Strengthening state capabilities and entrepreneurship: Causal evidence from Brazil

Thiago de Lucena

Diana Moreira

June 27, 2024

Abstract

Appropriate allocation of talent is important for firm creation and economic growth. In many countries, the public sector absorbs a large share of the most talent due to its size and high public sector premium, especially among high skilled occupations. Moreover, the most talented often spend years pursuing those chances, and once on the job their human capital skills might be underutilized . While this allocation will strengthen state capability, it may crowd out other important economic activity. In this paper, we investigate the extent of this crowd out focusing on business creation. Exploring a regression discontinuity design on a novel data set of a large public service entry exam with more than 1.5 million applicants, we find that being approved on a public sector entry exam reduces the probability of creating a firm in subsequent years by 33 % percent. To shed light on the extent that employees are not being pushed to its full potential, we examine attendance around the opening of a business in an event-study design. We show that the opening of a business has limited effects on absenteeism of public sector workers that are firm owners.

1 Introduction

The misallocation of talent is an important barrier to firm creation and economic growth (Hsieh et al, 2019) academic talent has been show to be associated with more successful firms . In numerous developing countries the public sector is likely a major magnet of academically talented people. This is because it usually pays very high wages and admission is through very competitive exams, hence positions end up being granted to extreme academic talented individuals. While employees in the public sector are potentially producing highly valuable public goods, there is little understanding of what is being missed out with this allocation. Moreover, the process is perceive to drain brains at least through two mechanisms: it is believed that candidates often stay multiple years dedicated to winning the job lottery and when admitted lacks the work environment to entice the candidate to reach its full potential. An alternative perspective is that the academically talented who self select into the public sector might be more risky averse than usual, with limited effects on firm creation.

In this paper we investigate the costs in terms of forgone form creation and economic activity of talent that is employed in the public sector in Brazil. We ask several related questions: does getting a job in the public sector affects the probability of the candidate to create a firm? How does not winning a job is sunk in terms of making individual stay outside of the labor force for multiple years? Is there evidence of talent underutilization?

our analysis is divided in three parts. in the first part we document

Theoretically, there are several channels through which the public sector may affect incentives for its members to open firms. First, an analysis considering the model Lucas Jr (1978) and assuming that an agent earns higher earnings in the public sector than he could otherwise earn in the private sector (Finan *et al.* (2015)) may lead one to believe that joining the public sector increases agents opportunity cost therefore reducing incentives of one becoming an entrepreneur. However, this does not have to be the case because, as noticed by Finan *et al.* (2015), the profile of public sector workers differ starkly from private sector ones. It is possible that public servants would not create firms even if they were in the private sector; in this case, effects may be non-existent. Furthermore, different from contemplated

by Lucas Jr (1978), a nascent body of literature suggests that individuals may switch from employment to entrepreneurship in a non-dichotomous way; that is, an individual can engage in both entrepreneurship and paid employment simultaneously (Demir *et al.* (2020)). In this case, it is unclear if switching from the public to the private sector would lead to differences in entrepreneurship engagement. Hence, a priori, it is impossible to know if joining the public sector will lead to a lower probability of becoming an entrepreneur or what the magnitude of this decrease would be.

This is, to the best of our knowledge, the first study to estimate the causal effects of joining the public sector on firm creation. Using a novel data set for two large-scale public service entry exams for the Brazilian Social Security System, one of Brazil's largest public institutions, we use a Regression discontinuity design to compare firm creation rates for applicants who were closely approved to those who closely failed to enter to enter public service. Our results show that entering public service in Brazil is associated with a decrease in the probability of opening a firm in the years post-exam by almost 33 %. This result is robust to a series of different bandwidths and functional forms. Crucial to our analysis and link to growth, our definition of firms excludes micro-firms, commonly associated with survivalist self-employment activities, hence focusing on opportunity-driven entrepreneurship that has been suggested to promote structural transformation and induce growth (Naudé (2008), Gries & Naudé (2010)).

Furthermore, we explore the fact that Brazil, similar to other countries, regulates the entrepreneurship activities of public servants. We show that the decrease in entrepreneurship is not driven by types of entrepreneurship in which Brazilian public servants face bans, suggesting that the effects can be interpreted as the public service providing lower non-legal incentives for firm creation.

Finally, using a novel data set with daily biometric attendance for more than 30,000 social security system workers, we employ an event study methodology to assess the impact of creating a firm on public servant attendance. We find no evidence that creating a firm increases a worker's probability of taking a vacation break or being absent for any other

reason. Put together, our results suggest that (1) Strengthening state capacity decreases entrepreneurship independently of imposing limitations on firm ownership (2) Imposing limitations on the ability of public servants to own firms has little cost; however, given that we do not observe any decrease in public servant attendance as a consequence of entrepreneurship, so are benefits.

This paper makes a few contributions to the literature. First it contributes to the literature of selection of public servants (Finan *et al.* (2015), Mocanu (2022), Deserranno (2019), Ashraf *et al.* (2020), Dal Bó *et al.* (2013)) and more specifically to the growing body of work that studies so on the context of public service exams (Dahis *et al.* (2023), Moreira & Pérez (2021)). It also speaks to the literature of determinants of entrepreneurship (Hamilton (2000), Parker *et al.* (2005), Kerr *et al.* (2018), Hvide & Oyer (2018)), especially studies that focus on hybrid entrepreneurship (Folta *et al.* (2010)) and to the literature that studies public policies aiming to increase private entrepreneurship (Caliendo & KÃŒnn (2011), Van Stel *et al.* (2007), Meager *et al.* (2003)).

This paper is organized as follows: Section provides reference about the context, Section 3 discusses data sets used in this study, Section 4 describes the methodology, Section 5.1 analyzes results and Section 6 concludes.

2 Context

2.1 Public Service Exam and INSS

Similar to several countries around the globe, the Brazilian constitution requires, since its creation in 1989, tenured-eligible public servants to be hired through public servant exams. While public service entry exams take place at municipal and state levels, federal-level public service entry exams tend to be the largest ones. One of the largest of these federal-level public service entry exams, with more than one million people combined over the three past editions, is the National Social Security Institute entry exam.

National Social Security Institute (*Instituto Nacional do Seguro Social*, henceforth, INSS) is a federal autarky responsible for managing the pension system and social security system in Brazil. INSS has more than 1,800 branches and a presence in more than 1,000 Brazilian municipalities; these branches are responsible for analyzing and operationalizing payments of the Brazilian public pension regime, which, as of 2023, has more than 40 million members. On its daily operations, branches assist pension members with requesting their benefits by providing verbal orientation, receiving physical documents, and creating requests on internal systems, among others.

In 2008, INSS's board elaborated a plan for the next 20 years of the institution, with one of its main goals being an expansion in the number of branches [1]. On that same year, a public service exam was created to hire workers for new and existing branches. Two other exams followed in 2012 and 2016. Applicants for available positions needed first to pay a fee and indicate a municipality and position that they were applying for. Two main categories of positions were possible: Technician and Analyst. The former category requires only a high-school degree and makes up more than 80 % of the available positions, and the latter is subdivided into several other areas and requires a tertiary degree.

Exams were held nationally, in different locations, on a single day and were composed of multiple choice questions divided into three different sections: basic, complimentary, and specific. Once graded, scores could fall into four categories: (1) Applicant can be disqualified based on score. (2) Applicant can be qualified but not approved (3) Applicant can be approved and waitlisted. (4) Applicant can be approved and be part of the first call. (1) Indicates that a candidate received a score so low that it is not even worth a ranking; this threshold is indicated prior to the exam. (2) Indicates that a candidate is ranked according to their preferred position and location but failed to obtain a high enough grade to place them on the waitlist. Hence, this candidate has zero probability of receiving an offer. (3) Approved-wait-listed candidates are those who did not obtain a score high enough to be part of the first call but have a score high enough to be part of the wait-list. The wait-list is

[1] https://www.jusbrasil.com.br/noticias/inss-20-anos-5-expansao-da-rede-de-agencias-leva-a-mais-cidadaos-os-servicos-previdenciarios/2256012

created based on the number of available positions per location. For example, in 2008, the number of approved candidates was twice the number of available positions, meaning that the number of wait-listed candidates was equal to the number of approved candidates. If applicants first called do not accept an offer or if active public servants in a given location retire, these applicants may be hired. Hence, these candidates have a probability of receiving an offer between zero and one. Finally, candidates whose grade awards them a ranking within the number of available positions are officially offered a contract to join INSS.

3 Data

3.1 Data Sets

Our data sets four public service exam comes from *Centro de Seleção e de Promoção de Eventos Universidade de Brasília* (Henceforth, CEBRASPE)a center associated with the Federal University of Brasilia, and one of the three major public service exam organizers in Brazil. CEBRASPE was responsible for organizing the exam in both 2008 and 2014. In both years, CEBRASPE made public, through their website, the grades of all candidates [2]. This information was released as a document in a PDF format, including each candidate's name, score, application number, job position, job location, and competition type. An example of these documents can be found in Figure 11. For both editions of the exam, CEBRASPE only released grades for candidates who achieved the qualification threshold. While this does not affect our LATE because candidates who are disqualified have grades that leave them too far away from the approval threshold, this only allows us to observe scores of 50,014 applicants on the 2008 exam and 320,281 on the 2015 exam. Table 1 gives a dimension of the selectivity of the exam, with more than 1.5 million people applying for less than three thousand positions.

Our data set for firm creation comes from Ministério da Fazenda (Ministry of Eco-

[2]For thee 2008 exam, grades of candidates that did not reach the minimum threshold for qualification were not released.

nomics), and it has information on all firms created in Brazil over the past 40 years. Importantly, this data set does not include firms that are considered Microfirms (MEI) by the Ministry of Economics. This classification was created by the Brazilian government to reduce red tape around self-employed workers and allows firms that have revenues less than R$ 81,000.00 per year and less than two workers to have a simplified tax scheme. Hence, our data set does not include self-employed workers.

We have official information on each company's tax ID number, initial start-up capital, industry, and partners. Furthermore, for each partner, we observe the type of partnership, their name, part of their social security number, and the year that they established a partnership.

Next, we proceed to merge both data sets. However, while data for firm creation has part of individual tax ID, which can serve as a unique identifier, data from CEBRASPE does not have this information, forcing us to use names to conduct the match. We explain the challenges of doing so and develop a mechanism to increase the accuracy of matches in Section 3.2

Lastly, our data set for the presence of public sector workers comes from SISREF. SISREF is a biometric system that records the presence of INSS workers in all branches of the country with a daily frequency. While this system was implemented in 2018 and data on the daily frequency of INSS workers has been available since then, we focus on the year 2019 since INSS moved to remote work for most of 2020 and 2021.

This data set contains the branch code, name, and daily frequency status for each INSS worker on each day of the year 2019. First, we define absence as not being present to work for any reason, including legal ones. For example, vacation leave is considered absence. We aggregate frequency to the weekly level. Next, using the method described in Section 3.2, we estimate the number of similarly named individuals in the Brazilian population. Finally, we match this data set with our firm creation data set.

3.2 Data Matching

Given that matches across data sets are done through full names, it is crucial for the quality of the match that we restrict our matches to cases where it is reasonable to assume that someone's full name is a unique identifier. This challenge can be mitigated if we have additional information about individuals in each data set; however, different from historical data where researchers usually have information on personal characteristics like date and region of birth but name strings that contain errors, we have very little information about characteristics of the applicants yet a very high-quality name string variable given that he is extracted from official records associated with tax id's [3].

In order to restrict our analysis to cases where the name is a unique identifier, we estimate the frequency of that full name in the Brazilian population for each full name in our data set. To do so, we first assume independence between surnames that compose the full name of an individual. For example, consider the name "Romário de Souza Faria":

$$P(\text{Full name} = \text{Romário de Souza Faria}) = P(\text{1st n.} = \text{Romário}) * P(\text{2nd n.} = \text{de Souza}) * P(\text{3rd n.} = \text{Faria})$$

We estimate the probability of someone having a certain first name using an official API from the Brazilian Bureau of Statistics (IBGE). This API returns the frequency of a given first name in the Brazilian population, according to the 2010 census. Due to privacy issues, the API does not disclose name frequencies for first names whose national frequency is below 20. In those cases, we input 20 as the estimated frequency. We then divide the frequency by the total Brazilian population (196 million) to estimate the probability of observing that name in the Brazilian population.

Unfortunately, IBGE does not provide any information about the frequency of last names. Hence, in order to estimate the frequency of each last name, we use private data

[3] For example see Abramitzky *et al.* (2021)

extracted from an internet website specializing in family history research. This data set contains the 1,000 most frequent surnames in Brazil.[4] For each individual surname in our data set, we first retrieve the frequency associated with it in the private data set, then divide the frequency by the total Brazilian population to get an estimate for the probability. If the surname is not among the top 1,000 most frequent surnames, we associate it with the frequency of the 1,000th most frequent least name of the sample. Hence, our estimator tends to overstate the frequency of uncommon surnames. We then divide each surname frequency by 196 million and multiply all first and last names. Doing so gives us an estimate of the probability of observing a name in the Brazilian population. Finally, multiplying this probability by the total of the Brazilian population yields the expected frequency of that name in Brazil. A list of the most frequent names can be found in Table 8.

4 Methodology

Our aim is to evaluate the impact of passing a public service entry exam on firm creation. We exploit the credibly exogenous source of variation of scores in a national public service entry exam. As previously mentioned, the probability of joining the public sector is a discontinuous function of the score of the applicant, enabling the use of a sharp regression discontinuity design (RDD) approach. Following Lee & Lemieux (2010), we estimate the following model:

$$FirmCreate_i = \beta_0 + \beta_1 * Approved_i + f(score_i - c_{b,y,p}) + \theta_{b,y} + \gamma_p + u_i \quad (1)$$

Where $Approved_i$ is a dummy which indicates whether individual i was approved on the exam or not ($Approved_i$=1) or not ($Approved_i$=0). $Score_i$ is the score of individual i on the exam. $c_{b,y}$ is the score cutoff of at branch b, and exam-year y, and f(.) are continuous functions of our running variable centered at the cutoff value. $FirmCreate_i$ indicates that the individual was listed as a partner in a firm in any period subsequent to the exam.

[4] https://forebears.io/brazil/surnames

The parameter of interest is β_1, captures the causal effect on $FirmCreate_i$ of being accepted into the public service. We estimate our equation assuming that f(.) is a flexible polynomial on both sides of the threshold. The running variable in our setting is the centered score of the individual, represented by $score_i - c_{m,y}$, which takes negative values for individuals with $Approved_i = 0$ and positive values for citizens with $Approved_i = 1$. $\lambda_{,yb}$ an branch-year fixed effect and γ_p position fixed effects. Following Gelman & Imbens (2019), we estimate only zero and first polynomials for different bandwidths, focusing on the [-10,+10] bandwidth. However, we also show that our results are robust in terms of both magnitude and significance to most other choices of bandwidth, including the optimal bandwidth calculated using the Calonico-Cattaneo-Titiunik (CCT) procedure from Calonico *et al.* (2014).

Since participants are graded on a discrete natural scale from 0 to 120, we place additional care to reduce bias that may arise due to the few mass points of our running variable around the threshold. Some strains of the literature suggest clustering standard errors on our running variable in order to account for the imperfect fit of the parametric function away from the threshold Lee & Card (2008). However, Kolesár & Rothe (2018) show that clustering standard errors by the running variable have poor coverage properties and argue there is no need to distinguish sharply between the case of a discrete and a continuous running variable, suggesting the use of heteroskedasticity-robust standard error for inference. We follow this approach by clustering our standard errors by location-year. Nevertheless, in our main tables, the approach suggested by Lee & Card (2008) leads to smaller standard errors; hence, our results are also robust to it.

To identify β_1 as a causal parameter, our design must satisfy two key assumptions. The first is that the selection criteria for public service is not subject to manipulation. In practice, this is very unlikely, given the objective, multiple-choice, and independent grading criteria of the exam. Nevertheless, we evaluate this possibility more formally by first presenting visual evidence of no manipulation around the cutoff (Figure 1). Next, we implement the procedure suggested by Frandsen (2017), which tests bunching of the running variable in a similar way to McCrary (2008) but allows for a discrete running variable. The results

for this test can be seen in Table 2; we fail to reject the null for all values of k. For the most conservative case, k = 0, we barely fail to reject the null at 10 %; however, assuming k = 0 here implies assuming a perfectly linear function around the cutoff. This assumption seems unreasonable, given the non-linearity around the cutoff. For any parameter value that allows for some flexibility of the polynomial, we fail to reject the null by a comfortable margin.

The second assumption requires baseline covariates to not change because of the treatment. Hence, we should not observe a treatment effect of passing the exam on a set of predetermined attributes of the applicants. We perform this analysis by estimating our preferred specification (Equation 1) using as dependent variable socioeconomic variables that we do not expect to change as a consequence of being accepted into public service, such as race, gender, disability status, and name frequency. The results can be found in Figure 4 (we focus on names with an estimated frequency of less than 0.01). There is no robust discontinuity around the threshold, increasing the internal validity of our design.

4.1 Absenteeism

The analysis of the impact of firms on absenteeism is still a work in progress and should only be interpreted as suggestive at this stage. Focusing on the four months before and four months after the creation of a firm, we estimate two different models making use of the daily absenteeism data aggregated at the weekly level described in Section 3.

$$Presence_{i,w} = \beta_0 + \beta_1 PostCreation_{i,w} + \theta_i + \lambda_w + u_{i,w}$$

We also estimate an event study model following the equation below:

$$Presence_{i,w} = \beta_0 + \beta_1 \sum_{\substack{j=-16 \\ j \neq -1}}^{16} WeekCreation_{i,w} + \theta_i + \lambda_w + u_{i,w}$$

Where $Presence_{i,w}$ is the average weekly presence of individual i on week w. $PostCreation_{i,w}$ is an indicator variable that takes a value equal to one for all weeks post-creation of a firm. θ_i denotes an individual fixed effects and λ_w is a week fixed effects. Finally, on the event study specification, $WeekCreation_{i,w}$ denotes a dummy for weeks relative to the week of creation of a firm.

5 Results

5.1 Firm Creation

We start our results section by focusing broadly on how acceptance into the public sector affects firm creation. We estimate Equation 1 on the sub-sample of full names that have an estimated frequency smaller than 0.01 and for a dependent variable that denotes firm creation with any type of ownership. Results for this analysis can be found in Table 3 and a visual representation of it in Figure 5. On average, being approved for public service is associated with a decrease in the probability of starting a new firm by 2.3 percentage points. This corresponds to a 29 % decrease relative to the mean. This result is robust to a zero-degree polynomial and robust in magnitude to a second-degree one. Figure 5 suggests that, indeed, a first-degree polynomial actually best captures the relationship between both variables around the cutoff. Figure 6 displays robustness tests relative to the size of the bandwidth.

Furthermore, we also analyze the impacts of passing the exam on starting a firm years prior to the exam. If the impact of passing a public service entry exam here is really causal, we should not observe any impact of being approved by the public sector on firm

creation in years prior to the exam. Table 4 displays this analysis, focusing again on individuals whose names have an estimated frequency below 0.01. On average, being accepted into the public sector is associated with a decrease in the probability of opening a firm by 0.6 percentage points. As expected, this estimate is non-significant, increasing the confidence that our main specification captures the causal impact of being accepted into the public service on firm creation.

Put together, these results indicate that simply being accepted into the private sector impacts the probability of firm creation later. However, less is known about the driver of these results. In particular, if these results are driven by types of ownership that are forbidden for public servants in Brazil. Table 5 addresses this question. Columns (1) and (3) display estimates for firms created where the creator is also a manager of the firm; Brazilian public servants are forbidden to take part in this type of firm ownership. Columns (2) and (4) display estimates for all other types of ownership.

Results for this analysis show that the restriction is not the main driver of the results. Decreases for both managerial and non-managerial positions are non-significant, with larger point estimates and lower p-values for non-managerial partners. If anything, non-managerial partners, a management type that does not face any restriction, seem to be the main driver of results suggesting that restrictions against firm ownership are not binding.

5.2 Absenteeism

We start our analysis of the impacts of firm creation on absenteeism by analyzing the most common statuses on SISREF. In practice, SISREF has more than 40 different codes for a daily status, with more than 35 of them implying some type of absence. The ten most common categories are displayed in Table 6. The most frequent entry is working overtime, with 43 % of the time. This means that the worker was present and clocked out after 8 hours of work. The code for presence follows being the input for around 13 % of the time.

Next, we define a worker as present if their daily entry was either present, worked overtime, or compensated for a holiday break, and we look at means of presence and absence

by month and days of the week. Presence peaks in May, with workers being present on an average of 72 % of the days. The months with the lowest average presence are December, January, and July. This is consistent with the anecdotal evidence of the vacation months in Brazil and overlaps with school breaks. [5]

We proceed to break down absence into two mutually exclusive categories: Vacation absences and non-vacation absences. We observe that vacation absences are one of the main drivers of low presence levels in January and July. Non-vacation absences peak in December and increase throughout the year. We can then compare these patterns with data on firm creation by month in Figure 9. We observe that firm creation peaks in May and October, roughly two months before vacation break months (December and July). Furthermore, firm creation is lowest in January, a month that also displays low presence and high vacation frequency. Nonetheless, without further analysis, it is impossible to conclude that firm creation causes a lower absence of public servants.

With the intent to deepen our understanding of this relationship, we perform the analysis described in Section 4. Table 7 displays this analysis. Odd columns display results for models without weeks fixed effects and even columns for models with it. Our preferred specification, column two, indicates that firm creators are, on average, 2.7 percentage points less likely to be present in the four-month period post-firm creation compared to four months before; this result is not statistically significant. We further break down the analysis by entrepreneur management type since it's possible that manager entrepreneurs perform higher levels of time-consuming tasks, hence having larger point estimates. Columns three and four display this analysis, and once again, point estimates are non-significant. The point estimate for the model without week fixed effects, although non-significant, has a large economic magnitude, 8.2 % of the mean. However, once we include week fixed effects the sign of the point estimate changes suggesting that the creation of firm by these types of management take place around weeks and months with a general high level of absence. Either way, this analysis does not find any evidence that firm creation leads to higher absenteeism. We hope

[5] With Brazil being located in the Southern Hemisphere, the "long" summer break takes place in December-January, while the short winter break takes place in July.

to improve the methodology for the panel analysis on further versions of this study.

6 Conclusion

This work analyzed the impact of passing a public sector entry exam on the probability of being an entrepreneur. Using a Regression Discontinuity design on a novel data set of two of the largest public sector exams in Brazil's history, we first show that being approved on a public service exam is associated with a 29 % decrease in the probability of being an entrepreneur. This result is robust to functional forms and bandwidths. Next, we show that this effect is not driven only by ownership types in which public servants in Brazil face legal restrictions. Finally, we analyze if creating firms is costly for the public sector in Brazil. Namely, we investigate if opening a firm is associated with higher absenteeism. While we find some seasonality in firms' openings owned by public sector workers, we find no evidence that these workers are more absent due to these firms' openings, a result that holds for both managerial and non-managerial firm partners.

References

Abramitzky, Ran, Boustan, Leah, Eriksson, Katherine, Feigenbaum, James, & Pérez, Santiago. 2021. Automated linking of historical data. *Journal of Economic Literature*, **59**(3), 865–918.

Ashraf, Nava, Bandiera, Oriana, Davenport, Edward, & Lee, Scott S. 2020. Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services. *American Economic Review*, **110**(5), 1355–1394.

Caliendo, Marco, & KÃŒnn, Steffen. 2011. Start-up subsidies for the unemployed: Long-term evidence and effect heterogeneity. *Journal of Public Economics*, **95**(3), 311–331. New Directions in the Economics of Welfare: Special Issue Celebrating Nobel Laureate Amartya Sen's 75th Birthday.

Calonico, Sebastian, Cattaneo, Matias D, & Titiunik, Rocio. 2014. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, **82**(6), 2295–2326.

Dahis, Ricardo, Schiavon, Laura, & Scot, Thiago. 2023. Selecting top bureaucrats: Admission exams and performance in brazil. *Review of Economics and Statistics*, 1–47.

Dal Bó, Ernesto, Finan, Frederico, & Rossi, Martín A. 2013. Strengthening state capabilities: The role of financial incentives in the call to public service. *The Quarterly Journal of Economics*, **128**(3), 1169–1218.

Demir, Cemre, Werner, Arndt, Kraus, Sascha, & Jones, Paul. 2020. Hybrid entrepreneurship: A systematic literature review. *Journal of Small Business & Entrepreneurship*, **34**(1), 29–52.

Deserranno, Erika. 2019. Financial incentives as signals: experimental evidence from the recruitment of village promoters in Uganda. *American Economic Journal: Applied Economics*, **11**(1), 277–317.

Finan, Frederico, Olken, Benjamin A, & Pande, Rohini. 2015. The personnel economics of the state.

Folta, Timothy B, Delmar, Frédéric, & Wennberg, Karl. 2010. Hybrid entrepreneurship. *Management science*, **56**(2), 253–269.

Frandsen, Brigham R. 2017. Party bias in union representation elections: Testing for manipulation in the regression discontinuity design when the running variable is discrete. *Pages 281–315 of: Regression discontinuity designs: Theory and applications*. Emerald Publishing Limited.

Gelman, Andrew, & Imbens, Guido. 2019. Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, **37**(3), 447–456.

Gries, Thomas, & Naudé, Wim. 2010. Entrepreneurship and structural economic transformation. *Small Business Economics*, **34**, 13–29.

Hamilton, Barton H. 2000. Does entrepreneurship pay? An empirical analysis of the returns to self-employment. *Journal of Political economy*, **108**(3), 604–631.

Hvide, Hans K, & Oyer, Paul. 2018. *Dinner table human capital and entrepreneurship*. Tech. rept. National Bureau of Economic Research.

Kerr, Sari Pekkala, Kerr, William R, Xu, Tina, *et al.* 2018. Personality traits of entrepreneurs: A review of recent literature. *Foundations and Trends® in Entrepreneurship*, **14**(3), 279–356.

Kolesár, Michal, & Rothe, Christoph. 2018. Inference in regression discontinuity designs with a discrete running variable. *American Economic Review*, **108**(8), 2277–2304.

Lee, David S, & Card, David. 2008. Regression discontinuity inference with specification error. *Journal of Econometrics*, **142**(2), 655–674.

Lee, David S, & Lemieux, Thomas. 2010. Regression discontinuity designs in economics. *Journal of economic literature*, **48**(2), 281–355.

Lucas Jr, Robert E. 1978. On the size distribution of business firms. *The Bell Journal of Economics*, 508–523.

McCrary, Justin. 2008. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics*, **142**(2), 698–714.

Meager, Nigel, Bates, Peter, & Cowling, Marc. 2003. An Evaluation of Business Start-Up Support for Young People. *National Institute Economic Review*, **186**, 59–72.

Mocanu, Tatiana. 2022. *Designing gender equity: Evidence from hiring practices and committees*. Tech. rept. Working paper.

Moreira, Diana, & Pérez, Santiago. 2021. *Civil service exams and organizational performance: Evidence from the pendleton act*. Tech. rept. National Bureau of Economic Research.

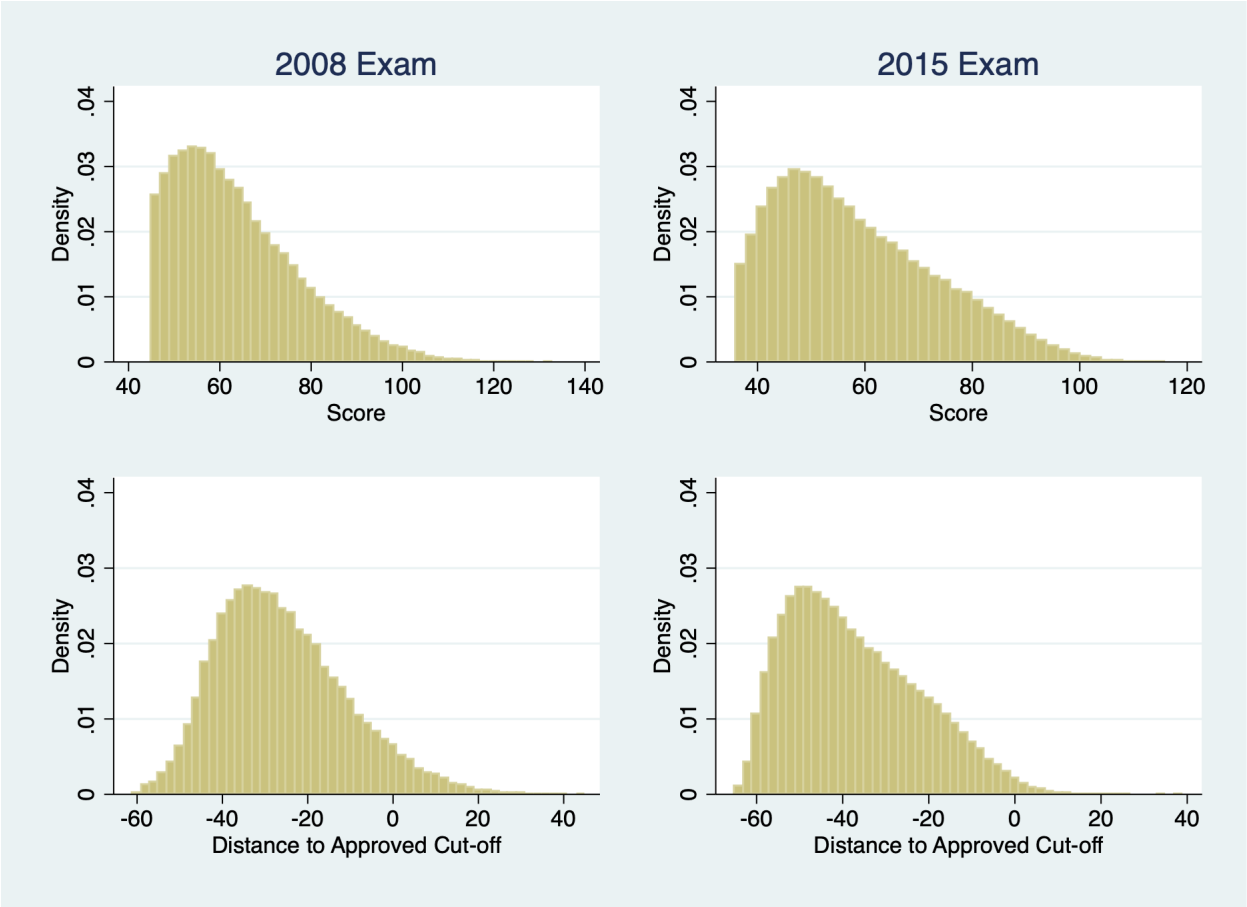
Naudé, Wim. 2008. Entrepreneurship in economic development.

Parker, Simon C, *et al.* 2005. The economics of entrepreneurship: What we know and what we don't. *Foundations and Trends® in Entrepreneurship*, **1**(1), 1–54.

Van Stel, André, Storey, David J, & Thurik, A Roy. 2007. The effect of business regulations on nascent and young business entrepreneurship. *Small business economics*, **28**, 171–186.

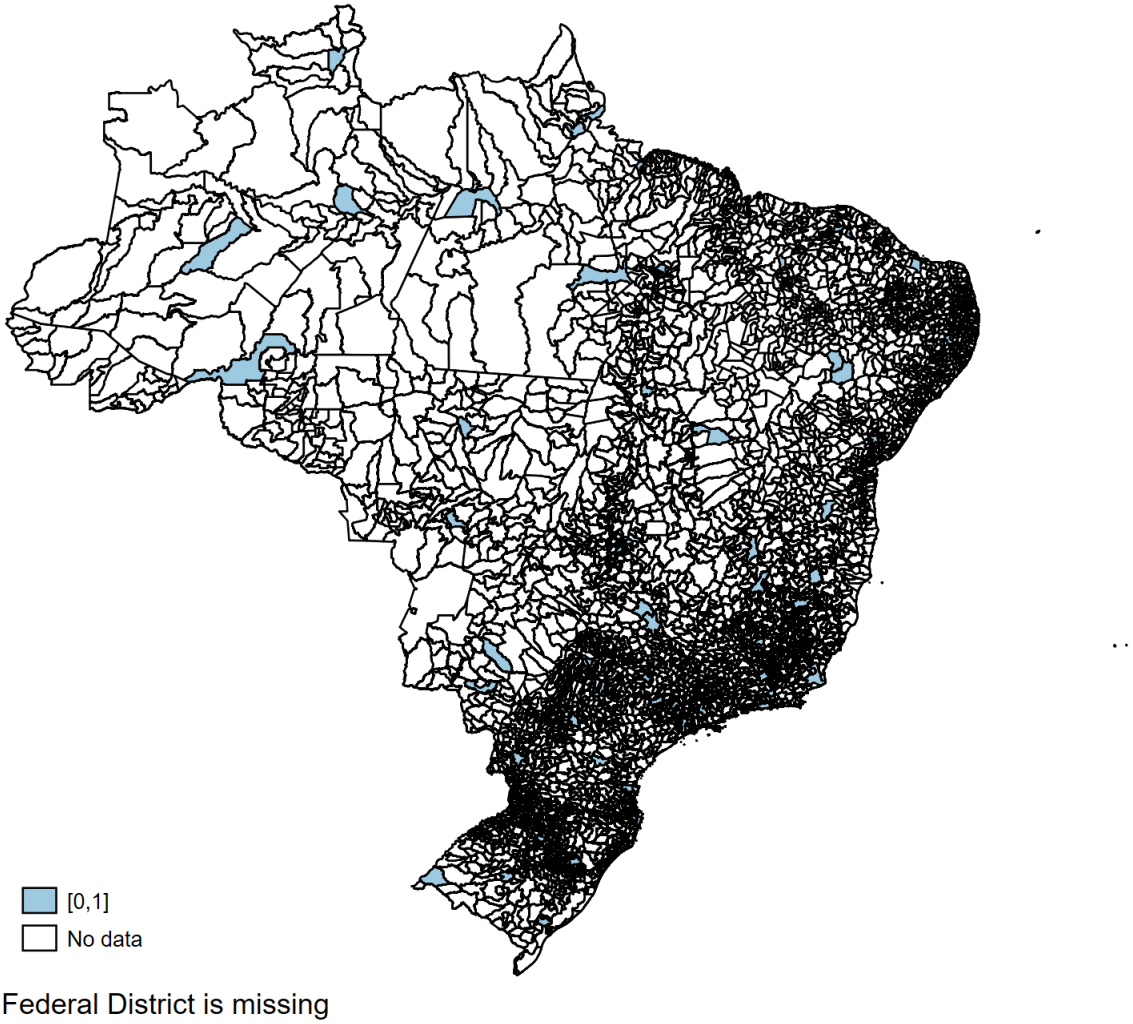
7 Figures

Figure 1: Scores and Distance to Approved Cutoff



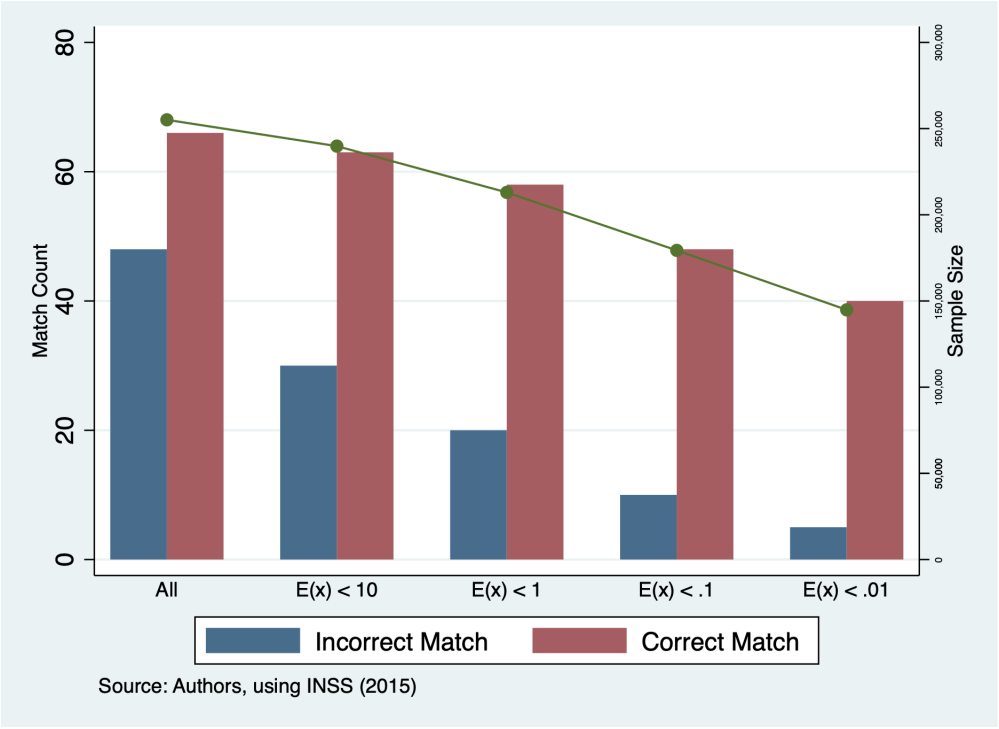
Note: Histogram plot of scores of qualified candidates. For a candidate to qualify for the exam, he/she must have had a score higher than 45 points on the 2008 exam and 35 points on the 2015 exam. Scores of applicants who failed to qualify were not reported by CEBRASPE. Distance to approved cutoff indicates the score distance to the lowest score approved applicant for non-approved applicants, and distance to the highest score non-approved applicant for approved candidates

Figure 2: Municipalities with an INSS open position in 2015



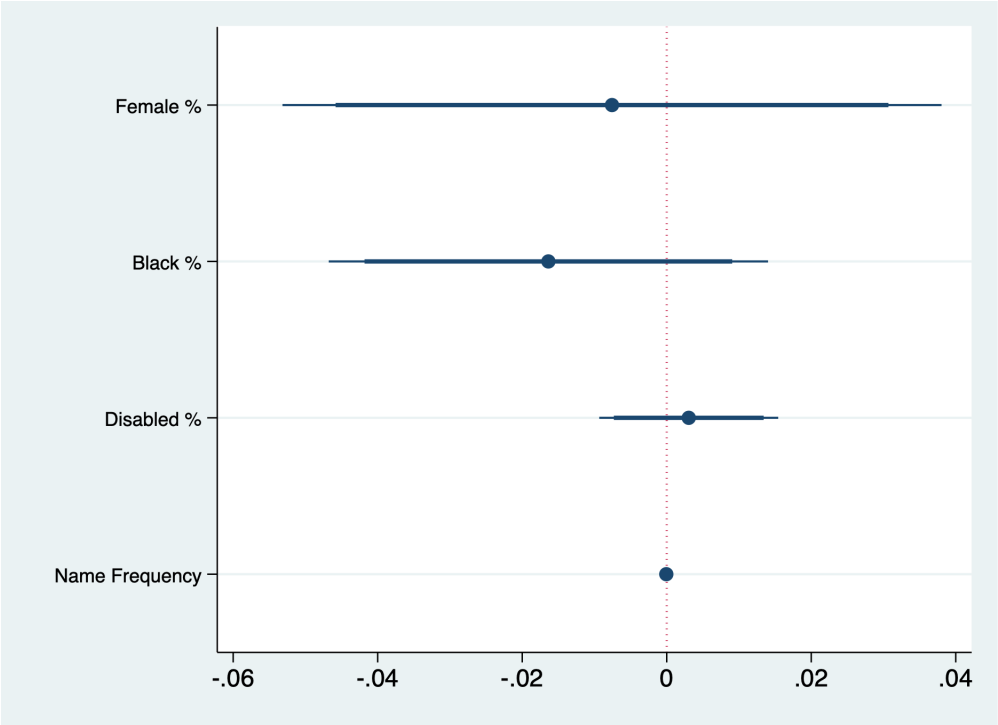
Note: Municipalities that hired INSS public servants on the 2015 exam, not including the Federal District. Includes both technicians and analysts.

Figure 3: Matches by Estimated Frequency



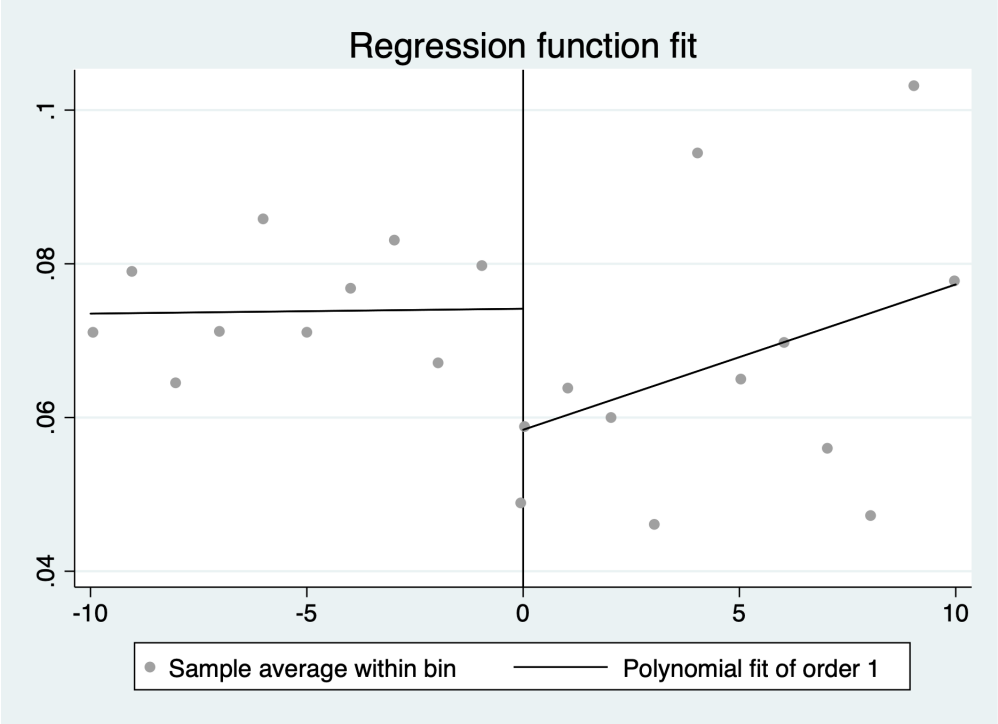
Note: The left vertical axis represents the total count of correct/incorrect matches based on the sample of 850 applicants approved on the first call of the 2015 exam. These numbers differ slightly from the ones in our sample (838). Correct match is defined as when, for a match, the six non-censured digits of CPF on the Receita Federal data set match six digits on the same position on the the result call. The horizontal axis displays different cutoffs of estimated name frequency. The right horizontal axis displays the total number of observations in 2015.

Figure 4: Balance test on Characteristics



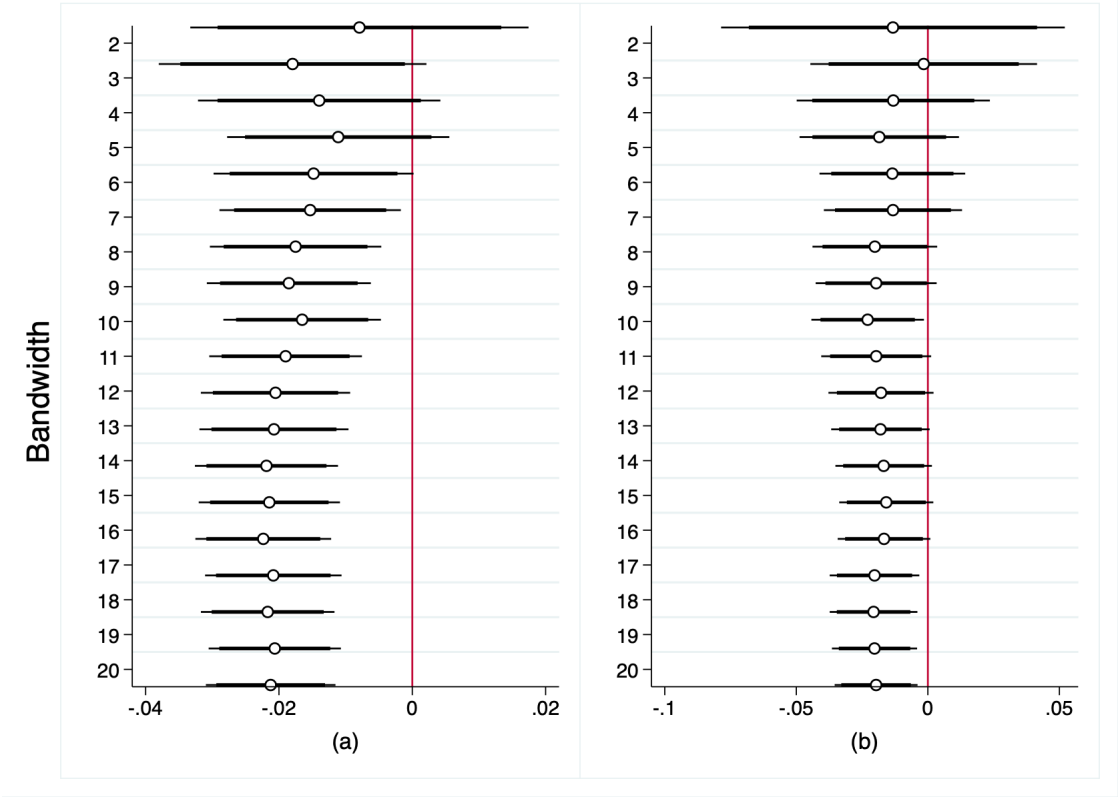
Note: Balanced tests using our preferred specification of first-degree flexible polynomial fit based on a 10-point bandwidth and a sample restricted to observations where $\hat{F} < 0.01$. Race and disability status is only available for 2015 and is defined as if the candidate applied for a position with that type of quota. Gender extracted from first name frequency, based on the 2010 Census, using GenderBR command on R.

Figure 5: Regression discontinuity plot



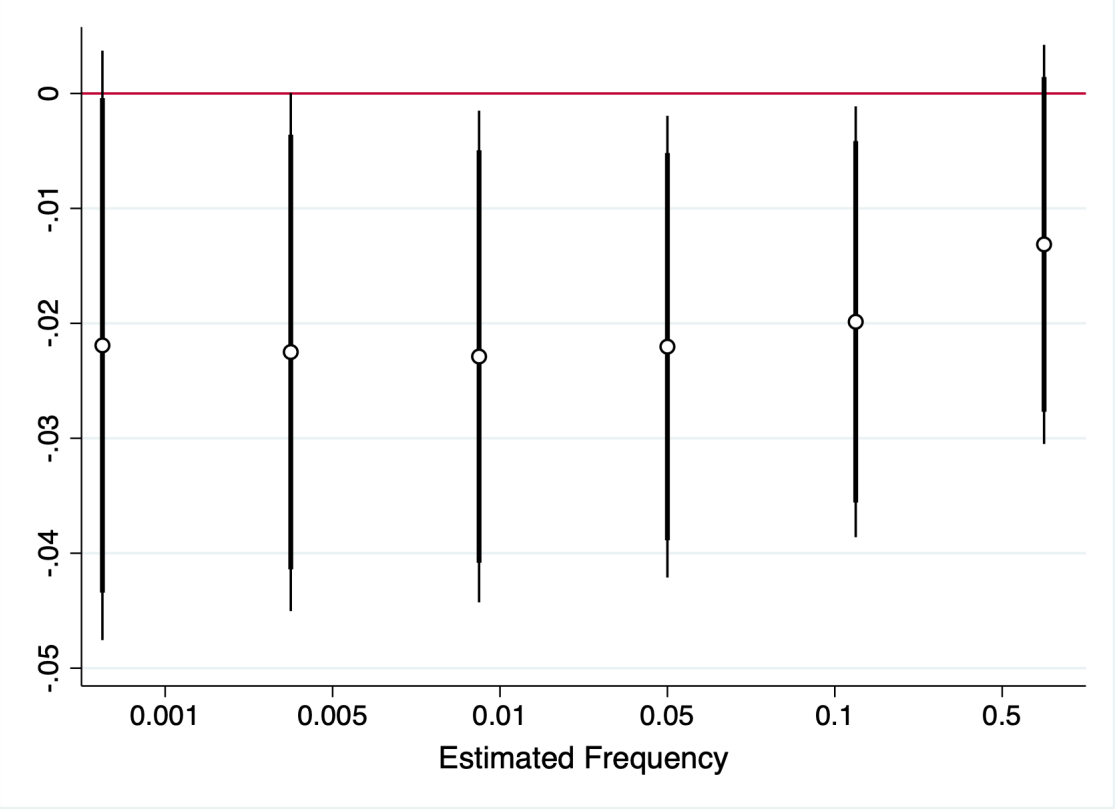
Note: First-degree flexible polynomial fit plot based on a 10-point bandwidth and a sample restricted to observations where $\hat{F} < 0.01$. Excludes competition for positions that were opened through quotas for either black or disabled applicants.

Figure 6: Bandwidth Robustness



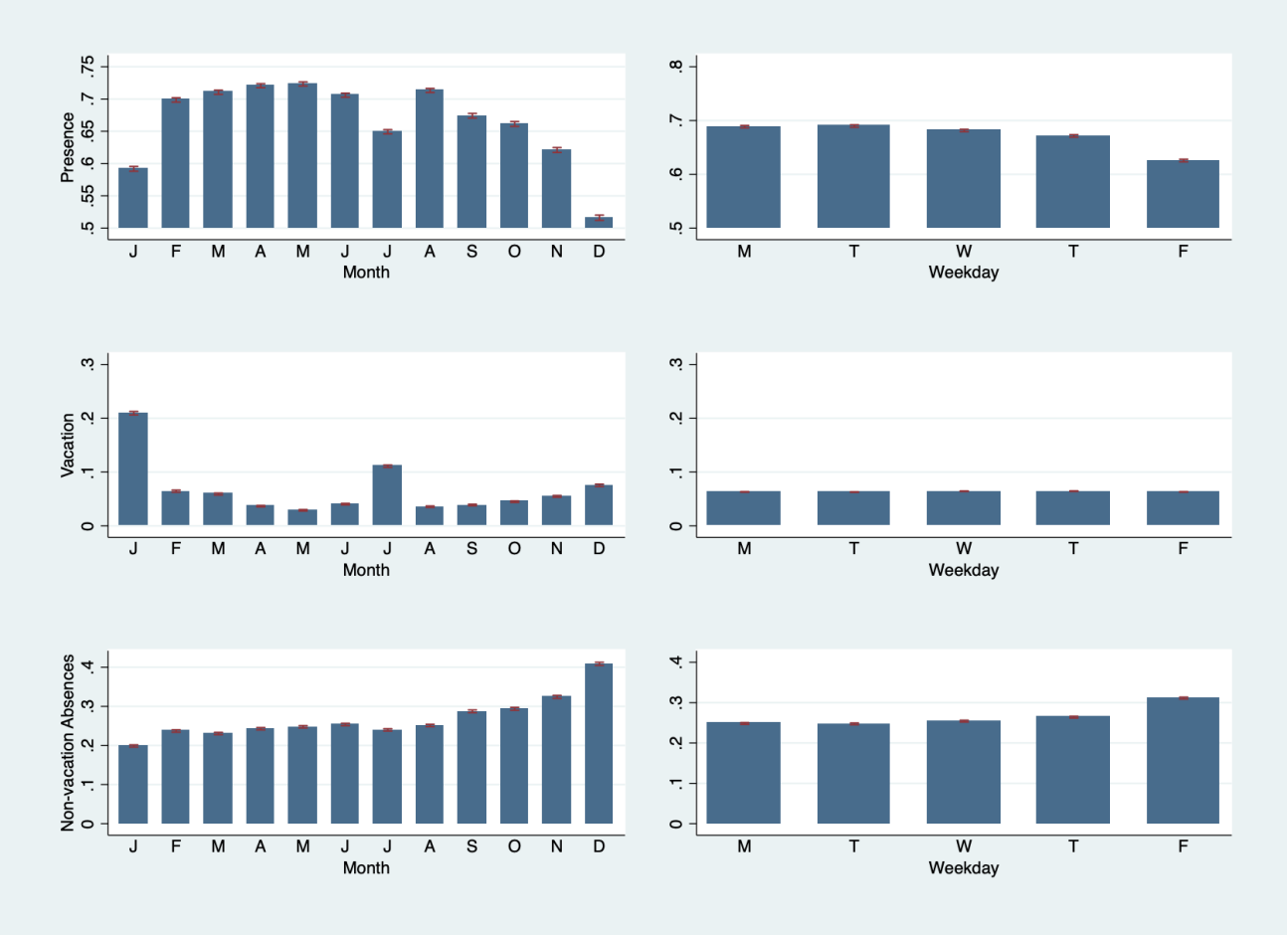
Note: (a) Displays estimates based on a zero-degree polynomial and (b) based on a first-degree flexible polynomial. Sample restricted to observations where $\hat{F} < 0.01$. Excludes competition for positions that were opened through quotas for either black or disabled applicants.

Figure 7: Frequency Cut-off Robustness



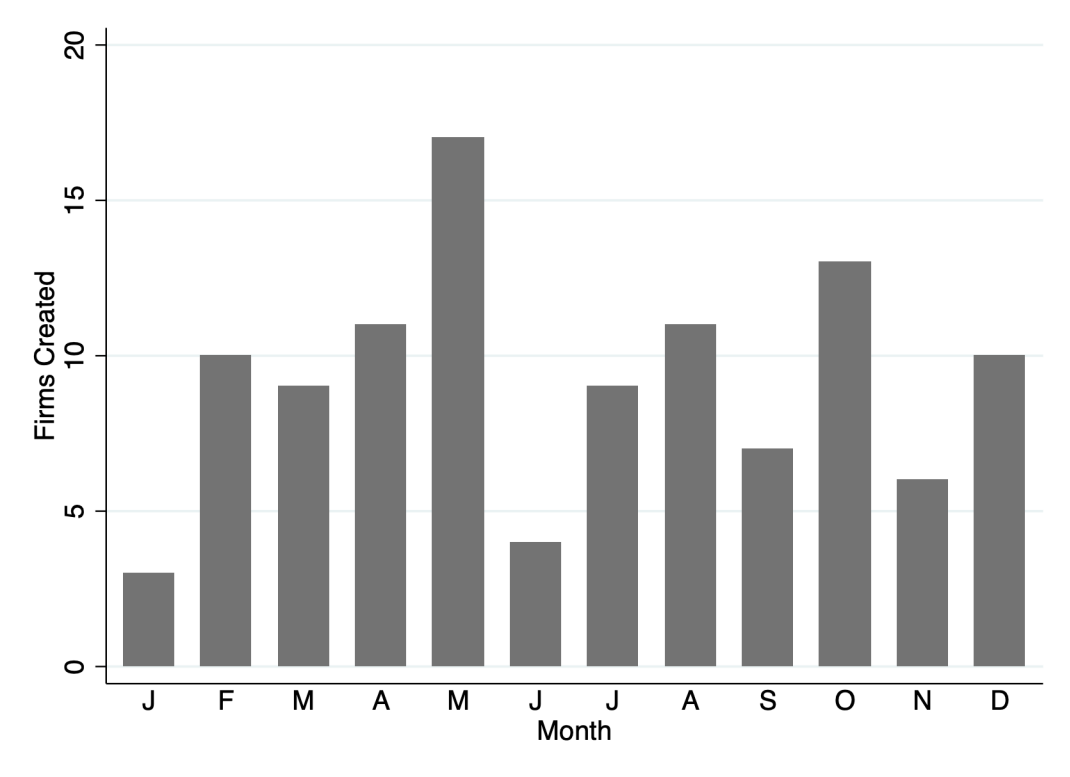
Note: (a) Displays estimates based on a zero-degree polynomial and (b) based on a first-degree flexible polynomial. Sample restricted to observations where $\hat{F} < 0.01$. Excludes competition for positions that were opened through quotas for either black or disabled applicants.

Figure 8: Worker presence by month and weekday



Note: Average worker frequency of workers by month and weekday excluding weekends and national holidays for the year of 2019. We also exclude workers who were always absent from our sample (1.26%). Presence on a given day is defined as a day in which the code present, holiday break compensation, or late was inputted. Non-vacation absence is defined as all other categories that are not absent due to vacation or presence.

Figure 9: Firm Creation by month



Note: Firms created by month by INSS employees for the year 2019. The sample was restricted to names with estimated frequency < 0.01.

8 Tables

Table 1: Total number of applicants by year

	Applicants	Qualified	Approved	Positions
Year: 2008				
Analyst	93,631	.	.	599
Technician	498,757	50,014	3,197	1,400
Year: 2015				
Analyst	43,639	9,737	678	147
Technician	1,035,834	310,544	2,798	791

Note: Analyst data for the 2008 Exam not available yet. Future versions of this work will include this data. The total number of positions reported according to extracted from documents differs slightly from official calls.

Table 2: Frandsen test

	k = 0.00	k = 0.02	k = 0.04	k = 0.06
p-values	0.104	0.143	0.264	0.444

Note: Table displays for p-values for Frandsen Test of manipulation of cutoff on Regression Discontinuity designs for select degree of departure from linearity around the threshold, k. k = 0 assumes a perfectly linear function, an unreasonable assumption in this case. Since the p-value is monotonic, increasing in k, and k $\in$ [0,1], we fail to reject the null for all values of k.

Table 3: Impact of Passing Public Service Exam on Firm Creation

	(1) 0 dgr.	(2) 1st dgr.	(3) 1st dgr. Flex	(4) 2nd dgr.
Approved	-0.017***	-0.023**	-0.023**	-0.021
	(0.006)	(0.011)	(0.011)	(0.016)
Dist. from Cut-off		0.001	0.001	-0.001
		(0.001)	(0.001)	(0.005)
Approved*Dist. from Cut-off			0.000	0.002
			(0.002)	(0.008)
Dist. from Cut-off ^{2}				-0.000
				(0.000)
Approved*Dist. from Cut-off^2				0.000
				(0.001)
Mean	0.072	0.072	0.072	0.072
N	11305	11305	11305	11305

Note: Standard errors under parenthesis. The dependent variable is an indicator variable that takes value one if the applicant is a firm partner post-exam (2009 onward for the 2008 exam and 2016 onward for the 2015 exam). Includes position and branch-year fixed effect. Standard errors clustered at the branch-year. The sample was restricted to individuals with expected name frequency < 0.01 and bandwidth of 10 points to cutoff. Excludes racial and disability quota positions.
* $p < .10$, ** $p < .05$, *** $p < .01$.

Table 4: Robustness test - Placebo Regression

	(1) 0 dgr.	(2) 1st dgr.	(3) 1st dgr. Flex	(4) 2nd dgr.
Approved	0.002	-0.004	-0.006	-0.003
	(0.005)	(0.011)	(0.011)	(0.016)
Dist. from Cut-off		0.001	0.000	-0.003
		(0.001)	(0.001)	(0.005)
Approved*Dist. from Cut-off			0.002	0.008
			(0.002)	(0.007)
Dist. from Cut-off ^{2}				-0.000
				(0.000)
Approved*Dist. from Cut-off^2				-0.000
				(0.001)
Mean	0.064	0.064	0.064	0.064
N	11305	11305	11305	11305

Note: Standard errors under parenthesis. The dependent variable is an indicator variable that takes value one if the applicant is a firm partner pre-exam. Includes position and branch-year fixed effect. Standard errors clustered at the branch-year. The sample was restricted to individuals with expected name frequency < 0.01 and bandwidth of 10 points to cutoff. Excludes racial and disability quota positions.
* $p < .10$, ** $p < .05$, *** $p < .01$.

Table 5: Impact of Passing Public Service Exam - Managerial Heterogeneity

	b = 10 pts		b = 20 pts	
	(1)	(2)	(3)	(4)
	Mngr. Partner	Non-M. Partner	Mngr. Partner	Non-M. Partner
Approved	-0.008	-0.013	-0.009	-0.012*
	(0.008)	(0.008)	(0.006)	(0.006)
Dist. Cut-off	0.000	0.000	-0.000*	0.000
	(0.001)	(0.001)	(0.000)	(0.000)
Approved*Dist. Cut-off	-0.003**	0.003	-0.002***	0.002*
	(0.002)	(0.002)	(0.001)	(0.001)
Mean	0.036	0.042	0.042	0.040
N	11305	11305	30444	30444

Note: Standard errors under parenthesis. The dependent variable is an indicator variable that takes value one if the applicant is a firm partner post-exam. Managerial partners are self-declared in IRS data. Includes position and branch-year fixed effect. Standard errors clustered at the branch-year. The sample was restricted to individuals with an expected name frequency < 0.01. Excludes racial and disability quota positions.
* $p < .10$, ** $p < .05$, *** $p < .01$.

Table 6: SISREF entry frquencies

	Freq.	Pct.	Cum. Pct.
Worked overtime	3,057,701	43.89	43.89
Present	966,667	13.87	57.76
Late or early leave	671,011	9.63	67.39
Holliday break compensation	564,739	8.11	75.50
On vacation	462,827	6.64	82.14
On business trip	125,262	1.80	83.94
Not present	116,296	1.67	85.61
No Contract	79,387	1.14	86.75
Others	923,437	13.25	100.00
Total	6,967,327	100.00	

Note: Frequencies of different entries for all INSS public servants for 2019. Frequencies exclude weekends and federal holidays. "Holiday break compensation" refers to workers working additional hours in order to take additional vacation days during the end-of-the-year break.

Table 7: Impact of firm creation on presence

	(1) All	(2) All	(3) Mngr. Partner	(4) Mngr. Partner	(5) Non-M. Partner	(6) Non-M. Partner
Post Creation	-0.005	-0.027	-0.056	0.020	-0.015	-0.010
	(0.026)	(0.020)	(0.046)	(0.056)	(0.020)	(0.029)
Indiv. FE	X	X	X	X	X	X
Week. FE		X		X		X
Indiv.	0.672	0.672	0.681	0.681	0.657	0.657
Observations	101	101	33	33	68	68
N	2706	2706	812	812	1894	1894

Note: Analysis using the presence of all INSS public servants for 2019. Presence is defined as codes for "Present", "Worked overtime", or "Holiday break compensation". Frequencies exclude weekends and federal holidays.

Table 8: Names by highest and Lowest Frequencies

Full Name	$\hat{f}$
Radhakunda Devi Dasi de Maria Moraes Mesiano	1.862e-19
Didio Raul Americo Jimenes Alvarenga Neto	1.969e-19
Hevelin Cristine Aparecida San Juan Baltieri	4.143e-19
Lizie Gabriela Maria Marina Zambolini Vicente	6.088e-19
Hannah Huss Leon Denizard Boudet Portes	6.193e-19
⋮	⋮
José de Souza	97335.745
Antonio da Silva	151241.23
José dos Santos	198027.9
José da Silva	337812.29
Maria da Silva	688837.09

Note: Estimated frequencies of most common names and least common names in INSS exam sample.

Figure 10: example of the official document containing scores for INSS exams

INSTITUTO NACIONAL DO SEGURO SOCIAL (INSS)
CONCURSO PÚBLICO PARA PROVIMENTO DE VAGAS NOS CARGOS DE ANALISTA DO SEGURO SOCIAL E DE TÉCNICO DO SEGURO SOCIAL
EDITAL Nº 7 – INSS, DE 20 DE JUNHO DE 2016

A Presidente substituta do Instituto Nacional do Seguro Social (INSS) torna públicos o **resultado final nas provas objetivas**, a **convocação para a perícia médica dos candidatos que se declararam com deficiência** e a **convocação dos candidatos para a verificação da condição declarada para concorrer às vagas reservadas aos candidatos negros**, referentes ao concurso público para provimento de vagas nos cargos de analista do seguro social e de técnico do seguro social.

1 DO RESULTADO FINAL NAS PROVAS OBJETIVAS

1.1 Resultado final nas provas objetivas, na seguinte ordem: cargo/gerência-executiva, número de inscrição, nome do candidato em ordem alfabética e nota final nas provas objetivas.

1.1.157 CARGO 1: ANALISTA DO SEGURO SOCIAL COM FORMAÇÃO EM SERVIÇO SOCIAL – MANAUS/AM

10331611, Ana Carolina de Abreu Albuquerque, 60.00 / 12583866, Ana Carolina Goncalves Gomes, 44.00 / 12747357, Aurilene Noronha Vieira, 70.00 / 11814047, Azevalda Monteiro Machado, 36.00 / 12264713, Bianca Carvalho Pinto, 48.00 / 13381046, Bianca Ferreira Brandao Farias, 46.00 / 13070416, Brunella Liberato, 38.00 / 11785846, Claudia Farias de Oliveira, 50.00 / 12701708, Claudia Leonor Gomes Barros, 44.00 / 13333716, Daniela Cerdeira Barbosa, 40.00 / 12751982, Danielle Dias de Souza, 49.00 / 12378572, Danielle do Socorro Cavalcante dos Santos, 43.00 / 12248948, Dayene Montenegro Mendonca, 50.00 / 13122986, Dilson Leandro Pinto do Nascimento, 36.00 / 13317738, Dulce Claudia Rodrigues de Melo Araujo, 42.00 / 13113550, Elaine Maria Ribeiro Lopes, 36.00 / 12730696, Elene Lunieres Santiago Pegas, 36.00 / 12641040, Elma de Oliveira Chagas, 59.00 / 12155969, Eneas Pimentel Pinto Filho, 37.00 / 12626529, Erik Oris Martins de Lima, 56.00 / 10271415, Erika Martins Beleza, 44.00 / 11765196, Fabiola Cristina Lopes de Andrade, 38.00 / 12370124, Fabiola Monica Sousa de Oliveira, 50.00 / 12371415, Francisca Santos Ramos, 55.00 / 10462570, Francisco Chagas dos Santos Filho, 41.00 / 13254948, Gil Cesar

Figure 11: Example of INSS branch