Chasing Government Jobs: How Aggregate Labor Supply Responds to Public Sector Hiring Policy in India

Kunal Mangal*

(Job Market Paper)

This Version: April 4, 2021
Latest Version: click here

Abstract

Many countries allocate government jobs through a system of highly competitive exams. For example, in India, civil service exams regularly attract over a half a million applications, with an acceptance rate of less than 0.1%. This paper studies whether the intense competition for these jobs affects aggregate labor supply. To answer this question, I study how the labor market responded to a civil service hiring freeze in the state of Tamil Nadu. I find that candidates responded by spending more time studying, not less. A decade after the hiring freeze was lifted, the cohorts that were most impacted also have lower earnings, suggesting that participation in the exam process did not build human capital. Finally, I provide evidence that structural features of the testing environment—such as how well candidates are able to forecast their own performance, and the underlying returns to study effort—help explain the observed response. Together, these results indicate that public sector hiring policy has the potential to move the whole labor market.

*PhD candidate in Public Policy at Harvard University. Email: kmangal@g.harvard.edu. I am grateful to my advisors Emily Breza, Asim Khwaja, and Rohini Pande for their support. Robert Townsend provided much appreciated initial encouragement. Kranti Mane provided outstanding research assistance. I also thank Augustin Bergeron, Shweta Bhogale, Michael Boozer, Nikita Kohli, Tauhidur Rahman, Sagar Saxena, Utkarsh Saxena, Niharika Singh, Perdie Stilwell, Nikkil Sudharsanan, and seminar participants at Harvard for thoughtful discussions and comments. This work would not have been possible without the support of K. Nanthakumar and S. Nagarajan of the Tamil Nadu Government, the staff at the R&D Section of TNPSC, and Dr. Anand Patil and Mrs. Vaishali Patil in Maharashtra. I am also grateful to the many candidates for government jobs who were willing to take the time to share their world with me. Of course, any errors are my own. All human subjects data collection has been approved by the Harvard University IRB.

1 Introduction

Government employees in developing countries tend to enjoy substantial rents: not only are wages typically higher than what comparable workers would earn in the private sector, but these jobs also come with many valuable and rare amenities, such as lifetime job security.[1]

What costs do these rents impose on the rest of the economy? A key concern is that rents could induce losses orders of magnitude larger than the fiscal costs, due to behavioral responses. Many countries have been particularly sensitive to the possibility that the competition for rents will lead to the selection of less qualified candidates, either due to patronage or bribery, and have responded by implementing rigid systems of civil service exams, in which selection is based on objective, transparent criteria.

Although competitive exams usually succeed in minimizing political interference in the selection process,[2] economists have long been concerned that they do not fully mitigate the costs of rent-seeking behavior. In particular, one worries that the prospect of a lucrative government job encourages individuals to divert time away from productive activity towards unproductive preparation for the selection exam.[3] However, it is unclear whether enough candidates respond in this way to affect the aggregate economy;[4] and it is possible that the effect may even be positive, if studying for the exam builds general human capital.[5] Thus, it is still an open question whether rent-seeking through

[1] Finan, Olken, and Pande (2017) show that public sector wage premia decline with GDP per capita. Wage premia likely understate the ex-post rents that government employees enjoy because of the amenities.

[2] For example, Colonnelli, Prem, and Teso (2020) show that connections generally matter for selection into the Brazilian bureaucracy, but not for positions that are filled via competitive exam.

[3] This exact concern has found mention in the literature from Krueger (1974) to, more recently, Muralidharan (2015) and Banerjee and Duflo (2019). There are other potential costs which I do not address in this paper. For example, another strand of the literature discusses how these rents could starve the private sector of talented individuals, which would in turn affect aggregate productivity and investment (Murphy, Shleifer, and Vishny, 1991; Geromichalos and Kospentaris, 2020).

[4] A competitive exam is a tournament, and tournament theory predicts that only candidates on the margins of selection should be responsive to the prize amount (Lazear and Rosen, 1981).

[5] An increase in general human capital is just one potential social benefit of exam preparation. For example, learning about how government works (which is a commonly found on the syllabus of these exams) might create more engaged citizens, who have a stronger belief in democratic ideals or who

competitive exams imposes meaningful social costs.

In this paper, I provide, to my knowledge, the first empirical evidence on how the competition for rents through the competitive exam system affects the rest of the economy. I address three related questions: First, do rents in the public sector affect individuals' labor supply decisions? Second, are investments in exam preparation productive in the labor market, or are they mostly unproductive signaling costs? And finally, what factors affect how individuals respond to the availability of rents?

To answer these questions, I study the labor market impact of a partial hiring freeze in the state of Tamil Nadu in India. India is a country where rents in public sector employment are particularly large, and where competitive exams are commonplace.[6] In 2001, while staring down a fiscal crisis, the Government of Tamil Nadu suspended hiring for most civil service posts for an indefinite period of time. The hiring freeze was ultimately lifted in 2006. Although civil service hiring fell by 85% during this period, because these jobs constitute a small share of the overall government hiring, the hiring freeze had a negligible impact on aggregate labor demand.[7] Thus, how the labor market equilibrium shifted during the hiring freeze tells us how labor supply responded, which in turn helps us better understand the nature of the competition for rents in the civil service.

My analysis draws on data from nationally-representative household surveys, government reports that I digitized, and newly available application and testing data from the government agency that conducts civil service examinations in Tamil Nadu. I focus on college graduates, who are empirically the demographic group most likely to apply for civil service positions. To identify the impact of the hiring freeze, my main results use a difference-in-differences design that compares: i) Tamil Nadu with the rest of India; and ii) exposed cohorts to unexposed cohorts. For identification, I rely on the fact that the

are better able to advocate for themselves and others. This paper will not be able to speak to those considerations.

[6] In the sample of 32 countries that Finan et al. (2017) include in their cross-country comparison, India has the largest (unadjusted) public sector wage premium, both in absolute terms, and relative to its GDP per capita. Consistent with government employees enjoying rents, surveys of representative samples of Indian youth consistently find that about two-thirds prefer government employment to either private sector jobs or self-employment (see Appendix Figure A.1). Among the rural college-educated youth population, the preference for government jobs stands at over 80% (Kumar, 2019).

[7] See Appendix B for details.

college graduation rates of men remain stable across cohorts. Unfortunately, because the same is not true of women, I restrict the sample to men.[8]

First, I show that aggregate labor supply does in fact respond to the availability of government jobs. Using data from the National Sample Survey, I find that men who were expected to graduate from college during the hiring freeze are 30% more likely to be unemployed in their 20s than men in cohorts whose labor market trajectories were measured before the start of the hiring freeze. The increase in unemployment corresponds to a nearly equal decrease in employment rates.

Why are fresh college graduates more unemployed? The most likely answer is that candidates spent extra time preparing for the competitive exam. During the hiring freeze, the application rate for civil service exams skyrocketed to 10-20 times its normal rate. It is unlikely, then, that college graduates responded to the hiring freeze by seeking employment in the private sector instead.

If college graduates spent more time preparing for the exam, did they build general human capital in the process? My next set of results suggest that the answer is no. If exam preparation builds general human capital, we should expect to see higher labor market earnings in the long-run among cohorts that spent more time preparing. To test this hypothesis, I use data from the Consumer Pyramids Household Survey, which measures labor market earnings about a decade after the hiring freeze ended. I find that, if anything, earnings declined among those cohorts that spent more time in unemployment.

Lastly, I try to understand why candidates responded the way they did. I focus on how the testing environment shapes the incentives for applicants in a way that helps explain their response. There are at least two aspects to the response that we observe that are puzzling. First, it is unclear why candidates were willing to spend more time studying when the probability of obtaining a civil service job declined (at least in the short run). Second, given that the hiring freeze did not have a definite end date, it is unclear why candidates did not take up private sector jobs until the uncertainty was resolved.

To answer the first question, I propose that candidates are generally over-optimistic

[8] Women are well-represented among civil service exam applicants in Tamil Nadu. Between 2012 and 2016, women represented 49% of all applicants in competitive exams for state-level jobs in Tamil Nadu.

about their own probability of selection, and only revise their beliefs downwards through the process of making attempts. I provide suggestive evidence to support this hypothesis. I first show that candidates are generally over-optimistic about their exam performance, drawing on an incentivized prediction task I conducted with 88 civil service aspirants in Maharashtra, a state with a similar civil service examination system as the one in Tamil Nadu. Next, using civil service exam application and testing data from between 2012 and 2016 in Tamil Nadu, I show that candidates respond to prior test scores when deciding whether to make re-application decisions. A key empirical challenge in estimating this relationship is that re-application decisions may be endogenous to ability. I therefore draw from Item Response Theory, a branch of psychometrics, to construct an instrument. The instrument isolate the "luck" component of the test score from variation in ability. Consistent with candidates learning about ability, I show that this luck component predicts re-application decisions. Under some assumptions, the effect of past test scores on re-application decisions is large enough to account for the increase in unemployment that we observe in response to the hiring freeze.

Next, I turn to why candidates may choose not to wait to resume studying until the hiring freeze is over. One reason this might be the case is if the returns to exam preparation are convex in the amount of time spent studying. In that case, candidates who start to prepare early can "out-run" candidates who prepare later, inducing an incentive to start as early as possible. I then use the application and testing data from Tamil Nadu to provide empirical evidence that the returns to additional attempts are in fact convex.

This paper contributes to several distinct strands of the literature. First, it helps us understand why unemployment is high among college graduates in a developing country setting. On average, college graduates are relatively more likely to be unemployed in poorer countries (Feng, Lagakos, and Rauch, 2018), but why this is so is not well understood. Previous literature has largely focused on frictions within the private sector labor market (Abebe, Caria, Fafchamps, Falco, Franklin, and Quinn, 2018; Banerjee and Chiplunkar, 2018). In this paper, I provide evidence for an alternative mechanism that explains why: the unemployed are searching for government jobs.

This paper also has implications for understanding optimal public sector hiring policy. Motivated by a focus on improving service delivery, much of the existing literature has focused on the effects of these policies on the set of people that are ultimately selected (Dal Bó, Finan, and Rossi, 2013; Ashraf, Bandiera, and Jack, 2014; Ashraf, Bandiera, Davenport, and Lee, 2020). By contrast, this paper redirects focus towards the vast majority of candidates who apply but are not selected. In a context where this population is large—such as in India—the effect on this latter population appears to be large enough that is worth considering this population explicitly when designing hiring policy.

More broadly, this paper helps us understand how workers respond to demand shocks within highly desirable and salient sectors of the economy. When these shocks occur, incumbent workers face a choice between doubling down, or cutting their losses. In the United States, evidence from the manufacturing sector (a desirable and salient sector for less-educated men) suggests that men tend to double down (Autor, Dorn, Hanson, and Song, 2014). In this paper, I provide evidence from a different context for a similar pattern of responses.

These results suggest that public sector hiring policy in India has the potential to affect the entire labor market. This represents both an opportunity and a challenge: hiring policy is a relatively unexplored policy lever for combating unemployment in this context, but that also means that the chance that hiring policy decisions have unintended consequences in the economy are also relatively high.

This paper proceeds as follows. Section 2 describes the competitive exam system in India and provides details about the hiring freeze policy. Section 3 presents evidence on the short-run labor supply impacts of the hiring freeze. Section 4 presents evidence on the long-run impact of the hiring freeze on earnings. Section 5 discusses how the testing environment influences candidates' response to the hiring freeze. Section 6 concludes.

2 Setting

2.1 The Competitive Examination System

In India, most administrative positions—such as clerk, typist, and section officer—are filled through a system of competitive exams.[9] All competitive exams include a multiple choice test. For more skilled positions, the exam may also include an essay component and/or an oral interview. The exam typically covers a wide range of academic subjects, including history, geography, mathematics and logical reasoning, languages, and science. The government conducts a single set of exams for batches of vacancies with similar job descriptions and required qualifications. After the results are tabulated, candidates then choose their preferred posting according to their exam rank.[10]

Government jobs advertised through competitive exams have eligibility requirements. In Tamil Nadu, all posts require candidates to be at least 18 years of age and have a minimum of a 10th standard education. Unlike other states, Tamil Nadu does not have upper age limits for most applicants, and candidates can make an unlimited number of attempts. In addition to 10th standard, some posts require require college degrees and/or degrees in specific fields. For recruitments completed between 1995 and 2010, 43% of posts and 25% of vacancies required a college degree.

These exams are heavily over-subscribed. Table 1 highlights a typical example from Tamil Nadu for a recruitment advertised in 1999, a few years before the hiring freeze was implemented. In this case, the Tamil Nadu government notified 310 vacancies through its Group 4 examination, which recruits for the most junior category of clerical workers. It received 405,927 applications. Relative to the entire eligible population ages 18-40, this corresponds to an application rate of about 5.6%. Because the application rate for state-level government jobs is so high, it is plausible that changes in candidate behavior could be reflected in aggregate labor market outcomes.

[9] The government also conducts exams for specialized positions, such as surgeons, scientists, statisticians, and university lecturers.

[10] In general, the exam process has enough integrity (especially in Tamil Nadu) that cheating is rare. In cases where cheating is detected it is usually punished severely. For example, in Tamil Nadu 99 candidates were caught in a cheating scandal in January 2020 and were subsequently banned from applying for government jobs for life (Rajan, 2020).

2.2 The Hiring Freeze

In November 2001, the Government of Tamil Nadu publicly announced that it would suspend recruitment for "non-essential" posts for an indefinite period of time. Doctors, police, and teachers were explicitly exempted from the hiring freeze. This meant that the freeze applied mostly to administrative posts. In case a department wanted to make an exception to the hiring freeze, it had to submit a proposal to a panel of senior bureaucrats for approval.[11] This policy was ultimately rescinded in July 2006.[12]

According to the World Bank, the proximate cause of the hiring freeze appears to be a fiscal crisis, triggered by a set of pay raises that the Government implemented in the late 1990s (Bank, 2004). Although other states experienced fiscal crises around the same time, to the best of my knowledge they did not implement a hiring freeze.[13] I therefore use the set of states excluding Tamil Nadu as a control group in the empirical analysis. I test the sensitivity of the results to the choice of states included in the control group. To the extent that other states also implemented hiring freezes at the same time, I expect the estimated effects to be attenuated.

At the time of the hiring freeze, there were three government agencies in Tamil Nadu responsible for recruitment: the Tamil Nadu Public Service Commission (which recruited both administrative and medical posts); the Tamil Nadu Uniformed Services Board (which recruited police); and the Tamil Nadu Teacher Recruitment Board (which recruited primary and secondary teachers).[14] Because the hiring freeze exempted teachers, doctors, and police, the effect of hiring freeze thus fell entirely on recruitments conducted by the Tamil Nadu Public Service Commission (TNPSC, hereafter). In Appendix Figure A.2, I present evidence that recruitment in the exempt positions continued as usual.

Over the course of the five years of the hiring freeze, the Government made few

[11] Specifically, proposals were vetted by a committee consisting of the Chief Secretary, the Finance Secretary, and the Secretary (Personnel and Administrative Reforms).

[12] The hiring freeze was announced in Tamil Nadu Government Order 212/2001. The freeze was lifted in Government Order 91/2006.

[13] To make this determination precisely, I would need to collect information from each of the state governments. These requests for information are often denied on the grounds that they would require too much time of the department's staff.

[14] In 2012, the Tamil Nadu Government established the Tamil Nadu Medical Recruitment Board, which took over responsibilities of recruiting medical staff from TNPSC.

exceptions. There were only 15 exams conducted during the entire course of the hiring freeze at TNPSC (of which 6 were for medical personnel), as opposed to an average of 28 *per year* when TNPSC was fully functional. As a result, as we see in Figure 1, the number of available vacancies advertised by TNPSC fell by approximately 85% during the hiring freeze. After the hiring freeze was lifted, the number of vacancies notified returned to roughly the same level it was at before the hiring freeze was announced.

The number of vacancies that were abolished due to the freeze was small relative to the overall size of the labor force. A back-of-the-envelope calculation suggests that the hiring freeze caused the most exposed cohorts of male college graduates to lose about 600 fewer vacancies over five years. Meanwhile, these same cohorts have a population of about 100,000. So even if the hiring freeze caused a one-to-one loss in employment (which is dubious, since family business is common), *at most* only about 0.6% the cohort's employment should be affected. Even accounting for the large wage premium, the drop in average earnings due to the aggregate demand shock is on the order of 0.4% of cohort-average earnings. (See Appendix B for the details of these calculations). I therefore treat the direct demand effect of the hiring freeze (i.e. the reduction in labor demand due to less government hiring) as negligible, and ascribe any observable shifts in labor market equilibrium to an endogenous supply response.

3 Short-Run Impacts of the Hiring Freeze

In this section, I assess whether and by how much the hiring freeze affected aggregate labor market outcomes both during and in the years immediately following the hiring freeze.

3.1 Changes in Labor Supply

3.1.1 Data

For this analysis, I use data from the National Sample Survey (NSS), a nationally representative household survey conducted by the Government of India. I combine all rounds of the NSS conducted between 1994 and 2010 that included a module on employment. This includes two rounds conducted before the hiring freeze; three rounds conducted during the hiring freeze; and two rounds conducted after the end of the hiring freeze.[15] By stacking these individual rounds, I obtain a data set of repeated cross-sections.

My key outcome variable is employment status. I consider three categories: employed, unemployed, and out of labor force. These variables are constructed using the NSS's Usual Principal Status definition. Household members' Usual Principal Status is the activity in which they spent the majority of their time over the year prior to the date of the survey. In accordance with the NSS definition, I consider individuals to be employed if their principal status included any form of own-account work, salaried work, or casual labor. Individuals are marked as unemployed if they were "available" for work but not working.[16] Relevant to this setting, individuals who are enrolled in school are considered unemployed if they would consider leaving in order to take up an available job opportunity (NSS Handbook). This means that individuals who continue to collect degrees while they prepare for government exams—as documented in Jeffrey (2010)—would be marked as unemployed. Being out of the labor force is the residual category among those who are neither employed nor unemployed.

Unless otherwise noted, I adjust all estimates according to the sampling weights provided with the data. I normalize weights so that observations have equal weight across rounds relative to each other.[17]

3.1.2 Empirical Strategy

The key empirical challenge is to estimate how labor market outcomes would evolve in Tamil Nadu in the absence of the hiring freeze. To construct this counterfactual, I use a difference-in-differences (DD) design that compares Tamil Nadu with the rest of India,

[15] Specifically, I use data from the 50th, 55th, 60th, 61st, 62nd, 64th, and 66th rounds. These surveys were conducted during the following months, respectively: July 1993 - June 1994; July 1999 - June 2000; January 2004 - June 2004; July 2004 - June 2005; July 2005 - June 2006; July 2007 - June 2008; and July 2009 - June 2010. For simplicity, I refer to each round by the year in which it was completed.

[16] Note that this definition does not include explicit criteria for active search.

[17] That is: if w_{ir} are NSS-provided weights for individual i in round r, and there are N_r observations in round r, then the weights I use are: $N_r * w_{ir} / \sum_r w_{ir}$.

and compares more affected cohorts with less affected cohorts.[18]

Who is likely to be affected by the hiring freeze? The hiring freeze policy will likely only affect a specific segment of Tamil Nadu's labor market. In general, Indian states require that candidates who appear for competitive examinations have at least a 10th standard education. In the year 2000, the year before the hiring freeze was first implemented, this requirement excluded about 70% of the population between the ages of 18 to 40 in Tamil Nadu. Moreover, as we will see, application rates are very heterogeneous within the eligible population. For these reasons, even though the *total number* of applicants is large, the *share* of the overall population of Tamil Nadu that would have been actively making application decisions during the hiring freeze is likely to be relatively small. The NSS unfortunately does not provide me with enough statistical power to measure the impact at an aggregate level. I therefore need to zoom in on the segment of the population that is most likely to consider applying during the hiring freeze.

To estimate how application rates vary across demographic groups, I use administrative data from the Tamil Nadu Public Service Commission for exams conducted between 2012 and 2014, and Census data from 2011. I estimate the application rate by dividing counts of the average number of applications received by age by the population estimate from the Census. The results of this calculation are presented in Figure 2. Note that application rates vary widely by age and education. Application rates are highest among college graduates around age 21, which is the year right after a typical student completes an undergraduate degree.[19]

Based on the observed variation in application rates, we should expect the largest effect for cohorts that turned 21 during the hiring freeze. That is because this group was most likely to make application decisions under usual conditions. This is my primary "treatment" group of interest. It is possible that cohorts that were older than 21 at

[18]Throughout the analysis, I include observations from Puducherry in Tamil Nadu. Puducherry is a small federally-administrated enclave entirely surrounded by Tamil Nadu, which shares the same language as Tamil Nadu, and which does not have a Public Service Commission of its own. Residents of Puducherry commonly apply for positions through the Tamil Nadu Public Service Commission.

[19]A typical undergraduate degree starts at age 18 and lasts 3 years, which makes a typical fresh graduate 21 years old.

the time the hiring freeze was announced were affected. However, we would also expect smaller effects sizes for this group relative to the group that graduated from college, since many individuals from the former group would have exited exam preparation already.

Sample Restrictions. I restrict the sample in three ways: 1) I restrict the sample to men. This is because, as we will see in Section 3.1.3, college graduation rates for women shift after the hiring freeze, which makes it difficult to disentangle the impact of the hiring freeze from violations of the parallel trends assumption. 2) I restrict the sample to individuals between the ages of 21 to 27 at the time the survey was completed, thereby focusing on the sample that is most likely to apply for government jobs.[20] 3) I further restrict the sample to cohorts who were between the ages of 17 to 30 in the year 2001. The lower bound corresponds to the youngest individuals who are expected to have graduated from college before the end of the hiring freeze.[21]

Regression Specification. Figure 3 summarizes the variation that I use. In each survey year, I plot the cohorts that are included in the sample after implementing the restrictions described in the preceding paragraph. I define cohorts by their age in 2001, the year in which the hiring freeze was announced.[22] The comparison group includes all individuals whose outcomes were measured before the start of the hiring freeze. The treatment group includes all observations measured after the implementation of the hiring freeze belonging to individuals who were expected to complete college before the end of the hiring freeze. Throughout, I use age 21 as the expected age of college graduation. The treatment group therefore includes seven cohorts, i.e. those between the ages of 17 to 24 in 2001.[23] I divide the treatment group into two groups: 1) those who are expected to have graduated from college during the hiring freeze (i.e. age 17-21 in 2001); and 2) those who are expected to

[20]For most rounds, the NSS was conducted over the course of a year. Assuming that birthdays are roughly uniformly distributed, this means that about half of the sample will have aged another year during the course of the survey. I can break the tie either way. I choose to break it by adding one to the reported age for each individual in the sample.

[21]There is no conceptual reason to include the upper bound. Its primary purpose is to exclude the cohort that was age 31 in 2001, which is a severe outlier relative to all the other cohorts in the sample (see Appendix Figure A.3). All the results hold if the cohorts older than 31 are included in the sample as well.

[22]Specifically, I compute [Age in 2001] = [Age] + (2001 − [NSS Round Completion Year]).

[23]Given the sample restrictions and the timing of the NSS rounds, cohorts that were older than 24 of age in 2001 were only surveyed before the hiring freeze was announced.

have already graduated from college before the hiring freeze (i.e. age 22 to 24 in 2001). My empirical strategy compares each of these groups to the comparison group in Tamil Nadu and to its counterpart in the rest of India.

I implement these comparisons using the following regression specification:

$$y_i = \beta_1[TN_{s(i)} \times During_{c(i)} \times Freeze_{t(i)}] + \beta_2[TN_{s(i)} \times Before_{c(i)} \times Freeze_{t(i)}] \\ + \zeta TN_{s(i)} + \gamma_{c(i)} + \delta Freeze_{t(i)} + \Gamma' X_i + \epsilon_i \quad (1)$$

Because the data consists of repeated cross-sections, each observation is a unique individual. Cohorts $c(i)$ are indexed according to their age in 2001. $TN_{s(i)}$ is an indicator for whether state s is Tamil Nadu. $During_{c(i)}$ and $Before_{c(i)}$ are indicators for whether cohorts were expected to graduate either *during* or *before* the hiring freeze, respectively. That is, $During_{c(i)} = \mathbf{1}[17 \leq c(i) \leq 21]$ and $Before_{c(i)} = \mathbf{1}[c(i) \geq 22]$. $Freeze_{t(i)}$ is an indicator for whether the individual was surveyed in a year $t(i)$ after the hiring freeze was implemented, i.e. $Freeze_{t(i)} = \mathbf{1}[t(i) \geq 2001]$. Finally, the vector X_i includes a set of control variables, including: 1) dummy variables for the individual's age at the time of the survey, interacted with $TN_{s(i)}$; and 2) caste and religion dummies.[24]

The primary coefficients of interest are β_1 and β_2. These parameters identify the impact of the hiring freeze under a parallel trends assumption. Before the hiring freeze was announced, Tamil Nadu and the rest of India had similar average rates of unemployment and employment within the analysis sample (see Appendix Table A.3). The parallel trends assumption requires that Tamil Nadu and the rest of India would continue to have similar average outcomes in this sample across time if not for the hiring freeze.

To assess the validity of the parallel trends assumption it is standard practice to compare trends before the implementation of the policy change. Unfortunately, the paucity of data before the hiring freeze does not allow me to estimate pre-trends with enough precision for this test to be informative.[25] Instead, I implement an alternative over-

[24]Both caste and religion are coded in groups of three. Caste is either ST, SC, or Other. Religion is either Hindu, Muslim, or Other.

[25]The sample sizes in state x cohort cells are often less than a hundred observations, especially for older cohorts. See Appendix Table A.1.

identification test made available by the institutional context. Recall, individuals with less than a 10th standard education are not eligible to apply for government jobs through competitive exams (henceforth, I refer to this group as the ineligible sample). Therefore, if the rest of India serves as a valid counterfactual, we should expect $\beta_1 = \beta_2 = 0$ when the specification in equation (1) is run on the ineligible sample.[26] As with the pre-trends test, this test is neither necessary nor sufficient for valid identification in the college-educated sample. However, because employment status tends to be correlated between the two samples across years and states (see Appendix Figure A.4), it is plausible that shocks to employment status are common across both samples, and hence this test should be informative.

I also explicitly compare the coefficients from the college sample with the coefficients from the ineligible sample using a triple difference design. The full estimating equation for this specification is:

$$\begin{aligned} y_i = College_i \times \Big[& \beta_1 [TN_{s(i)} \times During_{c(i)} \times Freeze_{t(i)}] + \beta_2 [TN_{s(i)} \times Before_{c(i)} \times Freeze_{t(i)}] \\ & + \gamma_{c(i),1} + \delta_1 Freeze_{t(i)} + \Gamma_1' X_i + \alpha \Big] \\ + \Big[& \eta_1 [TN_{s(i)} \times During_{c(i)} \times Freeze_{t(i)}] + \eta_2 [TN_{s(i)} \times Before_{c(i)} \times Freeze_{t(i)}] \\ & + \gamma_{c(i),0} + \delta_0 Freeze_{t(i)} + \Gamma_0' X_i \Big] + \epsilon_i \end{aligned} \tag{2}$$

Across both specifications, I cluster standard errors at the state-by-cohort level.[27] In doing so, I treat clustering as a design correction that accounts for the fact that the treatment (i.e. exposure to the hiring freeze) varied across cohorts within Tamil Nadu (Abadie, Athey, Imbens, and Wooldridge, 2017). Since cohorts are tracked across multiple survey rounds, state-by-cohort clusters will also capture serial correlation in error terms across years.[28] Although the total number of clusters is large, traditional

[26] This assumes that general equilibrium effects on the ineligible sample are negligible.

[27] Several states split during this time period. I ignore these splits when assigning observations to states, maintaining consistent state definitions across the 8 rounds of the NSS.

[28] This approach is standard in the literature on the effects of graduating during a recession, which also features shocks that vary in intensity across states and cohorts (Kahn, 2010; Oreopoulos, Von Wachter, and Heisz, 2012; Schwandt and Von Wachter, 2019).

clustered standard errors are still too small because the number of clusters corresponding to the coefficients of interest is also small (Donald and Lang, 2007; MacKinnon and Webb, 2018). I therefore report confidence intervals using the wild bootstrap procedure outlined in Cameron, Gelbach, and Miller (2008). My own simulations indicate that these confidence intervals are likely to have nearly the correct coverage rate in this setting.[29]

The validity of restricting to the analysis to a sample of college graduates depends on whether college graduation rates moved in parallel in Tamil Nadu and the rest of India. In Appendix Table A.4, I assess whether this is the case. For men, I observe no statistically significant changes in college completion after the hiring freeze. By contrast, I see a large increase for women. It is unclear whether this shift reflects a violation of the parallel trends assumption or is an endogenous outcome of the hiring freeze.[30] To simplify the analysis, I therefore restrict the analysis to men.

3.1.3 Results

I begin by presenting the DD results for each treatment cohort using unadjusted cell means. Although these estimates are imprecise, they allow us to more transparently assess the underlying variation that informs the estimates from the parametric specifications. To compute these estimates, I first compute unweighted averages of unemployment and employment by state x cohort x year cells. For each treatment cohort, I compute a simple DD estimate by subtracting the Tamil Nadu mean from the simple average of the remaining states, and then comparing that difference with the comparison group.[31]

The results of this exercise are presented in Figure 4. Note that unemployment is consistently higher (and employment consistently lower) among all cohorts that were

[29] I construct the simulation as follows. For each of 500 iterations, I replace the dependent variable with a set of random 0/1 draws that are i.i.d. across observations. I then tabulate the fraction of confidence intervals that include zero. The results are reported in Appendix Table A.2.

[30] One reason why we might expect to see a parallel trends violation for women in particular is that this period coincided with a large expansion in the set of available respectable work opportunities (especially business process outsourcing work), which both affected educational attainment and were not uniformly available across Indian states (Jensen, 2012).

[31] In more precise terms: Let s index states, c index cohorts, $t \in \{0, 1\}$ be an indicator for whether outcomes were measured after the hiring freeze, and TN_s be an indicator for Tamil Nadu. Then for each outcome y, I present estimates of: $\big(E\big[y\,|\,c, t = 1, TN_s = 1\big] - E\big[y\,|\,t = 0, TN_s = 1\big]\big) - \big(E\big[y\,|\,c, t = 1, TN_s = 0\big] - E\big[y\,|\,t = 0, TN_s = 0\big]\big)$.

expected to graduate from college during the hiring freeze. Meanwhile, among cohorts that we expect to have already graduated, the estimates tend to hover around zero. The consistency of the estimates across these two groups of cohorts suggets that it is unlikely that estimates of β_1 and β_2 in equation (1) are driven by individual cohorts.

Panel A of Table 2 summarizes the estimates for equation (1). The coefficients in Column (1) indicate that, after the implementation of the hiring freeze, unemployment among cohorts that were expected to graduate from college during the hiring freeze increased by a statistically significant 6.2 percentage points relative to the rest of India (95% CI: 0.4 - 11.9 p.p.). This corresponds to a $6.2/20.7 = 30\%$ increase in the likelihood of unemployment. Meanwhile, we observe opposite-signed and statistically insignificant effects on cohorts that were already expected to have graduated. These coefficients capture the *average* shift in cohorts' labor market trajectory between the ages of 21 and 27. The increase in unemployment could therefore reflect both an intensive margin effect (i.e. individuals spending more time in unemployment) and an extensive margin effect (i.e. more people ever experiencing unemployment). In Columns (2) and (3) we see that the increase in unemployment is almost entirely accounted for by a decline in employment. Changes in labor force participation are negligible.

Panel B re-estimates equation (1) on the ineligible sample. For all three outcome variables, the coefficients are small and statistically insignificant. The null effect indicates that, in the ineligible sample, men tended to follow the same early career trajectories during the hiring freeze as their predecessors did prior to the hiring freeze. This provides some reassurance that the parallel trends assumption is reasonable for men in this context.

Finally, in Panel C, I present estimates from the triple difference specification in equation (2), which effectively estimates the difference between the coefficients of Panels A and B. Since the coefficients in Panel B are all close to zero, the point estimates in Panel C are very similar to those in Panel A. The confidence intervals are also nearly identical.

3.1.4 Robustness

I probe the robustness of these results in three ways:

Choice of comparison group. I test whether the results in Table 2 are sensitive to the choice of states to include in the comparison group. I try several variations. In Appendix A.5 I use only the states that neighbor Tamil Nadu in the comparison group (namely Karnataka, Kerala, and undivided Andhra Pradesh). In Appendix Table A.6, I only use other large states, which I define as those with at least 500 observations per state in the sample of male college graduates.[32] In both cases, the point estimates of β_1 remain very similar. As we would expect, the confidence intervals are tighter when I use more comparison states.

This lack of sensitivity to the choice of comparison states generalizes: I find that on average I obtain the same estimate of β_1 when I use a *random* subset of states in the comparison group. That is, if I randomly sample 10 states from the set of comparison states and re-estimate equation (1), the mean of this distribution nearly coincides with the estimates of β_1 reported in Table 2 (see Appendix Figure A.5). This is exactly what we would expect if states experience common shocks across time and state-specific trends are largely absent in this context.

Specification. I probe robustness to dropping the caste and religion controls. In case the types of individual completing college responded to the hiring freeze policy, these controls may no longer be exogenous. These results are presented in Appendix Table A.7. The estimates remain similar.

Alternative interpretations. So far, I have interpreted the estimates in Table 2 as reflecting shifts in labor supply. Here, I consider alternative interpretations of these coefficients.

In particular, one might be concerned about the direct effects of the conditions that precipitated the hiring freeze in the first place. As discussed in Section 2, the Tamil Nadu

[32] The states included in this sample are: undivided Andhra Pradesh, Bihar, Gujarat, Karnataka, Kerala, undivided Madhya Pradesh, Maharashtra, Odisha, Punjab, Rajasthan, Uttar Pradesh (including Uttarakhand), and West Bengal.

government appears to have implemented the hiring freeze because it faced a fiscal crisis. In 2001, the same year as the implementation of the hiring freeze, Tamil Nadu experienced a drop in GDP growth relative to the rest of the country (see Appendix Figure A.6). This fact raises the possibility that the increase in unemployment is a result of the more well-understood cost of graduating during a recession (Kahn, 2010; Oreopoulos et al., 2012; Schwandt and Von Wachter, 2019). Furthermore, the labor market may be affected by contemporaneous changes in service delivery or budget re-allocations.

The triple difference specification addresses these concerns to some extent: in general, it is hard to conceive of a mechanism in which macroeconomic shocks only affect the cohorts of college graduates most likely to apply for government jobs during the hiring freeze. Still, it is possible that demand shocks for less educated individuals are not reflected in employment status, since their labor supply tends to be less elastic (Jayachandran, 2006).

To aid in distinguishing between demand- and supply-based interpretations of the data, I study the impacts on earnings. Consider a simple supply and demand model of the aggregate labor market, in which both curves have finite elasticity. If the increase in unemployment reflects a reduction an aggregate labor supply, then we would expect to observe an *increase* in average wages among the remaining participants in the labor market. Conversely, if the increase in unemployment reflects a drop in aggregate labor demand, then see a *decrease* in wages.

To assess how wages responded to the hiring freeze, I use earnings data in the NSS. Household members report the number of days employed in the week prior to the survey, and their earnings in each day. I compute average wages by dividing weekly earnings by the number of days worked in the week. I use the same specification from the main analysis (i.e. equation (1)), with the sample restricted to male college graduates who reported any days of employment. Appendix Table A.8 summarizes these results. Members of the cohorts age 17 to 21 in 2001 who chose to stay in the labor market had higher earnings. This evidence, combined with the evidence on employment status, is consistent with aggregate labor supply falling after the implementation of the hiring freeze.

3.2 Linking Unemployment to Exam Preparation

After the implementation of the hiring freeze, the cohorts that were most likely to be affected spent more time unemployed. Why is this the case? In this section, I present evidence that the most likely account is that they spent more time preparing full-time for the exam. Unfortunately, in India there are no existing datasets that directly measure exam preparation during this time period. However, if exam preparation did increase, then we should observe an increase in the application rate during the hiring freeze.[33] Recall that not all recruitments were frozen during the hiring freeze. I can therefore test whether recruitments conducted during the hiring freeze received more or less applications than similar recruitments conducted before the hiring freeze.

3.2.1 Data

I digitized all the annual reports of the Tamil Nadu Public Service Commission that were published between 1995 and 2010. These reports provide statistics for all recruitments *completed* during the report year.[34]

In the report, vacancies are classified into "state" and "subordinate" positions. The former include the highest level positions for which TNPSC conducts examinations. It turns out that the only state-level recruitments conducted by TNPSC during the hiring freeze were for specialized legal and medical positions (such as judge, surgeon, and veterinarian), which were exempt under the hiring freeze (and thus these applicants should be unaffected). Therefore I focus the analysis on the sample of subordinate positions, which reflects 57% of posts and 75% of vacancies in this period.

3.2.2 Empirical Strategy

I compare recruitments conducted during the hiring freeze against those with similar number of vacancies before the hiring freeze. The regression I estimate takes the following

[33] Of course, it is possible that candidates appear for the exam without preparing. But, as we will see, the fact that tougher exams receive fewer applications suggests that candidates tend to consider their preparedness when deciding whether to apply.

[34] On average, it takes 475 days between the date of last application and the date when the result is announced. The maximum observed in the sample is 2998 days.

form:

$$\log y_i = \alpha + \beta_1 freeze_{t(i)} + \beta_2 after_{t(i)} + \gamma \log(vacancies)_i + \epsilon_i \tag{3}$$

where i indexes recruitments, and $t(i)$ measures the year in which recruitment i was notified. The outcome of interest is application intensity, for which I observe two distinct measures: 1) the number of applications received; and 2) the number of candidates who appeared for the exam. The variable $freeze_{t(i)}$ is a dummy for whether the last date to apply occurred while the hiring freeze was still in effect, and $after_{t(i)}$ is a dummy for whether the last date to apply occurred after the freeze was lifted. The variable $vacancies_i$ tracks the number of vacancies advertised in the recruitment. Recruitments with fewer vacancies are typically those for more senior positions, which involve more difficult exams. Controlling for the number of advertised vacancies therefore proxies for a range of different features of the advertised position.

The coefficients β_1 and β_2 identify the impact of the hiring freeze under the assumption that recruitments conducted before the hiring freeze are valid counterfactuals. To assess the validity of this assumption, I explicitly check for trends by estimating the following specification:

$$\log y_i = \alpha_{t(i)} + \gamma \log(vacancies)_i + \epsilon_i \tag{4}$$

The parameter of interest in this specification is $\alpha_{t(i)}$. In the absence of meaningful pre-trends, we should expect to see roughly constant values of of $\alpha_{t(i)}$ before the hiring freeze, and a sharp change in $\alpha_{t(i)}$ during the hiring freeze.

3.2.3 Results

Figure 5a provides an approximate visual illustration of the regression in equation (3), using applications received as an outcome variable. There were four recruitments conducted during the hiring freeze. Those recruitments are labeled with the year in which the recruitment was conducted. Relative to the number of advertised vacancies, the number of applications received is much higher than usual. Table 3 summarizes estimates of the magnitude of this difference. The estimates indicate that applications increased by 301 log points during the hiring freeze; that is equivalent to a 20-fold increase in the usual application rate. The effect on the number of candidates that actually appeared for the exam is even larger.

The increase in the application rate during the hiring freeze is not part of a long-run trend. After the hiring freeze, the vacancy-application curve falls, but it does not entirely return to the same level it was at before the hiring freeze. The point estimate (β_2) suggests that the number of candidates appearing for exams is still 173 log points higher than the period before the hiring freeze. Still, this is meaningfully smaller than the application level observed during the hiring freeze: a test of the equality of the β_1 and β_2 coefficients rejects at the 1% level.

In Figure 5b, I plot year-by-year estimates of the change in the hiring freeze using equation (4), which confirms that the increase in the application rate during the hiring freeze is not continuous with pre-trends.

4 Long-run Effects of the Hiring Freeze

In this section, I assess whether the hiring freeze had an impact on the earnings of cohorts between 2014-2019, about a decade after the end of the hiring freeze.

4.1 Data

I use the Centre for Monitoring the Indian Economy's Consumer Pyramids Household Survey (CMIE-CPHS). This survey follows a nationally-representative panel of approximately 160,000 households every four months, starting in January 2014. I use all waves of data collected between January 2014 and December 2019.

In each wave, households report earnings for the *previous* four months.[35] My primary outcome is labor market earnings, which includes, wages, overtime, bonuses, and income from self-employment.[36] The CMIE-CPHS reports nominal income figures. I deflate all

[35]In the first wave (January to April 2014), respondents reported income for months that extended into 2013. I drop observations that reflect income from 2013.

[36]The CMIE-CPHS documentation provides the following details on how this variable is constructed:

> This is the income earned by each working member of the household in the form of wages.

reported income values to their 2014 values using annual inflation rates reported by the World Bank.

As in Section 3, I identify cohorts by their age in 2001.[37] I correct for measurement error in the observed age using an imputation procedure detailed in Appendix C. I weight all estimates using the sampling weights provided by CMIE.

4.2 Empirical Strategy

I adapt the cohort-based approach from Section 3 to study the impact on long run outcomes. As in Section 3, I restrict the main analysis to male college graduates between the ages of 17 to 30 in 2001.[38] The main difference is that I do not observe outcomes measured before the hiring freeze. To accommodate, I will treat the cohorts age 27 to 30 in 2001 as the comparison group, assuming they are relatively unaffected by the hiring freeze. This is consistent with the evidence from Section 3 that older cohorts appear to be unaffected by the hiring freeze.

I estimate the following difference-in-differences specification:[39]

$$y_{it} = \beta_1 \big[TN_{s(it)} \times \mathbf{1}(17 \leq c_{it} \leq 21)\big] + \beta_2 \big[TN_{s(it)} \times \mathbf{1}(22 \leq c_{it} \leq 26)\big] + \zeta TN_{s(it)} + \gamma_{c(it)} + \alpha_t + \epsilon_{it} \quad (5)$$

This is the salary earned at the end of a month by the salaried people in India. If a businessman takes a salary from the business, it is included here as wages. A salaried person may earn a salary from his employers and may also work as a home-based worker (for example, by giving tuitions). In such cases, the income earned from home-based work is also added into wages. Wages can be paid at the end of a month, a week, a fortnight or any other frequency. All of these are added into a monthly salary appropriately during the capture of data. Wages includes over-time payments received. Wages also includes bribes. If an employee is given a part of his/her income in the form of food or other goods, the value of these is included in wages with a corresponding entry in the expenses of the respective item head. If rent expenses are reimbursed by the employer, then it is included in wages. A corresponding entry in expenses is be made only if such an expense is made.

[37] Specifically, I compute [Age in 2001] = $floor$([Age]−[Months between Survey Date and Jan 2001]/12).

[38] Education is measured independently in each wave of the survey. Due to measurement error, the measured education of individuals sometimes fluctuates. I assign individuals the maximum modal observed education level.

[39] Compared to equation (1), this specification drops the age-specific dummies. That is because these coefficients would largely be collinear with the cohort effects, both because the CMIE-CPHS is a panel and because there are no outcomes measured before the implementation of the hiring freeze.

where y_{it} is the outcome for individual i measured in month t. Note that while all the variation in exposure to the hiring freeze is across individuals, the panel structure allows us to measure individual outcomes more precisely. Cohorts are indexed according to their age in 2001. The group of cohorts $\mathbf{1}(17 \leq c_i \leq 21)$ is intended to capture individuals expected to graduate from college during the hiring freeze; the group $\mathbf{1}(22 \leq c_i \leq 26)$ captures the individuals expected to graduate from college before the hiring freeze.

To assess the parallel trends assumption, I run the same specification on the ineligible sample, i.e. men with less than a 10th standard education. I also use this sample as part of a triple difference specification that compares the difference-in-differences coefficients between the college and ineligible samples:

$$\begin{aligned} y_{it} = College_i \times \Big[&\beta_1 \big[TN_{s(it)} \times \mathbf{1}(17 \leq c_{it} \leq 21)\big] + \beta_2 \big[TN_{s(it)} \times \mathbf{1}(22 \leq c_{it} \leq 26)\big] \\ &+ \zeta_1 TN_{s(it)} + \gamma_{c(it),1} + \alpha_{t,1} \Big] \\ + \Big[&\eta_1 \big[TN_{s(it)} \times \mathbf{1}(17 \leq c_{it} \leq 21)\big] + \eta_2 \big[TN_{s(it)} \times \mathbf{1}(22 \leq c_{it} \leq 26)\big] \\ &+ \zeta_0 TN_{s(it)} + \gamma_{c(it),0} + \alpha_{t,0} \Big] + \epsilon_{it} \qquad (6) \end{aligned}$$

As before, for both specifications I cluster errors at the state x cohort level and report 95% confidence intervals using the wild bootstrap.

4.3 Results

The same cohorts of men who spend more time in unemployment in the short run appear to have lower labor market earnings in the long run. This result is summarized in Table 4. I am unable to detect effects on average earnings with any precision (Column (1)). However, if I find suggestive evidence that affected men are less likely to appear in the top of the earnings distribution. I find suggestive evidence that men who were 17 to 21 in 2001 are 6 percentage points more likely to earn less than Rs. 20,000 per month in 2014 INR (95% CI: -0.004 - 0.129). Note that Rs. 20,000 corresponds to the 75th percentile of the earnings distribution in the rest of India for this cohort.

Reassuringly, we do not see any of these same impacts for the ineligible sample (see Panel B), or for cohorts who were already expected to have graduated from college before the hiring freeze. The triple difference estimates suggest that the effects on the college sample are not driven by common shocks.

Taken seriously, the estimates in Column (2) imply a large drop in earnings. Recall, the evidence in Section 3 suggests that about 6% of the population that graduated during the hiring freeze spent an average of an extra year in unemployment. Assuming no impact on the remaining population, these estimates imply that a single additional year of studying caused all individuals studying for the exam to drop below the Rs. 20K threshold. It is unlikely that we can account for this effect just based on the returns to labor market experience, when Mincer regressions suggest that those returns are on the order of 3% per year. One possible explanation is that affected men remain less attached to the workforce and more reliant on family for financial support. Jeffrey (2010) documents how men who prepare for government jobs tend to delay household formation, which could increase financial dependence in the long-run. The evidence in Columns (3) and (4) hints at that story: Although there is no indication of a shift in household income (Column (4)), affected mens' earnings are a smaller share of total household income (Column (3)). Further work is needed to assess this hypothesis more robustly.

5 Mechanisms

The results so far suggest that college graduates increased the amount of time they spent preparing for the exam in response to the hiring freeze. There are two aspects of this response that are puzzling.

First, why would candidates be willing to study *longer* because of a hiring freeze? As we will see in this section, during normal testing years, almost all candidates in Tamil Nadu drop out voluntarily. This suggests that time spent preparing for the exam is costly, and that these costs bind. For some reason, the hiring freeze made candidates more willing to incur the costs associated with additional exam preparation. What factors explain why those additional costs were more worthwhile after the hiring freeze than they were before?

Second, why did candidates not take up employment until the hiring freeze was over? Recall, the hiring freeze policy did not include an end date when it was announced. This meant that there was substantial uncertainty over how long the hiring freeze would last. Candidates who continued to study during the hiring freeze ran the risk that their investments would have a lower return than they expected.[40] What prevented candidates from waiting to make these investment decisions until after the uncertainty was resolved?

In this section, I propose several plausible answers based on the ways in which structural features of the testing environment—namely, the kind of information that candidates have about their own probability of selection, and the underlying returns to studying—shape candidates' incentives. I then put together a collage of evidence to demonstrate the plausibility of the proposed hypotheses.

This discussion is meant to be suggestive, and does not rule out alternative mechanisms. In Appendix D, I outline a broader set of hypotheses that are consistent with the observed response. However, without detailed data on what candidates believed or were doing during the hiring freeze, it is difficult to assess many of these alternative hypotheses. One of the key advantages of focusing on the testing environment is that I can study the implications of the incentives it generates without relying on data that are necessarily from the same population as the one that experienced the hiring freeze.

5.1 Why might candidates be willing to study for longer?

One reason why candidates might be willing to study for longer is if they are generally over-optimistic about their chances of success, and they only revise their beliefs downwards after making unsuccessful attempts. Because the hiring freeze reduced the number of prior attempts that candidates had a chance to make before deciding whether to persist, they were more likely to make these decisions on the basis of over-optimistic beliefs. This, in turn, made candidates more willing to spend extra time on the exam track.

Implicit in this model of behavior are the following two testable hypotheses:

[40] This would happen if, for example, the hiring freeze lasted longer than expected, or if the number of vacancies announced after the end of the freeze was lower than expected.

- *Hypothesis 1*: Most candidates are over-optimistic about their exam performance, especially in early attempts
- *Hypothesis 2*: Re-application decisions are responsive to prior test scores.

I provide empirical evidence for each of these hypotheses in turn.

5.1.1 Candidates are over-optimistic about exam performance

Data. I use data from a survey of candidates preparing for the Maharashtra state civil service exam. Across India, state civil service exams ask similar types of questions, so candidates from Maharashtra should be able to forecast their own performance about as well as candidates from Tamil Nadu.[41]

The survey was conducted in September 2019. The sample consists of 88 candidates who were currently preparing for the state civil service exam in the city of Pune, Maharashtra. To target candidates who were preparing full time, I conducted the survey in three separate libraries in the city that are well-known to host a high-proportion of candidates preparing full time for state civil service exams.

As a part of the survey, respondents were asked to predict their score before they took a practice test. The practice test followed the same format as one of the most popular civil service exams conducted in Maharashtra.[42] This prediction was incentivized, following an adaption of the "crossover" mechanism used in Mobius, Niederle, Niehaus, and Rosenblat (2011). Respondents were asked whether they would prefer to obtain a reward of 200 INR via either a coin flip or if their score was higher than some threshold X.[43] The maximum possible score on the exam is 100 points, so X varied from 5 to 95 by 5. After the test was completed, I randomly selected one of the observed choices to determine whether respondents received the reward.[44] As discussed in Mobius et al. (2011), under a

[41] If anything, the candidates in Maharashtra are likely to have better information than candidates preparing during the hiring freeze, since the market for practice tests and coaching classes has expanded dramatically since the early 2000s.

[42] Namely, the practice test replicates the format of the Maharashtra Public Service Commission's "Group B" preliminary exam.

[43] This is a large reward. Respondents reported receiving on average Rs. 184 per day from home to support their living expenses while they were studying.

[44] The rewards were delivered to respondents within an hour of completing the practice test.

wide range of utility functions, the point at which respondents switch from preferring to bet on their score to betting on a coin flip identifies the median of the respondent's prior distribution of their test performance.[45] I use this switching point as my measure of the respondent's prediction.[46] I measure the bias of the prediction as follows: $Bias = (\hat{y} - y)$, where y is the realized test score and $\hat{y}$ is the prediction.

Before they made their prediction, respondents were provided with information that helped them calibrate the difficulty of the exam they were about to take. Specifically, they were told that a separate group of candidates had already taken the same practice test, and they were told what the average score was in this calibration sample was. Respondents were also told the number of participants in this “calibration sample" and how well this group had performed on various civil service exams conducted in the past year.

The survey also asked respondents to report their history with either the Maharashtra state civil service exams or the Union Public Service Commission exam.[47] I measure the number of prior years spent preparing as the number of years in which the respondent appeared for either a state exam or the UPSC exam. Because I only ask about exams since 2015, this variable is censored at 4.

Results. Figure 6a plots a scatter plot of the predicted score against the realized score. Note that almost all values are above the 45 degree line, which is consistent with respondents having positively biased beliefs about their performance.

The bias appears to diminish with prior experience. In Figure 6b, we see that respondents who have taken at least one prior test have less biased beliefs than candidates without any prior exam experience. The variation in the amount of prior experience is not exogenous, so it possible that factors that vary with experience also affect bias.

[45] In case a respondent did not naturally provide a switch point in their responses, the task was explained to that respondent again.

[46] Because there is a gap of 5 points each alternative, I do not observe the point at which respondents switch exactly. I therefore take the midpoint between the X value at which the respondent last prefers betting on his score and the value of X at which they prefer as a measure of the switch point.

[47] The latter is an exam for more prestigious central government exams. Some candidates who prepare for the state service exams simultaneously prepare for these posts as well. In this sample, about 23% have given a UPSC exam.

However, controlling for variables that plausibly affect both experience and bias does not attenuate the relationship (see Appendix Table A.9).

Discussion. Why are candidates over-optimistic in the first place? For reasons that Lazear (2016) cites, we should not expect this phenomenon to be unique to civil service exam preparation: as long as individuals observe their relative ability across occupations imperfectly, then sorting effects imply that on average people have positively biased beliefs about their ability in their chosen occupation. The more difficult it is for candidates to predict their performance on civil service exams, the stronger this effect should be. A key prediction of the Lazear model is that the bias should be largest for individuals at the beginning of their career, and diminish with experience. The evidence presented above is consistent with these predictions.

5.1.2 Candidates respond to prior test scores

Data. I use data on applications and test scores from the Tamil Nadu Public Service Commission. I observe the universe of general-skill civil service exams that were scheduled between 2012 and 2016.[48] For each of these exams, I observe the universe of applicants. In total, there were the 16 exams conducted during this period.

The estimation strategy relies on linking candidates across attempts. To do so, I match candidates using a combination of name, parents' names, and date of birth.[49] Overall, this method works very well: less than 0.1% of applications are marked as duplicates. I drop the handful of duplicates from the dataset. Because applications are official documents, it is costly for candidates to make mistakes in either spelling names or writing an incorrect date of birth. Therefore these fields tend to be consistent across time for the same candidate.[50]

[48] TNPSC also conducts exams for a wide variety of government positions that require specialized degrees, such as Chemist, or Geologist. Posts that require specialized skills attract far fewer applicants. During this period, the median notification for a general skill post attracted about 640,000 applicants, while the median notification for specialized skills attracted about 2500 applicants.

[49] To protect the identity of the candidates, TNPSC anonymized all names. In order to match names, I therefore compared a set of numeric IDs across examinations.

[50] To the extent that candidates are mismatched across attempts, the effects that we observe should be attenuated.

Empirical Strategy. To estimate the causal effect of prior test scores on re-application decisions, an ideal experiment would look something like the following. Imagine if, after all candidates completed their test, the scores that were reported back to them were perturbed by some random amount, unbeknownst to them. In that case, we would be able to compare two candidates who actually obtained the same actual test score, but who observed different signals based on the random perturbation. This way, two otherwise identical candidates would obtain different signals, and we can directly compare their responses.

The key idea behind my empirical strategy is that the measurement error that is inherent in standardized tests allows us to approximate this ideal experiment. In general, two candidates with the same ability will obtain different test scores due to luck. Using insights from Item Response Theory, a branch of the psychometrics literature, I can isolate the "luck component" of the test score from variation due to ability. This residual variation approximates a random shock that causes two otherwise identical candidates to observe different test scores. As long as candidates do not know their true ability, they should react to the variation induced by luck.

Estimating Luck. The total number of correct responses on a test does not necessarily incorporate all the available information. We may be able to obtain more precise estimates of ability by accounting for which candidates answer which questions correctly. This is because not all test questions *discriminate* to the same degree. A test question's discrimination is the rate at which higher ability candidates are more likely to answer the question correctly. A question with low discrimination is one in which high ability candidates are guessing about the correct answer nearly as much as low ability candidates.

I use a three-parameter logistic model to estimate the ability of each candidate. For each candidate, I observe $x_{ij} \in \{0, 1\}$, a matrix that tracks whether each candidate i answered question j correctly. According to the model, the probability of a correct answer is given by:

$$Pr(x_{ij} = 1|\theta_i, a_j, b_j, c_j) = c_j + (1 - c_j)\left[\frac{\exp(a_j(\theta_i - b_j))}{1 + \exp(a_j(\theta_i - b_j))}\right] \tag{7}$$

where θ_i is the individual's ability, a_j reflects the discrimination coefficient, b_j captures the difficulty of the question, and c_j captures the probability of guessing the question correctly.

Assuming that responses across questions are independent, we can estimate the model parameters by maximizing the following likelihood function:

$$\mathcal{L}(X) = \prod_i \prod_j Pr(x_{ij} = 1|\theta_i, a_j, b_j, c_j) \tag{8}$$

Because of the high dimensionality of the parameter space and a lack of a closed form solution, the estimates are obtained using an Expectation Maximization algorithm.[51]

A key output of the model is the *score residual.* For each candidate, the model generates an estimate $\hat{\theta}_i$ of his ability. It also generates estimates of the question-specific parameters $(\hat{a}_j, \hat{b}_j, \hat{c}_j)$. Using these estimated parameters, I can use the model to generate a predicted score, $\hat{T} = \sum_j Pr(x_{ij} = 1|\hat{\theta}_i, \hat{a}_j, \hat{b}_j, \hat{c}_j)$. Let us call $T - \hat{T}$ the *score residual.* This will be my measure of the "luck component" of the test score.

Interpreting the residual as a measure of luck depends critically on whether the model is correctly specified. To assess the fit of the model, in Appendix Figure A.7 I plot the average score residual against estimated ability. If the model is well-specified, then the average residual should be zero across the distribution of ability. We see that the model tends to fit the data reasonably well for ability estimates between -2 and 1, but the fit deteriorates towards the extremes of the distribution. Thus, unless noted otherwise, in all of the subsequent analysis I restrict the sample to candidates with estimated ability in between -2 and 1, where the fit is better.

Because the Item Response Theory model uses more information than the total correct responses, it produces ability estimates that are not perfectly correlated with the total test score. In Figure 7a, I plot test scores against the estimated ability parameters for a particular exam from 2013, the 2013/09 Group 4 exam. The variation in actual test scores among candidates with the same estimated ability generates variation in the score

[51] This algorithm is implemented in the *mirt* package. The documentation for this package is available in Chalmers et al. (2012).

residual. In Figure 7b, I plot the histogram of the score residual in this sample. Note that distribution is wide: candidates with the same ability can experience fluctuations in test scores of up to 30 points in either direction, out of a total of 300 possible points. About 9.6% of the total variance in test scores can be accounted for by variance in the score residual.

Estimating how candidates respond to variation in test scores. I estimate an OLS regression of the following form:[52]

$$applied_{i,t} = \alpha + \beta residual_{i,0} + f(\theta_i) + \epsilon_i \tag{9}$$

where $applied_{i,t}$ is an indicator for whether candidate i applied for exam t, and $residual_{i,0}$ is the score residual calculated from a baseline exam. For this analysis I use the 2013/09 Group 4 exam as the baseline exam. I control flexibly for the effect of ability, $f(\theta)$, by splitting the distribution of ability into ventiles and including a dummy for each bin in the regression.[53] I assume the errors terms are independent across candidates. Identification of β depends on a conditional independence assumption: ϵ_i should be independent of $residual_{i,0}$, conditional on θ_i.

To match the population that I study in Sections 3 and 4, I restrict the sample male, college graduates, who were younger than 30 at the time of the baseline exam, and who had minimal prior testing experience (which I proxy by dropping individuals who made any application in 2012).

Results. Figure 8 summarizes the main result. It presents estimates of β from equation (9) for each of the 11 exams that were conducted after the baseline exam. We see two patterns that are consistent with the hypothesis that candidates learn from past experience. First, we see large positive coefficients for the exams that were conducted shortly after the baseline exam. This tells us that candidates with higher than expected

[52] Note that because percentile rank vary linearly with the rank residual, this specification produces identical estimates as one that instruments percentile rank with the residual.

[53] To increase the security of the exam, there are several different versions of the exams that are administered at the same time. Each version has the same set of questions but presents them in a different order. The ability parameters are estimated *within* the set of candidates that take a given version of the test. The specification therefore also interacts the ability ventile with the exam group ID.

test scores were more likely to re-apply in the future. Second, we see that the effect of the information shock from the baseline exam decays over time. For exams conducted more than a year after the baseline exam, the effect of the information shock is close to zero. This also makes sense. As candidates gain additional experience, prior information should become less relevant.

Discussion. A back-of-the-envelope calculation suggests that the marginal effect of lower-than-expected test scores is not quite enough to explain the increase in unemployment that we saw in Section 3.[54] However, this marginal effect holds constant being unsuccessful on the exam in the first place, which may have a direct effect in its own right. For example, candidates who are unsuccessful may get demoralized, or have a harder time convincing their family to continue to support their studies. Indeed, the fact that about 60% candidates drop out permanently after what is likely their first year of attempting the exam (see Appendix Figure A.8) suggests that re-application after a single failed attempt is particularly costly.[55] If we assume that the observational drop-out rate is causal—that is, that the hiring freeze causes 60% of candidates to persist instead of dropping out—then it is possible to reasonably account for the unemployment effect of the hiring freeze.[56]

[54] The population for which we observed the largest impact of the hiring freeze was male recent college graduates. According to the 2001 Census, there were 240,512 male college graduates between the ages of 20 to 24 in the population of Tamil Nadu. Meanwhile, there were 405,927 applicants in the last large group exam conducted before the hiring freeze (see Table 1), of which current application data suggests that 12% belong to this demographic. On average candidates score 70 points below the passing cutoff mark. If we assume that all candidates have very naive priors and believe they would score at the cutoff prior to taking the test, then the test score shock is 70 points. The estimates imply that an information shock of 70 points would have decreased persistence by $0.00164 \times 70 = 11.5\%$. In other words, the absence of this information shock might have increased persistence by 11.5%. This would only lead to an increase in the fraction of the population applying of about $(240,512 \times 12\% \times 11.5\%) / 405,927 = 2.3\%$. Since not all candidates who apply are unemployed, the effect on unemployment is likely smaller.

[55] Because TNPSC does not track attempts directly, I proxy for the first attempt by restricting attention to a sub-sample: i) between the ages of 20-22 at the time of the attempt; that ii) did not take appear for any exam prior to the first observed exam.

[56] Continuing from Footnote 54, an increase in persistence of 60% implies that the effect on the fraction of the population applying would be about $(240,512 \times 12\% \times 60\%) / 405,927 = 12.2\%$. It is unknown what fraction of applicants prepare full time, but if that number is close to 50% (which is plausible for this particular demographic) then the overall effect on unemployment would be about 6%.

5.2 Why not work until the hiring freeze is over?

One of the main reasons why we might expect to see employment rates increase during the hiring freeze is that candidates had the option of studying again after the hiring freeze was lifted. This is true even if they remained over-optimistic about their probability of success. The fact that most candidates choose to remain unemployed during the hiring freeze suggests that interrupting exam preparation is costly. One reason that might be the case is if the returns to exam preparation are convex. In that case, candidates who drop out would not be able to easily catch up with those who kept studying during the hiring freeze. This generates a strategic incentive for candidates to keep studying.[57]

5.2.1 The returns to exam preparation are convex

Data. I continue using application and test score data from TNPSC, with some of the same sample restrictions: I limit the analysis to individuals with i) with college degrees; ii) who participated in the 2013/09 Group 4 exam (which I again use as the baseline exam); and iii) who have an estimated ability measure θ_i between -2 and 1 (see Section 5.1.2 for more details). A key difference is that I no longer restrict the sample to men or those without prior test-taking experience. This is done to increase the sample size, which maximizes statistical power for the non-linear IV that I rely on to estimate the returns to exam preparation.

Empirical Strategy. The goal is to estimate how the amount of time spent preparing for civil service exams affects the probability of success. Ideally, I would use a direct measure of the amount of time spent studying for the civil service exam. However, in the absence of direct measures I proxy for study time using attempts. This proxy is reasonable if either: i) candidates study for a fixed amount before each test; or ii) candidates study continuously but the exams are roughly evenly spaced.

The instrumental variable strategy that I use in Section 5.1.2 also provides a first stage for estimating the convexity of the returns to exam preparation. The instrument introduces random variation in the number of subsequent attempts that candidates make.

[57] A formal model of this logic is provided in Appendix D.

I can then use this variation to estimate how test scores and the probability of success depend on the number of attempts made.

I estimate a two-stage least squares regression. The second-stage specification is:

$$selection_{i,1} = \beta_0 + \beta_1 \widehat{attempts}_{i,1} + \beta_2 \widehat{attempts^2}_{i,1} + f(\theta_i) + \epsilon_i \tag{10}$$

where $selection_{i,1}$ is an indicator for whether the candidate was successful in any of the three exams that were notified after the baseline exam.[58] The dependent variable, $attempts_{i,1}$, measures the number of attempts made on the three exams that were notified after the baseline exam. The fitted values for this regression come from:

$$attempts_{i,1} = \gamma_0 + \gamma_1 residual_{i,0} + \gamma_2 residual_{i,0}^2 + g(\theta_i) + \nu_i \tag{11}$$

$$attempts_{i,1}^2 = \delta + \delta_1 residual_{i,0} + \delta_2 residual_{i,0}^2 + h(\theta_i) + \eta_i \tag{12}$$

The parameters of interest are β_1 and β_2, which are just identified.

Results. Table 5 presents the results. Column (1) presents coefficients for the OLS estimate of equation (10); column (2) presents the IV estimates. In both cases, the coefficient on (Additional Attempts)^2 is positive, consistent with the returns to additional attempts being convex.

6 Conclusion

The public sector hiring freeze that the government of Tamil Nadu enacted between 2001 and 2006 had far-reaching consequences for the labor market. Cohorts that graduated from college during the hiring freeze spent substantially more time unemployed. The extra time spent in unemployment among college graduates appears to reflect increased investments in exam preparation. A decade after the end of the hiring freeze, these same cohorts have lower earnings.

[58] Multiple selection is uncommon. Of the 1,679 candidates in the sample selected in these three exams, 91% were selected only once.

These findings suggest that, in the current environment, reductions in government hiring can have substantial adverse affects. As long as candidates face strong incentives to continue studying, a regular and timely testing policy may help reduce the unemployment rate on the margin.[59] The importance of the convexity of the returns to exam preparation invites us to consider ways to “de-convexify" the selection process. One possible mechanism would be to randomly select among candidates who clear a certain minimum threshold, instead of selecting candidates based on their rank order in each exam.

The implications of these findings for public sector wages is not obvious. Although in theory reducing the size of the prize would, in theory, reduce the incentive to prepare as intensely, doing so may also reduce the caliber of applicants (Dal Bó et al., 2013), or affect morale within the civil service (Mas, 2006), which would then affect service delivery. These costs are uncertain, and may offset any potential gains from reducing rent-seeking.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge (2017), “When should you adjust standard errors for clustering?” Technical report, National Bureau of Economic Research.

Abebe, Girum Tefera, Stefano Caria, Marcel Fafchamps, Paolo Falco, Simon Franklin, and Simon Quinn (2018), “Anonymity or distance? job search and labour market exclusion in a growing african city.”

Ashraf, Nava, Oriana Bandiera, Edward Davenport, and Scott S Lee (2020), “Losing prosociality in the quest for talent? sorting, selection, and productivity in the delivery of public services.” *American Economic Review*, 110, 1355–94.

Ashraf, Nava, Oriana Bandiera, and B Kelsey Jack (2014), “No margin, no mission? a field experiment on incentives for public service delivery.” *Journal of public economics*, 120, 1–17.

Autor, David H, David Dorn, Gordon H Hanson, and Jae Song (2014), “Trade adjustment: Worker-level evidence.” *The Quarterly Journal of Economics*, 129, 1799–1860.

Banerjee, Abhijit and Gaurav Chiplunkar (2018), “How important are matching frictions in the labor market? experimental & non-experimental evidence from a large indian firm.” Technical report, Working paper.

Banerjee, Abhijit V and Esther Duflo (2019), *Good economics for hard times: Better answers to our biggest problems*. Penguin UK.

[59]Delays in the testing process are very common at the state level, especially in states like Uttar Pradesh and Bihar that also tend to have high levels of unemployment.

Bank, The World (2004), "State fiscal reforms in india: Progress and prospects. report no. 28849-in." Technical report.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller (2008), "Bootstrap-based improvements for inference with clustered errors." *The Review of Economics and Statistics*, 90, 414–427.

Chalmers, R Philip et al. (2012), "mirt: A multidimensional item response theory package for the r environment." *Journal of statistical Software*, 48, 1–29.

Colonnelli, Emanuele, Mounu Prem, and Edoardo Teso (2020), "Patronage and selection in public sector organizations." *American Economic Review*, 110, 3071–99.

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

Donald, Stephen G and Kevin Lang (2007), "Inference with difference-in-differences and other panel data." *The review of Economics and Statistics*, 89, 221–233.

Feng, Ying, David Lagakos, and James E Rauch (2018), "Unemployment and development." Technical report, National bureau of economic research.

Finan, Frederico, Benjamin A Olken, and Rohini Pande (2017), "The personnel economics of the developing state." In *Handbook of Economic Field Experiments*, volume 2, 467–514, Elsevier.

Geromichalos, Athanasios and Ioannis Kospentaris (2020), "The unintended consequences of meritocratic government hiring." Technical report.

Jayachandran, Seema (2006), "Selling labor low: Wage responses to productivity shocks in developing countries." *Journal of political Economy*, 114, 538–575.

Jeffrey, Craig (2010), "Timepass: Youth, class, and time among unemployed young men in india." *American Ethnologist*, 37, 465–481.

Jensen, Robert (2012), "Do labor market opportunities affect young women's work and family decisions? experimental evidence from india." *The Quarterly Journal of Economics*, 127, 753–792.

Kahn, Lisa B (2010), "The long-term labor market consequences of graduating from college in a bad economy." *Labour economics*, 17, 303–316.

Krueger, Anne O (1974), "The political economy of the rent-seeking society." *The American economic review*, 64, 291–303.

Kumar, Sanjay (2019), *Youth in India: Aspirations, Attitudes, Anxieties*. Taylor & Francis.

Kumar, Sanjay and Pranav Gupta (2018), "What young india wants: 'sarkari naukri'." *Livemint*.

Lazear, Edward P (2016), "Overconfidence and occupational choice." Technical report, National Bureau of Economic Research.

Lazear, Edward P and Sherwin Rosen (1981), "Rank-order tournaments as optimum labor contracts." *Journal of political Economy*, 89, 841–864.

MacKinnon, James G and Matthew D Webb (2018), "The wild bootstrap for few (treated) clusters." *The Econometrics Journal*, 21, 114–135.

Mas, Alexandre (2006), "Pay, reference points, and police performance." *The Quarterly Journal of Economics*, 121, 783–821.

Mobius, Markus M, Muriel Niederle, Paul Niehaus, and Tanya S Rosenblat (2011), "Managing self-confidence: Theory and experimental evidence." Technical report, National Bureau of Economic Research.

Muralidharan, Karthik (2015), "A new approach to public sector hiring in india for improved service delivery." In *India Policy Forum*, volume 12, Brookings-NCAER.

Murphy, Kevin M, Andrei Shleifer, and Robert W Vishny (1991), "The allocation of talent: Implications for growth." *The quarterly journal of economics*, 106, 503–530.

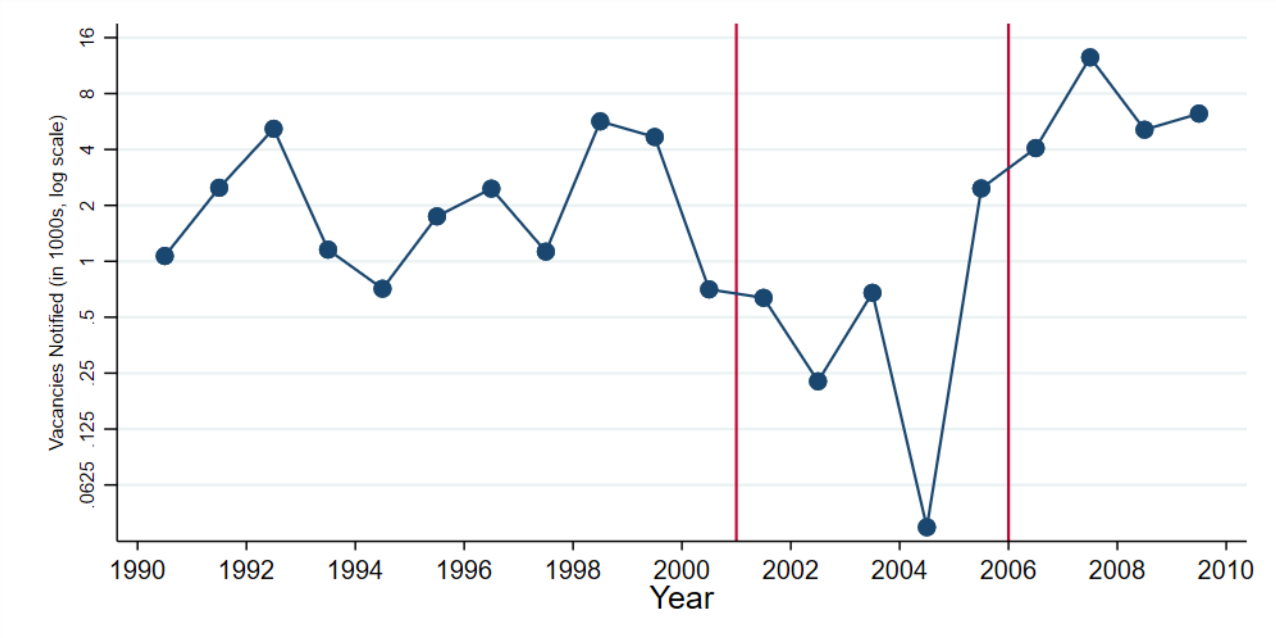
Oreopoulos, Philip, Till Von Wachter, and Andrew Heisz (2012), "The short-and long-term career effects of graduating in a recession." *American Economic Journal: Applied Economics*, 4, 1–29.

Rajan, MC (2020), "Tamil nadu disqualifies, bans 99 cheating group iv examinees for life." URL https://www.hindustantimes.com/india-news/tamil-nadu-disqualifies-bans-99-cheating-group-iv-examinees-for-life/story-Bkzk7iYWTk8GoRsyifYUCM.html.

Schwandt, Hannes and Till Von Wachter (2019), "Unlucky cohorts: Estimating the long-term effects of entering the labor market in a recession in large cross-sectional data sets." *Journal of Labor Economics*, 37, S161–S198.

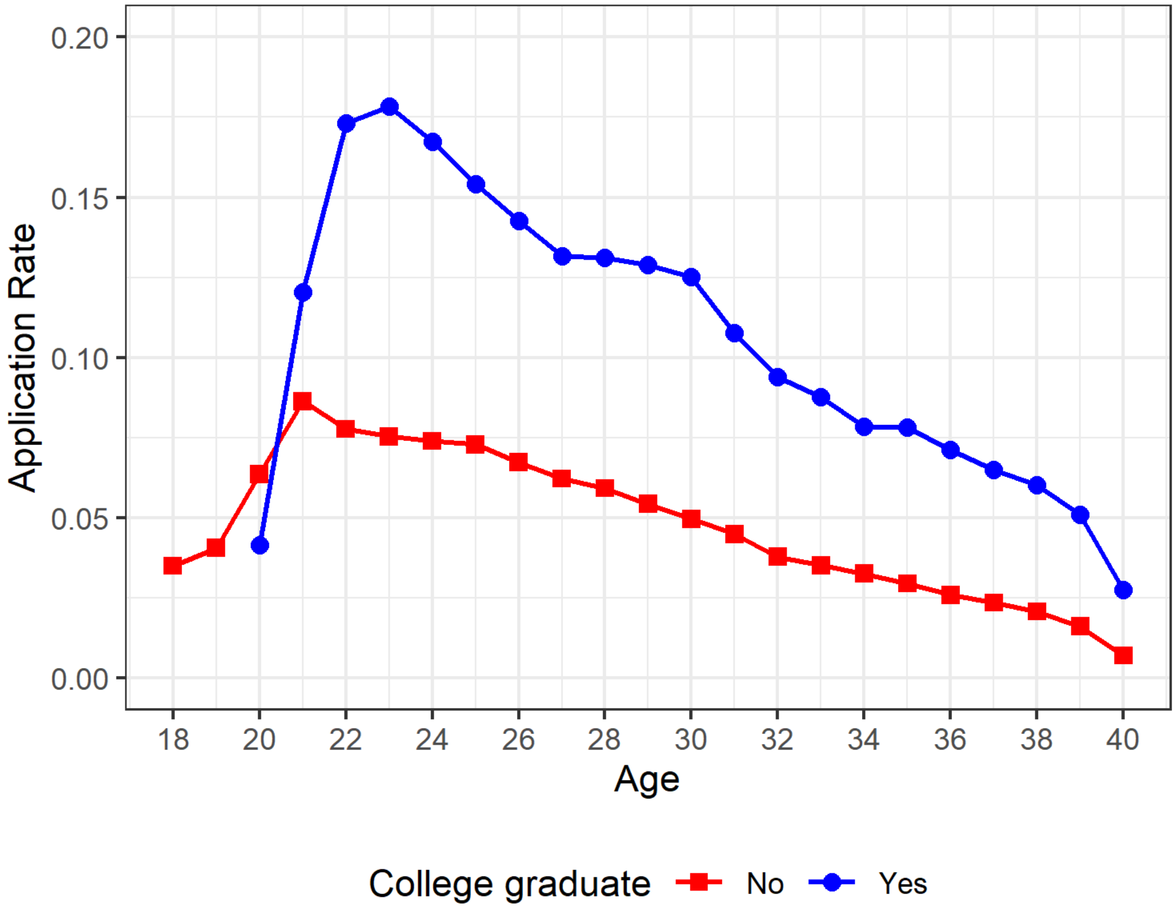
7 Figures

Figure 1: Available Vacancies Fall Dramatically During the Hiring Freeze



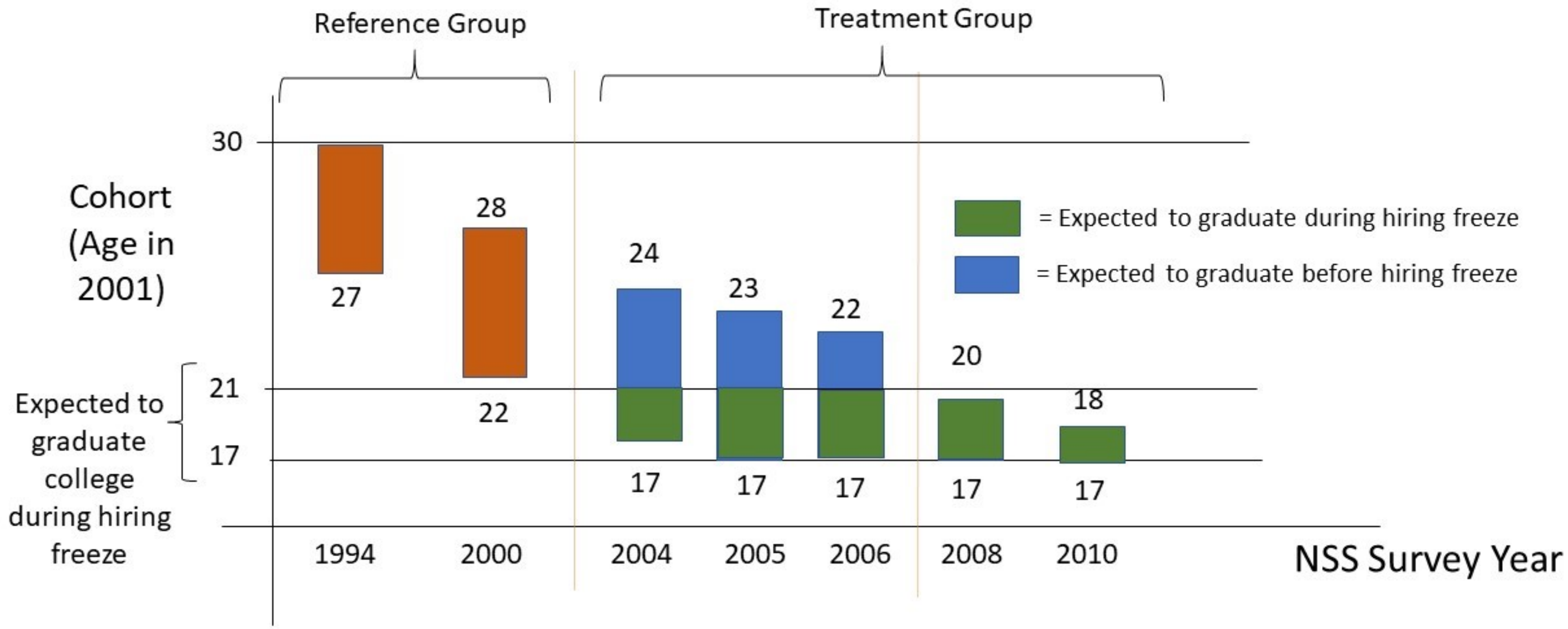
Notes: Data sourced from the Annual Reports of the Tamil Nadu Public Service Commission, 1995 to 2010. The figure plots the total number of vacancies advertised for the given fiscal year. Red lines mark the beginning and end of the hiring freeze. The Y-axis is in log scale.

Figure 2: What Fraction of Eligible Men Apply for Posts through TNPSC?



Data: 1) Administrative data from the Tamil Nadu Public Service Commission for the exams conducted according to the following notifications: 2012/14; 2012/26; 2013/09; and 2014/07; and 2) the 2011 Census of India.
Notes: I use the date of birth included in the application to calculate each candidate's age on the last date to apply for the exam. I then divide the number of applications in each age/education bin with an estimate of the corresponding population size in Tamil Nadu from the 2011 Census. Applications from outside of Tamil Nadu are a negligible share of the overall application pool. The Census reports population estimates by educational attainment according to age ranges, e.g. 20-24, 25-29, and so on. I divide the total reported population level by the size of the age bin, and then compute a three-year moving average to obtain a smoothed population series by age. The graph plots the average application rate across the four exams included in the sample.

Figure 3: Empirical Strategy



Notes: Figure illustrates the empirical strategy used in Section 3.

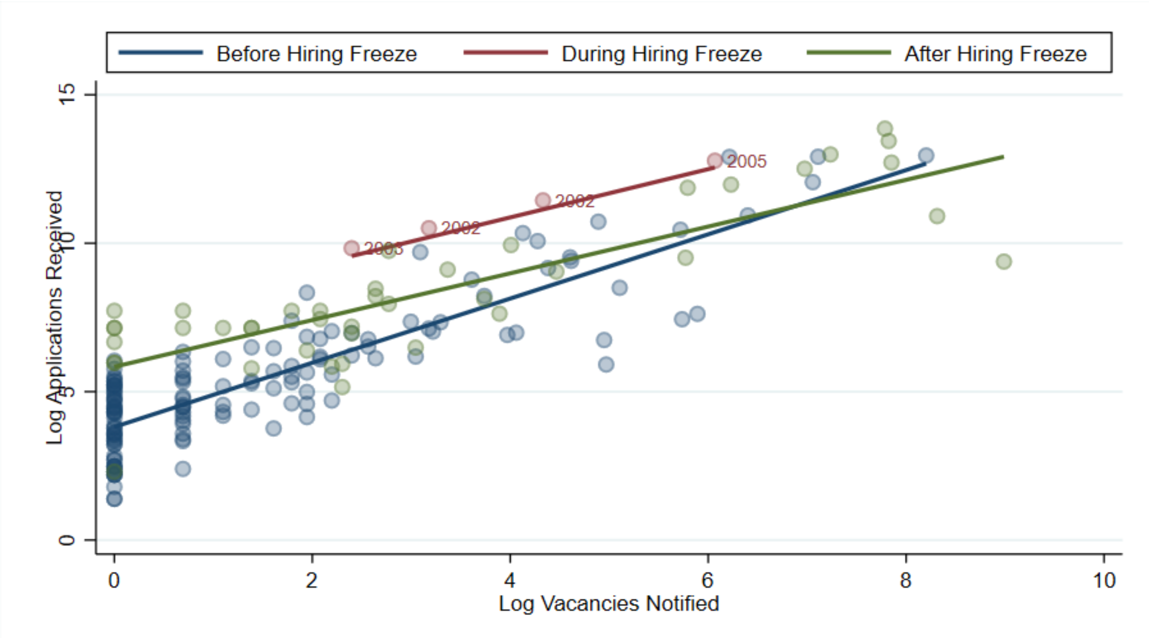
Figure 4: Unadjusted Difference-in-Differences Estimates of Short-Run Impacts on Labor Supply

(a) Unemployment

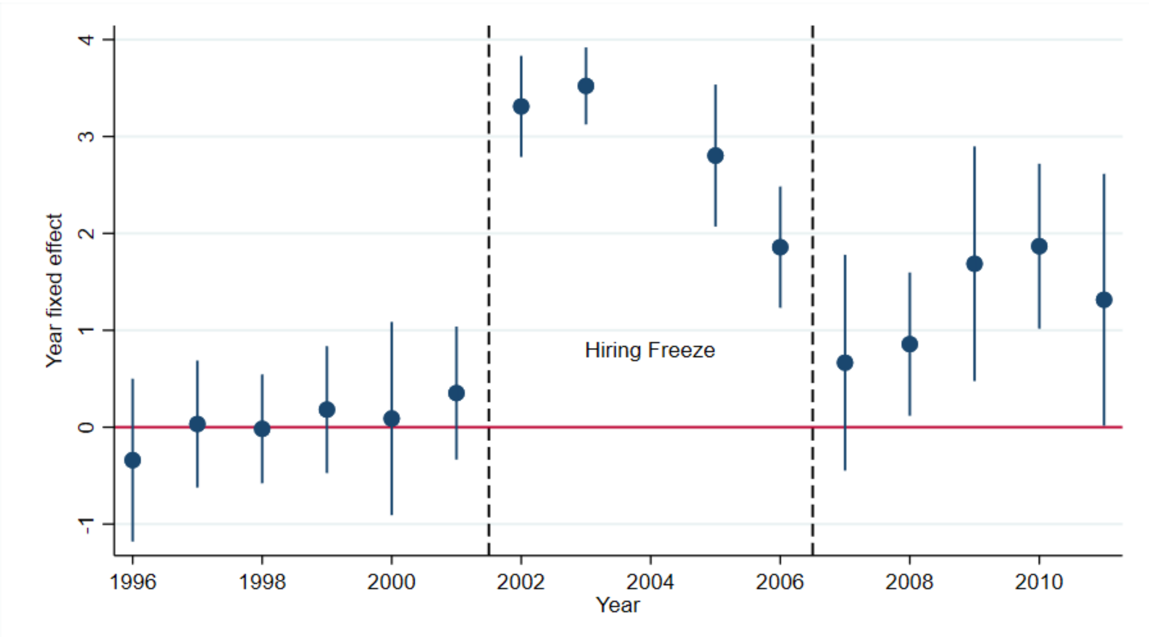
(b) Employment

Data: National Sample Survey, 1994 to 2010.
Notes: Sample restricted to college-educated men between the ages of 21 to 27 at the time of the survey. For each cohort whose outcomes were measured after the implementation of the hiring freeze, I compute a simple difference-in-differences estimate. I first compute unweighted average outcomes by state x cohort x year cells. Let s index states, c index cohorts, $t \in \{0,1\}$ be an indicator for whether outcomes were measured after the hiring freeze, and TN_s be an indicator for Tamil Nadu. Then for each outcome y, I present estimates of: $\big(E[y \,|\, c, t = 1, TN_s = 1] - E[y \,|\, t = 0, TN_s = 1]\big) - \big(E[y \,|\, c, t = 1, TN_s = 0] - E[y \,|\, t = 0, TN_s = 0]\big)$.

Figure 5: The Application Rate Increases During the Hiring Freeze



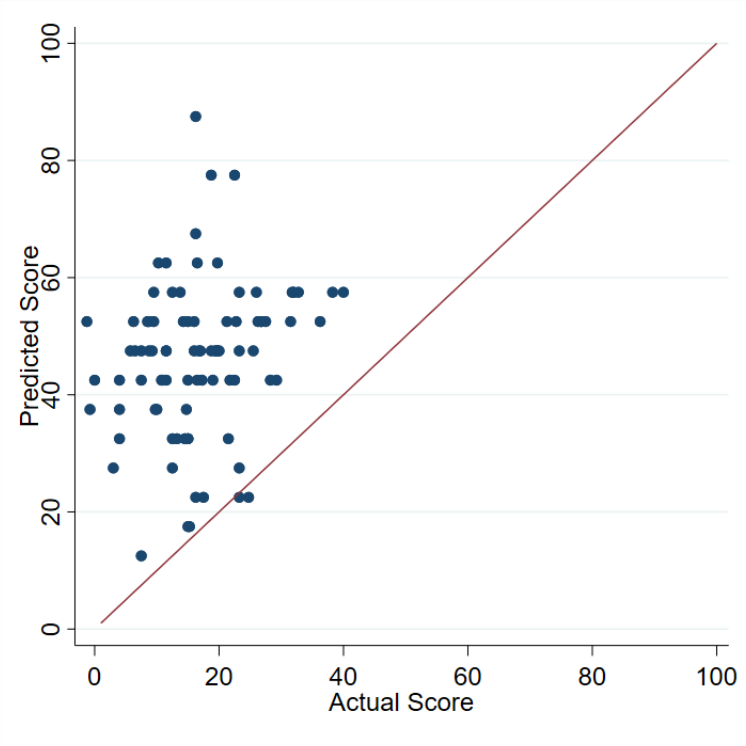
(a) Aggregate effect



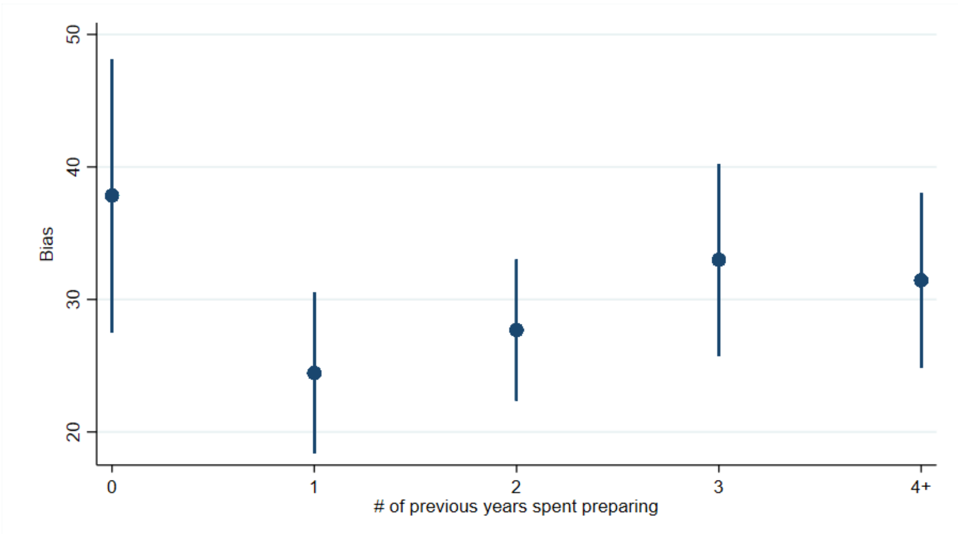
(b) Year-by-year effects

Data: Annual Reports of the Tamil Nadu Public Service Commission, 1995-2010.
Notes: The unit of observations is a recruitment that was notified and completed between 1995 and 2010. The sample excludes posts classified as "state" level. Recruitments are dated according to the year in which applications were last accepted. Recruitments were marked as occuring during the hiring freeze if the last date to apply occurred during the freeze. Panel A: For each recruitment published in the report, the figure plots the log of the applications received against the log of the vacancies advertised. Panel B: plots the α_t coefficients from equation (4), where 1995 is the base year.

Figure 6: Candidates are over-optimistic about exam performance, especially on early attempts



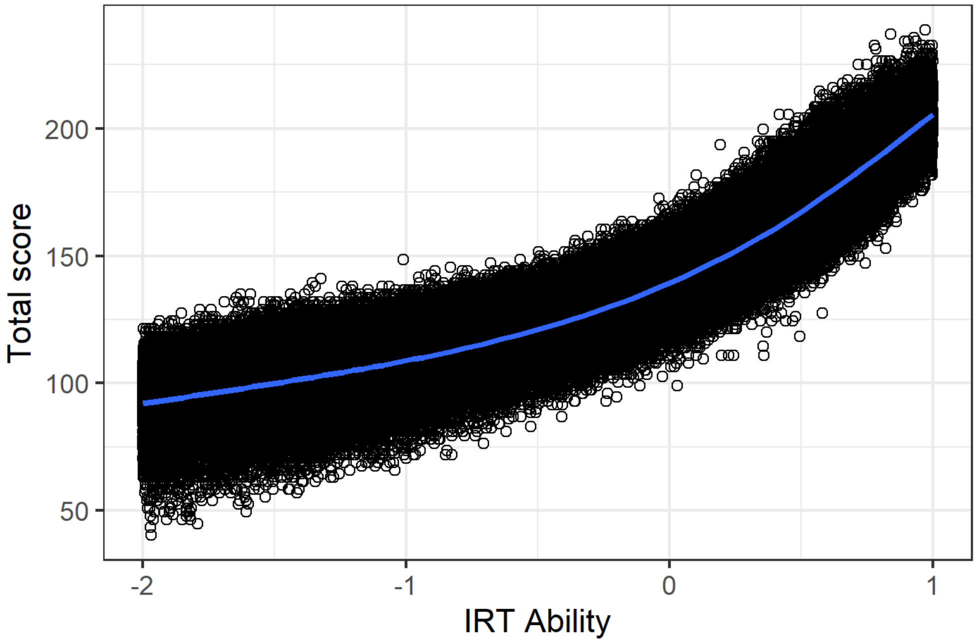
(a) Predicted vs. Realized Test Scores



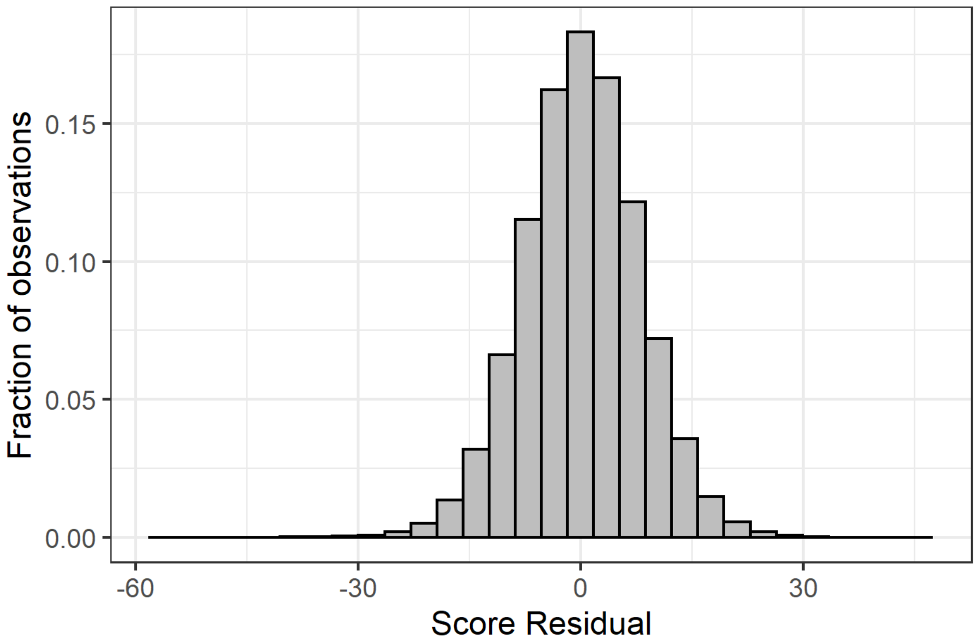
(b) Bias vs. prior experience

Data: Survey data from Pune, Maharashtra.
Notes: Respondents were asked to predict their performance on a practice test right before they took it. The prediction was done in an incentivized manner. See Section 5.1.1 of the main text for more details on how this prediction was elicited. The top figure shows the discrepancy between the prediction and the realized test score. The red line plots the 45 degree line. The bottom figure shows average bias (i.e. predicted score - realized score) depending on the number of prior years in which the respondent participated in a civil service exam. Error bars show 95% confidence intervals under the assumption that errors are independent across respondents.

Figure 7: Extracting the Luck Component of the Test Score



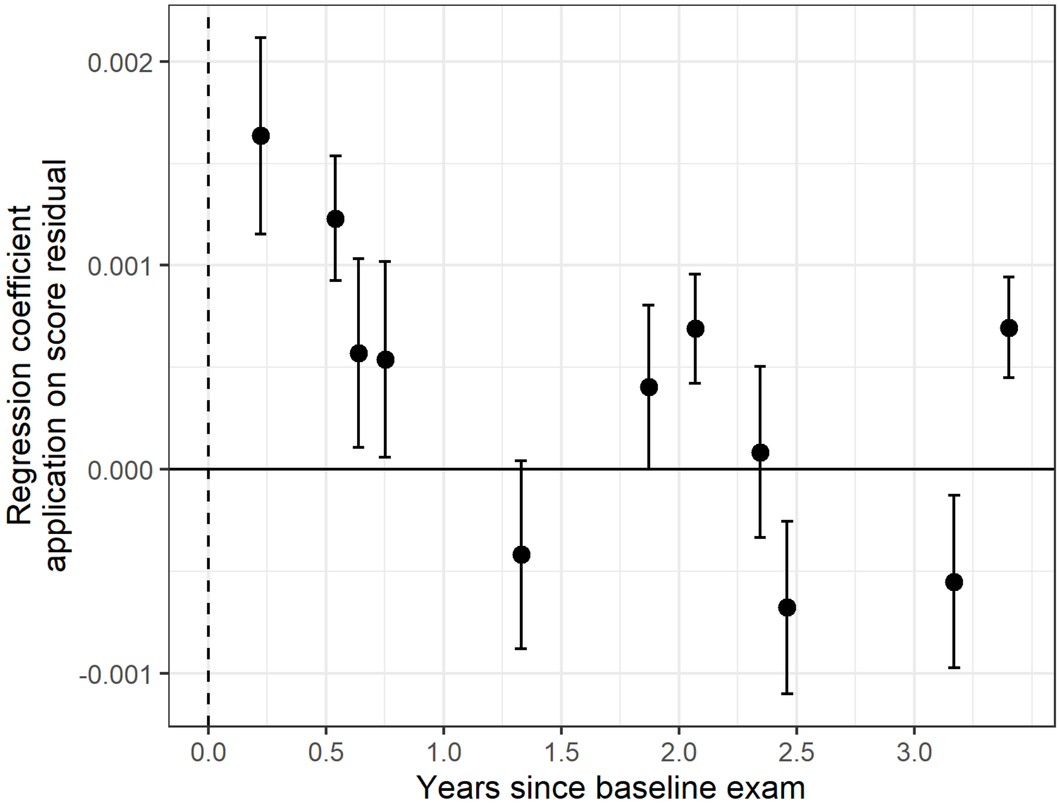
(a) Correlation between IRT ability and test scores



(b) Distribution of the score residual

Data: Administrative data from the Tamil Nadu Public Service Commission, 2013/09 exam.
Notes: IRT Ability is a measure of ability that accounts for the specific questions that candidates answered correctly. The ability parameter is estimated using the model described in Equation 7 of the main text. Figure a) plots a scatter plot; the blue line plots a local linear regression. Figure b) plots the score residual. This is the difference between the candidate's actual score and the predicted score according to the model. In both cases, the sample is restricted to individuals where the fit of the IRT model is reasonably good, which is for individuals with an estimated ability greater than -2 and less than 1.

Figure 8: Candidates base re-application decisions on past test scores



Data: Administrative data from the Tamil Nadu Public Service Commission.
Notes: The figure plots the impact of score variation in the baseline exam (the 2013 Group 4 exam) on subsequent exam-taking. The x-axis plots the years between a particular exam and the baseline exam (which is the 2013 Group 4 exam). The y-axis plots the estimate of the β coefficient from the regression specified in equation (9). This is a regression of a dummy of whether the candidate applied for a particular exam and the rank residual on the baseline exam. The error bars plot 95% confidence intervals.

8 Tables

Table 1: Application Intensity in Tamil Nadu

Group 4 Recruitment Notified in 1999	
Vacancies	310
Applications Received	405,927
Application to Vacancy Ratio	1,309
Eligible Population (18-40)	7,169,276
Share of eligible population applying	5.6%

Notes: This table summarizes statistics for a particular recruitment conducted by the Tamil Nadu Public Service Commission (TNPSC). All data sourced from TNPSC except data on the eligible population, which is calculated from the 2001 Indian Census. The eligible population refers to the total number of Tamil Nadu residents with at least a 10th standard education between the ages of 18-40.

Table 2: Short-Run Impacts of the Hiring Freeze on Male College Graduates

	(1) Unemployed	(2) Employed	(3) Out of labor force
Panel A: Diff-in-diff estimates, college sample			
TN $\times$ Age 17-21 in 2001, Post-freeze (β_1)	0.065**	-0.063*	-0.002
	[0.010, 0.127]	[-0.125, 0.004]	[-0.088, 0.074]
TN $\times$ Age 22-24 in 2001, Post-freeze (β_2)	-0.035	0.071	-0.036
	[-0.206, 0.078]	[-0.072, 0.263]	[-0.168, 0.096]
Mean, reference group in TN	0.211	0.491	0.298
Observations	19,299	19,299	19,299
Panel B: Diff-in-diff estimates, ineligible sample			
TN $\times$ Age 17-21 in 2001, Post-freeze ($\tilde{\beta}_1$)	-0.001	0.005	-0.004
	[-0.020, 0.018]	[-0.029, 0.035]	[-0.026, 0.022]
TN $\times$ Age 22-24 in 2001, Post-freeze ($\tilde{\beta}_2$)	-0.004	0.006	-0.002
	[-0.030, 0.021]	[-0.025, 0.039]	[-0.045, 0.034]
Mean, reference group in TN	0.033	0.930	0.037
Observations	90,284	90,284	90,284
Panel C: Triple difference estimates			
$\beta_1 - \tilde{\beta}_1$	0.066**	-0.068*	0.002
	[0.009, 0.135]	[-0.148, 0.014]	[-0.092, 0.081]
$\beta_2 - \tilde{\beta}_2$	-0.032	0.065	-0.034
	[-0.189, 0.090]	[-0.069, 0.231]	[-0.149, 0.106]
Observations	109,583	109,583	109,583

Data: National Sample Survey, 1994 to 2010.
Notes: Panel A presents difference-in-differences estimates of the impact of the hiring freeze on employment status for the main sample of interest. This sample is: 1) men; 2) who are college graduates; 3) between the ages of 21 to 27 at the time of the survey; and 4) who were between the ages of 17 to 30 in 2001. Coefficients correspond to β_1 and β_2 from equation (1) in the main text. Panel B presents an over-identification test of the parallel trends assumption, estimating equation (1) on the sample of individuals ineligible for government jobs, i.e. those with less than a 10th standard education. Panel C presents triple difference estimates, differencing the coefficients from Panels A and B. 95% confidence intervals in brackets, computed via wild bootstrap with 999 replications, clustered by state x cohort level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Shifts in the Average Application Rate Over Time

	(1) Log Applications	(2) Log Attended Exam
During Hiring Freeze (β_1)	3.01***	3.61***
	(0.33)	(0.33)
After Hiring Freeze (β_2)	1.28***	1.73***
	(0.26)	(0.27)
Log Vacancies	0.97***	0.98***
	(0.06)	(0.06)
p-value: $\beta_1 = \beta_2$	0.000	0.000
Observations	181	181

Data: Annual Reports of the Tamil Nadu Public Service Commission, 1995-2010.
Notes: The unit of observations is a recruitment that was notified and completed between 1995 and 2010. The sample excludes posts classified as "state" level. Recruitments are dated according to the year in which applications were last accepted. Recruitments were marked as occuring during (after) the hiring freeze if the last date to apply occurred during (after) the freeze. Columns present coefficient estimates of equation (3) from the main text. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Long-run Impacts of the Hiring Freeze on Labor Market Earnings

	(1) Log Earnings	(2) Earnings < 20K	(3) Share of HH Income	(4) Log HH Income
Panel A: Diff-in-diff estimates, college sample				
TN × Age 17-21 in 2001 (β_1)	-0.020	0.062*	-0.054***	0.063
	[-0.153, 0.121]	[-0.012, 0.133]	[-0.090, -0.018]	[-0.131, 0.284]
TN × Age 22-26 in 2001 (β_2)	0.049	0.006	-0.012	0.050
	[-0.078, 0.184]	[-0.090, 0.100]	[-0.068, 0.047]	[-0.149, 0.264]
Mean, reference group in TN	9.761	0.589	0.718	10.101
Unique individuals	50,742	53,472	52,924	52,924
Observations	677,974	752,592	744,298	744,298
Panel B: Diff-in-diff estimates, ineligible sample				
TN × Age 17-21 in 2001 ($\tilde{\beta}_1$)	0.029	-0.002	-0.010	0.032
	[-0.043, 0.097]	[-0.012, 0.006]	[-0.036, 0.016]	[-0.036, 0.094]
TN × Age 22-26 in 2001 ($\tilde{\beta}_2$)	0.038	0.001	0.006	0.030
	[-0.025, 0.103]	[-0.009, 0.010]	[-0.019, 0.033]	[-0.039, 0.100]
Mean, reference group in TN	8.945	0.978	0.601	9.472
Unique individuals	106,571	114,606	112,872	112,872
Observations	1,586,145	1,763,558	1,734,713	1,734,713
Panel C: Triple difference estimates				
$\beta_1 - \tilde{\beta}_1$	-0.048	0.065*	-0.044*	0.031
	[-0.176, 0.087]	[-0.013, 0.134]	[-0.088, 0.007]	[-0.152, 0.221]
$\beta_2 - \tilde{\beta}_2$	0.011	0.006	-0.019	0.020
	[-0.118, 0.139]	[-0.087, 0.100]	[-0.083, 0.042]	[-0.167, 0.231]
Unique individuals	157,313	168,078	165,796	165,796
Observations	2,264,119	2,516,150	2,479,011	2,479,011

Data: CMIE-Consumer Pyramids Household Survey, 2014-2019.
Notes: Sample restricted to: 1) men; 2) between the ages of 17 to 30 in 2001. Panels A and B report estimates from equation (5) from the main text, where cohorts ages 27 to 30 are the reference category. Panel A uses the sample of college graduates; Panel B uses the ineligible sample, i.e. those with less than a 10th standard education. Panel C reports estimates from equation (6). Earnings reported in real 2014 INR. Earnings include all labor market earnings, including self-employment income and in-kind income. Household income includes both labor market earnings, and earnings from assets and transfers. 95% confidence intervals in brackets constructed via wild bootstrap with 999 replications, clustered at the state x cohort level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Estimating the Convexity of the Returns to Additional Attempts

	(1)	(2)
Additional Attempts	-0.0044^{***}	-0.32^{***}
	(0.0003)	(0.07)
(Additional Attempts)^2	0.0031^{***}	0.12^{***}
	(0.0001)	(0.02)
Specification	OLS	IV
Kleibergen-Paap F	-	18.2
Controls	None	Estimated Ability
Observations	518,256	518,256

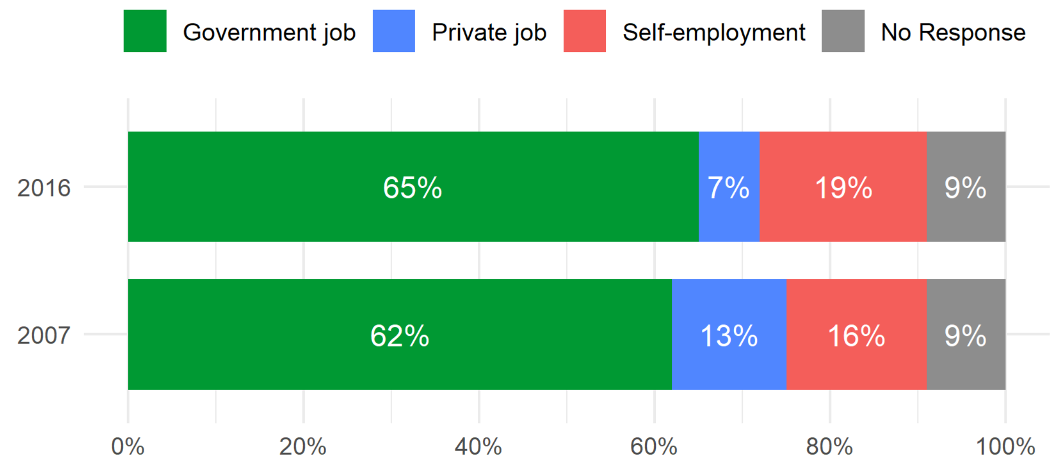
Data: Tamil Nadu Public Service Commission Administrative Data, 2013-2014.
Notes: The sample consists of candidates who: 1) appeared for the 2013 Group 4 exam; 2) are college graduates; and 3) whose estimated ability is between -2 and 1 (see Section 5.1.2 for more details). The dependent variable is whether the candidate was successful in any of the three exams that were notified following the 2013 Group 4 exam. "Additional Attempts" measures the number of additional attempts made in the same time period. In the second column, I instrument for the endogenous variables using the score residual. See equation (10) for the specification. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Appendix

Table of Contents

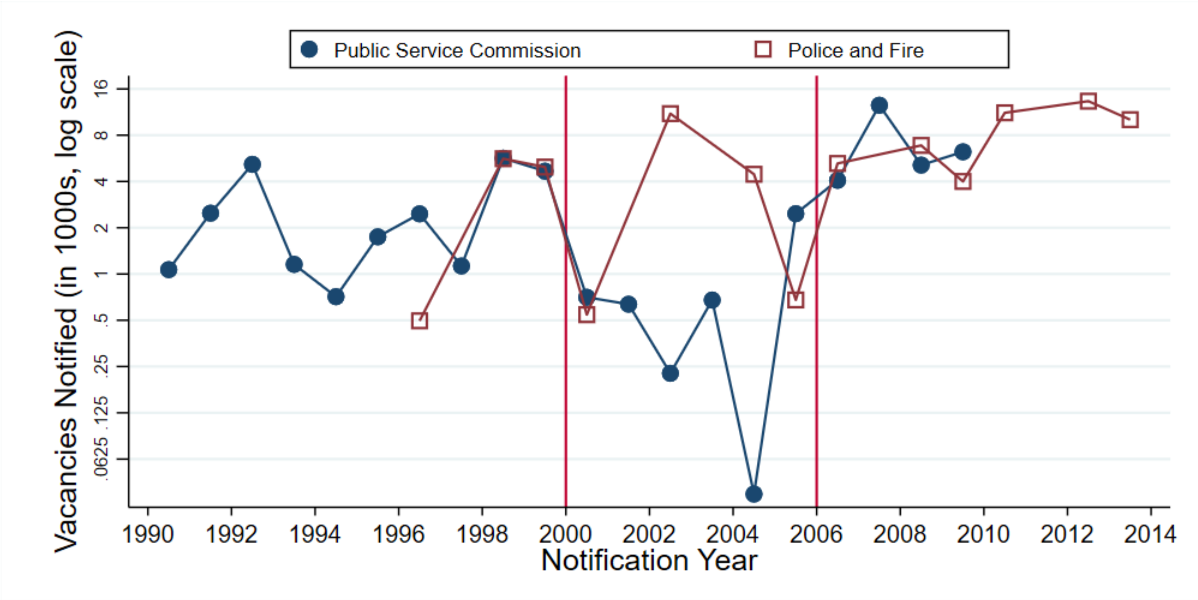
A Additional Figures and Tables

Figure A.1: Indian Youth Career Aspirations

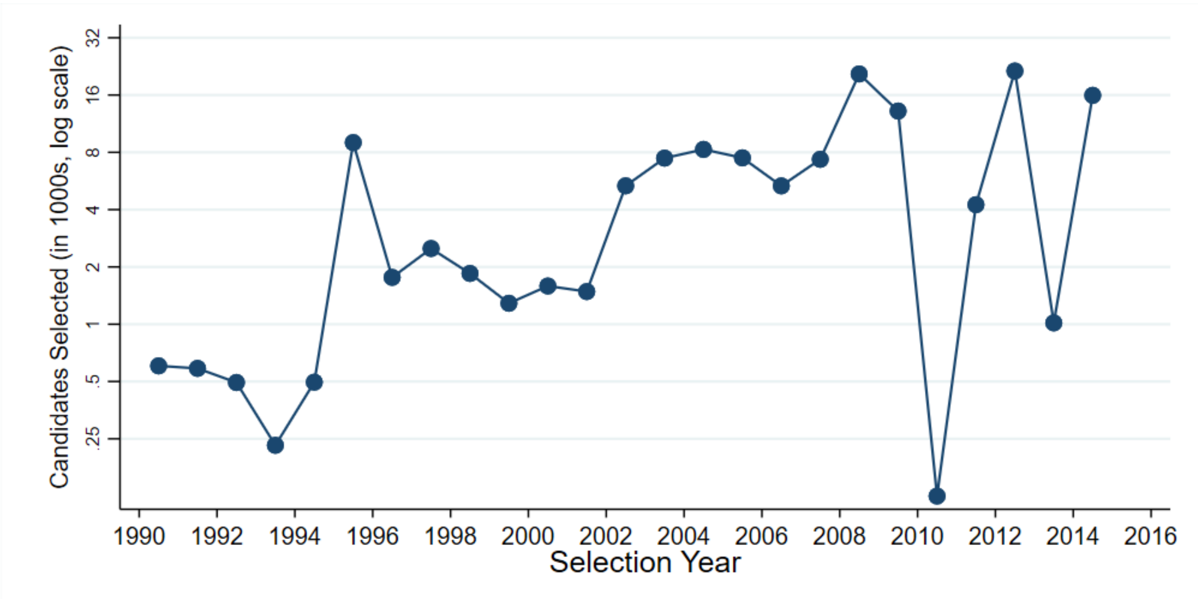


Notes: Data comes from the Lokniti-CSDS-KAS Youth Survey, conducted in 2007 and 2016. The survey was conducted with a representative sample of 5,513 and 6,122 individuals, respectively. This graph reproduces the first figure of Kumar and Gupta (2018).

Figure A.2: Hiring Outside of TNPSC is Relatively Unaffected



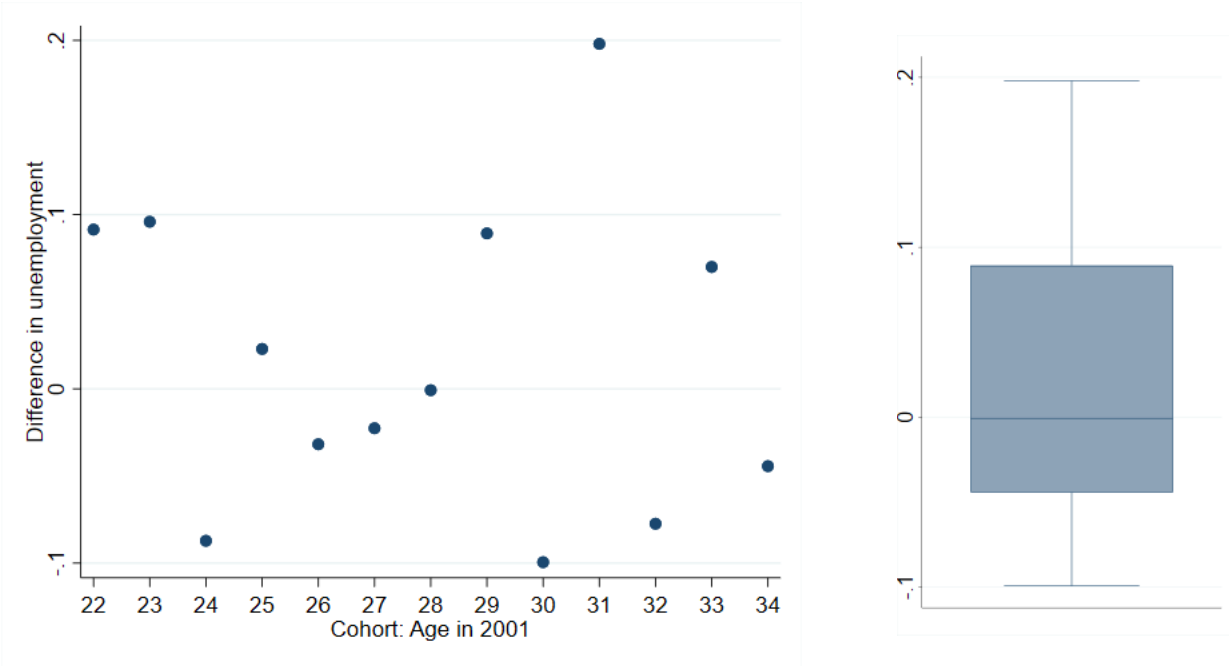
(a) Uniformed Services Board



(b) Teacher Recruitment Board

Notes: Data for the top figure provided by the Tamil Nadu Uniformed Services Board. Data for the bottom figure provided by the Tamil Nadu Teacher Recruitment Board. A caveat in the bottom figure is that it shows the number of teachers recruited by the year in which the exam was completed, not the year in which the post was advertised.

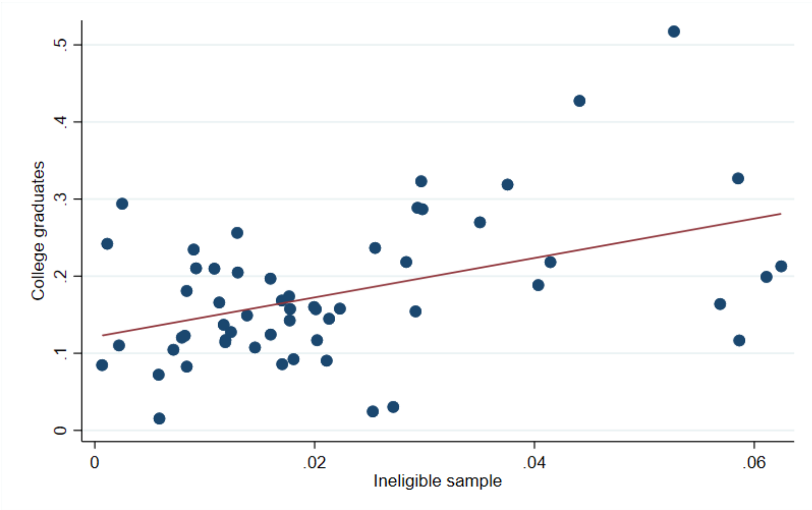
Figure A.3: Outliers in Outcomes Measured Before the Hiring Freeze



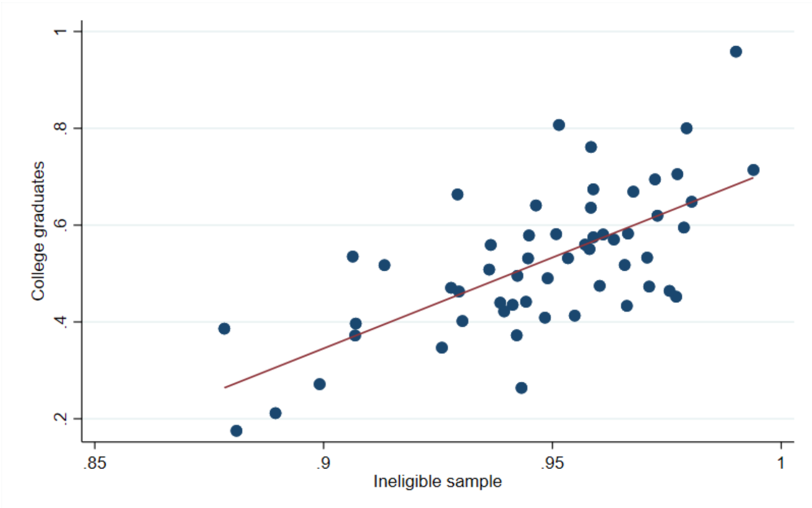
Notes: Figure plots the average difference in unemployment in the main analysis sample between Tamil Nadu and the rest of India for each cohort whose outcomes were measured before the hiring freeze, i.e. before 2001. The main analysis sample is: 1) men; 2) who are college graduates; 3) between the ages of 21 to 27 at the time of the survey.

Figure A.4: Employment status is correlated between the college-educated and ineligible samples across states and years

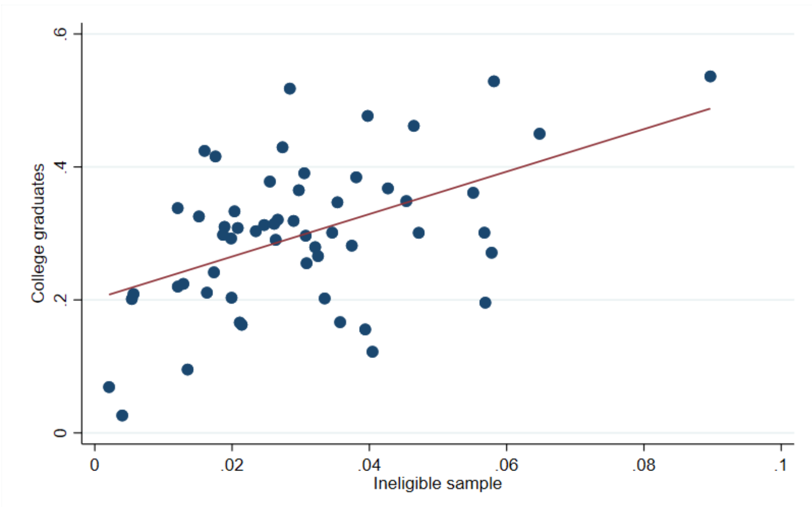
Unemployed



Employed

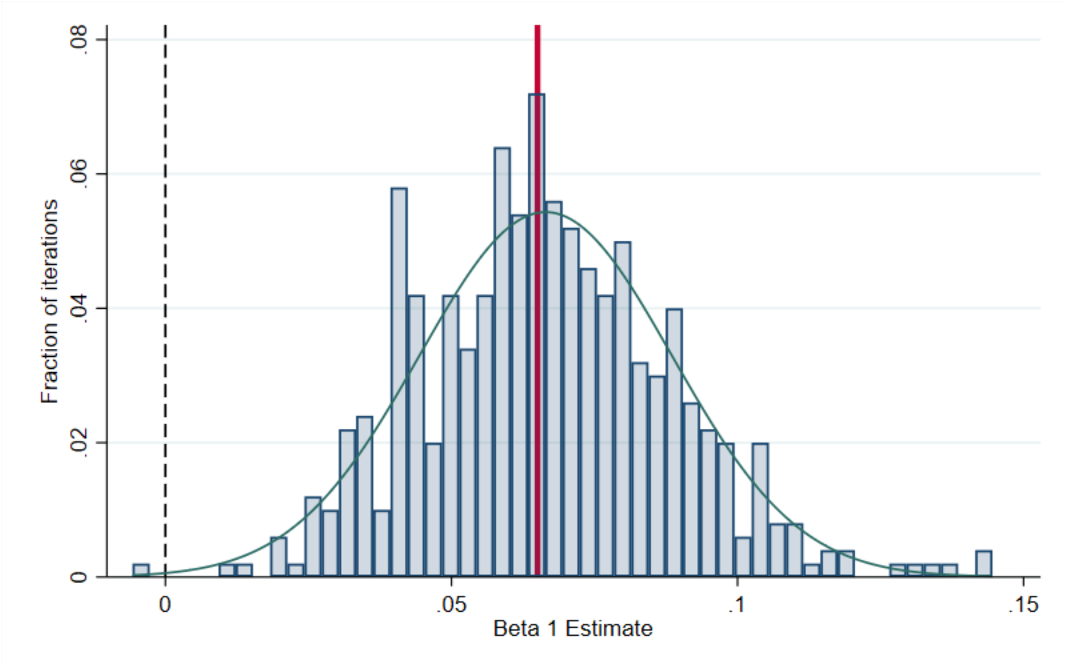


Out of the labor force

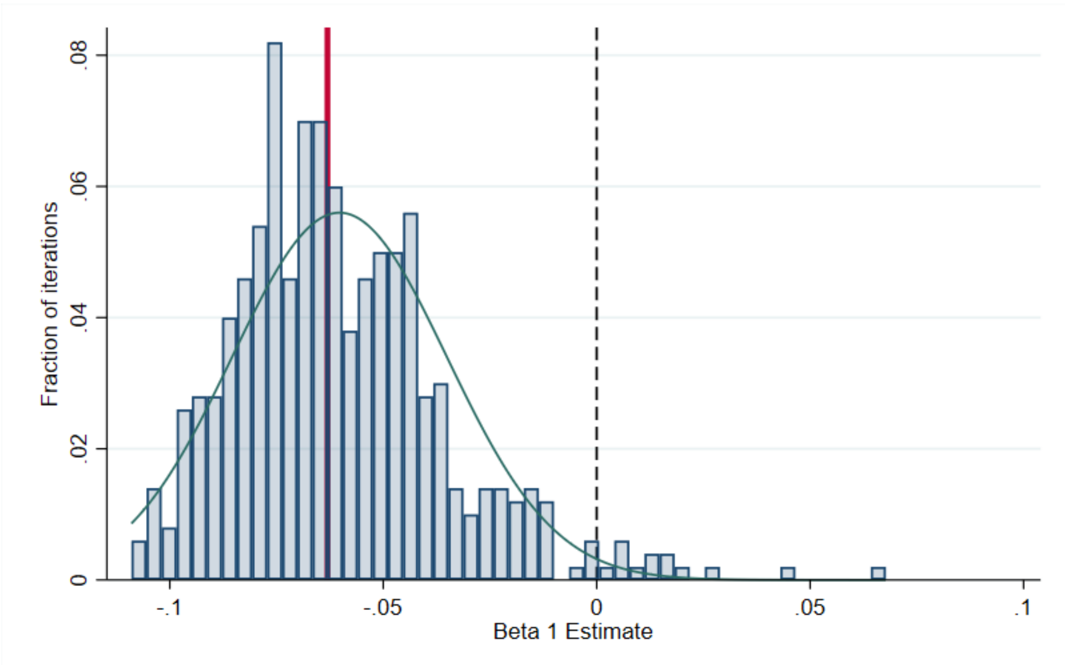


Notes: The figure uses data from 7 rounds of the National Sample Survey (NSS), collected between 1993 and 2010. Sample restricted to: 1) men between the ages of 21 to 27 at the time of the survey; 2) states with at least 750 observations in the college-educated sample; 3) not Tamil Nadu. Employment status is defined according to the NSS's Usual Principal Status definition. Each observation is a state-year. The x-axis plots the mean of the employment outcome for the ineligible sample, i.e. those with less than a 10th standard education. The y-axis plots the mean for the college-educated sample. The red line plots the regression line.

Figure A.5: Estimates of the Short-Run Impact of the Hiring Freeze are not Sensitive to the Choice of Comparison States



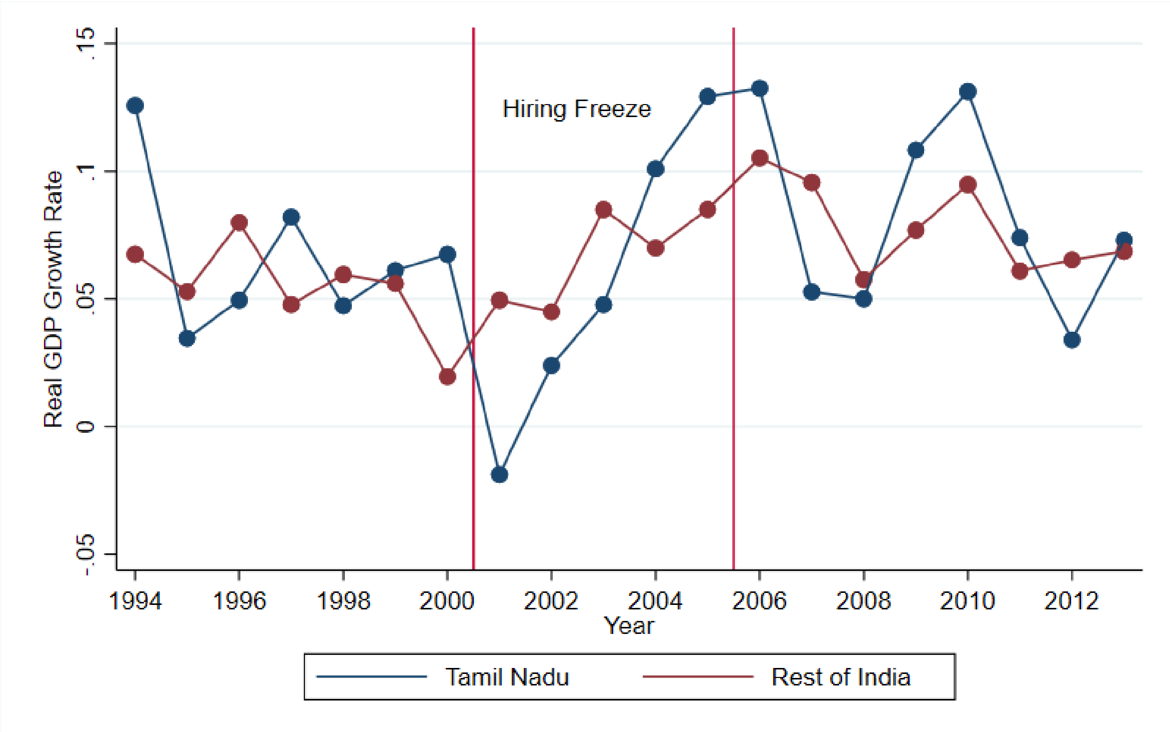
(a) Outcome: Unemployed



(b) Outcome: Employed

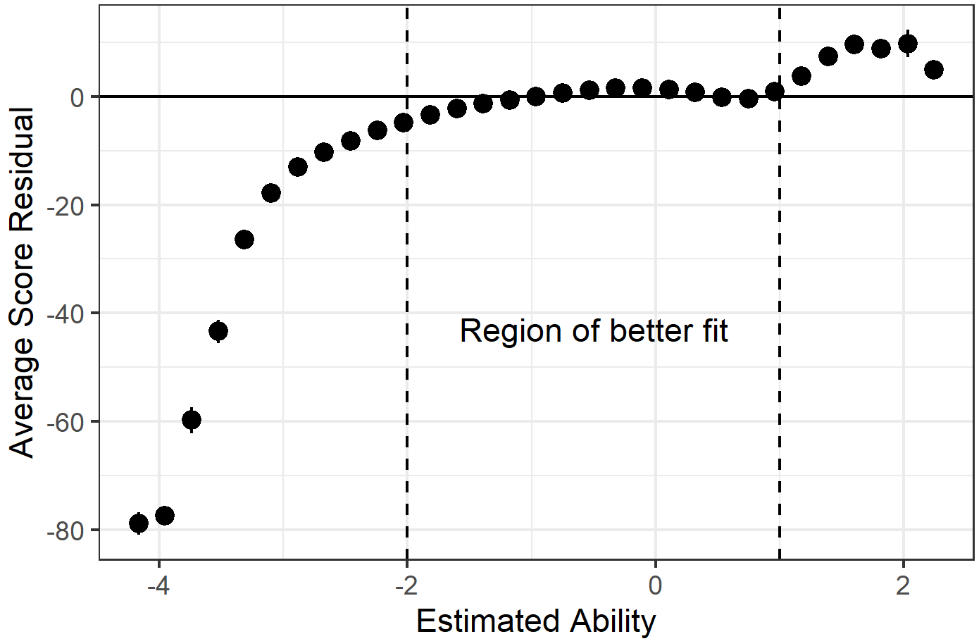
Notes: I randomly sample 10 states from the set of 30 available comparison states. In each of 500 iterations, I re-estimate equation (1) using only the sampled comparison states and Tamil Nadu. The figures plot histograms of the estimates of β_1; in the top panel, the outcome variable is unemployment, and in the bottom panel it is employment. A normal distribution is super-imposed. The thick red line marks the estimate from Table 2. The dashed black line marks zero.

Figure A.6: Comparison of the GDP Growth Rate in Tamil Nadu and the Rest of India



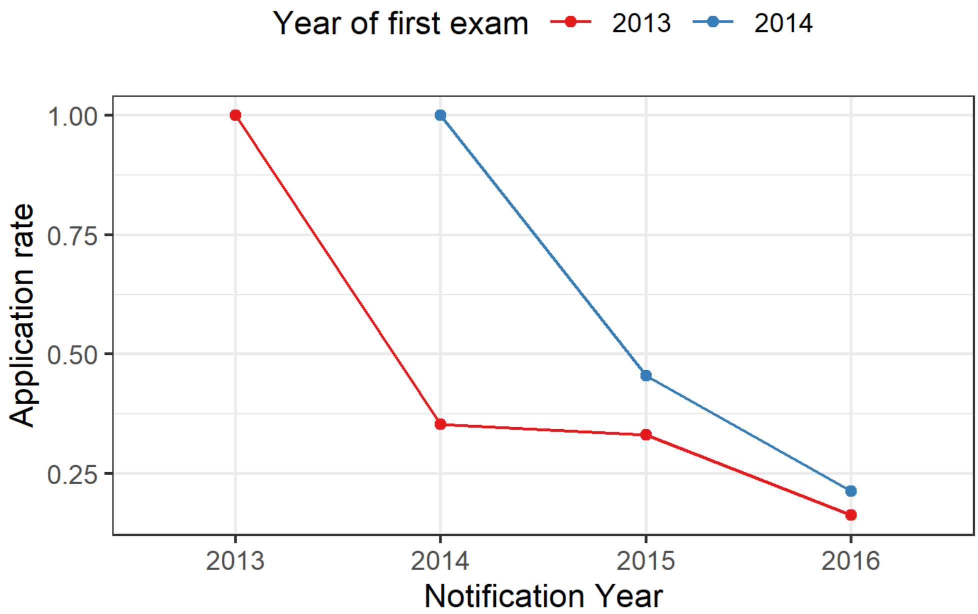
Notes: Data sourced from the website of Niti Aayog. For this time period, the government calculates three different series: the 1993-1994 series, the 1999-2000 series, and the 2004-2005 series. In some cases these series overlap, in which case I average the estimates of the growth rates across the series.

Figure A.7: Assessing the Fit of the IRT Model



Data: Administrative data from the Tamil Nadu Public Service Commission, 2013/09 exam
Notes: The figure presents estimates of the average score residual conditional on estimated the estimated ability parameter. If the model is correctly specified, then the average residual should be zero across the distribution. The dashed lines at -2 and 1 demarcate the boundary of the region where the model has a better fit.

Figure A.8: Most candidates drop out after their first attempts



Data: Administrative data from the Tamil Nadu Public Service Commission
Notes: The figure presents re-application rates over time. The sample is restricted to male college graduates were between the ages of 20 to 22 in the year in which they are likely to have made their first attempt. The year of the first attempt is inferred by restricting the sample to individuals who did not make an attempt in any prior year. The y-axis plots the fraction of applicants in the cohort who applied to any exam that was notified in the given year.

Table A.1: Sample Size in Tamil Nadu Cohorts by Education Level, National Sample Survey

Age in 2001	College Graduates	Ineligible Sample
17	122	489
18	125	539
19	134	468
20	140	402
21	115	349
22	122	591
23	89	350
24	52	275
25	48	160
26	40	155
27	49	266
28	47	417
29	26	179
30	29	202

Notes: Dataset combines all rounds of the National Sample Survey conducted between 1994 and 2010 that included a module on employment. This includes the 50th, 55th, 60th, 61st, 62nd, 64th, and 66th rounds. I impose the same sample restrictions that I use in the main analysis in Section 3, namely: 1) men; 2) between the ages of 21 and 27 at the time of the survey; 3) between the ages of 17 to 30 in 2001. Sample further restricted to Tamil Nadu. Ineligible sample refers to individuals with less than a 10th standard education.

Table A.2: Coverage Rate of 95% Confidence Intervals for Main Specification

	Parameter	
Inference Method	β_1	β_2
Stata Clustered SE	0.856	0.784
Wild Bootstrap	0.928	0.93

Notes: Table reports the results of simulations that test the coverage rate of different inference methods for the data and main specification used in Section 3. In each of 500 iterations, the outcome variable is changed to a new draw of a Bernoulii random variable that is i.i.d. across observations with a mean of 0.5. The coverage rate measures the fraction of confidence intervals that contain zero.

Table A.3: Tamil Nadu vs. the Rest of India Before the Hiring Freeze

	(1) Unemployment	(2) Employment	(3) Out of Labor Force
Tamil Nadu	-0.002	0.038	-0.036
	(0.036)	(0.044)	(0.038)
Mean, rest of India	0.213	0.453	0.334
Observations	5,087	5,087	5,087

Notes: Sample restricted to: 1) men; 2) who are college graduates; 3) between the ages of 21 to 27 at the time of the survey; 4) who were less than 30 years of age in 2001; and 5) whose outcomes were measured before 2001. Standard errors clustered at the year x NSS primary sampling unit level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: College Completion Rates Across Cohorts by Sex

	(1)	(2)
TN $\times$ Age 17-21 in 2001, Post-freeze (β_1)	0.012	0.023**
	[-0.007, 0.030]	[0.006, 0.038]
TN $\times$ Age 22-24 in 2001, Post-freeze (β_2)	0.022	0.018
	[-0.020, 0.061]	[-0.008, 0.044]
Sample	Men	Women
Mean, reference group in TN	0.072	0.064
Observations	167,722	163,784

Notes: 95% confidence intervals in brackets, computed via wild bootstrap with 999 replications, clustered by state x cohort level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Short-Run Impacts of the Hiring Freeze, Comparison States Restricted to South India

	(1) Unemployed	(2) Employed	(3) Out of labor force
Panel A: Diff-in-diff estimates, college sample			
TN $\times$ Age 17-21 in 2001, Post-freeze (β_1)	0.064	-0.051	-0.013
	[-0.021, 0.168]	[-0.138, 0.038]	[-0.124, 0.084]
TN $\times$ Age 22-24 in 2001, Post-freeze (β_2)	-0.043	0.073	-0.030
	[-0.254, 0.183]	[-0.223, 0.317]	[-0.156, 0.096]
Mean, reference group in TN	0.211	0.491	0.298
Observations	3,832	3,832	3,832
Panel B: Diff-in-diff estimates, ineligible sample			
TN $\times$ Age 17-21 in 2001, Post-freeze ($\tilde{\beta}_1$)	0.000	-0.001	0.000
	[-0.019, 0.019]	[-0.032, 0.027]	[-0.017, 0.020]
TN $\times$ Age 22-24 in 2001, Post-freeze ($\tilde{\beta}_2$)	-0.002	0.007	-0.005
	[-0.025, 0.021]	[-0.029, 0.044]	[-0.043, 0.026]
Mean, reference group in TN	0.033	0.930	0.037
Observations	17,461	17,461	17,461
Panel C: Triple difference estimates			
$\beta_1 - \tilde{\beta}_1$	0.063	-0.051	-0.013
	[-0.020, 0.157]	[-0.146, 0.045]	[-0.127, 0.092]
$\beta_2 - \tilde{\beta}_2$	-0.041	0.066	-0.025
	[-0.254, 0.169]	[-0.202, 0.329]	[-0.164, 0.132]
Observations	21,293	21,293	21,293

Notes: 95% confidence intervals in brackets, computed via wild bootstrap with 999 replications, clustered by state x cohort level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: Short-Run Impacts of the Hiring Freeze, Comparison States Restricted to Large States

	(1) Unemployed	(2) Employed	(3) Out of labor force
Panel A: Diff-in-diff estimates, college sample			
TN × Age 17-21 in 2001, Post-freeze (β_1)	0.072**	-0.075**	0.003
	[0.017, 0.134]	[-0.139, -0.014]	[-0.081, 0.075]
TN × Age 22-24 in 2001, Post-freeze (β_2)	-0.023	0.059	-0.036
	[-0.183, 0.109]	[-0.084, 0.249]	[-0.151, 0.081]
Mean, reference group in TN	0.211	0.491	0.298
Observations	15,417	15,417	15,417
Panel B: Diff-in-diff estimates, ineligible sample			
TN × Age 17-21 in 2001, Post-freeze ($\tilde{\beta}_1$)	0.072**	-0.075**	0.003
	[0.015, 0.129]	[-0.138, -0.013]	[-0.081, 0.074]
TN × Age 22-24 in 2001, Post-freeze ($\tilde{\beta}_2$)	-0.023	0.059	-0.036
	[-0.200, 0.089]	[-0.081, 0.238]	[-0.156, 0.083]
Mean, reference group in TN	0.211	0.491	0.298
Observations	15,417	15,417	15,417
Panel C: Triple difference estimates			
$\beta_1 - \tilde{\beta}_1$	0.073**	-0.080**	0.008
	[0.017, 0.136]	[-0.156, -0.006]	[-0.082, 0.081]
$\beta_2 - \tilde{\beta}_2$	-0.019	0.052	-0.032
	[-0.176, 0.106]	[-0.091, 0.234]	[-0.147, 0.102]
Observations	87,865	87,865	87,865

Notes: 95% confidence intervals in brackets, computed via wild bootstrap with 999 replications, clustered by state x cohort level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.7: Short-Run Impacts of the Hiring Freeze, Dropping Caste and Religion Controls

	(1) Unemployed	(2) Employed	(3) Out of labor force
Panel A: Diff-in-diff estimates, college sample			
TN × Age 17-21 in 2001, Post-freeze (β_1)	0.060**	-0.057*	-0.003
	[0.004, 0.122]	[-0.125, 0.008]	[-0.088, 0.072]
TN × Age 22-24 in 2001, Post-freeze (β_2)	-0.028	0.064	-0.035
	[-0.188, 0.088]	[-0.084, 0.253]	[-0.158, 0.084]
Mean, reference group in TN	0.211	0.491	0.298
Observations	19,303	19,303	19,303
Panel B: Diff-in-diff estimates, ineligible sample			
TN × Age 17-21 in 2001, Post-freeze ($\tilde{\beta}_1$)	-0.002	0.005	-0.004
	[-0.022, 0.018]	[-0.026, 0.036]	[-0.027, 0.022]
TN × Age 22-24 in 2001, Post-freeze ($\tilde{\beta}_2$)	-0.004	0.007	-0.003
	[-0.041, 0.025]	[-0.022, 0.038]	[-0.045, 0.036]
Mean, reference group in TN	0.033	0.930	0.037
Observations	90,300	90,300	90,300
Panel C: Triple difference estimates			
$\beta_1 - \tilde{\beta}_1$	0.062**	-0.063	0.001
	[0.001, 0.130]	[-0.144, 0.028]	[-0.084, 0.077]
$\beta_2 - \tilde{\beta}_2$	-0.024	0.057	-0.033
	[-0.202, 0.093]	[-0.110, 0.234]	[-0.149, 0.107]
Observations	109,603	109,603	109,603

Notes: 95% confidence intervals in brackets, computed via wild bootstrap with 999 replications, clustered by state x cohort level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.8: Short-Run Impacts of the Hiring Freeze on Wages

	(1) Log Weekly earnings	(2) Log Average daily wage
TN $\times$ Age 17-21 in 2001, Post-freeze (β_1)	0.249*	0.234*
	[-0.052, 0.599]	[-0.063, 0.535]
TN $\times$ Age 22-24 in 2001, Post-freeze (β_2)	0.613	0.613
	[-0.415, 1.181]	[-0.446, 1.134]
Mean, reference group in TN	6.291	4.398
Observations	4,818	4,818

Notes: 95% confidence intervals in brackets, computed via wild bootstrap with 999 replications, clustered by state x cohort level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.9: Candidates with no prior experience with the exam have more positively biased beliefs about exam performance

	(1) Bias	(2) Bias
No prior experience	10.180**	14.911***
	(3.241)	(3.408)
Has taken practice test in prior 8 months		0.945
		(3.735)
Any nuclear family in government job		6.162
		(14.978)
Any family in government job		-0.640
		(15.097)
Age		0.621
		(0.494)
Average bias, candidates with prior experience	29.1	28.8
Library fixed effects	Yes	Yes
R^2	0.037	0.103
Observations	88	85

Data Source: Survey data from Pune, Maharashtra.
Notes: Bias is the difference between a candidate's predicted score on a practice test and their actual score. Positive values correspond to over-estimates. "No prior experience" is an indicator for whether the candidate reported appearing for a Maharashtra state or Union Public Service Commission exam in the past 4 years. Specification includes fixed effects for the library in which the candidate was surveyed. Heteroskedasticity-robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

B Estimating the Direct Demand Effect

Summary. In this section I assess the size of the effect that the hiring freeze may have had arising just from the government's reduced expenditure in the labor market. I term this effect the *demand effect*. I estimate that this effect is small, and an order of magnitude smaller than the shifts in labor market equilibrium that we observe.

Estimation Strategy.

1. How many vacancies were lost as a result of the hiring freeze?

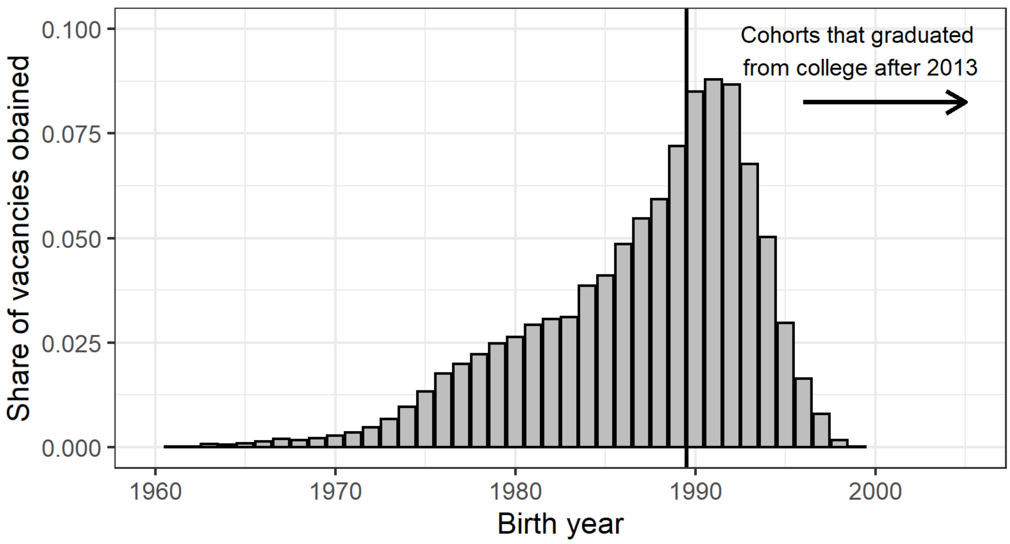
   Because the hiring freeze order exempted teachers and police, I assume all losses were incurred in TNPSC. Estimating the loss in vacancies in TNPSC requires making assumptions about the data generating process. I assume that, in the absence of the freeze, vacancies in a given year are generated according to $\log(vacancy)_t = \mu + \epsilon_t$, where $E[\epsilon_t] = 0$. This assumption is roughly consistent with the observed variation in vacancies in Figure 1.

   I then estimate a regression of the form $\log(vacancy)_t = \alpha + \beta freeze_t + \epsilon_t$, where $freeze_t = 1$ for $2001 \geq t \geq 2005$. This model implies that loss in the number of vacancies *per year* can be estimated as $\exp(\hat{\alpha}) - \exp(\hat{\alpha} + \hat{\beta})) = 2422$. Over the 5 years of the hiring freeze, the estimated loss in vacancies is: 12110.

2. How many vacancies were lost by each cohort?

   The overall loss in vacancies is distributed across cohorts. Using TNPSC administrative data from 2012-2016, in Figure B.1 I plot the fraction of all available posts that accrue to each cohort.

   Figure B.1: Fraction of vacancies accruing to each cohort

   We see that no cohort captures more than 8.2% of the available vacancies over 5 years. Because my analysis focuses on male college graduates, we would also want to know what share of that 8.2% is captured by them. I find that it's about 60%. I thus estimate the loss in vacancies to individual cohorts of male college graduates to be at most:

$$12,110 \times 0.082 \times 0.6 = 595$$

3. How does the loss in vacancies compare to the size of the labor force in each cohort?

   The 2011 Census indicates that were 484,027 male college graduates between the ages of 30-34. This is the age category that is closest to the group on which I focus my analysis (i.e. recent college graduates in 2001 would be ≈ 21). This tells us that there were about 484,027 / 5 = 96,805 male college graduates in each individual cohort.

   Thus, the reduction in vacancies means that about

   $$595/96,805 = 0.006$$

   or 0.6% of the most affected cohorts were delayed in getting or did not get a government job. Note, this is an upper bound, assuming a cohort lost 5 years worth of vacancies. Among the cohorts that were expected to graduate from college during the hiring freeze (the focus of the analysis), cohorts lost between 1 to 5 years of vacancies, so the average affect across this group is on the order of 0.3%.

4. What is the average loss in income for each vacancy?

   The estimate of the public sector wage premium in Finan et al. (2017) is 71.2 log points (Table 1, Column 3). Thus the earnings effect should be about

   $$71.2 \times 0.006 = 0.43 \text{ log points}$$

   which is also about 0.43%.

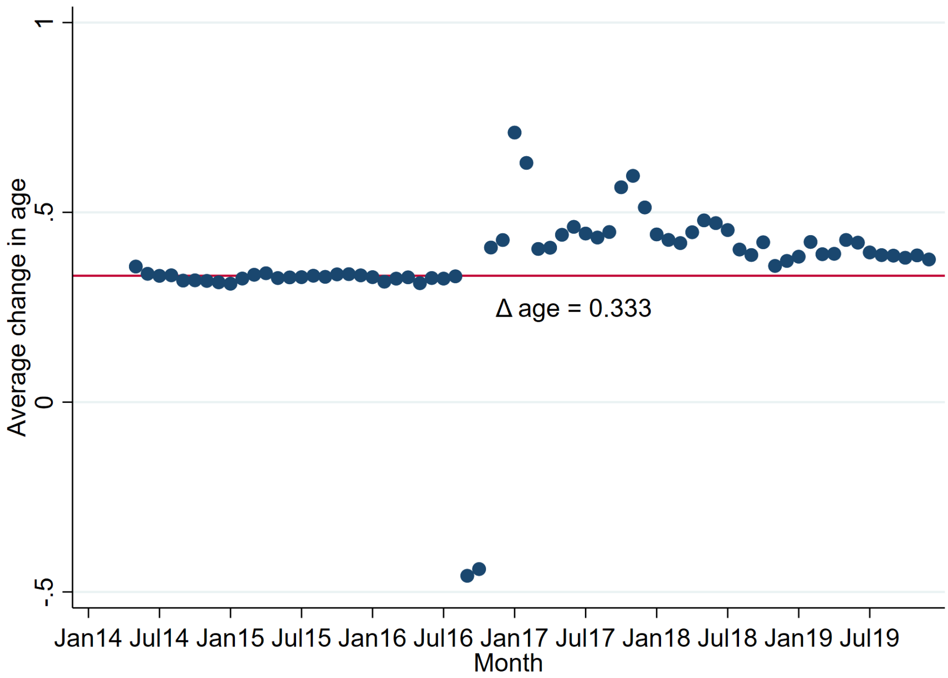
C Measurement Error in Age in the CMIE

Summary. Age is a critical variable in the analysis in Section 4, since it defines which individuals belong to which cohort. In this appendix, I first present evidence that there is substantial measurement error in this variable after September 2016. I then discuss the imputation procedure that I use to adjust for this error.

Evidence of measurement error. In each wave of the survey, CMIE captures the age of each household member. This allows me to track how the age of each individual in the sample evolves over the course of the panel. Since birthdays are roughly uniformly distributed, and since CMIE conducts three survey waves per year, then roughly one-third of the sample should complete a birthday between each wave.

To check whether this is the case, I compute, for the sample collected in each month, the average difference in age for each respondent from the previous wave. These results are presented in Figure C.1 below. The red line marks 1/3, which is where the average should lie if measurement error is *on average* close to zero. It appears this is the case until September 2016. In October and November of 2016, age increments too slowly; thereafter, the age increments too fast. I'll refer to the period from January 2014 until September 2016 as the "Good Period," since the measurement error appears to be zero on average in this time frame.

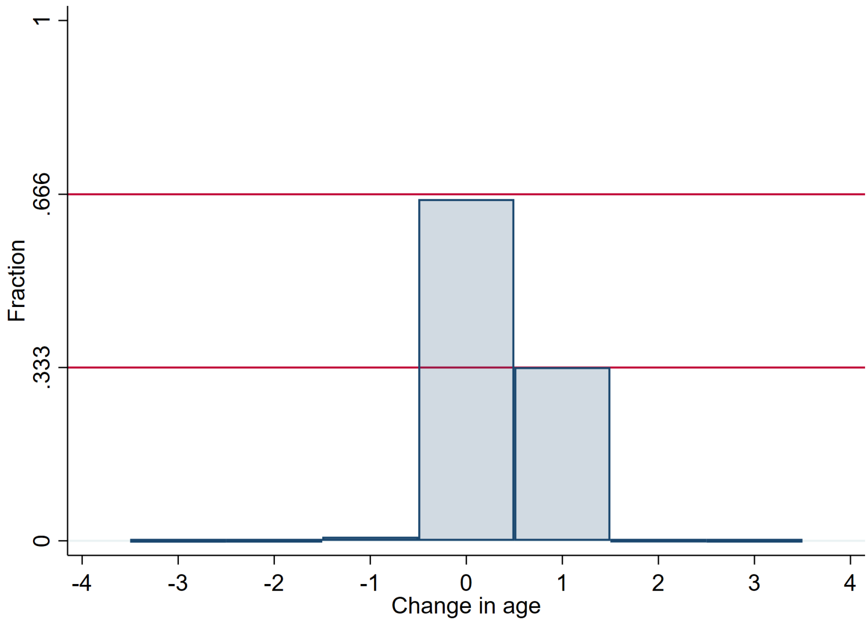
Figure C.1: Average change in age between waves



Characterizing measurement error during the Good Period. For our purposes, even if measurement error is zero on average across the whole sample, we may still be concerned about measurement error at the individual level. In particular, we might worry that: 1) the size of the measurement error is still substantial for individuals; 2) measurement error is correlated with age; and 3) errors are serially correlated. I present evidence that suggests that none of these concerns apply.

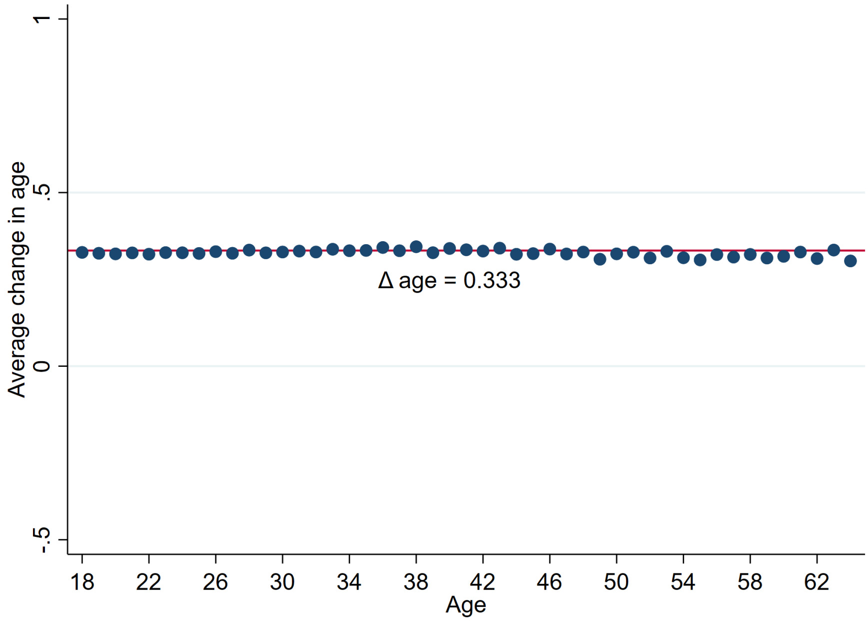
First, given that age is measured in whole numbers, if individual age is correctly measured, then we should see that about 2/3 of the sample has the same age across waves, and 1/3 of the sample increments by 1. In Figure C.2 below, I confirm that this is the case. Only about 1% of observations due not fit into this expected pattern.

Figure C.2: Average change in age between waves



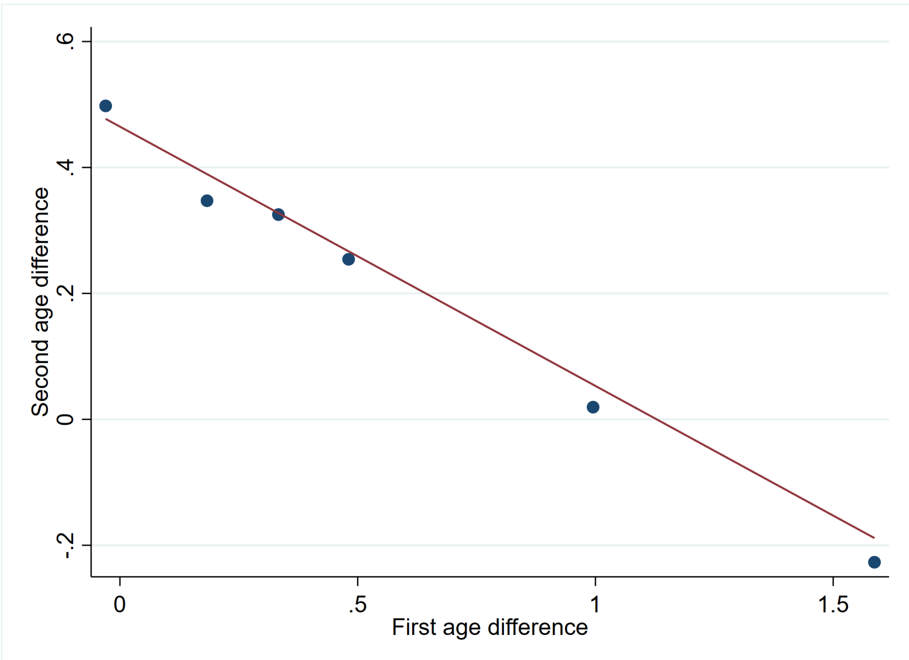
Second, I test whether some positive errors in some age groups cancel out negative errors for other age groups. This does not appear to be the case either. To illustrate, the figure below plots a candidate's age in a given wave on the X-axis (Age_t), and the average change in age between subsequent waves on the Y-axis (i.e. $E[Age_{t+1} - Age_t]$). We see that people of all age groups seem to be aging at the correct speed.

Figure C.3: Average change in age between waves



Lastly, to test for serial correlation in the errors, I correlate the difference in ages between two successive waves ($Age_{t+1} - Age_t$) with their lagged equivalents ($Age_t - Age_{t-1}$). The strong negative correlation we see in Figure C.4 is consistent with mean reversion, which is what one would expect if the errors were serially uncorrelated.

Figure C.4: Average change in age between waves



Imputation. One possible solution to the measurement error problem is to restrict the analysis the Good Period, i.e. observations collected before October 2016. The advantage of this approach is that it imposes minimal assumptions on the structure of the measurement error. However, the lack of assumptions imposes heavy costs: 1) restricting the sample in this way would result in a loss in 54% of the available observations; and 2) it misses an opportunity to reduce measurement error at the individual level by exploiting the panel structure of the data.

An alternative approach—and the one I use in the paper—is to find, for each individual, a sequence of ages that increments correctly according to calendar time and fits the data best. To fit the data, we have to have some notion of the structure of the measurement error. Based on the evidence in the previous section, I assume that, *for each individual*, the measurement error is zero on average. I can therefore compute the best fitting age series by minimizing quadratic loss.

Formally: I suppose that in each wave of the survey, an individual's true age is given by a vector $\mathbf{a}_i = (a_{i1}, a_{i2}, \ldots, a_{iT})$. Due to measurement error, in each wave we only observe $\hat{a}_{it} = a_{it} + \epsilon_{it}$. The evidence presented above suggests that $E[\epsilon_{it}] = 0$ in the Good Period. If this assumption holds, then the true age minimizes a quadratic loss function. Thus, within the Good Period, we can calculated an imputed age series $\bar{\mathbf{a}}_i$ as follows:

$$\bar{\mathbf{a}}_i = \arg\min_{\mathbf{a}_i} \frac{1}{T} \sum_{t=1}^{T} (\hat{a}_{it} - a_{it})^2 \qquad s.t.\ \forall t, \quad a_{i,t+1} - a_{i,t} = 1/3 \tag{C.1}$$

I implement this algorithm by computing, for each individual, many different age streams. I take each observed age and then add a perturbation $\Delta \in \{-11/12, -10/12, \ldots, 10/12, 11/12\}$. I then compute the calendar-consistent age stream using that age in the observed month as a starting point. In other words, given T data points for an individual, I compute $23T$ plausible age streams. For each of these potential age streams, I then choose the one that minimizes quadratic loss.

Once I have imputed age, I can extrapolate into the Bad Period by adding 1/3 to the last imputed age for each additional wave, i.e. if I observe T observations in the Good Period, then the imputed ages in the bad period will be $\bar{a}_{it} = (1/3)(t - T) + \bar{a}_{iT}$. I drop from the analysis any individuals that were not surveyed during the Good Period.

D Additional Discussion on Mechanisms

Summary. This section provides more detailed discussions on points raised in Section 5 of the main paper.

Alternative mechanisms not considered in the main text

I will use the following general theoretical framework to summarize the range of mechanisms that could be operative in this setting.

Think of the decision to continue studying as a dynamic discrete choice problem. Individuals decide whether to continue on the "exam track" or switch to the "private job" track. They stay on the exam track as long as:

$$p_t g + V^g_{t+1} > w_t + V^p_{t+1} - c_t \tag{D.1}$$

where

- p_t is the probability of obtaining a government job in period t
- g is the value of a government job
- V^g_{t+1} is the continuation value of staying on the exam track
- w_t is the value of the private job in period t
- V^p_{t+1} is the continuation value of the private job track
- c_t is the cost of switching tracks

The hiring freeze unambiguously reduces p_t, so if candidates stay on the exam track it must be because either: 1) the hiring freeze simultaneously raises V^g_{t+1}; or 2) switching costs are very large.

In the main text, I discuss a mechanism that raises V^g_{t+1}: a lack of testing implies that candidates over-estimate the probability of success in the future for longer than they otherwise would.

However, it is possible that there are other reasons why V^g_{t+1} would increase. One class of explanations relates to beliefs about future vacancy availability. For example, candidates may have thought that the government would compensate for the hiring freeze by increasing vacancy availability in the future. Alternatively, candidates may believe that if enough of them stayed on the exam track, then the government would feel pressure to provide more vacancies at the end of the freeze and/or end the hiring freeze sooner.

If switching costs are large, it is likely not because searching for a private sector job is expected to take years (in fact, for many candidates, the outside option is likely to be self-employment). Instead, it seems more likely that these costs are psychological or social. Candidates may have a hard time giving up on their dreams, or "admitting defeat" to their friends and family.

Why the convexity of the returns to exam preparation matters

This section provides a formal model of the proposition that candidates will have an incentive to study *during* the hiring freeze if the returns to exam preparation are convex.

Let us examine a situation in which two identical candidates (indexed by A and B) need to make a decision about whether to study during the hiring freeze. Both candidates know that the hiring freeze will last t_1 years and that the time between the end of the hiring freeze and the resumption of exams will last t_2 years.

Test scores are a function of the time spent preparing for the exam, plus an error term, i.e. $T_i = h(s_i) + \epsilon_i$, where s_i is the total time spent preparing by candidate i. The cost per unit of time spent studying is c. The value of the government job is g.

Candidates are not able to coordinate their decisions with each other. By assumption, both candidates find it valuable to study at least after the vacancies are announced. In that case both candidates will study for $s_i = t_2$ years, and have the same average score. The winner will be determined by who obtains the larger shock to their score. Write $F(x) = Pr(\epsilon_B - \epsilon_A \leq x)$. This implies that for candidate A the payoff to both studying after the vacancies are announced is $F(0)g - ct_2$.

Depending on the returns to studying, candidates may have an incentive to "deviate" and study during the hiring freeze as well. It's worth deviating as long as

$$\big(F(h(t_1 + t_2) - h(t_2))\big)g - c(t_1 + t_2) > F(0)g - ct_2 \tag{D.2}$$

which is equivalent to

$$\frac{P(t_1) - P(0)}{t_1} > c/g \tag{D.3}$$

where $P(x) \equiv F(h(x + t_2) - h(t_2))$. Note that P captures the marginal returns to study effort. The more convex is P, the larger the term on the left hand side will be.