3.3

Empirical Investigation

This section presents our initial investigation. First, we show some descriptive statistics about labor market patterns: sector, geographical location and wages. Ideally, we would like to analyze the evolution in three distinct periods: (i) before the diploma issued, (ii) after the diploma, but before the scandal and (iii) after the scandal. At this point, we still lack the data on the year of the diploma issued, but we do have information that they were issued between 2011 and 2016. In next steps of this research we will be able to include the years of 2018 and 2019, and also the year that the diploma was issued. This will allow us to apply an event study setting to properly analyze dynamic effects of our two events of interest: the issuance of the diploma and the posterior cancellation and juridical litigation.

⁵We also tried to merge using only the CPF number as the key identifier, but this results in a large number of incorrect matches. On the other hand, using the CPF number and the full name practically guarantees the accuracy of the matching process, but leaves a lot of names out, as per abbreviations and name changes due to marriages, for example.

3.3.0.1 Descriptive statistics

Our sample reveals interesting patterns. Figure 3.1 shows the distribution of formal employees in Brazil (green bars) and diploma holders (red bars) among Brazilian states. We can see that these individuals are dispersed across the whole country, but four states have shares of treated individuals more than double of the general workforce: Pará, Tocantins, Maranhão and Piauí. All of these are poor states in North or Northeastern regions.

Figure 3.1: Distribution across Brazilian states



Figure 3.2(a) shows that, compared to the general sectoral distribution of formal employees, the diploma holders (green bars) are disproportionately placed at the public sector: the share of treated individuals classified as general public administration was 70% of the population, against 25% of the general population. 3.2(b) shows that this pattern is apparently not a result of the obtainment of the diploma, but was already present before. This might suggest - contrary to our initial prior - that acquiring the diploma was not driven by the need to fulfill a condition required to be hired in the public sector, though more research is needed.

Coherently, the same pattern shows in the classification of *contract type* (Figures 3.3(a) and (b)). Over two-thirds of treated individuals are employed by the public sector, while around one-sixth of the overall individuals. Again, we see similar numbers before the period of diploma acquisitions.

Figure 3.2: Sectoral Distribution

(a) All years

(b) Before 2010

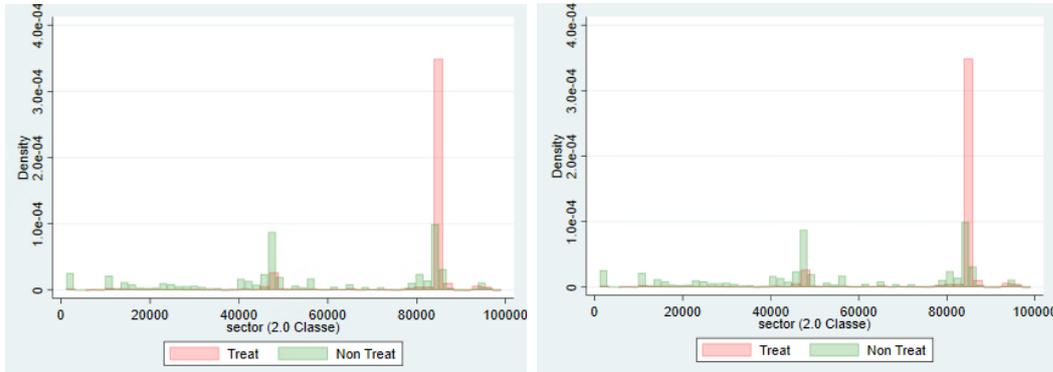
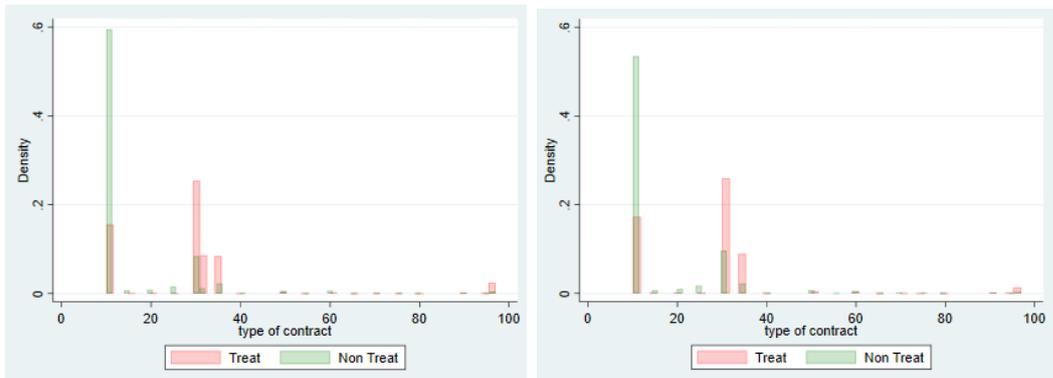


Figure 3.3: Distribution of contract types

(a) All years

(b) Before 2010

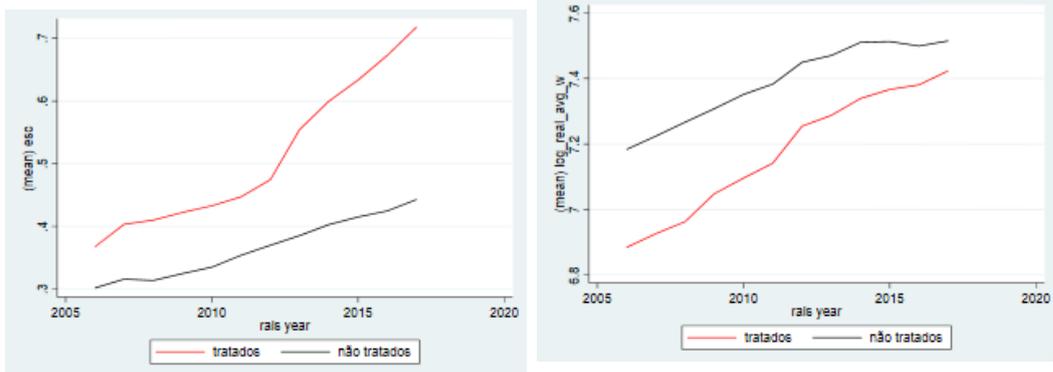


Consistent with the period of diplomas acquisition, Figure 3.4 shows the evolution of the share of workers classified as completed secondary education (a) and the evolution of wages (b).

Figure 3.4: Evolution of secondary education and wages

(a) Share classified as secondary education

(b) Wages



3.3.1

Econometric strategy

Since we lack the individual information on the year that the diploma was issued, we proceed our analysis running the following regression:

$$w_{i,t} = \alpha_i + \lambda_t + \sum_{\substack{l=2006 \\ l \neq 2010}}^{l=2017} \mu_l T_{i,t}^l + \beta X_{i,t} + v_{i,t} \quad (3-1)$$

In which $w_{i,t}$ are wages, $T_{i,t}^l$ is a dummy variable that equals one for observations of the treated individuals in the year $t = l$ and $t = 2010$ is our baseline year. $X_{i,t}$ are a set of controls which includes age and state, α_i are individual fixed effects and λ_t are dummies that capture annual fixed effects. If we believe that the diploma issuance has an impact on wages, we expect this effect starting around 2011, when these irregular diplomas started been issued. Furthermore, the possible impacts of the scandal might already be shown in 2017, although probably they would be more visible in 2018 and 2019, years that we will soon include in the analysis.

3.3.2

Results

This section show the preliminary results of this work. Figure 3.5 plots the estimated coefficients of the dummies in equation 3.1. As explained, 2010 is omitted since we consider this year as the baseline period, and each dummy might be interpreted as the differential in mean wages between treated and control groups. Before 2010 we see some variation with no clear trend. From 2011 on we can observe a positive trend for the treated individuals, which amount to 7% of wage gain versus nontreated workers in the years 2016-17. We also note that the coefficient of 2017 does not drop significantly, suggesting there was no immediate large effect of the scandal on wages, though there is still room to speculate on the effects in the following years. Figure 3.6 then shows that both workers in the private sector as well as in the public sector enjoy positive trajectories on their wages compared to the control group. We interpret this as preliminary evidence as diploma effects in the case of secondary education. Table 3.1 presents the estimation results.

Figure 3.5: Diploma Effects

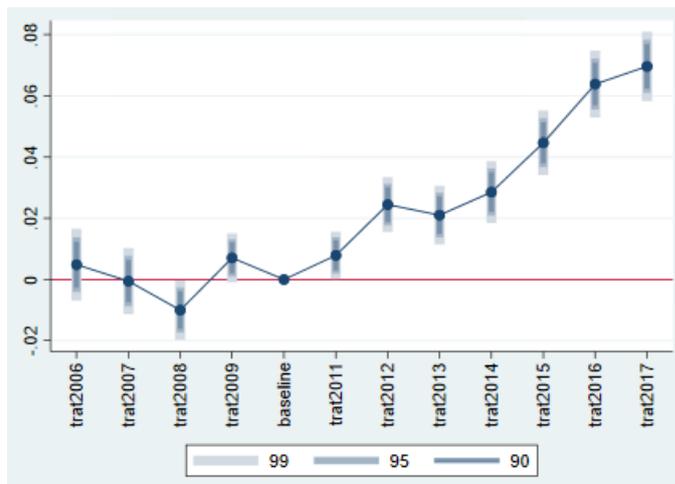
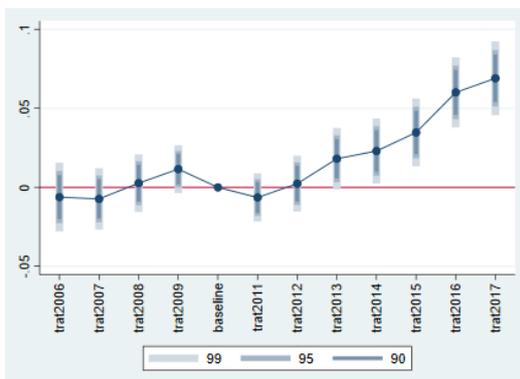


Figure 3.6: Diploma Effects - public versus private sectors

(a) Private sector



(b) Public

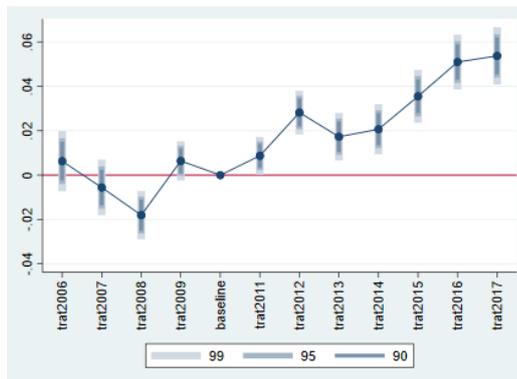


Table 3.1: Regression results

	Secondary Education	Wage		
		Full Sample	Private Sector	Public Sector
2006	-0.039 (0.004)***	0.007 (0.005)	0.006 (0.005)	-0.006 (0.008)
2007	-0.023 (0.004)***	0.000 (0.004)	-0.006 (0.005)	-0.007 (0.008)
2008	-0.013 (0.003)***	-0.009 (0.004)**	-0.018 (0.004)***	0.003 (0.007)
2009	-0.003 (0.003)	0.008 (0.003)**	0.006 (0.003)*	0.012 (0.006)**
2010 (baseline)	-	-	-	-
2011	-0.005 (0.003)**	0.008 (0.003)***	0.009 (0.003)***	-0.006 (0.006)
2012	0.007 (0.003)**	0.024 (0.003)***	0.028 (0.004)***	0.002 (0.007)
2013	0.054 (0.003)***	0.019 (0.004)***	0.017 (0.004)***	0.018 (0.008)**
2014	0.076 (0.004)***	0.026 (0.004)***	0.021 (0.004)***	0.023 (0.008)***
2015	0.099 (0.004)***	0.041 (0.004)***	0.036 (0.005)***	0.035 (0.008)***
2016	0.132 (0.004)***	0.060 (0.004)***	0.051 (0.005)***	0.060 (0.009)***
2017	0.158 (0.004)***	0.063 (0.004)***	0.054 (0.005)***	0.069 (0.009)***
Constant	0.018 (0.016)	6.443 (0.018)***	6.585 (0.027)***	6.414 (0.025)***
Individual FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Age dummies	Yes	Yes	Yes	Yes
Observations	2,079,177	2,079,177	904,869	1,174,308
R-squared	0.134	0.318	0.351	0.210
Number of cpf	368,155	368,155	142,602	258,295

Robust standard errors in parentheses

3.4

Final remarks

This chapter introduces a novel database of irregular university diplomas and proposes an initial investigation that might contribute to the literature on a few dimensions. First, the analysis of the wage trajectories of the workers that acquired these irregular diplomas could reveal whether diploma effects persist or diminish over time, informing about the dynamics of statistical discrimination and employer learning in the market of low quality secondary education. Second, the dynamics after 2017 can shed light on whether this large information shock has reached the labor market and if possible impacts can be accounted for a revelation of work ethics, ability or other. We will briefly be able to expand the analysis as we introduce data on 2018 and 2019 labor market outcomes.

Last, we are also able to analyze differential responses by the private and public sector, which might inform about relevant public sector dynamics. Although very preliminary, our results show that workers employed in both public and private sector enjoy lasting wage premiums with similar trajectories, leaving us the question whether these will persist after the revelation of the scandal.

4

References

- [1] Altonji, J. G.; Pierret, C. R. (2001): "Employer Learning and Statistical Discrimination," *The Quarterly Journal of Economics*, vol. 116(1), pp. 313-350.
- [2] Arteaga, C, (2018): "The effect of human capital on earnings: Evidence from a reform at Colombia's top university," *Journal of Public Economics*, Elsevier, vol. 157(C), pages 212-225.
- [3] Bordón, P.; Braga, B. (2020): "Employer learning, statistical discrimination and university prestige," *Economics of Education Review*, Elsevier, vol. 77(C).
- [4] Borusyak, K.; Jaravel, X. (2016): *Revisiting Event Study Designs, with an Application to the Estimation of the Marginal Propensity to Consume*. Working Paper.
- [5] Braga, B. (2018): "Earnings dynamics: The role of education throughout a worker's career," *Labour Economics*, Evol. 52(C), pages 83-97.
- [6] Bennett, P.; Blundell, R.; Salvanes, K. (2020) : "A Second Chance? Labor Market Returns to Adult Education Using School Reforms," IZA Discussion Paper No. 13592.
- [7] Ferraz, C.; Finan, F. (2008): "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes *". *Quarterly Journal of Economics* 123, nº 2: 703-45.
- [8] Kahn, L. B. (2013): "Asymmetric Information between Employers" *American Economic Journal: Applied Economics*, American Economic Association, vol. 5(4), pages 165-205, October.
- [9] Lange, F. (2007): "The Speed of Employer Learning," *Journal of Labor Economics*, University of Chicago Press, vol. 25, pages 1-35.
- [10] Mazrekaj, D.; Cabus, S. (2019): "Does a High School Diploma Matter? Evidence Using Regression Discontinuity Design". Working Paper.
- [11] Heckman, J.; Humphries, J.; Mader, N. (2010): *The GED*. NBER Working Paper n. 16064.
- [12] Heckman, J.; Urzua, S.; Vlytacil, E. (2006): "Understanding Instrumental Variables in Models with Essential Heterogeneity" *Review of Economics and Statistics* 88(3):389-432 Hussey, A. (2011): "The effect of ethics on labor market success: Evidence from MBAs". *Journal of Economic Behavior Organization*, Elsevier, vol. 80(1), pages 168-180.
- [13] Tyler, John H.; Murnane, Richard J.; Willett, John B. (2000): "Estimating the Labor Market Signaling Value of the GED". *The Quarterly Journal of Economics*, 2000, vol. 115, issue 2, 431-468

- [14] Sun, L.; Abraham, S. (2018): "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," Papers 1804.05785, arXiv.org, revised Sep 2020.

A

Appendix to Chapter 1

A.1

Evidence gathered

Figure A.1: Email exchanged with proposal

diploma ensino médio Caixa de entrada x 🔍 🖨️ 📧

Marcio Firme <marciogoldfirme@gmail.com> 17 de jun. de 2019 15:20 ☆ ↶ ⋮
para

Boa tarde Aline,
Queria saber como emitir diploma de ensino médio no estado do Rio. Os prazos são aqueles do site historicoescolarquente.com.br? Queria também saber as escolas disponíveis, posso escolher? Aguardo resposta! Vai me ajudar muito!
abraços,

. @gmail.com> 18 de jun. de 2019 10:28 ☆ ↶ ⋮
para mim

Bom dia.

Para você que gostaria de adquirir uma documentação de procedência do Ensino Fundamental, Ensino Médio, Técnico ou Ensino Superior essa é a hora!

Trabalhamos a anos no ramo somente com documentação original, todos os documentos são reconhecidos pelo MEC com publicação no Diário Oficial e cadastro no GDAE, podendo ingressar em qualquer faculdade, cursos, adquirir CREA ou CRECI, prestar concursos, apresentar em empresas e etc...

Emitimos para vários estados brasileiros (com data retroativa caso haja preferência), portanto os documentos serão emitidos por uma instituição de ensino do mesmo estado residente do cliente ou por uma instituição específica de sua preferência caso haja disponibilidade, consulte-nos!

O prazo para a entrega dos documentos é de 5 dias úteis via Sedex ou entregamos em mãos no centro de São Paulo.

Para sua maior comodidade o pagamento pode ser feito via depósito ou transferência bancária, trabalhamos com diversos bancos para melhor atendê-lo.

Todo o processo é feito sobre o mais absoluto sigilo, após a emissão dos documentos todos os dados encaminhados serão apagados para sua segurança.

Não caia em golpes e receba papéis sem valor algum! Adquirindo conosco, o cliente estará recebendo uma documentação de procedência conforme dito anteriormente.

Vamos aos valores!

Documentação Ensino Fundamental R\$ 600,00;
Ensino Médio R\$ 600,00;
Ensino Fundamental + Ensino Médio R\$ 800,00;
Curso Técnico ou Superior a consultar!

Para adquirir agora, por favor nos encaminhe os dados solicitados abaixo:

- Nome completo;
- Data de nascimento;
- Número RG;
- Estado emissor do RG;
- Data de expedição do RG;
- Nome da mãe;
- Nome do pai;
- Endereço completo;
- Ano de conclusão desejado (pode ser retroativo);

Figure A.2: Two diplomas with the exact same grades

DISCIPLINAS CURSO <u>Ensino Médio - Educação de Jovens e Adultos</u>			Ensino Fundamental Instituto Latino de Ciência e Tecnologia - 1º Sem. 2010 Rio de Janeiro - RJ	O PRESENTE FOI REGISTRADO SOB O Nº <u>23564</u> EM FLS. <u>047</u> DO LIVRO Nº <u>010</u> DESTA ESTABELECIMENTO, CONFORME LISTAGEM PUBLICADA NO D.O. DE <u>31 / 08 / 2012</u> FLS. <u>06 e 07</u>
DISCIPLINAS MÉDIA OBSERVAÇÃO			OBSERVAÇÕES: De acordo com a deliberação nº 285/2003 e nº 297/2006 - CEE/RJ, artigo 3º, parágrafo único, o(a) aluno(a) consta com 75% de carga horária e frequência mínima devido nossa Instituição de Ensino ser a Distância. O(a) aluno(a) teve sua vida escolar regularizada no Ensino Fundamental, com base nas deliberações 225/1998, art. 2º / 253/2000, art. 5º, Parágrafo 2º - Item II do CEE/RJ de acordo com a LDB nº. 9394/96, art. 22, 23 e 24, obtendo aprovação, com documentação arquivada nesta instituição de Ensino. O(a) aluno(a) está apto(a) a ingressar no Ensino Superior.	Rio de Janeiro - RJ 25/09/2012 LOCAL DATA
Língua Portuguesa 76 Habilitado 2011 Literatura 81 Habilitado 2011 Artes 84 Habilitado 2011 Redação 79 Habilitado 2011 Matemática 71 Habilitado 2011 Química 72 Habilitado 2011 Física 76 Habilitado 2011 Biologia 73 Habilitado 2011 História 84 Habilitado 2011 Sociologia 78 Habilitado 2011 Filosofia 78 Habilitado 2011 Geografia 75 Habilitado 2011 Língua Estrangeira 83 Habilitado 2011	OUTRAS HABILITAÇÕES:			DIRETOR CREDENCIADO SECRETÁRIA ESCOLAR Secretária Escolar - Parecer 003/2007
DISCIPLINAS CURSO <u>Ensino Médio - Educação de Jovens e Adultos</u>			Ensino Fundamental Sumaré - SP	O PRESENTE FOI REGISTRADO SOB O Nº <u>25415</u> EM FLS. <u>22</u> DO LIVRO Nº <u>011</u> DESTA ESTABELECIMENTO, CONFORME LISTAGEM PUBLICADA NO D.O. DE <u>28 / 09 / 2012</u> FLS. <u>12 e 13</u>
DISCIPLINAS MÉDIA OBSERVAÇÃO			OBSERVAÇÕES: De acordo com a deliberação nº 285/2003 e nº 297/2006 - CEE/RJ, artigo 3º, parágrafo único, o(a) aluno consta com 75% de carga horária e frequência mínima devido nossa Instituição de Ensino ser a Distância.	Rio de Janeiro - RJ 20/12/2012 LOCAL DATA
Língua Portuguesa 76 Habilitado 2011 Literatura 81 Habilitado 2011 Artes 84 Habilitado 2011 Redação 79 Habilitado 2011 Matemática 71 Habilitado 2011 Química 72 Habilitado 2011 Física 76 Habilitado 2011 Biologia 73 Habilitado 2011 História 84 Habilitado 2011 Sociologia 78 Habilitado 2011 Filosofia 78 Habilitado 2011 Geografia 75 Habilitado 2011 Língua Estrangeira 83 Habilitado 2011	OUTRAS HABILITAÇÕES:			DIRETOR CREDENCIADO SECRETÁRIA ESCOLAR

Figure A.3: Example of a payment for a diploma, with the same value as proposed in the email received

Plano de Pagamento

CURSO DE EDUCAÇÃO PARA JOVENS E ADULTOS

NÍVEL: _____

CRONOGRAMA:

Parcelas	Vencimentos	Valores
1ª	08/03/2012	R\$100,00 pago 100%
2ª	10/04/2012	R\$100,00 pago 100%
3ª	10/05/2012	R\$100,00 pago 100%
4ª	10/06/2012	R\$100,00 pago 100%
5ª	10/07/2012	R\$100,00 pago 100%
6ª	10/08/2012	R\$100,00 pago 100%
D.O	10/09/2012	R\$ 80,00 pago 20/12/12

O valor total do curso é de R\$600,00 (Seiscentos Reais)
 O valor total do D. O. é de R\$ 80,00 (Oitenta Reais)
 As parcelas deverão ser pagas até a data acordada, caso contrário, entrará em vigor a cláusula QUARTA deste CONTRATO DE PRESTAÇÃO DE SERVIÇO EDUCACIONAL.

Rio de Janeiro, 08 de março de 2012

 Funcionário do Pólo

 Aluno

A.2

Probability of treatment

We estimate probability scores of being treated using a probit model in which the probability that individual i acquired a diploma, depends on year, state and age.

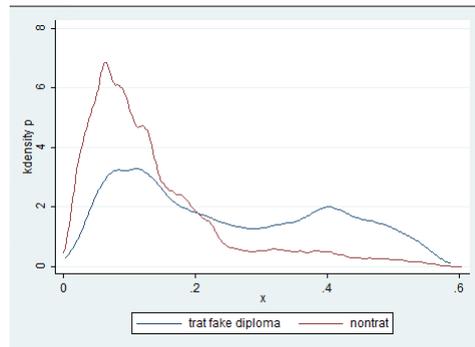
$$p(i) = \alpha_i * X^T * \beta \quad (\text{A-1})$$

Where $p(i)$ is the probability that observation i acquired a diploma, which depends on year, state and age. We then use inverse probability weighting to estimate the Average Treatment on the Treated (ATT) effect.

$$\mathbf{w}_i = \begin{cases} 1 & T_i = 1 \\ \frac{p_i}{1-p_i} & T_i = 0 \end{cases} \quad (\text{A-2})$$

Figure bellow shows the histogram of probability scores of treated and nontreated individuals. It shows that we have common support and that probabilities of the treated group being treated are higher than of the probabilities of the nontreated group.

Figure A.4: Kernel



A.3

Results

Table A.1: Diploma Effects on Wages - other samples

	Restricted years, with missing wages	Restricted years, all ages	Restricted years, all ages, with missing wages	All years, all ages	All year, all ages, with missing wages
<-6	0.005 (0.019)	0.025 (0.026)	0.015 (0.014)	0.026 (0.027)	0.014 (0.014)
-5	-0.006 (0.016)	0.019 (0.022)	-0.002 (0.012)	0.020 (0.022)	-0.002 (0.012)
-4	-0.020 (0.013)	-0.007 (0.017)	-0.015 (0.010)	-0.006 (0.017)	-0.016 (0.010)
-3	-0.033 (0.011)***	-0.014 (0.014)	-0.021 (0.008)**	-0.014 (0.014)	-0.021 (0.008)**
-2	-0.009 (0.009)	-0.000 (0.010)	-0.001 (0.007)	0.002 (0.010)	0.001 (0.007)
baseline	-	-	-	-	-
0	-0.005 (0.008)	0.029 (0.010)***	0.008 (0.006)	0.027 (0.010)***	0.007 (0.006)
1	0.011 (0.011)	0.034 (0.012)***	0.022 (0.008)***	0.036 (0.012)***	0.022 (0.008)***
2	0.024 (0.012)**	0.052 (0.013)***	0.038 (0.009)***	0.052 (0.013)***	0.038 (0.009)***
3	0.033 (0.014)**	0.058 (0.016)***	0.046 (0.011)***	0.055 (0.014)***	0.046 (0.010)***
4	0.062 (0.016)***	0.072 (0.017)***	0.064 (0.012)***	0.064 (0.015)***	0.059 (0.011)***
5	0.046 (0.020)**	0.074 (0.020)***	0.064 (0.015)***	0.064 (0.017)***	0.056 (0.012)***
>6	0.086 (0.030)***	0.092 (0.027)***	0.088 (0.022)***	0.058 (0.018)***	0.050 (0.014)***
male	0.060 (0.023)***	0.041 (0.028)	0.042 (0.017)**	0.062 (0.027)**	0.051 (0.017)***
black race/color	0.003 (0.008)	0.012 (0.009)	0.006 (0.006)	0.012 (0.008)	0.005 (0.005)
Constant	6.589 (0.021)***	6.420 (0.124)***	6.297 (0.025)***	6.389 (0.072)***	6.287 (0.024)***
Sample: all years	No	No	No	Yes	Yes
Sample: all ages	No	Yes	Yes	Yes	Yes
Sample: workers with missing wages	Yes	No	Yes	No	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Age dummies	Yes	Yes	Yes	Yes	Yes
Observations	124,104	76,325	206,041	92,745	233,534
R-squared	0.284	0.367	0.379	0.354	0.382
Number of cpf	33,439	19,151	50,470	21,024	53,106

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Figure A.5: FE, with missing, all ages, low pscores

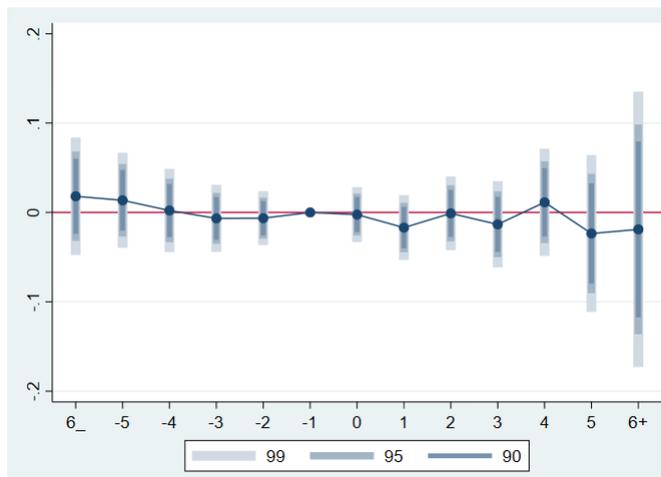
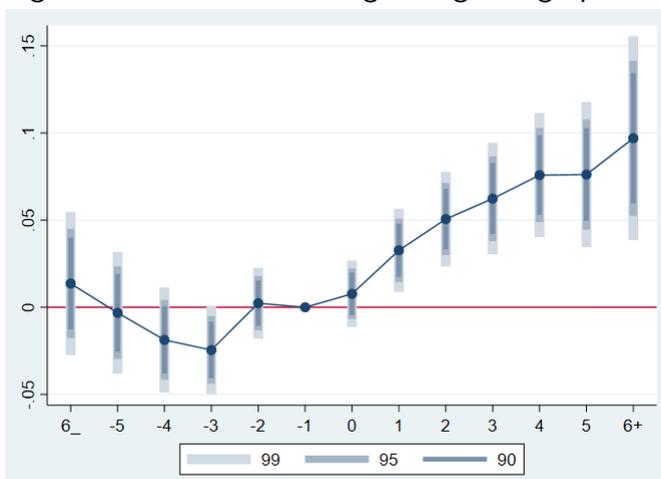


Figure A.6: FE, with missing, all ages, high pscores



B Appendix to Chapter 2

B.1 Treatment probabilities

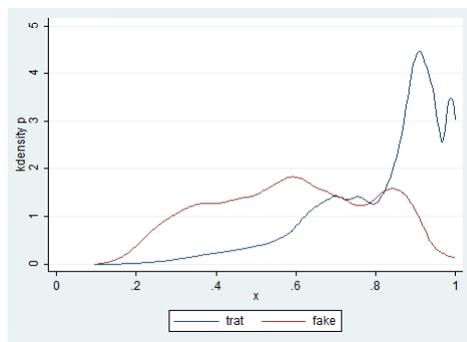
We estimate probability scores of being treated using a probit model in which the the probability that the individual graduated in an AHS depends on year, state and age, as bellow.

$$p(i) = \psi_i * X' * \beta \quad (\text{B-1})$$

Where $p(i)$ is the probability that observation i graduated in a AHS, which depends on year, state and age. We then weight each observation by the inverse probability, where w_i is defined as bellow, corresponding to the estimating the Average Treatment effect on the Treated (ATT).

$$w_i = \begin{cases} 1 & T_i = 1 \\ \frac{p_i}{1-p_i} & T_i = 0 \end{cases} \quad (\text{B-2})$$

Figure B.1: Pcores estimated: real AHS versus diploma buyers



C

Appendix to Chapter 3

C.0.1

List of courses

IES: (1877) Faculdade Atual, Curso de (100598) Pedagogia, ingressantes 2011/2012. (4005) Faculdade Cenecista de Senhor do Bonfim, Curso de (108463) Letras (Português - Literatura). (2633) Instituto de Educação e Tecnologias - INET, Curso de (100532) Pedagogia, ingressantes 2009 a 2012. (801) Faculdade Montenegro, Curso de (19836) Secretariado Executivo, ingressantes 2000/2009/2010. (2929) Faculdade de Milagres Ceará - FAMICE, Cursos de (81104) Administração, ingressantes 2007 a 2011; (113206) Ciências Contábeis, ingressantes 2006/2008/2010; (90610) Letras (Português - Inglês), ingressantes 2009 a 2011; (90609) Letras (Português - Literatura), ingressantes 2008/2010. (1501) Faculdade Latino Americana de Educação, Cursos de (60598) Administração, ingressantes 2009/2010; (37250) Pedagogia, ingressantes 2008 a 2011. (1172) Faculdade AD 1 - UNISABER/AD1, Cursos de (120265) Pedagogia, ingressantes 2009/2010 (706) Faculdade de Artes Dulcina de Moraes, Curso de (24098) Artes, ingressantes 2014/2015. (740) Faculdade de Ciências Humanas de Vitória - FAVIX, Curso de (57408) Pedagogia. (3375) Instituto de Ensino Superior Múltiplo, (1332126) Educação Física - Bacharelado, ingressantes 2009; (103036) Letras (Português - Literatura), ingressantes 2010. (1863) Faculdade de Mantena (nome atual Faculdade Mantense dos Vales Gerais - INTERVALE), Curso de (73778) Serviço Social, ingressantes 2010. (2151) Instituto Superior de Educação Berlaar, Cursos de (105382) (122608) Serviço Social, ingressantes 2007 a 2010. (4446) Faculdade Cidade de Guanhães, Cursos de (96987) História, ingressantes 2006 a 2013; (96985) Matemática, ingressantes de 2008 a 2013; (96989) Serviço Social, ingressantes 2010 a 2012. (3394) Faculdade Vale do Ipiranga, Curso de (82397) Administração, ingressantes 2010 (1678) Faculdade de Ciências de Wenceslau Braz, Curso de (46479) Pedagogia, ingressantes 2008 a 2012. (12597) Faculdade Passionista de Educação de Curitiba, Curso de (1043544) Letras (Português - Literatura). (1034) - União das Escolas Superiores da FUNESO, Cursos de (18221) Administração, ingressantes 2010/2011; (33560) Pedagogia, ingressantes 2010/2011. (3376) Faculdade de Teologia Integrada, Curso de (86436) Teologia, ingressantes 2010/2011. (2012) Sociedade de Educação Cultura e Esportes de Pesqueira - SECEP, Curso de (57988) Filosofia, ingressantes 2009. (3995) Faculdade Evangélica Cristo Rei, Cursos de (91112) Ciências Biológicas - Licenciatura, ingressantes 2006/2008/2010/2011/2012; (91110) Ciências Contábeis, ingressantes 2010/2011; (91105) História, ingressantes 2006 a 2013; (91104) Letras (Português - Literatura), ingressantes 2006 a 2013; (91107) Pedagogia ingressantes 2008 a 2013. (2677) Instituto Superior de Educação São Judas Tadeu, Cursos de (73842) Pedagogia, ingressantes 2009/2010/2012. (2384) Faculdade Integrada do Brasil, Curso de (90614) Pedagogia, ingressantes 2010 a 2012. (2832) Instituto Superior de Educação Programus, Curso

de (80234) Pedagogia, ingressantes 2010/2012. (14914) Faculdade Teológica Evangélica do Rio de Janeiro - FATERJ, Cursos de (5000007) Teologia, ingressantes 2009 a 2012. (5023) Faculdades de Santo Augusto, Cursos de (114901) Administração, ingressantes 2011; (1110541) Educação Física - Licenciatura, ingressantes 2011 a 2013. (1756) Faculdade Associada Brasil, Curso de (100580) Pedagogia, ingressantes 2010. (2289) Faculdade União Cultural do Estado de São Paulo, Curso de (105407) Pedagogia, ingressantes 2009/2010/2011. (2341) Faculdade da Aldeia de Carapicuíba, Curso de (5000223) Pedagogia, ingressantes 2010/2011/2013. (509) Faculdades Integradas de Cruzeiro, Curso de (39058) Pedagogia, ingressantes 2011/2012. (2332) Faculdade Corporativa CE-SPI, Curso de (74272) Pedagogia, ingressantes 2010/2013. (1865) Instituto Superior de Educação Alvorada Plus, Cursos de (49863) Letras (Português - Espanhol); (113022) Pedagogia. (363) Faculdade Mozarteum de São Paulo - FAMOSP, Curso de (33541) Artes Visuais, Ingressantes 2011 a 2014. (17322) Faculdade Integrada de Araguatins, Curso de (2500060) Pedagogia, ingressantes 2008 a 2012.