The Selection of Talent

Experimental and Structural Evidence from Ethiopia*

Girum Abebe[†], Stefano Caria[‡], Esteban Ortiz-Ospina[§]

November 3, 2017

JOB MARKET PAPER

[Latest Version]

Abstract

We study how search frictions in the labour market affect firms' ability to recruit talented workers. In a field experiment in Ethiopia, we show that a small monetary incentive for making a job application enables an employer to attract more talented applicants. The effect is driven by workers with lower incomes and weaker outside options. It is similar in size to the increase in applicant quality generated by a second intervention that doubles the wage offer. Our findings are consistent with a model in which talented jobseekers face large application costs and credit constraints. We structurally estimate this model and find that the cost of making an application is large (on average 9-13 percent of the monthly wage), and is positively correlated with jobseeker ability. An estimated 30 percent of individuals are unable to pay this cost because of credit constraints. For the average firm in this market, we find that the application incentive has an internal rate of return of 11 percent. However, in a second experiment, we show that local firm managers underestimate these positive impacts, explaining why the use of application incentives is limited.

*We are grateful to Abi Adams, Arun Advani, Nava Ashraf, Oriana Bandiera, Vittorio Bassi, Stefano DellaVigna, Marcel Fafchamps, Erica Field, Simon Franklin, Douglas Gollin, Clement Imbert, Supreet Kaur, Philipp Kircher, Jeremy Magruder, David McKenzie, Guy Michaels, Paul Niehaus, Imran Rasul, Chris Roth, Simon Quinn, Yared Seid, Alemayehu Seyoum Taffesse, Chris Woodruff for helpful comments and to Koen Maskaant and Alemayehu Woldu for outstanding research assistance. We acknowledge funding from the IGC (Project No. 1-VCC-VETH-VXXXX-32304). The project would not have been possible without the support of Rose Page, Simon Quinn, CSAE and EDRI.

[†]Ethiopian Development Research Institute. Email: `girum.abebe@edri.org.et`.

[‡]University of Oxford. `stefano.caria@qeh.ox.ac.uk`, Web: www.stefanocaria.com.

[§]University of Oxford. Email: `esteban.ortiz-ospina@bsg.ox.ac.uk`.

1 Introduction

Hiring talented workers is key for firm productivity and growth. However, attracting the best workers can often be difficult for firms. When there are frictions in the labour market, the pool of available talent is limited because jobseekers do not have the information, time and resources to apply for all suitable jobs. Offering a competitive wage may thus not guarantee that the best candidate will apply for the position. Unless firms can find alternative ways to attract and select the best workers, talent will be misallocated. This can generate large costs for firms and for the whole economy (Hsieh et al., 2013; Hoffman et al., 2015; Algan et al., 2017).

We study how labour market frictions affect firms' ability to make good hires. Our setting is a developing country where jobseekers have limited financial resources and often take informal, short-term jobs to generate income (Abebe et al., 2016). In this context, the opportunity cost of the time and money required to secure a job in a formal firm can be substantial.[1] Recent research also shows that financial constraints are common among poor jobseekers in high-income countries (Card et al., 2007, 2010; Phillips, 2014). Yet, we have no evidence, from either developed or developing economies, on what firms can do to recruit the most talented individuals in these markets.

In a field experiment in Addis Ababa, Ethiopia we document how the number and quality of applicants for a clerical job changes when applications costs are decreased by offering a monetary incentive. In a second treatment, we double the wage offer but do not provide any financial incentive for applications. The incentive is worth 4 .5 USD, while, for the average jobseeker, the expected value of the high wage offer is worth about 100 USD. We randomise the offer of these two treatments over the sample of individuals who call to inquire about the position. We collect data on all potential applicants during this first phone call and in a follow-up phone interview. Further, we measure the quality of the individuals who apply for the job through a battery of personnel selection tests that capture cognitive ability, non-cognitive ability and relevant work experience. These tests are reliable predictors of work performance and are used by firms worldwide (Heckman et al., 2006; Autor and Scarborough, 2008; Hoffman et al., 2015).[2]

We find that the application incentive increases application rates by a significant

[1] This will typically entail preparing the application materials and visiting the firm, possibly multiple times, to deposit the materials and complete screening tests and interviews.

[2] We use the Raven and Stroop tests for cognitive ability (Schmidt and Hunter, 1998). For non-cognitive skills we administer the Big-5 personality test and the Grit scale (John and Srivastava, 1999; Duckworth et al., 2007). We also collect detailed data on work experience and economic preferences.

11.5 percentage points.[3] This effect corresponds to a 28 percent increase over a control group application rate of 41 percent. It amounts to about two thirds of the increase in applicants that we observe when we double the wage for the same position.

Our most important finding is that the application incentive *improves* the quality of the applicant pool. In particular, cognitive ability is significantly higher, at the mean, at the top (90th and 75th percentile) and at the bottom (25th percentile) of the distribution. The results of a statistical test comparing the two distributions indeed suggest that cognitive ability in the application incentive group stochastically dominates cognitive ability in the control group. The magnitude of the effect is also substantial. The number of top applicants almost doubles.[4] Further, the average Raven test score increases by about .1 of a standard deviation, which corresponds to 1.2 additional correct answers in the test. This effect is similar to those found in related studies. For example, Dal Bó et al. (2013) estimate that a 30 percent wage increase raises the Raven score of Mexican applicants for a public-sector job by about .5 correct answers. In our experiment, doubling the wage also improves the cognitive ability of the applicant pool, raising the Raven score by about .7 correct answers. We do not find significant changes in applicants' non-cognitive ability or work experience related to the position.

The improvements in quality are stronger among jobseekers who are currently unemployed, less experienced, and among women. These are groups who have, on average, worse outcomes in the labour market and lower incomes. This suggests that the application incentive does not increase quality at the cost of attracting individuals who do not value the position highly. On the contrary, this intervention mostly taps from the pool of talent of low-income jobseekers who stand to benefit the most from the job. To explore this point further, we generate an individual measure of the net present value of the experiment's job using a simple dynamic framework of job search and a forecast of the wage that each individual would be paid if employed in the market. We obtain this forecast with a Post-LASSO estimator to avoid overfitting (Belloni et al., 2014). We find that the increase in quality is significantly larger for the group of respondents that values the job the most.

We rule out several potential explanations for our findings that are not related to application costs. First, we show that test effort is not significantly different across treatment groups. To study this, we administer a test that requires effort, but very little ability. We find no significant differences in performance in this test. Second, we provide evidence that the application incentive does not make the position more salient in the

[3] We registered the study and the analysis in a pre-analysis plan.

[4] We define top applicants as individuals with cognitive ability above the 90th percentile of the distribution in the control group.

mind of prospective applicants. We proxy salience using the accuracy with which jobseekers recollect information about the job (Botta et al., 2010; Santangelo and Macaluso, 2013). We show that, one month after the initial phone call, treated individuals recollect this information with similar accuracy as control individuals, suggesting that the job has similar salience in the minds of control and incentive group individuals. Third, we show that the application incentive does not affect subjects' expectations about how long it will take them to find a new job, or the wage that this new job would pay.[5] It is thus unlikely that subjects (wrongly) infer information about themselves or about the labour market from the offer of the incentive. Finally, we find that the incentive is associated with only minor changes in beliefs about various attributes of the job, such as days of holiday. We estimate that these changes can account for only 5 percent of the total effect of the application incentive.

Using a simple model of application decisions, we show formally that the incentive attracts better applicants only in markets where higher quality jobseekers face larger application costs. We structurally estimate this model to quantify the size of application costs and their correlation with jobseeker ability. We identify these key parameters using the exogenous variation generated by the experiment. We use a classical minimum distance estimator and we bootstrap the estimation to perform inference (Wooldridge, 2010). The fit between the simulated and empirical moments is good. For example, we fit all application rates with less than one percentage point of error. Further, the model can match a key non-targeted moment – jobseekers' assessment of the probability of receiving a job offer – and replicates closely non-targeted patterns of the data, for example the uniform shift in the distribution of quality.

We find that application costs are large and strongly correlated with jobseeker quality. For the group of individuals who value the job the most, the correlation between cost and quality is .46. The magnitude of application costs is also substantial. At the mean, application costs amount to 13 percent of the monthly wage and 38 percent of the estimated net present value of of the job, for the high value group. We also estimate that a large share of the sample – about 30 percent – is credit constrained. This is likely to result in a large misallocation of talent in the economy. Using an estimate of the average value of cognitive ability for firms (Bowles et al., 2001), we calculate that for the average firm in this market the internal rate of return (IRR) of the application incentive is 11 percent. This is above market interest rates and passes standard hurdle rates. Through counterfactual policy analysis, we also show that the IRR increases sub-

[5]We also show that the incentive does not affect jobseekers' assessment of the probability of being offered the experiment's job. We discuss the implications of this finding in Section 3.

stantially when the incentive is either (i) targeted to marginal applicants or (ii) offered conditional on a good performance in the selection test.

These results leave open two questions. First, what drives the correlation between costs and ability? Second, if applications incentives have positive returns for firms, why are they not commonly used in this market? To answer the first question, we present evidence for a selection mechanism that can generate a positive correlation between costs and jobseeker ability. We rely on a unique high-frequency panel dataset on young jobseekers in Addis Ababa collected by Abebe et al. (2016). This dataset has information on labour market outcomes, cognitive ability (measured through a Raven test), and two measures of costs: distance from the city centre (which proxies for the monetary cost of making an application) and savings (which proxy for the opportunity cost of money). We show that high quality subjects living near the city centre or having high savings ('low-cost') are more likely to stop looking for work than individuals with the same quality who live further away or have lower saving. The average quality of low-cost individuals who search for employment thus deteriorates over time, generating a positive correlation between cost and quality.

To answer the second question, we run a second experiment with a sample of firms in Addis Ababa that are recruiting clerical workers. First, we confirm that the application incentive attracts better workers by showing that firm managers rank the anonymised CVs of applicants from the incentive group above those of control group applicants. The task is incentivised: the higher the rank, the higher the probability that we will invite that individual to make an application at the manager's firm. Second, we show that firm managers substantially underestimate the positive effect of the incentive on the quality of applicants. To do this, we elicit managers' incentivised forecasts of the effects of this intervention (DellaVigna and Pope, 2016). On average, managers expect this intervention to *decrease* applicant quality. Misinformation about the positive effect of the incentive is thus a plausible explanation for why the use of application incentives is limited in this context (Hanna et al., 2014).

Our results make several contributions to the literature. First, we provide the first worker selection experiment that manipulates application costs. Some recent experiments in developing and developed countries have manipulated the wage, or workers' expectations about the wage (Dal Bó et al., 2013; Deserranno, 2014; Ashraf et al., 2014; Belot et al., 2017). These studies find that higher wages attract more and better applicants. In the US, Flory et al. (2014) show that pay schemes that rely on competition among workers discourage female jobseekers, while Mas and Pallais (2016) infer subjects' willingness to pay for flexible work schedules from application decisions in a field experiment. Our results highlight that, when jobseekers find it costly to participate in the labour market, firms may may hire better workers if they reduce application costs.

Second, we provide a structural estimate of the magnitude of search costs in an urban labour market and highlight a new mechanism that can lead to the misallocation of talent. This contributes to a recent, growing literature that studies the allocation of talent. Previous studies have focused on the role of discrimination (Hsieh et al., 2013), migration costs (Bryan and Morten, 2015; Imbert and Papp, 2016; Lagakos et al., 2017), housing market failures (Hsieh and Moretti, 2015), and corruption (Weaver, 2016). We provide original empirical evidence on the importance of search frictions – in particular, high application costs and credit constraints at the top of the ability distribution. These frictions have been the focus of several theoretical papers, but direct evidence on their magnitude has been limited to date (Marimon and Zilibotti, 1999; Rogerson et al., 2005; Galenianos et al., 2011).

Finally, our results have important implications for active labour market policies, in particular job search assistance. The recent literature has focused on the impacts of these policies on workers' employment outcomes.[6] Our findings suggest that these policies have the potential to improve the pool of talent available to firms. This motivates the design of new evaluations that assess whether job search support improves the allocation of talent in frictional labour markets. Further, our results highlight that managers do not have accurate beliefs about the returns of different recruitment practices and may thus fail to optimise firms' recruitment policies (Hanna et al., 2014; DellaVigna and Pope, 2016). Providing information to managers may thus be a cost-effective intervention in this context.

The rest of the paper is organised as follows. Section 2 describes Addis Ababa's labour market. We present a model of job application decisions in Section 3. Section 4 describes the experimental design and the data. Section 5 discusses the impacts of the two interventions. We present the structural estimation in Section 6. Section 7 studies what drives of the correlation between costs and quality, and analyses the data from the second experiment.

2 Context

Ethiopia is the second most populous country in Sub-Saharan Africa and its capital Addis Ababa has a total population of approximately three million people. The country is undergoing a fast process of structural transformation, characterised by rapid urbani-

[6] Abebe et al. (2016) and Franklin (2016) provide experimental evidence for Ethiopia. Crépon and Van den Berg (2016) and McKenzie (2017) offer recent reviews of this literature.

sation and sustained economic growth. Addis Ababa is at the forefront of this process. The number of jobs in the city has grown from 740,000 to 1,245,000 between 1999 and 2013 (CSA, 2000, 2014). At the same time, a large number of migrants and young people have joined the labour market. Employment rates have thus stayed constant, at just above 50 percent.

In this section we describe the labour market in Addis Ababa, from the point of view of both firms and workers. To do this, we complement existing datasets with an original survey that we collect for this study. The sample is composed of 196 firms that advertised a vacancy for a clerical job during a period of six weeks in 2017.[7] In each firm, we request to interview the head of the selection committee – typically the head of the HR department or the firm's CEO. We use this sample of managers to run the second experiment reported in the paper.[8]

2.1 Finding a worker in Addis Ababa

Finding the right worker can be challenging for firms in Addis Ababa. In our survey, we ask managers to report the most important HR problem experienced by their firm. Finding workers with adequate skills is the most frequently mentioned challenge. As shown in Figure 1, about 35 percent of managers considers this to be the most pressing HR problem for their firm. Retention, absenteeism, motivation and conduct are all mentioned less frequently than hiring. We then ask managers what would be the best strategy to improve hiring outcomes. About 60 percent of them answer that offering higher wages would be the most effective way to improve the quality of recruits. Application incentives are mentioned rarely and, in practice, they are not frequently used by firms in the city.

< Figure 1 here. >

The firms in our sample hire workers on a frequent basis. In the two months preceding the interview, the average hiring rate among these firms was between 2.2 and 2.8 percent. Similarly, Abebe et al. (2016) document annual hiring rates of about 19 percent for a sample of 496 firms in Addis Ababa. Hiring occurs both to expand the workforce and to replace workers who leave the firm. In our sample, separation rates

[7] We screen all vacancies advertised on the main job-vacancy boards or in a popular newspaper insert. To identify clerical jobs, we categorise each vacancy according to the 2010 Standard Classification of Occupations of the US Bureau of Labor Statistics. For the full list of occupations included in the survey see Table A.1.

[8] During the interview, each manager first completes the CV-ranking and forecast tasks, which we describe in detail in the sections 5 and 7, and then answers the survey questions about his or her firm.

were between 1 and 2.2 percent in the two months preceding the interview. These flows are moderately smaller than those reported by firms in the US. Over the period 2007-2016, the average monthly hiring rate among US firms was 3,4 percent and the average monthly separation rates was 3,3 percent.[9]

Hiring is also costly for firms, in terms of both money and time. Among firms in our sample, average recruitment costs amount to about 104 USD and 18 hours of staff time (worth about 40 USD when valued at the mean wage of HR managers in the same firms). Total costs corresponds approximately to one month of salary for one of the high-wage jobs in the experiment. These costs do not vary substantially with the number of applicants (many of the costs, such as those related to advertising and developing tests and interviews, are fixed). Managers estimate that considering one more application entails no further monetary costs and would not require more than one hour of staff time.

Firms screen workers by assessing their CVs and by administering written tests and interviews. Educational qualifications, GPA and previous work experience are the most important variables that managers consider when they assess candidates' CVs. Firms often require applicants to deposit their CV and the other application materials in person. Written tests and interviews are also used frequently. Both interviews and written tests are used to assess general cognitive ability, specific technical knowledge, and personality traits.

2.2 Finding a job in Addis Ababa

Jobseekers spend substantial amounts of money and time to find work in Addis Ababa. Using self-reported expenditure data, Abebe et al. (2016) estimate that the monetary cost of searching and applying for jobs amounts to one quarter of weekly expenditure for individuals who are actively looking for employment. To pay for these costs, jobseekers need to frequently take up informal, short-term jobs, which are relatively easier to secure. These challenges are described in detail in Abebe et al. (2016). Here we report one additional piece of descriptive evidence: jobseekers apply to only a small fraction of available vacancies. In our sample, for example, the average unemployed person in the control group completes 1.8 job applications in 30 days. On the other hand, when we screened job boards and newspaper to sample firms that were hiring clerical workers, we were able to find at least 30 vacancies per week. This low number of applications is consistent with the existence of financial constraints that limit job search intensity.

[9]These figures can be retrieved from the online data repository of the Job Openings and Labour Turnover Survey.

3 A simple model of job application decisions

We propose a simple model of application decision that captures two key frictions in job search: application costs and uncertainty about the probability of being offered the job. The model characterises the effects of the interventions on application rates and the quality of the applicant pool. It predicts that both interventions will increase application rates. Further, it shows that the application incentives increases the quality of the applicant pool only when ability and application costs are positively correlated among jobseekers.

3.1 Set up

Jobseeker Characteristics. We consider a set of individuals deciding whether to apply for the experiment's job. For tractability, we focus on the large-number case and assume that these jobseekers form a continuum of unit measure.

Jobseekers differ in terms of their quality (noted T in what follows), as well as in terms of the benefit that they derive from being offered the job (noted B). Heterogeneity in T captures differences in productivity, while heterogeneity in B captures differences in outside options. To fix ideas, it is helpful to think of T as the score on the Raven test (a reliable predictor of worker performance) and of B as the monetary net present value of being offered the job (where a negative net present value translates into $B = 0$, since being offered the job does not require jobseekers to *take* the job). Indeed, these are the empirical counterparts that we use for estimation, as described in Section 6.

Jobseekers who wish to apply must incur a cost (noted C) which we allow to be heterogeneous across the population. C is the net opportunity cost of applying for the job, that is, the economic value of all the things that jobseekers have to give up in order to apply (typically both money and time). This cost is heterogeneous for two reasons. First, the time and money required to make the application differ across jobseekers (e.g. jobseekers who live farther away from the application centre have to pay a more expensive bus fare). Second, the value of time and money differs according to the circumstances of the jobseeker (e.g. poorer jobseekers will find it relatively more expensive to pay the same bus fare compared to jobseekers with better financial resources). As we discuss in more detail in Section 6 and 7, this will play an important role in our interpretation of the experiment's outcomes. We also allow C to be negative. This captures the fact that some people may derive a net benefit from attending the testing sessions, independently of getting the job (e.g. because of the value of networking, or because they learn something valuable about the market).

Selectivity. Jobseekers make application choices on the understanding that they will get the job if $T > a$, where a captures the perceived selectivity of the application process. Here we treat a as a fixed parameter, which we are later going to estimate, and which we assume is common to the whole relevant population of jobseekers.[10] Our assumption is equivalent to saying that a jobseeker will get the job if they score sufficiently high on the Raven test. This is consistent with the fact that cognitive ability is the main criterion for worker selection in the experiment (see Section 4 for more details).

There are two important implications that follow from this assumption. First, we allow workers to have an incorrect perception of selectivity. This is in line with the empirical literature on overconfidence and biased beliefs (Malmendier and Tate, 2015; Spinnewijn, 2015; Hoffman and Burks, 2017; Abebe et al., 2016). It is also consistent with the data on jobseeker beliefs which we collect as part of the experiment: jobseekers in our sample hold overly optimistic beliefs about the probability of a job offer (we discuss these beliefs in Section 5).

Second, we assume that a does not change with treatment. This is motivated by the empirical observation that the interventions do not change jobseekers' assessment of the probability of getting the job. This failure to predict the increased competitiveness of the selection process is consistent with the results of a beauty contest task which we administer to all applicants. This task shows that 80 percent of applicants are not strategically sophisticated (see Crawford et al. (2013) for a detailed discussion of strategic sophistication).

Information. Jobseekers are uncertain about whether they will be offered the experiment's job if the make an application. This uncertainty stems from the fact that they do not directly observe T. However, jobseekers do observe their other characteristics, including costs C and the benefit of getting the job B, which are informative about T.[11] Jobseekers' beliefs about the probability of getting the job are thus a function of these characteristics and of perceived selectivity a.

[10] In an alternative framework, a could be derived from equilibrium conditions pinned down by beliefs about the number of vacancies. This is indeed an approach that has been explored in the literature on selection from endogenous applications (Alonso, 2016; Jewitt and Ortiz-Ospina, 2016).

[11] B is *net* present value of the experiment's job. Thus, to calculate B, jobseekers need to know (i) the wage offered for the experiment's job and (ii) the wage that they would be paid in expectation if they got another job. We assume that jobseekers are aware of (ii) because they can acquire information about what other people with similar characteristics (education, experience, etc...) earn in the market. The same characteristics, however, are insufficient to determine with certainty whether the jobseeker is going to pass any particular selection test.

Application choices. We stipulate that a jobseeker will apply for the job if and only if the expected value of the application is greater than the cost:

$$\Pr(T > a|C = c, B = b) \times b \geq c. \tag{1}$$

Distributional Assumptions. We make the following assumptions about the distribution of the variables in the model.

Assumption 1 The benefit from receiving a job offer is given by

$$B \in \{b_1, b_2, ..., b_n\} \text{ where } b_z \geq 0 \text{ for } \{z = 1, 2, ..., n\}.$$

Assumption 2 Conditional on $B = b_z$, quality T and application costs C follow a bivariate normal distribution characterised by

$$\begin{pmatrix} T_z \\ C_z \end{pmatrix} \sim \mathcal{N}\left[\left(\begin{array}{c} \mu_{T_z} \\ \mu_{C_z} \end{array}\right), \left(\begin{array}{cc} \sigma_{T_z} & \sigma_{CT_z} \\ \sigma_{CT_z} & \sigma_{C_z} \end{array}\right)\right] \; for \; \{z = 1, 2, ..., n\}.$$

Throughout the rest of the paper we use the same notation introduced in Assumption 2. That is, we use sub-indices to denote quality and costs conditional on B-types.

3.2 Solving the model

In this model, application choices are fully characterised by application costs. First, for all $B = b_z$, individuals with cost $c_z \leq 0$ apply with probability one as the benefit from receiving an offer is (weakly) positive. Second, for all $B = b_z$, individuals with $c_z > b_z$ do not apply for the position. Finally, if $\frac{c_z}{b_z} \in (0, 1)$, then there is a level of cost for which jobseekers are indifferent between making an application or not. That is, there is a level c_z^* such that

$$\Pr(T_z > a|C_z = c_z^*) = \frac{c_z^*}{b_z} \tag{2}$$

We provide two propositions that show the existence and uniqueness of this cut-off level of cost c_z^*, for $\{z = 1, 2, ..., n\}$. These propositions rely on the following two definitions, which we use throughout:

(i) The relative cost curve, $k(c_z) \equiv \frac{c_z}{b_z}$

(ii) The job offer curve, $\alpha(c_z) \equiv \Pr(T_z > a|C_z = c_z)$.

Proposition 1 (Cut-off existence)**.** *For $B = b_z > 0$, there is at least one cost level c_z^* such that $0 < c_z^* < b_z$ and $\alpha(c_z^*) = k(c_z^*)$*

We provide formal proofs for all propositions in the Appendix. Here, we summarise the intuition of each proof. Proposition 1 can be established by comparing the job offer curve and the relative cost curve at low and high values of c, for all $B = b_z$. When costs are close to zero, the job offer curve lies *above* the relative cost curve. Workers with costs close to zero thus apply for the job. On the other hand, when the cost of making the application approaches the value of the job b_z, the job offer curve lies *below* the relative cost curve. Workers facing this level of costs do not apply for the job. Since both curves are continuous, this reasoning implies that they cross at least once in the interval between 0 and b_z. In other words, there is at least one level of costs at which workers stop applying for the position.

Proposition 2 (Cut-off uniqueness)**.** *Suppose $\rho_z < \frac{\sqrt{2\pi}\sqrt{1-\rho_z^2}\sigma_{C_z}}{b_z}$ for $B = b_z > 0$. Then there is exactly one cost level c_z^* such that $0 < c_z^* < b_z$ and $\alpha(c_z^*) = k(c_z^*)$*

Proposition 2 follows from the fact that, provided ρ is sufficiently small, increasing costs always reduces the expected value of making the application. Thus, there will be only one level of costs at which the expected value of the application is zero and the two curves cross. The condition on ρ rules out cases in which the expected value of the application turns positive for people with very high application costs, because they have a high probability of getting the job. In the empirical analysis, we confirm that the structural estimates of the correlation ρ satisfy this condition.

Propositions 1 and 2 enable us to characterise application choices on the basis of application costs. If application costs are negative, then workers apply with probability one; and if application costs are larger than the benefit from being offered the job, then workers apply with probability zero. Otherwise, if application costs are positive but smaller than the benefit of being offered the job, then workers apply if and only if their costs are below the threshold c_i^* for which $\alpha(c_z^*) = k(c_z^*)$. In other words, for each $z = \{1, 2, ..., n\}$, individuals apply as follows:

$$\Pr(Apply|C = c_z) = \begin{cases} 1 & if \ c_z \leq 0 \\ 1 & if \ c_z \leq c_z^* \ and \ \frac{c_z}{b_z} \in (0,1) \\ 0 & if \ c_z > c_z^* \ and \ \frac{c_z}{b_z} \in (0,1) \\ 0 & if \ \frac{c_z}{b_z} \geq 1 \end{cases} . \quad (3)$$

In what follows we focus on the parameter space for which the conditions in Proposition 2 hold, which allows us to model applications with Equation (3).

Assumption 3 For each $z = \{1, 2, ..., n\}$, the correlation between T_z and C_z is such that:

$$\rho_z < \frac{\sqrt{2\pi}\sqrt{1-\rho_z^2}\sigma_{C_z}}{b_z}$$

3.3 The effects of the interventions

The model enables us to capture the distinct effects of the two interventions in a parsimonious way. Specifically, the application incentive can be modelled as a shock that lowers application costs, shifting the distribution of C to the left by an amount τ'. This changes the application cut-off from c_z^* to $c_z^{*\prime}$, affecting application rates and the quality of the applicant pool. Similarly, the high wage offer can be modelled as a shock that raises the value of the job, shifting the distribution of B to the right by an amount τ'' and thus moving the application cut-off from c_z^* to $c_z^{*\prime\prime}$. We provide two propositions that characterise the effects of these shocks on application rates and applicant quality.

Proposition 3. *The interventions increase application rates.*

Proposition 3 follows from the fact that the treatments reduce relative costs without affecting the probability of getting the job (because perceived selectivity a is fixed). This raises the cutoff c_z^* at which the expected value of the application is zero, motivating a group of marginal applicants to apply for the job. Figure 2 illustrates.

< Figure 2 here. >

Proposition 4. *For each $B = b_z > 0$, the interventions increase the quality of the applicant pool if and only if $\rho_z > 0$.*

Proposition 4 presents the core insight of the model. For each level of B, the interventions attract a group of marginal applicants who face larger application costs compared to control group applicants. If costs and quality are positively correlated, these marginal applicants will have, on average, higher quality than the applicants in the control group. We illustrate this graphically, for the case where $\rho > 0$, in Figure 3.

Numerical simulations suggest that the increase in quality that we obtain in this case is fairly uniform across the distribution. Further, in Figures A.1 and A.2 in Appendix we show the effects of the interventions when the correlation between cost and quality is negative. In this case, application rates increase, while applicant quality decreases.

< **Figure 3 here.** >

3.4 Credit constraints

We introduce a third source of friction in the model: credit constraints. Following the previous literature (e.g Banerjee and Newman (1993)), we model these constraints as a maximum cost $\bar{c}$ that individuals are able to pay to apply for the job. An individual who faces costs above $\bar{c}$ is not able to apply for the job, even if the expected return is greater than the cost. The key implication of adding credit constraints to our model is that the cutoff c^* is censored at $\bar{c}$. This is going to decrease the effect of the high wage offer on application rates and applicant quality if the new, uncensored cut-off point is beyond $\bar{c}$. On the other hand, the application incentive relaxes the credit constraint by exactly τ'. The impact of this intervention is thus not affected by the presence of credit constraints. We are going to use this intuition in order to estimate the magnitude of credit constraints in Section 6.

4 Design and data

4.1 Design

We study the recruitment of workers for clerical jobs in Addis Ababa. These positions are based at the Ethiopian Development Research Institute (EDRI). They are advertised for eight fortnights. On the Sunday at the beginning of each fortnight, the positions are advertised in a local newspaper and in the main job vacancy boards of the city. The advertisement describes the position as a three-months fixed term appointment based in Addis Ababa and specifies that candidates must hold a university degree or a vocational diploma. Interested individuals are invited to call a specified phone number to get more information about the position and the application process. The deadline for applications is on the Friday of the same week.

A small team of enumerators answers the phone calls of interested jobseekers following a standardised script. First, they ask a short number of questions capturing callers' socio-demographic characteristics and work experience. Second, they give some information about the position. Third, they explain that, in order to apply for the position the jobseeker has to attend a testing session at our application centre, on a specified day. Jobseekers have to bring to the session a CV, a cover letter and proof of identity.

We randomly vary two features of the description of the position across callers: the wage and whether we offer an application incentive.[12] Callers assigned to the *control group* are informed that the position pays a monthly wage of 1,600 ETB (74 USD), before tax, and are not offered the application incentive. Callers assigned to the *application incentive* group are also told that the position pays a wage of 1,600 ETB per month. In addition, these callers are informed that, if they complete the testing session, they will receive a monetary payment of 100 ETB (4.5 USD). This payment is presented as a reimbursement of the costs jobseekers may incur in the application process. Finally, callers assigned to the *high wage* group are told that the position pays a wage of 3,200 ETB (148 USD) per month and are not offered the application incentive. We calibrated these wages at the 35th and 75th percentile of the distribution of earnings for similar positions using data from Abebe et al. (2016). Using jobseekers' assessment of the probability of getting the job, we calculate that the expected value of the high wage offer is worth about 100 USD for the average subject.

All jobseekers who call before the application deadline of a given fortnight are assigned a testing day.[13] This can be from Monday to Friday of the second week of that fortnight, or on the first Monday of the following fortnight. To reduce the risk of contamination across experimental conditions, individuals assigned to different treatment groups are invited to take the test on different days.[14] Two of these six testing days in each fortnight of the experiment are assigned to each treatment group. The assignment of testing days to treatment is randomly varied every fortnight. If a jobseeker cannot attend the testing session on the proposed day, we allow them to attend the other testing session assigned to his or her treatment group for that fortnight.

We call back all jobseekers four weeks after the first phone call. In this second inter-

[12]We describe the randomisation procedure in section 4.4 below.

[13]We do not allow jobseekers to call on more than one fortnight. After each phone call, enumerators check our database and disqualify the person if they have called in a previous fortnight.

[14]To further reduce the risk of contamination, we tell callers that we are hiring for multiple positions. If callers assigned to different treatment groups discuss about the nature of the position, this feature should help them explain why different callers are offered different terms. Specifically, callers in the control group are told that they have been assigned to a position called 'position A'. Callers in the application incentive and high wage groups are informed that they have been assigned to positions 'B' or 'C', respectively. We do not give any information about why a jobseeker is assigned to a particular position. If asked, the enumerator will respond that (i) the enumerator is not authorised to disclose the exact criteria we use to assign callers to positions, (ii) that one major factor is to keep the number of applicants across positions constant.

view, we ask a set of questions about the job applications that individuals have made in the 30 days after the first phone call. Completion of this second phone interview is incentivised with a monetary payment of 20 ETB (.85 USD).

We offer three jobs per fortnight – one per treatment group.[15] For each position, the five applicants with the highest score on an index of cognitive ability (which combines the scores on the Raven and Stroop tests) are invited for an interview. EDRI decides who among these interviewees is given the job.

4.2 Measuring applicant quality

We measure the quality of the individuals that apply for the experiment's job with a number of popular personnel selection tests. These tests are good predictors of worker productivity and are thus routinely used by firms worldwide (Chamorro-Premuzic and Furnham, 2010). We also collect information about relevant work experience and GPA scores – two variables that local employers use to screen applicants. Finally, we also collect information about economic preferences.

We measure cognitive ability with the widely used Raven and Stroop tests. The Raven test measures fluid intelligence, the ability to make meaning out of complex information and to reproduce this information. Several meta-analyses have identified the Raven test as the single best predictor of worker productivity (Schmidt and Hunter, 1998; Raven, 2000). This test has been widely used in the recent economics literature to measure worker quality (Dal Bó et al., 2013; Beaman et al., 2013; Abebe et al., 2016). The Stroop test is a popular test of cognitive control, the ability to direct and discipline attention which is required to perform complex tasks (Diamond, 2013). We use a version of the Stroop task developed by Mani et al. (2013).

For non-cognitive skills we use two widely used and validated scales: the big five inventory (BFI-44) and the grit scale (John and Srivastava, 1999; Duckworth et al., 2007). We focus on three facets on non-cognitive ability which have been identified as particularly relevant to work performance: conscientiousness, neuroticism and grit. These respectively capture a careful and vigilant attitude at work, the ability to deal with stressful situations, and the capacity to persevere through challenges (Chamorro-Premuzic and Furnham, 2010). We perform standard validity checks for the psychometric measures and satisfy accepted thresholds (e.g. see Table A.2 for Cronbach α). Laajaj and Macours (2017) emphasise the value of performing validity tests when psychometric scales are used in new contexts. We also administer scales measuring locus of control

[15] In a small number of instances, we combine two fortnights of the same treatment group together. In this case, we offer only one job to the applicants assigned to that treatment group in these two fortnights.

and confidence.

Further, we collect information about relevant work experience. For this purpose, we use the classification of tasks developed by Autor and Handel (2013). This includes the following categories: physical, routine, problem-solving, managerial, mathematical, and client-interaction tasks. For each of these, we ask participants to report the number of months of experience in jobs that required them to perform that task on a regular basis. We focus on routine, problem-solving and managerial tasks, as these were identified by firms during preliminary qualitative work as the most relevant types of experience.

We aggregate the individual measures in indices of cognitive ability, non-cognitive ability and experience. Each index is constructed as the sum of the standardised values of three measures reported in Table A.3 in the Appendix (Anderson, 2008).[16]

Finally, we measure four types of economic preferences: an incentivised measure of time preferences, and non-incentivised measures of risk preferences, social preferences and level-k rationality. The task to measure time preferences is an adapted version of the game by Augenblick et al. (2015). In this task, participants have to allocate units of work across two work sessions. For risk preferences and social preferences we use questions from the Global Preferences Survey (Falk et al., 2016). Finally, we administer a simplified and non-incentivised version of the beauty contest game to elicit level-k rationality (Crawford et al., 2013).

4.3 The sample, randomisation and attrition

Over the eight fortnights of the experiment, 4,689 jobseekers made a phone call to inquire about the position. A average of 590 individuals made this phone call each week. This stayed constant over the course of the experiment, suggesting that that the positions generated sustained interest among jobseekers. Table 1 reports summary statistics for the population of callers. The typical caller is young, male and has some work experience. The average age is 26. 15 percent of the sample is 30 or older. Women account for 21 percent of the sample. On average, callers have 28 months of wage-work experience. This masks substantial heterogeneity, as 47 percent of the sample has no work experience. Callers also have a variety of educational backgrounds.

< Table 1 here. >

[16] We think of the three components of the index as representing three distinct facets of that particular ability. We thus give each component of the index equal weight. Results, however, are qualitatively unchanged if we weight by the inverse of the covariance matrix.

This sample is fairly representative of the population of jobseekers in the city. Using data from the 2013 Labour Force Survey, we find that average age of jobseekers in Addis Ababa is 28 and that about 52 percent of jobseekers have no work experience. Further, using the new firm data we collected, we find that the average GPA in this sample (2.97) is similar to the average GPA of the applicants at other firms in this market (2.75).

We randomise the offer of the two treatments using a stratification rule in order to improve covariate balance (Bruhn and McKenzie, 2009). We create strata of six consecutive callers of the same gender and same level of work experience.[17] In each stratum, we randomly allocate two callers to the control group, two callers to the application incentive group and two callers to the high wage group. These callers are invited to a testing session at our application centre during the following week. There are two testing sessions per treatment group, per fortnight. We randomise the allocation of testing sessions to days of the week. We do this in a single draw for all eight fortnights and re-randomise until we have an allocation that is balanced across days of the week.[18]

We find that covariates are balanced across treatment groups and that attrition is modest and uncorrelated with treatment. 1,557 callers are assigned to the control condition, 1,559 to the incentive condition, and 1,573 to the high wage condition. Table 1 reports means and balance tests for the characteristics of callers that we measure during the first phone interview. Overall, we do not find evidence of imbalances across treatment groups. In the second phone survey, we interview 93.5 percent of the sample (attrition is thus 6.5 percent). This is consistent with recent studies with similar populations in urban East Africa (Abebe et al., 2016). Figure A.3 shows that attrition is not systematically related to treatment status.

4.4 Empirical strategy

Our objective is to study the impacts of the interventions on application rates and the quality of applicants. We estimate effects on application rates using a regression model of the following form:

$$\texttt{apply}_i = \beta_0 + \beta_1 \cdot \texttt{incentive}_i + \beta_2 \cdot \texttt{high wage}_i + \kappa_b + u_i, \tag{4}$$

[17] We define an experience dummy using the median number of months of work experience of callers in the pilot.

[18] The experiment is implemented over eight fortnights and there are six testing days per fortnight. The randomisation rule is that (i) each treatment should be allocated two testing days each fortnight, and (ii) no treatment should be allocated, overall, more than three or less than two sessions on the same day of the week. For this exercise, we consider the Monday session on the the following fortnight as being a distinct 'day of the week'.

where $\texttt{apply}_i$ is a dummy that captures whether person i has applied for the job, $\texttt{incentive}_i$ and $\texttt{high wage}_i$ identify individuals who have been offered the application incentive and the high-wage treatment, and κ_b are stratum dummies (Bruhn and McKenzie, 2009). The coefficients β_1 and β_2 capture the change in application rates generated by the application incentive and the high wage offer. We use a similar model to study the effects of the interventions on expectations and other job-search activities.

We study impacts on the quality of applicants by measuring changes in the conditional mean and conditional quantiles of applicant quality. Standard quantile regression models (Koenker and Hallock, 2001) enable us to estimate a conditional quantile function of the following form:

$$Q_\theta(y_i|X_i) = \gamma_0 + \gamma_1 \cdot \texttt{incentive}_i + \gamma_2 \cdot \texttt{high wage}_i. \tag{5}$$

For each measure of worker quality y, γ_1 and γ_2 capture the change in conditional quantile θ caused by the treatments. For example, suppose that we are studying the 90th percentile of the distribution of cognitive ability and that we obtain an estimate of γ_1 of 1. This would say that an applicant at the 90th percentile of the distribution in the incentive group has a cognitive ability score that is one point higher than an applicant at the 90th percentile of the control distribution. A key implication of this quantile shift is that the proportion of applicants who score above the 90th percentile of the control distribution increases. This suggests that to study changes in applicant quality we can also compare the probability that an applicant scores above a given threshold across the two groups. In the results section, we show that our findings are robust to the use of this alternative empirical strategy.

We focus the quantile analysis on five percentiles: 90th, 75th, 50th, 25th, 10th. We also present a test of stochastic dominance first proposed by Barrett and Donald (2003). Stochastic dominance occurs when the CDF of one distribution is weakly smaller than the CDF of the other distribution at all points of the support *and* strictly smaller at least at one point. The null hypothesis of the Barrett and Donald (2003) test is that the CDF of one distribution is weakly smaller than the CDF of the other distribution. To have evidence that distribution A dominates distribution B, we should thus both (i) reject that B is weakly smaller than A and (ii) fail to reject that A is weakly smaller than B. In the results section, we report and interpret the findings of both tests.

We perform inference using robust standard errors in all regressions and we correct for multiple comparisons. In general, we are unable to find evidence of heteroskedasticity in the quantile models (Machado and Silva, 2000). The use of robust standard errors is thus conservative.[19] To deal with multiple comparisons, we calculate q-values

[19]For quantile regressions, robust standard errors are computed using the Stata command developed

obtained with the sharpened procedure proposed by Benjamini et al. (2006). These give us the expected proportion of false discoveries that we need tolerate if we want to reject a particular hypothesis. We control, in turn, for multiple comparisons for the same index, and multiple comparisons across indices. To use q-values we need to assume that the test statistics related to the hypotheses in a family are positively regression dependent (Benjamini and Yekutieli, 2001). This would fail, for example, if a positive treatment effect on one quantile was associated with a null treatment effect on a different quantile. Our model suggests that this should not be the case.

5 Results

5.1 Impacts on application rates

We find that the incentive has a large and significant effect on applications. Individuals in the incentive group are 11.5 percentage points more likely to apply for the position than individuals in the control group. 41 percent of subjects in the control group apply for the position, so this treatment effect amounts to a 27 percent increase in application rates. Further, we find that individuals in the high wage group are 18.7 percentage points more likely to apply to the position. Thus the application incentive generates an increase in applications that is about two thirds of the increase in applications that can be obtained by doubling the wage. The two effects are statistically different from each other. We report these results in Table 2.

< Table 2 here. >

5.2 Impacts on the quality of the applicant pool

The application incentive improves the quality of the applicant pool. This is our most important finding. The incentive raises average cognitive ability among applicants by .25 points, or .12 of a standard deviation (Table 3). This effect is significant at the 5 percent level and is robust to the correction for multiple comparisons. Applicants in the incentive group perform significantly better in both the Raven and the Stroop tests. Compared to applicants in the control group, they answer correctly 1.2 additional questions in the Raven test and they require 2.6 fewer seconds to complete the Stroop task. These treatment effects compare favourably to those documented in previous worker selection experiments. For example, Dal Bó et al. (2013) document an increase in performance on the Raven test of about half a correct answer. We report the full results

by Machado et al. (2011).

for the individual tests in Table A.4 in the Appendix. We also find that the applicants attracted by the incentive have GPA scores that are a significant .1 standard deviation higher than control applicants (Table A.6). This is an important result as many firms in Addis Ababa use GPA scores as a key signal about candidate quality during the recruitment process. Thus the applicant pool improves also in terms of the screening criteria used by firms in this setting.

The increase in quality occurs both at the top and at the bottom of the distribution. The cognitive ability scores at the 90th, 75th and 25th percentiles improve significantly (Table 3). These effects are robust to the correction for multiple comparisons: q-values are generally below .1 and always below .15. We also estimate positive, but insignificant effects at the 50th and 10th percentiles. We assess the magnitude of these effects in two ways. First, we note that the increase in quality at the 90th and 75th percentile corresponds to about .1 of a standard deviation of the cognitive ability index. Second, we document a large effect on the number of top applicants (defined as individuals above the 90th percentile of the cognitive ability score in the control group). Top applicants nearly double from 63 in the control group to 117 in the incentive group. This effect is generated by a combination of higher application rates, and a significant, 4.4 percentage points increase in the proportion of top applicants in the applicant pool (see Table A.5 in the Appendix). At the same time, the number of applicants at the bottom of the distribution is fairly stable. For example, compared to the control condition, the application incentive attracts only nine additional applicants who score below the 10th percentile of the control distribution.

Consistently with the results for specific quantiles, we find suggestive evidence that the cognitive ability distribution among treated applicants stochastically dominates the control distribution. This is an attractive feature if the firm's objective is maximise the ability of its hires.[20] We see the characteristic pattern of stochastic dominance when we plot the cumulative distributions of cognitive ability for the two groups (Figure 4). Using the formal test of Barrett and Donald (2003) we find no evidence to reject the

[20] Stochastic dominance makes it possible to unambiguously rank distributions for objective functions that are increasing in the value of the random variable (Deaton, 1997; Barrett and Donald, 2003). Thus, in our setting, the dominant distribution would be preferred both by firms who maximise the expected quality of hires, and by 'risk-averse' firms with an objective function that is increasing and concave in quality. The comparison would not be unambiguous, however, if firms value having a smaller pool of applicants or if acceptance rates are lower in the dominant group. We consider the first point in Section 6. Regarding the second point, we show below that the increase in quality generated by the incentive is concentrated among those jobseekers with the weakest outside options. These jobseekers are likely to have the highest acceptance rates. This further increases the value of the applicant pool attracted by the application incentive.

hypothesis that the CDF of the incentive distribution is weakly smaller than the CDF of the control distribution (p=.949). This result is consistent with dominance of the incentive distribution over the control distribution. However, it also consistent with the equality of the two distributions. We thus also test the null hypothesis that the CDF of the control distribution is weakly lower than the CDF of the incentive group. For this test we obtain a p-value of .136, giving us suggestive evidence of stochastic dominance.

The high wage offer also attracts an applicant pool with higher cognitive ability. We estimate significant positive effects at the mean, and at the 90th, 75th and 25th percentiles. The magnitude of these point estimates is smaller than those we obtained for the application incentive, but we cannot reject the null hypothesis that the two treatments have the same effect. The significant estimates of the impact of the high wage offer are associated with q-values above .1 (and in two cases above .2). This suggests that the statistical significance of the results on the high wage offer is not robust to the correction for multiple comparisons.

< Table 3 here. >

< Figure 4 here. >

Lastly, we are unable to find significant differences in non-cognitive ability or experience between applicants in the incentive group and applicants in the control group. The high wage offer significantly increases median non-cognitive ability, but does non significantly affect the other percentiles of the distribution. Tables A.7 and A.8 report the results from these regressions.

5.3 Impacts on search for other jobs and job-search outcomes

We do not find evidence that the application incentive distorts individuals' search for other jobs or impacts their labour market outcomes. This is not surprising, as the small cash incentive ensures that applications for the experiment's job do not crowd out other search effort. To study the search for other jobs, we use the data collected during the second phone interview, 30 days after the initial phone call, and a regression model with same form as model (4). We investigate whether the interventions change the number of applications made, the amount of money and time spent on job search, the number of interviews and job offers obtained, and whether the jobseeker is currently working in a new job. We report results in Table A.9 in the Appendix. For the application incentive, we consistently estimate small and insignificant coefficients.

On the other hand, we find that individuals in the high wage group have significantly worse outcomes than the controls: they obtain .04 fewer interviews, .03 fewer offers and are about 2 percentage points less likely to be working in a new job. The last of these estimates is also statistically different from the estimate of the effect of the application incentive. One possible explanation for this result is that the additional applicants that are attracted in this treatment run out of resources to search for other jobs. The effects of the high wage offer on search effort are indeed negative: the intervention is associated with a 4 percent decline in the number of applications to other jobs and a 3 percent decline in the time spent on job applications. The magnitude of these effects is however relatively small and the estimates are not statistically significant. In the next section, we show that these average effects masks considerable heterogeneity with respect to credit constraints.

5.4 Who drives the increase in quality?

We study the heterogeneity of treatment effects along several dimensions. These include demographic characteristics (gender and age), labour market variables (employment status and work experience), a measure of credit constraints and a variable capturing how much subjects value of the job. We detect credit constraints by quantifying the interest rate at which individual are able to borrow.[21] Further, we estimate the value of the job by forecasting the wage that each individual can expect to be to paid in the market and incorporating this in a simple calibrated model of job search. We describe in detail the procedure that we use in Appedix A.3. For each dimension of heterogeneity x, we estimate a model of this form:

$$\begin{aligned} \mathtt{y}_i = &\beta_0 + \beta_1 \cdot \mathtt{incentive}_i \cdot I(x = 1) + \beta_2 \cdot \mathtt{high\ wage}_i \cdot I(x_i = 1) \\ &+ \beta_3 \cdot \mathtt{incentive}_i \cdot I(x_i = 0) + \beta_4 \cdot \mathtt{high\ wage}_i \cdot I(x_i = 0) \\ &+ I(x_i = 1) + \kappa_b + u_i. \end{aligned} \tag{6}$$

[21] Credit constrained individuals are only able to borrow at a very high interest rate (infinitely high, if credit is strictly rationed). To quantify this rate, we ask individuals to consider a hypothetical scenario where they have to borrow a small amount of money. Individuals then report whether they would like to borrow this sum from a lender who offers a known interest rate or from their usual source of credit. We vary the interest rate offered by the lender (from 30 percent to 5 percent per month). By looking at the rate at which individuals start to borrow from the lender, we can put bounds on the interest rate that each individual is offered by their usual source of credit. The question works well and 91 percent of individuals give consistent answers (they switch from their usual source of credit to the lender no more than once). In this section, we define as credit-constrained individuals who prefer to borrow at a 30 percent monthly interest rate rather than borrow from their usual source of credit. This includes about 30 percent of the sample. About 51 percent of the sample can borrow at less than 5 percent per month, which is roughly consistent with market rates and thus at most minor credit constraints.

Model (6) gives us separate estimates of the effect of treatment for individuals for whom $x = 1$ and individuals for whom $x = 0$. When a variable is continuous, x is dummy that splits the sample at the median of that variable. For each regression and each treatment, we present an F-test of the hypothesis that there is no heterogeneity in the effect of that treatment ($H_0 : \beta_1 = \beta_3$ for the incentive, and $H_0 : \beta_2 = \beta_4$ for the high wage offer). Results are reported in Tables A.10 to A.14 in the Appendix.

We find that the increase in cognitive ability caused by the incentive is significantly stronger among women, the unemployed, the less experienced, and for those individuals whom we estimate to value the job the most. These are groups that on average fare worse in the labour market and that respond more strongly to job search support (Card et al., 2010; Abebe et al., 2016). Further, with the exception of work experience, we cannot document heterogenous impacts of the high wage offer with respect to these dimensions. The magnitude of the heterogeneity in the effects of the incentive on quality is large. For example, among males, the effect of the incentive on average cognitive ability is close to zero. Among women, on the other hand, the cognitive ability score more than doubles (and the Raven test score increases from about 36 to about 40). We also document significantly larger effects for women at the 90th and 75th percentiles. We illustrate these results graphically in Figure A.4, where we show that the proportion of female top-applicants grows from 18 percent in the control group to 31 percent in the application incentive group. Lastly, we can geolocate a share of our sample (we are currently working on geolocating the full sample), and find *suggestive* evidence that the increase in quality is higher among jobseekers who reside in neighbourhoods that are farther away from the application centre. We illustrate this using non-parametric plots in Figure A.5.

We also find evidence suggesting that the effect of the high wage treatment differs depending on the jobseeker's credit constraints. The increase in application rates for the experiment's job is similar for individuals who experience high and low constraints. However, highly constrained individuals concomitantly reduce the number of applications to other jobs (by a significant 10 percent), while less constrained individuals do not change their other search behaviour. Further, we find that the high wage offer is significantly less effective at increasing quality from constrained applicants. Highly constrained applicants, on the other hand, do not experience a fall in other job search when offered the incentive, and have similar impacts on quality as their less constrained peers. We highlight that the measure of credit constraints we use is collected during the second phone call, after the treatments have been offered. To moderate the concerns that arise from this, we note that the phrasing of the question refers to the 'usual' source of credit (which is unlikely to have changed in a period of 30 days) and that we cannot find any effects of the interventions on the level of credit constraints reported. We thus consider these results as suggestive evidence on the role of credit constraints.

5.5 Alternative explanations

In this section we consider four alternative explanations for our results that are unrelated to the cost of making an application. We do not find evidence suggesting that these channels drive the effect of the application incentive.

Do the interventions change test effort? We test whether the treatments change effort in the selection tests. For this purpose, we administer a task that requires effort, but virtually no ability. The task requires applicants to transcribe ten string of meaningless letters. Dal Bó et al. (2013) used a similar strategy to control for differential test effort. In Table A.15 in the Appendix we show that the number of transcribed strings and the number of mistakes in transcription are not significantly different across the three groups. This suggests that the treatments do not change test effort.

Do the interventions make the job more salient in the mind of jobseekers? We study whether the treatments induce individuals to pay more attention to the experiment's job. Inattentive or cognitively-loaded jobseekers may forget to apply for the job. The treatments increase the cost of this mistake. This may encourage jobseekers to pay more attention to the position and thus reduce the probability that they forget to make the application (Gabaix and Laibson, 2005; DellaVigna, 2009). First, we note that a mechanism of this type is likely to work against the direction of our findings, as cognitive load temporarily decreases cognitive ability (Mani et al., 2013). Second, we directly test this hypothesis by exploiting the fact that salient information is more likely to be remembered (Botta et al., 2010; Santangelo and Macaluso, 2013). In particular, we investigate whether treated individuals recollect information about the position more accurately than control individuals. In the second phone interview we ask respondents to recall the wage that was offered to them in the first phone call. In the control group, about 70 percent of individuals report the correct figure. The remaining subjects either report an incorrect figure, or declare that they do not remember. The average report has an absolute mistake of 167 ETB. Importantly, we cannot find statistically significant differences between the recalls of individuals in the incentive group and those of individuals in the control group. However, we find that individuals in the high wage group recall the wage more accurately. They are both more likely to report the correct figure (by 3.8 percentage points), and they make smaller absolute mistakes on average (by 46 ETB). We report these results in Table A.16.

Do the interventions change jobseekers' beliefs about their prospects in the labour market? We study whether individuals update their beliefs about their *general prospects in the labour market* in response to treatment. This could be the result of a revision in the beliefs that individuals hold about their own employability, or in the beliefs about the labour market. For this test we use two questions from the second phone interview. In the first question, we ask subjects to forecast the number of weeks that it would take them to find a job that paid at least their reservation wage. In the second question, we ask respondents to report the wage that they expect this job will pay.[22] We find that the application incentive does not have a significant effect on either of these beliefs. The high wage offer, on the other hand, significantly increases expected wages by about 9 percent. Table A.17 in the Appendix reports these results.

We also do not find evidence that the interventions change jobseekers' beliefs about the *probability of getting the experiment's job*. We show this result in Table A.18 in the Appendix. This finding is consistent with the low levels of strategic sophistication that are documented in a simplified beauty contest task administered to all applicants at the end of the testing session Crawford et al. (2013). In general, applicants are overconfident about their likelihood of getting the experiment's job. This is consistent with recent research showing that beliefs about individual performance are characterised by overconfidence in several contexts, including job search (Malmendier and Tate, 2015; Spinnewijn, 2015).

Do the interventions change jobseekers' beliefs about the attributes of the job? We test whether the treatments affect the beliefs that individuals hold about the *characteristics of the experiment's job*. To test for this, in the second phone call we collect jobseekers' beliefs about several attributes of the job: holidays, non-standard working hours, the degree of autonomy, how satisfying the work will be, whether they will learn new skills, etc... We regress each of these beliefs on the two treatment dummies and report results

[22] To elicit expectations about the wage, we follow the method of Attanasio and Kaufmann (2009). We ask respondents to report the minimum and maximum wage that the job can pay. We then identify the midpoint between these two values and ask respondents to report the probability that the job will pay more than the midpoint. Following Attanasio and Kaufmann (2009), we assume that beliefs follow a triangular distribution. This distribution is fully characterised by an upper bound, a lower bound and a mode. The maximum and minimum wage reported by respondents identify the upper and lower bounds. Given the two bounds, the value of the CDF at the midpoint identifies the mode of the distribution.

in Table A.19 in the Appendix. We find that the application incentive has a modest significant effect on two of these dimensions: the proportion of people who think the job will have more than four days of holidays per month goes up by 2 percentage points, and the proportion of people who think that the job will help them to find a job in the future goes up by 3 percentage points. These two expectations are weak predictors of the decision to apply for the experiment's job. Among control group individuals, the belief that the job has long holidays raises the probability of making an application by 7.8 percentage points, while the belief that this job will help with job search in the future raises the probability of making an application by 8.2 percentage points. To assess the potential effect of this channel on application rates, we multiply the treatment effects on the beliefs by the effects that these beliefs have on application rates and add up. The result is that this channel can explain a change in application rates of about half a percentage point. In other words, net of the effect of expectations, the application incentives would raise applications by 11 percentage points (as opposed to 11.5 percentage points).

Mediation analysis. We use mediation analysis to quantify the contribution of the channels above to the treatment effects on application rates. We focus on application rates as the potential mediators – the salience of the job and the various dimensions of jobseeker beliefs[23] – are correlated with application rates, but do not seem to have a systematic influence on the types of workers who apply for the experiment's job.[24] As recommended in the recent literature, we use sequential g estimation (Vansteelandt, 2009; Acharya et al., 2016) to identify the *average controlled direct effect* (ACDE) of the interventions. This quantity refers to the effect that the interventions would have on an outcome if the mediators are fixed at some particular value.[25] We find that the ACDE of

[23]We focus on the dimensions which were significantly affected by treatment. These are the expected wage and an indicator of expected job attributes obtained as the sum of all seven binary beliefs reported in Table A.19.

[24]These variables are significant predictors of application rates in the control group. However, their effect on application rates is not significantly different depending on whether a jobseeker has work experience or not, or whether a jobseeker has above-median GPA or not.

[25]In order to identify the ACDE we have to assume *sequential unconfoundedness*. In a case where treatment is randomly assigned, this amounts to assuming that there are no omitted variables which confound the effect of the mediator on the outcome, conditional on treatment and a set of pre-treatment controls (Acharya et al., 2016). Given this assumption, we can identify the ACDE with a simple two-step procedure. In the first step, we regress the outcome on the mediator, the treatment dummies, a set of controls, and the interaction between the mediator and all other variables. We then obtain the predicted value of the outcome fixing all mediators to zero. This is the 'demediated' outcome. In the second step, we regress the demediated outcome on the treatment dummies. The coefficients from this regression give

the high wage offer on application rates is 9 percentage points (with a 95% confidence interval ranging from 3 to 13 percentage points). This is significantly smaller than the original estimate reported in table 2 (which was 18.7 percentage points, with a 95% confidence interval ranging from 15 to 22 percentage points). The effects on the mediators reported in this section thus have a quantitatively large influence on application rates in the high wage group. The controlled direct effect of the incentive, on the other hand, is quantitatively similar and statistically indistinguishable from the original treatment effect (the two estimates are 11.5 and 13 percentage points). This is not surprising, as we only find evidence of large and significant effects on the mediators for the high wage offer.

6 Structural analysis

In this section we discuss the identification and estimation of the structural model. We then present and interpret the estimates of the structural parameters. We find that application costs are large, heterogeneous and positively related with quality. Further, a large share of individuals are credit constrained according to the model's estimates.

6.1 Identification and estimation

Our objective is to estimate the following parameters: perceived selectivity (a), the parameters that characterise the joint distribution of costs (C) and quality (T), for each level of the value of job (B), and the magnitude of the shocks to costs and benefits ($\tau_{\texttt{incentive}}$ and $\tau_{\texttt{wage}}$). This last parameter – $\tau_{\texttt{wage}}$ – differs from the discounted value of the wage increase when individuals have credit constraints.

We use direct measures of T and B. We proxy T with the score on the Raven test. We predict B by specifying a simple dynamic framework of job search. They key parameter that generates heterogeneity in B is the wage that individuals would be paid in the market. We predict this wage using the Post-LASSO estimator recommended by Belloni et al. (2014). For the structural estimation, we discard individuals with a negative B and we split the remaining individuals (about 65 percent of the sample) at the median level of B. On average, an individual in the high-B group gets a net, discounted benefit from the experiment's job of about 548 ETB (23.5 USD). For the low B group, the benefit is about 377 ETB (16 USD). We describe our procedure in detail in Appedix A.3. Ten parameters describe the joint distribution of T and C for the high and low B groups.[26] This means that we have a total of 13 parameters to estimate.

us the estimate of the ACDE.

[26]These parameters are: μ_{Tl}, μ_{Cl}, σ_{Tl}, σ_{Cl}, σ_{TCl}, μ_{Th}, μ_{Ch}, σ_{Th}, σ_{Ch}, σ_{TCh}.

To identify these structural parameters we use fourteen empirical moments. We use control group application rates and the average and standard deviation of the Raven score among control group applicants (3 moments). Further, we use the change in application rates and the change in the average applicant Raven score induced by the two treatments (4 moments).[27] We compute these moments separately for the high and low B groups, giving us fourteen moments in total. The thirteen parameters are jointly identified by these fourteen moments.

The intuition for identification is as follows. The six moments from the control group describe the truncated distribution of T. Thus these moments enable us to identify the mean and standard deviations of quality (μ_T, σ_T) and the mean of application costs (μ_C), which carries information about the point of truncation. Conditional on these parameters, the changes in application rates induced by the two interventions identify the severity of the shocks $\tau_{\texttt{incentive}}$ and $\tau_{\texttt{wage}}$ (which have a first-order influence on the shift in cutoff c^*) and the standard deviations of costs σ_{Ch} and σ_{Cl} (which, conditional μ_C, determine the number of people that lie between the two cutoffs). Further, the change in average quality induced by the two treatments identifies the covariance between cost and quality and perceived selectivity a. Table 4 summarises.

< Table 4 here. >

We study credit constraints by comparing the cutoff point on c implied by $\tau_{\texttt{wage}}$ ($c^{*\prime\prime}_{\tau}$) to the cut-off point when the size of the shock is the actual value of the wage increase w ($c^{*\prime\prime}_{\texttt{w}}$). If credit constraints $\bar{c}$ bind, then the second cutoff point will be larger than the first cutoff point ($c^{*\prime\prime}_{\texttt{w}} > c^{*\prime\prime}_{\tau}$). Further, the first cutoff will be found exactly at the point where credit constraints start binding: $c^{*\prime\prime}_{\tau} = \bar{c}$. This enables us to identify the proportion of jobseekers that are credit constrained.

To estimate the model we use a classical minimum distance estimator (Wooldridge, 2010). We save the fourteen empirical moments in a vector $\boldsymbol{m}$. For a 13×1 parameter vector $\boldsymbol{\theta}$, we solve the model and calculate fourteen simulated moments $\boldsymbol{m}_S(\boldsymbol{\theta})$. We update $\boldsymbol{\theta}$ in order to solve:

$$\hat{\boldsymbol{\theta}} = \min_{\boldsymbol{\theta}} \left[\boldsymbol{m}_S(\boldsymbol{\theta}) - \boldsymbol{m}\right]' \cdot J(\boldsymbol{m})^{-1} \cdot \left[\boldsymbol{m}_S(\boldsymbol{\theta}) - \boldsymbol{m}\right],$$

$J(\boldsymbol{m})$ is a diagonal matrix that contains the variance of each moment, ensuring that more precisely estimated moments get a greater weight in estimation. In line with the recent literature, we use this simple weighting matrix instead of the theoretically optimal weighting matrix, which may suffer from small sample bias (Altonji and Segal,

[27] For the high wage group, we use the demediated change in application rates, as calculated in Section 5.5.

1996). We calculate $J(\boldsymbol{m})$ using a bootstrap with 1,000 replications. We then use a second bootstrap to perform inference (keeping $J(\boldsymbol{m})$ fixed). We include the estimation of B and the demediation procedure in both bootstrap procedures.

6.2 Results

Fit with targeted and non-targeted moments and other validity checks. The estimation is successful and we obtain a good fit between empirical and simulated moments. We report parameter estimates in Table 5 below and we compare empirical and simulated moments in Table A.20 in the Appendix. All simulated application rates are within one percentage point of the empirical moment. The mean and standard deviation of the Raven test in the control group are matched almost exactly (e.g. for the low B group, the difference between the empirical and simulated moment is of about .06 correct answers). Finally, we also fit fairly precisely the simulated change in the Raven score induced by the treatments. The difference between the simulated and the empirical treatment effects is always less than half a correct answer (with the exception of the effect of the incentive treatment for the high B group, which is in the right direction, but quantitatively somewhat under-predicted).

We further validate the model by showing that it has a reasonable fit with a key non-targeted moment: subjects' assessment of the probability of getting an offer for the experiment's job.[28] Subjects are widely overconfident and the average probability reported is 46 percent. To put this in context, we give one job about every 115 applicants and participants have reasonable expectations about this number. Our model's estimates are consistent with this level of overconfidence. In particular, the average jobseeker in the model forecasts that the probability of getting the job is about 40 percent.

We also find that two key predictions of the model are consistent with the data. First, the model predicts that the effects of both treatments produce a uniform rightward shift of the quality curve. In the previous section, we showed that this is indeed the case for our treatments (see Figures 4 and A.6 and the discussion on stochastic dominance). Second, the model predicts that average quality among the jobseekers who do not apply for the position is higher than among those who apply. We check this prediction by looking at individuals' GPA, which we observe for both applicants and not applicants and is correlated with cognitive ability. We find that non-applicants' GPA is about a significant 8 percent higher than applicants' GPA. This confirms the prediction of the model.

[28] We elicit this probability during the second phone call. However, we ask subjects to report the forecast that they made at the time of deciding whether to apply for the position or not.

Finally, we support the intuition for identification given above by studying the elasticity of the simulated moments with respect to the parameters of the model. As in Kaboski and Townsend (2011) and Lagakos et al. (2017), we first compute all moments using the structural estimates of the parameters. We then shock by one percent the value of each parameter at a time, and compute the percent change in the simulated moments. This is illustrative of what drives identification very close to the structural estimates. We report the results in Table A.21 in the Appendix. The estimated elasticities are consistent with the intuitions reported above. For example, the elasticity of the change in applicant quality with respect to the covariance between cost and quality is close to 2 (a 1 percent change in the covariance leads to a 2 percent change in the simulated moment). For the other moments, the elasticity with respect to this parameter is much lower.

Parameter estimates. We estimate that application costs are large and heterogeneous. For the high value group, the mean of application costs is 207 ETB. This amounts to 13 percent of the monthly salary offered to individuals in the control group, or to about 38 percent of the value of the job. For the low value group, mean costs are about 140 ETB, or 9 percent of the salary and 37 percent of the value of the job. We also estimate that application costs have a large dispersion, in both groups. The standard deviation of application costs is about 254 ETB for the high B group and 234 for the low B group. This implies that 80 percent of individuals in the high B group and 73 percent of individuals in the low B group have positive application costs.

Our estimates confirm that application costs are positively correlated with worker quality. This correlation is about 0.47 for the high B group and 0.62 for the low B group.[29] These estimates imply a large increase in average Raven scores as we move along the cost distribution. For example, a jobseeker with costs one standard deviation above the mean has a Raven score that is about 6.5 scores higher than the average jobseeker (a 15 percent increase). Using the average Mincerian return to cognitive ability reported in the review paper by Bowles et al. (2001), we estimate that the value of this additional ability would be 208 ETB per month, similar to the size of mean application costs for the high B group.

Our parameter estimates also suggest that credit constraints are widespread. We calculate that the wage shock implies a cutoff point at about 1,800 ETB.[30] We estimate

[29] These estimates satisfy Assumption 3, which guarantees a unique crossing point.

[30] This is a conservative estimate, which is derived by adjusting the value of the wage shock downwards. If the shock was set to the full discounted value of the wage increase, the implied cutoff point would be even larger. The adjustment reflects the fact that a number of factors other than credit con-

that jobseekers stop applying at a much lower cost. For the high value group this cost is between 270 and 350 ETB.[31] This implies that credit constraints start binding from 350 ETB. About 29 percent of the sample faces costs above this figure and is thus predicted to be credit constrained. This estimate is very similar to the self-reported measure we discussed in Section 5.3. According to subjects' self-reports, about 30 percent of the sample is willing to borrow at extremely high interest rates which suggest credit rationing. As discussed in that section, the response to treatment of this group is also consistent with credit constraints.

< Table 5 here. >

The returns of the interventions and policy simulations. Finally, we assess the returns of the interventions and of two counterfactual policies. Each intervention enables the firm to recruit workers with higher cognitive ability and hence higher productivity (cognitive ability is a strong predictor of productivity). This generates a stream of profits for the firm since the wage is fixed to the level that was originally posted. Each intervention also entails two types of costs. First, the firm has to pay the direct cost of the intervention (the cost of the incentive or the higher wage). Second, the firm has to employ staff time to review the additional applications.

We calibrate costs and benefits in order to assess the effect of the interventions on an average firm recruiting a clerical worker in this market. For this purpose, we use the data that we collected from local firm managers. First, we quantify recruitment costs using managers' assessment of the time required to review one more application. On average managers report that this requires about one hour of work.[32] We price this hour at the mean salary of the HR staff who review applications in these firms. Second, we calibrate the number of applicants in the control group and the number of jobs on offer using the average of these variables among the firms in our sample. Third, we compute worker turnover rates and use these to assess the expected number of months that the

straints may reduce the effective value of the high wage offer. For example, some jobseekers may not believe that a higher wage will be paid. We estimate the magnitude of these factors by taking the ratio of $\tau_{\texttt{incentive}}$ to the monetary value of the application incentive. Credit constraints should in principle not affect the value of the incentive. However, other factors (mistrust, memory, etc..) may push $\tau_{\texttt{incentive}}$ below its nominal monetary value of 100 ETB. We thus interpret the ratio of $\tau_{\texttt{incentive}}$ to 100 as an estimate of the importance of these non-modelled factors that change the value of the interventions. We multiply the discounted value of the wage increase by this ratio in order to control for these factors.

[31] This figure changes depending on whether we fit the raw or the demediated moment.

[32] We also ask whether there are any financial costs involved in reviewing one more application. The great majority of managers report that this is not the case. The majority of financial costs are fixed costs related to items such as advertising the position.

worker is going to spend in the firm.[33] Finally, we calibrate the productivity gains from higher worker quality using the average Mincerian return to cognitive ability reported in Bowles et al. (2001).

We design two counterfactual policies that reduce the upfront costs of the application incentive. One drawback of the application incentive is that the firm subsidises a large group of infra-marginal individuals who would have applied for the job in the absence of the incentive. To decrease transfers to infra-marginal applicants we propose the following two policies: (i) an application incentive that is offered to all individuals who would not apply without the incentive (this assumes that the firm can develop an accurate targeting device based on worker observables); (ii) an application incentive that is offered only to the applicants who score above a threshold in the test (we set this threshold to the level that fills the positions on offer in expectation, so in practice under this policy the incentive is offered to all hires). These interventions reduce transfers to infra-marginal individuals by exploiting, in turn, the information available to firms and the information available to workers. However, it is unlikely that the firm will be able to identify marginal applicants without error. Hence intervention (i) should be considered as an upper bound of what a targeted incentive may deliver.

< **Table 6 here.** >

We find that the application incentive has a positive internal rate of return (IRR) of about 11 percent. This is above market interest rates (which are about 5 percent in Ethiopia), and in line with the hurdles rates commonly reported by firms.[34] The two counterfactual incentive schemes have very large IRRs, above 100 percent. Finally, the high wage offer has a large negative IRR. We present these results in Table 6 below. In the second part of Table 6 we give a break down of how each intervention changes costs and benefits. When the incentive is offered to all hires, the cost of the intervention decreases by about 90 percent, but benefits also decrease substantially. When the incentive is offered to all marginal applicants, the cost of the intervention decreases by about 80 percent and benefits are unchanged.

[33]The expected spell of employment in the firm is 42 months. We assume, conservatively, that the high wage offer is only valid for the first three months. In subsequent months the firm reverts to the baseline wage.

[34]We are not aware of data on the hurdle rates used by firms in Ethiopia or in countries with similar macroeconomic conditions. A recent survey by the Bank of England finds that most firms in the UK adopt hurdle rates between 5 and 15 percent (Saleheen et al., 2017).

7 Discussion

In this section we address two important questions that follow from our findings. First, what is the mechanism that drives the correlation between the size of the application costs faced by a jobseeker and his or her cognitive ability? Second, why are application incentives not used more frequently by firms given that they have a large positive return? We provide some answers to these questions by leveraging a high-frequency panel dataset on young jobseekers and a second experiment with firm managers in Addis Ababa.

7.1 Why are costs and quality positively correlated?

We provide evidence for a selection mechanism that can generate a correlation between application costs and applicant quality. We hypothesise that low-cost, high-quality individuals stop searching for work faster than high-cost, high-quality individuals. This is both because they are more likely to secure a job and because they can afford to remain inactive if they do not find a suitable position.[35] Thus, over time, the average quality of the low-cost jobseekers who keep looking for employment decreases in comparison to the quality of high-cost jobseekers. This results in a positive correlation between costs and quality among those individuals who look for employment at any given point in time.

To provide evidence for this mechanism, we use a fortnightly panel dataset that tracks a sample of young adults in Addis Ababa for one year. This dataset has information about job-search decisions and employment outcomes. It also includes a Raven test score obtained close to the beginning of the panel. Further, it contains two variables which proxy for search costs in labour markets: a measure of direct costs (distance from the city centre) and a measure of financial resources (savings at baseline). The dataset was collected by and is described in detail in Abebe et al. (2016).

We find clear support for the selection mechanism. Over the course of the year, average quality among low-cost jobseekers declines markedly, while average quality among high-cost jobseekers is roughly constant. We present this result in Figure 5. Further, using regression analysis, we show that the trends for high and low-cost jobseekers are statistically different from each other. To produce this result, we create a dataset of average Raven scores among jobseekers without work, by fortnight and by individual

[35]The reverse may happen among low-quality types. For this group the chances of being offered a position are relatively low. So those workers who face high costs of search are more likely to stop searching for stable work and take up casual employment in the informal sector.

type (high and low cost). Changes in this variable are due to selection of individuals in and out of the group of jobseekers. We report the results of our analysis in Table 7. We find that, irrespective of which measure of costs we use, there is no significant trend in the average quality of high-cost jobseekers. On the other hand, there is a negative trend in the average quality of low-cost jobseekers, which is both significantly different from zero and significantly different from the trend of high-cost jobseekers. Reassuringly, we are unable to find differential trends if we split the sample using two 'Placebo' variables that are not directly related with the cost of job search: being married, and reporting high life satisfaction at baseline (Table A.22).

Finally, we present evidence suggesting that the differential trend is mostly due to transitions from search to inactivity. In particular, when we use savings to proxy for search costs, we find that low-cost jobseekers with above-average Raven scores are a significant 10 percentage points more likely to stop searching in the following period, compared to high-cost jobseekers with similar Raven scores. This effect is a combination of two separate types of transitions. Low-cost, high-Raven jobseekers are (a significant) 7.6 percentage points more likely to become inactive next period, and (an insignificant) 2.5 percentage points more likely to become employed. Among jobseekers with below-average Raven scores there are no significant differences in transitions. When we define costs using distance from the centre of the city, we find effects that are qualitatively in the same direction, but of a smaller magnitude and generally insignificant. We report these results of this analysis in Tables A.23 and A.24 in the Appendix.

< **Table 7 here.** >

< **Figure 5 here.** >

7.2 Why are application incentive not commonly used in this context?

We conclude by reporting the results of a second experiment that studies the preferences and expectations of managers at firms recruiting for clerical positions.[36] This experiment enables us to explore two possible reasons why firms in Addis Ababa do not use application incentives. First, firm managers may not value general cognitive ability and thus may not rank applicants from the incentive group above control applicants. Alternatively, managers would like to recruit workers with higher cognitive ability, but do not expect that the application incentive will attract these workers.

In the first task, we study whether firm managers rank treatment group applicants above control applicants. To incentive this task, we offer to invite one person from our

[36]We describe how we sample these firms in section 2.

sample of applicants to make a new job application at the manager's firm. The manager can determine who this person is going to be by ranking the standardised CVs of three selected applicants. We sample one applicant from each experimental group. At this point of the experiment, however, the manager has not been informed about the two interventions nor about how the three applicants have been selected. On the CVs, we report applicants' education, age, work experience, GPA and the results from the Raven and conscientiousness tests (Figure A.7 shows a sample CV). We select triplets of applicants that reproduce as closely as possible the average differences in these characteristics between groups.[37] After the manager ranks the CVs, we randomly draw two of the three CVs and invite the person with the higher rank to make an application at the manager's firm.

We find that both interventions improve the quality of the applicant pool in the eyes of local firm managers. We show this result in Table 8 using a series of linear probability models. In the first two columns, the dependent variable is a dummy for individuals who are ranked first. In the third column, the dependent variable is a dummy for being ranked first or second. We find that applicants from the incentive group are a significant 36.9 percent more likely to be ranked first than control applicants, and a significant 37.4 percentage points more likely to be ranked first or second. In column two, we only consider applicants from the control and incentive groups. We find that incentive group applicants are ranked above control group applicants about 70 percent of the times.

< **Table 8 here.** >

In the second task, we test whether managers are misinformed about the effects of the application incentive. We first give managers detailed information on the experiment and then ask them to forecast the impacts of the application incentive on application rates and applicant quality (measured with the Raven test). To measure quality at different points of the distribution, we obtain forecasts of (i) the average Raven score and (ii) the average Raven score among the 100 highest-scoring applicants. Further, before forecasts are made, we disclose the application rates and Raven test scores of applicants in the control and high wage groups in order to anchor managers' priors on the correct level of these variables. We reward managers for the accuracy of one randomly drawn forecast.

[37] In total, we sample sixteen triplets of applicants and randomly allocate a triplet to each manager. Across triplets, we randomly allocate the order with which the candidates from the three groups are presented.

We find that managers make considerable forecasting errors and generally underestimate the impacts of the application incentive on applicant quality. In Figure 6 we report a box plot of the distribution of forecasting errors for the three forecasts. On average, managers expect that the application incentive will increase application rates and decrease applicant quality. In particular, they predict that performance in the Raven test will fall by about one correct answer, both at the mean and at the top of the distribution. In reality, performance in the Raven test improves by about 1 correct answer in both cases. Overall, about 75 percent of managers underestimate the level of cognitive ability of the applicants in the incentive group. Misinformation about the returns of the application incentive thus gives us a plausible explanation for why this intervention is not used more frequently by firms in Addis Ababa.

< **Figure 6 here.** >

8 Conclusion

In a worker recruitment experiment in Addis Ababa, Ethiopia we show that firms can use application incentives to attract applicants with higher cognitive ability. We estimate a structural model of applications decisions and find that the positive effect of application incentives follows from the fact that application costs are large, heterogeneous and, surprisingly, positively correlated with jobseeker ability. Using a high-frequency panel dataset on job search decisions, we show evidence that this correlation is driven by selection into the pool of jobseekers. Our estimates suggest that for the average firm in this market the application incentive generates large positive returns. However, in a second experiment, we show that local firm managers underestimate these returns. This can explain why application incentives are not commonly used by firms in this context.

The gains in applicant quality generated by the incentive are driven by groups of jobseekers that have low incomes and weak outside options in the labour market. These are the jobseekers for whom the net present value of the experiment's job is largest. Enabling these jobseekers to participate more effectively in the labour market would benefit both firms *and* workers. This suggests that well-targeted active labour market policies may have positive effects on allocative efficiency in the labour market.

Our experimental evidence on how application costs affect firms' ability to recruit talented workers is new in the literature, and generates a number of specific leads for future research First, it would be important to study the interaction between interventions that incentivise applications and interventions that improve the quality of screening

(Autor and Scarborough, 2008). As more detailed and informative tests may discourage prospective applicants (Alonso, 2016), improved screening may need to be bundled with application incentives in order to be effective. Second, it would be interesting to study how firms adjust investment when they hire more talented workers. If personnel ability is complementary to capital and technology (Bender et al., 2016), the dynamic gains from relaxing labour constraints could be very large. Finally, it would be important to understand whether behavioural factors such as overconfidence can distort jobseekers' portfolio of applications and job-entry decisions. For example, overconfident individuals may wait too long in unemployment, or may overestimate earnings in occupations where wages are volatile. These factors could have large repercussions on the allocation of workers' talent in the economy.

References

Abebe, G., S. Caria, M. Fafchamps, P. Falco, S. Franklin, and S. Quinn (2016). Curse of Anonymity or Tyranny of Distance? The Impacts of Job-Search Support in Urban Ethiopia. *NBER Working Paper No. 22409*.

Abebe, G., S. Caria, M. Fafchamps, P. Falco, S. Franklin, S. Quinn, and F. Shilpi (2016). Job Fairs: Matching Firms and Workers in a Field Experiment in Ethiopia. *Working Paper*.

Acharya, A., M. Blackwell, and M. Sen (2016). Explaining Causal Findings Without Bias: Detecting and Assessing Direct Effects. *American Political Science Review 110*(3), 512–529.

Algan, Y., B. Crépon, and D. Glover (2017). The Value of a Vacancy: Evidence from a Randomised Experiment with the French Employment Agency. *Working Paper*.

Alonso, R. (2016). Recruitment and Selection in Organizations. *Working Paper*.

Altonji, J. G. and L. M. Segal (1996). Small-Sample Bias in GMM Estimation of Covariance Structures. *Journal of Business & Economic Statistics 14*(3), 353–366.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association 103*(484).

Ashraf, N., O. Bandiera, S. S. Lee, et al. (2014). Do-Gooders and Go-Getters: Career Incentives, Selection, and Performance in Public Service Delivery. *Working Paper*.

Attanasio, O. and K. Kaufmann (2009). Educational Choices, Subjective Expectations, and Credit Constraints. *NBER Working Paper No. 15087*.

Augenblick, N., M. Niederle, and C. Sprenger (2015). Working over Time: Dynamic Inconsistency in Real Effort Tasks. *The Quarterly Journal of Economics*, 1067–1115.

Autor, D. H. and M. J. Handel (2013). Putting Tasks to the Test: Human Capital, Job Tasks, and Wages. *Journal of Labor Economics 31*(S1), S59–S96.

Autor, D. H. and D. Scarborough (2008). Does Job Testing Harm Minority Workers? Evidence from Retail Establishments. *The Quarterly Journal of Economics 123*(1), 219–277.

Balakrishnan, U., J. Haushofer, and P. Jakiela (2015). How Soon Is Now? Evidence of Present Bias from Convex Time Budget Experiments. *Working Paper*.

Banerjee, A. V. and A. F. Newman (1993). Occupational Choice and the Process of Development. *Journal of political economy 101*(2), 274–298.

Barrett, G. F. and S. G. Donald (2003). Consistent Tests for Stochastic Dominance. *Econometrica 71*(1), 71–104.

Beaman, L., N. Keleher, and J. Magruder (2013). Do Job Networks Disadvantage Women? Evidence from a Recruitment Experiment in Malawi. *Working Paper*.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). High-Dimensional Methods and Inference on Structural and Treatment Effects. *The Journal of Economic Perspectives 28*(2), 29–50.

Belot, M., P. Kircher, and P. Muller (2017). How Wage Announcements Affect Job Search Behaviour - A Field Experimental Investigation. *Working Paper*.

Bender, S., N. Bloom, D. Card, J. Van Reenen, and S. Wolter (2016). Management Practices, Workforce Selection and Productivity. *NBER Working Paper 22101*.

Benjamini, Y., A. M. Krieger, and D. Yekutieli (2006). Adaptive Linear Step-up Procedures that Control the False Discovery Rate. *Biometrika 93*(3), 491–507.

Benjamini, Y. and D. Yekutieli (2001). The Control of the False Discovery Rate in Multiple Testing under Dependency. *Annals of statistics*, 1165–1188.

Botta, F., V. Santangelo, A. Raffone, J. Lupiáñez, and M. O. Belardinelli (2010). Exogenous and Endogenous Spatial Attention Effects on Visuospatial Working Memory. *The Quarterly Journal of Experimental Psychology 63*(8), 1590–1602.

Bowles, S., H. Gintis, and M. Osborne (2001). The Determinants of Earnings: A Behavioral Approach. *Journal of economic literature 39*(4), 1137–1176.

Bruhn, M. and D. McKenzie (2009). In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics 1*(4), 200–232.

Bryan, G. and M. Morten (2015). Economic Development and the Spatial Allocation of Labor: Evidence from Indonesia. *Working Paper*.

Card, D., R. Chetty, and A. Weber (2007). Cash-on-Hand and Competing Models of Intertemporal Behavior: New Evidence from the Labor Market. *The Quarterly journal of economics 122*(4), 1511–1560.

Card, D., J. Kluve, and A. Weber (2010). Active Labour Market Policy Evaluations: A Meta-Analysis. *The economic journal 120*(548).

Chamorro-Premuzic, T. and A. Furnham (2010). *The Psychology of Personnel Selection*. Cambridge University Press.

Crawford, V. P., M. A. Costa-Gomes, and N. Iriberri (2013). Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications. *Journal of Economic Literature 51*(1), 5–62.

Crépon, B. and G. J. Van den Berg (2016). Active Labor Market Policies. *Annual Review of Economics 8*, 521–546.

CSA (2000). Report on the 1999 labour force survey. *Federal Democratic Republic of Ethiopia, Central Statistical Agency*.

CSA (2014). Key findings on the 2014 urban employment unemployment survey. Technical report, Federal Democratic Republic of Ethiopia, Central Statistical Agency.

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

Deaton, A. (1997). *The Analysis of Household Surveys: a Microeconometric Approach to Development Policy*. World Bank Publications.

DellaVigna, S. (2009). Psychology and Economics: Evidence from the Field. *Journal of Economic Literature 47*(2), 315–372.

DellaVigna, S. and D. Pope (2016). What Motivates Effort? Evidence and Expert Forecasts. *NBER Working Paper No. 22193*.

Deserranno, E. (2014). Financial Incentives as Signals: Experimental Evidence from the Recruitment of Health Workers. *Working Paper*.

Diamond, A. (2013). Executive Functions. *Annual review of psychology 64*, 135–168.

Duckworth, A. L., C. Peterson, M. D. Matthews, and D. R. Kelly (2007). Grit: Perseverance and Passion for Long-Term Goals. *Journal of Personality and Social Psychology 92*(6), 1087.

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2016). Global Evidence on Economic Preferences. *Working Paper*.

Falk, A., A. Becker, T. Dohmen, D. Huffman, and U. Sunde (2016). The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences. *Working Paper*.

Flory, J. A., A. Leibbrandt, and J. A. List (2014). Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job Entry Decisions. *The Review of Economic Studies 82*(1), 122–155.

Franklin, S. (2016). Location, Search Costs and Youth Unemployment: A Randomized Trial of Transport Subsidies in Ethiopia. *Economic Journal, Forthcoming*.

Gabaix, X. and D. Laibson (2005). Bounded Rationality and Directed Cognition. *Working Paper*.

Galenianos, M., P. Kircher, and G. Virág (2011). Market Power and Efficiency in a Search Model. *International Economic Review 52*(1), 85–103.

Hanna, R., S. Mullainathan, and J. Schwartzstein (2014). Learning through Noticing: Theory and Evidence from a Field Experiment. *The Quarterly Journal of Economics 129*(3), 1311–1353.

Heckman, J. J., J. Stixrud, and S. Urzua (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics 24*(3), 411–482.

Hoffman, M. and S. V. Burks (2017). Worker Overconfidence: Field Evidence and Implications for Employee Turnover and Returns from Training. *NBER Working Paper. No 23240*.

Hoffman, M., L. B. Kahn, and D. Li (2015). Discretion in Hiring. *NBER Working Paper No. 21709*.

Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow (2013). The Allocation of Talent and US Economic Growth. *NBER Working Paper No. 18693*.

Hsieh, C.-T. and E. Moretti (2015). Why do Cities Matter? Local Growth and Aggregate Growth. *NBER Working Paper No. 21154*.

Imbert, C. and J. Papp (2016). Short-term Migration Costs: Evidence from India's Employment Guarantee. *Working Paper*.

Jewitt, I. and E. Ortiz-Ospina (2016). Selection in Universities. *Working Paper*.

John, O. P. and S. Srivastava (1999). The Big Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives. *Handbook of Personality: Theory and Research 2*(1999), 102–138.

Kaboski, J. P. and R. M. Townsend (2011). A Structural Evaluation of a Large-Scale Quasi-Experimental Microfinance Initiative. *Econometrica 79*(5), 1357–1406.

Koenker, R. and K. Hallock (2001). Quantile Regression: An Introduction. *Journal of Economic Perspectives 15*(4), 43–56.

Laajaj, R. and K. Macours (2017). Measuring Skills in Developing Countries. *Working Paper*.

Lagakos, D., M. Mobarak, and M. E. Waugh (2017). The Welfare Effects of Encouraging Rural-Urban Migration. *Working Paper*.

Machado, J. A. and J. S. Silva (2000). Glejser's Test Revisited. *Journal of Econometrics 97*(1), 189–202.

Machado, J. A. F., P. Parente, and J. M. C. Santos Silva (2011). QREG2: Stata Module to Perform Quantile Regression with Robust and Clustered Standard Errors.

Malmendier, U. and G. Tate (2015). Behavioral CEOs: The Role of Managerial Overconfidence. *The Journal of Economic Perspectives 29*(4), 37–60.

Mani, A., S. Mullainathan, E. Shafir, and J. Zhao (2013). Poverty Impedes Cognitive Function. *Science 341*(6149), 976–980.

Marimon, R. and F. Zilibotti (1999). Unemployment vs. Mismatch of Talents: Reconsidering Unemployment Benefits. *The Economic Journal 109*(455), 266–291.

Mas, A. and A. Pallais (2016). Valuing Alternative Work Arrangements. *NBER Working Paper No. 22708*.

Mas, A. and A. Pallais (2017). Labor Supply and the Value of Non-Work Time: Experimental Estimates from the Field. *Working Paper*.

McKenzie, D. (2017). How Effective are Active Labor Market Policies in Developing Countries? A Critical Review of Recent Evidence. *The World Bank Research Observer*.

Phillips, D. C. (2014). Getting to work: Experimental evidence on job search and transportation costs. *Labour Economics 29*, 72–82.

Raven, J. (2000). The Raven's Progressive Matrices: Change and Stability over Culture and Time. *Cognitive psychology 41*(1), 1–48.

Rogerson, R., R. Shimer, and R. Wright (2005). Search-theoretic models of the labor market: A survey. *Journal of economic literature 43*(4), 959–988.

Saleheen, J., I. Levina, M. Melolinna, and S. Tatomir (2017). The Financial System and Productive Investment: New Survey Evidence. *Bank of England Quarterly Bulletin Q1*.

Santangelo, V. and E. Macaluso (2013). Visual Salience Improves Spatial Working Memory via Enhanced Parieto-Temporal Functional Connectivity. *Journal of Neuroscience 33*(9), 4110–4117.

Schmidt, F. L. and J. E. Hunter (1998). The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings. *Psychological bulletin 124*(2), 262.

Spinnewijn, J. (2015). Unemployed but Optimistic: Optimal Insurance Design with Biased Beliefs. *Journal of the European Economic Association 13*(1), 130–167.

Vansteelandt, S. (2009). Estimating Direct Effects in Cohort and Case–Control Studies. *Epidemiology 20*(6), 851–860.

Weaver, J. (2016). Jobs for Sale: Bribery and Misallocation in Hiring. *Working Paper*.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT press.

Figures and tables for inclusion in the main text

Figure 1: Most important HR problem

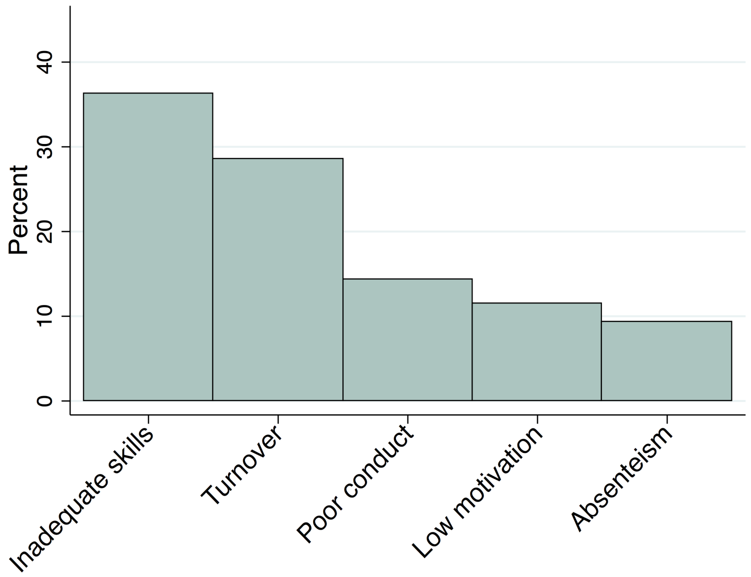
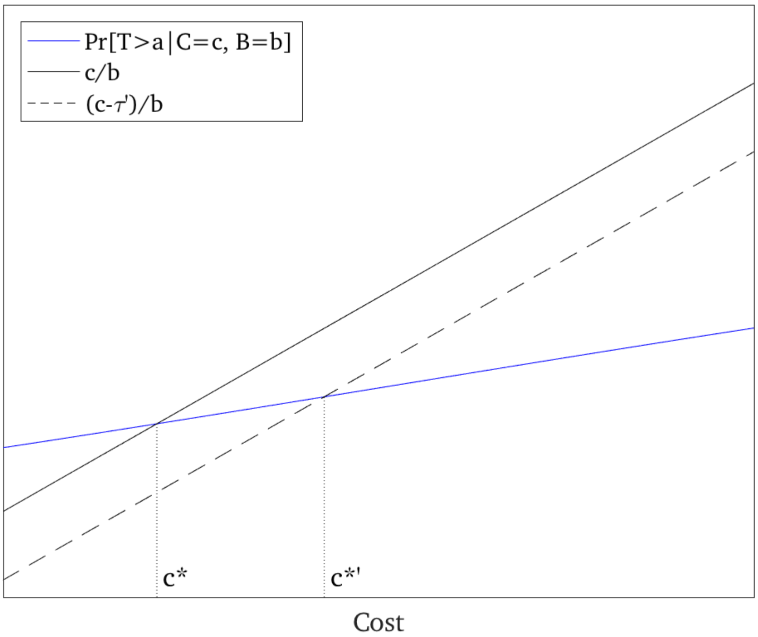
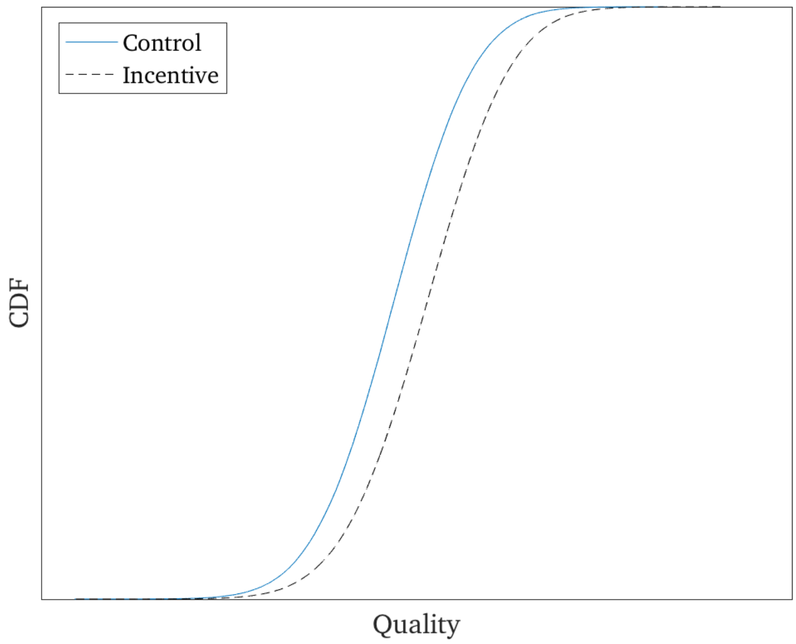


Figure 2: The application decision



Note: $\rho_{TC} > 0$

Figure 3: Predicted impacts on the distribution of applicant quality



Note: $\rho_{TC} > 0$

Figure 4: Impacts on the distribution of applicant cognitive ability
Incentive treatment

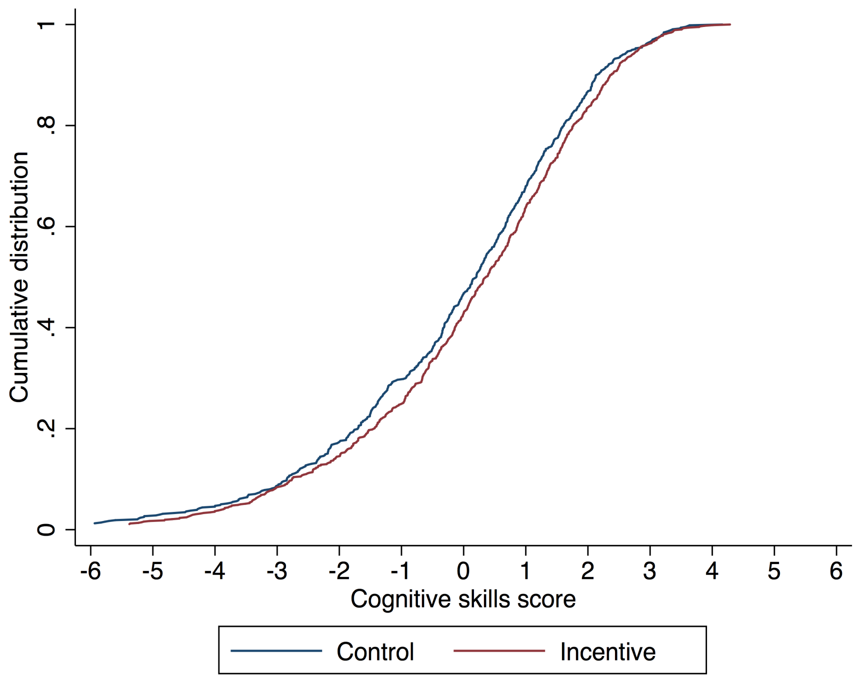


Figure 5: The selection mechanism

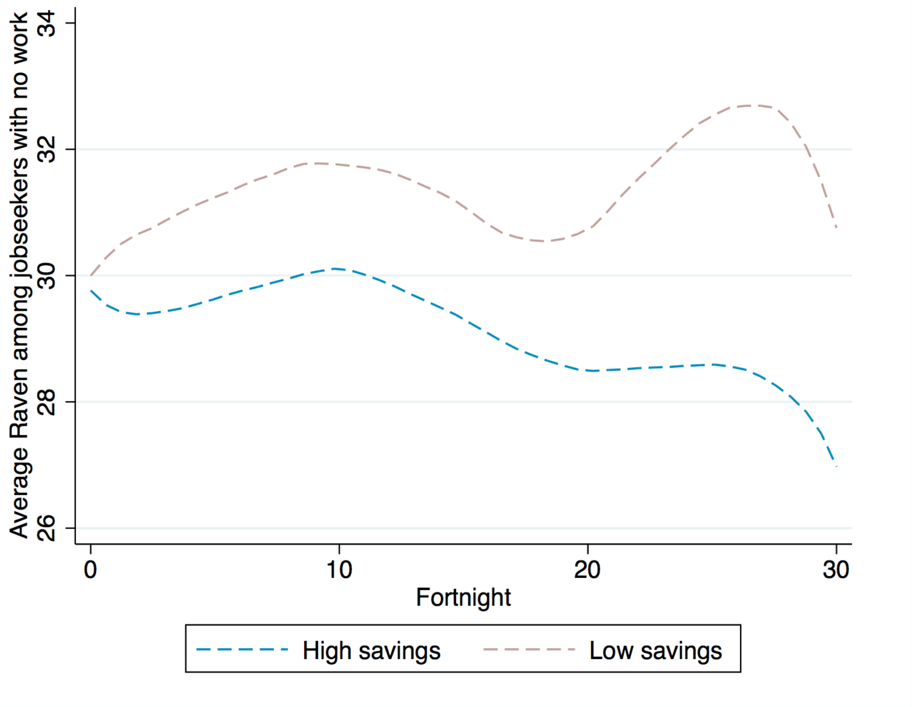
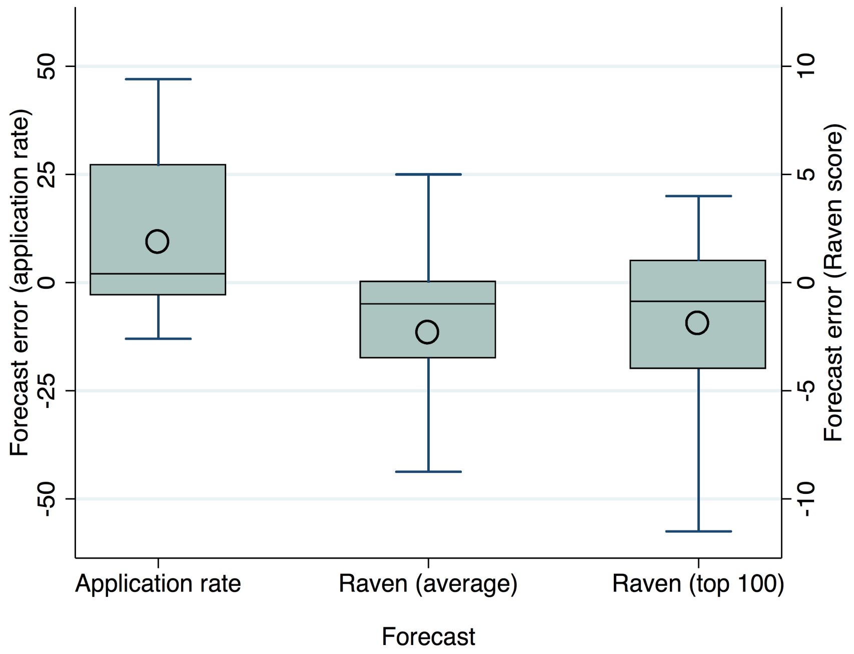


Figure 6: Forecast accuracy of firm managers



Note: The circle shows the mean of the variable, the box shows the interquartile range and the horizontal line inside the box shows the median.

Table 1: Summary statistics and balance

	Mean			N	F-test (p)
	Incentive	High wage	Control		
Female	0.21	0.21	0.21	4689	0.98
Age	26.08	25.95	26.24	4686	0.21
Born in Addis Ababa	0.24	0.24	0.23	4689	0.53
First language is Amharic	0.68	.7	0.67	4689	0.21
Heard about job in newspaper	0.55	0.58	0.56	4689	0.33
Engineering or hard science	0.50	0.49	0.48	4689	0.46
Economics	0.15	0.16	0.17	4689	0.53
Other social science	0.15	0.16	0.14	4689	0.15
Wage work experience (dummy)	0.53	0.53	0.53	4689	0.97
Wage work experience (months)	28.12	28.45	29.06	4689	0.84
Self-employed experience	0.33	0.35	0.35	4689	0.59
Currently unemployed	0.67	0.65	0.64	4689	0.18
Currently wage employed	0.24	0.26	0.27	4689	0.18

The last column shows the p-value for an F-test of the null hypothesis that the characteristics of applicants are balanced across treatments.

Table 2: Application rates

	Application
	(1)
Incentive	.115
	(.016)***
High wage	.187
	(.016)***
Control mean	.411
Incentive = Wage (p)	.000
Obs.	4689

Notes: OLS regression. The second to last row reports the p-value of an F-test of the null hypothesis that the two treatments have the same effect. Robust standard errors reported in parenthesis. * p<0.1, ** p<0.05, *** p< 0.01.

Table 3: Cognitive ability

	Mean	Percentile				
		90th	75th	50th	25th	10th
	(1)	(2)	(3)	(4)	(5)	(6)
Incentive	.248	.229	.229	.17	.412	.079
	(0.112)**	(0.110)**	(0.117)**	(0.133)	(0.173)**	(0.250)
	[0.074]*	[0.074]*	[0.074]*	[0.243]	[0.074]*	[0.751]
	[0.081]*	[0.115]†	[0.148]†	[0.607]	[0.053]*	[1.000]
High Wage	.194	.202	.227	.075	.28	.155
	(0.110)*	(0.108)*	(0.112)**	(0.131)	(0.165)*	(0.227)
	[0.136]†	[0.136]†	[0.136]†	[0.568]	[0.136]†	[0.568]
	[0.225]	[0.182]	[0.130]†	[0.852]	[0.271]	[0.743]
Incentive = Wage (p)	0.574	0.795	0.983	0.448	0.371	0.741
Obs.	2386	2386	2386	2386	2386	2386

Notes: Estimates from OLS (Column 1) and quantile (Columns 2-6) regressions. The second to last row reports the p-value of an F-test of the null hypothesis that the treatments have the same effect. Robust standard errors reported in parenthesis. Sharpened q-values (Benjamini et al., 2006) reported in brackets. The first q-value controls the false discovery rate for multiple comparisons related to the same index. The second q-value controls the false discovery rate for multiple comparisons across indices. † q<0.15, * q<0.1, ** q<0.05, *** q< 0.01.

Table 4: Identification

Structural parameters (13)		Moments (14)
Quality and costs: μ_{T_l}, σ_{T_l}, μ_{C_l}, μ_{T_h}, σ_{T_h}, μ_{C_h}	$\Leftrightarrow$	$\mathrm{E}[T\|C < c^*, B = b_z, \texttt{control}]$, $\mathrm{SD}[T\|C < c^*, B = b_z, \texttt{control}]$, $\Pr[C < c^*\|B = b_z, \texttt{control}]$ for $z \in \{l, h\}$
Shocks and st. dev. of costs: σ_{C_l}, σ_{C_h}, $\tau_{\texttt{incentive}}$, $\tau_{\texttt{wage}}$	$\Leftrightarrow$	$\Delta\text{Applications}[B = b_z, \texttt{incentive}]$ $\Delta\text{Applications}[B = b_z, \texttt{wage}]$ for $z \in \{l, h\}$
Covariance and selectivity: a, σ_{TC_l}, σ_{TC_h}	$\Leftrightarrow$	$\Delta\text{ApplicantQuality}[B = b_z, \texttt{incentive}]$, $\Delta\text{ApplicantQuality}[B = b_z, \texttt{wage}]$ for $z \in \{l, h\}$

Table 5: Parameter estimates

	Low B	High B
μ_T	45.248	45.040
	(1.641)	(1.069)
σ_T	13.615	13.743
	(0.673)	(0.659)
μ_C	144.960	206.690
	(42.650)	(58.464)
σ_C	233.860	254.410
	(49.410)	(60.168)
ρ_{TC}	0.624	0.467
	(0.087)	(0.080)
a	48.923	
	(3.879)	
$\tau_{incentive}$	33.025	
	(16.973)	
τ_{wage}	89.519	
	(118.527)	

Notes: Estimates from classical minimum distance estimator. Standard errors obtained through a bootstrap of the structural estimation reported in parenthesis. The bootstrap includes the estimation of B and the demediation procedure.

Table 6: The returns of the interventions

	Incentive given to...			High wage offer
	All applicants	Marginal applicants	All hires	
Internal Rate of Return	11.0	166.9	330.0	<0
		Costs		
Recruitment costs (month 0)	- 391	- 149	- 391	- 468
Cost of incentive (month 0)	-5868	- 600	-1009	0
Wage costs (months 1-3)	0	0	0	-9600
		Benefits		
Value of higher ability (months 1-42)	182	66	182	215

Table 7: The selection mechanism

	Average Raven score among jobseekers	
	(1)	(2)
Fortnight	.035	.010
	(.028)	(.018)
Low cost	-.957	3.254
	(.562)*	(.516)***
Low cost * Fortnight	-.101	-.096
	(.037)***	(.038)**
Const.	31.060	28.564
	(.397)***	(.237)***
Low cost =	High savings	Low Distance
Fortnight + Low cost * Fortnight = 0 (p)	0.007	0.034
Obs.	52	52

Notes: OLS regression. Robust standard errors reported in parenthesis. * p<0.1, ** p<0.05, *** p< 0.01.

Table 8: Firm managers' ranking of workers

	Ranked first		Ranked first or second
	(1)	(2)	(3)
Incentive	0.369	0.456	0.374
	(0.053)***	(0.064)***	(0.055)***
High wage	0.154		0.287
	(0.048)***		(0.062)***
Control mean	0.159	0.272	0.446
Incentive = Wage (p)	0.001		0.069
No. managers	195	195	195
Obs.	585	390	585

Notes: OLS regression. The unit of observation is an applicant-manager pair. We thus have three observations per manager. In column two we drop applicants from the high wage group. Standard errors clustered at the manager level reported in parenthesis. * p<0.1, ** p<0.05, *** p< 0.01.

Appendix

A.1 Proofs

Proposition 1 (Cut-off existence). *For $B = b_z > 0$, there is at least one cost level c_z^* such that $0 < c_z^* < b_z$ and $\alpha(c_z^*) = k(c_z^*)$.*

Proof. Note that $\alpha(0+\epsilon) > k(0+\epsilon)$, for some small positive ϵ. And similarly, $\alpha(b_z - \epsilon') < k(b-\epsilon')$ for some positive ϵ'. Hence, given that both $k(c)$ and $\alpha(c)$ are continuous, it must be the case that they cross at least once as c traverses the interval $(0, b)$. Naturally, this reasoning applies to all B-types; so dropping the subscripts is without loss of generality. $\square$

Proposition 2 (Cut-off uniqueness). *Suppose $\rho_z < \frac{\sqrt{2\pi}\sqrt{1-\rho_z^2}\sigma_{C_z}}{b_z}$ for $B = b_z > 0$. Then there is exactly one cost level c_z^* such that $0 < c_z^* < b_z$ and $\alpha(c_z^*) = k(c_z^*)$.*

Proof. Proposition 1 shows that $\alpha(c)$ and $k(c)$ cross at least once as c traverses the interval $(0, b)$. Hence, to show that there is one and only one cost level c^* for which $\alpha(c^*) = k(c^*)$, it suffices to check that the derivative of $H(c) \equiv \alpha(c) - k(c)$ with respect to c is negative in the interval $(0, b)$.

Since

$$T|C = c \sim \mathcal{N}\left(\mu_T + \frac{\sigma_T}{\sigma_C}\rho(c - \mu_C), (1 - \rho^2)\sigma_T^2\right)$$

We can write

$$H(c) \equiv \Pr\{T > a|C = c\} - \frac{c}{b} = 1 - \Phi\left(\frac{a - \mu_T - \frac{\sigma_T}{\sigma_C}\rho(c - \mu_C)}{\sqrt{1-\rho^2}\sigma_T}\right) - \frac{c}{b}$$

Differentiating with respect to c we get

$$\alpha'(c) - k'(c) = \frac{\rho}{\sqrt{2\pi}\sqrt{1-\rho^2}\sigma_C}\exp\left\{-\frac{\left[a - \mu_T - \frac{\sigma_T}{\sigma_C}\rho(c - \mu_C)\right]^2}{2(1-\rho^2)\sigma_T^2}\right\} - \frac{1}{b}$$

When $\rho < 0$, the derivative is always negative. So, by Proposition 1, $H(c)$ has at least one root. By monotonicity, this root is unique. When $\rho = 0$, $\alpha(c)$ is horizontal. Here a similar argument applies, showing that the root is unique. When $\rho > 0$, note that the exponential function is bounded above by 1. Hence, the derivative is negative whenever

$$\frac{\rho}{\sqrt{2\pi}\sqrt{1-\rho^2}\sigma_C} - \frac{1}{b} < 0 \iff \rho < \frac{\sqrt{2\pi}\sqrt{1-\rho^2}\sigma_C}{b}$$

As before, the reasoning above applies to all B-types; so dropping the subscripts is without loss of generality. $\square$

Proposition 3 (Treatment effect on applications)**.** *The interventions increase application rates.*

Proof. Note that c^* is defined as the level of C for which the cost and job offer curves cross. Hence, we have that

$$H(c^*;\tau') = \alpha(c^*) - k(c^*;\tau') = \Pr(T > a|C = c^*) - \frac{c^* - \tau'}{b} = 0$$

So we can then use implicit differentiation on $H(\cdot)$ to establish the direction of the shock. This gives:

$$\frac{dc^*}{d\tau'} = -\frac{\partial H(c^*;\tau')/\partial\tau'}{\partial H(c^*;\tau')/\partial c^*} > 0$$

To see why this object is positive, note that (i) the numerator in the fraction is positive, since $\partial H(c^*;\tau')/\partial\tau' = 1/b$; and (ii) the denominator, as shown in the proof of Proposition 2, is negative by Assumption 3. This, in turn, means that $c^* < c^{*\prime}$ – the cutoff moves to the right. Since $\Pr(C < c^*) < \Pr(C < c^{*\prime})$, this shows that the application incentive leads to more applications.

A similar reasoning applies to the wage offer, for which we have

$$H(c^*;\tau'') = \alpha(c^*) - k(c^*;\tau'') = \Pr(T > a|C = c^*) - \frac{c^*}{b + \tau''} = 0$$

where

$$\frac{dc^*}{d\tau''} = -\frac{\partial H(c^*;\tau'')/\partial\tau''}{\partial H(c^*;\tau'')/\partial c^*} > 0$$

Figure 2 illustrates. $\square$

Proposition 4 (Treatment effect on quality)**.** *For each $B = b_z > 0$, the interventions increase the quality of the applicant pool if and only if $\rho_z > 0$.*

Proof. The following are two well known results for Normal random variables:

(i) $E(T \mid C) = \mu_T + \frac{\sigma_T}{\sigma_C}\rho(C - \mu_C)$

(ii) $E(C \mid C < c^*) = \mu_c - \sigma_c \frac{\phi\left(\frac{c^* - \mu_C}{\sigma_C}\right)}{\Phi\left(\frac{c^* - \mu_C}{\sigma_C}\right)}$

These two results can be used in conjunction with the law of iterated expectations to derive an expression for the quality of applicants:

$$
\begin{aligned}
E(T|C < c^*) &= E(E(T|C)|C < c^*) \\
&= E\left(\mu_T + \frac{\sigma_T}{\sigma_C}\rho(C - \mu_C) \mid C < c^*\right) \\
&= \mu_T - \frac{\sigma_T}{\sigma_C}\rho\ \left(\mu_c - E\left(C \mid C < c^*\right)\right) \\
&= \mu_T - \rho\,\sigma_T\,\frac{\phi\left(\frac{c^* - \mu_C}{\sigma_C}\right)}{\Phi\left(\frac{c^* - \mu_C}{\sigma_C}\right)} \\
&= \mu_T - \rho\,\sigma_T\,\lambda\left(\frac{c^* - \mu_C}{\sigma_C}\right) \qquad \text{(A.1)}
\end{aligned}
$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the PDF and CDF of the standard normal distribution, and $\lambda(\cdot)$ is often called the Inverse Mills Ratio.

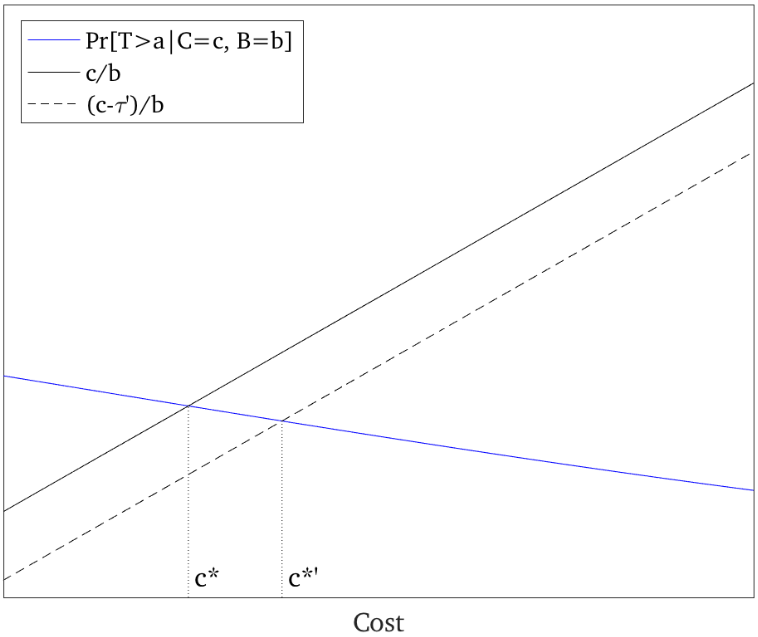
From Proposition 3 we know that both interventions ($\tau\prime$ and $\tau\prime\prime$) operate via shifts in application cut-offs—and we know that for both interventions, the shifts go in the same direction. Hence we complete the proof by differentiating with respect to τ.

$$
\frac{d}{d\tau}E(T|C < c^*(\tau)) = -\frac{\rho\sigma_T\frac{dc^*(\tau)}{d\tau}\frac{d\lambda(c)}{dc}}{\sigma_C}
$$

The sign of the derivative is positive if and only if ρ is positive. This follows from the fact that $\frac{d\lambda(c)}{dc} < 0$ (a result that is easy to check for the Normal distribution) and $\frac{\partial c^*(\tau)}{\partial\tau} > 0$ (Proposition 3). □

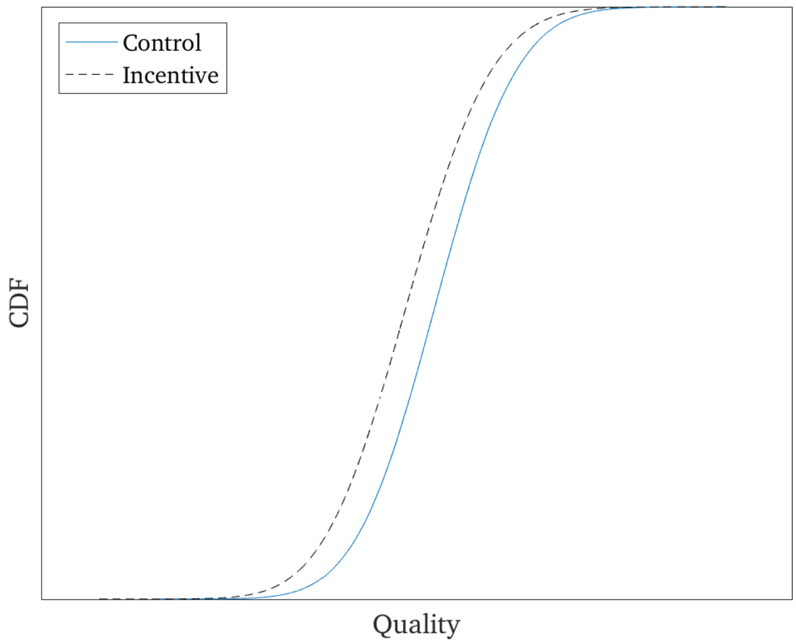
A.2 Figures and Tables

Figure A.1: The application decision



Note: $\rho_{TC} < 0$

Figure A.2: Predicted impacts on the distribution of applicant quality



Note: $\rho_{TC} < 0$

Figure A.3: Attrition

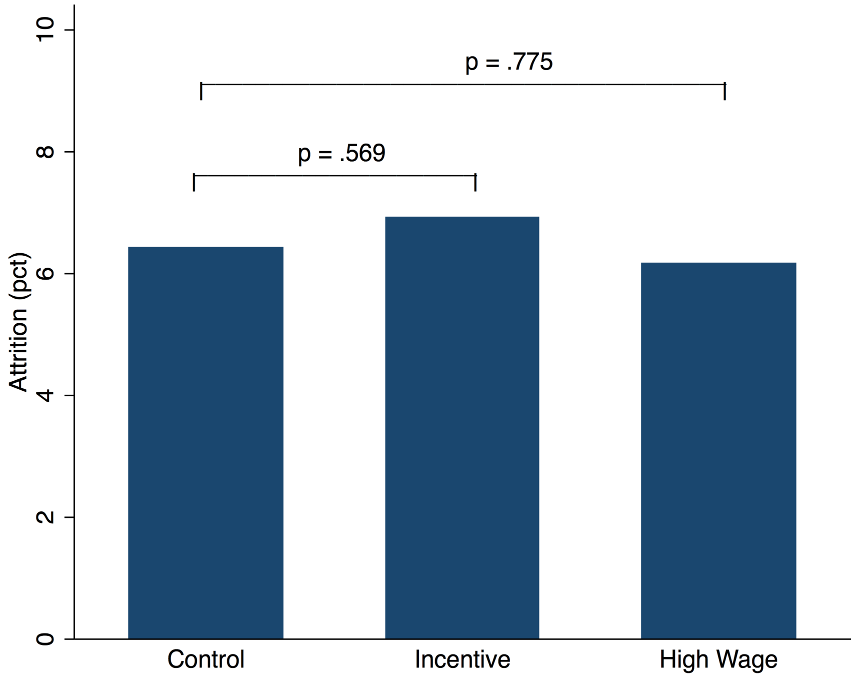


Figure A.4: The proportion of female top applicants

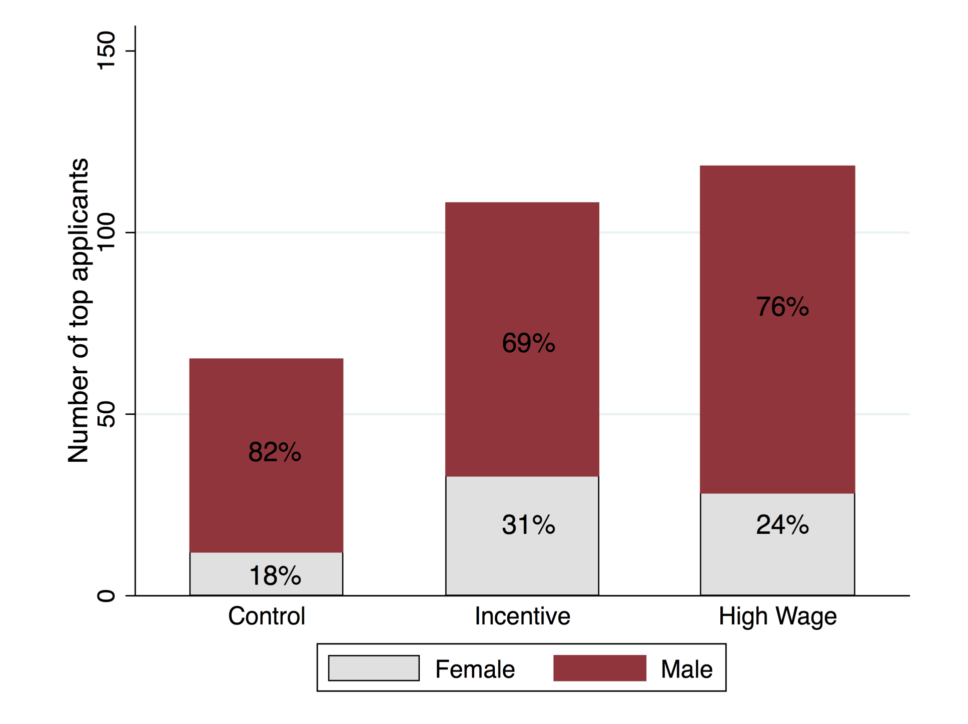
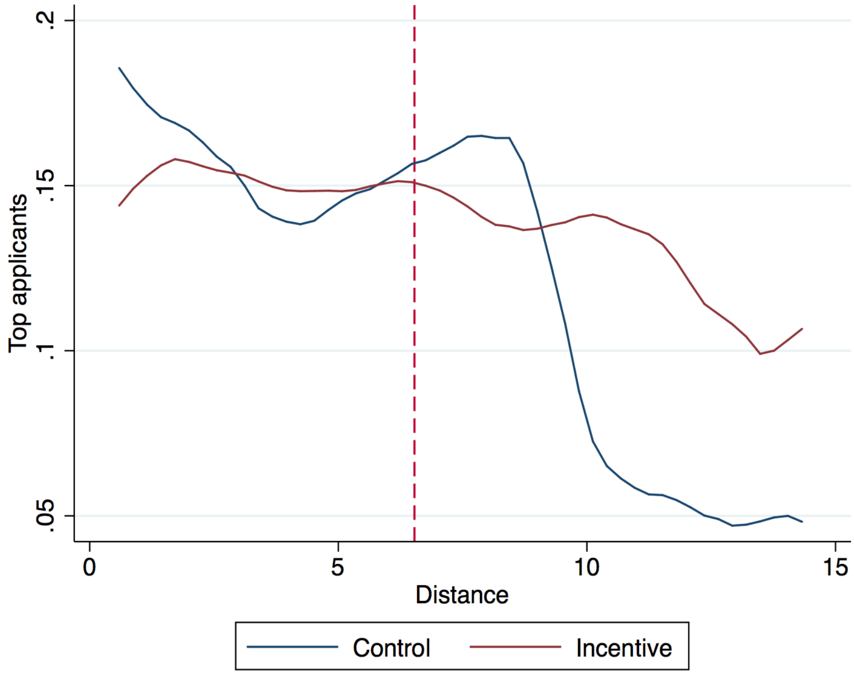


Figure A.5: Heterogeneity by distance to the application centre



Note: the vertical line represents the median distance to the application centre.

Figure A.6: Impacts on the distribution of applicant cognitive ability. High Wage treatment

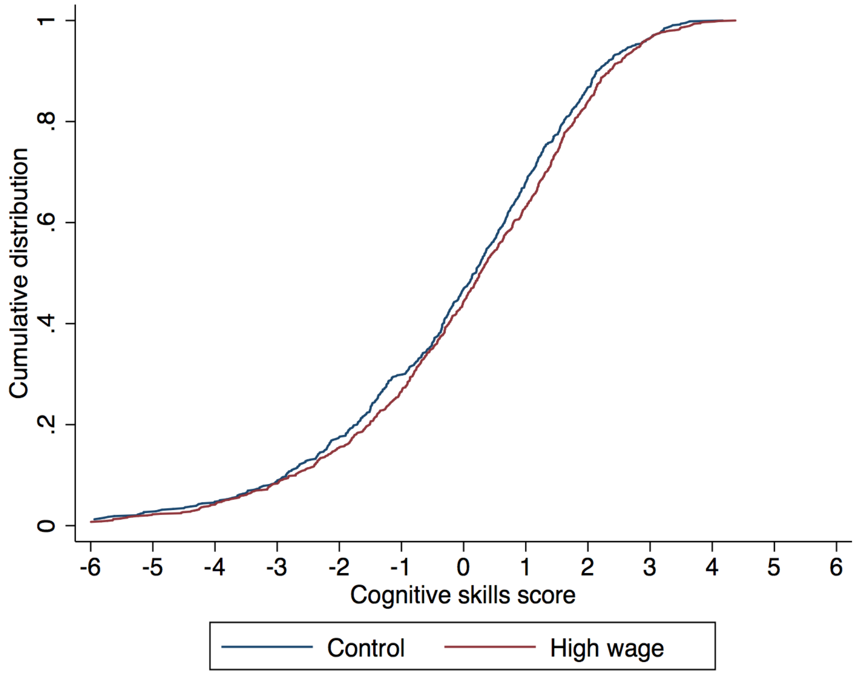


Figure A.7: Sample CV

Candidate 1

Age: 33

Education

Highest level of education: BA (BSc) degree

Field of study: Natural and Computational Sciences

Average grade/GPA: 2.57

Work Experience

Has work experience? Yes

Last employer: MAYLEKO LOAGE

Type of employer: Private business

Test scores

Cognitive ability score: 410

Non-cognitive ability score: 600

Table A.1: Occupations included in the firm survey

Job code	Description	Examples
43-1	Supervisors of Support Workers	First-Line supervisors of office support workers
43-2	Communications Equipment Operators	Telephone operators
43-3	Financial Clerks	Bill and Account Collectors; Bookkeeping, accounting, and auditing clerks
43-4	Information and Record Clerks	Correspondence clerks, credit checkers, customer service representatives
43-9	Other Administrative Support Workers	Computer and data entry operators, claims processing clerks
13-1	Business Operations Specialists	Buyers and purchasing agents, cost estimators, claim checkers, logisticians
13-2	Legal Support Workers	Legal assistants, court workers

Table A.2: Psychometrics Validity Checks: Cronbach α

	Raw	Ipsatised	Laajaj and Macours (2017)
Conscientiousness	.59	.70	.51
Neuroticism	.60	.62	.31
Grit	.53	.72	

Table A.3: Indices of applicant quality

Index	Variable	Measure
Cognitive ability	Raven	No. of correct answers
	Stroop	Time in seconds
	Stroop	No. mistakes
Non-cognitive ability	Conscientiousness	BFI44 score
	Neuroticism	BFI44 score
	Grit	Score on grit scale
Experience	Routine tasks	No. months
	Managerial tasks	No. months
	Problem solving tasks	No. months

Table A.4: Components of index

	Raven	Stroop time	Stroop mistakes
	(1)	(2)	(3)
Incentive	1.155	-2.601	-0.050
	(0.618)*	(1.046)**	(0.188)
	[0.092]	[0.039]	[0.791]
High wage	0.591	-0.982	-0.302
	(0.618)	(1.046)	(0.188)*
	[0.337]	[0.337]	[0.297]
Control mean	38.593	117.304	3.854
F-test Incentive = Wage (p)	0.307	0.078	0.098
Obs.	2397	2386	2388

Notes: OLS regression. The second to last row reports the p-value of an F-test of the null hypothesis that the treatments have the same effect. 'Raven' is the number of correctly answered questions on the Raven test. 'Stroop time' is the number of seconds required to complete the Stroop test. 'Stroop mistakes' is the number of mistakes made in the Stroop task. The negative coefficients on 'Stroop time' and 'Stroop mistakes' indicate better performance. Sharpened q-values (Benjamini et al., 2006) reported in brackets. q-values control the false discovery rate for the multiple comparisons reported in the same row of the table. Robust standard errors reported in parenthesis. * p<0.1, ** p<0.05, *** p< 0.01.

Table A.5: Proportion of applicants who score above a threshold

	Threshold (percentile in control group distribution)				
	90th	75th	50th	25th	10th
	(1)	(2)	(3)	(4)	(5)
Incentive	.044	.053	.029	.04	.009
	(0.017)**	(0.024)**	(0.026)	(0.022)*	(0.016)
High Wage	.027	.052	.016	.032	.006
	(0.016)*	(0.023)**	(0.026)	(0.022)	(0.015)
Incentive = Wage (p)	0.295	0.966	0.596	0.654	0.865
Obs.	2386	2386	2386	2386	2386

Notes: OLS regression. The second to last row reports the p-value of an F-test of the null hypothesis that the treatments have the same effect. Robust standard errors reported in parenthesis. ** p<0.1, ** p<0.05, *** p< 0.01.

Table A.6: GPA

	Mean	Percentile				
		90th	75th	50th	25th	10th
	(1)	(2)	(3)	(4)	(5)	(6)
Incentive	.049	.070	.050	.060	.040	.080
	(.025)**	(.044)	(.034)	(.033)*	(.034)	(.039)**
High wage	.012	-.010	-.010	.030	1.60e-15	.040
	(.024)	(.042)	(.034)	(.032)	(.033)	(.038)
Control group value	2.943	3.560	3.270	2.930	2.600	2.320
F-test Incentive = Wage (p)	.088	.042	.056	.32	.195	.263
Obs.	2285	2285	2285	2285	2285	2285

Notes: Estimates from OLS (Column 1) and quantile (Columns 2-6) regressions. The second to last row reports the p-value of an F-test of the null hypothesis that the treatments have the same effect. Robust standard errors reported in parenthesis. * p<0.1, ** p<0.05, *** p< 0.01.

Table A.7: Non-cognitive ability

	Mean	Percentile				
		90th	75th	50th	25th	10th
	(1)	(2)	(3)	(4)	(5)	(6)
Incentive	-.095	-.155	-.161	-.004	-.141	-.137
	(0.125)	(0.134)	(0.132)	(0.157)	(0.199)	(0.229)
	[0.658]	[0.658]	[0.658]	[0.980]	[0.658]	[0.658]
	[0.566]	[0.369]	[0.336]	[1.000]	[0.717]	[1.000]
High Wage	.17	0	-.039	.257	.162	.26
	(0.118)	(0.127)	(0.129)	(0.147)*	(0.203)	(0.222)
	[0.450]	[1.000]	[0.917]	[0.450]	[0.637]	[0.483]
	[0.225]	[1.000]	[0.764]	[0.245]	[0.637]	[0.725]
Incentive = Wage (p)	0.015	0.235	0.288	0.038	0.095	0.091
Obs.	2373	2373	2373	2373	2373	2373

Notes: Estimates from OLS (Column 1) and quantile (Columns 2-6) regressions. The second to last row reports the p-value of an F-test of the null hypothesis that the treatments have the same effect. Robust standard errors reported in parenthesis. Sharpened q-values (Benjamini et al., 2006) reported in brackets. The first q-value controls the false discovery rate for multiple comparisons related to the same index. The second q-value controls the false discovery rate for multiple comparisons across indices. † q<0.15, * q<0.1, ** q<0.05, *** q< 0.01.

Table A.8: Experience

	Mean	Percentile				
		90th	75th	50th	25th	10th
	(1)	(2)	(3)	(4)	(5)	(6)
Incentive	-.091	.163	-.044	0	0	0
	(0.158)	(0.850)	(0.117)	(0.008)	(0.004)	(0.002)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
	[0.566]	[0.848]	[0.704]	[1.000]	[1.000]	[1.000]
High Wage	-.063	.302	-.088	0	0	0
	(0.157)	(0.733)	(0.147)	(0.008)	(0.004)	(0.002)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
	[0.687]	[1.000]	[0.764]	[1.000]	[1.000]	[1.000]
Incentive = Wage (p)	0.808	0.811	0.718	1.000	1.000	1.000
Obs.	2311	2311	2311	2311	2311	2311

Notes: Estimates from OLS (Column 1) and quantile (Columns 2-6) regressions. The second to last row reports the p-value of an F-test of the null hypothesis that the treatments have the same effect. Robust standard errors reported in parenthesis. Sharpened q-values (Benjamini et al., 2006) reported in brackets. The first q-value controls the false discovery rate for multiple comparisons related to the same index. The second q-value controls the false discovery rate for multiple comparisons across indices. † q<0.15, * q<0.1, ** q<0.05, *** q< 0.01.

Table A.9: Outcomes in 30 days after first phone call

	Applications	Money	Time	Interviews	Offers	Has job
	(1)	(2)	(3)	(4)	(5)	(6)
Incentive	-.01	1.236	27.37	-.012	-.013	-.002
	(0.072)	(4.889)	(25.202)	(0.024)	(0.013)	(0.007)
High Wage	-.078	.29	-11.607	-.04	-.028	-.019
	(0.066)	(4.800)	(23.427)	(0.023)*	(0.012)**	(0.004)***
Control group mean	1.574	60.914	392.362	.31	.103	.048
F-test Incentive = Wage (p)	0.313	0.852	0.107	0.231	0.233	0.012
Obs.	4329	4329	4329	4329	4329	4329

Notes: OLS regression. Robust standard errors reported in parenthesis. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A.10: Heterogeneous impacts on applications

Heterogeneity by	Gender	Age	Employment	Experience	Value job	Credit constraints
	(1)	(2)	(3)	(4)	(5)	(6)
Impacts for	Female	Young	Employed	High experience	High value	High constraint
Incentive	.073	.101	.158	.136	.091	.128
	(0.039)*	(0.027)***	(0.030)***	(0.025)***	(0.025)***	(0.022)***
High wage	.19	.196	.147	.156	.206	.184
	(0.038)***	(0.027)***	(0.030)***	(0.025)***	(0.024)***	(0.022)***
Control mean	.429	.464	.27	.329	.504	.373
Incentive = Wage (p)	0.002	0.000	0.704	0.436	0.000	0.013
Impacts for	Male	Old	Not employed	Low experience	Low value	Low/No constraint
Incentive	.125	.121	.097	.092	.137	.083
	(0.020)***	(0.023)***	(0.021)***	(0.025)***	(0.025)***	(0.034)**
High wage	.185	.179	.214	.217	.171	.183
	(0.020)***	(0.023)***	(0.021)***	(0.024)***	(0.024)***	(0.033)***
Control mean	.407	.372	.473	.492	.318	.511
Incentive = Wage (p)	0.003	0.015	0.000	0.000	0.179	0.002
F-test No het. incentive (p)	0.235	0.567	0.098	0.209	0.186	0.265
F-test No het. wage (p)	0.891	0.631	0.065	0.080	0.302	0.980
Obs.	4689	4686	4689	4686	4686	4274

Notes: OLS regressions. Column headings indicate the dimension of heterogeneity studied. For each treatment, the last panel reports the p-value of an F-test of the null hypothesis that the effect of treatment is not heterogenous across the dimension under study. Robust standard errors reported in parenthesis. * p<0.1, ** p<0.05, *** p< 0.01.

Table A.11: Heterogeneous impacts on other job search

Heterogeneity by	Gender	Age	Employment	Experience	Value job	Credit constraints
	(1)	(2)	(3)	(4)	(5)	(6)
Impacts for	Female	Young	Employed	High experience	High value	High constraint
Incentive	-.014	-.01	-.027	-.073	.013	-.005
	(0.142)	(0.120)	(0.100)	(0.102)	(0.131)	(0.103)
High wage	-.128	-.174	-.107	-.129	-.133	-.161
	(0.137)	(0.117)	(0.101)	(0.105)	(0.110)	(0.090)*
Control mean	1.407	1.78	1.093	1.322	1.821	1.611
Incentive = Wage (p)	0.383	0.110	0.398	0.499	0.205	0.093
Impacts for	Male	Old	Not employed	Low experience	Low value	Low/No constraint
Incentive	.005	.002	.02	.063	-.012	.085
	(0.094)	(0.107)	(0.105)	(0.120)	(0.087)	(0.118)
High wage	-.083	-.025	-.059	-.067	-.04	.125
	(0.083)	(0.089)	(0.092)	(0.097)	(0.090)	(0.119)
Control mean	1.617	1.41	1.78	1.812	1.312	1.44
Incentive = Wage (p)	0.300	0.791	0.405	0.259	0.742	0.744
F-test No het. incentive (p)	0.908	0.940	0.744	0.387	0.873	0.566
F-test No het. wage (p)	0.781	0.314	0.724	0.661	0.515	0.055
Obs.	4328	4325	4329	4326	4325	4217

Notes: OLS regressions. Column headings indicate the dimension of heterogeneity studied. For each treatment, the last panel reports the p-value of an F-test of the null hypothesis that the effect of treatment is not heterogenous across the dimension under study. Robust standard errors reported in parenthesis. * p<0.1, ** p<0.05, *** p< 0.01.

Table A.12: Heterogeneous impacts on cognitive ability

Heterogeneity by	Gender	Age	Employment	Experience	Value job	Credit constraints
	(1)	(2)	(3)	(4)	(5)	(6)
Impacts for	Female	Young	Employed	High experience	High value	High constraint
Incentive	1.153	.411	-.116	-.006	.407	.27
	(0.270)***	(0.169)**	(0.214)	(0.154)	(0.152)***	(0.147)*
High wage	.447	.244	-.064	-.044	.297	.058
	(0.272)	(0.171)	(0.214)	(0.152)	(0.150)**	(0.145)
Incentive = Wage (p)	0.000	0.241	0.794	0.798	0.388	0.084
Impacts for	Male	Old	Not employed	Low experience	Low value	Low/No constraint
Incentive	.008	.096	.377	.428	.033	.155
	(0.121)	(0.148)	(0.129)***	(0.155)***	(0.160)	(0.196)
High wage	.123	.152	.272	.354	.051	.474
	(0.118)	(0.142)	(0.127)**	(0.152)**	(0.157)	(0.191)**
Incentive = Wage (p)	0.295	0.672	0.343	0.562	0.903	0.076
F-test No het. incentive (p)	0.000	0.161	0.049	0.047	0.091	0.638
F-test No het. wage (p)	0.274	0.679	0.175	0.064	0.258	0.083
Obs.	2386	2384	2386	2385	2384	2182

Notes: OLS regressions. Column headings indicate the dimension of heterogeneity studied. For each treatment, the last panel reports the p-value of an F-test of the null hypothesis that the effect of treatment is not heterogenous across the dimension under study. Robust standard errors reported in parenthesis. * p<0.1, ** p<0.05, *** p< 0.01.

Table A.13: Heterogeneous impacts on top applicants (75th percentile)

Heterogeneity by	Gender	Age	Employment	Experience	Value job	Credit constraints
	(1)	(2)	(3)	(4)	(5)	(6)
Impacts for	Female	Young	Employed	High experience	High value	High constraint
Incentive	.24	.094	.005	.018	.075	.047
	(0.052)***	(0.035)***	(0.046)	(0.035)	(0.032)**	(0.030)
High wage	.097	.063	.041	.04	.047	.03
	(0.047)**	(0.034)*	(0.047)	(0.035)	(0.030)	(0.029)
Incentive = Wage (p)	0.004	0.351	0.377	0.503	0.345	0.552
Impacts for	Male	Old	Not employed	Low experience	Low value	Low/No constraint
Incentive	.003	.014	.073	.081	.029	.048
	(0.026)	(0.031)	(0.027)***	(0.031)**	(0.034)	(0.041)
High wage	.039	.044	.057	.061	.065	.111
	(0.026)	(0.031)	(0.026)**	(0.030)**	(0.034)*	(0.041)***
Incentive = Wage (p)	0.142	0.311	0.535	0.494	0.256	0.117
F-test No het. incentive (p)	0.000	0.090	0.204	0.187	0.320	0.981
F-test No het. wage (p)	0.279	0.670	0.770	0.661	0.693	0.110
Obs.	2386	2384	2386	2385	2384	2182

Notes: OLS regressions. Column headings indicate the dimension of heterogeneity studied. For each treatment, the last panel reports the p-value of an F-test of the null hypothesis that the effect of treatment is not heterogenous across the dimension under study. Robust standard errors reported in parenthesis. * p<0.1, ** p<0.05, *** p< 0.01.

Table A.14: Heterogeneous impacts on quality at the top of the distribution

Heterogeneity by	Gender		Age		Employment		Experience		Value job		Credit constraints	
	p90	p75	p90	p75	p90	p75	p90	p75	p90	p75	p90	p75
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Impacts for	Female		Young		Employed		High Experience		High value job		High constraint	
Incentive	.582	.848	.131	.131	.283	.033	.252	.124	.191	.118	.358	.24
	(0.241)**	(0.230)***	(0.161)	(0.168)	(0.217)	(0.258)	(0.189)	(0.168)	(0.134)	(0.145)	(0.154)**	(0.143)*
High wage	.247	.257	.164	.103	.167	.189	.098	.262	.117	.035	.016	.129
	(0.246)	(0.225)	(0.168)	(0.175)	(0.222)	(0.252)	(0.185)	(0.161)	(0.134)	(0.141)	(0.146)	(0.129)
Incentive = Wage (p)	0.074	0.002	0.821	0.835	0.549	0.513	0.320	0.358	0.563	0.495	0.010	0.414
Impacts for	Male		Old		Not employed		Low experience		Low value job		Low/No constraint	
Incentive	.081	-.069	.266	.086	.227	.185	.244	.176	.337	.124	.088	.163
	(0.126)	(0.131)	(0.160)*	(0.150)	(0.122)*	(0.134)	(0.136)*	(0.155)	(0.190)*	(0.177)	(0.182)	(0.242)
High wage	.175	.126	.143	.244	.136	.139	.165	.136	.269	.333	.408	.579
	(0.121)	(0.130)	(0.152)	(0.139)*	(0.120)	(0.129)	(0.129)	(0.153)	(0.179)	(0.183)*	(0.182)**	(0.244)**
Incentive = Wage (p)	0.420	0.075	0.408	0.260	0.416	0.682	0.534	0.756	0.658	0.170	0.036	0.020
F-test No het. incentive (p)	0.066	0.001	0.553	0.842	0.821	0.602	0.973	0.821	0.530	0.977	0.259	0.786
F-test No het. wage (p)	0.793	0.614	0.927	0.528	0.903	0.860	0.768	0.572	0.496	0.197	0.094	0.102
Obs.	2386	2386	2384	2384	2386	2386	2385	2385	2384	2384	2182	2182

Notes: Quantile regressions. Column headings indicate the dimension of heterogeneity studied. For each treatment, the last panel reports the p-value of an F-test of the null hypothesis that the effect of treatment is not heterogenous across the dimension under study. Robust standard errors reported in parenthesis. * p<0.1, ** p<0.05, *** p< 0.01.

Table A.15: Test effort

	Mistakes	Unfinished
	(1)	(2)
Incentive	.096	-.019
	(.082)	(.038)
High wage	.067	-.028
	(.078)	(.034)
Control mean	.711	.081
F-test Incentive = Wage (p)	.712	.782
Obs.	2316	2332

Notes: OLS regression. The second to last row reports the p-value of an F-test of the null hypothesis that the treatments have the same effect. 'Mistake' is the count of mistakes the applicant has made in the ten parts of the task. 'Unfinished' is the count of the parts of the task that the applicant has not completed. Robust standard errors reported in parenthesis. * p<0.1, ** p<0.05, *** p< 0.01.

Table A.16: Salience of the position

	Correct answer	Absolute mistake
	(1)	(2)
Incentive	.025	-30.614
	(.015)	(18.745)
High wage	.039	-44.817
	(.015)**	(18.100)**
Control mean	.687	167.162
F-test Incentive = Wage (p)	.349	.361
Obs.	4376	3635

Notes: OLS regression. The second to last row reports the p-value of an F-test of the null hypothesis that the treatments have the same effect. 'Correct answer' is a dummy capturing whether the respondent recalled the wage offered correctly. 'Absolute mistake' is the absolute difference between the wage recalled by the respondent and the wage actually offered. The number of observation changes because some individuals report that they do not remember the wage offered. These individuals are included in the regression reported in this first column, but not in the regression reported in the second column. Robust standard errors reported in parenthesis. * p<0.1, ** p<0.05, *** p< 0.01.

Table A.17: Beliefs about labour-market prospects

	Weeks unemployment	Wage
	(1)	(2)
Incentive	-.043	102.065
	(.483)	(173.808)
High wage	-.642	445.483
	(.533)	(128.670)***
Control mean	8.687	5190.663
F-test Incentive = Wage (p)	.235	.051
Obs.	3850	3818

Notes: OLS regression. The second to last row reports the p-value of an F-test of the null hypothesis that the treatments have the same effect. 'Weeks unemployment' captures the number of weeks that the respondent expect she or he would need in order to be offered a job they would be willing to work at. 'Wage' captures the wage that the respondent expects this job will pay. Beliefs about the wage are elicited through the method of Attanasio and Kaufmann (2009), as explained in footnote 22. Robust standard errors reported in parenthesis. * p<0.1, ** p<0.05, *** p< 0.01.

Table A.18: Beliefs about the probability of getting the experiment's job

	Probability of getting experiment's job
	(1)
Incentive	.112
	(0.916)
High Wage	.655
	(0.540)
Control mean	48.537
F-test Incentive = Wage (p)	0.617
Obs.	3447

Notes: OLS regression. The data on beliefs is truncated at the 95th and 5th percentiles of the distribution. The second to last row reports the p-value of an F-test of the null hypothesis that the treatments have the same effect. . Robust standard errors reported in parenthesis. * p<0.1, ** p<0.05, *** p< 0.01.

Table A.19: Beliefs about the attributes of the job

	Holidays	Overtime	Satisfaction	Autonomy	Career	Opportunities	New Skills
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Incentive	.023	.018	.007	-.012	.035	.006	-.005
	(.012)*	(.017)	(.010)	(.017)	(.013)***	(.010)	(.004)
High Wage	.015	.044	.028	-.006	.043	.003	.001
	(.011)	(.017)***	(.009)***	(.017)	(.012)***	(.010)	(.004)
Control group mean	.127	.412	.904	.486	.810	.908	.986
F-test Inc. = Wage (p)	.523	.123	.019	.728	.495	.769	.118
Obs.	4367	4363	4365	4362	4364	4365	4369

Notes: OLS regression. The second to last row reports the p-value of an F-test of the null hypothesis that the treatments have the same effect. 'Holiday' is a dummy variable capturing whether the respondent believes the job has more than four days of holiday per month. 'Overtime' is a dummy variable capturing whether the respondent believes the job will require work in the evenings. 'Satisfaction' is a dummy variable capturing whether the respondent believes the job will be satisfying. 'Autonomy' is a dummy variable capturing whether the respondent believes he or she will have freedom to organise their own schedule at work. 'Career' is a dummy variable capturing whether the respondent believes the experience in this job will help them find other jobs in the future. 'Opportunity' is a dummy variable capturing whether the respondent believes there will be further work opportunities at the Ethiopian Development Research Institute. 'New Skills' is a dummy variable capturing whether the respondent believes they will learn new skills in this job. Robust standard errors reported in parenthesis. * p<0.1, ** p<0.05, *** p< 0.01.

Table A.20: Fit between empirical and simulated moments

Moment	Empirical	Simulated
Low B		
$\Pr[C < c^* \| B = b_l, \texttt{control}]$	47.561	47.573
$\mathrm{E}[T \| C < c^*, B = b_l, \texttt{control}]$	38.198	38.137
$\mathrm{SD}[T \| C < c^*, B = b_l, \texttt{control}]$	11.793	11.768
ΔApplications$[B = b_l, \texttt{incentive}]$	11.169	11.030
ΔApplications$[B = b_l, \texttt{wage}]$	13.442	13.533
ΔQuality$[B = b_l, \texttt{incentive}]$	1.903	1.463
ΔQuality$[B = b_l, \texttt{wage}]$	1.515	1.781
High B		
$\Pr[C < c^* \| B = b_h, \texttt{control}]$	49.667	49.616
$\mathrm{E}[T \| C < c^*, B = b_h, \texttt{control}]$	39.813	39.878
$\mathrm{SD}[T \| C < c^*, B = b_h, \texttt{control}]$	12.715	12.750
ΔApplications$[B = b_h, \texttt{incentive}]$	8.337	9.147
ΔApplications$[B = b_h, \texttt{wage}]$	11.450	10.586
ΔQuality$[B = b_h, \texttt{incentive}]$	2.331	0.909
ΔQuality$[B = b_h, \texttt{wage}]$	0.581	1.047

Table A.21: Elasticity of simulated moments

	μ_T	σ_T	μ_C	σ_C	σ_{TC}	a	τ_{inc}	τ_{wage}
$\mathrm{E}[T\|C < c^*, B = b_h, \texttt{control}]$	1.678	0.058	0.148	0.143	0.145	0.587	0.000	0.000
$\mathrm{SD}[T\|C < c^*, B = b_h, \texttt{control}]$	0.149	1.176	0.039	0.157	0.173	0.157	0.000	0.000
$\mathrm{Pr}[C < c^*\|B = b_h, \texttt{control}]$	4.341	0.468	1.147	0.119	0.113	4.535	0.000	0.000
ΔApplications$[B = b_h, \texttt{incentive}]$	0.466	0.905	0.015	2.671	1.002	1.164	0.995	0.000
ΔApplications$[B = b_h, \texttt{wage}]$	7.340	0.491	0.794	3.061	1.171	8.505	0.000	1.143
ΔQuality$[B = b_h, \texttt{incentive}]$	0.922	1.036	0.382	3.601	2.022	0.514	0.967	0.000
ΔQuality$[B = b_h, \texttt{wage}]$	5.728	0.630	0.391	3.972	2.186	6.764	0.000	1.108

Note: Only moments for the high B group are included.

Table A.22: The selection mechanism: Placebo

	Average Raven score among jobseekers	
	(1)	(2)
Fortnight	-.041	-.037
	(.024)*	(.032)
Placebo	-2.050	2.225
	(.716)***	(.644)***
Fortnight * placebo	.012	.001
	(.050)	(.042)
Const.	30.761	29.374
	(.423)***	(.550)***
Placebo =	Married	High life satisfaction
Obs.	52	52

Notes: OLS regression. ***p< 0.01, **p<0.05, *p<0.1. Robust standard errors are reported in parenthesis.

Table A.23: Transitions, by savings

	Stops Search	Search-to-Work	Search-to-Inactive
	(1)	(2)	(3)
High saving	-.017	-.005	-.012
	(.034)	(.017)	(.031)
Above-average Raven	-.071	-.017	-.054
	(.036)*	(.018)	(.033)
High saving * Above-average Raven	.100	.024	.076
	(.043)**	(.022)	(.039)*
Const.	.362	.086	.276
	(.029)***	(.015)***	(.027)***
Obs.	4795	4795	4795

Notes: OLS regression. The sample comprises all fortnights when an individual is out of work and searching for employment. 'High savings' is a dummy variable that identifies individuals with above-median levels of savings at baseline. 'Above-average Raven' is a dummy variable that identifies individuals who have obtained an above-average Raven score. This average is calculated among those individuals who are out of work and searching for employment in the first fortnight of the panel. Standard errors clustered at the level of the individual jobseeker are reported in parenthesis. * p<0.1, ** p<0.05, *** p< 0.01.

Table A.24: Transitions, by distance from the city centre

	Stops Search	Search-to-Work	Search-to-Inactive
	(1)	(2)	(3)
Low distance	-.015	-.016	.0007
	(.031)	(.015)	(.027)
Above-average Raven	-.016	-.013	-.003
	(.028)	(.015)	(.025)
Low distance * Above-average Raven	.026	.025	.001
	(.040)	(.020)	(.035)
Const.	.357	.090	.266
	(.022)***	(.012)***	(.019)***
Obs.	4795	4795	4795

Notes: OLS regression. The sample comprises all fortnights when an individual is out of work and searching for employment. 'High savings' is a dummy variable that identifies individuals with above-median levels of savings at baseline. 'Above-average Raven' is a dummy variable that identifies individuals who have obtained an above-average Raven score. This average is calculated among those individuals who are out of work and searching for employment in the first fortnight of the panel. Standard errors clustered at the level of the individual jobseeker are reported in parenthesis. * p<0.1, ** p<0.05, *** p< 0.01.

A.3 The value of the job

In this section, we describe how we calculate the value of the experiment's job. This value is given by the stream of utility that the worker obtains if they get the job, minus the stream of utility that the worker would have obtained otherwise:

$$b = \begin{cases} V(j) - V(u) & \text{if currently unemployed} \\ V(j) - V(e) & \text{if currently employed} \end{cases} \tag{A.2}$$

where $V(j)$ is the gross value of the experiment's job, $V(u)$ is the value of being unemployed, and $V(e)$ is the value of being employed at the wage that the market currently pays for the worker's skills (we will refer to this as the 'market wage').

We proceed in two steps. First, we characterise $V(j)$, $V(u)$ and $V(e)$ as functions of the wage paid by the experiment's job, market wages, worker impatience and the probability of finding and losing a job. Second, we forecast the market wage of each worker using a Post-LASSO estimator (Belloni et al., 2014) and make informed assumptions about the other parameters. Throughout this section, we assume that time is discrete and measured in months. We also assume that workers have a time-separable, linear utility function of the following form:

$$U_t = \sum_{k=0}^{T} \delta^{t+k} E[w_{t+k}] \tag{A.3}$$

We start by calculating the value of unemployment. We assume that the worker values non-work time at c (Mas and Pallais, 2017). This includes transfers, the value of leisure, etc... We assume that c is give to the worker at the end of the month. Further, we assume that the worker will find a job in the next period with probability p. The value of being in unemployment is thus given by:

$$\begin{aligned} V(u) &= \delta c + \delta^2 p V(e) + \delta^2 (1-p) V(u) \\ &= \frac{\delta c + \delta^2 p V(e)}{1 - \delta^2 (1-p)} \end{aligned} \tag{A.4}$$

The value of being employed, on the other hand, is given by:

$$\begin{aligned} V(e) &= \delta w + \delta^2 (1-q) V(e) + \delta^2 q V(u) \\ &= \frac{\delta w + \delta^2 q V(u)}{1 - \delta^2 (1-q)} \end{aligned} \tag{A.5}$$

where w is the market wage and q is the probability of losing the job in any given period of time. We can substitute $V(e)$ into (A.4) to derive an expression that defines

$V(u)$ only as a function of the parameters c, w, δ, p and q:

$$\begin{aligned} V(u) &= \frac{\delta c}{1-\delta^2(1-p)} + \frac{\delta^2 p}{1-\delta^2(1-p)} \frac{\delta w + \delta^2 q V(u)}{1-\delta^2(1-q)} \\ &= \left(1 - \frac{\delta^4 pq}{(1-\delta^2(1-p))\,(1-\delta^2(1-q))}\right)^{-1} \times \left(c' + \frac{\delta^2 p}{1-\delta^2(1-p)} w'\right) \end{aligned} \tag{A.6}$$

where $c' = \frac{\delta c}{1-\delta^2(1-p)}$ and $w' = \frac{\delta w}{1-\delta^2(1-q)}$. The value of being employed can be obtained by substituting (A.6) into (A.5).

Finally, the gross value of getting the experiment's job for a worker in treatment group f is given by:

$$V(j) = \sum_{k=0}^{3} \delta^k w_f + \delta^4 \left(pV(e) + (1-p)V(u)\right)$$

The worker will obtain wage w_f for three consecutive months and will then return to unemployment. For simplicity, we assume that work experience in the experiment's job does not affect future wages and that the worker will only hear about new job opportunities in the last month of the job. These assumptions make our estimates of $V(j)$ conservative.

We can now write an expression for the value of the job for an unemployed person. This is given by:

$$V(j) - V(u) = \sum_{k=0}^{3} \delta^k w_f + \delta^4 \left(pV(e) + (1-p)V(u)\right) - V(u)$$

Further, the value of the job for an employed person is given by:

$$V(j) - V(e) = \sum_{k=0}^{3} \delta^k w_f + \delta^4 \left(pV(e) + (1-p)V(u)\right) - V(e)$$

In our second step we forecast market wages. To do this, we use the Post-LASSO estimator recommended by Belloni et al. (2014). This estimator is obtained in two stages. First, we regress individual wages on a large set of covariates, using the LASSO estimator. This allows us to select a sub-set of covariates that can be used for forecasting. Second, we run an OLS regression of wages on the covariates selected in the first stage and use the OLS coefficients to derive a forecast of w for each worker.

The Post-LASSO estimator is recommended to produce forecasts when a large number of potentially informative covariates are available. In these settings, estimators that maximise in-sample fit often have poor out-of-sample properties, as they tend to fit some of the noise in the data. The original LASSO estimator reduces over-fitting by imposing a penalty on non-zero coefficients. More precisely, for a canonical model:

$$y_i = \sum_{j=1}^{p} x_{i,j}\beta_j + u_i \tag{A.7}$$

the LASSO estimator of the parameter vector β is obtained by minimising the following function:

$$\widehat{\beta} = \arg\min_{\beta} \sum^{n} \left(y_i - \sum_{j=1}^{p} x_{i,j}\beta_j \right)^2 - \lambda \sum_{j=1}^{p} |\beta_j|\gamma_j$$

where λ is a penalty parameter and γ_j are penalty loadings. One problem with this estimator is that the non-zero coefficients tend to be biased towards zero. The Post-LASSO estimator reduces this bias by re-estimating the coefficients with OLS.

We use a rich set of variables in order to forecast wages. These variables describe the socio-demographic characteristics of workers, their educational achievements, and their labour market experience. We report the full list of variables in table A.25 below. To maximise the flexibility of our empirical model, we discretise continuous variables and include dummies for each possible discretised value of the variable. Finally, our measure of wages refers to the jobs that subjects found between the two phone interviews. If a subject has not found a new job but was employed at the time of their first phone call, we use the wage of this job.[38] We report the coefficients estimates obtained with the Post-Lasso estimator in Table A.26 below. The first column shows the estimates obtained by using the theoretically optimal penalty and the second column shows the estimate obtained with a manually-set lower penalty, which allows us to capture a number of additional plausible predictors. The predicted values we obtain from these two models are highly correlated. In what follows, we use the predicted values obtained with the optimal loadings.

We make the following assumptions on the remaining parameters. First, we assume that the monthly discounting factor is $\delta = 0.786$. To determine this figure, we use the daily discounting factor estimated in a recent experiment in Nairobi (Balakrishnan et al., 2015). The estimates of Balakrishnan et al. (2015) suggest relatively high levels of impatience, which is consistent with the cross-country survey evidence reported by Falk et al. (2016) for sub-Saharan Africa. Second, we set the probability of finding a job to 14.5 percent and the probability of losing a job to 17.7 percent, respectively. These figures reflect monthly transition rates from non-employment to employment, and vice-versa, which we calculate using the high-frequency panel data collected by Abebe et al.

[38] We make an adjustment to the forecasted wages to ensure that the mean of the forecast matches that of representative data for workers in Addis Ababa of comparable age and level of education.

Table A.25: Variables used to forecast wages

Variable	Description
female	Female
age	Age
age_sq	Age squared
born_aa	Individual was born in Addis Ababa
newspaper	Individual has found out about the vacancy in the newspaper
amharic	First language is Amharic
oromo	First language is Oromifa
engineer	Engineering or hard science background
economics	Economics background
social_scientist	Degree in social science (other than economics)
GPA_dummy_	Dummies for GPA score (1 point intervals)
wexperience	Wage work experience (number of months)
wexperience_sq	Wage work experience squared
e_type_	Dummies for type of employer in last job
wage_dummy_	Dummies for wage earned in last job (2,000 ETB intervals)
sexperience	Individual has experience in self-employment

(2016). Finally, we assume that the value of c is 1,230 ETB. We calculate this figure by using estimates of the value of non-work time from Mas and Pallais (2017) and mean forecasted wages. This figure seems realistic in our context, as unemployed jobseekers report an average monthly expenditure of about 1,000 ETB.

We estimate that the position has positive value for about 61 percent of the individuals in our sample. To confirm that our estimates are informative, we regress the application dummy on our estimate of the value of the job. We find a large and significant correlation: a standard deviation increase in the value of the job is associated with a 10 percentage points increase in application rates. We report the estimates in Table A.27 below.

Table A.26: Post-LASSO regression of wages (control group observations)

	Optimal penalty	Manual penalty
	(1)	(2)
Heard of job on newspaper	108.186	53.964
	(190.313)	(181.409)
Economics background	719.240	683.966
	(226.937)***	(216.406)***
Work experience (months)	18.188	16.961
	(3.241)***	(3.140)***
Worked for private foreign business		170.528
		(672.758)
Age (years)	45.814	23.353
	(35.137)	(34.037)
GPA dummy (2-3)		234.117
		(167.059)
Previous wage dummy (less than 2000)		-153.670
		(233.614)
Previous wage dummy (2000-4000)		966.191
		(383.637)**
Previous wage dummy (4000-6000)	1862.386	2055.467
	(430.861)***	(418.041)***
Previous wage dummy (6000-8000)		4127.487
		(734.743)***
Previous wage dummy (8000-10000)	3682.031	3981.823
	(781.358)***	(754.920)***
Const.	1388.831	1923.725
	(852.186)	(835.074)**
Obs.	361	361

Table A.27: Regression of applications on the value of the job

	Application
	(1)
Value of the job (z score)	.107
	(.007)***
Const.	.512
	(.007)***
Obs.	4686