The World Bank Economic Review, 00(0), 2024, 1–23
https://doi.org10.1093/wber/lhad047
Article

OXFORD

How Much Are Government Jobs in Developing Countries Worth?

Kunal Mangal

Abstract

Government jobs in developing countries are valuable not just because they pay relatively higher wages, but also because they provide many valuable amenities. How does the value of these amenities compare with the nominal wage itself? The observed search behavior of candidates preparing for competitive exams for government jobs is used to infer a lower bound on the total value of a government job, including amenities. Based on a sample of 147 candidates preparing for civil service exams in Pune, India, the amenity value of a government job is estimated to comprise at least two-thirds of total compensation. The high amenity value is not driven by misinformed beliefs about the nominal wage, nor by a high value placed on the process of studying itself. Insights from focus group discussions help explain which government job amenities are most valued in this setting.

JEL classification: J32, J45, J64, O17

Keywords: public sector, amenities, competitive exam

1. Introduction

How do public- and private-sector compensations compare? The answer to this question is an important policy parameter that determines which people seek public-sector jobs, how much they invest in obtaining them, and how much effort they expend once employed. However, a complete answer to this question for developing countries remains elusive. This is because government jobs in developing countries offer many amenities that are hard to price. For example, government employees in developing countries typically obtain lifetime job security, have ready access to bribe payments, and enjoy some measure of celebrity status. The extent to which the amenity value of public-sector jobs contributes to total compensation is an open question.

The absence of order-of-magnitude estimates of the value of public-sector job amenities is limiting. Because wages are the most visible component of total compensation, they are often the focus of research on public–private compensation differentials (Finan, Olken, and Pande 2017; Campos et al. 2017; Gindling

Kunal Mangal is an independent researcher. His email address is kmangal@alumni.harvard.edu. The Weiss Fund for Development Economics provided funding for data collection. The author is grateful to Robert Townsend, Dr. Anand Patil, Mrs. Vaishali Patil, Mahesh Bade, and Ram for their advice and support; to Kranti Mane and Deepak Palhal for sincere research assistance; to Nikita Kohli and Perdie Stilwell for help preparing the draft; to the editor, Eric Edmonds, and the three referees for constructive comments; and to the MPSC candidates who took the time to participate in the survey, despite their many stresses. The research protocol was approved by Harvard University's Institutional Review Board.

et al. 2020; Araujo 2020). But if the amenity value of government jobs is large, we may be ignoring the part of compensation that is most responsible for allocating individuals and effort.

The standard method of valuing job amenities uses variation in job characteristics within a candidate's choice set (Stern 2004; Mas and Pallais 2017). However, in developing countries, such comparisons are difficult to obtain. In these contexts, the value of public-sector jobs is likely far beyond what most candidates could realistically expect to obtain from the private sector. It is thus unlikely that for most candidates we would be able to observe or induce a choice set in which a private-sector offer competes with a public-sector offer.

This paper therefore develops an alternative strategy for estimating the value of public-sector jobs. The analysis focuses on India, which has one of the largest public-sector wage premia in the world (Finan, Olken, and Pande 2017). The empirical strategy makes use of the fact that in India most government jobs are allocated through a system of highly structured competitive exams. In a typical exam, the government receives several thousand applications for each vacancy. In order to remain competitive, many candidates spend years studying full time. The value candidates place on government jobs can be inferred from the amount of time they are willing to spend studying for these exams. Imposing some parametric structure on a model of exam preparation prices this time in monetary terms.

The model is estimated using data from a novel survey of 147 candidates preparing for a set of state-level civil service exams known as the Maharashtra Public Service Commission (MPSC) exams. The sample is drawn from four clusters in Pune, a city in western India in the state of Maharashtra. Given the small number of clusters, the reported standard errors only account for sampling uncertainty within clusters and treat the clusters themselves as fixed.

The survey targeted a neighborhood in Pune in which candidates from all over the state come to study. This is therefore a population of highly motivated applicants who are likely to value government jobs more than the average applicant. Focusing on this population helps us understand whether there is a link between the amenity value of government jobs and the heavy investments that some candidates make in exam preparation. This question is especially relevant for policymakers concerned about the social cost of candidates enduring periods of prolonged unemployment while preparing for government job exams.

The model is estimated in three distinct ways, each of which imposes different restrictions and assumptions on the data and the data-generating process. The estimates, though imprecise, consistently indicate that candidates in the sample value government jobs at a minimum of Rs. 250,000 per month.[1] By comparison, the annuity value of the nominal salary of a Tehsildar—one of the highest-paying jobs offered through these exams—is about Rs. 81,000 per month.[2] These estimates suggest that the amenity value of a government job is at least 66 percent of total compensation.

There are two potential alternative explanations for why individuals may appear to have a high value for government jobs. First, candidates may be misinformed about the salary structure. Second, candidates may derive utility from the process of studying itself, above and beyond the instrumental value of succeeding. However, upon further inspection, neither of these hypotheses can account for the large implied amenity value of government jobs.

Finally, focus-group discussions with candidates help shed light on which specific government job amenities are valuable. The discussions largely centered on the importance of how candidates' work is perceived in their villages, and the kinds of status, authority, and power that government officials have in rural India. These responses suggest that candidates' exposure to social hierarchies where government officials are at or near the top may have played an important role in socializing them to value government jobs.

1 All monetary figures are reported in nominal pre-COVID 2020 Indian Rupees (INR), which corresponds to the time period when the survey was conducted.

2 The annuity value calculation accounts for the schedule of salary increments that government employees can expect to obtain once in service.

Conceptually, this paper builds on a long literature in labor economics that uses queues for particular employment opportunities as evidence of rents (Krueger 1988; Holzer, Katz, and Krueger 1991). This paper takes that insight one step further to price the value of those rents in the public sector. For the private sector, estimating the value of rents from queuing behavior would require taking a perhaps unjustifiably specific stand on jobseekers' search behavior across firms. However, in this context, modeling this behavior is more feasible because the exam process is already highly structured.

This paper proceeds as follows. First, an optimal stopping model of exam preparation is presented. Next, the data that will be used to estimate the model is described. Next, the estimation strategy and results are presented. Next, alternative explanations for the high estimated value of government jobs are discussed. Next, findings from focus-group discussions are summarized to help explain the high amenity value of government jobs. Finally, the paper concludes with a discussion of the implications of a large amenity value of government jobs for personnel policy and future research.

2. A Model of Exam Preparation

Consider a model of exam preparation as an optimal stopping problem. The model incorporates several features specific to the context. Candidates maximize their expected lifetime earnings over a finite horizon. In each period, candidates decide whether to prepare for the exam or not. If yes, then they obtain a government job with some probability. If not, then they take their outside option in the private sector. A key prediction of this model is that for each candidate there should be an age at which they drop out of exam preparation and take up their outside option in the private sector. This dropout age is monotonically increasing in the value of a government job. The value of the government job is the unobserved model parameter that rationalizes the observed dropout behavior.

2.1. Setup

In each year t, an agent decides whether to study for a government job. If yes, then the agent is unemployed. If not, then the agent takes their outside offer in the private sector, which yields an annual income of w. Consistent with the context of this paper, search costs for private-sector jobs are treated as negligible, so agents do not need to spend time searching to obtain them.[3]

Candidates that are studying obtain a government job in the next period with probability p. The government job is worth w' per year. This term incorporates both the wage and amenity values of a government job, which are defined *relative* to the outside option. While studying, the agent receives income b through transfers from family members. Agents have a finite horizon T and discount the future at rate β.

The agent's search problem can be summarized with the following value functions. For $t = 0, 1, 2, \ldots, T-1$,

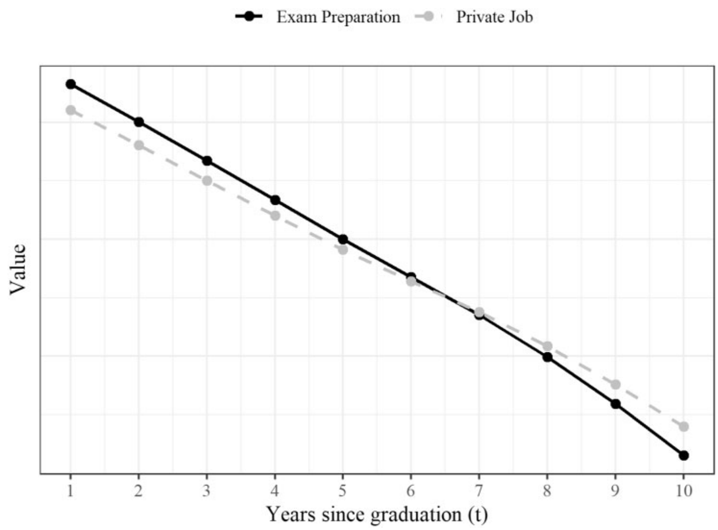
$$\begin{aligned} G_t &= u(w') + \beta \max\{G_{t+1}, P_{t+1}, U_{t+1}\}, \\ P_t &= u(w) + \beta \max\{P_{t+1}, U_{t+1}\}, \\ U_t &= u(b) + \beta[pG_{t+1} + (1-p)\max\{P_{t+1}, U_{t+1}\}], \end{aligned} \tag{1}$$

where G_t is the value of working in a government job at time t, P_t is the value of working in a private-sector job, and U_t is the value of preparing for the government job exam. In the final period, the value of each state is just the flow value, i.e. $G_T = u(w')$, $P_T = u(w)$, and $U_T = u(b)$.

The structure that the model imposes on intra-household decision-making is important. Specifically, the model assumes that household transfers are gifts conditional on preparing for the exam. This is a

3 For 75 percent of the sample, the outside option is either farming or business.

Figure 1. An Illustration of the Optimal Stopping Model.



Source: Illustration based on simulated data generated by the author.
Note: The figure illustrates how in an optimal stopping model of exam preparation, a candidate will eventually switch from exam preparation to private-sector employment if they are not successful in obtaining a government job. The figure plots the value functions for U_t (exam preparation) and P_t (private job) from the model described in equation (1) for a specific set of model parameters: $b = 3,000$, $w = 8,000$, $w' = 30,000$, $p = 0.085$, $\beta = 0.9$. Note that (a) at $t = 6$, the value of exam preparation no longer exceeds the value of a private job and (b) the value functions only cross once.

reasonable assumption in a setting where parents often feel obligated to support their children until they are settled in their careers. To the extent that there are other intra-household arrangements for these transfers, the interpretation of the parameters would need to adjust accordingly.[4]

2.2. Optimal Stopping

This model only has meaningful content when $w' > w > b$. If $w \leq b$ then a candidate would never give up preparing. If $w' \leq w$ then a candidate would never prepare for a government job in the first place. But when $w' > w > b$ the model sets up an interesting optimal stopping rule:

Proposition 1. *Someone who starts unemployed will eventually switch to private-sector work if not employed by the government, i.e. if $U_0 > P_0$, then $U_t < P_t$ for some t. Furthermore, private employment is an absorbing state, i.e. $U_t < P_t \Rightarrow U_{t+s} < P_{t+s}$ for all s.*

Proof. See the appendix.

Thus, for someone starting from unemployment, the optimal path is to keep trying for a government job, and if that doesn't work out, to switch to a private job at some time t^* for the remainder of their career. The switching point t^* will be referred to as the *dropout age*.

Figure 1 provides intuition for the dropout rule. The figure plots the value of unemployment and the private-sector job. Both are declining in time because of the finite time horizon. However, the value of unemployment declines faster than the value of a private-sector job. This is because the more time

4 In particular, if the transfer is a loan that parents expect students to pay back, then w' is measured net of unobserved loan repayments, i.e. the value of government job amenities will be *higher* than what the model otherwise suggests, in order to compensate for this extra unobserved cost. If, instead, parents and students are maximizing joint household income, then the direction of the bias will depend on several unmodeled factors, including the distribution of household wealth.

one spends in unemployment, the less time there is available to enjoy the government job, even if one is successful in obtaining it. Consequently, there is some point at which the value functions cross. This crossing point is the dropout age.

The dropout age can be expressed a function of the model parameters. It is the value t^* at which the value of a private job just equals the value of unemployment.

Proposition 2. *When $b < w < w'$, the optimal dropout age is given by*

$$t^* = \begin{cases} 0 & \textit{if } u(w) - u(b) \geq \dfrac{\beta(1-\beta^T)}{1-\beta} \\ & \quad \times\, p[u(w') - u(w)], \\ T - \dfrac{1}{\ln\beta}\ln\left[1 - \dfrac{(1-\beta)[u(w)-u(b)]}{\beta p[u(w')-u(w)]}\right] & \textit{otherwise.} \end{cases} \tag{2}$$

Proof. See the appendix.

Equation (2) is piecewise for the following reason. The government job has to be sufficiently valuable to make studying worthwhile. If the gain in utility from switching from unemployment to the private job in the first period is larger than the value of the possibility of obtaining a government job for all remaining periods, then there is no incentive to study.

3. The Peth Area Library Survey

The data for this analysis come from a survey fielded by the author in a neighborhood in the center of the city of Pune known as the Peth Area. Within the state of Maharashtra, Pune is well known as a hub for preparation for government job exams. Students from all over the state migrate to Pune to study. In particular, the 411030 zip code—the Peth Area—is the epicenter of exam preparation.[5] This zip code has a high concentration of both candidates preparing for competitive exams, and businesses that cater to their needs, including book shops, photocopy shops (which maintain ready catalogs of practice tests and study materials), coaching classes, libraries, canteens, and hostels.

3.1. Setting

In India, the wage component of public-sector compensation is relatively high. Finan, Olken, and Pande (2017) estimate the public-sector wage premium to be 105 percent (see their table 1, column 3). Compared to the 34 other countries in their sample, this estimate makes India an outlier—both in absolute terms, and relative to countries with comparable GDP per capita.

Most government-job aspirants in the Peth Area prepare for state-level civil service jobs. The exams for these jobs are conducted by the Maharashtra Public Service Commission (MPSC), a state-level government agency. In a typical year, candidates participate in 1 to 3 exams. These exams attract about 1,500 applicants for each available vacancy.[6]

Peth Area candidates have substantially higher pass rates than the average candidate. The survey asked candidates to report their score on two recent preliminary exams.[7] These exams are qualifying exams that determine whether a candidate passes on to the next stage of the selection process. On average only 2 percent and 4 percent of candidates passed these exams, respectively.[8] However, in the Peth Area sample, among those who took the exam, 19 percent and 42 percent passed.

5 Appendix Figure B1 includes a map of this area.

6 For example, in the 2016 State Service Exam, there were 191,536 applicants and 135 were ultimately selected.

7 Specifically, it asked for their score in the 2019 State Services Preliminary Exam and the 2019 Combined Group B Preliminary exam, which took place 10–12 months prior to the survey.

8 These figures are sourced from the MPSC's 2018–2019 Annual Report.

Studying in the Peth Area is among the most expensive options available for a prospective MPSC candidate, relative to either studying at home or in a smaller town. For some candidates, though, the cost is worth it. Living in the Peth Area provides access to a concentration of motivated peers, study material, and high-quality coaching classes. Candidates in the Peth Area sample are therefore both highly motivated and willing to invest heavily. For this reason, it is likely that selection into the sample will result in relatively high estimates of the value of amenities compared to estimates based on the average candidate. These estimates are also more likely to reflect the preferences of candidates on the margin of selection, compared to estimates based on the average candidate.

3.2. Sampling

The sampling frame is a random sample of candidates preparing for state-level government jobs in libraries in the Peth Area. Libraries serve as the primary sampling unit. These libraries are private businesses that offer candidates a quiet space to study for a fee. Libraries are in high demand because out-of-town students generally do not find their rooms conducive to studying.

Before sampling libraries, the author and a research assistant first conducted a census of all libraries in the Peth Area. This was done by physically traversing the entire zip code and verifying the presence of each library in person.[9] The census recorded data on the size of the library (measured in terms of the approximate number of desks available), the fee structure, and the availability of amenities. The census yielded a total of 166 libraries in the Peth Area.

The survey was conducted in six libraries drawn from a stratified random sample of the census list. After dropping six libraries that had restrictions on the types of students that could join, the remaining 160 libraries were divided into six groups based on their size and their monthly fee.[10] One library from each stratum was sampled for the survey. In the case that the survey was not feasible at the sampled library, the survey was conducted at a randomly chosen alternative from the same stratum.

Finally, students were sampled within libraries. Sampled students then received a paper survey form. Those who agreed to participate in the survey filled out the form and returned it to a research assistant, who then verified that it had been filled out correctly and answered follow-up questions in the case of confusion. The sampling strategy was designed in a way that allowed the research assistant sufficient time to attend to each sampled student, while also accounting for the fact that the population in the library is constantly moving, as students enter and leave throughout the day. To account for the possibility that the population of students varies across the day, the student sample is stratified by time. For each library, the day was divided into 7–16 time slots, ranging from 9:30 a.m. to 6:00 p.m. to account for the changing composition of students over the course of the day. The research assistant divided the set of available desks in the library into roughly equal-sized groups. Each group of desks was then randomly matched to a time slot. The matching was done in a way that allowed for gaps in the survey schedule to ensure that the probability that a time slot was selected was independent of the library size. At the designated time, the research assistant would visit the section and provide a copy of the survey to all students who were (a) present in the desk at the start of the session and (b) currently preparing for a state-level government job. In case a student sat down at a desk in that section after the start of the session, that student would be excluded from the sample.

Appendix Table B1 summarizes details of the response rate at each of the six libraries included in the survey. The survey was conducted between February 11, 2020 and March 12, 2020. The response rate

9 Even still, it is possible to have missed some libraries. To increase the coverage of the census, the author developed a website that allowed members of the public (particularly MPSC students) to suggest a library that was missing from the list. Respondents were offered compensation for each library that they found that was not on the list. The census was complete only after the public stopped sending new suggestions.

10 These six bins were constructed using the Cartesian product of three bins for size (dividing the marginal distribution by terciles) and two bins for fees (dividing the marginal distribution by the median).

Table 1. Summary Statistics

	Full sample			Restricted sample		
	Mean	Std. dev.	*N*	Mean	Std. dev.	*N*
Persistence						
Age	24.8	2.7	186	24.7	2.6	116
Dropout age	26.9	3.7	114	26.9	3.7	114
Demographic characteristics						
Male (0/1)	0.809	0.394	194	0.752	0.434	121
From Pune district (0/1)	0.052	0.222	194	0.033	0.180	121
Caste group: General category (0/1)	0.186	0.390	194	0.190	0.394	121
Caste group : Scheduled Caste/Scheduled Tribe (0/1)	0.103	0.305	194	0.116	0.321	121
Work experience						
Currently working (0/1)	0.015	0.124	194	0.025	0.156	121
Ever worked (0/1)	0.134	0.342	194	0.149	0.357	121
Alternative occupation						
Alt. occ.: Business (0/1)	0.557	0.498	194	0.595	0.493	121
Alt. occ.: Farming (0/1)	0.258	0.439	194	0.273	0.447	121
Alt. occ.: Wage employment (0/1)	0.278	0.449	194	0.281	0.451	121
Expected monthly income in alt. occ. after 1 year of experience	45,160	42,890	157	43,040	37,490	99
Expected monthly income in alt. occ. after 10 years of experience	279,500	1,005,980	149	236,780	741,830	94
Income support						
Total monthly expenses	8,290	1,940	194	8,600	2,060	121
Monthly transfer from home	7,920	2,010	189	8,170	1,980	117
Subjective beliefs						
Subjective yearly pass probability	0.053	0.074	154	0.058	0.081	99
Expected monthly income as a Tehsildar after 1 year of experience	76,590	100,090	170	72,970	86,610	107
Expected monthly income as a Tehsildar after 10 years of experience	306,200	1,204,570	161	397,710	1,502,020	101

Source: Author's analysis based on data from the Peth Area Library Survey.
Note: This table presents summary statistics from the survey. The sample consists of Maharashtra Public Service Commission (MPSC) exam candidates studying in libraries located in the 411030 zip code of Pune. The Restricted sample keeps only those survey rounds in which data on the preferred dropout age are available. All monetary values are reported in nominal pre-COVID 2020 Indian Rupees (INR), and rounded to the closest tens place. The alternative occupation categories are not mutually exclusive.

fell dramatically in the last library because the onset of the COVID-19 pandemic caused most students to return to their hometown.

3.3. Defining the Analysis Sample

The analysis references two distinct samples. The *full sample* uses the set of observations who have non-missing values for all the variables used for structural estimation. Next, the *restricted sample* further restricts the sample to observations for whom the anticipated dropout age data are available. Due to an error in survey implementation, this variable is not available for individuals in the first two libraries that were surveyed.[11] On the whole, the full sample and restricted sample report similar averages for a wide range of survey responses (see table 1), which suggests that these samples are comparable.

11 In the first two libraries, respondents mistakenly thought that the question asked about the maximum allowable age instead of their own personal preference. In subsequent surveys, the research assistant explicitly clarified the meaning of the question with respondents.

The reported standard errors do not adjust for clustering. There are only four clusters in the restricted sample, and six in the full sample; with so few clusters, clustered standard errors will largely be uninformative. As discussed in Abadie et al. (2023), unclustered standard errors only tell us about the uncertainty induced by sampling variation *within* the specific libraries in the sample, and not about the uncertainty arising from which libraries were included in the sample. This means that these standard errors are valid for inference about the population of students that attend the specific libraries that appear in the sample, but they are not correct for inference about the population of MPSC students in the Peth Area as a whole. However, since students in the Peth Area are already a highly selected group within the population of MPSC students, the thrust of the conclusions of this study does not meaningfully change if the students who study at the sampled libraries is taken to be the population of interest.

3.4. Measurement of Model Parameters

The survey captures variables that proxy for five main parameters that relate to the model: (a) b, the level of consumption that candidates have while preparing for the exam; (b) w, earnings in the outside option; (c) p, the probability of success; (d) t, the candidate's current age; and (e) t^*, the age at which the candidate drops out. Summary statistics for each of these parameters are included in table 1.

The parameter b is measured by asking respondents to report the amount of income they receive from home every month. On average, candidates receive Rs. 8,000 per month. Almost all candidates are supported by their family. Candidates were also asked to report their monthly expenditure across a range of standard categories. On average, transfers from home fund 97 percent of total expenses.

The parameter w is measured by asking respondents to estimate their monthly earnings in their outside option. This was done by first asking candidates to consider the specific career they would choose if they were to drop out of exam preparation right away. Candidates were then asked to report the income they expect to earn in a typical month in that career. This question was asked over two different time horizons—within 1 year of starting and within 10 years of starting—to account for the possibility that some careers have lower initial earnings but higher lifetime earnings. Consistent with the model's assumption of minimal search effort in the outside option, most candidates (about 75 percent) anticipate that their outside option is either farming or business.

How reliable are the self-reported outside earnings? The main threat to interpreting the parameter estimates correctly is that respondents may misreport their true beliefs about outside earnings.[12] One way to assess the importance of this concern is to compare the reported distribution of earnings with the actual distribution of earnings of similar individuals. If respondents' beliefs align with the actual distribution, it is more likely that they are reporting truthfully. To this end, Appendix Figure B2 compares candidates' reported outside earnings with data from CMIE's Consumer Pyramids Household Survey (CPHS), which provides data on a sample of households across India. The analysis uses CPHS data from waves conducted in 2019, restricting the sample to individuals who are comparable to the Peth Area Survey sample: college graduates in Maharashtra between the ages of 25 and 30. The figure compares the distribution of total household earnings with the distribution of expected earnings in the outside option within 1 year of leaving exam preparation (from the full sample). In the CPHS data, observations are re-weighted so that business, farm, and wage earners are represented with the same frequency as in the Peth Area Survey. The figure shows that the two distributions are very similar. Average earnings in the CPHS are Rs. 44,800 per month, compared to an expected Rs. 46,900 per month in the Peth Area Survey.

The parameter p is measured by asking candidates to provide subjective estimates of the average probability of success for candidates in the Peth Area. For the purposes of estimating the value of a government job, candidates' subjective beliefs about the probability of success matter more than the objective prob-

12 It is also possible that respondents have biased beliefs about outside earnings, but as long as they report the information that they act upon then the parameter estimates should still have the correct interpretation.

ability. To elicit these beliefs, candidates were informed that about 12,000 students study for the MPSC in the Peth Area, and were then asked to estimate how many of the 12,000 candidates they expect to be successful in any given year.[13] The respondent's subjective assessment of p is the recorded response divided by 12,000. In a few cases, respondents provide values greater than 50 percent. This appears to be a result of misunderstanding the question, and therefore these responses are removed from the analysis.

The parameter t, the respondent's current age, is measured by asking respondents for their date of birth. The difference between the date of the survey and the date of birth gives a precise estimate of the respondent's age.

Finally, the parameter t^* is measured by asking respondents to report the maximum age at which they would be willing to prepare for the exam. As mentioned above, this outcome is only available in the restricted sample. Note that this a preference parameter, and not a belief. Candidates may drop out sooner than their preferred dropout age due to constraints (e.g. because of a shock to household income), but if the self-reported data are reliable then they should stay no longer than the observed dropout age.

How reliable are the self-reported preferred dropout ages? The main threats to reliability are that (a) candidates may not be truthful in their reports (e.g. because they are embarrassed about stating their true preferences) and (b) candidates may not be time-consistent in their preferences. One test of reliability is that the number of candidates studying at a given age should not be larger than the number of candidates who expect to study that long. In other words, the value of the cumulative distribution function (CDF) of the current age should never be smaller than the value of the CDF of the preferred dropout age. This consistency check holds in the data (see appendix fig. B3). Moreover, the preferred dropout age is not out of line with the time it actually takes for candidates to get selected. Appendix Figure B4 plots the distribution of the preferred dropout age in the sample against the age at which candidates are actually selected. The distributions nearly overlap, which suggests that candidates' expectations are realistic about the level of investment required to succeed.

There is substantial variation in the preferred dropout age (see fig. 2). At the 10th percentile, respondents report not being willing to continue studying past age 22, or just one to two years after completing college. At the 90th percentile, respondents report being willing to study until age 31. The model helps make sense of this variation.

4. Structural Estimation

4.1. Estimation Strategy

Agents are assumed to be risk averse with a Bernoulli utility of $u(c) = \ln(c)$. This assumption accords with the available evidence on risk aversion in labor supply (Chetty 2006). The robustness of this assumption is checked by setting $u(c) = (c^{1-\eta} - 1)/(1-\eta)$ and perturbing η around a neighborhood of 1.

The model depends on two constants. First, the discount factor β is fixed using the prevailing interest rate. The State Bank of India provided interest rates of 6.8 percent for one year deposits at the time of the survey.[14] The discount factor is therefore fixed at $1/(1 + 0.068) \approx 0.936$. Second, candidates' last anticipated working year T is fixed at age 60. This is both the standard mandatory retirement age for government employees and the age at which male college graduates in Maharashtra typically retire.[15]

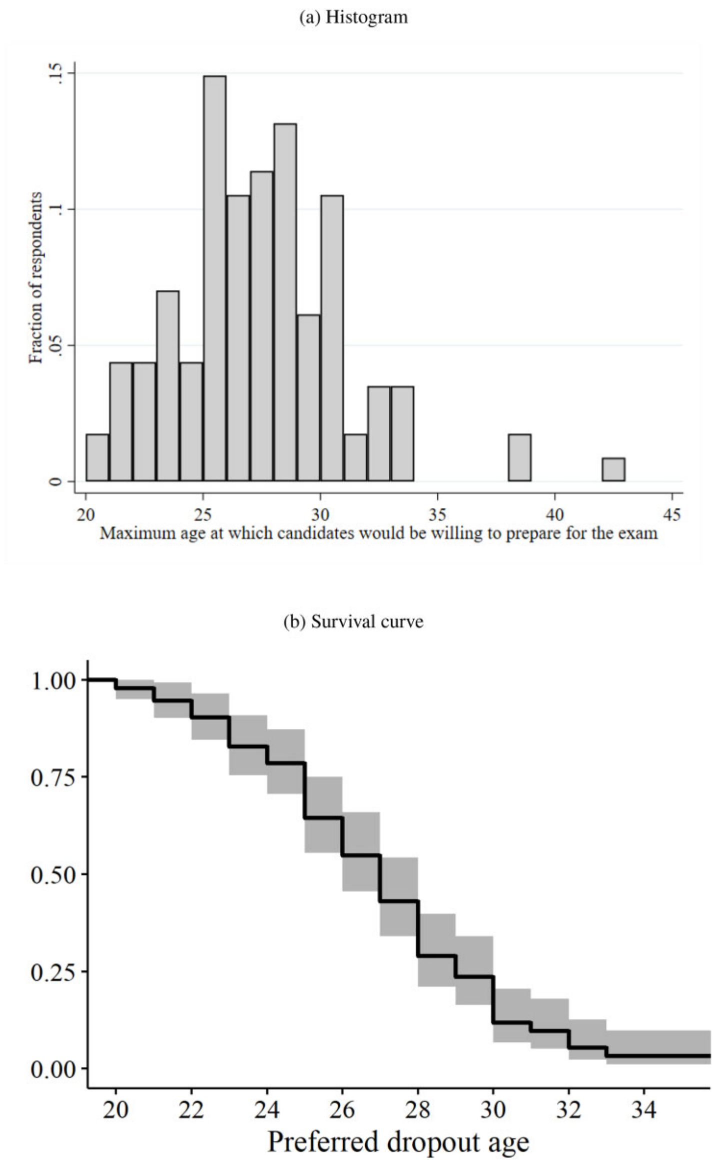
In the absence of an obviously superior method of estimating the model, three distinct approaches are reported, each of which imposes different kinds of assumptions on the data and the data-generating pro-

13 The figure of 12,000 candidates studying in the Peth Area is based on the library census. The census provides data on total capacity at each library. The total number of students is estimated to be total capacity multiplied by the average attendance rate at 9 a.m., when attendance typically was the highest.

14 The source for the interest rate is the SBI website: https://sbi.co.in/web/interest-rates/interest-rates/deposit-rates.

15 Consistent with this assumption, there is a steep rise in reitrement rates at this age, which is visible in household survey data. See appendix fig. B5.

Figure 2. Distribution of Candidates' Preferred Dropout Age.



Source: Author's analysis based on data from the Peth Area Library Survey.

Note: The survey asked candidates to report the maximum age at which they would be willing to continue studying for Maharashtra Public Service Commission (MPSC) exams, i.e. their preferred dropout age. The figure plots two distinct views of the distribution of candidates' responses. Panel A plots a histogram. Panel B plots a Kaplan–Meier estimate of the survival curve, with a 95 percent confidence interval in the shaded region.

cess. To the extent that these approaches yield similar results, they should help triangulate the underlying parameter of interest: the money-equivalent value of a government job.

Estimator 1: Moment inequality. The first approach, which imposes the weakest assumptions, uses a partial identification strategy. This approach addresses the concern that the actual dropout age is not observed. However, by virtue of appearing in the sample, it must be that candidates' dropout age is at least as large as their current age.

According to the model, candidate i persists as long as the value of unemployment U_t exceeds the value of obtaining a private-sector job P_t. This is true as long as age t_i satisfies

$$\frac{u(w_i) - u(b_i)}{u(w') - u(w_i)} \leq p_i\left[\frac{\beta(1-\beta^{T-t_i})}{1-\beta}\right].$$

Suppose that all unobserved heterogeneity is due to unobserved variation in p_i, i.e. $p_i = \bar{p} + \epsilon_i$ where $E[\epsilon_i] = 0$. In that case, rearranging this inequality and taking the expectation of both sides yields the following moment inequality:

$$E\left[\frac{u(w_i) - u(b_i)}{u(w') - u(w_i)} \cdot \frac{1-\beta}{\beta(1-\beta^{T-t_i})} - \bar{p}\right] \leq 0. \quad (3)$$

Since the left-hand side is strictly decreasing in w' for $w' > w_i$, the moment inequality identifies a lower bound on the value of w' that is consistent with the data.

Estimator 2: GMM. The parameter w' can be point identified by replacing the inequality in equation (3) with an equality at the preferred dropout age.[16] The validity of this estimate requires stronger assumptions, namely that candidates are time-consistent and report their preferences truthfully.

Estimator 3: Maximum likelihood. Alternatively, suppose that all individuals have the same subjective probability of selection, but value the government job differently. In particular, suppose that $\ln w_i' \sim N(\mu, \sigma^2)$, and that all individuals have the same subjective probability of selection $\bar{p}$. The parameters of the distribution of $\ln w_i'$ can be estimated using maximum likelihood. This model implies that a particular function of the data z_i is normally distributed. Therefore, the maximum likelihood estimate admits the following closed-form expressions:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} z_i,$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(z_i - \hat{\mu})^2,$$

where in the case of log utility z_i is given by[17]

$$z_i = \frac{u(w_i) - u(b_i)}{\phi(T, t_i^*) \cdot \bar{p}} + u(w_i). \quad (4)$$

Here, $\phi(T, t_i) = [\beta(1-\beta^{T-t_i})]/(1-\beta)$. Because w_i' is log-normally distributed, I estimate $E[w_i']$ with $\exp(\hat{\mu} + \frac{1}{2}\hat{\sigma}^2)$, and I estimate the median of w_i' with $\exp(\hat{\mu})$. As is well known, the maximum likelihood estimator of σ^2 is biased downwards, but it has lower mean square error than the unbiased estimator $n/(n-1)\hat{\sigma}^2$.

Inference. For all three estimation strategies, standard errors are calculated based on 1,000 bootstrap samples. The 95 percent confidence intervals report the range between the 2.5th percentile to the 97.5th percentile of the bootstrap distribution.

4.2. Results

All three estimation approaches suggest that candidates place a high value on government jobs (see table

16 In theory, this method would still estimate a lower bound if the maximum age eligibility requirement was binding. However, this does not appear to occur in the data. Only a handful of candidates report a preferred dropout age at the age limit.

17 For more general CRRA utility functions, one can show that

$$z_i = \frac{1}{1-\eta}\ln\left[(1-\eta)\left(\frac{u(w_i) - u(b_i)}{\phi(T, t_i^*) \cdot \bar{p}} + u(w_i)\right) + 1\right].$$

Table 2. Estimates of the Value of a Government Job

Estimator	Source of unobserved heterogeneity	Parameter	Parameter estimate (Rs. lakh/month)	
1 – Moment Inequality	Probability of selection	Lower bound on w'	5.8	4.6
			[3.6, 11.7]	[2.6, 10.7]
2 – GMM	Probability of selection	w'	–	4.7
			–	[2.7, 11.8]
3 – MLE	Value of government job	μ	–	0.9
				[0.2, 1.8]
		σ^2	–	3.6
				[2.3, 6.0]
		$E[w_i']$	–	14.7
				[4.1, 107.7]
		Median w_i'	–	2.4
				[1.3, 6.2]
Sample			Full	Restricted
N			147	93

Source: Author's analysis based on data from the Peth Area Library Survey.
Note: The table presents estimates of the value of a government job using the three different estimation strategies desribed in the Estimation Strategy section. GMM refers to the Generalized Method of Moments estimator. MLE refers to the Maximum Likelihood Estimator. For each estimate, a 95 percent confidence interval based on 1,000 bootstrap samples is provided in brackets. The confidence intervals do not adjust for clustering. The restricted sample excludes observations where data on the preferred dropout age is unavailable, which is necessary to compute Estimators 2 and 3.

2). The table reports estimates in lakhs of rupees (equal to 100,000 rupees), a natural unit of account for wages in India. In the full sample, the partial identification approach yields a lower bound of Rs. 5.8 lakh per month [95 percent CI: 3.6–11.7]. This estimate falls in the restricted sample, which likely reflects normal sampling variation. As in table 1, the preferred dropout age is not much higher than candidates' current age. Accordingly, the point-identified estimate of the value of a government job is not much higher than the lower bound in the same sample (4.7 vs. 4.6).

The maximum likelihood estimator requires a wide degree of dispersion in the valuation of amenities in order to fit the data. Moreover, the implied distribution of the value of a government job exhibits a strong right skew. The estimates imply that the majority of candidates in the sample value government jobs at more than Rs. 2.4 lakh per month [95 percent CI: 1.3–6.2].

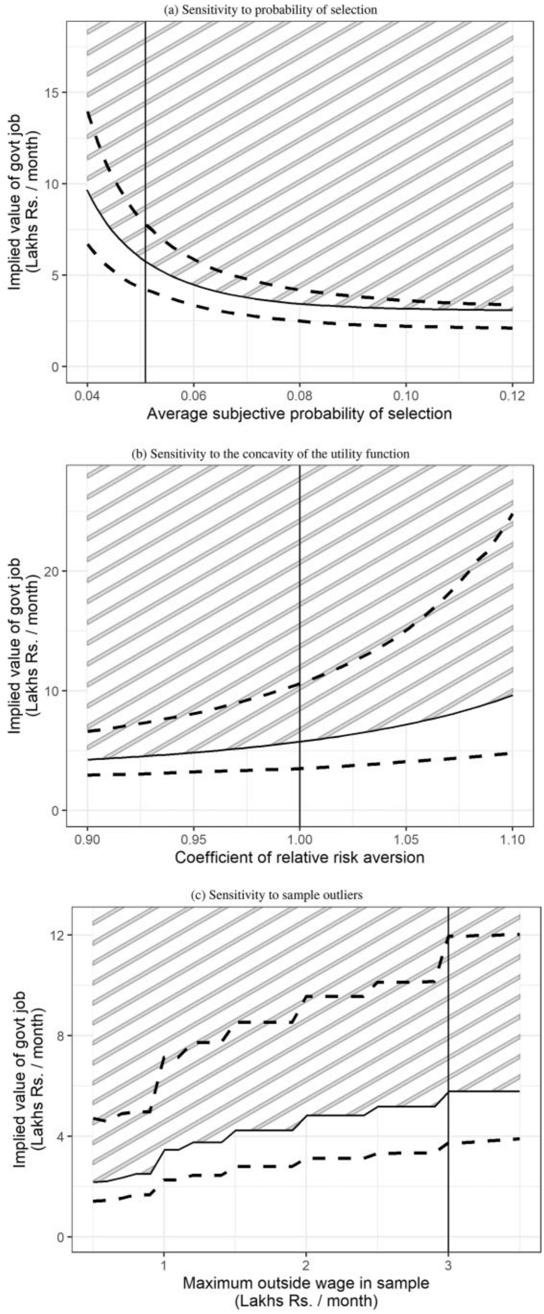
How do these estimates compare to the nominal salary? Table 3 presents a calculation of the nominal salary of a Tehsildar, one of the highest-paid posts recruited through MPSC competitive exams. To account for the fact that government salaries increase every year, the net present value of the income stream is converted into the equivalent annuity. This requires fixing a discount factor, which is kept the same as the value used to estimate the model. This calculation yields an annuity value of a Tehsildar post of about Rs. 0.8 lakh per month. The gap between the nominal salary and the private valuation most likely reflects the amenity value of government jobs. This implies that at least 66 percent of the value of a government job is due to unobserved amenities.

4.3. Robustness

This section assesses the robustness of the main findings to local perturbations of key features of model parameters and the data. Each of these robustness checks focuses on how the estimate derived from Estimator 1 (i.e. the Moment Inequality) changes, since this is the estimate that makes the least assumptions of the data. The results are summarized in fig. 3.

The first robustness check examines the sensitivity of the estimates to $\bar{p}$. This accounts for the possibility that the estimates of the probability of selection are not the same as the ones that account for their behavior, e.g. because individuals over-weight low probabilities, as in prospect theory. Even if $\bar{p}$ were

Figure 3. Sensitivity of the Estimated Lower Bound on the Value of a Government Job.



Source: Author's analysis based on data from the Peth Area Library Survey.

Note: The figure plots the sensitivity of the estimates of the value of a government job to model parameters. In each figure, a given model parameter varies along the *x*-axis. The *y*-axis plots the estimated value of a government job using the moment inequality estimator (estimator 1). The shaded area marks the region in which the parameter values are consistent with the observed search behavior as a function of the variation in that model parameter. The solid vertical line marks the values of the parameter used in constructing the main estimate. The dashed line mark the 95 percent confidence interval of the boundary of the shaded region, obtained by via 1,000 bootstrap samples.

Table 3. Maharashtra Tehsildar Salary Calculation

	As of 2019
Government policy	
Salary group	Pay band 3 with grade pay 5000
Starting pay	Rs. 55,100 per month
Annual growth rate	3%
Retirement age	60
Model parameters	
Annual discount factor	0.936
Value calculations	
Total Net Present Value	Rs. 14,242,460
Annuity equivalent	Rs. 81,429 per month

Source: Maharashtra 7th Pay Commission. See (a) the Maharashtra Government Resolution No. RPS 2019/CR-1/SER–9 dated January 30, 2019 (Finance Department) and (b) Maharashtra Government Resolution No. RPS 2019/CR-01/SEVA-9, dated January 30, 2019 (Finance Department). These documents are available at https://finance.maharashtra.gov.in/Sitemap/finance/pdf/Gazette_Seventh_Pay.pdf.

Note: The Tehsildar post is among the highest paying posts available through the Maharashtra Public Service Commission (MPSC) exam. The table presents a calculation of the expected annuity equivalent of a Tehsildar salary over the lifetime of their expected career. The net present value calculation is based on career trajectory in which one starts as a Tehsildar at age 20, the earliest age at which one is eligible to be selected for this post. All monetary values are reported in nominal pre-COVID 2020 Indian Rupees (INR).

twice as large as the value observed in the data, the estimated lower bound on w' would not fall below Rs. 3 lakh per month, and the 95 percent confidence interval would still exclude valuations less than Rs. 2 lakh per month. This is still substantially more than the nominal value of a government job.

The second robustness check considers how sensitive the estimates are to the functional form of the utility function by varying the risk aversion parameter. The implied value of a government job decreases as the risk aversion coefficient decreases. But even with a 25 percent reduction (from 1 to 0.75) in the parameter, the estimate of w' still stays above Rs. 3.5 lakh per month.

The final robustness check considers how the estimate falls when individuals with the highest reported outside wage offers are excluded from the sample. The model implies a single common value of a government job across all candidates. To fit the data, the model may place substantial weight on ensuring that the estimate falls above these values. The figure shows that the estimated lower bound on w' is indeed sensitive to excluding these observations. However, even when individuals who anticipate an outside option of more than Rs. 1 lakh per month are excluded, the estimated lower bound on w' remains above Rs. 3.5 lakh per month.

5. Alternative Explanations

The model requires candidates to place a high value on government jobs in order to rationalize their search behavior. This section examines two alternative explanations for the high estimated value of government jobs: (a) that candidates are misinformed about the nominal wage in government jobs and (b) that candidates derive value from the search process itself. Neither of these alternative explanations are compelling in this context.

5.1. Are Candidates Misinformed about the Salary in Government Jobs?

Candidates may persist simply because they overestimate the salary offered in government jobs. It is not unreasonable to believe this to be the case. Information about the wage offered in government jobs is not necessarily easy to obtain. The notifications advertising government jobs generally do not list the nominal monthly wage. Instead, it lists the "pay band." One then needs to look up the nominal wage in a table that the government continually revises.

To assess beliefs about wages, the survey asked respondents to guess the monthly wage of a Tehsildar after one year of experience.[18] In general, candidates tend to have accurate beliefs about the initial salary. The median belief is Rs. 60,000 per month, which is close to the true value of Rs. 55,000 per month. Moreover, about 60 percent of respondents guessed within Rs. 20,000 of the true value, and 85 percent of individuals provided an estimate of less than Rs. 1 lakh per month. The average estimate (as seen in table 1) is much higher than the median because a few individuals provide very large estimates. However, there is no evidence to suggest that individuals systematically entertain beliefs about compensation that are out of line with the official salary.

5.2. Does Exam Preparation Also Have Amenities?

Having a government job may not be the only state with unobserved amenities. It is possible that there is an amenity value associated with exam preparation, independent of the instrumental value of obtaining a government job. These amenities could be either net positive or net negative relative to the outside option. To the extent that exam preparation is associated with positive amenities, we can rationalize the observed search behavior with a lower value for government jobs.

There are two main sources of amenities in exam preparation. First, there are amenities that are intrinsic to the search process. On the positive side, candidates may value delaying starting work, or the lifestyle of being a student. On the negative side, candidates may find studying stressful, tedious, or unpleasant.

The other source of amenities is the city of Pune itself. Candidates in the sample study in the city but often plan to take up employment in smaller towns or villages. The cost of living in Pune is higher, so the transfers that candidates receive from home have less purchasing power in the city. However, the amenities associated with urban life may partially or fully compensate for the higher cost of living. The net effect on persistence depends on the balance of these factors.

One way to address this concern within the scope of the model is to multiply the observed values of b_i by some positive constant α, where values of $\alpha < 1$ correspond to net negative amenities of unemployment and values of $\alpha > 1$ correspond to net positive amenities. Even with large positive amenities, the conclusions do not change substantially. For example, the estimated lower bound on w' still remains above Rs. 3.5 lakh per month even when $\alpha = 2$.

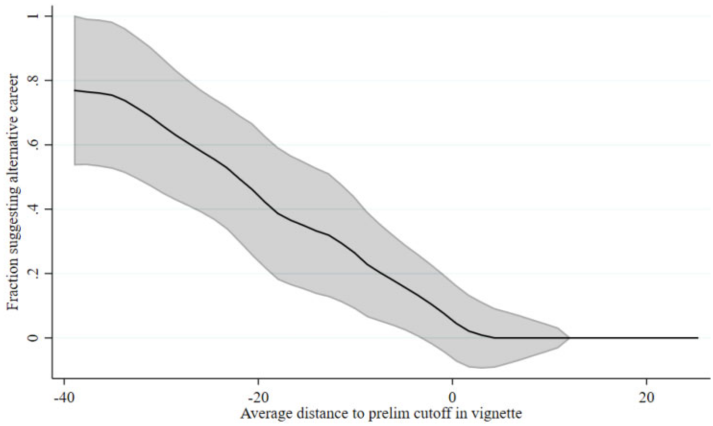
Evidence from a survey experiment further indicates that the amenity value of exam preparation is not so large as to encourage candidates to continue preparing for its own sake. If candidates did in fact have these non-instrumental motivations, then we would expect to see candidates express a preference to persist even as the probability of passing vanishes. This logic can be expressed formally as follows. Suppose candidates persist in period t as long as $p_t w'_t + b_t > w_t$, where p_t is the probability of success, w'_t is the value of a government job, b_t is the value of searching, and w_t is the value of the outside option. As long as $w_t > b_t$, then there is some value of p_t below which candidates will prefer to drop out.

The survey experiment tested this hypothesis with a convenience sample of 50 MPSC candidates in the Peth Area. Respondents were presented with a vignette in which they were asked whether they would recommend a hypothetical friend to take the test next year, given their score history over the past three attempts. In each iteration of the survey, the score history of this hypothetical friend was randomly varied.[19] The vignette was designed so that the hypothetical friend described came from the same district as the respondent and had the same gender. This was done to increase the likelihood that the respondents'

18 As reported in table 1, the survey also asked candidates about their salary beliefs after 10 years of experience. However, the accuracy of these beliefs is difficult to assess because these beliefs incorporate uncertainty about interim government policy changes.

19 The score in the 2016 exam X_{2016} is randomly chosen from the set {10, 30, 50}. Scores for 2017 and 2018 were generated using the following auto-regressive process: $X_t = 0.33X_{2016} + 0.67X_{t-1} + \epsilon_t$, where $\epsilon_t \sim N(0, \sigma = 4.5)$. This generates a set of realistic scores that have a fixed mean but vary in trajectory. The exam has two stages. If the randomly generated score crosses the cutoff, then the score history included a randomly generated second stage score from a uniform distribution between 30 and 50.

Figure 4. Are Candidates Willing to Persist No Matter How Low Their Score Is?



Source: Author's analysis based on data from the Peth Area Vignette Survey.
Note: The figure summarizes the results of a vignette experiment that measured how likely respondents were to recommend a hypothetical candidate persist in exam preparation based on their test score history. The *x*-axis plots the average of the three prior test scores shown in the vignette. The *y*-axis plots the fraction of respondents recommending that the hypothetical friend choose another career.

recommendation reflects how they would make the same decision for themselves. Respondents were able to provide one of three recommendations: (a) continue preparing for the MPSC only, (b) prepare for the MPSC, but also prepare a backup option, and (c) focus on an alternative career. Candidates are marked as recommending the friend drop out if they respond with option (c).

Figure 4 plots the fraction of individuals recommending dropout as a function of the average distance to the preliminary exam cutoff score across the three scores shown in the vignette.[20] If individuals had strong non-instrumental reasons for studying, then they would likely not recommend dropping out no matter how low the scores presented in the vignette were. The share of candidates not recommending drop out even at very low scores is indicative of the share of candidates who have non-instrumental reasons for studying. The fact that most candidates recommend dropping out at very low test scores suggests that non-instrumental reasons for studying are uncommon.

6. Unpacking the Amenity Value

What, specifically, are the unobserved amenities that candidates value so highly? This section summarizes findings from a series of focus-group discussions that help us better understand candidates' motivations for investing heavily in exam preparation. The focus-group discussions were conducted by the author with about 10 candidates across three libraries in the Peth Area.[21] Candidates were selected from specific libraries where someone from the MPSC community could facilitate an introduction.[22] The small, non-

20 Note that clearing the preliminary cutoff score only allows one to progress to the next stage of the exam, at which point the odds against selection are still substantial.

21 The focus-group discussions were conducted in January 2018, two years before the Peth Area Survey. None of the libraries from the focus-group sample re-appear in the Peth Area Survey sample.

22 Candidates were reluctant to speak candidly without the encouragement of someone they knew personally. The introductions permitted deeper, more honest conversations than would have been possible if participants had been randomly sampled.

random sample precludes generalization,[23] but the responses provide a useful starting point for developing a deeper understanding.

6.1. Candidate Responses

Focus group participants were asked whether they would ever consider focusing their job search on the private sector instead of continuing on their current path of exam preparation. Specifically, participants were asked whether they would prefer (a) a Tehsildar post paying about Rs. 50,000 per month (i.e. the best outcome in the exam process) or (b) a private-sector job paying Rs. 3 lakh per month. All participants responded that they would take the government job, consistent with the estimates from the model.

The consensus explanation for this preference was that government officers were unique in their ability to command "respect" in their home villages. Candidates believed that while private employment "only provides money," government employment confers a kind of social status that money cannot buy. They provided examples to illustrate the ways in which government officials receive special treatment: being addressed in the local language with the "formal you" (*aap*) instead of the "informal you" (*tumhi*), being invited as a special guest for social gatherings (e.g. for weddings, religious festivals, or public celebrations of national holidays), being solicited for help or counsel, or being a common topic of conversation. Candidates were skeptical that someone who works in the private sector could achieve the same level of local celebrity.[24]

Why are government officials so highly respected in villages? Candidates highlighted two main reasons. First, government work is considered to be intrinsically valuable. Candidates described government work as a form of *samaj seva* (social service).[25] By contrast, they thought of people who work in the private sector as working only for themselves. Second, government officials are respected because people depend on them in order to access the state. For example, candidates described how it was useful to know someone in government in order to fix a pot hole, or to get a piped water connection to one's house.[26] These relationships are important enough to most people that even if they thought poorly of an official they would continue to treat them with deference. Candidates described this kind of power as one of the perks of government employment.

6.2. Discussion

Focus-group participants place consistent emphasis on the importance of "respect," which seems to comprise some mix of access to power, authority, and status.

There seem to be two main reasons why respect is important. First, candidates articulated a belief system that allows for limited means of gaining respect outside of government employment. Because respect is scarce relative to opportunities to make money, its implicit value is high. Second, candidates seem to be highly sensitive to how their work is perceived in their community. Candidates describe government officials being a shown a level of deference in day-to-day life that is not common in cities. Exposure to rural life where government officials are at or near the top of the social hierarchy therefore appears to have played a meaningful role in socializing these candidates to value government jobs highly. This may

23 The focus-group participants represented several districts, and included both men and women. The participants generally had at least a year or two of experience preparing for MPSC exams, were studying full time, and all came from rural Maharashtra.

24 As one candidate put it, no one in the village knows who Satya Nadella is (i.e. the current CEO of Microsoft, and this candidate's stand-in for the pinnacle of private-sector success), but everyone knows who the local government official is.

25 Candidates were asked whether either corruption or complacency would affect their ability to serve the public as officers. In their responses, candidates were remarkably sanguine about their ability to resist the pressures of the system.

26 Even if that official does not work in the concerned department, their recommendation can ensure the request is taken seriously.

help explain why the highly motivated candidates we find in the Peth Area are often from rural areas, even though the sample itself is located in the heart of the city.

7. Conclusion

Candidates invest more in preparing for government job exams than what can be justified by the nominal wage of the jobs themselves. This behavior cannot be fully explained by candidates deriving value from the process of studying for these exams, or by biased beliefs about compensation in government jobs. The most likely explanation that rationalizes candidates' behavior is that unobserved positive amenities comprise a large share of total compensation in government employment.

Future research should aim to develop a better understanding of which specific amenities are valuable, and how changing them would affect selection into and performance in the bureaucracy. The answers to these questions have important policy implications. For example, the fact that amenities are an important part of total compensation suggests that tying amenities to public-sector employees' behavior can be a powerful source of incentives (see, e.g., Khan, Khwaja, and Olken (2019)). But in order for this strategy to be effective, it is important to identify which amenities matter the most.

Data Availability Statement

An anonymized version of all data used in the analysis is available in the supplementary online material.

References

Abadie, A., S. Athey, G. W. Imbens, and J. Wooldridge, 2023. “When Should You Adjust Standard Errors for Clustering?” *Quarterly Journal of Economics* 138(1): 1–35.

Araujo, F., 2020. "Selection Bias and the Returns to Public-Sector Employment." *Essays on Labor and Behavioral Economics*, 3–32. University of Pittsburgh.

Campos,, M.M., *et al.* (2017) "Understanding the public sector pay gap." *IZA Journal of Labor Policy* .6,doi: 10.1186/s40173-017-0086-0

Chetty, R., 2006. “A New Method of Estimating Risk Aversion.” *American Economic Review* 96(5): 1821–34.

Finan, F., B. A. Olken, and R. Pande, 2017. “The Personnel Economics of the Developing State.”In *Handbook of Economic Field Experiments*, 467–514. Elsevier.

Gindling, T.H. , *et al.* (2020) "Are public sector workers in developing countries overpaid? Evidence from a new global dataset." *World Development*.126 doi:10.1016/j.worlddev.2019.104737.

Holzer, H. J., L. F. Katz, and A. B. Krueger, 1991. “Job Queues and Wages.” *Quarterly Journal of Economics* 106(3): 739–68.

Khan, A. Q., A. I. Khwaja, and B. A. Olken, 2019. “Making Moves Matter: Experimental Evidence on Incentivizing Bureaucrats through Performance-Based Postings.” *American Economic Review* 109(1): 237–70.

Krueger, A. B., 1988. “The Determinants of Queues for Federal Jobs.” *ILR Review* 41(4): 567–81.

Mas, A., and A. Pallais, 2017. “Valuing Alternative Work Arrangements.” *American Economic Review* 107(12): 3722–59.

Stern, S, 2004. “Do Scientists Pay to Be Scientists?” *Management Science* 50(6): 835–53.

Appendix

A. Theory Appendix

This section presents proofs of the propositions from the main text.

Lemma 1. *The variables G_t, P_t, and U_t are strictly decreasing in t.*

Proof. We will start with G_t. Since $G_t > \max\{U_t, P_t\}$ for all t by assumption, it is an absorbing state. Therefore, G_t is just a finite geometric sum for $T - t + 1$ periods. Thus

$$G_t = u(w')\frac{1-\beta^{T-(t-1)}}{1-\beta}, \tag{A1}$$

which is clearly decreasing in t.

The next step is to verify the lemma for both P and U simultaneously, working backwards from period T. Since $P_T = u(w)$ and $U_T = u(b)$, we can write

$$\begin{aligned} P_{T-1} &= P_T + \beta \max\{P_T, U_T\}, \\ U_{T-1} &= U_T + \beta[pG_T + (1-p)\max\{P_T, U_T\}], \end{aligned}$$

or, equivalently,

$$\begin{aligned} P_T - P_{T-1} &= -\beta \max\{P_T, U_T\} < 0, \\ U_T - U_{T-1} &= -\beta[pG_T + (1-p)\max\{P_T, U_T\}] < 0. \end{aligned}$$

Now assume the induction hypothesis, i.e. $P_t - P_{t-1} < 0$ and $U_t - U_{t-1} < 0$ for some t. First we want to show that

$$P_{t-1} - P_{t-2} < 0,$$

which is true if and only if

$$\max\{P_t, U_t\} - \max\{P_{t-1}, U_{t-1}\} < 0.$$

There are four cases. Note that $U_t - U_{t-1} < 0$ by assumption and

$$P_t - P_{t-1} < 0 \Rightarrow P_t - \max\{P_{t-1}, U_{t-1}\} < 0.$$

Therefore, the only remaining case is $U_t - P_{t-1}$. This case occurs when $P_{t-1} > U_{t-1}$. By the induction hypothesis we also know that $U_{t-1} > U_t$. Putting these inequalities together we get $U_t - P_{t-1} < 0$.

Next we want to show the similar case for U, i.e.

$$U_{t-1} - U_{t-2} < 0.$$

This expression holds if and only if

$$\beta p(G_t - G_{t-1}) + \beta(1-p)(\max\{U_t, P_t\} - \max\{U_{t-1}, P_{t-1}\}) < 0.$$

This is clearly true since (a) we established that $G_t - G_{t-1} < 0$ since G_t is decreasing and (b) we just showed that $\max\{U_t, P_t\} - \max\{U_{t-1}, P_{t-1}\} < 0$.

Proposition 1. *Someone who starts unemployed will eventually take private-sector work if not employed by the government, i.e. if $U_0 > P_0$, then $U_t < P_t$ or some t. Furthermore, taking private employment is an absorbing state, i.e. $P_t > U_t \Rightarrow P_{t+s} > U_{t+s}$ for all s.*

Proof. If someone starts unemployed, then $U_0 > P_0$. Since $U_T < P_T$ by construction, U_t must cross P_t at some point t^*. Furthermore, since both U_t and P_t are strictly decreasing (by Lemma A1), they must cross at a single point. Therefore after accepting private employment, the agent will never choose to remain unemployed.

Proposition 2. *When $b < w < w'$, the optimal dropout age is given by*

$$t^* = \begin{cases} 0 & \text{if } u(w) - u(b) \geq \dfrac{\beta(1-\beta^T)}{1-\beta} \\ & \quad \times\, p[u(w') - u(w)], \\ T - \dfrac{1}{\ln\beta}\ln\left[1 - \dfrac{(1-\beta)[u(w)-u(b)]}{\beta p[u(w')-u(w)]}\right] & \text{otherwise.} \end{cases}$$

Proof. Since there is a single crossing point between U_t and P_t, the optimal stopping point is given by the t at which $U_t = P_t$.

Since P and G are both absorbing states, we can write their value functions as

$$G_t = u(w')\frac{1-\beta^{(T-(t-1))}}{1-\beta},$$

$$P_t = u(w)\frac{1-\beta^{(T-(t-1))}}{1-\beta}.$$

Setting P_t equal to U_t at t^* yields

$$u(w) + \beta P_{t^*+1} = u(b) + \beta p G_{t^*+1} + \beta(1-p)P_{t^*+1}.$$

Solving for t^* by substituting in the formulas for G_t and P_t gives us

$$t^* = T - \frac{1}{\ln\beta}\ln\left[1 - \frac{(1-\beta)[u(w)-u(b)]}{\beta p[u(w')-u(w)]}\right].$$

Given the assumption that $b < w < w'$, the second term is positive, so t^* is always strictly less than T. However, it is possible that t^* will fall below zero, which is outside the domain. Solving for $t^* \leq 0$ yields the condition $u(w) - u(b) \geq \frac{\beta(1-\beta^T)}{1-\beta} p[u(w') - u(w)]$.

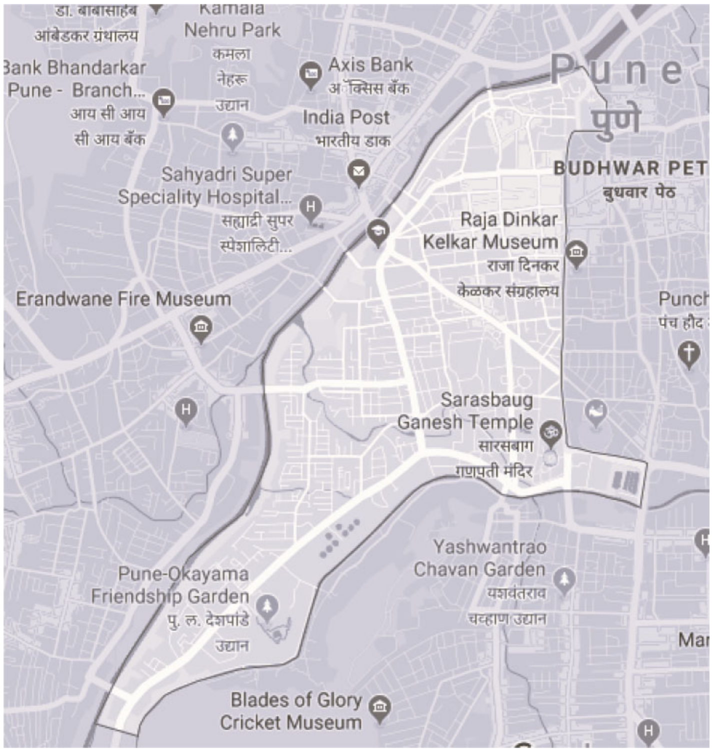
B. Additional Figures and Tables

Table B1. Peth Area Survey Response Rates

Library	Fee group	Capacity group	Survey rounds	Present (P)	Eligible (E)	Completed (C)	Eligibility rate (E/P)	Response rate (C/E)
1	Low	Low	2	53	44	36	83%	82%
2	High	Low	2	68	38	30	56%	79%
3	High	High	2	126	76	57	60%	75%
4	Low	High	1	63	48	33	76%	69%
5	Low	Medium	1	51	40	28	78%	70%
6	High	Medium	1	24	16	6	67%	38%
Total				385	262	190	68%	73%

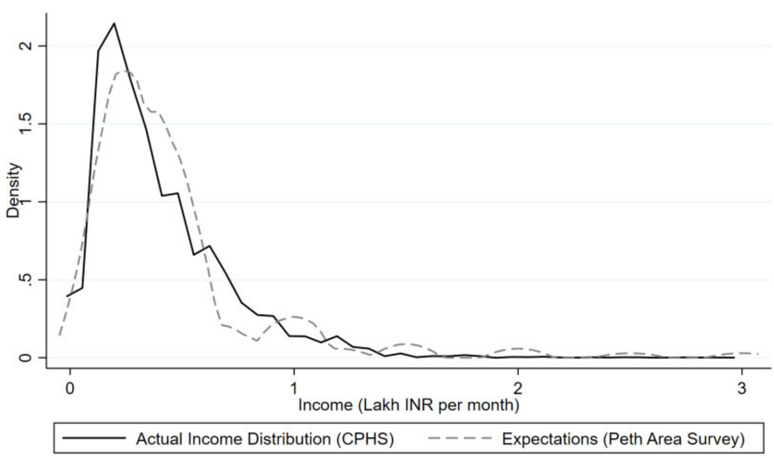
Source: The table summarizes statistics from the implementation of the Peth Area Library Survey conducted by the author.
Note: The Present count is the number of students who were sitting in sampled desks at the time of the survey. The Eligible count is the number of students in those sampled desks who were currently studying for an exam conducted by the Maharashtra Public Service Commission (MPSC). The Completed count is the number of students who returned a non-blank copy of the survey. The number of survey rounds is the number of times each cluster of desks was sampled in the library.

Figure B1. A Map of the Peth Area, Pune.



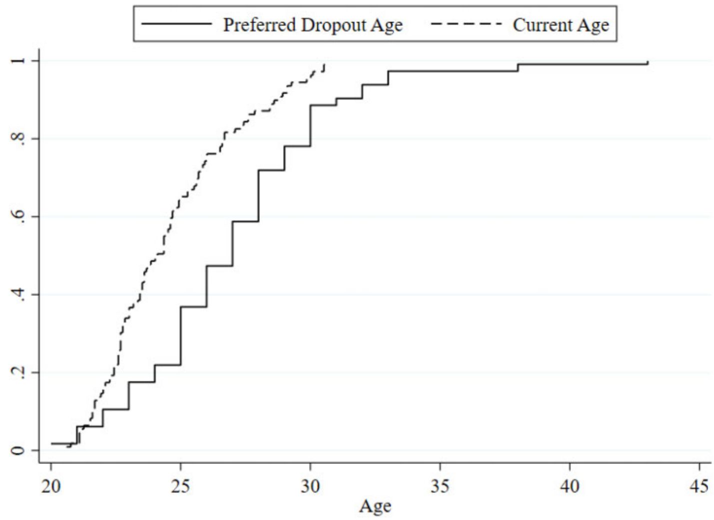
Source: Google Maps.
Note: The shaded region highlights the 411030 zip code of the city of Pune, India. This area is also known as the Peth Area. The sample for the Peth Area Survey is drawn from the population of candidates studying in libraries within this area.

Figure B2. Comparing the Actual and Expected Distribution of Earnings.



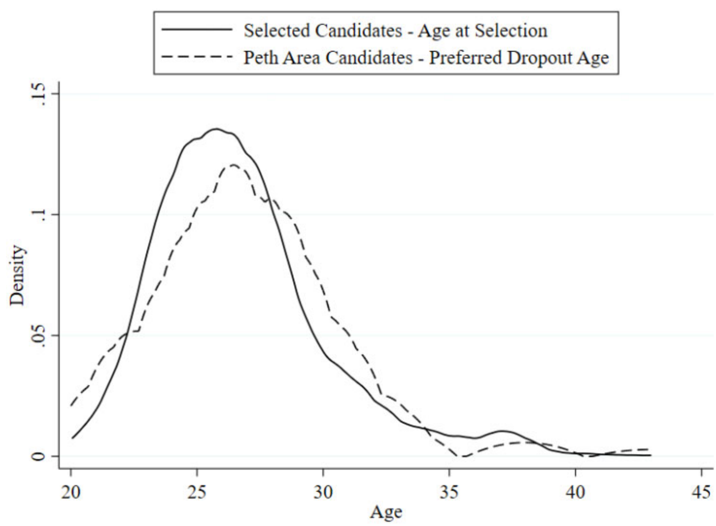
Source: Consumer Pyramids Household Survey (CPHS); Peth Area Library Survey.
Note: The figure plots kernel densities of both the empirical income distribution and the expectations recorded in the full sample of the Peth Area Survey. Expectations capture candidates' expected earnings in their outside option one year after leaving exam preparation. The CPHS sample is restricted to college graduates between the ages of 25 and 30 who live in Maharashtra. These observations are reweighted so that they reflect the same share of individuals in business, farming, and wage earnings as we see in the Library Survey data. For legibility, the long right tail of the CPHS distribution has been truncated.

Figure B3. A Test of Bias in the Preferred Dropout Age.



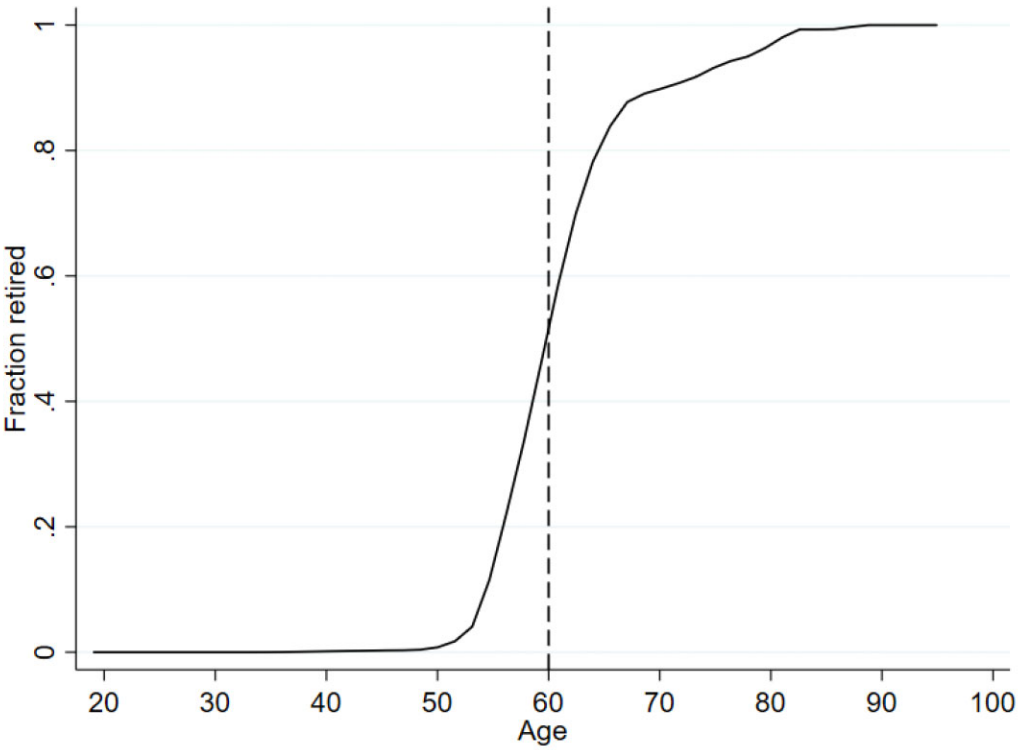
Source: Peth Area Library Survey
Note: The figure compares candidates' current age (dashed line) with the maximum age they would be willing to keep studying for the exam (solid line). Each curve plots a cumulative distribution function (CDF). The sample is restricted to candidates who provided a valid measure of their preferred dropout age.

Figure B4. Are Candidates Willing to Persist as Long as It Takes to Get Selected?.



Source: Author's analysis based on data from (a) the Peth Area Library Survey and (b) Maharashtra Public Service Commission (MPSC) 2019 List of Candidates Eligible for Recommendation for Groups A, B, and C (Advertisement Nos. 7/2019, 8/2019, 9/2019, 14/2019, 15/2019, and 16/2019).
Note: The figure compares the age at which candidates are selected into Maharashtra Public Service Commission (MPSC) posts (solid line) with the maximum age at which candidates say they would be willing to prepare for the exam (dashed line). For the Selected Candidate sample, age is calculated as the gap between candidates' date of birth and the day on which the preliminary exam was held. For the Peth Area Candidate sample, age is calculated as the gap between candidates' date of birth and the date of the survey.

Figure B5. Age of Retirement for College Graduates in Maharashtra.



Source: Consumer Pyramids Household Survey (CPHS), Wave 1, 2019 round.
Note: The figure plots a local linear regression of the fraction of college graduates in Maharashtra who are retired as a function of their age. The dashed line marks age 60, the age of retirement used in the estimation of the model.