How much are government jobs in developing countries worth?

Kunal Mangal*

December 12, 2022

Abstract

Government jobs in developing countries are valuable not just because they pay relatively higher wages, but also because they provide many valuable amenities. Does the value of these amenities compete with the nominal wage itself? I use the observed search behavior of candidates preparing for competitive exams for government jobs to infer a lower bound on the total value of a government job, including amenities. Based on a sample of 120 male candidates preparing for civil service exams in Pune, India, I estimate that the amenity value of a government job is *at least* 81% 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. I conclude with a discussion of the implications of these findings for policy and the questions it raises for future research.

---

*Email: kunal.mangal@apu.edu.in. The Weiss Fund for Development Economics provided funding for data collection. I am grateful to Robert Townsend for helping me get started; to Kranti Mane and Deepak Palhal for sincere research assistance; to Mr. Anand Patil and Mrs. Vaishali Patil for their invaluable support; to Mahesh Bade for his expertise; to Ram for his friendship and advice during the project; and to the MPSC candidates who willing to take the time to participate in the survey, despite their many stresses. The research protocol was approved by the Harvard IRB. All errors are my own.

1 Introduction

How do public and private sector compensation 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 typically offer amenities that are hard to price, and which may constitute a large share of total compensation. For example, government employees in developing countries typically obtain lifetime job security, have ready access to bribe payments, and enjoy some measure of fame and celebrity status. It is an open question whether the value of these amenities competes with the wage premium itself.

The absence of even order-of-magnitude estimates of the value of public sector job amenities limits the conversation on public sector compensation. Because wages are the most visible component of total compensation, most work on public sector compensation has focused on wages (Finan, Olken, and Pande, 2017; Araujo, 2020; Kuipers, 2022). But if amenities are a large component of total compensation, we may be ignoring the part of compensation that is most responsible for allocating individuals and effort.

The standard method to value job amenities is to use variation in the characteristics of jobs within a candidate's choice set (Stern, 2004; Mas and Pallais, 2017). But 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 valuing government jobs. I focus on India, which has one of the largest public sector wage premia in the world (Finan et al., 2017). My empirical strategy makes use of the fact that in India most government jobs are allocated through a system of highly structured system of 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. I infer how much candidates value government jobs from the amount of time they are willing to spend studying for these exams. By imposing some parametric structure on a model of exam preparation, I can price this time in monetary terms.

To estimate the model, I collected data from a sample of 120 candidates preparing for civil service exams in Pune, a city in western India in the state of Maharashtra. I targeted the survey to a neighborhood of the city in which candidates from all over the state come to study. This is therefore a population of highly motivated applicants. My sample consists of individuals for state-level civil service exams, known as the Maharashtra Public Service Commission (MPSC) exams.

I first provide evidence for the relevance of the model in this context. I focus my analysis on men, for whom the assumption of maximizing expected lifetime income is more realistic. The model makes predictions about how the exogenous parameters should correlate with dropout age. I provide evidence that these correlations appear in the analysis sample.

Next, I use the model to infer the value of a government job. I use three estimators, each of which imposes different restrictions and assumptions on the data and the data generating process. The estimates indicate that, in this sample, candidates value a government job at least 425,000 INR per month. By comparison, the annuity value of the nominal salary of a Tehsildar—one of the highest paying jobs offered through these exams—is about 81,000 INR per month.[1] This suggests that amenity value of a government job is at least 81% of total compensation.

I consider two alternative explanations for why individuals may appear to have a high value for government jobs. First, I consider the possibility that candidates are misinformed about the salary structure. Second, I consider whether candidates derive process utility from the process of studying for a government job, which encourages them to participate above

[1] The annuity value calculation takes into account the fact that government salaries are scheduled to increase by 3% per year of service.

and beyond the instrumental value of obtaining a government job. I find that neither of these hypotheses can account for the large implied amenity value of government jobs.

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. For the private sector, estimating the value of rents from queuing behavior would require taking a perhaps unjustifiably specific stand on the structure of jobseekers' search behavior across firms. However, in this context, because the exam process is already highly structured, modeling this behavior is more feasible.

This paper proceeds as follows. Section 2 presents a model of exam preparation. Section 3 describes the survey data that will be used to estimate this model. Section 4 presents the estimation strategy and the results. Section 5 discusses alternative explanations for the high observed value of government jobs. Section 6 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

I model exam preparation as an optimal stopping problem. The model incorporates specific features of 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. I infer the value of government jobs from the unobserved model parameter that rationalizes the observed dropout behavior.

2.1 Set-Up

In each year t, the agent decides whether to study for a government job. If yes, then the agent is unemployed. If not, then he takes his outside offer in the private sector, which yields an annual income of w. Consistent with the context of this paper, I treat search costs for the private sector jobs as negligible, so agents do not need to spend time searching to obtain it.[2]

Candidates that are studying obtain a government job in the next period with a 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 $t = 0, 1, 2, \ldots, T-1$:

$$G_t = u(w') + \beta \max\{G_{t+1}, P_{t+1}, U_{t+1}\} \tag{1}$$

$$P_t = u(w) + \beta \max\{P_{t+1}, U_{t+1}\} \tag{2}$$

$$U_t = u(b) + \beta \Big[pG_{t+1} + (1-p)\max\{P_{t+1}, U_{t+1}\} \Big] \tag{3}$$

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 $U_T = u(b)$.

2.2 Optimal Stopping

This model only has meaningful content when $w' > w > b$. If $w < b$ then a candidate would never give up preparing. If $w' < w$ then a candidate would never prepare for a government

[2] For about 78% of the sample, the outside option is either farming or business.

job in the first place. But when $w' > w > b$ the model set up an interesting optimal stopping rule:

Proposition 1. *Someone who starts unemployed will eventually take private sector work if not employed by the government, i.e. if* $U_1 > P_1$*, then* $P_t > U_t$ *for some* t*. Furthermore, taking private employment is an absorbing state, i.e.* $P_t > U_t \implies P_{t+s} > U_{t+s}$ *for all* s*.*

Proof. See 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 the career. I will refer to t^* as the *dropout age*.

Appendix 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 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*

$$t^* = \begin{cases} 0 & \text{if } u(w) - u(b) \geq \dfrac{\beta(1-\beta^T)}{1-\beta} p[u(w') - u(w)] \\ T - \dfrac{1}{\ln \beta} \ln \left[1 - \dfrac{(1-\beta)[u(w) - u(b)]}{\beta p [u(w') - u(w)]} \right] & \text{otherwise} \end{cases} \tag{4}$$

Proof. See Appendix. □

This function 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.

Note that this expression yields the following intuitive predictions for the exogenous variables p, w, and b affect t^*:

- $\partial t^* / \partial b > 0$, i.e. the dropout age covaries positively with income during exam preparation
- $\partial t^* / \partial w < 0$, i.e. the dropout age covaries negatively with income during exam preparation
- $\partial t^* / \partial p > 0$, i.e. the dropout age covaries positively with the (subjective) probability of obtaining a government job

These are predictions that I can take to the data to assess the validity of the model.

3 The Peth Area Library Survey

3.1 Setting

This study is based in India, a country in which the wage component of public sector compensation is relatively high. The evidence from Finan et al. (2017) indicates that the public wage premium in India is large, at about 105% (see their Table 1, Column 3). Compared to the 34 other countries in their sample, India stands out as an outlier, both in absolute terms and relative to its GDP per capita.

This paper uses data from a survey I fielded in the city of Pune. 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—known as the Peth Area—is the epicenter of exam preparation. Appendix Figure B.1 includes a map of this area. This zip code has a high concentration of both candidates preparing for competitive exams and businesses that cater to their needs, including book shops, photo copiers (which maintain ready catalogs of practice tests and study materials), coaching classes, libraries, canteens and hostels.

Most government job aspirants in the Peth Area prepare for state-level civil service jobs. These 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 1-3 exams, and there are about 1500 applicants for each available vacancy.[3] It is commonly understood that many candidates who give the test do not prepare nearly as intensely as candidates in the Peth Area. We should therefore expect selection rates in this group to be much higher on average.

3.2 Sampling

I conducted the survey with 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 business 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, a research assistant and I 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.[4] In each library that we spotted, we collected data on the following items: 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.

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

[4] Even still, it is possible that we may have missed some libraries. To increase the coverage of the census, we developed an online app that allowed members of the public (in particular MPSC students) to suggest a library that was missing from our list. The website indicated that students would receive a compensation for each library that they found that wasn't on our list. We proceeded only after we stopped receiving new suggestions.

To sample libraries, we drew a stratified random sample from the census. After dropping six libraries that had restrictions on the types of students that could join, we divided the remaining 160 into 6 groups based on their size and their monthly fee. To construct these six bins, I took the Cartesian product of three bins for size (dividing the marginal distribution by terciles) and two bins for fees (dividing the marignal distribution by the median). Within each strata, I created a random re-sampling ordering list. I also randomly varied the order in which we visited libraries from each strata.

Finally, I sampled students 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 answers and answered follow-up questions in 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, I stratified the sample by time. For each library, I divided the day into 7-16 time slots in which we would conduct a session, ranging from 9:30am to 6:00pm 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. We 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 designed time, the research assistant would visit the section and provide a copy of the survey to all students who were: 1) present in the desk at the start of the session; and 2) 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 B.1 summarizes details of the response rate at each of the six libraries included in the survey. The survey was conducted between February 11th, 2020 and March

12th, 2020. The response rate 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

Throughout the analysis, I restrict the analysis to men. Given the low female labor force participation rate in India, it is unclear whether a model based on maximizing earning potential would be appropriate for women. In principle, this is a testable hypothesis. For example, I can test whether the reduced form relationships that the model predicts hold for women as well as for men. Unfortunately, given the small sample of women, I do not have enough statistical power to run this test. I therefore drop women from the sample based on this *a priori* assumption about the context.

There are two distinct samples that I use for this analysis. The *full sample* uses the set of observations who have non-missing values for all the variables used in the structural analysis. Next, the *restricted sample* further restricts the sample to the observations for whom the anticipated dropout age data is available. Due to an error in survey implementation, this variable is not available for individuals in the first two libraries that were surveyed.[5] 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.

Throughout the analysis, I present standard errors that do not adjust for clustering. As discussed in Abadie, Athey, Imbens, and Wooldridge (2017), these standard errors are valid for inferences about the population of students that attend the specific libraries that appear in my sample, but they are not correct for inferences about the population of MPSC students in the Peth Area as a whole. With only 4-6 clusters, clustered standard errors will largely be uninformative; this study is thus not well suited to say much about the overall population of MPSC students in the Peth Area as a whole. However, since students in the Peth Area are

[5]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, we explicitly clarified the meaning of the question with respondents.

already a highly selected group within the population of MPSC students, the thrust of the conclusions of this study does not meaningfully change if we treat the students who study at the sampled libraries as 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: 1) b, the level of consumption that candidates have while preparing for the exam; 2) w, earnings in the outside option; 3) p, the probability of success; and 4) t, the candidate's current age; and 5) t^*, the age at which the candidate drops out. Summary statistics for each of these parameters is included in Table 1.

I measure b 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. I asked candidates to report separately their monthly expenditure across a range of standard categories. On average, the transfers from home total to 97% of total expenses. It is therefore fairly costly for candidates to come to Pune to study. Candidates have told me that what makes it worthwhile is that they are able to see how well prepared the competition is, which motivates them to study further.

I measure w 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. We then asked candidates to consider the income they expect to earn in a typical month in that career. We ask this question 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 by higher lifetime earnings. Consistent with the model's assumption of minimal search effort in the outside option, most candidates (about 75%) 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.[6] 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 make this comparison, I use data from CMIE's Consumer Pyramids Household Survey (CPHS), which provides data on a sample of households across India. I use data from the waves conducted in 2019. I restrict the sample to individuals who are comparable to the Peth Area Survey sample: male college graduates in Maharashtra between the ages of 25 to 30. I then compare 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, I reweight observations so that business, farm and wage earners are represented with the same frequency as in the Peth Area Survey. Appendix Figure B.2 compares the two earnings distributions. The distributions are very similar. Average earnings in the CPHS are 44,800 INR per month, compared to an expected 46,900 INR per month in the Peth Area Survey.

I measure p 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, what matters is the subjective probability and not the objective probability of success. To elicit these beliefs, we told candidates that about 12,000 students study for the MPSC in the Peth Area, and we asked them to estimate how many of them they expect to be successful in any given year.[7] 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%. This appears to be a result of misunderstanding the question, and I therefore remove these responses from the analysis.

[6] 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.

[7] The figure of 12,000 candidates studying in the Peth Area is based on the library census. For each library we estimated the total capacity and then multiplied the total observe capacity across all libraries by the average attendance rate at 9am, when attendance typically was the highest.

I measure t, the respondent's current age, by asking for their date of birth. Because I know the date of the survey and the date of birth, I can estimate age with a high degree of precision.

Finally, I measure t^* 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: 1) candidates may not be truthful in their reports (e.g. because they are embarrassed about stating their true preferences); and 2) candidates may not be time-consistent in their preferences. One test of reliability is that we should not see more candidates studying at a given age then 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 larger than the value of the CDF of the preferred dropout age. This consistency check holds in the data (see Appendix Figure B.3).

There is substantial variation in the preferred dropout age (see Figure 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, respondent report being willing to study until age 31. The model helps us understand this variation.

3.5 Assessing the Validity of the Model

The validity of the parameter estimates depend on how well the optimal stopping model outlined in Section 2 describes candidates' search behavior. In this section, I assess whether the assumptions of the model fit the men studying for MPSC jobs in the Peth Area.

To do so, I test whether the reduced form correlations that the model predicts also show up in the data. Table 2 presents the correlations between the preferred dropout age and measures of the main model parameters. All specifications include reservation category fixed effects. The parameters correlate with the preferred dropout age in the expected way. In Column (1) we see a weak with transfers from home. In Column (2), we see a negative correlation with expected outside earnings. In Column (3), we see a positive correlation between with the candidate's subjective assessment of the pass probability. If we combine all the predictors together, we see in Column (4) that the coefficients maintain the correct sign. These correlations suggest that male candidates are in fact thinking about their persistence decisions in a way that is aligned with the model.

4 Structural Estimation

4.1 Estimation Strategy

I suppose that agents are risk averse with a Bernoulii utility of $u(c) = \ln(c)$. This assumption accords with the available evidence on risk aversion in labor supply (Chetty, 2006). I test the sensitivity of this assumption by setting $u(c) = (c^{1-\eta} - 1)/(1-\eta)$ and perturbing η around a neighborhood of 1.

I supply two constants to the model. First, I fix the discount rate β using the prevailing interest rate. The State Bank of India provided interest rates of 6.8% for one year deposits at the time of the survey.[8] I therefore fix the discount factor at $1/(1 + 0.068) \approx 0.936$. Second, I assume that candidates' last anticipated working year T is 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.[9]

In the absence of an obviously superior method of estimating the model, I apply three

[8] Data obtained from the SBI website: https://sbi.co.in/web/interest-rates/interest-rates/deposit-rates

[9] Using data from the CMIE Consumer Households Pyramids Survey, I verify this assumption empirically. See Appendix Figure B.4.

distinct approaches, which impose different kinds of assumptions on the data and the data generating process. To the extent that these approaches yield similar results, they should help us 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 I do not observe the dropout age. However, by virtue of appearing in the sample, I know 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:

$$\frac{u(w_i) - u(b_i)}{u(w') - u(w_i)} \leq p_i \left[\frac{\beta(1 - \beta^{T - t_i})}{1 - \beta} \right] \tag{5}$$

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

$$E\left[\frac{u(w_i) - u(b_i)}{u(w') - u(w_i)} \cdot \frac{1 - \beta}{\beta(1 - \beta^{T - t_i})} - \overline{p} \right] \leq 0 \tag{6}$$

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. I can point identify w' by replacing the inequality in equation (6) with an equality at the preferred dropout age.[10] The validity of this estimate requires stronger assumptions, namely that candidates are time-consistent and report their preferences truthfully.

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

Estimator 3: Maximum likelihood. Alternatively, I can assume that all individuals have the same subjective probability of selection, but value the government job differently. In particular, I suppose that $\ln w_i' \sim N(\mu, \sigma^2)$, and that all individuals have the same subjective probability of selection $\overline{p}$. I then estimate the parameters of the distribution of $\ln w_i'$ via 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:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} z_i \tag{7}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} (z_i - \hat{\mu})^2 \tag{8}$$

where in the case of log utility z_i is given by:[11]

$$z_i = \frac{u(w_i) - u(b_i)}{\phi(T, t_i^*) \cdot \overline{p}} + u(w_i) \tag{9}$$

Here, $\phi(T, t_i) = \left[\beta(1 - \beta^{T-t_i})\right] / (1 - \beta)$. Because w_i' is lognormally 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, I use a bootstrap procedure to calculate confidence intervals. All standard errors calculations are based on 1000 repetitions. I report 95% confidence intervals that are given by the range between the 2.5th percentile to the 97.5th percentile of the bootstrap distribution.

[11] For more general CRRA utility functions, one can show that

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

4.2 Results

Table 3 summarizes the results. Across all three estimation approaches, I consistently estimate a very high valuation of government jobs. For parsimony, I report 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. 5.491 lakh per month [95% CI: 3.402-11.072]. This estimate falls when I focus on the restricted sample, which likely reflects normal sampling variation. As we saw 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 (4.333 vs. 4.251).

How does these estimates compare to the nominal salary? In Table 4, I present 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, I convert the net present value of the income stream into the equivalent annuity. This requires fixing a discount factor, for which I use the same value I used to estimate the model. This calculation yields an annuity value of a Tehsildar post of about Rs. 0.81 lakh per month. I assume that the gap between the nominal salary and the private valuation reflects the amenity value of government jobs. This implies that at least 81% of the value of a government job is due to unobserved amenities.

4.3 Robustness

I assess the robustness of these findings to local perturbations of key features of model parameters and the data. For each of these robustness checks, I focus on the estimate with minimal requirements of the data, i.e. the lower bound estimate on w' using the full sample. The results are summarized in Figure 3.

First, I study the sensitivity of the estimate to $\overline{p}$. 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 we suppose that $\overline{p}$ is twice as large as the value observed in the data, the estimated lower bound on w' does not fall below Rs. 3 lakhs per month, and the 95% confidence interval excludes valuations less than Rs. 2 lakhs per month. This is still substantially more than the nominal value of a government job.

Next, I consider how perturbations in the risk aversion parameter affects the estimate. One might worry that the sample of individuals who selects into exam preparation is less risk averse than the average member of the population. The implied value of a government job is declines as the risk aversion coefficient also declines. But even with a 20% reduction (from 1 to 0.8), the estimate still stays above Rs. 3.5 lakhs per month.

Finally, I consider how the estimate falls when I exclude individuals with the highest reported outside wage offers 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 indeed the estimated lower bound on w' is sensitive to excluding these observations. However, even when individuals who anticipate an outside option of more than 1 lakh INR per month are excluded, the estimated lower bound on w' remains above 3 lakh INR per month.

5 Alternative Explanations

Thus far, I have interpreted the high monthly wage that rationalizes search behavior as reflecting a high amenity value of government jobs. In this section, I consider two alternative explanations for the high estimated value of government jobs: 1) that candidates are misinformed about the nominal wage in government jobs; and 2) that candidates derive value from the search process itself. I conclude that neither of these alternative explanations are compelling in this context.

5.1 Are candidates misinformed about the salary in government jobs?

Candidates may persist in studying simply because they overestimate the salary offered in government jobs. It it 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, I include a question in the survey that asks respondents to guess the monthly wage of a Tehsildar after 1 year of experience.[12] In general, candidates tend to have accurate beliefs about the initial salary. The median belief is 60,000 INR per month, which is close to the true value of 55,000 INR per month. Moreover, about 64% of respondents guessed within Rs. 20,000 of the true value, and 90% 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. But there is no evidence to suggest that individuals systematically entertain beliefs about compensation that are out of line with the official salary.

5.2 Do candidates derive value from the search process?

So far, I have assumed that the government job is the only state that has unobserved amenities. However, it is possible that candidates derive value from the search process itself, independent of the instrumental value of obtaining a government job. Candidates may value delaying starting work, living in Pune, or the lifestyle of being a student. In that case, exam preparation is not as costly as I have supposed, in which case we can rationalize the search

[12] As reported in Table 1, I also asked candidates about their salary beliefs after 10 years of experience. However, it is difficult to assess the correctness of these beliefs, since the true value also incorporate uncertainty about interim government policy changes.

behavior with a lower value for government jobs.

One way to address this concern within the scope of the model is to multiply the observed values of b_i by some constant $\alpha > 1$. This does not affect the conclusions substantially. The estimated lower bound on w' still remains above 3 lakh per month, even if I set $\alpha = 2$.

I also provide evidence from a survey experiment 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 that they would still 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 $b_t < w_t$, then there is some value of p_t below which candidates will prefer to drop out.

To test whether this is the case, I constructed a vignette experiment in which I asked a convenience sample of 50 MPSC candidates in the Peth Area 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, I randomly varied the score history of the hypothetical friend.[13] I ensured that the hypothetical friend described came from the same district as the respondent and had the same gender. This was done to maximize the likelihood that the respondents' 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. I treat responses of either A or B as a recommendation to persist.

[13] The score in the 2016 exam X_{2016} is randomly chosen from the set $\{10, 30, 50\}$. I then generate scores for 2017 and 2018 using the following AR 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 I randomly generate a main score exam from a uniform distribution between 30 and 50.

Figure 4 plots the fraction of individuals recommending that the friend drop-out as a function of the average distance to the preliminary exam cutoff score across the three scores showed in the vignette.[14] If individuals had strong non-instrumental reasons for studying, we would expect the fraction of candidates recommending dropping out to plateau at some value less than 1. This does not appear to be the case.

6 Conclusion

The value of a government job in India far exceeds the nominal wage, indicating that amenities comprise a large share total compensation. This finding has implications for policy and raises several important questions for future research.

First, even if not everyone in the economy can earn government salaries, it may be possible to democratize access to some of the amenities provided by the public sector. The estimates from this paper suggest that doing so might result in large total welfare gains. For example, if a good share of the amenity value is derived from the insurance value of a government job, then the government can increase access to this valuable amenity by providing social insurance. There is therefore substantial value in unpacking the components of the large amenity value of government jobs to identify potentially high-value policies. In this way, the rush for government jobs may provide a window into the determinants of welfare for a large section of the population.

Finally, given that government employees appear to derive much of the value of their job from the amenities, the results of this paper suggest that tying public sector employee amenities' to behavior could be a powerful source of incentives. There is some research that has begun to explore this space (e.g. Khan, Khwaja, and Olken (2019)), but given how multidimensional amenities are, we have perhaps only begun to scratch the surface of the possibilities here.

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

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.

Araujo, Felipe (2020), "Selection bias and the returns to public-sector employment." Technical report, LeHigh University.

Chetty, Raj (2006), "A new method of estimating risk aversion." *American Economic Review*, 96, 1821–1834.

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.

Holzer, Harry J., Lawrence F. Katz, and Alan B. Krueger (1991), "Job Queues and Wages*." *The Quarterly Journal of Economics*, 106, 739–768, URL https://doi.org/10.2307/2937925.

Khan, Adnan Q, Asim Ijaz Khwaja, and Benjamin A Olken (2019), "Making moves matter: Experimental evidence on incentivizing bureaucrats through performance-based postings." *American Economic Review*, 109, 237–70.

Krueger, Alan B (1988), "The determinants of queues for federal jobs." *ILR Review*, 41, 567–581.

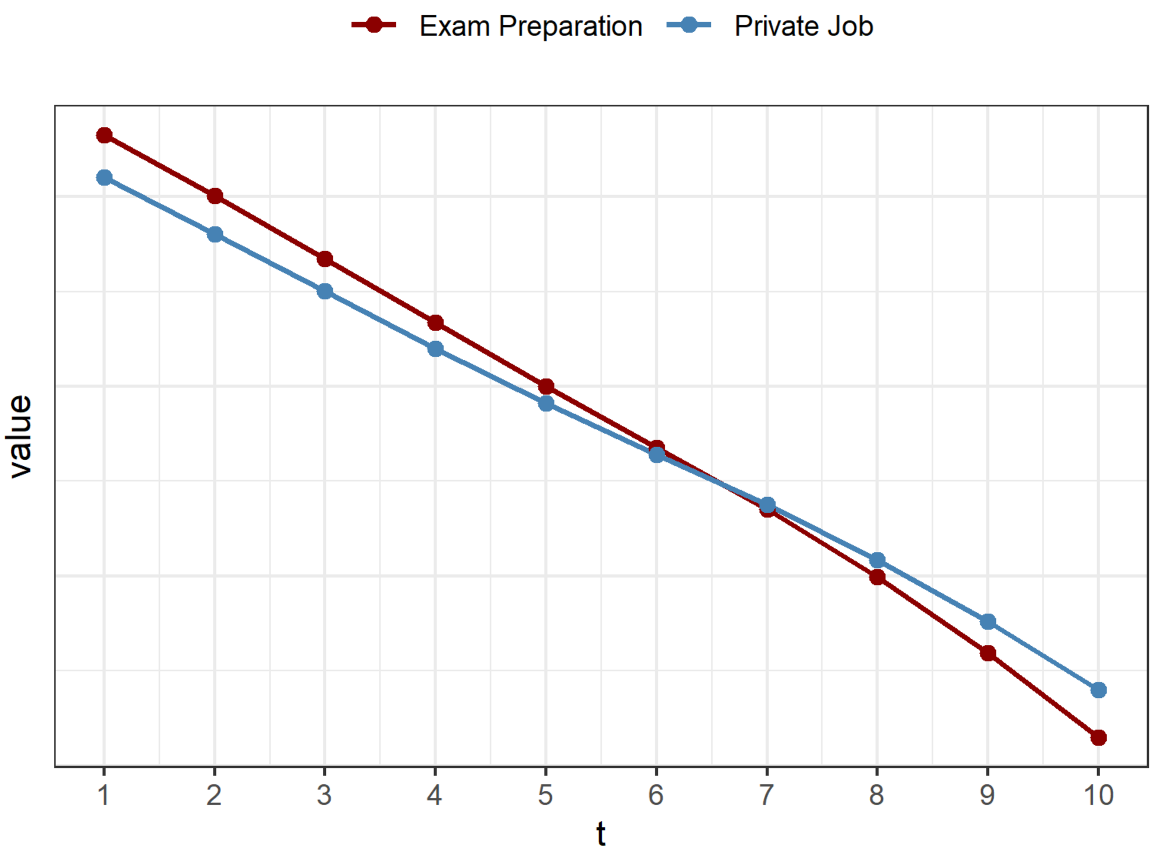
Kuipers, Nicholas (2022), "The indonesian public sector wage premium." Technical report, UC Berkeley.

Mas, Alexandre and Amanda Pallais (2017), "Valuing alternative work arrangements." *American Economic Review*, 107, 3722–59.

Stern, Scott (2004), "Do scientists pay to be scientists?" *Management science*, 50, 835–853.

7 Figures

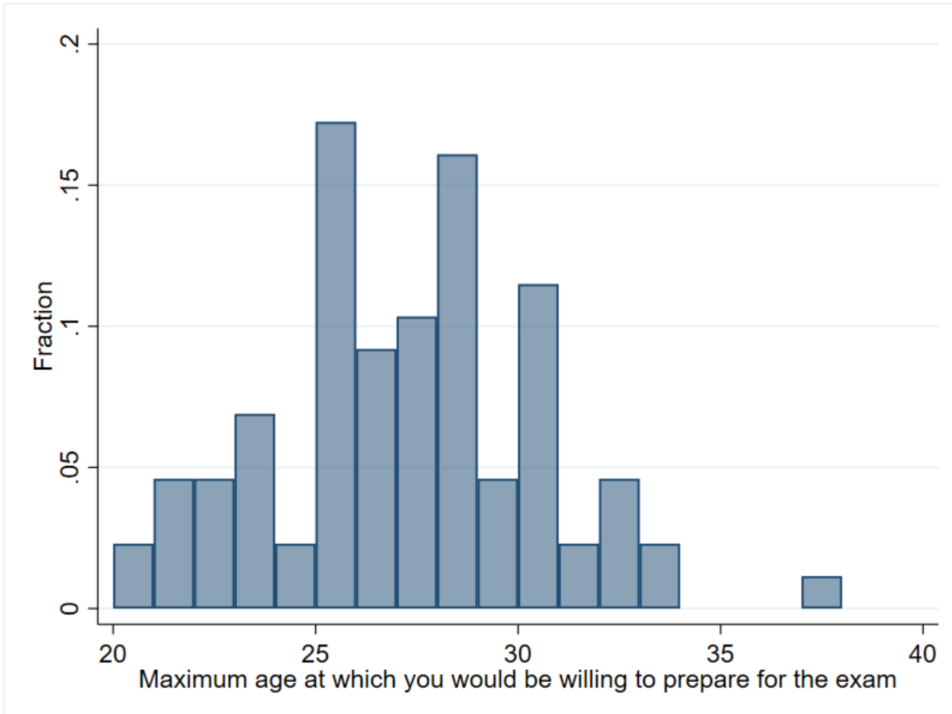
Figure 1: An illustration of the optimal stopping model



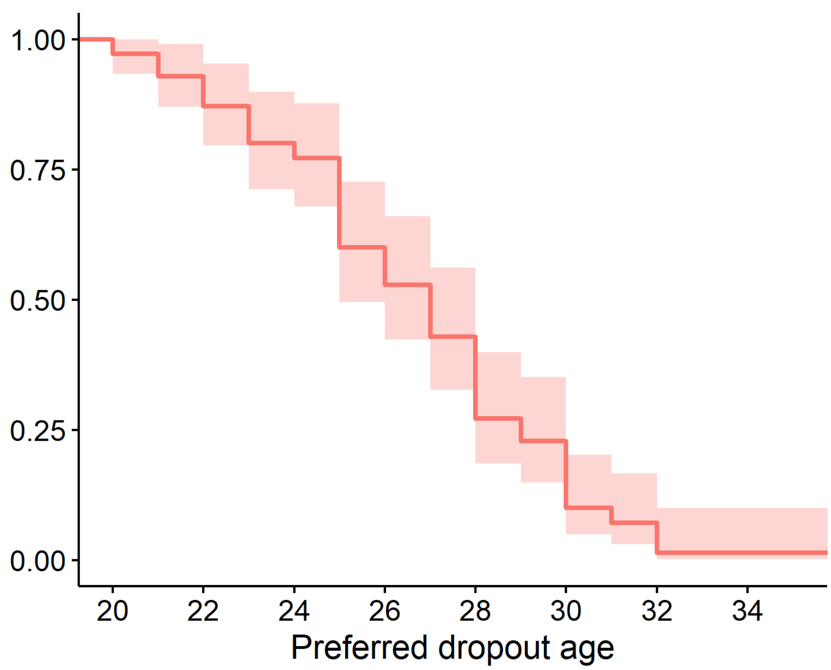
Notes: The figure plots the value functions for U_t (exam preparation) and P_t (private job) from the model for a specific set of model parameters: $b = 3000$, $w = 8000$, $w' = 30000$, $p = 0.085$, $\beta = 0.9$. The figure is meant to illustrate the optimal stopping rule: at $t = 6$, the value of exam preparation no longer exceeds the value of a private job. The value functions only cross once.

Figure 2: Distribution of the preferred dropout age

(a) Histogram

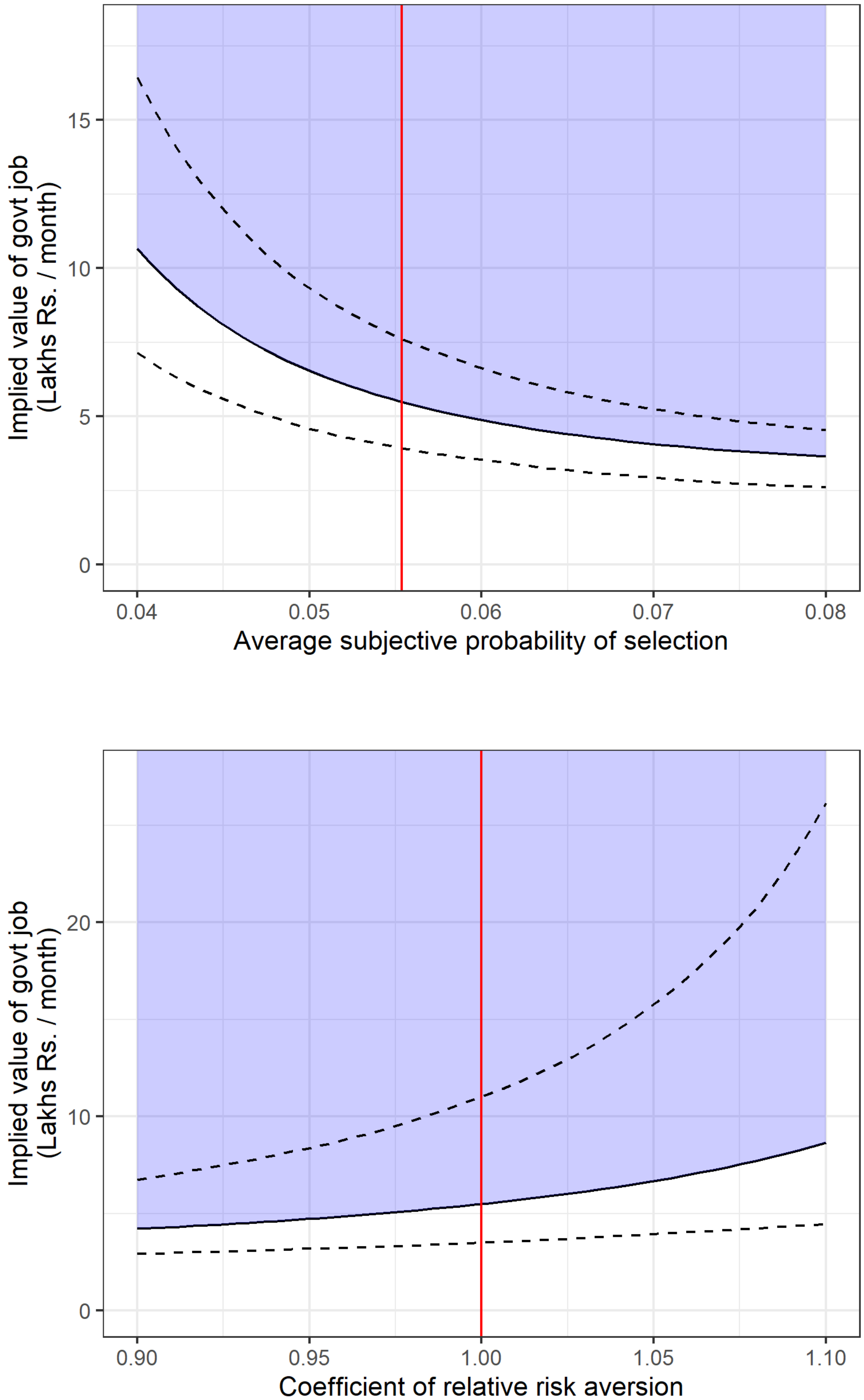


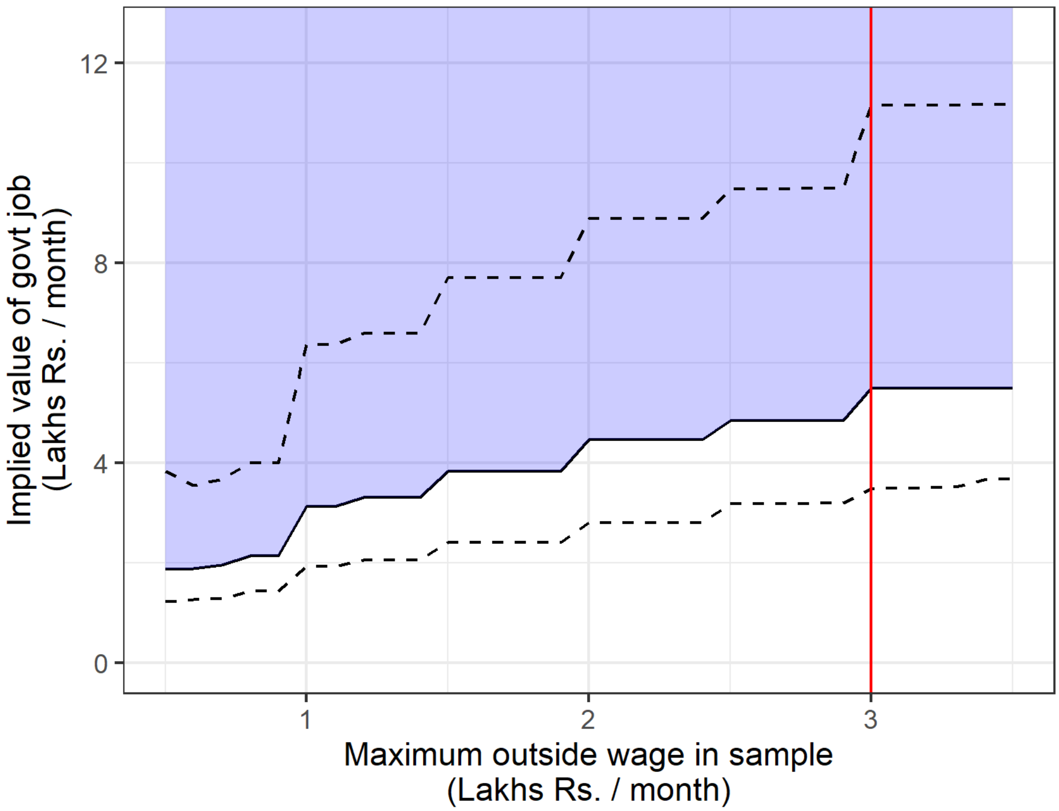
(b) Survival curve



Notes: Panels A plots a histogram of the preferred dropout age. Panel B plots a Kaplan-Meier estimate of the survival curve. The red bands show 95% confidence intervals.

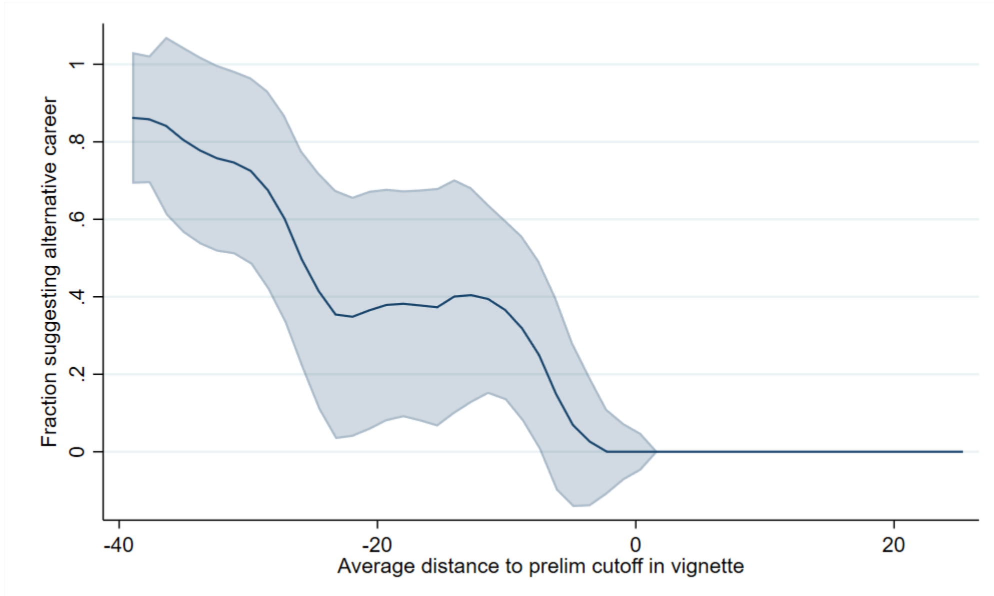
Figure 3: Sensitivity of the Estimated Lower Bound on w'





Notes: In each figure, the value of a model parameter varies along the x axis. The blue 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. This region is estimated using the moment inequality in equation (6) of the main text. The solid red line marks the values of the parameter used in constructing the main estimate. The dashed line mark the 95% confidence interval, obtained by via 1000 bootstrap samples.

Figure 4: Are candidates willing to persist no matter how low their score is?



Notes: This figure summarizes the results of the vignette experiment described in Section 5.2. 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.

8 Tables

Table 1: Summary Statistics

	Full Sample			Restricted Sample		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Persistence						
Age	24.8	2.7	148	24.6	2.7	85
Dropout Age	26.7	3.3	87	26.7	3.3	87
Demographic Characteristics						
From Pune District (0/1)	0.032	0.176	157	0.022	0.147	91
Caste Group: General Category (0/1)	0.172	0.379	157	0.176	0.383	91
Caste Group : SC / ST (0/1)	0.108	0.312	157	0.121	0.328	91
Work Experience						
Currently working (0/1)	0.006	0.080	157	0.011	0.105	91
Ever worked (0/1)	0.134	0.341	157	0.143	0.352	91
Alternative Occupation						
Alt. Occ.: Business (0/1)	0.554	0.499	157	0.582	0.496	91
Alt. Occ.: Farming (0/1)	0.299	0.459	157	0.330	0.473	91
Alt. Occ.: Wage Employment (0/1)	0.229	0.422	157	0.231	0.424	91
Expected monthly income in alt. occ. after 1 year of experience	46,850	45,750	131	44,320	40,710	77
Expected monthly income in alt. occ. after 10 years of experience	308,920	1,096,540	124	264,970	830,730	73
Income Support						
Total Monthly Expenses	8,210	1,690	157	8,570	1,760	91
Monthly transfer from home	7,810	1,920	153	8,030	1,810	88
Subjective Beliefs						
Subjective yearly pass probability	0.059	0.077	126	0.065	0.085	78
Expected monthly income as a Tehsildar after 1 year of experience	91,980	162,090	140	96,070	184,720	82
Expected monthly income as a Tehsildar after 10 years of experience	271,710	1,010,490	135	353,090	1,297,470	79

Notes: This table presents summary statistics from the Peth Area survey. The sample consists of male MPSC candidates studying in libraries located in the 411030 zip code of Pune. The restricted sample is restricted to the survey rounds in which the question on the preferred dropout age was asked correctly. All income figures are reported in INR, and rounded to the closest tens place. The alternative occupation categories are not mutually exclusive.

Table 2: Reduced Form Correlations

	(1)	(2)	(3)	(4)
$\ln b$	1.756			1.190
	(1.418)			(1.684)
$\ln w$		-0.862*		-1.000**
		(0.450)		(0.398)
Subjective p			8.925*	9.074*
			(5.026)	(5.015)
Current age	0.663**	0.653**	0.713**	0.660**
	(0.126)	(0.118)	(0.108)	(0.104)
Reservation FE	X	X	X	X
Observations	79	71	73	63

Notes: Table presents correlations between the preferred dropout age (the dependent variable) and the exogenous model parameters. All specifications include fixed effects for the respondent's reservation category. Heteroskedasticity-robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Estimates of the Value of a Government Job

Estimator	Source of unobserved heterogeneity	Parameter	Parameter Estimate (Rs. Lakhs / month)	
1 - Moment inequality	Probability of selection	Lower bound on w'	5.491	4.251
			[3.402, 11.072]	[2.310, 9.918]
2 - GMM	Probability of selection	w'	–	4.333
			–	[2.404, 10.277]
3 - MLE	Value of government job	μ	–	0.685
				[0.001, 1.671]
		σ^2	–	3.478
				[2.195, 5.854]
		$E[w_i']$	–	11.289
				[3.291, 88.295]
		Median w_i'	–	1.984
				[1.001, 5.320]
Sample			Full	Restricted
N			120	70

Notes: Table presents estimates using three different estimation strategies. For each estimate, I provide 95% confidence intervals based on 1000 bootstrap samples in brackets. The confidence intervals do not adjust for clustering. The restricted sample excludes individuals with a missing value of the preferred dropout age.

Table 4: 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 NPV, starting from age 20	Rs. 14,242,460
Annuity equivalent	Rs. 81,429 per month

Notes: These calculations are based on the Maharashtra 7th Pay Commission pay matrix and policies.

Appendix

A Theory Appendix

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

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

Proof. We'll 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

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

which is clearly decreasing in t.

Next, I will 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

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

or, equivalently,

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

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 iff

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

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

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

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 iff

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

This is clearly true since: 1) we established that $G_t - G_{t-1} < 0$ since G_t is decreasing, and 2) 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.* $P_t > U_t$ *for some* t. *Furthermore, taking private employment is an absorbing state, i.e.* $P_t > U_t \implies P_{t+s} > U_{t+s}$ *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 1), 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*

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

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

$$G_t = u(w') \frac{1 - \beta^{(T-(t-1))}}{1-\beta} \tag{A.3}$$

$$P_t = u(w) \frac{1 - \beta^{(T-(t-1))}}{1-\beta} \tag{A.4}$$

Setting P_t equal to U_t at t^* yields:

$$u(w) + \beta P_{t^*+1} = u(b) + \beta p G_{t^*+1} + \beta(1-p) P_{t^*+1} \tag{A.5}$$

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

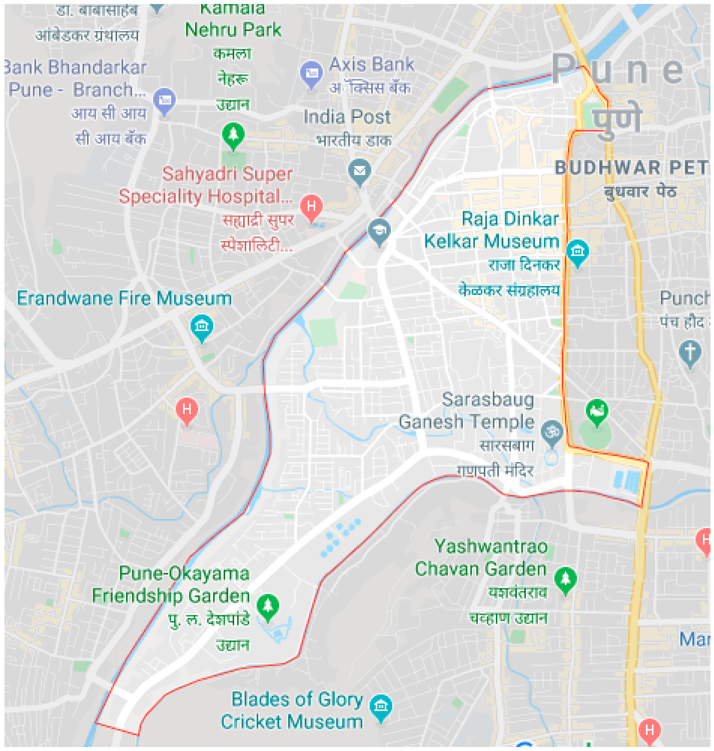
$$t^* = T - \frac{1}{\ln \beta} \ln \left[1 - \frac{(1-\beta)[u(w) - u(b)]}{\beta p[u(w') - u(w)]}\right] \tag{A.6}$$

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 less than zero, which is outside the domain. Solving for when $t^* \leq 0$ yields the condition $u(w) - u(b) \geq \frac{\beta(1-\beta^T)}{1-\beta} p[u(w') - u(w)]$.

□

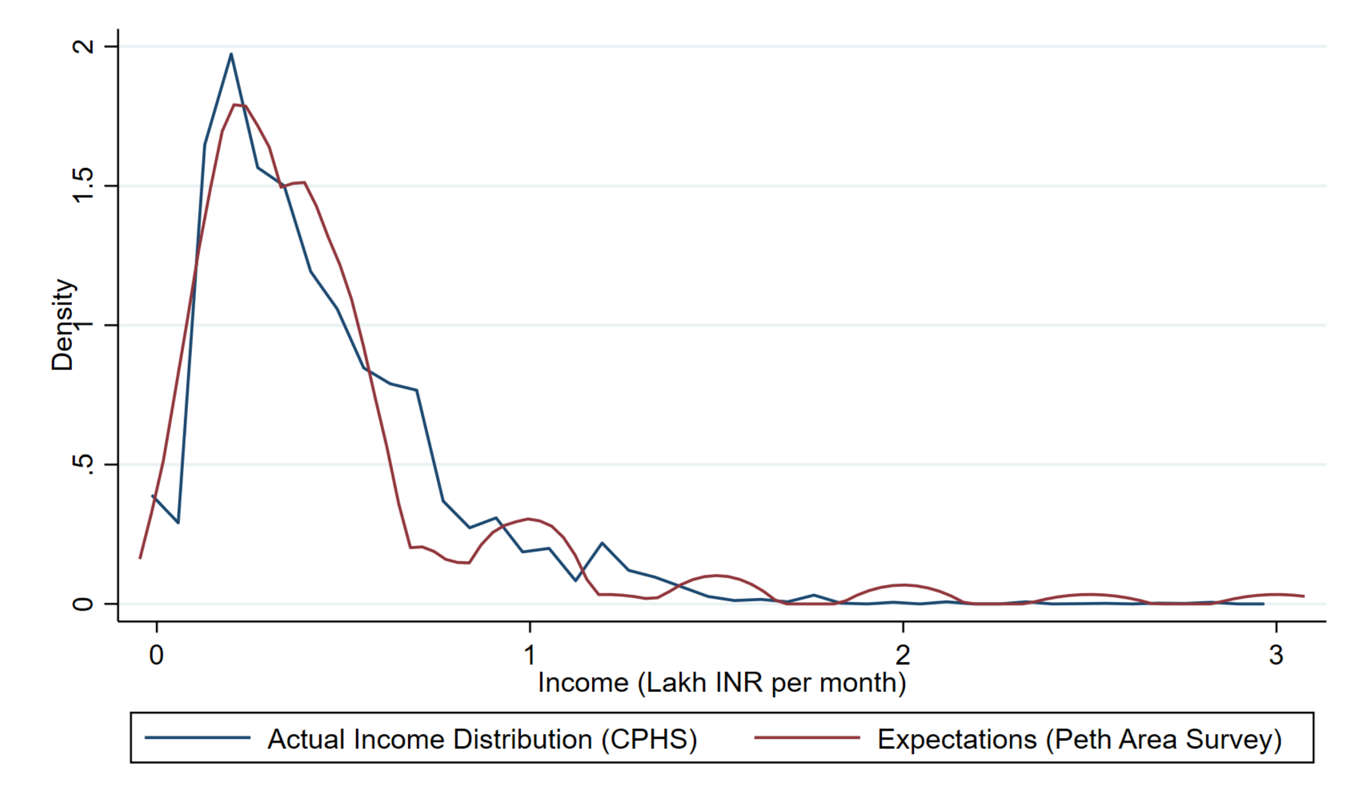
B Additional Figures and Tables

Figure B.1: A Map of the Peth Area, Pune



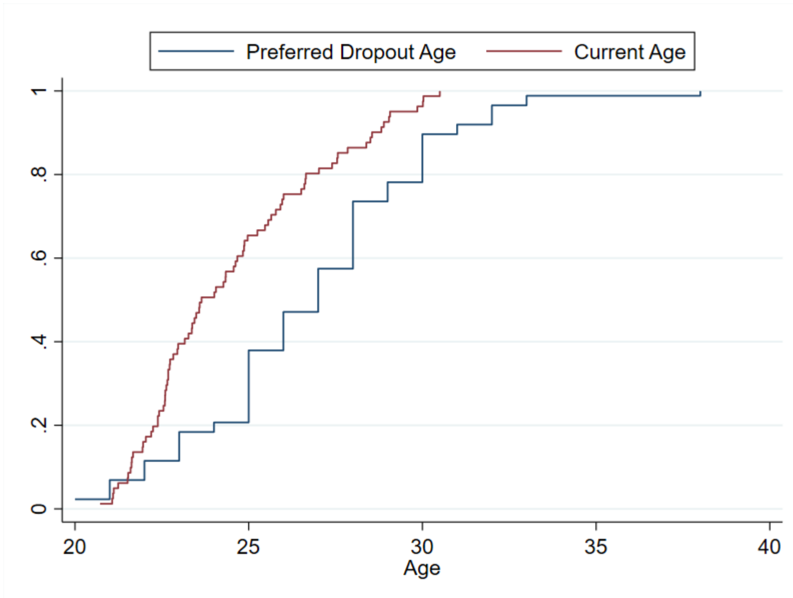
Notes: The Peth Area is marked by the shaded region. The boundary marks the edges of the 411030 pin code, which define the boundaries of the sampled area.

Figure B.2: Comparing the actual and expected distribution of earnings



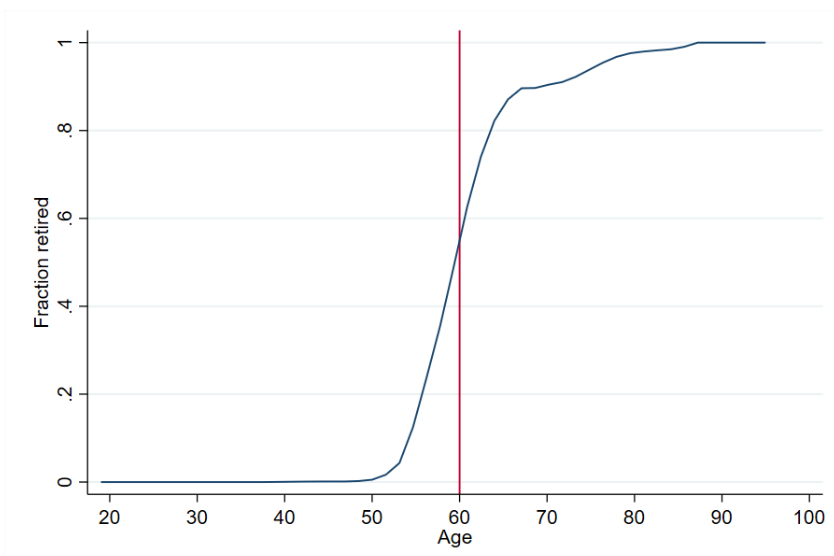
Notes: The figure plots kernel densities of both the empirical income distribution and the expectations recorded in the Peth Area Survey. The expectations data comes from the Full Sample. It is the expected income in the outside option 1 year after leaving exam preparation. The actual income distribution comes from the CMIE Consumer Pyramids Household Survey. The CPHS sample is restricted to male college graduates between the ages of 25 to 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 expectations data. For legibility, the long right tail of the CPHS distribution has been truncated.

Figure B.3: A test of bias in the preferred dropout age



Notes: The red line plots the cumulative distribution function (CDF) of the respondents' current age. The blue line plots the CDF of the stated dropout age. Sample restricted to men in the restricted sample who provided a valid measure of their preferred dropout age.

Figure B.4: Age of retirement for male college graduates in Maharashtra



Notes: Figure is based on data from the CMIE Consumer Households Pyramids Survey. I use the 2019 Wave 1 round. The figure plots a non-parametric estimate of the fraction of male college graduates in Maharashtra that are retired conditional on their age. The red line marks the age of retirement used in the estimation of the model.

Table B.1: 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%

Notes: The table summarizes statistics from the implementation of the survey. 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 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.