Employer Learning, Statistical Discrimination and University Prestige

Paola Bordón and Breno Braga*

*University of Wisconsin and University of Michigan

April, 2014

Abstract

This paper investigates whether firms use university prestige to statistically discriminate among college graduates. The test is based on the employer learning literature which suggests that if firms use a characteristic for statistical discrimination, this variable should become less important for earnings as a worker gains labor market experience. In this framework, we use a regression discontinuity design to estimate a 19% wage premium for recent graduates of two of the most selective universities in Chile. However, we find that this premium decreases by 3 percentage points per year of labor market experience. These results suggest that employers use college selectivity as a signal of workers’ quality when they leave school. However, as workers reveal their productivity throughout their careers, they become rewarded based on their true quality rather than the prestige of their college.

JEL Classification: I21, J31, J71

Keywords: Employer Learning, Statistical Discrimination, College Selectivity.

*We would like to thank John Bound, Charlie Brown, Brad Hershbein, John Kennan, David Lam, Fabian Lange, Brian McCall, Jeff Smith, Christopher Taber, Lesley Turner, Sergio Urzua and participants at EconCon, LACEA, the University of Michigan and University of Wisconsin labor workshops for their valuable comments. The authors can be contacted by email: bordon@wisc.edu and bgbraga@umich.edu.
1 Introduction

Labor markets are characterized by incomplete information on workers’ productivity (Spence (1973)). There are some characteristics of workers, such as labor market ability, that are important for performance on the job but are not easily observable by employers. In this context, firms often have to make judgments on workers’ unobservable quality on the basis of the available information. Within this framework, statistical discrimination is defined as employers using a group identity of workers to infer their unobservable quality.

The most traditional group identity studied in the statistical discrimination context is race (Phelps (1972) and Aigner and Cain (1977)). In this literature, the racial wage gap is justified not because employers are prejudiced against a particular race but because they use race identity to predict the unobservable quality of workers. More recently, evidence was found that firms use schooling (Farber and Gibbons (1996), Altonji and Pierret (2001) and Lange (2007)) or information on lay-offs (Gibbons and Katz (1991) and Hu and Taber (2011)) to statistically discriminate workers.

In this paper we study a new dimension of statistical discrimination: we investigate if firms use the prestige of the university attended by a worker to predict his or her unobservable labor market quality. We believe that college prestige satisfies the typical features of group identity that might be used for statistical discrimination for two main reasons. First, this information is easily accessible to firms: workers use the university name in their resumes and prestigious universities are widely recognized in the labor market. Second, there is evidence that more talented individuals attend more prestigious universities (Hoxby (1998) and Dale and Krueger (2002)). Overall, elite universities have a very competitive application process and tend to select higher quality candidates.\(^1\) Within this framework, it is natural to believe that firms use university prestige in order to infer the unobservable labor market quality of workers.

In order to test if employers use university prestige as a signal of workers’ unobservable quality, we rely on the statistical discrimination and employer learning (EL-SD) literature (Altonji and

\(^1\)As it will become clear later, the underlying assumption is that universities are better at screening candidates than firms.
The underlying assumption is that the imperfect information about a worker’s quality tends to disappear with time. At the early stages, firms assess workers on the basis of easily observable variables that are correlated with their unobservable quality. As a worker gains experience in the labor market, employers weigh these characteristics with other information that becomes available, such as references and on-the-job performance. If employers use a characteristic to statistically discriminate a worker in the early stage of his career, this information should become less important for earnings as a worker reveals his true productivity with time.

This paper uses data from *Futuro Laboral* of the Chilean Ministry of Education. This data satisfies the purpose of the paper for several reasons: first, it follows different cohorts of college graduate workers from Chile in their first years in the labor market, the period in which most of the employer learning happens (Lange (2007)). Second, the data presents information on labor market outcomes such as earnings from administrative data and we can identify workers that graduate from elite universities. Finally, the data contains information on the scores of the centralized admission test to universities in Chile. As it will be clear later, this information will be used to construct the running variable in the regression discontinuity test we suggest.

We take advantage of the centralized admission process to college in Chile to propose a statistical discrimination test based on regression discontinuity design (RD). Using information on the admission test scores we are able to identify workers who were just above or just below the admission thresholds to the two most prestigious universities in Chile. We suggest an EL-SD test that compares the earnings’ dynamics between these two groups of workers as they gain experience in the labor market. The test predicts that if firms use university prestige to statistically discriminate workers: i) individuals barely admitted to the most selective universities in Chile should be paid substantially more than those barely rejected when they graduate from college; ii) the wage differential between these two groups of workers should shrink as individuals progress in their career.

---

2Other important papers in this literature include Lange (2007), Schönberg (2007), Arcidiacono et al. (2010), and Mansour (2012).

3Kaufmann et al. (2012) and Hastings et al. (2013) are two recent papers that have also explored the regression discontinuities generated by the centralized admission process to universities in Chile. Kaufmann et al. (2012) looks at effect of graduating from a elite university on marriage outcomes and Hastings et al. (2013) studies labor market returns to college admission. None of these papers explore how the selective university wage premium changes throughout a worker’s career, that is the main object of interest of this paper.
The idea for the test is similar to the one presented by Altonji and Pierret (2001). Employers do not observe admission test scores but they know that prestigious universities admit on average better candidates. If employers use the selectivity of a university as a signal of worker’s inherent ability, individuals just above the admission cutoff must be better paid than those just below the admission cutoff. Nevertheless, the EL-SD model proposes that employers learn a worker’s unobservable quality with time. In consequence, the signal associated with graduating from a prestigious university should become less important for earnings and the wage differential between workers similar pre-college characteristics should shrink with time.

We find evidence for statistical discrimination on the basis of university prestige. We estimate that workers just above the admission cutoff to the two most prestigious universities in Chile earn on average 12% more than those just below the cutoff in the first year after graduation. However, this wage premium tend to decrease by 2 percentage points by year of experience in the labor market, to the point that we cannot reject a zero earnings differential between these two groups of workers 4 years after their graduation. We also take into consideration the fuzziness of the regression discontinuity design, meaning that not all students admitted to a prestigious university in Chile attend such university, to estimate the local average effect of graduating from a prestigious university on earnings. We estimate a 19% wage premium for recent graduates of the two most prestigious university in Chile. However, this wage premium decreases by 3 percentage points per year of experience.

Based on these findings, this paper contributes to different dimensions of the existing literature. First, this paper is a contribution to the EL-SD literature because we study statistical discrimination on the basis of a different group identity. While there is an extensive literature that analyzes the use of race, gender, and schooling, we are one of the first papers to study whether firms use prestige of college to statistically discriminate workers. Furthermore, to the best of our knowledge, our paper is the first to propose an employer learning-statistical discrimination test based on a regression

--

4To the best of our knowledge, Lang and Siniver (2011) and Hershbein (2013) are the two other papers that have addressed this issue. Lang and Siniver have a similar approach to estimate how returns to attending a elite university in Israel changes with labor market experience. However, the authors are unable to properly exploit the regression discontinuity in the college admission.
discontinuity design. We present issues with the traditional test proposed by Altonji and Pierret (2001) if employers statistically discriminate workers on the basis of characteristics that are not present in the data and are correlated with graduating from a prestigious university, such as family social-economic background. We also demonstrate that the regression discontinuity test we propose is robust to such bias.

Second, we contribute to the literature which studies the effect of graduating from an elite university on labor market outcomes. There is an extensive series of papers that estimate the returns to graduating from a selective university on earnings (Brewer et al. (1999), Hoxby (1998), Dale and Krueger (2002), and Black and Smith (2006)), including papers that have used a regression discontinuity design (Saavedra (2008) and Hoekstra (2009)). The overall finding is that there is a positive effect of graduating from a prestigious college on earnings. While there is big effort in the literature to overcome the selection bias associated to attending a prestigious university, little attention has been given to the mechanisms that generate the college selectivity wage premium.

Different from past work, in this paper we shed some light on the reasons for why workers from prestigious universities receive higher wages after graduation. On one hand, attending a selective university could be associated with receiving better instruction and having more accomplished peers. In this context, prestigious universities have an advantage of increasing a worker’s productivity in comparison to less prestigious universities. On the other hand, the main effect of attending a selective university might be to signal to employers an unobservable inherent ability of a worker. In this context, the extra value added from a selective college education might not be significantly higher than that from a less prestigious university. Using the regression discontinuity test we propose we are able to disentangle the signaling effects form the value added effect of graduating from a prestigious university. Our finding of a rapid decrease in the elite college premium for workers

---

5The only exception is Dale and Krueger (2002) who find no wage premium from attending a selective college. It is interesting to note that the authors estimates the wage premium approximately 15-19 years after a worker’s graduation from college. The zero effects for individuals with similar pre-college characteristics later in advanced age does not contradict the empirical findings of this paper.

6Pop-Eleches and Urquiola (2013) is one of the few papers that have discussed the benefits from attending a higher quality school. Their paper address the behavior effect of students, parents and teachers in response to a student admission to better secondary school. Nevertheless, the paper has little to say about the impact of attending a higher quality school on labor market outcomes.
with similar pre-university characteristics is evidence that signaling mechanisms are stronger than productivity mechanisms. In particular, the fact that we cannot reject a wage differential between workers just above and below the admission cutoff after 4 years in the labor market suggests that the value added from the two most prestigious university in Chile is not significantly different from the less prestigious schools.

2 Institutional Framework

Higher education in Chile comprises three types of institutions: Universities, Professional Institutes (IPs), and Technical Formation Centers (CFTs). Universities provide the highest degree of learning, combining teaching, research and outreach activities; they teach accredited degree programs (2.5 to 4 years) and award academic degrees (5 to 7 years). Professional Institutes are in charge of granting professional degrees other than those awarded by universities, and they are also authorized to grant higher education technical degrees in areas where this is required. Technical Formation Centers are intended to equip higher level technicians with the competencies and skills needed to respond to the needs of industry in the public and private sectors.

Universities can be divided into two main categories: traditional and non-traditional institutions. Traditional institutions comprise the oldest and most prestigious universities created before 1981, and those institutions that derived from the old universities (created after 1980). Traditional establishments consist of 25 fully autonomous universities coordinated by the Council of Chancellors of Chilean Universities (CRUCH) and are eligible to obtain partial funding from the state. They employ a single admission process: the University Selection Test (PAA)\(^7\). This test is made up of three compulsory sub-tests including language, mathematics, and history and geography of Chile. Additionally, depending on which programs they are planning to apply to, students may be required to take the following specific PAA tests: advanced mathematics, physics, chemistry, biology, and history.

\(^7\)In 2004 the university selection test was modified and it is now called PSU.
1. First, students take the PAA test and after receiving their score they make their application choices. Students apply to a major and university (or program) simultaneously and can only apply to 8 programs, ranking them up by preferences. The only criterion used for admission in the traditional universities is the score in the PAA. This final admission scores consists of a weighted average of the compulsory and major specific tests and high school GPA, with each program setting its specific PAA weights.\(^8\) The number of vacancies for each program is announced before the application process and programs fill their vacancies solely based on the final weighted scores. The admission score cutoff is defined by the score of the last student admitted into a program and it is not known before the application decisions and therefore students cannot manipulate which side of the cutoff on they fall on.\(^9\)

Non-traditional universities were created after 1981, have no state financial support and might not necessarily use the PAA score to select their incoming students. Nevertheless, the anecdotal evidence is that the majority of students willing to attend higher education in Chile take the PAA at the end of high school independent of the university they are planning to attend. The test is relatively inexpensive and administrated throughout the country.

All higher education institutions charge tuition and fees. However, for those students enrolled in one of the traditional universities, solidarity credits and scholarships are available. Competition in these markets, particularly for undergraduates, is often geographically circumscribed to local and regional markets, and it can be more or less intense depending on the institution. As of 2001, the Chilean higher education system consisted of 60 universities (25 traditional universities and 35 new private universities without direct public subsidy), 42 professional institutes (all of them private), and 117 private technical formation centers.

The increasing enrollment in higher education has led to an increasing number of graduates in the last two decades. In 1995, 24,400 graduates entered the labor market, whereas in 2000 around

\(^8\)For example, engineering in a prestigious university requires 20% of mathematics, 10% of language, 10% of history, 20% high school GPA, 30% specifics mathematics, and 10% physics. The final score to the same major in a different university might requires different weights.

\(^9\)Students could use the admission score cutoff of previous years as a reference. Given the variation of the admission cutoff overtime and the possibility to apply to 8 different programs, we believe that students with marginal scores to be admitted in prestigious university tend to apply to these competitive programs.
42,000 graduates did, and in 2005, 71,170 new graduates were entering the job market. This means than in ten years the number of graduates has almost tripled. Traditional universities have more than doubled the number of graduates they produce, but private universities have increased by 6.7 times their number of graduates.

3 Data

The data to be used in the study comes from Futuro Laboral, a project of the Ministry of Education of Chile that follows individuals over the first years of labor market experience after graduating from higher education programs. The panel data set matches tax returns with transcripts of students’ majors and the institutions they graduated from. The unit of analysis concerns only those who graduate from both traditional and non-traditional universities; those who have stopped studying or did not continue their studies after graduating from high school are not in the sample. Income information is available between the years 1996 and 2005. We have data for the 1995, 1998, 2000 and 2001 graduating classes.\(^{10}\)

The information provided by the Internal Revenue Service (SII) comprises age, sex, name of the institution that individuals graduated from, major, the year of graduation, annual income reported in tax returns, city or cities of employment, number of employers and economic sector. The raw data contains every worker in Chile that had positive earnings between 1996 and 2005, even those who exempt from tax.\(^{11} \)\(^{12}\) For a random sub-sample, the Ministry of Education gathers information about the PAA score, high school grades and the institutions students graduated from high school. As the PAA scores have an important role in both the traditional EL-SD and regression discontinuity analysis, we restrict our study to this sub-sample.

\(^{10}\)Note that the cohorts are observed for different length of time. For example, while we observe 10 years of labor market experience for the 1995 graduation class, we only 4 years of labor market experience for the 2001. Unfortunately, the project was deactivated and the income data for more recent years was not not collected.\(^{11}\)Note that in Chile, married couples must fill their taxes separately.\(^{12}\)A concern is that part of the individual from prestigious universities might go to graduate school after finishing their baccalaureate studies and therefore would be omitted in the earnings sample. However, the fraction of workers that go to graduate school in Chile is very low. Using data from the National Socioeconomic Characterization Survey in the year 2000, we find that only 0.65% of 25-34 years old individuals with a bachelor degree were enrolled in graduate school or had obtained a graduate degree.
The wage measured in the sample is the annual income that comes from jobs and services provided by the individual.\textsuperscript{13} We use consumer price index (IPC) as a deflator to compute real wages. The experience variable is computed as the number of years an individual has income and has paid taxes after graduation. The final sample consists of 58,477 individuals and 322,688 observations.

We divide universities into two groups: selective and non-selective universities. The selective universities comprises two of the oldest and most prestigious universities in the country. These schools attract students with the highest PAA scores and therefore are the most selective schools in the country. The programs of these two universities have also been consistently ranked among the highest in Chile and their prestige is well recognized nationwide.\textsuperscript{14} See Table 1 for descriptive statistics regarding these two groups. As expected, selective universities have on average higher scores in Math and Language components of the PAA tests, and their students have higher high school grades. We also see that 11\% of selective universities students went to a private high school, compared to 7\% from non-selective universities. We also plot in the distribution of language and math PAA scores for college graduates from selective and non-selective universities on figures 2 and 3 respectively. One can see from the figures that the language and math scores of graduates from selective universities are concentrated at the higher end of the distribution. Finally, we show in Table 2 that workers from the two selective universities have on average higher earnings than those from the less prestigious schools.

4 Regression Discontinuity Test

In order to provide evidence for statistical discrimination based on college prestige, we use a regression discontinuity (RD) design. The basic idea is to compare how the earnings of those just above

\begin{footnotesize}
\textsuperscript{13}We do not have information on weeks or hours worked in the sample and for this reason we cannot explore how much of the annual income increase is due to changes in hours or week of work. Nevertheless, workers with a bachelor degree in Chile present both a high employment attachment and the majority work full time. Using the National Socioeconomic Characterization Survey in the year 2000, we find that 86.7\% of 25-34 years old individuals with a bachelor degree work are employed in the period of the interview and from those, 88\% work more than 35 hours per week.

\textsuperscript{14}Due to a confidentiality agreement with the Ministry of Education, we cannot provide the name of these two institutions.
\end{footnotesize}
and just below the cutoff for admission to the most selective universities in Chile change as workers accumulate experience in the labor market. The identification assumption is that other factors that could affect earnings are continuous at the admission cutoff and students have limited power to manipulate on which side of the admission cutoffs they might fall. Furthermore, we assume that employer do not have access to the test scores that a prestigious university uses in their admission process.

4.1 Employer Learning Statistical Discrimination Model

The standard employer learning model specifies the log-productivity of a college graduate worker $i$ with experience level $t$:

$$ y_{it} = r s_i + \alpha_1 q_i + \lambda z_i + \eta_i + H(t) $$

where $s_i$ captures information that is available to both employers and researchers. In this paper, $s_i$ is defined as an indicator if a worker graduated from a prestigious university or not. The variable $q_i$ describes information available to employers and not present in the data, such as family social economic background, $z_i$ is a characteristic present in the data but not available to employers and $\eta_i$ is a measure of a worker’s inherent ability that is not available in the data or to employers. Finally, $H(t)$ describes the relation between log-productivity and experience and does not depend on the other variables of the model.

In the absence of information on $z_i$ and $q_i$, employers form expectations based on other observed characteristics of workers. Altonji and Pierret (2001) assume that these conditional expectations are linear on $s$ and $q$:\textsuperscript{16}

$$ z = \mathbb{E}[z|s, q] + v = \gamma_1 q_i + \gamma_2 s + v $$

$$ \eta = \mathbb{E}[\eta|s, q] + e = \alpha_2 s + e $$

\textsuperscript{15}Students can retake the test next year, but they cannot retake the test after they got their results the same year, which decreases the probability of manipulation.

\textsuperscript{16}A normalization allows suppressing $q$ in the second expectation.
where \( v \) and \( e \) are scalar with mean zero and uncorrelated with \( s \) and \( q \) by construction. Under this assumption, one can characterize the expected value of \( y \) given information on \( s \) and \( q \):

\[
E[y|s, q] = (r + \lambda \gamma_2 + \alpha_2)s + (\alpha_1 + \lambda \gamma_1)q + H(t)
\]

In the traditional EL- SD model, employers have access to a noisy measure of a worker’s productivity after each period that an individual spend in the labor market:

\[
\tilde{y}_{i\tau} = y_{i\tau} + \epsilon_{i\tau}
\]

where the noise \( \epsilon_{i\tau} \) is independent of all the variables of the model. As in Altonji and Pierret (2001), employers share equal information about workers, labor markets are competitive and there is a spot market for labor services. As a consequence, wages are equal to the expected productivity of a worker, given the information available to employers at each period.

\[
W_{it} = E[\exp(y_{i\tau})|s, q, \tilde{y}_{i0}, ..., \tilde{y}_{i(t-1)}]
\]

Lange (2007) assumes that \( \epsilon_{i\tau} \) is independently, identically and normally distributed with a finite variance. Under this assumption, the process of updating the expectations of employers have a very simple structure an the log-wage process can be represent by:

\[
w_{it} = (1 - \theta_t)E[y|s, q] + \theta_t \frac{1}{t} \sum_{\tau=0}^{t-1} \tilde{y}_{i\tau} + \tilde{H}(t) \tag{2}
\]

where \( \tilde{H}(t) \) is a linear transformation of \( H(t) \) and \( \theta_t \) is a function of the variances of and \( \epsilon_{i\tau} \), \( s \) and \( q \). Furthermore \( \theta_0 = 0 \) and \( \theta_t \) strictly increases with \( t \) converging to 1 as \( t \) goes to infinite.\(^{17}\) This expression demonstrates that as a worker progress in his career, employer weight less their initial believe on a worker’s productivity based on \( s \) and \( q \), and weight more the new information that becomes available during a worker’s career.

\(^{17}\)See Lange (2007) for the formal derivations of these parameters.
4.1.1 Traditional EL-SD Test

The object of interest in the traditional employer learning model is the linear projection of the log-wage \( w_{it} \) on \( s, z \) and \( t \).

\[
\mathbb{E}^*[w_{it}|s, z, x] = b_{sx}s + b_{zx}z + \bar{H}(t)
\]

Without lost of generality, one can define the the projections of the unobservable variables \((q, \eta)\)
on the observable variables \((s, z)\):

\[
q = \gamma_3s + \gamma_4z + u_1
\]

\[
\eta = \gamma_5s + \gamma_6z + u_2
\]

Using the independence of \( \varepsilon_{it} \) to all the variables of the model, Lange (2007) show that the coefficients of the projections:

\[
b_{st} = (1 - \theta_t)b_{s0} + \theta_t b_{s\infty}
\]

where, as discuss before, \( \theta_0 = 0 \) and \( \lim_{t \to \infty} \theta_t = 1 \). The traditional EL-SD test consists in estimating how \( b_{st} \) and \( b_{zt} \) change with experience level \( t \). Indeed, Altonji and Pierret (2001) propose that if firms statistically discriminate workers on the basis of \( s \) and if \( z \) is positively related to \( s \), one should observe that \( b_{st} \) falls with \( t \) and \( b_{zt} \) should rise with \( t \).

Furthermore, under the assumptions above Lange (2007) shows that:

\[
b_{s0} = \frac{r}{A} + \frac{\alpha_1 \gamma_3 + \alpha_2 + \lambda (\gamma_2 + \gamma_1 \gamma_3)}{B} \]

\[
b_{z0} = \frac{\alpha_1 + \lambda \gamma_1 \gamma_4}{D}
\]
where the coefficient $b_{s0}$ represents the relation between graduating from a prestigious university and wages in the beginning of a workers career. The first term $A$ captures the direct effect of attending a prestigious university on productivity. The second term $B$ represents the direct impact of $q$ on wages and the fact that $q$ is not present in the data but it is correlated to $s$. This can be interpreted as the traditional omitted variable problem associated with estimating the returns to graduating from a prestigious university (Dale and Krueger (2002)). It captures the relation between any variable that affects wages, is correlated to graduating from a prestigious university and it is not present in the data. Finally, the term $C$ reflects the fact that employers do not observe $\eta$ and $z$ in the beginning of a worker’s career, but are aware of their relation with $s$. Therefore, employer use $s$ as a signal of unobservable components of a worker’s productivity. In the same way, the relation between $z$ and the log wages of a worker in the beginning of his career is given by the coefficient $b_{z0}$. As employers do not observe $z$, this coefficient only captures the fact that we are omitting $q$ from the linear prediction and that $z$ and $q$ are correlated.

$$b_{s\infty} = \frac{r}{E} + \frac{\alpha_1 \gamma_3 + \gamma_5}{F}$$ \hspace{1cm} (7)$$

$$b_{z\infty} = \frac{\lambda}{G} + \frac{\alpha_1 \gamma_4 + \gamma_6}{H}$$ \hspace{1cm} (8)$$

The coefficients $b_{s\infty}$ and $b_{z\infty}$ represent the relation between $s$ and $z$ respectively with wages as $t \to \infty$ and $\theta_t \to 1$. As before, $E$ represents the direct effect of graduating from a prestigious university on wages. The coefficient $F$ captures the fact that $\eta$ and $q$ have an impact on long-run wages, are related to $s$ but they are omitted in the linear prediction because they are not observed in the data. Note that $F$ is different from the term $B$ because firms only learn $\eta$ with time. In the same way, the term $G$ captures the direct impact of $z$ on productivity and $H$ captures the correlation of $z$ to the omitted variables $\eta$ and $q$.

One important issue that has been omitted from the employer learning literature (Altonji and Pierret (2001) and Lange (2007)) is how the correlation between $s$ and the unobservable factor $q$
can affect the conclusions of statistical discrimination test. This issue arises if firms statistically discriminate workers on the basis of variables that are not observed in the data, such as family social economic background, that are correlated to graduating from prestigious university. In this situation, the traditional employer learning test might suggest that employers statistically discriminate a worker on the basis of university prestige, when in fact firms might be using family social economic status as a signal of a worker’s unobservable characteristics.

In order to give some perspective of the issue, we analyze the extreme case where \( s \) is not correlated to \( \eta \) and \( z \) (\( \alpha_2 = 0, \gamma_2 = 0 \) and \( \gamma_5 = 0 \)). In this situation, employers should not use \( s \) as a signal of a worker’s unobservable characteristics, and therefore, workers are not statically discriminated on the basis of university prestige. Furthermore, assuming that \( q \) is correlated with \( \eta \) and \( z \) (\( \gamma_4 \neq 0 \)), and therefore \( q \) is used by employer to statistically discriminate workers. Under this assumption, the traditional employer learning test would suggest that firms statistically discriminate workers on the basis of university prestige because \( b_s \infty < b_{s0} \) and \( b_z \infty > b_{z0} \). Note, however that this conclusion is being driven by the correlation of \( s \) and \( q \), and the fact that employers use \( q \) to predict \( z \), which is capture by the term \( \lambda \gamma_1 \gamma_3 \) in equation (6).

### 4.1.2 Regression Discontinuity EL-SD Test

The object of interest of the EL-SD test we propose is how the difference between average log-wages of individuals just above and just below the admission cutoff to a prestigious university changes with experience. Precisely, we define \( Dist.Cutoff_i \) as the distance between a student’s test score and the admission threshold of a prestigious university. For simplicity, we assume that all students admitted to a prestigious university graduate from this university, such that \( s_i = 1 \) if \( Dist.Cutoff_i \geq 0 \) and \( s_i = 0 \) otherwise.\(^{18}\)

\(^{18}\)As it will be clear later, this assumptions is not confirmed in the data because some students admitted to a prestigious university decide to attend a less prestigious university (fuzzy regression discontinuity). For simplicity, we ignore this possibility here.
\[ \tau_t = \lim_{\text{Dist.Cutoff} \downarrow 0} \mathbb{E}[w_{it}|\text{Dist.Cutoff}_i] - \lim_{\text{Dist.Cutoff} \uparrow 0} \mathbb{E}[w_{it}|\text{Dist.Cutoff}_i] \]  

(9)

that represents local average difference of log-wages by experience levels at the admission cutoff. The employer learning statistical discrimination consists in testing if \( \tau_t \) decreases with \( t \).

Note that by definition, we have that:

\[ \lim_{\text{Dist.Cutoff} \downarrow 0} \mathbb{E}[s|\text{Dist.Cutoff}_i] = 1 \text{ and } \lim_{\text{Dist.Cutoff} \uparrow 0} \mathbb{E}[s|\text{Dist.Cutoff}_i] = 0 \]

Furthermore, we assume that the distribution of the other variables of the model \( \{z_i, q_i, \eta_i\} \) is continuous around the admission cut-offs. In this case, the expected values of these variables just above and just below the admission cutoff are the same:

\[ \lim_{\text{Dist.Cutoff} \downarrow 0} \mathbb{E}[X|\text{Dist.Cutoff}_i] = \lim_{\text{Dist.Cutoff} \uparrow 0} \mathbb{E}[X|\text{Dist.Cutoff}_i] \]

for \( X = q, z, \eta \). Using these two conditions, the assumption that employer do not have access to \( \text{Dist.Cutoff}_i \), and the the log-wage process derived in (2), one can show that:

\[ \tau_t = (1 - \theta_t) (r + \lambda \gamma_2 + \alpha_2) + \theta_t r \]

\[ = \frac{r}{I} + (1 - \theta_t) (\alpha_2 + \lambda \gamma_2) \]

(10)

where \( \theta_t \) is defined in the same way as above. The regression discontinuity effect of graduating from a prestigious university on wages at experience level \( x \) is composed by two terms. The first term \( I \) represents the direct effect of \( s \) on the workers productivity. The second term \( L \) represents the fact that employers do not observe \( \eta \) and \( z \) and use \( s \) as a signal for these two variables. In other words, if firms statistically discriminate among workers on the basis of university prestige, we have that \( L > 0 \). However, the signaling term \( L \) becomes less important for earnings as firms learn about a workers true productivity, \( \tau_t \) decreases with \( t \) and converges to \( r \) as \( \theta_t \) goes to 1.

There are is an important difference between the regression discontinuity test we propose and the traditional employer learning test: the parameter \( \tau_t \) does not depend on the relation between \( s \) and
In other words, the regression discontinuity test is robust to the existence of characteristics that could be used for statistic discrimination that are related to graduating from a prestigious university and that are not present in the data. This difference is important because, as discussed above, the traditional EL-SD test might confound statistical discrimination based on family socioeconomic status and statistical discrimination based on college prestige since these factors are intrinsically related and we do not observe family socioeconomic status in the data.

4.2 The Admission Process and the RD Design

Our data contains information on the year a student took the PAA test, his or her scores on each component of the test, the college he or she graduated from and the major. We do not observe application decisions and therefore have to make extra assumptions and sample restrictions to perform the regression discontinuity design. Precisely, we restrict the data to individuals who graduated with engineering, business, medical and law degrees (competitive majors) and assume that these workers would prefer to graduate with these majors in a less prestigious university rather than study a different major in a prestigious university. Under this assumption, we can interpret that workers just above the admission cutoff (competitive major at prestigious universities) are those who were accepted to the highest program of their preference and those below the threshold (competitive major in less prestigious college) are those who were accepted to the second highest program of their preference. We find evidence that this is a plausible assumption. First, these are the programs with highest admission cutoffs and therefore should be top choices of applicants. Second, there is a positive wage differential between workers with the competitive majors in less prestigious university and workers with less competitive major in prestigious university. We interpret this as evidence that students have incentives to study engineering, business, medical or law degree at a less prestigious rather than other major in a prestigious university.

Using additional data on the PAA weights used by these programs in the two prestigious universities we are able to reconstruct the final weighted score for all individuals in the restricted
sample. As a result, we derive $Univ_1.Score_i$ and $Univ_2.Score_i$ that represents the PAA weighted score of individual $i$ at prestigious university 1 and 2 respectively.

Given the possibility that a student can be accepted in two, one or neither of the prestigious universities, we define the running variable used in the RD as follows:

$$Dist.Cutoff_i = \max\{Univ_1.Score_i - Univ_1.Cutoff_i, Univ_2.Score_i - Univ_2.Cutoff_i\}$$

where $Univ_1.Cutoff_i$ and $Univ_2.Cutoff_j$ are the admission score cutoffs used by universities 1 and 2 for individual $i$’s major in the year of application to college. Note that individuals with $Dist.Cutoff_i$ slightly greater than zero were barely admitted to at least one of the two prestigious universities and individuals with slightly lower than zero were barely reject by both schools.

In the RD design we will be interested in the following object:

$$\tau_t = \lim_{Dist.Cutoff_i \downarrow 0} E[w_{it} | Dist.Cutoff_i] - \lim_{Dist.Cutoff_i \uparrow 0} E[w_{it} | Dist.Cutoff_i]$$

$$- \lim_{Dist.Cutoff_i \downarrow 0} E[g_i | Dist.Cutoff_i] + \lim_{Dist.Cutoff_i \uparrow 0} E[g_i | Dist.Cutoff_i]$$

where $g_i$ is an indicator if worker $i$ graduated from an elite university, $t$ measures years of experience in the labor market, and $w_{it}$ is the log(wages) after $t$ years of experience. Note that the parameter $\tau_t$ represents the local average treatment effect on earnings after $t$ years of experience for workers around the admission cutoffs that would enroll in a prestigious university if they were admitted (intent-to-treat effect).

The employer learning-statistical discrimination RD test we propose consists of estimating if $\tau_t$ decreases with $t$. The test is based on the assumption that the unobserved ability ($\eta_i$) is positively correlated to graduating from a selective university but is continuous around the admission cutoff.

In this framework, assuming that firms do not observe $Dist.Cutoff_i$, they will use information on

---

19 We were only able to obtain PAA weights for years starting in the year 2000. In order to construct final scores for individuals that too the PAA prior to 2000, we assume that programs used the same weights for previous years. The evidence is that programs do not change weights over time.

20 Information on program admission cutoffs were collected at the universities websites (late application years) and newspapers (early application years). We find that 4% of individuals in our restricted sample with a prestigious university degree have weighted scores lower than the admission cutoffs. This could be justified by measurement errors in the admission cutoffs and weighs used in the paper or transfers from less prestigious universities. We drop these individuals from the sample used in the RD analysis.

21 For a discussion of the relationship between regression discontinuity design and treatment effects, see Lee and Lemieux (2010)
college prestige to predict that workers just above the admission cutoff have a higher $\eta_i$. However, the wage differential between those above and below the cutoff should decline if firms learn the true distribution of $\eta_i$ as workers gain experience and therefore should rely less on college prestige to set wages.

4.3 Results

We first address the empirical question if the probability of graduating from one of the two prestigious universities in Chile is discontinuous at the admission cutoff. Note that it is possible that individuals with a higher score than the admission cutoffs decided to attend a less prestigious university, which implies that we have a fuzzy regression discontinuity design. Figure 4 shows graphically the discontinuity in the probability of graduating from a prestigious university at the cutoff. From the figure, we find that the discontinuity in graduation from a prestigious university is approximately 60 percentage. This means that around 60% of the individuals with PAA scores just sufficiently high for admission choose to attend an elite university. Consequently, being just above the admission cutoff causes a large increase in the probability of graduating from a prestigious university in Chile, which is a necessary condition for the validity of the RD design.

Next, in figure 5 we present further evidence for the validity of the RD design. The basic idea is to test if there is a jump at the discontinuity for per-treatment variables that should not be affected by the treatment. Precisely, if being above or below the cut-off is random, we should observe a zero treatment effect on the probability of being female or graduating from a private high school (Imbens and Lemieux (2008)). The figure suggests that there is no discontinuity of these variables around the cutoff. In fact, from a formal test using the same specification in columns (1) to (3) of table 3 but using female or private high school indicator as dependent variable, we cannot reject at reasonable levels of significance that there are zero effects of being above the cut-off on these per-treatment outcomes.\footnote{Due to space constraint we omit the tests here, but they are available under request.}

\footnote{Note that in section 4 we also assume that firms cannot observe $Dist.Cutoff_i$. Furthermore, screening workers is expensive and firms learn fast (Lange (2007)), therefore it is not economically attractable to perform ability tests on recent college graduates.}
In order to present evidence of the effects of admission to a selective university on earning, we plot in figure 6 unconditional means of log annual earnings on the vertical axis and the distance from the admission cutoff on the horizontal axis for the first 4 years of labor market experience. The open circles represent 16 points local average and the lines represent linear fits of the data below and above the admission cutoff. The figure shows that there is a jump in earnings in the first year of labor market experience for workers who are just above the cutoff. This discontinuity is consistent with previous literature that finds a significant effect on earnings for being just above the admission cutoff of recent college graduates (Saavedra (2008)). However, as workers gain labor market experience, the discontinuity in earnings tend to decrease to the point that there is no apparent difference in terms of earnings between workers just above and just below the cutoffs four years after graduation. In addition to that, we observe that workers tend to be paid more in accordance with their weighted score as they accumulate experience in the market.

Table 3 presents further statistical evidence for discontinuity in earnings at the admission cutoff. In columns (1) to (3) of panel A of the table, we show that workers above the admission cutoff have on average 6-8% higher earnings than just below the admission cutoff in their first 10 years of labor market experience (varying little with bandwidth). In columns (4) to (7) we present specification that allows that the return to being approved at a selective university to change along a worker’s career. Under this specification, we estimate a 10%-14% of wage premium for those above the cutoff in their first year of labor market experience, but this differential decreases by 1.5 to 2.7 percentage points per year of experience.

In Panel B of Table 3 we present the earnings discontinuity estimates taking into consideration that not all applicants with sufficiently high scores enroll in the top universities. For this purpose, we estimate an earnings equation using a two-stage least square method, where both graduating from a prestigious university and its interaction with experience are instrumented with an indicator for PAA scores above the admission cutoff and its interaction with experience. We estimate a 16-22% effect of graduating from a selective university on earnings of recent college graduates. However, this gap decreases by 2.1-3.7 percentage points per year of experience in the labor market. Note
that these estimates should be interpreted as the casual effect only for those applicants that would enroll in a prestigious university and graduate in the event of achieving a sufficiently high score (intent-to-treat effect).

In order to provide a robustness checks for the main RD findings, we present in table 4 estimates for the earnings discontinuity at the admission cutoff and its interaction with experience for different model specifications. Precisely, we show in row (1) that our estimates are not sensitive to the exclusion of controls, which is expected if treatment is random around the admission cutoff. In rows (2) and (3) we test how our estimates change with different specifications for the distance from the admission cutoff. Finally we estimate our preferred model for males and females separately. While we estimate similar coefficients for these two groups, we do not find a significant change in the returns to being approved by a prestigious university with experience for women. We notice however that this result is due to large standard errors that might be explained by the fact that we have a smaller fraction of women in the restricted sample.

5 Conclusion

This paper tests whether firms statistically discriminate based on the selectivity of the university attended by workers. We first follow the employer learning statistical discrimination test suggested by Altonji and Pierret (2001) and show that the returns to graduating from an elite university in Chile decreases with experience and that the returns to hard-to-observe ability correlates increase with experience. These results are interpreted as evidence for statistical discrimination based on university selectivity.

Furthermore, we take advantage of the centralized admission process of traditional universities in Chile to propose a statistical discrimination test based on a regression discontinuity design. We show that recent graduates just above the admission cutoff to the most prestigious universities in Chile have significantly higher earnings than those just below the cutoff. However, as workers gain labor market experience, the earnings gap between these two groups decreases to the point that we cannot reject zero wage differentials 4 years after graduation. We interpret this result as firms
paying workers in accordance with the selectivity of their college when they graduate from school, but rewarding them based on their true productivity as they reveal their quality to employers.

Our results shed some light on the benefits of graduating from a selective university. We interpret our findings as evidence that attending a prestigious university has a significant impact on signaling to firms a worker’s unobservable quality. However, employers learn fast and individuals tend to be paid in accordance with their true ability as they gain experience in the labor market.
References


Hoxby, Caroline M (1998), ‘The return to attending a more selective college: 1960 to the present’, Unpublished manuscript, Department of Economics, Harvard University, Cambridge, MA.


Kaufmann, Katja Maria, Matthias Messner and Alex Solis (2012), Returns to elite higher education in the marriage market: Evidence from Chile, Technical report, Bocconi University working paper.


Table 1: Descriptive Statistics for Selective and Non-Selective Universities

<table>
<thead>
<tr>
<th>Variables</th>
<th>Selective Universities</th>
<th>Non-selective Universities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>Female</td>
<td>0.53</td>
<td>0.5</td>
</tr>
<tr>
<td>Language PAA Score</td>
<td>680.6</td>
<td>61.3</td>
</tr>
<tr>
<td>Math PAA Score</td>
<td>715.9</td>
<td>68.6</td>
</tr>
<tr>
<td>High School Grade</td>
<td>644.5</td>
<td>78.7</td>
</tr>
<tr>
<td>Private High School</td>
<td>0.11</td>
<td>0.32</td>
</tr>
</tbody>
</table>

Number of Individuals: 11,554, 46,923

Note: Math and Language PAA scores are components of the centralized test for admission in University in Chile. See section 3 for definition of selective university.

Table 2: Earnings for Selective and Non-Selective Universities

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Selective Universities</th>
<th>Non-selective Universities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log of Annual Wage (in 1999 Pesos)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>15.58</td>
<td>15.19</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>1.09</td>
<td>1.16</td>
</tr>
<tr>
<td>Observations</td>
<td>61,844</td>
<td>251,233</td>
</tr>
</tbody>
</table>

Note: See section 3 for definition of selective university.
Table 3: **EL-SD Regression Discontinuity Test**

**Panel A**

<table>
<thead>
<tr>
<th>Model</th>
<th>Bandwidth (Points from Cutoff)</th>
<th>Reduced Form</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>125</td>
<td>100</td>
</tr>
<tr>
<td>Approved at Selective University</td>
<td>0.081</td>
<td>0.073</td>
</tr>
<tr>
<td></td>
<td>(0.0287)** (0.0308)** (0.0342)*</td>
<td>(0.0330)** (0.0348)** (0.0382)**</td>
</tr>
<tr>
<td>Approved at Selective Univ.* Experience</td>
<td>-0.027</td>
<td>-0.020</td>
</tr>
</tbody>
</table>

| Observations | 39,748 | 36,639 | 31,843 | 39,748 | 36,639 | 31,843 |
| R-squared    | 0.135  | 0.128  | 0.120  | 0.135  | 0.128  | 0.120  |

**Panel B**

<table>
<thead>
<tr>
<th>Model</th>
<th>Bandwidth (Points from Cutoff)</th>
<th>Reduced Form</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>125</td>
<td>100</td>
</tr>
<tr>
<td>Graduated from Selective University</td>
<td>0.133</td>
<td>0.122</td>
</tr>
<tr>
<td></td>
<td>(0.0470)** (0.0512)** (0.0576)*</td>
<td>(0.0525)** (0.0565)** (0.0629)**</td>
</tr>
<tr>
<td>Graduated from Selective Univ.* Experience</td>
<td>-0.037</td>
<td>-0.028</td>
</tr>
</tbody>
</table>

| Observations | 39,748 | 36,639 | 31,843 | 39,748 | 36,639 | 31,843 |
| R-squared    | 0.140  | 0.133  | 0.125  | 0.140  | 0.134  | 0.126  |

**Approved at Selective Univ.: Points from the Cutoff= 0**

**Controls:** Points from the Cutoff, and Interaction of Points from the Cutoff with Approved at Prestigious Univ., Female, Cubic Experience Polynomial, Major Dummies, and Year Dummies.

**Instrument in Panel B:** In columns (1)-(6) the endogenous variables are instrumented with Approved at Prestigious University and in columns (4)-(6) also with its interaction with experience.

White/Huber standard errors accounting clustered at the individual level are reported in parentheses

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

Note: The sample is restricted to individuals with engineering, business, medical and law degree (see section 5 for details).

---

26
Table 4: **EL-SD Regression Discontinuity Test - Robustness Checks**

<table>
<thead>
<tr>
<th>Regression Specification</th>
<th>Additional Controls</th>
<th>Function of Points from the Cutoff</th>
<th>Flexible Coefficient?</th>
<th>Sample</th>
<th>Estimated Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Approved at Selective University</td>
</tr>
<tr>
<td>(1)</td>
<td>No</td>
<td>Linear</td>
<td>Yes</td>
<td>All</td>
<td>0.130</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0354)***</td>
</tr>
<tr>
<td>(2)</td>
<td>Yes</td>
<td>Cubic</td>
<td>No</td>
<td>All</td>
<td>0.103</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0426)**</td>
</tr>
<tr>
<td>(3)</td>
<td>Yes</td>
<td>Cubic</td>
<td>Yes</td>
<td>All</td>
<td>0.109</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0425)**</td>
</tr>
<tr>
<td>(4)</td>
<td>Yes</td>
<td>Linear</td>
<td>Yes</td>
<td>Males</td>
<td>0.100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0427)**</td>
</tr>
<tr>
<td>(5)</td>
<td>Yes</td>
<td>Linear</td>
<td>Yes</td>
<td>Females</td>
<td>0.153</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0599)**</td>
</tr>
</tbody>
</table>

All specifications include Cubic Experience Polynomial.

**Approved at Selective Univ.:** Points from the Cutoff $\geq 0$

**Additional Controls:** Female, Major Dummies, and Year Dummies.

**Flexible coefficient** indicates whether the estimated coefficients of points from cutoff was allowed to differ on each side of the admission cutoff

White/Huber standard errors clustered at the individual level are reported in parentheses

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

Note: The sample is restricted to individuals with engineering, business, medical and law degree (see section 5 for details).
Figure 1: Application Process to Traditional Universities

<table>
<thead>
<tr>
<th>Taking the PAA Test</th>
<th>Result of the PAA Test</th>
<th>Application to Programs</th>
<th>Results and Enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Week of Dec.</td>
<td>1st Week of Jan.</td>
<td>2nd Week of Jan.</td>
<td>3rd Week of Jan.</td>
</tr>
</tbody>
</table>

Figure 2: Smoothed Language PAA Score Distribution

Note: Language PAA is a component of the centralized test for admission to university in Chile. See section 3 for definition of selective university.
Figure 3: Smoothed Math PAA Score Distribution

Note: Math PAA is a component of the centralized test for admission to university in Chile. See section 3 for definition of selective university.
Figure 4: **Graduation from Selective University Discontinuity**

Note: Open circles represent 16 points local averages and the lines are local linear fits below and above the admission cutoff. The sample is restricted to individuals with engineering, business, medical and law degrees (see section 5 for details).
Figure 5: Discontinuity at Pre-treatment Outcomes

Note: Open circles represent 16 points local averages and the lines are local linear fits below and above the admission cutoff. The sample is restricted to individuals with engineering, business, medical and law degrees (see section 5 for details).
Figure 6: **Earnings Discontinuity by Experience**

Note: Earnings are defined as log annual wages measured in real Chilean pesos. Open circles represent 16 points local averages and the lines are local linear fits below and above the admission cutoff. The sample is restricted to individuals with engineering, business, medical and law degree (see section 5 for details).