Fake it till you (don't) make it: Fake diplomas and signaling in labor markets

Eric Bettinger[*] **Marcio Firmo**[†] **Gustavo Gonzaga**[‡] **Vitor Pereira**[§¶]

August 27, 2025

Abstract

Does schooling create new skills or does it merely signal individuals' underlying abilities? We provide new and innovative evidence on this long-standing question in the context of adult education in Brazil. In 2018, police in Rio de Janeiro abruptly shut down 21 high schools, presenting evidence that these schools sold diplomas without providing any educational instruction whatsoever - thus providing a signal but not human capital. We match student records with historical employment data and find that initially, the labor market returns to high school degrees are similar for individuals with legitimate and fake credentials. However, within a few years, the value of fake diplomas begins to decline. The initial gains for fake diploma holders seem to be closely associated with a spike in the probability of changing jobs and occupations shortly after obtaining the diploma. However, over time, these individuals experience substantially greater volatility in formal labor market participation: the likelihood of leaving a job at least three

[*]Stanford University, email: ebetting@stanford.edu.

[†]State University of Rio de Janeiro/UERJ, email: marciogoldfirmo@gmail.com.

[‡]Pontifical Catholic University of Rio de Janeiro/PUC-Rio, email: gonzaga@econ.puc-rio.br.

[§]Federal University of Rio de Janeiro/UFRJ, email: vitpereira@gmail.com.

[¶]We are especially grateful to Alessandro Sathler (SEEDUC/RJ) for invaluable help understanding the scheme and analysing forged diplomas. We also thank Gabriel Thomas for superb research assistance, as well as Arthur Vasconcellos, Carolina Moura, Manoela Rodrigues, and Raissa Oliveira for their support at various stages of the project. All remaining errors are our own.

times over the period is 10 percentage points higher. We also find additional evidence that the value of the fake credential depreciates more rapidly in contexts where worker productivity is more readily observable. Overall, the results provide strong evidence that, at least in the short term, schooling acts primarily as a signal in the labor market. However, the strength of that signal erodes quickly as employers learn about the true productivity of their employees.

1 Introduction

Do schools teach skills, or do they simply signal the underlying abilities of individuals? This fundamental question remains central in the economics of education. On the one hand, there is the human capital model in which workers acquire new skills through schooling and subsequently rent these skills to the labor market (Becker, 1994; Ehrenberg et al., 2021). On the other hand, there is the signaling model, where education serves primarily to reveal the underlying abilities of the students, with its monetary and non-monetary costs deterring students with low ability to pursue further education (Spence, 1973; Arrow, 1973). The signaling perspective has been extensively reviewed and empirically tested (Weiss, 1995). Empirical attempts to distinguish between the two theories have relied on various identification strategies, including compulsory schooling laws (Bedard, 2001), discontinuities in educational requirements (Clark and Martorell, 2014), and employer learning models (Lange, 2007). Although hybrid models that incorporate both skill acquisition and signaling are plausible (Hungerford and Solon, 1987), the literature has not yet reached a consensus on whether one framework clearly dominates the other.

Our work builds on previous efforts to distinguish whether schooling provides skills or serves as a signal (Jaeger and Page, 1996; Bedard, 2001; Tyler et al., 2000; Clark and Martorell, 2014). We focus on the adult education market in Brazil, which represents a significant segment of the education system, and take advantage of a unique natural experiment: On September 24, 2018, law enforcement authorities conducted raids on 11 private schools in the city as part of *Operação Nota Zero* (Operation Grade F) (O Globo (2018); G1 Rio (2018)). The investigation revealed that these schools, most of which had operated for a decade or more, provided no educational services other than conferring fake degrees. Later, as the investigation expanded, police uncovered ten more schools that also sold fraudulent diplomas.

Brazil requires all schools to publish a list of students who completed their diplomas. We scraped the online listings of degree recipients from all high schools in the state of Rio de Janeiro, the core of the fraudulent scheme, from 2005 to 2015. We then merged this data containing the student diploma lists (both fake and real) to

RAIS (*Relação Anual de Informações Sociais*), an employer-employee dataset that contains information on the complete labor market history of all formal workers in Brazil from 1995 to 2020.

We compare labor market outcomes across three groups: recipients of fraudulent diplomas from the 21 schools identified in the police investigation, adults who earned legitimate high school diplomas during the same period, and high school dropouts. By matching high school graduation records (both fraudulent and legitimate) with comprehensive employer-employee data, we reconstruct complete employment histories for these adult learners. This longitudinal approach enables within-person comparisons and allows us to identify precise discontinuities in earnings profiles that coincide with diploma receipt.

Our findings reveal several key patterns. Initially, both fraudulent and legitimate diplomas yield substantial and statistically similar returns, suggesting that the signaling value of credentials operates independently of underlying human capital acquisition. However, these returns diverge significantly over time. While employment probabilities continue rising for legitimate graduates, they begin declining for holders of fraudulent credentials. Similarly, mean wages diverge between the two groups, regardless of whether we condition on employment status.

These temporal patterns align closely with predictions from signaling theory and the employer learning literature. Consistent with signaling models, individuals who remain with the same employer after obtaining a diploma experience negligible wage gains—likely because incumbent employers already possess detailed information about worker productivity, rendering educational signals less informative. In contrast, many individuals strategically time diploma receipt with job transitions, where prospective employers have limited productivity information. These job changes create opportunities for diplomas to function as effective signals, generating substantially larger observed returns.

We also document important sectoral heterogeneity in our results. In industries with limited career advancement opportunities and compressed wage structures, employment dynamics remain remarkably similar between fraudulent and legitimate diploma holders. Significant divergences in log wages emerge only in sectors where career ladders provide substantial opportunities for upward mobility and wage growth—precisely the contexts where educational signaling effects should be more pronounced.

[WE SHOULD LATER THINK ABOUT THREATS TO IDENTIFICATION, PLACEBOS AND ROBUSTNESS CHECKS, AND THEN PLACE A PARAGRAPH HERE. ONE TEST IS TO USE THE LIST OF ONE OF THESE SCHOOLS THAT WAS FOUND BY THE POLICE. BOTH GROUPS BOUGHT THE DIPLOMA, BUT ONLY ONE OBTAINED. WE CAN ALSO TRY TO MERGE WITH TEST SCORES. I HAVE TEST SCORES FROM RIO, YEARS 2010-2013.]

Our research contributes to two primary strands of economic literature. First, we advance understanding of education's signaling value. Previous studies such as Clark and Martorell (2014) and Mazrekaj and Cabus (2019) employ regression discontinuity designs around diploma requirements, yielding mixed evidence on signaling effects. Our unique natural experiment—comparing fraudulent and legitimate credentials within identical institutional contexts provides a clear opportunity to estimate the signaling and human capital effects.

Second, we contribute novel evidence to the employer learning literature. Building on seminal work by Altonji and Pierret (2001), Lange (2007), Bordon and Braga (2020), and Kahn (2013), we document the dynamic evolution of credential values as employers acquire information about worker productivity. Our design allows us to observe learning processes for workers with identical signals but heterogeneous underlying human capital, providing fresh insights into signaling dynamics, temporal patterns of employer learning, and asymmetric information across labor market participants.

The study most closely aligned with our population—adult learners—is Tyler et al. (2000), who exploit interstate variation in GED passing standards to compare individuals with similar educational backgrounds and test scores. While their quasi-experimental design provides valuable insights into credentialing effects, our approach offers the opportunity to directly compare fraudulent and legitimate credentials within identical institutional settings. Second, our comprehensive employer-employee matched data allows us to track complete employment histories and examine how returns evolve as employers learn about worker productivity—dynamics that cannot be observed in their cross-sectional framework.

Our work also advances beyond existing regression discontinuity studies such as Clark and Martorell (2014) Graetz (2021) Jepsen et al. (2016) and recent research on skill certifications like Mancino et al. (2024). These studies, while methodologically rigorous, cannot separate signaling from human capital effects since all credentialed individuals in their samples possess the underlying skills. Our unique setting—where some individuals hold credentials without corresponding human capital—enables us to isolate pure signaling mechanisms and observe their temporal evolution.

Our study is also related to the work by Arteaga (2018), Hansen et al. (2024),who examines educational signaling in traditional academic contexts where some individuals have acquired more human capital than others because of curricular or GPA grading changes, but both obtain the same college diploma. Our focuses on adult high school education and provides novel insights into signaling dynamics in contexts where employer information asymmetries may be particularly pronounced and initial wages might be less related to workers human capital (Arcidiacono et al., 2010).

The remainder of the paper is organized as follows. Section 1 provides background information on the Brazilian context. Section 2 reviews the relevant literature and presents a basic signaling model. Section 3 describes the empirical strategy and data. Section 4 presents the baseline results. Section 5 discusses the findings and their implications for human capital and signaling theories. Section 6 concludes.

2 Institutional background

Since 1988, Brazil has undergone a remarkable educational transformation, bringing millions of previously excluded students into the school system. According to analysis of IBGE and INEP data (Pereira, 2022), this expansion has led to dramatic improvements on enrollment rates: in 1992, only 32% of Brazilians aged 17 had completed primary and middle education, rising to 86% by 2019. High school completion rates among 20-year-old individuals increased from less than 20% in 1992 to over 70% in 2019. However, substantial challenges remain in educational completion rates. At age 18, only 1 in 2 Brazilians had finished high school, and even at age 24, the completion rate reaches only 7 out of 10 individuals. School Census data shows that the dropout problem is particularly acute in high school, where nearly 10% of students drop out each grade every year (Pereira, 2022).

This increased demand for education has been matched by a rise in the number of schools offering programs to teach and retrain workers. Consequently, the number of adults returning to complete their high school education has significantly increased. In 2019, around 1.2 million adults were pursuing high school degrees.

In Brazil, adult education represents a significant segment of the education system: in 2019, approximately 3.2 million students were enrolled in adult high school education (*Educação de Jovens e Adultos - EJA*) programs (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2020). Of these, approximately 1.2 million were pursuing high school completion through adult education, compared to 7.5 million students enrolled in regular high school education. Private institutions operate in both markets, accounting for roughly 12.5% of total high school enrollments and playing a substantial role in Brazil's adult education sector (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2020).

Brazil's private school sector is highly decentralized, with hundreds of small private schools just within Rio de Janeiro, the setting for our research. The regulations for establishing a private school are relatively lax. A school can operate by presenting a coherent curriculum plan and having at least one room available for classes. Accreditation is relatively easy to obtain; schools simply need to register the diplomas they award with the State Department and the Ministry of Education. Before the pandemic, schools were allowed to conduct up to 20% of their classes online.

2.1 Operação Nota Zero

G1 Rio (2018))

On the morning of September 24, 2018, 110 police officers from Rio de Janeiro conducted a major operation targeting 11 private schools, as well as the homes of their owners, directors, and staff, to collect further evidence regarding the issuance of fake diplomas. These schools allegedly sold more than 350,000 fake diplomas over five years to individuals seeking a high school degree without completing the required coursework. This operation was widely covered by local newspapers and TV stations, including the country's largest media conglomerate, *Grupo Globo*[1] [2]. The entire scheme is estimated to have generated up to R$700 million (approximately 130 million USD) over five years (see Figure A1).

The investigation was carried out by the Rio de Janeiro Civil Police in collaboration with the State Department of Education. It uncovered a market for buying and selling fake high school diplomas. These diplomas were issued by legally recognized institutions that had authorization from the Department of Education, despite the fact that the students had never attended a single class. In reality, these schools enrolled far more students than the maximum number authorized by the Rio de Janeiro State Council of Education (*Conselho Estadual de Educação - CEE*). For example, in 2015, one such school (*Instituto Educacional Luminis*) had 12,021 students enrolled - 120 times its authorized capacity of 100 students[3].

The process of buying a fake diploma was straightforward. Individuals would contact intermediaries, provide some personal information, and make a payment. This information was then forwarded to a school participating in the scheme, which would issue a diploma within days. The diploma falsely certified that the student attended at least 75% of the classes and achieved the necessary grades to pass.

A brief Google search in Portuguese on how to buy a diploma in Rio de Janeiro reveals numerous "diploma sellers" (Figure **??**). On one of these websites, the message is straightforward:

"If you don't have time and have not finished your studies, you can easily and quickly get the documentation with us. We have been working for years in this

[1] https://globoplay.globo.com/v/7039781/

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

[3] https://oglobo.globo.com/rio/fabrica-de-diplomas-escola-com-capacidade-para-100-alunos-ja-teve-12-mil-matriculados-23116831

business, issuing original documents certified by the Ministry of Education"

We reached out to one of these merchants by email to inquire about purchasing a diploma, as shown in Figure A2. In response, we received detailed information about the scheme: a high school diploma would cost R$600 (around 110 USD) and could be issued within five days.

Although similar schemes were found across the country, the investigation revealed that the State of Rio de Janeiro was a major hub for diploma-selling due to lax enforcement by its State Department of Education. Since purchasing a diploma only required providing personal information and making a payment, individuals from other states could easily obtain diplomas from Rio de Janeiro. The police investigation found that most of these individuals came from the states of Minas Gerais, Paraná, and Mato Grosso do Sul.

3 Theoretical Background

Signaling Models and Employer Learning

Economists have long been interested in what schools actually do. Becker's Human Capital (Becker, 1994) focused on how schools help students develop skills. Each lesson imbues the student with a form of human capital, a tangible skill that the individual could use to improve his/her economic situation. While some classes might teach a specific skill (e.g. computer programming), others might teach lessons on punctuality or critical thinking. All of these classes and training form the human capital portfolio that a person can then rent to employers for a wage.

The signaling model, in contrast, does not involve the acquisition of skills (Spence, 1973). According to this model, schools impose costs on students, with lower-ability individuals facing disproportionately large costs. These large costs discourage them from pursuing additional education. Moreover, the model assumes that individuals' abilities are unobservable to employers, who rely on the schooling process to help identify those with the highest abilities. Consider the following simple formulation on the model.

Suppose that the proportion of high types in the economy is given by p. Let w_h and w_l be the productivity of the high (h) and low (l) types. Let C_h and C_l be the cost of schooling in any given period for high and low types, respectively. For simplicity, we assume this to be constant in each period in which an individual attends school. We assume that $c_h < c_l$. Let r be the discount rate.

High types will get Q years of schooling if

$$\sum_{t=Q}^{T} \frac{w_h}{(1+r)^t} - \sum_{t=0}^{Q} \frac{C_h}{(1+r)^t} > \sum_{t=0}^{T} \frac{w_l}{(1+r)^t}.$$

Low types will not get any schooling so long as

$$\sum_{t=Q}^{T} \frac{w_h}{(1+r)^t} - \sum_{t=0}^{Q} \frac{C_l}{(1+r)^t} < \sum_{t=0}^{T} \frac{w_l}{(1+r)^t}.$$

If a Q exists such that both inequalities are satisfied, the model produces a separating equilibrium where people choose the appropriate level of schooling to signal their true ability.

The introduction of fake diplomas presents a new complication to the standard signaling model. Let R_i be the reputation cost for person i if they are found to have a fake diploma. We assume that this is normally distributed across the entire population. This reputation cost could be associated with legal consequences, the subsequent treatment by employers of someone who was caught with a fake diploma, and/or the psychological costs associated with dishonesty. A distribution of reputation ensures some variation in the likelihood that individuals, particularly low types, take up the diploma. Let D_i be the expected time under which the individual i can avoid detection. Once an individual's fake diploma is detected, their subsequent wage would be w_{lt} for the rest of their careers.

A high type would pursue a fake diploma if

$$\sum_{t=0}^{D} \frac{w_h}{(1+r)^t} + \sum_{t=D}^{T} \frac{w_l}{(1+r)^t} - R_i > \sum_{t=Q}^{T} \frac{w_h}{(1+r)^t} - \sum_{t=0}^{Q} \frac{C_h}{(1+r)^t}.$$

If $Q < D$, we can rewrite this as

$$\sum_{t=0}^{Q} \frac{w_h}{(1+r)^t} + \sum_{t=0}^{Q} \frac{C_h}{(1+r)^t} > \sum_{t=D}^{T} \frac{(w_h - w_l)}{(1+r)^t} + R_i.$$

If $Q > D$, we can rewrite this as

$$\sum_{t=0}^{D} \frac{w_h}{(1+r)^t} + \sum_{t=D}^{Q} \frac{w_l}{(1+r)^t} + \sum_{t=0}^{Q} \frac{C_h}{(1+r)^t} > \sum_{t=Q}^{T} \frac{(w_h - w_l)}{(1+r)^t} + R_i.$$ [4]

The left-hand sides reflect the benefits of working immediately and the cost savings of not going to school. The right-hand side has the penalties from reputation and the change in post-detection wages resulting from being treated as a low-type after detection.

For the low types, assuming that Q is such that they do not pursue schooling, they will pursue a fake diploma if

$$\sum_{t=0}^{D} \frac{w_h - w_l}{(1+r)^t} > R_i.$$

The key parameters are the size of the reputation cost R_i and the expected time to detection D. In a pure signaling model, productivity is assumed to be unobservable

[4] If employers learn and are willing to offer high ability type workers w_h if they detect a fake diploma, then the two equations simplify dramatically. Any penalty in future wages is fully captured by R_i and individuals gauge the earnings they would gain from immediate work and the savings in schooling costs relative to the penalty:

$$\sum_{t=0}^{Q} \frac{w_h}{(1+r)^t} + \sum_{t=0}^{Q} \frac{C_h}{(1+r)^t} > R_i.$$

by firms. However, there is a rich literature in economics that considers how employers (and employees) learn and how vacancies are created when expected worker productivity is lower than the outside options that both firms and workers have (see, e.g. Menzio and Shi, 2011, Jovanovic 1979).

Suppose that there is a constant probability of detection for each period λ. The probability of detection in period t is given by

$$h(\lambda, t) = \lambda(1-\lambda)^t$$

and the expected time to detection D is then $\frac{1}{\lambda}$.

What influences λ? In the search literature, a few characteristics may increase the likelihood of detection. First, in some jobs, it is easy to observe productivity ("inspection goods" in Menzio and Shi (2011)), while in other jobs, it is much harder ("experience goods"). For example, productivity might be easy to observe in many manufacturing settings, and hence λ is high. In many service industries, it can be more difficult to observe. Additionally, tenure at a firm or observable work histories could facilitate employers' ability to detect productivity. Other characteristics such as firm size can also be associated with differing values of λ.

What does this imply in our setting? First, if employees learn but λ is low, firms cannot initially distinguish the quality of a signal, we should still see an immediate return to a signal; however, as employers and employees learn, the premium for fake diplomas would fall.[5] The presence of an immediate return is reflective of a signaling model, and its fading out suggests that the market is learning. Second, the pattern in the long run does not help us distinguish between a signaling or a human capital model. It could be that the short-run signal in the absence of skills creates a wedge between returns to a skill-based diploma and one completely lacking in skills that grows over time. It could also be that the short-run signal becomes more powerful in the long run as "imperfect" or "flawed" signals are removed. Third, we should see heterogeneity according to the degree to which the

[5] Alternatively, you can think of this as purifying the signal that schooling provided by weeding out pretenders.

employer can see additional signals of a worker's productivity (e.g., job history within or across firms). In sectors, occupations, and firms where productivity is more easily observed, the market should adjust sooner, eliminating any signaling premium from a fake diploma.

4 Data and Empirical Strategy

This section describes the datasets, the construction of the main sample, and the empirical strategy used to estimate the labor market impacts of obtaining a fake versus a legitimate high school diploma.

4.1 Data Sources and Sample Construction

We combine three data sources to construct the analysis sample: (i) a series of administrative records of investigations about schools suspected to grant fake diplomas, (ii) the *Diário Oficial do Estado do Rio de Janeiro* (DOERJ), an official registry that publishes the names of all high school graduates in the state of Rio de Janeiro and (iii) the *Relação Anual de Informações Sociais* (RAIS), an employer-employee matched administrative dataset collected by the Brazilian Ministry of Labor, which covers the universe of formal labor contracts in Brazil from 1995 to 2020.

Identifying fake diplomas

We worked in collaboration with the Rio de Janeiro State Department of Education to identify schools that issued fake diplomas. We were granted access to administrative records of preliminary investigations conducted by the State Department of Education, the (public) police records, lawsuits, and all decisions made by the State Council of Education to initiate investigations into diploma-selling schemes or to revoke a school's authorization to operate. These documents provide detailed information on how each scheme was carried out, the irregularities discovered during inspections, as well as the school's name, address, whether it offered adult education, and its opening and closing dates. We identified 21 schools involved in diploma-selling schemes.

Figure A3 illustrates the mechanical nature of the fraud: Two diplomas issued to different students reproduce an identical grade table line-by-line, confirming that schools relied on templates rather than real academic records. Figure A4 complements this evidence on the supply side, showing a formal 'payment plan' that prices the fake credential at exactly R600 − *theamountquotedbyintermediaries* − *paidinsixmonthlyinstallmentsandevenincludesanadditionalfeeforpublicationintheDiário Oficial*.

After understanding the scheme and identifying the schools, we scraped data from the Official Gazette (*Diário Oficial*) of the State of Rio de Janeiro (*DOERJ*) to collect the full names of all high school graduates in the state between 2008 and 2015. These data include the graduating semester and year and the schools' names and addresses.[6]. There are 710,792 rows in the raw DOERJ dataset. As this dataset is derived from a web-scrapping procedure, some entries are fuzzy and do not represent actual individuals, but rather other parts of the DOERJ text. After removing these observations, we still have 683,810 observations. Next, we identify individuals who are real diploma holders and those who bought fake diplomas based on the names of the 21 schools which have been reported as sellers of fake adult high school diplomas. This step results in 549,195 real diploma holders and 134,651 fake diploma holders. Finally, we restrict the sample to those individuals to those who are uniquely identified by their full names within this sample, resulting in 608,934 diploma holders.

Our third data source is RAIS (*Relação Anual de Informações Sociais*), a comprehensive employer-employee administrative dataset from the Ministry of Labor that covers the universe of workers and firms in the formal labor market in Brazil[7]. The RAIS dataset contains detailed information on all formal employees, such as personal characteristics (age, gender, race, and educational attainment), contract details (wage, tenure, hiring data, and separation date), and employer characteristics (sector of activity, establishment size, and location). Each employee is uniquely identified by their tax registry number (CPF). We use RAIS to gather outcome data for recipients of both real and fake diplomas. We also use RAIS to

[6] According to the State of Rio de Janeiro´s Education Council, the names of all high school graduates must be published in the State´s Official Gazette

[7] RAIS data has been used in various studies on the Brazilian labor market, including Gerard and Gonzaga (2021), Ulyssea (2018)

construct a sample of comparable adults.

In order to match students and create a comparable set of adults, we create a dataset on all individuals in RAIS who satisfy four conditions: (i) their full name should be reported in RAIS (i.e. they worked formally after 2002, the year when this variable started to be reported), (ii) they formally worked at least once in one of the four Brazilian states where schools that sold diplomas were more common (i.e. RJ, MG, PR and MS), (iii) they were born after 1955, and (iv) they are uniquely identified by their full name within the sample of workers who satisfy (i)-(iii). [8] This step results in 32,022,752 formal workers who are uniquely identified by their full name.

Identifying control groups

We compare students who obtained fake diplomas with two groups. The first one is composed of adults who earned degrees from legitimate high schools during the same period. We used the same web scraping of high school graduation registries described above to gather names, graduation dates, and student identification numbers for adults who obtained diplomas from legitimate schools during the same period. In order to be included in our analysis, the individual must have appeared in RAIS at least once in the five years before graduation. We include only observations in a window of five years before and five years after obtaining the credential.

We construct a second comparison group of high school dropouts composed of adults who completed primary education (grade 9) but did not obtain a high school degree. For each individual with a false or a true diploma, we match him or her with ten high school dropouts of the same sex, who were born in the same year and in the same region[9] In the vast majority of the cases, when there are more than ten possible matches, we randomly select the matching 10.

[8]Conditions (i) and (iv) are important as we identify individuals across administrative registries based on their full name. Conditions (ii) and (iii) are employed to focus solely on the population who were more likely involved in the diploma buying scheme and who are old enough to have an observed work history before the acquisition of the adult high school diploma (i.e. approximately aged 25-45).

[9]Southeast for *Rio de Janeiro* and *Minas Gerais*, South for *Paraná* and Center West for *Mato Grosso do Sul*

For both groups, we apply the same selection criteria used to construct our sample of diploma buyers: individuals who appeared at least once in the five-year period before graduation (or the matched individual's graduation), were at least 25 years old in the graduation year, graduated in the 2000-2015 period, and were working in the states of *Rio de Janeiro*, *Minas Gerais*, *Paraná*, and *Mato Grosso do Sul*.

Table 1 presents summary statistics for matched controls, fake-diploma holders, and legitimate-diploma holders, measured five years before and after graduation. Both treated groups start with somewhat higher formal-employment rates and wages than controls, and these gaps grow in the post-period. Fake and legitimate diploma recipients are quite similar to each other across most variables, whereas their differences with controls are larger and widen slightly after graduation. These descriptive patterns motivate the more formal event-study results that follow.

4.2 Empirical strategy

We use a dynamic difference-in-differences specification with two-way fixed effects, and we test the robustness of our results using the estimator from (Abraham and Sun, 2021). For each labor market outcome of interest, we estimate two separate regressions on different samples, as follows:

$$y_{i,t} = \alpha_i^A + \alpha_t^A + \sum_{l=-5}^{-2} \theta_l^{\text{lead}}\, Legit_{i,t}^l + \sum_{l=0}^{5} \theta_l^{\text{lag}}\, Legit_{i,t}^l + \beta^A X_{i,t} + v_{i,t}^A \quad \text{(A)}$$

$$y_{i,t} = \alpha_i^B + \alpha_t^B + \sum_{l=-5}^{-2} \mu_l^{\text{lead}}\, Fake_{i,t}^l + \sum_{l=0}^{5} \mu_l^{\text{lag}}\, Fake_{i,t}^l + \beta^B X_{i,t} + v_{i,t}^B \quad \text{(B)}$$

Equation (A) estimates the dynamic effects of obtaining a legitimate diploma using the sample of treated individuals and their matched controls. Equation (B) conducts an analogous analysis for individuals who received a fake diploma, using a separately matched sample. Since the regressions are estimated on different samples, the fixed effects α_i and α_t, as well as the control coefficients β, are indexed to indicate that their estimated values may differ across specifications.

In both regressions, the dependent variable $y_{i,t}$ denotes the labor market outcome for individual i at time t. The variables $Legit^l_{i,t}$ and $Fake^l_{i,t}$ are event-time indicators equal to one if individual i is l periods away from obtaining a legitimate or a fake diploma, respectively, and zero otherwise. The period $l = -1$ is omitted and serves as the reference category. The vector $X_{i,t}$ contains a set of individual covariates (gender, race, age and state of residence) and a full set of RAIS calendar-year dummies, so the estimates absorb time-varying composition effects as well as year-specific shocks common to all workers.

The coefficients θ^{lag}_l in Equation (A) and μ^{lag}_l in Equation (B) are the parameters of interest. They capture the causal effects of obtaining a legitimate or a fake diploma on $y_{i,t}$ at event time l, relative to the period immediately preceding diploma issuance. The lead coefficients θ^{lead}_l and μ^{lead}_l serve as placebo tests and help assess the plausibility of the parallel trends assumption.

In the following, we present the estimated event-time coefficients along with their 95% confidence intervals in a series of event-study graphs. These plots allow us to examine the dynamics of the treatment effects when comparing individuals with true or fake diplomas to high school dropouts of the same age.

As high school dropouts tend to enter the labor market earlier, results involving employment probabilities or outcomes not conditioned on employment often a negative but improving employment probability prior to graduation. Nonetheless, both groups exhibit very similar pre-treatment trends.

We now describe the panel's structure and how it affects the interpretation of the estimated coefficients.

4.3 Panel Structure: Balanced and Unbalanced Samples

The event-study analyses presented are estimated under two distinct data settings: unbalanced and balanced panels.

By default, panel datasets constructed from RAIS are unbalanced. Individuals appear in the data only if they are formally employed in a given year. As a result, an individual's presence in the dataset may vary year to year, depending on their formal labor market status. This means that a given individual may enter or exit the sample multiple times over the observation window.

When estimating the dynamic regressions on this unbalanced panel, the coefficients should be interpreted as conditional on the individual being observed in the formal labor market in each period—that is, conditional on formal employment [10].

In addition to these conditional analyses, we also estimate the same regressions using a balanced panel, allowing us to assess the unconditional effects of diploma acquisition. To construct this panel, we include all individuals in the sample for each year within the observation window of five years before graduation and five years after, imputing zeros for employment and earnings in the years when an individual is not observed in RAIS. This transformation ensures that each individual has a complete panel of observations, regardless of their labor market status. The resulting specification allows us to capture extensive-margin outcomes, such as participation in the formal labor market and transitions into and out of employment.

5 Results

5.1 Employer-Reported Educational Attainment

The RAIS dataset contains information reported by firms on the education levels of their employees [11]. We can use this data to examine whether workers signal high school completion to employers and whether employers record this information as legitimate in the RAIS administrative database.

We use these reports to construct a measure of the employer's assessment of the

[10]According to PNADC (Continuous National Household Survey), 75% of the working individuals of that age group, with primary school complete, high school incomplete or high school complete are in the formal sector

[11]Firms do not frequently verify or update the educational information in RAIS, This information is likely more accurate when workers begin a new job or change occupations.

worker's educational attainment—specifically, whether the employer classifies the worker as a high school graduate. This analysis allows us to assess the extent to which employers believe an employee has completed high school, potentially triggering a sequence of changes in employment outcomes.

Figure 1 shows the evolution of the likelihood of an individual being classified as having a high school diploma by their employer, conditional on being employed. We observe a sharp increase of 25 percentage points in this probability after obtaining the diploma, both for those who genuinely studied and for those who purchased a fake credential. [12]

We now move on to our main results, analyzing the effects of these diplomas on employment, wages, and labor market transitions.

5.2 Employment and wages

In this section, we analyze labor market results for our sample using a balanced panel, constructed as previously described. The left panel of Figure 2 illustrates the trajectory of formal employment probabilities for individuals who either complete an adult high school or acquire a fake diploma, compared to high school dropouts. During the five-year period prior to diploma issuance, two important patterns emerge: (i) the probability of employment for both real and fake diploma holders increases over time relative to high school dropouts, who tend to enter the labor market earlier (Ashenfelter, 1978), and (ii) the two groups of diploma holders exhibit similar pre-trends.

The five years following diploma issuance reveal distinct patterns. In the year the diploma is obtained, we observe a clear short-term gain for both groups—approximately a 2.5% increase in the probability of having a formal job. However, these gains diverge over time: individuals with fake credentials see the initial gain stabilize and then decline. In contrast, those who genuinely completed an adult high school continue to experience sustained improvements in formal labor market participation, with gains ranging from a 5% to 7.5% increase in employment probabilities.

[12] Many workers were already reporting having a high school diploma despite not having received a diploma.

Panel B of the same figure presents our first results in terms of earnings, using the same sample of individuals. The pattern closely mirrors that of employment: due to our imputation of zero wages for those not observed in the formal labor market in a given year (as described in Section 4), the extensive margin dominates the overall earnings trajectory. Following a similar initial increase, earnings gains for fake diploma holders gradually fade, while individuals with real diplomas sustain and even expand their gains over time.

The next natural question then arises: conditional on having a formal job, how do individuals who completed adult high school compare to those who acquired a fake diploma?

Figure 3 provides insight into this. In the left panel, conditional on having a formal job, both groups exhibit positive wage trajectories in the initial years following diploma issuance, with a slightly steeper path for individuals who genuinely completed high school. Two years after graduation (or diploma acquisition), both groups maintain similar gains: approximately 10.5% for real diploma holders and 8% for fake ones, with both 95% confidence intervals including the 9% mark. From that point onward, a divergence emerges: earnings gains for fake diploma holders plateau, while gains for true graduates continue to grow. This divergence stabilizes at a 5-percentage point difference, with real graduates achieving a 15% gain relative to the control group, compared to 10% for fake diploma holders.

Panel B, on the right, reveals an interesting pattern: fake diploma holders are as likely to hold positions in occupations that require a high school credential. This suggests that, among those who maintain formal employment, individuals with fake diplomas are relatively successful in securing jobs consistent with their (fake) credentials, even if they do not keep pace with the wage growth of those with genuine diplomas.

In summary, both types of degrees provide positive benefits in the first few years after diploma issuance. However, the benefits decline over time for fake diploma holders: some lose their jobs, while others remain employed but fail to keep pace with the wage and career trajectories of legitimate diploma holders. As we discuss below, these results align with a model where the signaling value of the credential erodes over time. In equilibrium, employers eventually learn that some workers lack the skills required to perform their jobs. In the next section, we show that the pattern of job turnover supports this interpretation.

5.3 Transitions

The left panel of Figure 4 plots the probability of having a different formal job compared to the previous year. Individuals who will eventually obtain either a legitimate or a fake diploma exhibit similar transition probabilities prior to diploma issuance. However, those with fake diplomas show a significantly higher probability of changing jobs immediately after obtaining the credential, a pattern that persists for up to three years. A similar dynamic is observed for occupational changes (right panel), although with more nuance. This is consistent with individuals using fake diplomas as a signal in the job market, particularly when targeting new employers who have less information about them. The effects are substantial: in the first year after obtaining the diploma, fake diploma holders are six percentage points more likely to change jobs than those who graduated from legitimate adult high schools.

We have shown in the previous section that a significant share of workers with fake diplomas is unable to maintain the same level of labor market participation as their counterparts with legitimate credentials. Combined with the findings of a higher probability of job switching for fake-diploma holders, this suggests the existence of a group of workers who attempt to "fake it" but ultimately fail to sustain stable employment.

We hypothesize that many of these individuals continue to rely on their fake credentials to access new jobs over time, resulting in persistently higher labor market turnover. To test this, we estimate a regression using our balanced panel, where the dependent variable is a dummy that equals one if the individual is either non-formally employed or has changed formal jobs more than three times during the observation period.

Figure 5 presents the corresponding results. It plots the accumulated probability of experiencing high job turnover or unemployment over five years after diploma issuance. We find that individuals with fake diplomas are approximately ten percentage points more likely to fall into this high-turnover category than those with genuine diplomas.

This pattern supports our interpretation of a signaling breakdown: workers with fake credentials may initially succeed in securing jobs, but their lack of underlying skills leads to repeated separations. However, it is unclear whether the separations result from employers learning about workers' productivity or from workers realizing that their skills do not match and subsequently leaving a particular job. We explore this further in the next section.

6 Implications for signaling and learning models

We have established that fake diplomas result in a significant short-term boost in employment and wages; however, some workers seem to be trapped in a high-turnover equilibrium. By contrast, the effects of genuine diplomas lead to sustained increases in both employment and earnings.

We now focus on how and why the boost increased, as well as the extent to which this boost reflects learning in the market by both employers and employees. We will achieve this by analyzing the heterogeneity in our results, as described below.

6.1 Employer learning

The employer learning literature shows that firms statistically discriminate based on schooling credentials when hiring but update their beliefs about worker productivity over time as they observe on-the-job performance. (Lange, 2007), for example, estimates that employers reduce their initial expectation errors about worker productivity by 50% within three years. Motivated by this framework, we split our sample into two groups: (i) workers employed at the same firm for at least three years at the time of diploma issuance, which we name *attached workers*, and (ii) *non-attached workers* with shorter tenure with their employers.

The left panel of Figure 6 displays interesting patterns in employment probabilities. Among attached workers—whose employers have accumulated substantial information about their productivity—both legitimate and fake diploma holders show similar increases in formal employment relative to the control group. This aligns with the idea that the diploma provides limited new information when the employer already knows the worker well.

In contrast, for non-attached workers, we observe a clear divergence in post-diploma trajectories. Those with legitimate diplomas experience persistent gains in formal employment probabilities, while those with fake diplomas display only short-term improvements that dissipate within a few years. This pattern aligns with a learning model where employers initially respond to the signal provided by the diploma but later revise their beliefs as the worker's true productivity is revealed.

While some fake diploma holders in the non-attached group eventually lose their jobs, others manage to remain employed. This raises the question of how wages evolve for those who retain employment. The right panel of Figure 6 provides insight into this question. Wage gains associated with diploma acquisition are larger for non-attached workers, reflecting the greater informational value of credentials when prior employer knowledge is limited. Within this group, individuals with legitimate diplomas experience substantially higher and more sustained wage growth than those with fake diplomas. This suggests that while the diploma serves as an initial signal for both groups, employers gradually uncover the true nature of fake diploma holders, limiting their wage growth over time.

For attached workers, wage effects are more modest, consistent with the idea that employers had already formed accurate beliefs about these workers' productivity before diploma acquisition. Nonetheless, wage gains are consistently higher for those with legitimate diplomas compared to fake diploma holders, reinforcing the enduring value of genuine credentials even after employer learning has occurred.

6.2 Job Ladders and the Value of Credentials

A growing body of literature in labor economics emphasizes that labor markets are segmented between occupations and firms that offer steep career and wage progression—so-called job ladders—and those that provide flat wage trajectories and limited upward mobility. In particular, (Cahuc et al., 2006) and (Lise et al., 2016) model labor markets as structured around heterogeneity in firm and job quality, where certain occupations offer substantial wage growth through tenure and on-the-job learning, while others do not.

Motivated by this framework, we classify occupations into two groups based on the average slope of wage growth with tenure, calculated from RAIS data. Occupations with steeper wage-tenure profiles are labeled as *high progression*, while those with flatter profiles are labeled *low progression*.This distinction reflects not only earnings dynamics but also the presence or absence of job ladders and internal career mobility opportunities within each occupational group.

Figure 7 shows how labor market returns to diploma acquisition (whether real or fake) differ sharply across these two occupational segments. A striking pattern emerges in both employment probabilities and conditional wages. In occupations characterized by high wage progression, individuals who obtain a legitimate diploma experience substantially larger and more persistent gains compared to any other group. Their employment and wage trajectories diverge significantly from those of fake diploma holders and all workers in low-progression occupations. This suggests that in job ladder environments, where career advancement heavily depends on productivity and learning, the labor market places significantly greater value on the human capital gained through education.

In contrast, for occupations with low wage progression—those closer to *dead-end jobs* [13]—the labor market reacts to diploma acquisition in a much more limited way. Gains in employment and wages are modest and remarkably similar regardless of whether the diploma is real or fake. This suggests that in these segments, the diploma functions more as a basic entry credential rather than a strong productivity signal, and the labor market has lower incentives and fewer mechanisms to differentiate between genuine and fraudulent credentials.

This heterogeneity is even more pronounced in wage outcomes. In high-progression occupations, the wage premium for legitimate diploma holders steadily increases over time, exceeding 20% after five years. Meanwhile, fake diploma holders—even

[13]as in the terminology of (Cahuc et al., 2006)

in high-progression occupations—fail to keep pace with their legitimate counterparts, as their wage trajectories flatten shortly after obtaining their diplomas. In low-progression occupations, all groups exhibit flatter wage trajectories, and the gap between real and fake credentials is minimal.

Overall, these results align with the job ladder models of (Cahuc et al., 2006) and (Lise et al., 2016), which predict that the informational role of credentials and the mechanisms of employer learning vary substantially based on the occupation's underlying structure. In jobs with steep wage progression and internal mobility, the value of a legitimate diploma is amplified, making the labor market more effective at distinguishing genuine human capital from mere credentials. Conversely, in low-progression occupations with limited advancement opportunities, the labor market's response to educational signals is weaker and less differentiated.

7 Conclusions

This paper investigates whether schooling primarily builds skills or serves as a signal of individuals' underlying abilities, using novel evidence from the adult education sector in Brazil. In 2018, a police operation uncovered a network of schools issuing fraudulent high school diplomas without providing any instruction. Exploiting this natural experiment, we compare labor market outcomes for individuals with fake diplomas to those with legitimate credentials and matched high school dropouts.

Our results show that initially, both fake and legitimate diploma holders experience similar earnings gains, suggesting a strong signaling effect. However, these gains for fake diploma recipients erode quickly over time. The early returns appear driven by job changes, during which employers are less informed about workers' productivity. In the longer run, individuals with fake diplomas experience greater volatility in formal employment, including a significantly higher likelihood of multiple job separations.

Overall, our findings provide strong support for the signaling model in the short run, while highlighting how the value of educational signals diminishes as employers learn about workers' true productivity. This erosion underscores the limits of purely signaling-based returns to education and the importance of actual skill acquisition in sustaining labor market outcomes.

References

Abraham, Sarah and Liyang Sun, "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*, 2021, *225* (2), 175–199.

Altonji, Joseph G. and Charles R. Pierret, "Employer Learning and Statistical Discrimination," *The Quarterly Journal of Economics*, 2001, *116* (1), 313–350.

Arcidiacono, Peter, Patrick Bayer, and Aurel Hizmo, "Beyond Signaling and Human Capital: Education and the Revelation of Ability," *American Economic Journal: Applied Economics*, October 2010, *2* (4), 76–104.

Arrow, Kenneth J, "Higher education as a filter," *Journal of Public Economics*, 1973, *2* (3), 193–216.

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

Ashenfelter, Orley, "Estimating the Effect of Training Programs on Earnings," *Review of Economics and Statistics*, 1978, *60* (1), 47–57.

Becker, Gary S, *Human capital: A theoretical and empirical analysis, with special reference to education*, University of Chicago press, 1994.

Bedard, Kelly, "Human capital versus signaling models: university access and high school dropouts," *Journal of Political Economy*, 2001, *109* (4), 749–775.

Bordon, Paola and Breno Braga, "Employer learning, statistical discrimination and university prestige," *Economics of Education Review*, 2020, *76*, 101968.

Cahuc, Pierre, Fabien Postel-Vinay, and Jean-Marc Robin, "Wage Bargaining with On-the-Job Search: Theory and Evidence," *Econometrica*, 2006, *74* (2), 323–364.

Clark, Damon and Paco Martorell, "The signaling value of a high school diploma," *Journal of Political Economy*, 2014, *122* (2), 282–318.

Ehrenberg, Ronald G, Robert S Smith, and Kevin F Hallock, *Modern labor economics: Theory and public policy*, 13th ed., Routledge, 2021.

G1 Rio, "Polícia do Rio faz operação de combate à emissão de diplomas escolares falsos," *G1*, September 2018.

Graetz, Georg, "On the interpretation of diploma wage effects estimated by regression discontinuity designs," *Canadian Journal of Economics/Revue canadienne d'Ã©conomique*, February 2021, *54* (1), 228–258.

Hansen, Anne Toft, Ulrik Hvidman, and Hans Henrik Sievertsen, "Grades and Employer Learning," *Journal of Labor Economics*, None 2024, *42* (3), 659–682.

Hungerford, Thomas and Gary Solon, "Sheepskin effects in the returns to education," *The Review of Economics and Statistics*, 1987, *69* (1), 175–177.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, "Censo Escolar da Educação Básica 2019: Resumo Técnico," Technical Report, INEP/MEC, Brasília 2020.

Jaeger, David A and Marianne E Page, "Degrees matter: New evidence on sheepskin effects in the returns to education," *The Review of Economics and Statistics*, 1996, *78* (1), 77–85.

Jepsen, Christopher, Peter Mueser, and Kenneth Troske, "Labor Market Returns to the GED Using Regression Discontinuity Analysis," *Journal of Political Economy*, 2016, *124* (3), 621–649.

Kahn, Lisa B., "Asymmetric Information between Employers," *American Economic Journal: Applied Economics*, 2013, *5* (4), 165–205.

Lange, Fabian, "The speed of employer learning," *Journal of Labor Economics*, 2007, *25* (1), 1–35.

Lise, Jeremy, Costas Meghir, and Jean-Marc Robin, "Matching, Sorting and Wages," *Review of Economic Dynamics*, 2016, *19*, 63–87. Special Issue in Honor of Dale Mortensen.

Mancino, M. Antonella, Leonardo Fabio Morales, and Diego F. Salazar, "Signaling worker quality in a developing country: Lessons from a certification program," CLEF Working Paper Series 69, Canadian Labour Economics Forum (CLEF), University of Waterloo 2024.

Mazrekaj, Deni and Sofie Cabus, "Does a High School Diploma Matter? Evidence Using Regression Discontinuity Design," 2019.

O Globo, "Operação enquadra escolas que emitiram 350 mil diplomas falsos e movimentaram R$ 700 milhões em 5 anos," *O Globo*, September 2018.

Pereira, Vitor, "Diagnóstico do abandono e da evasão escolar no Brasil," Artigo, Instituto Mobilidade e Desenvolvimento Social (IMDS) junho 2022. Série Artigos IMDS.

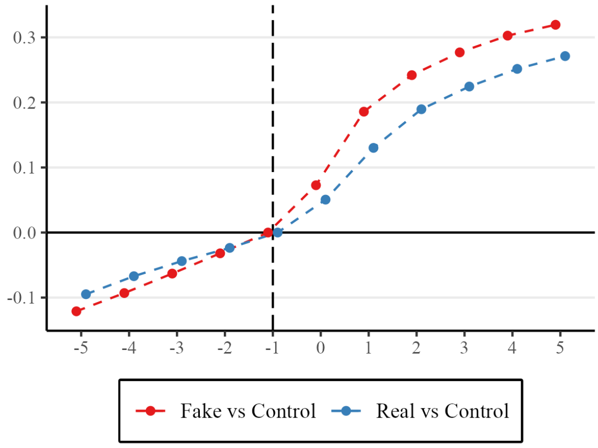
Spence, Michael, "Job market signaling," *The Quarterly Journal of Economics*, 1973, *87* (3), 355–374.

Tyler, John H, Richard J Murnane, and John B Willett, "Estimating the labor market signaling value of the GED," *The Quarterly Journal of Economics*, 2000, *115* (2), 431–468.

Weiss, Andrew, "Human capital vs. signalling explanations of wages," *Journal of Economic Perspectives*, 1995, *9* (4), 133–154.

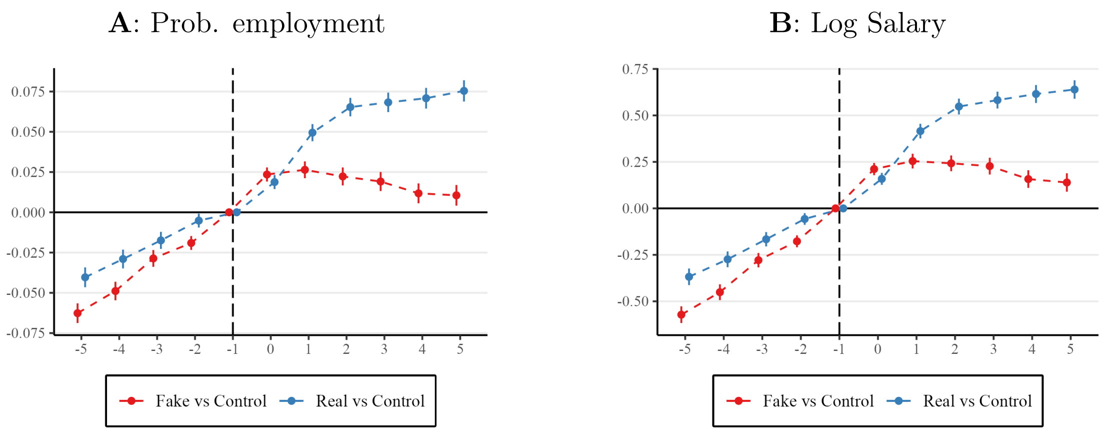
8 Figures

Figure 1: Probability of High School reported by employer



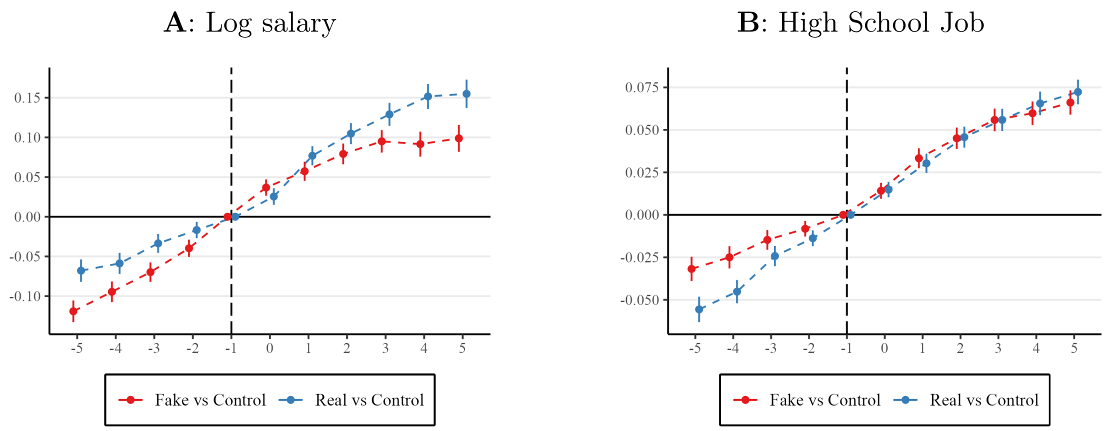
Note: This figure reports the dynamic event-study coefficients from Table 3 columns (5) and (6) for the two-way fixed-effects model, where we present the effects of having a fake high school diploma (Fake vs Control) and of having a legitimate high school diploma (Real vs Control). The outcome is the probability of having a high school diploma reported by the employer.

Figure 2: Unconditional effects on the probability of employment and wages (balanced panel)



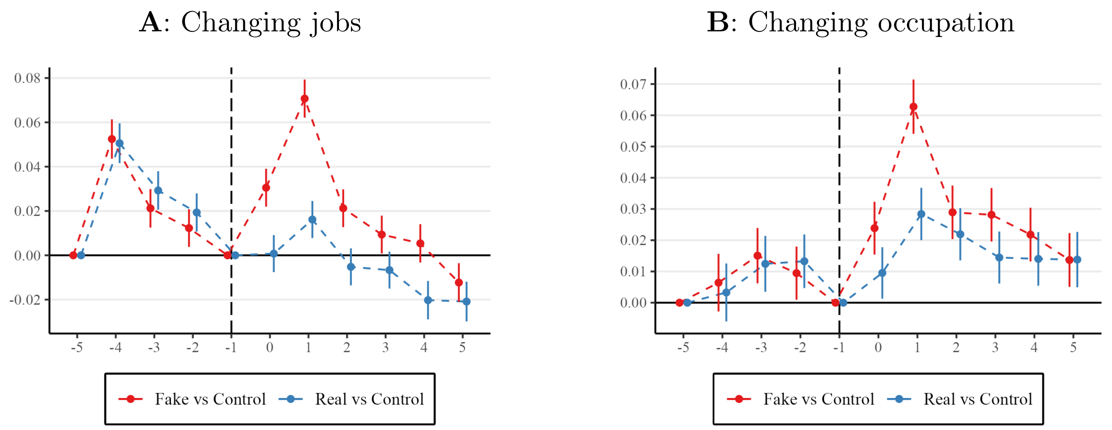
Note: This figure reports the dynamic event-study coefficients from Table 2 columns (1)-(4) for the two-way fixed-effects model, where we present the effects of having a fake high school diploma (Fake vs Control) and of having a legitimate high school diploma (Real vs Control). Panel A presents the effect on the probability of formal employment, while Panel B presents the effect on the Log of the monthly wage plus one, unconditional on formal employment.

Figure 3: Conditional effects on wages and high school jobs (unbalanced panel)



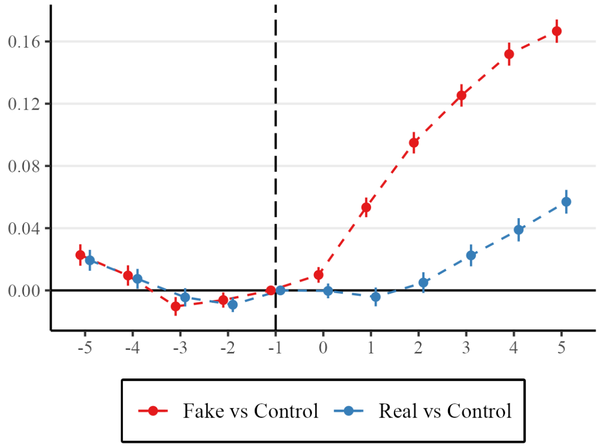
Note: This figure reports the dynamic event-study coefficients from Table 2 column (3)-(4) and Table 3 column (1)-(2) for the two-way fixed-effects model, where we present the effects of having a fake high school diploma (Fake vs Control) and of having a legitimate high school diploma (Real vs Control). Panel A presents the effect on the Log of the monthly salary plus one, conditional on formal employment, while Panel B reports the effect on the probability of being in an occupation that requires high school education.

Figure 4: Probability of changing jobs and occupation (balanced panel)



Note: This figure reports the dynamic event-study coefficients from Table 4 columns (3)-(6) for the two-way fixed-effects model, where we present the effects of having a fake high school diploma (Fake vs Control) and of having a legitimate high school diploma (Real vs Control). Panel A presents the effect on probability of changing jobs relative to the previous year, while Panel B reports the effect on the probability of changing occupations (4 digits of CBO) relative to the previous year.

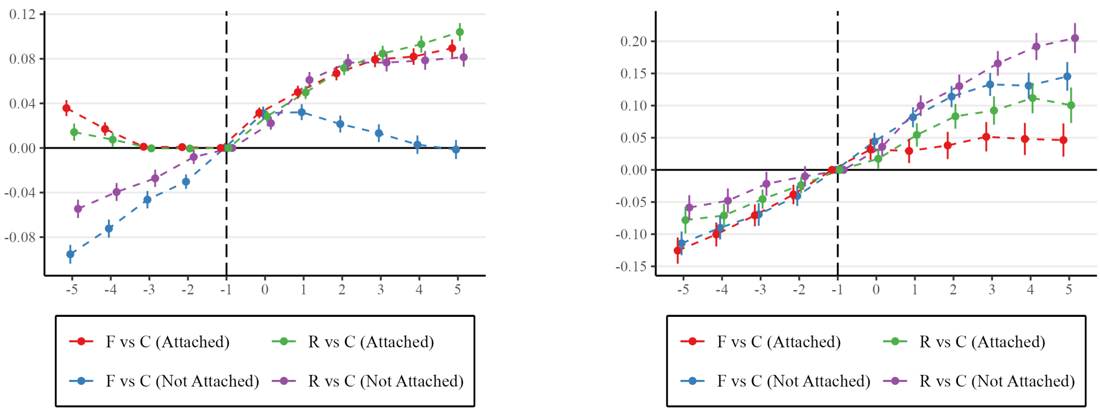
Figure 5: Probability of not being formally employed or being a frequent job switcher



Note: This figure reports the dynamic event-study coefficients from Table 5 columns (1)-(2) for the two-way fixed-effects model, where we present the effects of having a fake high school diploma (Fake vs Control) and of having a legitimate high school diploma (Real vs Control). The outcome is a dummy equal to one if the worker in not formally employed at a given year of if this worker is a frequent job switcher (i.e. has changed jobs at least 3 times since 5 years before graduating).

Figure 6: Heterogeneity by prior firm attachment

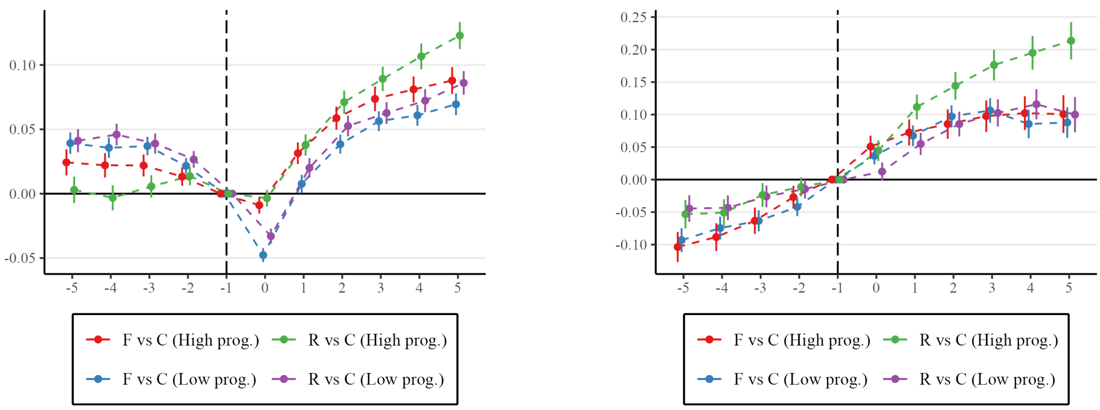
A: Employment probability **B**: Log wage (conditional)



Note: This figure reports the dynamic event-study coefficients for the two-way fixed-effects model, where we present the effects of having a fake high school diploma (Fake vs Control) and of having a legitimate high school diploma (Real vs Control) separately depending on prior firm attachment. A worker is considered to have prior firm attachment if she was employed by the same firm during the three years before graduation. Panel A reports the effect on the probability of formal employment, and Panel B reports the effect on the Log of the monthly wage plus one, conditional on formal employment.

Figure 7: Heterogeneity by mean occupation wage progression

A: Employment probability **B**: Log wage (conditional)



Note: This figure reports the dynamic event-study coefficients for the two-way fixed-effects model, where we present the effects of having a fake high school diploma (Fake vs Control) and of having a legitimate high school diploma (Real vs Control) separately depending on mean occupation wage progression. Panel A reports the effect on the probability of formal employment, and Panel B reports the effect on the Log of the monthly wage plus one, conditional on formal employment.

9 Tables

Table 1: Summary Statistics

	Mean					Difference			
	Controls	Fake		True		Fake - Controls		Fake - True	
Variable		Pre	Post	Pre	Post	Pre	Post	Pre	Post
Age	31.96	32.28	37.28	31.67	36.67	0.32	0.32	0.61	0.61
Male	0.61	0.69	0.69	0.53	0.53	0.08	0.08	0.15	0.15
White	0.55	0.65	0.65	0.59	0.59	0.10	0.10	0.06	0.06
Prob. employment	0.56	0.73	0.78	0.68	0.73	0.17	0.23	0.05	0.05
Num. contracts	0.66	0.96	1.06	0.87	0.94	0.30	0.41	0.09	0.13
Avg. salary	1599.50	2415.29	3066.98	2142.26	2679.13	815.79	1204.17	273.03	407.85
Prob. HS required	0.15	0.36	0.43	0.42	0.50	0.21	0.28	-0.06	-0.07
Prob. HS reported	0.02	0.43	0.73	0.48	0.72	0.41	0.71	-0.05	0.01
Tenure	42.52	37.50	39.51	37.04	44.06	-5.02	-13.93	0.46	-4.55
Log wage (cond.)	7.11	7.48	7.71	7.35	7.56	0.38	0.50	0.13	0.16
Prob. empl. 31/12	0.71	0.72	0.71	0.73	0.74	0.01	-0.02	-0.01	-0.03
Weekly hours	43.24	42.86	42.49	42.42	41.94	-0.38	-0.71	0.44	0.56
Prob. changed job	0.32	0.34	0.37	0.34	0.33	0.02	0.06	0.01	0.04
Prob. changed occup.	0.22	0.28	0.28	0.28	0.23	0.06	0.10	0.00	0.04
Prob. public sector job	0.03	0.04	0.06	0.06	0.09	0.01	0.03	-0.02	-0.03
Firm size (1-50)	0.56	0.46	0.33	0.46	0.37	-0.10	-0.15	0.00	-0.04
Firm size (50-250)	0.20	0.23	0.22	0.20	0.20	0.04	0.02	0.03	0.02
Firm size (>250)	0.25	0.31	0.45	0.34	0.43	0.06	0.14	-0.03	0.02
N individuals	616,739	29,887		31,799					
N observations	2,464,348	119,484	119,484	127,088	127,088				

Note: This table presents summary statistics for each group employed in the main analyses. Columns present the averages for each group, both 5 years before (Pre) and 5 years after (Post) graduation. Difference columns report regression coefficients for the differences in each variable across groups, using group dummies as explanatory variables. 'Age' is the age of the individual, 'Male' and 'White' are dummies equal to one if the individual is a male and white, respectively. 'Prob. employment' is a dummy equal to one if the individual is employed in RAIS at a given year, 'Num. contracts' reports the number of formal

Table 2: Summary statistics

	Controls		Fake		Real		Fake − Controls		Real − Controls		Fake − Real	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Age	31.96	36.96	32.28	37.28	31.67	36.67	0.32	0.32	-0.29	-0.29	0.61	0.61
Male	0.61	0.61	0.69	0.69	0.53	0.53	0.08	0.08	-0.07	-0.07	0.15	0.15
White	0.55	0.55	0.65	0.65	0.59	0.59	0.10	0.10	0.04	0.04	0.06	0.06
Prob. employment	0.56	0.56	0.73	0.78	0.68	0.73	0.17	0.23	0.12	0.17	0.05	0.05
Num. contracts	0.66	0.65	0.96	1.06	0.87	0.94	0.30	0.41	0.21	0.28	0.09	0.13
Avg. salary	1599.50	1882.82	2415.29	3086.98	2142.26	2679.13	815.79	1204.17	542.76	796.31	273.03	407.85
Prob. HS required	0.15	0.15	0.36	0.43	0.42	0.50	0.21	0.28	0.27	0.35	-0.06	-0.07
Prob. HS reported	0.02	0.03	0.43	0.73	0.48	0.72	0.41	0.71	0.46	0.69	-0.05	0.01
Tenure	42.52	53.44	37.50	39.51	37.04	44.06	-5.02	-13.93	-5.48	-9.38	0.46	-4.55
Log wage (cond.)	7.11	7.21	7.48	7.71	7.35	7.56	0.38	0.50	0.25	0.35	0.13	0.16
Prob. empl. 31/12	0.71	0.73	0.72	0.71	0.73	0.74	0.01	-0.02	0.02	0.01	-0.01	-0.03
Weekly hours	43.24	43.20	42.86	42.49	42.42	41.94	-0.38	-0.71	-0.82	-1.27	0.44	0.56
Prob. changed job	0.32	0.31	0.34	0.37	0.34	0.33	0.02	0.06	0.01	0.02	0.01	0.04
Prob. changed occup.	0.22	0.18	0.28	0.28	0.28	0.23	0.06	0.10	0.06	0.06	0.00	0.04
Prob. public sector job	0.03	0.03	0.04	0.06	0.06	0.09	0.01	0.03	0.03	0.06	-0.02	-0.03
Firm size (1-50)	0.56	0.48	0.46	0.33	0.46	0.37	-0.10	-0.15	-0.10	-0.11	0.00	-0.04
Firm size (50-250)	0.20	0.20	0.23	0.22	0.20	0.20	0.04	0.02	0.01	-0.01	0.03	0.02
Firm size (>250)	0.25	0.31	0.31	0.45	0.34	0.43	0.06	0.14	0.09	0.12	-0.03	0.02
N individuals	616,739		29,887		31,799							
N observations	2,464,348	2,464,348	119,484	119,484	127,088	127,088						

Note: This table presents summary statistics for each group employed in the main analyses. Columns (1)-(6) present the averages for each group, both 5 years before (Pre) and 5 years after (Post) graduation. Columns (7)-(12) report regression coefficients for the differences in each variable across groups, using group dummies as explanatory variables. 'Age' is the age of the individual, 'Male' and 'White' are dummies equal to one if the individual is a male and white, respectively, 'Prob. employment' is a dummy equal to one if the individual is employed in RAIS at a given year, 'Num. contracts' reports the number of formal labor contracts an individual had at a given year, 'Prob. HS required' is a dummy equal to one if the individual is employed at an occupation that requires high school education, 'Prob. HS reported' is a dummy equal to one if an employer reported that a worker had high school education, 'Log wage (cond.)' is the Log of the monthly salary plus one, 'Prob. empl. 31/12' is a dummy equal to one if a worker is employed at the end of the year at a given job, 'Prob. changed job' is a dummy equal to one if the worker has changed employers relative to the previous year (including going to non formal employment), 'Prob. changed occup.' is a dummy equal to one if the worker changed occupations (4 digits of the CBO) relative to the previous year, and 'Firm size (X-Y)' are dummies equal to one if the worker is employed at a firm with X-Y employees.

Table 3: Main results (TWFE) – Part 1/4

Dep. Var.:	Prob. employment		Log wage		Log wage (cond.)	
	Fake	Real	Fake	Real	Fake	Real
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
treat × dist = -5	-0.063***	-0.040***	-0.572***	-0.368***	-0.119***	-0.068***
	(0.003)	(0.003)	(0.023)	(0.023)	(0.007)	(0.007)
treat × dist = -4	-0.049***	-0.029***	-0.451***	-0.274***	-0.095***	-0.059***
	(0.003)	(0.003)	(0.022)	(0.022)	(0.007)	(0.007)
treat × dist = -3	-0.029***	-0.017***	-0.279***	-0.166***	-0.070***	-0.034***
	(0.003)	(0.003)	(0.020)	(0.020)	(0.006)	(0.006)
treat × dist = -2	-0.019***	-0.005**	-0.177***	-0.057***	-0.040***	-0.017***
	(0.002)	(0.002)	(0.016)	(0.016)	(0.006)	(0.005)
treat × dist = 0	0.024***	0.019***	0.211***	0.159***	0.037***	0.025***
	(0.002)	(0.002)	(0.017)	(0.016)	(0.005)	(0.005)
treat × dist = 1	0.026***	0.050***	0.254***	0.415***	0.057***	0.077***
	(0.003)	(0.003)	(0.020)	(0.020)	(0.006)	(0.006)
treat × dist = 2	0.022***	0.065***	0.242***	0.548***	0.079***	0.105***
	(0.003)	(0.003)	(0.022)	(0.022)	(0.007)	(0.007)
treat × dist = 3	0.019***	0.068***	0.227***	0.582***	0.095***	0.129***
	(0.003)	(0.003)	(0.023)	(0.023)	(0.007)	(0.007)
treat × dist = 4	0.012***	0.071***	0.158***	0.615***	0.091***	0.152***
	(0.003)	(0.003)	(0.024)	(0.025)	(0.008)	(0.008)
treat × dist = 5	0.011***	0.075***	0.139***	0.639***	0.099***	0.155***
	(0.003)	(0.003)	(0.025)	(0.025)	(0.009)	(0.009)
Fixed-effects						
Individual	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Dist. grad.	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	7,021,372	7,036,842	7,021,372	7,036,842	3,974,345	3,971,114
R^2	0.45387	0.45268	0.48065	0.47870	0.60474	0.60330

Note: This table reports the two-way fixed effects event-study results for the effect of having a fake (legitimate) high school diploma on the probability of formal employment ('Prob. employment'), on the Log of formal wage unconditional on employment ('Log wage'), and on the Log of formal wage conditional on formal employment ('Log wage (cond.)'). Columns (1), (3) and (5) report the results for fake diplomas and columns (2), (4) and (6) report the results for legitimate diplomas. We employ individual, calendar year, distance to graduation and age fixed-effects. Standard errors are clustered at the individual level. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Table 4: Main results (TWFE) – Part 2/4

Dep. Var.:	Prob. HS required		Prob. HS required (unc.)		Prob. HS reported	
	Fake	Real	Fake	Real	Fake	Real
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
treat × dist = -5	-0.032***	-0.056***	-0.061***	-0.080***	-0.121***	-0.095***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)
treat × dist = -4	-0.025***	-0.045***	-0.046***	-0.061***	-0.093***	-0.067***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
treat × dist = -3	-0.015***	-0.024***	-0.029***	-0.036***	-0.063***	-0.044***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
treat × dist = -2	-0.008***	-0.014***	-0.017***	-0.016***	-0.032***	-0.024***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
treat × dist = 0	0.014***	0.015***	0.024***	0.023***	0.073***	0.051***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
treat × dist = 1	0.033***	0.030***	0.045***	0.049***	0.186***	0.130***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
treat × dist = 2	0.045***	0.046***	0.054***	0.066***	0.242***	0.190***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)
treat × dist = 3	0.056***	0.056***	0.059***	0.071***	0.277***	0.224***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)
treat × dist = 4	0.060***	0.066***	0.055***	0.081***	0.303***	0.252***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)
treat × dist = 5	0.066***	0.072***	0.052***	0.083***	0.319***	0.271***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)
Fixed-effects						
Individual	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Dist. grad.	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	3,816,775	3,812,748	7,021,372	7,036,842	3,974,318	3,971,087
R^2	0.69709	0.70465	0.52848	0.53361	0.61949	0.63694

Note: This table reports the two-way fixed effects event-study results for the effect of having a fake (legitimate) high school diploma on the probability of having high school required conditional on formal employment ('Prob. HS required'), on the probability of having high school required unconditional on formal employment ('Prob. HS required (unc.)'), and on the probability of having high school reported ('Prob. HS reported'). Columns (1), (3) and (5) report the results for fake diplomas and columns (2), (4) and (6) report the results for legitimate diplomas. We employ individual, calendar year, distance to graduation and age fixed-effects. Standard errors are clustered at the individual level. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Table 5: Main results (TWFE) – Part 3/4

Dep. Var.:	Number of contracts		Prob. changed job		Prob. changed occup.	
	Fake	Real	Fake	Real	Fake	Real
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
treat × dist = -5	-0.107***	-0.057***				
	(0.006)	(0.005)				
treat × dist = -4	-0.080***	-0.037***	0.052***	0.051***	0.006	0.003
	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
treat × dist = -3	-0.050***	-0.007	0.021***	0.029***	0.015***	0.012***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.005)	(0.005)
treat × dist = -2	-0.028***	0.0009	0.012***	0.019***	0.009**	0.013***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
treat × dist = 0	0.081***	0.038***	0.030***	0.0008	0.024***	0.010**
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
treat × dist = 1	0.081***	0.085***	0.071***	0.016***	0.063***	0.028***
	(0.006)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
treat × dist = 2	0.056***	0.103***	0.021***	-0.005	0.029***	0.022***
	(0.006)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
treat × dist = 3	0.032***	0.100***	0.009**	-0.007	0.028***	0.014***
	(0.006)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
treat × dist = 4	0.015**	0.101***	0.005	-0.020***	0.022***	0.014***
	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)
treat × dist = 5	-0.0001	0.107***	-0.012***	-0.021***	0.014***	0.014***
	(0.006)	(0.006)	(0.004)	(0.005)	(0.004)	(0.005)
Fixed-effects						
Individual	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Dist. grad.	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	7,021,372	7,036,842	3,619,596	3,613,805	3,060,703	3,052,344
R^2	0.43144	0.42914	0.35154	0.35256	0.30364	0.30436

Note: This table reports the two-way fixed effects event-study results for the effect of having a fake (legitimate) high school diploma on the number of formal labor contracts ('Number of contracts'), on the probability of changing jobs ('Prob. changed job'), and on the probability of changing occupations ('Prob. changed occup.'). Columns (1), (3) and (5) report the results for fake diplomas and columns (2), (4) and (6) report the results for legitimate diplomas. We employ individual, calendar year, distance to graduation and age fixed-effects. Standard errors are clustered at the individual level. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Table 6: Main results (TWFE) – Part 4/4

Dep. Var.:	Prob. unempl/switcher		Prob. public sector		N. months employed	
	Fake	Real	Fake	Real	Fake	Real
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
treat × dist = -5	0.023***	0.019***	-0.003***	-0.027***	-0.755***	-0.607***
	(0.004)	(0.003)	(0.001)	(0.001)	(0.034)	(0.034)
treat × dist = -4	0.010***	0.007**	-0.003***	-0.026***	-0.596***	-0.447***
	(0.003)	(0.003)	(0.0009)	(0.001)	(0.032)	(0.032)
treat × dist = -3	-0.010***	-0.004	-0.002**	-0.017***	-0.352***	-0.288***
	(0.003)	(0.003)	(0.0008)	(0.001)	(0.029)	(0.029)
treat × dist = -2	-0.006**	-0.009***	-0.001*	-0.007***	-0.177***	-0.101***
	(0.002)	(0.002)	(0.0007)	(0.0008)	(0.023)	(0.023)
treat × dist = 0	0.010***	-0.0003	0.003***	0.005***	-0.081***	0.044*
	(0.003)	(0.002)	(0.0007)	(0.0008)	(0.024)	(0.023)
treat × dist = 1	0.053***	-0.004	0.008***	0.010***	0.150***	0.363***
	(0.003)	(0.003)	(0.0010)	(0.001)	(0.030)	(0.029)
treat × dist = 2	0.095***	0.005	0.011***	0.014***	0.210***	0.560***
	(0.004)	(0.003)	(0.001)	(0.001)	(0.032)	(0.032)
treat × dist = 3	0.125***	0.022***	0.013***	0.019***	0.127***	0.648***
	(0.004)	(0.004)	(0.001)	(0.001)	(0.034)	(0.034)
treat × dist = 4	0.152***	0.039***	0.015***	0.023***	0.071**	0.677***
	(0.004)	(0.004)	(0.001)	(0.001)	(0.035)	(0.036)
treat × dist = 5	0.167***	0.057***	0.016***	0.024***	0.013	0.695***
	(0.004)	(0.004)	(0.001)	(0.001)	(0.037)	(0.037)
Fixed-effects						
Individual	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Dist. grad.	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	7,021,372	7,036,842	7,021,372	7,036,842	7,021,372	7,036,842
R^2	0.39993	0.39947	0.79111	0.78870	0.55209	0.55161

Note: This table reports the two-way fixed effects event-study results for the effect of having a fake (legitimate) high school diploma on the probability of non-employment or being a frequent job switcher ('Prob. unempl/switcher'), on the probability of having a public sector job ('Prob. public sector'), and on the number of months a worker is employed at a given year ('N. months employed'). Columns (1), (3) and (5) report the results for fake diplomas and columns (2), (4) and (6) report the results for legitimate diplomas. We employ individual, calendar year, distance to graduation and age fixed-effects. Standard errors are clustered at the individual level. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Table 7: Robustness (Sun & Abraham): Part 1/4

Dependent Variables:	Prob. employment		Log wage		Log wage (Cond.)	
	Fake	Real	Fake	Real	Fake	Real
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
year = -5	-0.063***	-0.040***	-0.573***	-0.365***	-0.119***	-0.068***
	(0.003)	(0.003)	(0.023)	(0.023)	(0.007)	(0.007)
year = -4	-0.049***	-0.029***	-0.451***	-0.273***	-0.094***	-0.059***
	(0.003)	(0.003)	(0.022)	(0.022)	(0.007)	(0.007)
year = -3	-0.029***	-0.017***	-0.279***	-0.166***	-0.070***	-0.034***
	(0.003)	(0.003)	(0.020)	(0.020)	(0.006)	(0.006)
year = -2	-0.019***	-0.005**	-0.177***	-0.057***	-0.040***	-0.017***
	(0.002)	(0.002)	(0.016)	(0.016)	(0.006)	(0.005)
year = 0	0.024***	0.019***	0.211***	0.159***	0.037***	0.026***
	(0.002)	(0.002)	(0.017)	(0.016)	(0.005)	(0.005)
year = 1	0.026***	0.050***	0.254***	0.416***	0.058***	0.077***
	(0.003)	(0.003)	(0.020)	(0.020)	(0.006)	(0.006)
year = 2	0.022***	0.065***	0.242***	0.548***	0.080***	0.105***
	(0.003)	(0.003)	(0.022)	(0.022)	(0.007)	(0.007)
year = 3	0.019***	0.068***	0.227***	0.582***	0.096***	0.128***
	(0.003)	(0.003)	(0.023)	(0.023)	(0.007)	(0.007)
year = 4	0.011***	0.070***	0.152***	0.609***	0.093***	0.152***
	(0.003)	(0.003)	(0.024)	(0.025)	(0.008)	(0.008)
year = 5	0.011***	0.076***	0.142***	0.646***	0.100	0.151***
	(0.003)	(0.003)	(0.025)	(0.026)	(1.02)	(0.009)
Fixed-effects						
Individual	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Dist. grad.	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	7,021,372	7,036,842	7,021,372	7,036,842	3,974,345	3,971,114
R^2	0.45389	0.45273	0.48066	0.47875	0.60475	0.60333

Note: This table reports the Sun & Abraham event-study results for the effect of having a fake (legitimate) high school diploma on the probability of formal employment ('Prob. employment'), on the Log of formal wage unconditional on employment ('Log wage'), and on the Log of formal wage conditional on formal employment ('Log wage (cond.)'). Columns (1), (3) and (5) report the results for fake diplomas and columns (2), (4) and (6) report the results for legitimate diplomas. We employ individual, calendar year, distance to graduation and age fixed-effects. Standard errors are clustered at the individual level. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Table 8: Robustness (Sun & Abraham): Part 2/4

Dep. Var.:	Prob. HS required		Prob. HS required (unc.)		Prob. HS reported	
	Fake	Real	Fake	Real	Fake	Real
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
year = -5	-0.032***	-0.056***	-0.061***	-0.079***	-0.121***	-0.094***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)
year = -4	-0.025***	-0.045***	-0.046***	-0.061***	-0.093***	-0.067***
	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
year = -3	-0.015***	-0.024***	-0.029***	-0.036***	-0.063***	-0.044***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
year = -2	-0.008***	-0.014***	-0.017***	-0.016***	-0.032***	-0.024***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
year = 0	0.014***	0.015***	0.024***	0.023***	0.073***	0.051***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
year = 1	0.033***	0.030***	0.045***	0.049***	0.186***	0.130***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
year = 2	0.045***	0.046***	0.054***	0.066***	0.243***	0.190***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)
year = 3	0.056***	0.056***	0.059***	0.071***	0.278***	0.225***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)
year = 4	0.060***	0.066***	0.056***	0.082***	0.304***	0.255***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)
year = 5	0.067	0.073***	0.053***	0.085***	0.321**	0.276***
	(0.229)	(0.004)	(0.003)	(0.003)	(0.148)	(0.004)
Fixed-effects						
Individual	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Dist. grad	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	3,816,775	3,812,748	7,021,372	7,036,842	3,974,318	3,971,087
R^2	0.69712	0.70470	0.52855	0.53376	0.61977	0.63748

Note: This table reports the Sun & Abraham event-study results for the effect of having a fake (legitimate) high school diploma on the probability of having high school required conditional on formal employment ('Prob. HS required'), on the probability of having high school required unconditional on formal employment ('Prob. HS required (unc.)'), and on the probability of having high school reported ('Prob. HS reported'). Columns (1), (3) and (5) report the results for fake diplomas and columns (2), (4) and (6) report the results for legitimate diplomas. We employ individual, calendar year, distance to graduation and age fixed-effects. Standard errors are clustered at the individual level. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Table 9: Robustness (Sun & Abraham): Part 3/4

Dep. Var.:	Number of contracts		Prob. changed job		Prob. changed occup.	
	Fake	Real	Fake	Real	Fake	Real
Model:	(1)	(2)	(4)	(5)	(7)	(8)
Variables						
year = -5	-0.107***	-0.056***				
	(0.006)	(0.005)				
year = -4	-0.080***	-0.036***	0.052***	0.051***	0.006	0.003
	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
year = -3	-0.050***	-0.007	0.021***	0.030***	0.015***	0.012***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.005)	(0.005)
year = -2	-0.028***	0.0009	0.012***	0.020***	0.009**	0.013***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
year = 0	0.081***	0.038***	0.031***	0.0007	0.024***	0.009**
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
year = 1	0.082***	0.085***	0.071***	0.016***	0.063***	0.028***
	(0.006)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
year = 2	0.057***	0.103***	0.021***	-0.005	0.029***	0.021***
	(0.006)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
year = 3	0.033***	0.100***	0.010**	-0.007	0.029***	0.014***
	(0.006)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
year = 4	0.014**	0.100***	0.006	-0.019***	0.022***	0.014***
	(0.006)	(0.006)	(0.004)	(0.005)	(0.004)	(0.005)
year = 5	0.0006	0.110***	-0.011**	-0.018***	0.015***	0.015***
	(0.006)	(0.006)	(0.004)	(0.005)	(0.004)	(0.005)
Fixed-effects						
Individual	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Dist. grad.	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	7,021,372	7,036,842	3,619,596	3,613,805	3,060,703	3,052,344
R^2	0.43149	0.42919	0.35161	0.35263	0.30377	0.30446

Note: This table reports the Sun & Abraham event-study results for the effect of having a fake (legitimate) high school diploma on the number of formal labor contracts ('Number of contracts'), on the probability of changing jobs ('Prob. changed job'), and on the probability of changing occupations ('Prob. changed occup.'). Columns (1), (3) and (5) report the results for fake diplomas and columns (2), (4) and (6) report the results for legitimate diplomas. We employ individual, calendar year, distance to graduation and age fixed-effects. Standard errors are clustered at the individual level. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Table 10: Robustness (Sun & Abraham): Part 4/4

Dep. Var.:	Prob. unempl/switcher		Prob. public sector		N. months employed	
	Fake	Real	Fake	Real	Fake	Real
Model:	(1)	(2)	(4)	(5)	(7)	(8)
Variables						
year = -5	0.023***	0.019***	-0.003***	-0.027***	-0.756***	-0.605***
	(0.004)	(0.003)	(0.001)	(0.001)	(0.034)	(0.034)
year = -4	0.010***	0.007**	-0.003***	-0.026***	-0.596***	-0.446***
	(0.003)	(0.003)	(0.0009)	(0.001)	(0.032)	(0.032)
year = -3	-0.010***	-0.005	-0.002**	-0.017***	-0.353***	-0.287***
	(0.003)	(0.003)	(0.0008)	(0.001)	(0.029)	(0.029)
year = -2	-0.006**	-0.009***	-0.001*	-0.007***	-0.177***	-0.101***
	(0.002)	(0.002)	(0.0007)	(0.0008)	(0.023)	(0.023)
year = 0	0.010***	-0.0003	0.003***	0.005***	-0.081***	0.043*
	(0.003)	(0.002)	(0.0007)	(0.0008)	(0.024)	(0.023)
year = 1	0.053***	-0.004	0.008***	0.010***	0.150***	0.363***
	(0.003)	(0.003)	(0.0010)	(0.001)	(0.030)	(0.029)
year = 2	0.095***	0.005	0.011***	0.014***	0.211***	0.560***
	(0.004)	(0.003)	(0.001)	(0.001)	(0.032)	(0.032)
year = 3	0.126***	0.022***	0.013***	0.019***	0.128***	0.648***
	(0.004)	(0.004)	(0.001)	(0.001)	(0.034)	(0.034)
year = 4	0.153***	0.040***	0.015***	0.023***	0.065*	0.663***
	(0.004)	(0.004)	(0.001)	(0.001)	(0.036)	(0.037)
year = 5	0.167***	0.057***	0.016***	0.025***	0.017	0.709***
	(0.004)	(0.004)	(0.001)	(0.002)	(0.037)	(0.038)
Fixed-effects						
Individual	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Dist. grad.	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	7,021,372	7,036,842	7,021,372	7,036,842	7,021,372	7,036,842
R^2	0.39996	0.39952	0.79112	0.78888	0.55211	0.55165

Note: This table reports the Sun & Abraham event-study results for the effect of having a fake (legitimate) high school diploma on the probability of non-employment or being a frequent job switcher ('Prob. unempl/switcher'), on the probability of having a public sector job ('Prob. public sector'), and on the number of months a worker is employed at a given year ('N. months employed'). Columns (1), (3) and (5) report the results for fake diplomas and columns (2), (4) and (6) report the results for legitimate diplomas. We employ individual, calendar year, distance to graduation and age fixed-effects. Standard errors are clustered at the individual level. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

10 Appendix

10.1 Constructing the data

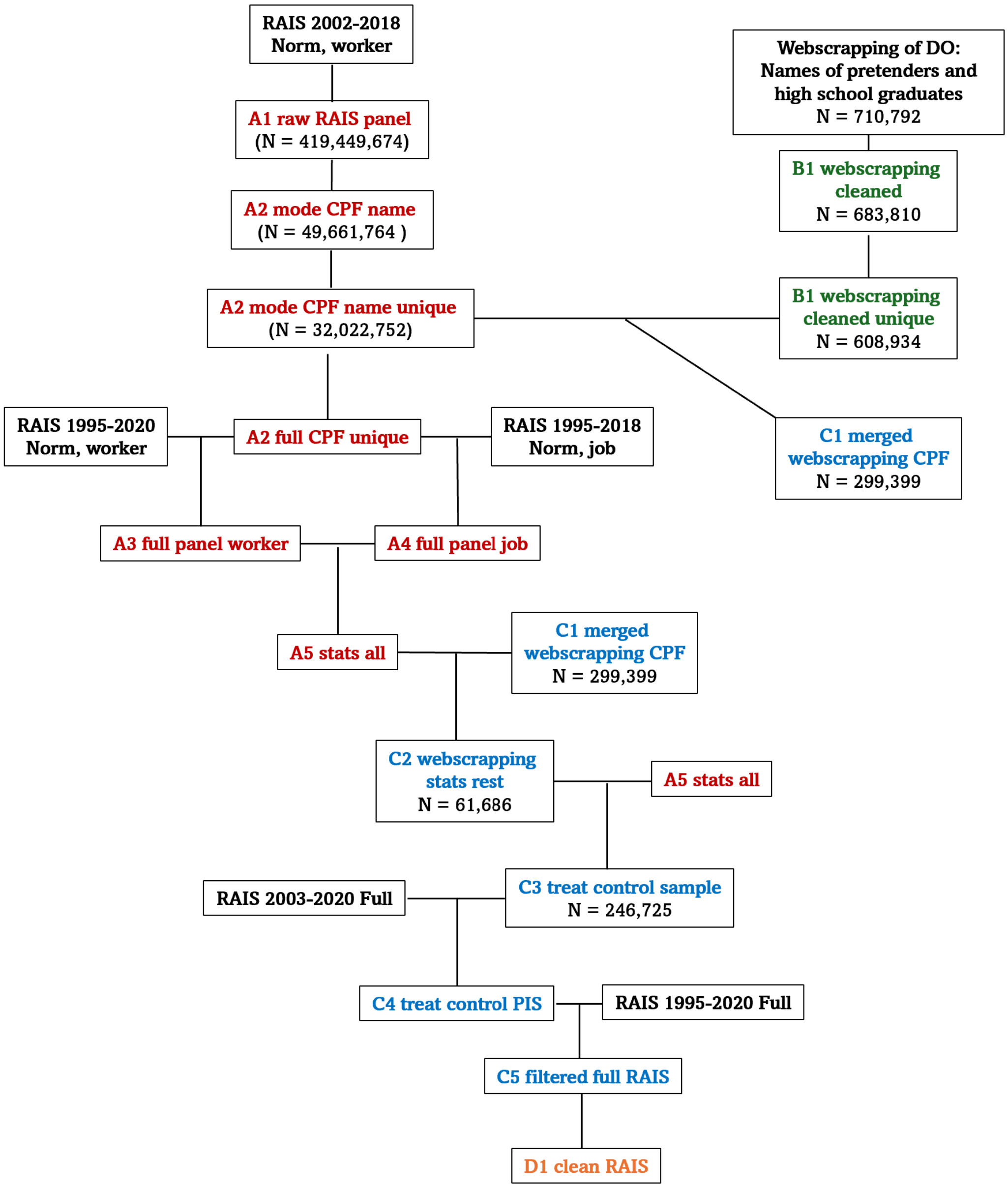
This appendix details the steps for the construction of the main sample. The two data sources are the Diário Oficial do Estado do Rio de Janeiro (DOERJ) cross-sectional registry and the Relação Anual de Informações Sociais (RAIS) panel registry. We use the DOERJ data to identify individuals who graduated from high school and their corresponding school names, which will be used to assign whether a graduate has a real or a fake diploma. The RAIS dataset allows us to follow the formal labor market trajectories of these high school diploma holders from 1995 to 2020. Importantly, both datasets have information about the full name, which is used in the merge procedure.

There are four main steps in the data construction process: Figure X1 illustrates the data construction process (the prefix X1 indicates the code where each step is conducted).

In the first step [A], we create a cross-section dataset derived from RAIS with information on all individuals who satisfy 4 conditions: (i) their full name should be reported in RAIS (i.e. they worked formally after 2002, the year when this variable started to be reported), (ii) they formally worked at least once in one of the four Brazilian states where schools that sold diplomas were more common (i.e. RJ, MG, PR and MS), (iii) they were born after 1955 and (iv) they are uniquely identified by their full name within the sample of workers who satisfy (i)-(iii). Conditions (i) and (iv) are important as we identify individuals across administrative registries based on their full name. Conditions (ii) and (iii) are employed to focus solely on the population who were more likely involved in the diploma buying scheme and who are old enough to have an observed work history before the acquisition of the adult high school diploma (i.e. approximately aged 25-45).

After filtering the RAIS 2002-2020 panel dataset to the 419,449,674 observations who satisfy conditions (i)-(iii) [A1], we calculate for each CPF number the mode of the full name variable. Although workers are uniquely identified by the Cadastro de

Figure 8: Diagram of steps for the data construction process



Pessoa Física (CPF) number, they are not uniquely identified by their full names. Additionally, there might be inconsistencies in the name variable across different years. Therefore, we calculate for each individual (identified by the CPF number) the mode of the name variable. This step results in 49,661,764 individual-level observations, which we further restrict to the 32,022,752 formal workers who are uniquely identified by the mode of their full name, therefore applying condition (iv) [A2]. At last, we filter RAIS 1995-2020 to these individuals once again and retrieve information on their gender, birth year, municipality, reported schooling [A3], employment, wage and tenure [A4], which are then summarized in the 'stats all' cross-sectional dataset [A5].

In the second step [B], we restrict the DOERJ administrative dataset to individuals who graduated from high school and are uniquely identified by their full names within this dataset. There are 710,792 rows in the raw DOERJ dataset. As this dataset is derived from a webscrapping procedure, some entries are fuzzy and do not represent actual individuals, but rather other parts of the DOERJ text. After removing observations with reported names such as 'aluno', 'diretor', 'modalidade' and 'cnpj', we still have 683,810 observations. Next, we identify individuals who are real diploma holders and those who bought fake diplomas based on the name of the school. There are 21 schools which have been reported as sellers of fake adult high school diplomas: Instituto Andreas Brunner, Centro Educacional Futura, Instituto Latino de Ciência e Tecnologia, Escola Triunfo, CECAMP, COBRA - Colégio Brasileiro, Colégio Joan Miro, CEC - Centro Carioca De Ensino, EPEC - AVM, Centro Educacional de Saude Ribeiro dos Santos, CEMOB, Colégio Jose Fonseca, Instituição De Ensino Sigma, Univerta – Sistemas Avançados de Educação, Colégio Estadual Frederico Azevedo, Centro de Formação Aplicação e Cultura, Centro Educacional Podio, Ciep Brizolão 119 - Austin, Sistema Augusto de Educação Integrada, Instituto Educacional Luminis, Nova Dimensão. This step results in 549,195 real diploma holders and 134,651 fake diploma holders. Finally, we restrict these individuals to those who are uniquely identified by their full names within this sample, resulting in 608,934 diploma holders [B1].

In the third step [C], we conduct three procedures: (C1) we merge the cross-sectional datasets from steps [A5] and [B1] using the full name, (C2) we filter the resulting dataset from (i) by key characteristics and (C3) we create a matched control group of individuals without high school diplomas, based on workers in the filtered dataset (C2).

In the first procedure [C1], we identify individuals with high school diplomas from the DOERJ registry (either fake or real) [B1] in the RAIS labor market registry [A5] using the full name variable. We find 299,399 individuals out of the 608,934 diploma holders. The accuracy of this merge procedure relies in the fact that we restricted both administrative datasets to individuals who are uniquely identified by their names within each dataset. However, an important caveat is that RAIS is not a universal registry of all Brazilians. Hence, it could be the case that there are 2 individuals with the same full name, but only one of them appears in the RAIS dataset, while only the other is in the DO dataset. In this case, we would perform an incorrect merge based on the full name, since both individuals are uniquely identified within their respective datasets.

In the second procedure [C2], we restrict the 299,399 diploma holders to those who (i) graduated from 2000 to 2015, (ii) were at least 25 years old at graduation and (iii) were formally employed in RAIS at least once in a 10-year window around the graduation year (either 5 years before or after this date). Restriction (iii) is applied to focus solely on individuals with some formal labor market attachment around the graduation year, while restriction (i) is employed to ensure that the 10-year window around graduation is feasible for all workers, as the RAIS dataset spans from 1995 to 2020. Finally, restriction (ii) implies that the resulting dataset refers to only adult high school students who had had a chance of prior work history before the having the diploma. Table 10 presents a balance check for the restriction procedure, where we apply each restriction separately in columns (2)-(4) and all restrictions at the same time in column (5). Only 61,686 out of the 299,399 diploma holders satisfy all restrictions. The share of individuals with fake diplomas increases from 17.4% to 48.5% once we apply all restrictions, such that 31,799 are real diploma holders and 29,887 are fake diploma holders.

In the third procedure [C3], we create a matched control group derived from [A5] of individuals who are similar in key characteristics to diploma holders [C2], but who

Table 11: Restricting the diploma holders dataset [C2]

	Full	Restricted by			Restricted	Diff.
		Grad. year	Age at grad.	Empl. window		
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Male	0.530	0.530	0.574	0.543	0.606	0.096
Birth year	1987.879	1987.646	1976.566	1987.098	1975.502	-15.588
Graduation year	2010.743	2009.946	2011.299	2010.130	2009.970	-0.973
Age at graduation	22.864	22.300	34.733	23.032	34.468	14.616
1(Fake diploma)	0.174	0.188	0.404	0.214	0.485	0.391
% High schooled (RAIS)	0.786	0.812	0.570	0.786	0.590	-0.247
Avg. salary	1898.089	1973.929	1919.692	1813.461	2043.473	186.031
Avg. tenure	17.776	18.124	28.652	18.735	30.749	16.600
Avg. salary 21-24 y.o.	1652.744	1696.171	1271.281	1608.473	1292.967	-428.049
Avg. tenure 21-24 y.o.	13.723	13.783	14.511	14.255	14.969	1.482
1(Empl in 1995)	0.113	0.103	0.556	0.132	0.676	0.595
1(Empl in 1996)	0.104	0.086	0.474	0.119	0.545	0.484
1(Empl in 1997)	0.115	0.104	0.447	0.139	0.505	0.444
1(Empl in 1998)	0.130	0.122	0.461	0.159	0.521	0.458
1(Empl in 1999)	0.141	0.133	0.454	0.168	0.507	0.445
1(Empl in 2000)	0.121	0.117	0.439	0.146	0.481	0.439
1(Empl in 2001)	0.127	0.125	0.494	0.161	0.529	0.495
1(Empl in 2002)	0.135	0.134	0.511	0.172	0.542	0.504
1(Empl in 2003)	0.158	0.157	0.554	0.199	0.580	0.538
1(Empl in 2004)	0.183	0.183	0.570	0.228	0.593	0.538
1(Empl in 2005)	0.200	0.200	0.585	0.248	0.607	0.535
1(Empl in 2006)	0.222	0.222	0.606	0.273	0.627	0.531
1(Empl in 2007)	0.257	0.257	0.629	0.313	0.651	0.519
1(Empl in 2008)	0.292	0.292	0.651	0.353	0.673	0.500
1(Empl in 2009)	0.314	0.314	0.656	0.376	0.678	0.478
1(Empl in 2010)	0.347	0.347	0.672	0.410	0.693	0.452
1(Empl in 2011)	0.388	0.391	0.683	0.461	0.710	0.417
1(Empl in 2012)	0.419	0.427	0.686	0.499	0.719	0.384
1(Empl in 2013)	0.455	0.466	0.689	0.539	0.724	0.341
1(Empl in 2014)	0.484	0.493	0.684	0.565	0.719	0.295
1(Empl in 2015)	0.496	0.502	0.664	0.570	0.704	0.255
1(Empl in 2016)	0.517	0.526	0.630	0.590	0.680	0.198
1(Empl in 2017)	0.517	0.527	0.584	0.581	0.639	0.146
1(Empl in 2018)	0.555	0.558	0.576	0.605	0.632	0.089
1(Empl in 2019)	0.090	0.099	0.122	0.106	0.177	0.098
1(Empl in 2020)	0.633	0.643	0.572	0.648	0.630	-0.003
N	299,399	265,266	78,737	224,160	61,686	

Note: This table details the process of restricting the diploma holders dataset to individuals who graduated from 2000 to 2015, who were at least 25 years old at the time of graduation and who were employed either 5 years before or 5 years after the graduation year. Column (1) presents the averages for all individuals who were matched with RAIS, ans columns (2)-(4) present the averages after applying each restriction separately. Column (5) reports the averages after applying all restrictions, and column (6) reports the differences between column (5) and column (1).

did not graduate from high school. We start by removing from [A5] individuals who are more likely to have a high school diploma. Thus, we drop those workers who appeared in the [C1] dataset (i.e. who have high school diplomas). Next, we calculate the share of RAIS observations where each individual had high school education reported by their employer and filter the dataset to those who had at most 10% observations with high school education reported. As a result, there are 5,632,275 candidates for the matching procedure.

For each worker among the 61,686 diploma holders, we match 10 control individuals by randomly selecting workers who share the same gender, birth year and mode of state (in RAIS 1995-2020), as well as the 10-year window around graduation dummy. Moreover, we conduct the matching procedure for each graduation year separately. Therefore, each control worker has a placebo graduation year, and there might be control workers who are matched to more than one real diploma holder. Table 11 presents a balance check for the real diploma holders and control individuals. There are 616,739 workers in the control group. Note that this is not precisely 10 times the number of real diploma holders since some gender – birth year – mode of state – 10-year window cells have less than 10 candidates for matched controls. After the matching procedure, we combine real diploma holders and controls into a single dataset and filter the RAIS 1995-2020 panel to these individuals [C4, C5].

Finally, the fourth step [D] consists of cleaning the resulting RAIS panel dataset of 678,425 workers. We set the panel dataset to the individual/year level by keeping only a worker's main job at a given year. Importantly, a worker may have more than one formal labor contract in the same year either because she had multiple jobs at the same time or because she changed employers in a given year. Therefore, we keep only the job that paid the highest amount of money for that worker by multiplying the average wage by the number of months the worker was employed at that firm. This results in an unbalanced panel with 7,818,852 individual/year observations spanning from 1995 to 2020, conditional on formal employment. Each observation reports the employer firm, the municipality of the firm, the firm sector (CNAE), the firm size, the worker's average wage at the firm, the occupation (CBO), whether this occupation requires high school education, the

Table 12: Balance check: Diploma holders and matched control group

Variable	Full (1)	Dip. holders (2)	Controls (3)	Diff. (4)
Graduation year	2009.97	2009.97	2009.97	0
Male	0.606	0.606	0.606	0
Birth year	1975.506	1975.502	1975.506	-0.004
Age at graduation	34.464	34.468	34.464	0.004
% High schooled (RAIS)	0.062	0.59	0.01	0.58***
Avg. salary	1436.275	2043.473	1375.473	667.999***
Avg. tenure	31.577	30.749	31.66	-0.91***
Avg. salary 21-24 y.o.	1110.839	1292.967	1087.808	205.159***
Avg. tenure 21-24 y.o.	15.326	14.969	15.371	-0.402***
1(Empl in 1995)	0.492	0.676	0.474	0.203**
1(Empl in 1996)	0.481	0.545	0.475	0.071
1(Empl in 1997)	0.469	0.505	0.466	0.039
1(Empl in 1998)	0.449	0.521	0.441	0.08***
1(Empl in 1999)	0.429	0.507	0.422	0.085***
1(Empl in 2000)	0.414	0.481	0.407	0.074***
1(Empl in 2001)	0.454	0.529	0.446	0.083***
1(Empl in 2002)	0.475	0.542	0.469	0.073***
1(Empl in 2003)	0.497	0.58	0.488	0.092***
1(Empl in 2004)	0.499	0.593	0.489	0.103***
1(Empl in 2005)	0.503	0.607	0.493	0.114***
1(Empl in 2006)	0.516	0.627	0.505	0.122***
1(Empl in 2007)	0.529	0.651	0.517	0.134***
1(Empl in 2008)	0.548	0.673	0.536	0.137***
1(Empl in 2009)	0.548	0.678	0.535	0.143***
1(Empl in 2010)	0.562	0.693	0.549	0.144***
1(Empl in 2011)	0.573	0.71	0.559	0.151***
1(Empl in 2012)	0.572	0.719	0.557	0.162***
1(Empl in 2013)	0.567	0.724	0.552	0.173***
1(Empl in 2014)	0.561	0.719	0.545	0.175***
1(Empl in 2015)	0.534	0.704	0.517	0.186***
1(Empl in 2016)	0.491	0.68	0.472	0.209******
1(Empl in 2017)	0.449	0.639	0.43	0.209***
1(Empl in 2018)	0.428	0.632	0.407	0.225***
1(Empl in 2019)	0.122	0.177	0.116	0.061***
1(Empl in 2020)	0.406	0.63	0.384	0.246***
N	678,425	61,686	616,739	

Note: This table presents the balance check between diploma holders and their matched control individuals. Column (1) reports the averages for all groups (controls and diploma holders), column (2) presents the averages only for diploma holders, and column (3) reports these averages only for the controls. Column (4) presents the differences between columns (2) and (3) derived from regressions where the explanatory variable is having a high school diploma.

worker's schooling as reported by the employer, the tenure at that firm, whether the contract was ended by the end of a given year and the separation reason, and the admission date.

Appendices

A Additional Figures

Figure A1: Newspaper headline on a large-scale fake-diploma scheme ("350 000 fake diplomas generated R $ 700 million in five years," O Globo)


globo.com | g1 | globoesporte | gshow | vídeos

O GLOBO RIO

350 mil diplomas falsos e movimentaram R$ 700 milhões em 5 anos

Note: Screenshot of a May 2019 article from the Rio edition of *O Globo*. The report describes a Civil Police operation that uncovered a nationwide network selling an estimated 350 000 forged school diplomas and transcripts between 2014 and 2019, generating roughly R$ 700 million (≈ US$ 175 million).

Figure A2: Inquiry and response from a forged-diploma provider

[redacted]@gmail.com> 17 de jun de 2019 15:20
para [redacted]

Boa tarde [redacted]
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 disponiveis, posso escolher? Aguardo resposta! Vai me ajudar muito!
abraços,

[redacted]@gmail.com> 18 de jun de 2019 10:28
para eu

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...

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 atende-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 papeis 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;

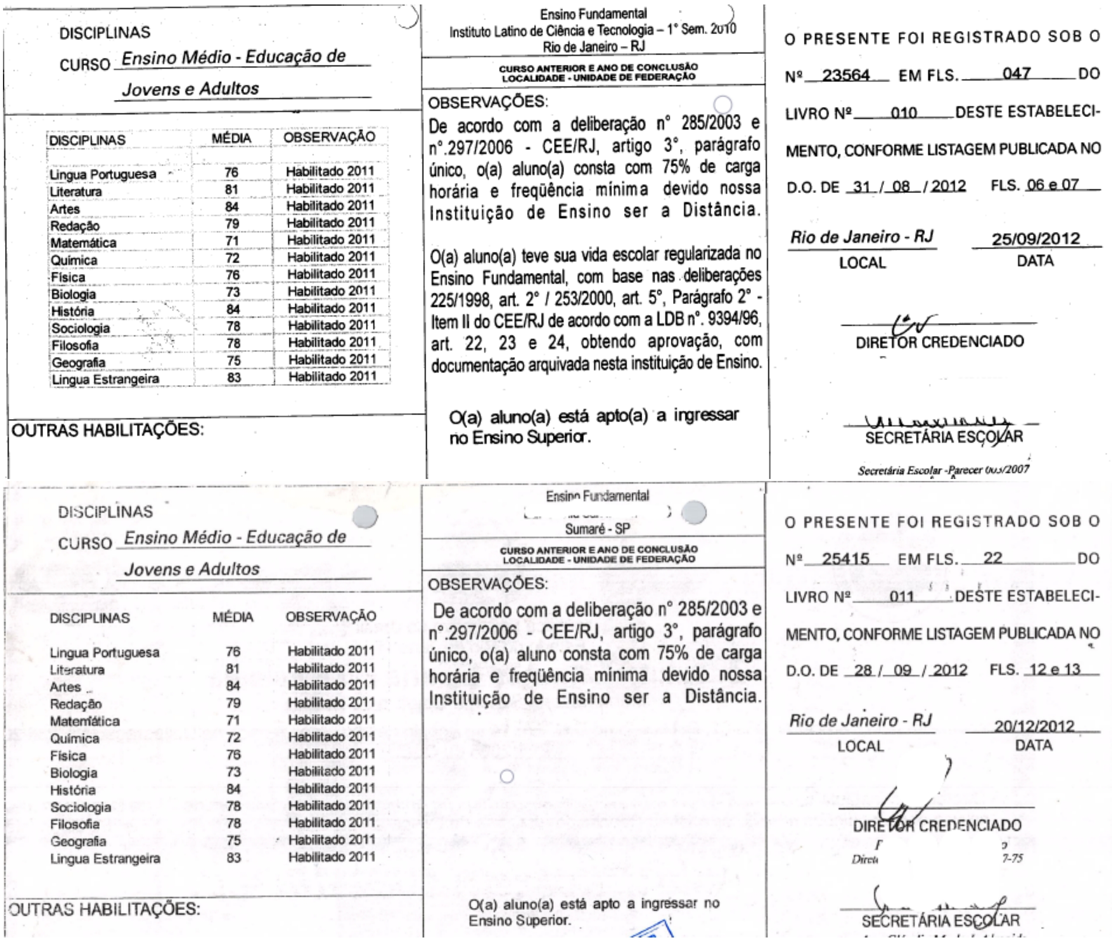
Ensino Médio R$ 600,00;

Ensino Fundamental + Ensino Médio R$ 800,00;

Curso Técnico ou Superior a consultar!

Note: The screenshots show the email sent by our research team asking how to "issue a high-school diploma in Rio," requesting delivery times and a list of available schools. The seller's reply offers forged documents for primary, secondary, technical, and even tertiary levels, claiming they are "original," registered with the Ministry of Education and published in the *Diário Oficial.* The vendor promises five-day delivery via courier, payment by bank transfer or deposit, "absolute confidentiality," and lists prices: R$ 600 for a high-school diploma, R$ 600 for a primary-school record, R$ 800 for both combined, with higher fees for college-level documents.

Figure A3: Two diplomas with the exact same grades



DISCIPLINAS

CURSO *Ensino Médio - Educação de Jovens e Adultos*

DISCIPLINAS	MÉDIA	OBSERVAÇÃO
Lingua Portuguesa	76	Habilitado 2011
Literatura	81	Habilitado 2011
Artes	84	Habilitado 2011
Redação	79	Habilitado 2011
Matemática	71	Habilitado 2011
Quimica	72	Habilitado 2011
Fisica	76	Habilitado 2011
Biologia	73	Habilitado 2011
História	84	Habilitado 2011
Sociologia	78	Habilitado 2011
Filosofia	78	Habilitado 2011
Geografia	75	Habilitado 2011
Lingua Estrangeira	83	Habilitado 2011

OUTRAS HABILITAÇÕES:

Ensino Fundamental
Instituto Latino de Ciência e Tecnologia – 1° Sem. 2010
Rio de Janeiro – RJ

CURSO ANTERIOR E ANO DE CONCLUSÃO
LOCALIDADE - UNIDADE DE FEDERAÇÃO

OBSERVAÇÕES:
De acordo com a deliberação n° 285/2003 e n°.297/2006 - CEE/RJ, artigo 3°, parágrafo único, o(a) aluno(a) consta com 75% de carga horária e freqüência 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. 2° / 253/2000, art. 5°, Parágrafo 2° - Item II do CEE/RJ de acordo com a LDB n°. 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.

O PRESENTE FOI REGISTRADO SOB O Nº 23564 EM FLS. 047 DO LIVRO Nº 010 DESTE ESTABELECIMENTO, CONFORME LISTAGEM PUBLICADA NO D.O. DE 31 / 08 / 2012 FLS. 06 e 07

Rio de Janeiro - RJ 25/09/2012
LOCAL DATA

DIRETOR CREDENCIADO

SECRETÁRIA ESCOLAR

Secretária Escolar - Parecer 003/2007

DISCIPLINAS

CURSO *Ensino Médio - Educação de Jovens e Adultos*

DISCIPLINAS	MÉDIA	OBSERVAÇÃO
Lingua 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
Lingua Estrangeira	83	Habilitado 2011

OUTRAS HABILITAÇÕES:

Ensino Fundamental
Sumaré - SP

CURSO ANTERIOR E ANO DE CONCLUSÃO
LOCALIDADE - UNIDADE DE FEDERAÇÃO

OBSERVAÇÕES:
De acordo com a deliberação n° 285/2003 e n°.297/2006 - CEE/RJ, artigo 3°, parágrafo único, o(a) aluno consta com 75% de carga horária e freqüência mínima devido nossa Instituição de Ensino ser a Distância.

O(a) aluno(a) está apto a ingressar no Ensino Superior.

O PRESENTE FOI REGISTRADO SOB O Nº 25415 EM FLS. 22 DO LIVRO Nº 011 DESTE ESTABELECIMENTO, CONFORME LISTAGEM PUBLICADA NO D.O. DE 28 / 09 / 2012 FLS. 12 e 13

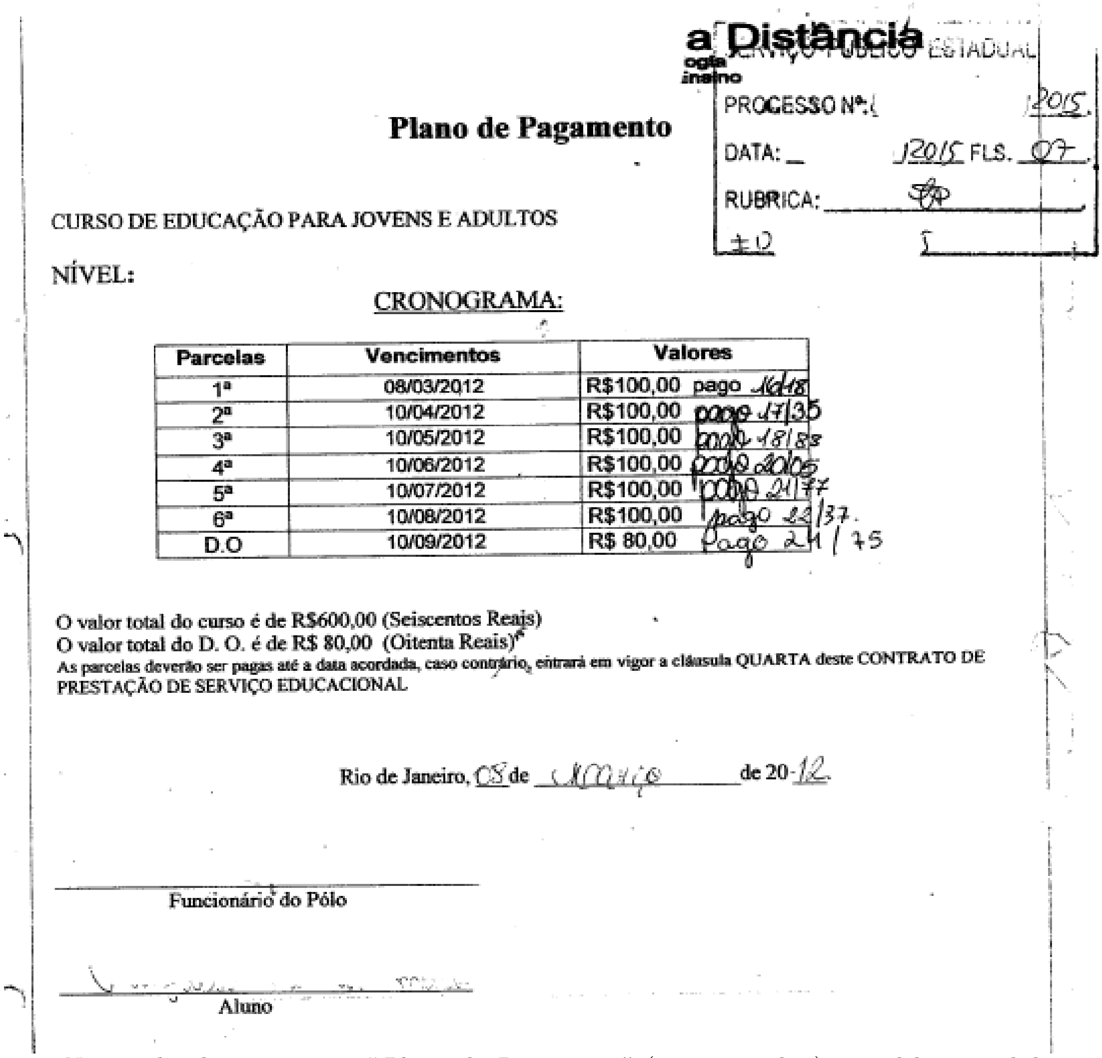
Rio de Janeiro - RJ 20/12/2012
LOCAL DATA

DIRETOR CREDENCIADO

SECRETÁRIA ESCOLAR

Note: Both panels show high-school completion certificates issued to *different students* (names redacted). Apart from the personal data and registration numbers, every element in the grade table is exactly the same: each subject lists the identical mark and the same attendance comment. The perfect duplication indicates that the diplomas were generated from a single template rather than derived from genuine coursework records.

Figure A4: Illustrative payment receipt for a fake high school diploma (total cost of $R\$$ 600, identical to the quote in Figure A2).



a Distância

SERVIÇO PÚBLICO ESTADUAL

PROCESSO Nº: /2015

DATA: /2015 FLS. 07

RUBRICA:

Plano de Pagamento

CURSO DE EDUCAÇÃO PARA JOVENS E ADULTOS

NÍVEL:

CRONOGRAMA:

Parcelas	Vencimentos	Valores
1ª	08/03/2012	R$100,00 pago 16/18
2ª	10/04/2012	R$100,00 pago 17/35
3ª	10/05/2012	R$100,00 pago 18/83
4ª	10/06/2012	R$100,00 pago 20/05
5ª	10/07/2012	R$100,00 pago 21/77
6ª	10/08/2012	R$100,00 pago 22/37.
D.O	10/09/2012	R$ 80,00 Pago 24/75

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 20-12.

Funcionário do Pólo

Aluno

Note: The document is a "*Plano de Pagamento*" (payment plan) issued by an adult-education school later implicated in the diploma-selling scheme. It schedules six monthly instalments of R$ 100 dated March–August 2012, summing to R$ 600—the same amount quoted by the vendor in the email shown in Figure A2. An additional line lists an *"D.O."* fee of R$ 80, referring to publication in the *Diário Oficial.*

B Constructing the Data

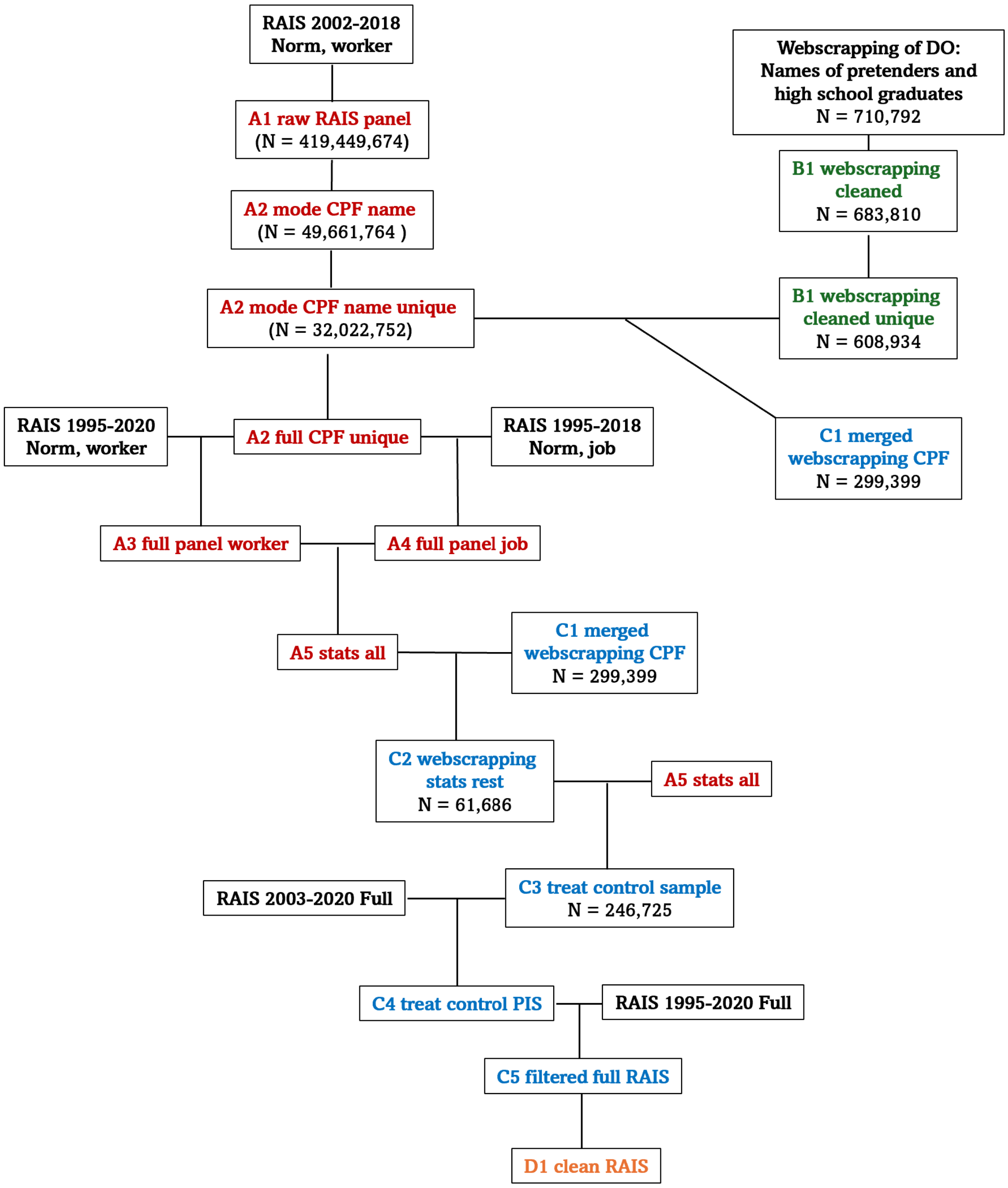
This appendix details the steps for the construction of the main sample. The two data sources are the Diário Oficial do Estado do Rio de Janeiro (DOERJ) cross-sectional registry and the Relação Anual de Informações Sociais (RAIS) panel registry. We use the DOERJ data to identify individuals who graduated from high school and their corresponding school names, which will be used to assign whether a graduate has a real or a fake diploma. The RAIS dataset allows us to follow the formal labor market trajectories of these high school diploma holders from 1995 to 2020. Importantly, both datasets have information about the full name, which is used in the merge procedure.

There are four main steps in the data construction process: Figure X1 illustrates the data construction process (the prefix X1 indicates the code where each step is conducted).

In the first step [A], we create a cross-section dataset derived from RAIS with information on all individuals who satisfy 4 conditions: (i) their full name should be reported in RAIS (i.e. they worked formally after 2002, the year when this variable started to be reported), (ii) they formally worked at least once in one of the four Brazilian states where schools that sold diplomas were more common (i.e. RJ, MG, PR and MS), (iii) they were born after 1955 and (iv) they are uniquely identified by their full name within the sample of workers who satisfy (i)-(iii). Conditions (i) and (iv) are important as we identify individuals across administrative registries based on their full name. Conditions (ii) and (iii) are employed to focus solely on the population who were more likely involved in the diploma buying scheme and who are old enough to have an observed work history before the acquisition of the adult high school diploma (i.e. approximately aged 25-45).

After filtering the RAIS 2002-2020 panel dataset to the 419,449,674 observations who satisfy conditions (i)-(iii) [A1], we calculate for each CPF number the mode of the full name variable. Although workers are uniquely identified by the Cadastro de Pessoa Física (CPF) number, they are not uniquely identified by their full names. Additionally, there might be inconsistencies in the name variable across different

Figure B1: Diagram of steps for the data construction process



years. Therefore, we calculate for each individual (identified by the CPF number) the mode of the name variable. This step results in 49,661,764 individual-level observations, which we further restrict to the 32,022,752 formal workers who are uniquely identified by the mode of their full name, therefore applying condition (iv) [A2]. At last, we filter RAIS 1995-2020 to these individuals once again and retrieve information on their gender, birth year, municipality, reported schooling [A3], employment, wage and tenure [A4], which are then summarized in the 'stats all' cross-sectional dataset [A5].

In the second step [B], we restrict the DOERJ administrative dataset to individuals who graduated from high school and are uniquely identified by their full names within this dataset. There are 710,792 rows in the raw DOERJ dataset. As this dataset is derived from a webscrapping procedure, some entries are fuzzy and do not represent actual individuals, but rather other parts of the DOERJ text. After removing observations with reported names such as 'aluno', 'diretor', 'modalidade' and 'cnpj', we still have 683,810 observations. Next, we identify individuals who are real diploma holders and those who bought fake diplomas based on the name of the school. There are 21 schools which have been reported as sellers of fake adult high school diplomas: Instituto Andreas Brunner, Centro Educacional Futura, Instituto Latino de Ciência e Tecnologia, Escola Triunfo, CECAMP, COBRA - Colégio Brasileiro, Colégio Joan Miro, CEC - Centro Carioca De Ensino, EPEC - AVM, Centro Educacional de Saude Ribeiro dos Santos, CEMOB, Colégio Jose Fonseca, Instituição De Ensino Sigma, Univerta – Sistemas Avançados de Educação, Colégio Estadual Frederico Azevedo, Centro de Formação Aplicação e Cultura, Centro Educacional Podio, Ciep Brizolão 119 - Austin, Sistema Augusto de Educação Integrada, Instituto Educacional Luminis, Nova Dimensão. This step results in 549,195 real diploma holders and 134,651 fake diploma holders. Finally, we restrict these individuals to those who are uniquely identified by their full names within this sample, resulting in 608,934 diploma holders [B1].

In the third step [C], we conduct three procedures: (C1) we merge the cross-sectional datasets from steps [A5] and [B1] using the full name, (C2) we filter the resulting dataset from (i) by key characteristics and (C3) we create a matched control group of individuals without high school diplomas, based on workers in the filtered dataset (C2).

In the first procedure [C1], we identify individuals with high school diplomas from the DOERJ registry (either fake or real) [B1] in the RAIS labor market registry [A5] using the full name variable. We find 299,399 individuals out of the 608,934 diploma holders. The accuracy of this merge procedure relies in the fact that we restricted both administrative datasets to individuals who are uniquely identified by their names within each dataset. However, an important caveat is that RAIS is not a universal registry of all Brazilians. Hence, it could be the case that there are 2 individuals with the same full name, but only one of them appears in the RAIS dataset, while only the other is in the DO dataset. In this case, we would perform an incorrect merge based on the full name, since both individuals are uniquely identified within their respective datasets.

In the second procedure [C2], we restrict the 299,399 diploma holders to those who (i) graduated from 2000 to 2015, (ii) were at least 25 years old at graduation and (iii) were formally employed in RAIS at least once in a 10-year window around the graduation year (either 5 years before or after this date). Restriction (iii) is applied to focus solely on individuals with some formal labor market attachment around the graduation year, while restriction (i) is employed to ensure that the 10-year window around graduation is feasible for all workers, as the RAIS dataset spans from 1995 to 2020. Finally, restriction (ii) implies that the resulting dataset refers to only adult high school students who had had a chance of prior work history before the having the diploma. Table 10 presents a balance check for the restriction procedure, where we apply each restriction separately in columns (2)-(4) and all restrictions at the same time in column (5). Only 61,686 out of the 299,399 diploma holders satisfy all restrictions. The share of individuals with fake diplomas increases from 17.4% to 48.5% once we apply all restrictions, such that 31,799 are real diploma holders and 29,887 are fake diploma holders.

In the third procedure [C3], we create a matched control group derived from [A5] of individuals who are similar in key characteristics to diploma holders [C2], but who did not graduate from high school. We start by removing from [A5] individuals who are more likely to have a high school diploma. Thus, we drop those workers

Table 13: Restricting the diploma holders dataset [C2]

	Full	Restricted by			Restricted	Diff.
		Grad. year	Age at grad.	Empl. window		
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Male	0.530	0.530	0.574	0.543	0.606	0.096
Birth year	1987.879	1987.646	1976.566	1987.098	1975.502	-15.588
Graduation year	2010.743	2009.946	2011.299	2010.130	2009.970	-0.973
Age at graduation	22.864	22.300	34.733	23.032	34.468	14.616
1(Fake diploma)	0.174	0.188	0.404	0.214	0.485	0.391
% High schooled (RAIS)	0.786	0.812	0.570	0.786	0.590	-0.247
Avg. salary	1898.089	1973.929	1919.692	1813.461	2043.473	186.031
Avg. tenure	17.776	18.124	28.652	18.735	30.749	16.600
Avg. salary 21-24 y.o.	1652.744	1696.171	1271.281	1608.473	1292.967	-428.049
Avg. tenure 21-24 y.o.	13.723	13.783	14.511	14.255	14.969	1.482
1(Empl in 1995)	0.113	0.103	0.556	0.132	0.676	0.595
1(Empl in 1996)	0.104	0.086	0.474	0.119	0.545	0.484
1(Empl in 1997)	0.115	0.104	0.447	0.139	0.505	0.444
1(Empl in 1998)	0.130	0.122	0.461	0.159	0.521	0.458
1(Empl in 1999)	0.141	0.133	0.454	0.168	0.507	0.445
1(Empl in 2000)	0.121	0.117	0.439	0.146	0.481	0.439
1(Empl in 2001)	0.127	0.125	0.494	0.161	0.529	0.495
1(Empl in 2002)	0.135	0.134	0.511	0.172	0.542	0.504
1(Empl in 2003)	0.158	0.157	0.554	0.199	0.580	0.538
1(Empl in 2004)	0.183	0.183	0.570	0.228	0.593	0.538
1(Empl in 2005)	0.200	0.200	0.585	0.248	0.607	0.535
1(Empl in 2006)	0.222	0.222	0.606	0.273	0.627	0.531
1(Empl in 2007)	0.257	0.257	0.629	0.313	0.651	0.519
1(Empl in 2008)	0.292	0.292	0.651	0.353	0.673	0.500
1(Empl in 2009)	0.314	0.314	0.656	0.376	0.678	0.478
1(Empl in 2010)	0.347	0.347	0.672	0.410	0.693	0.452
1(Empl in 2011)	0.388	0.391	0.683	0.461	0.710	0.417
1(Empl in 2012)	0.419	0.427	0.686	0.499	0.719	0.384
1(Empl in 2013)	0.455	0.466	0.689	0.539	0.724	0.341
1(Empl in 2014)	0.484	0.493	0.684	0.565	0.719	0.295
1(Empl in 2015)	0.496	0.502	0.664	0.570	0.704	0.255
1(Empl in 2016)	0.517	0.526	0.630	0.590	0.680	0.198
1(Empl in 2017)	0.517	0.527	0.584	0.581	0.639	0.146
1(Empl in 2018)	0.555	0.558	0.576	0.605	0.632	0.089
1(Empl in 2019)	0.090	0.099	0.122	0.106	0.177	0.098
1(Empl in 2020)	0.633	0.643	0.572	0.648	0.630	-0.003
N	299,399	265,266	78,737	224,160	61,686	

Note: This table details the process of restricting the diploma holders dataset to individuals who graduated from 2000 to 2015, who were at least 25 years old at the time of graduation and who were employed either 5 years before or 5 years after the graduation year. Column (1) presents the averages for all individuals who were matched with RAIS, ans columns (2)-(4) present the averages after applying each restriction separately. Column (5) reports the averages after applying all restrictions, and column (6) reports the differences between column (5) and column (1).

who appeared in the [C1] dataset (i.e. who have high school diplomas). Next, we calculate the share of RAIS observations where each individual had high school education reported by their employer and filter the dataset to those who had at most 10% observations with high school education reported. As a result, there are 5,632,275 candidates for the matching procedure.

For each worker among the 61,686 diploma holders, we match 10 control individuals by randomly selecting workers who share the same gender, birth year and mode of state (in RAIS 1995-2020), as well as the 10-year window around graduation dummy. Moreover, we conduct the matching procedure for each graduation year separately. Therefore, each control worker has a placebo graduation year, and there might be control workers who are matched to more than one real diploma holder. Table 11 presents a balance check for the real diploma holders and control individuals. There are 616,739 workers in the control group. Note that this is not precisely 10 times the number of real diploma holders since some gender – birth year – mode of state – 10-year window cells have less than 10 candidates for matched controls. After the matching procedure, we combine real diploma holders and controls into a single dataset and filter the RAIS 1995-2020 panel to these individuals [C4, C5].

Finally, the fourth step [D] consists of cleaning the resulting RAIS panel dataset of 678,425 workers. We set the panel dataset to the individual/year level by keeping only a worker's main job at a given year. Importantly, a worker may have more than one formal labor contract in the same year either because she had multiple jobs at the same time or because she changed employers in a given year. Therefore, we keep only the job that paid the highest amount of money for that worker by multiplying the average wage by the number of months the worker was employed at that firm. This results in an unbalanced panel with 7,818,852 individual/year observations spanning from 1995 to 2020, conditional on formal employment. Each observation reports the employer firm, the municipality of the firm, the firm sector (CNAE), the firm size, the worker's average wage at the firm, the occupation (CBO), whether this occupation requires high school education, the worker's schooling as reported by the employer, the tenure at that firm, whether the contract was ended by the end of a given year and the separation reason, and

Table 14: Balance check: Diploma holders and matched control group

Variable	Full (1)	Dip. holders (2)	Controls (3)	Diff. (4)
Graduation year	2009.97	2009.97	2009.97	0
Male	0.606	0.606	0.606	0
Birth year	1975.506	1975.502	1975.506	-0.004
Age at graduation	34.464	34.468	34.464	0.004
% High schooled (RAIS)	0.062	0.59	0.01	0.58***
Avg. salary	1436.275	2043.473	1375.473	667.999***
Avg. tenure	31.577	30.749	31.66	-0.91***
Avg. salary 21-24 y.o.	1110.839	1292.967	1087.808	205.159***
Avg. tenure 21-24 y.o.	15.326	14.969	15.371	-0.402***
1(Empl in 1995)	0.492	0.676	0.474	0.203**
1(Empl in 1996)	0.481	0.545	0.475	0.071
1(Empl in 1997)	0.469	0.505	0.466	0.039
1(Empl in 1998)	0.449	0.521	0.441	0.08***
1(Empl in 1999)	0.429	0.507	0.422	0.085***
1(Empl in 2000)	0.414	0.481	0.407	0.074***
1(Empl in 2001)	0.454	0.529	0.446	0.083***
1(Empl in 2002)	0.475	0.542	0.469	0.073***
1(Empl in 2003)	0.497	0.58	0.488	0.092***
1(Empl in 2004)	0.499	0.593	0.489	0.103***
1(Empl in 2005)	0.503	0.607	0.493	0.114***
1(Empl in 2006)	0.516	0.627	0.505	0.122***
1(Empl in 2007)	0.529	0.651	0.517	0.134***
1(Empl in 2008)	0.548	0.673	0.536	0.137***
1(Empl in 2009)	0.548	0.678	0.535	0.143***
1(Empl in 2010)	0.562	0.693	0.549	0.144***
1(Empl in 2011)	0.573	0.71	0.559	0.151***
1(Empl in 2012)	0.572	0.719	0.557	0.162***
1(Empl in 2013)	0.567	0.724	0.552	0.173***
1(Empl in 2014)	0.561	0.719	0.545	0.175***
1(Empl in 2015)	0.534	0.704	0.517	0.186***
1(Empl in 2016)	0.491	0.68	0.472	0.209******
1(Empl in 2017)	0.449	0.639	0.43	0.209***
1(Empl in 2018)	0.428	0.632	0.407	0.225***
1(Empl in 2019)	0.122	0.177	0.116	0.061***
1(Empl in 2020)	0.406	0.63	0.384	0.246***
N	678,425	61,686	616,739	

Note: This table presents the balance check between diploma holders and their matched control individuals. Column (1) reports the averages for all groups (controls and diploma holders), column (2) presents the averages only for diploma holders, and column (3) reports these averages only for the controls. Column (4) presents the differences between columns (2) and (3) derived from regressions where the explanatory variable is having a high school diploma.

the admission date.