Delegated Recruitment and Hiring Distortions*

Jacob Kohlhepp† Stepan Aleksenko‡

September 10, 2021

Abstract

We analyze how delegating recruitment influences the search for talent. During search, the recruiter does not learn worker productivity but only forms a belief characterized by an expectation and a variance. We demonstrate that delegation is equivalent to making the search technology less accurate. Delegation results in moral hazard with a multitasking flavor, where the recruiter wastes effort finding low-variance workers at the expense of high-expectation workers. As workers become more homogeneous with respect to productivity variance delegation becomes more efficient. Our model provides a theoretical connection between delegation and variance-based statistical discrimination.

Keywords: moral hazard, delegation, contracting, sequential search, recruit, discrimination in hiring

JEL Codes: D83, D86, J7

*We are grateful for comments and suggestions from Simon Board, Moritz Meyer-ter-Vehn, Tara Sinclair, participants in the Eastern Economic Association Annual conference, participants in the UCLA Theory Proseminar, and three anonymous recruiters.

†Department of Economics, UCLA. Contact: jkohlhepp@ucla.edu.

‡Department of Economics, UCLA. Contact: saleksenko@ucla.edu.

1 Introduction

Talent allocation has always been an economic force at the center of research about inequality, discrimination and productivity. The internet and popular press is brimming with inspiring quotes, white papers and advice all conveying a similar message: people are everything. And yet, the actual search for talented people is frequently a delegated responsibility. Between 2010 and 2018, the fraction of job postings for recruiting roles more than doubled. As of 2020, 18 percent of employed American workers had found their current job through a recruiter or a headhunter (Black, Hasan, and Koning 2020). Delegated recruitment is now a major feature of the labor market landscape, but the impact of this new feature remains unclear. Are recruiters well-aligned agents of the firm, or does delegation introduce distortions into the hiring process?

To understand this question, we propose a model of delegated recruitment. In our model, the principal of the firm (she) employs a recruiter (he) to sequentially search for a worker. The recruiter does not know the exact value of a searched worker's productivity; rather, he holds a belief about worker productivity, which we will assume throughout the paper is characterized by a variance and expectation. Contracts take a refund form, in which the recruiter is paid an amount upfront but must refund a portion of the payment if the hired worker is fired. Throughout the paper, we compare this model to a first-best benchmark, where the firm searches for a worker directly.

In this paper, we characterize the optimal contract through a single first-order condition. The characterization shows that delegation via refund contracts is equivalent to making the search technology less accurate. We provide a necessary and sufficient condition under which there is less search effort exerted under delegation than there is in the first-best. Additionally, the expected productivity distribution among selected workers under delegation is lower than under first-best. Finally, we demonstrate that social surplus and search effort increase and converge to the first-best benchmark as heterogeneity in productivity variance decreases. We show that our results apply naturally to two common parametric distributions.

To fix ideas, consider two candidates for a data science position. Candidate A is traditional: they graduated from a four-year college with a degree in statistics and interned with a prominent firm. Candidate B is nontraditional: they only have a high school degree and are self-taught; however, B won a popular machine-learning competition. When comparing A and B prior to hire, B's productivity might have a higher expectation, but also higher variance than A's. We will show that delegation results in a bias against candidate B in favor of candidate A.

We restrict attention to refund contracts for two reasons. First, anecdotal and survey evidence suggests that this is the main contract form used in practice.[1] Second, search is private to the recruiter and productivity is private to the firm. Refund contracts recognize this reality, and only condition payment on *employment* which is public and easily measured. We take the termination threshold as given and independent of the contract. For many firms, this is reasonable. The person deciding to terminate an employee is generally not the same person who hires the recruiter. Additionally, large firms are generally required by anti-discrimination law to treat employees fairly or equally (Carlsson, Fumarco, and Rooth 2014). This makes it unlikely that a firm can terminate employees differently based on the channel through which they were hired.[2]

Our first result characterizes the general contracting problem through a single first-order condition. The characterization shows that refund contracts induce a special type of moral hazard which is equivalent to making the search technology less accurate in a Blackwell sense. It is as also as if the firm and the recruiter face a canonical multitasking problem, in which the task the firm cares about (maximizing productivity expectation) can only be encouraged together with a wasteful task (minimizing productivity variance). We compare delegation to a first-best benchmark where the firm searches directly for an applicant. Except in knife-edge cases, first-best social surplus and recruitment strategy are not achieved.

We then ask how delegation influences search strategy, and show that when the delegation-induced degradation of the search technology satisfies a weak condition, there is less search effort exerted under delegation. We further show that this implies the expected productivity distribution among selected workers will be lower (in a first-order stochastic dominance sense) under delegation than the first-best. Under an independence assumption, the productivity variance distribution among selected workers is also lower. Finally, we demonstrate social surplus and search effort increase and converge to the first-best benchmark as heterogeneity in productivity variance decreases. This implies that agency loss in this setting crucially depends on the magnitude of differences in productivity uncertainty across candidates. When some workers exhibit less uncertainty (higher variance after search) than others, then the recruiter will waste search effort seeking such workers at the expense of expected productivity.

We explore two testable implications of the model. First, the specific way in which the refund contract warps incentives implies that, even if two workers have the same productivity expectation, the one with lower unobserved productivity variance will be hired. Further, we

1. We interviewed three recruiters, and they reported that such a contract was common and that the trial period is typically around 90 days. A survey by Top Echelon found that 96% of recruiters offer some sort of guarantee that a candidate will stay. Among those, 61% provided a replacement - but not money back - if the candidate failed to stay, while 26% offered a partial or full refund.

2. For a detailed review of the recruiting industry we refer the reader to Cowgill and Perkowski (2020).

show that statistical discrimination against a high-variance group is amplified by delegation: it is higher in a world with delegation than in a world with only direct search. Thus, our model provides a microfoundation for variance-based statistical discrimination. Second, our model predicts that firms are more likely to outsource recruitment during periods and in occupations where observable differences in productivity variance are small across workers. If variance cannot be well-predicted by observables, or productivity variance is indeed similar across worker types, then delegation is more likely.

Taken together, these two mechanisms suggest an interesting vicious cycle. Delegation induces bias against high-variance groups. If low labor market success causes these groups to exit the occupation, then the labor market in the future becomes more homogeneous, with respect to productivity variance. This reduction in heterogeneity makes delegation more appealing, thus increasing the share of firms utilizing recruiters as well as the average bias against high-variance groups.

Our modeling framework is general in that we do not specify an information structure. Instead, we take as primitives the posterior means and variances that result from some updating process. This is similar in spirit to the approach taken in the Bayesian persuasion literature, particularly Gentzkow and Kamenica (2016). A researcher can specify an information structure, derive the implied posterior mean and variance distributions, and apply our results. In our application section, we illustrate this process using a model of hiring which is popular in the discrimination literature that utilizes normal priors and normal signals. Our results can also be applied to models that do not have an information structure at all. Such models include those with match-specific effects and complementarities between types of firms and workers. We describe such an example in the Appendix.

The paper is structured as follows. First, we describe how our work contributes to the existing literature. Second, we introduce the model. Third, we characterize the general solution to the delegated and first-best problems. Fourth, we compare delegation to the first-best benchmark and present comparative statics. Fifth, we apply our results to specific parametric examples and economic situations. Finally, we discuss the implications and conclude.

2 Literature

Our main contribution is to labor market intermediary literature. Within this literature, Cowgill and Perkowski (2020) is the only other paper explicitly focused on agency issues stemming from the firm-recruiter relationship. The authors investigate agency issues arising from balancing firm and worker preferences. Their paper grapples with the idea that recruiters maximize imperfectly aligned objective functions, and to that extent it shares many similarities with our paper. However the papers are otherwise complements: Cowgill and Perkowski (2020) investigates how worker and firm preferences are balanced in light of recruiter reputation while we explore how firms balance misalignment and effort provision when worker productivity is uncertain.

Our paper is most theoretically related to Ulbricht (2016), which explores a general delegated sequential search problem. Like in our setting, Ulbricht considers the case when search is unobserved by the principal, and shows that in an unrestricted contract space, the first-best can be achieved. Unlike Ulbricht (2016), we restrict the set of feasible contracts to what we call refund contracts. We consider the case when searched objects are uncertain and differ in their mean and variance. This combination of a contract restriction and two dimensions of heterogeneity prevents the firm from achieving the first-best.

Our paper is also related to the more general delegated choice literature. The models in this literature feature a principal who must trade-off the comparative advantage of the agent (the agent usually has better information) with the agent's bias. Within this literature, two relevant papers are Armstrong and Vickers (2010) and Alexander Frankel (2014). In both, the preference misalignment between the principal and agent are primitives of the economic model. The authors then focus on optimal delegation schemes given this misalignment. Our paper is different in that we are concerned with how these preferences are misaligned in the first place, and we show that contract restrictions can generate misalignment like that which is described in this literature. For our specific setting with sequential search and refund contracts, we find that the agent overvalues productivity variance and undervalues productivity expectation relative to the principal.

Because our paper extends the delegated choice literature to a specific context, it is similar in spirit to Alex Frankel (2021) and Che, Dessein, and Kartik (2013). Similar to us, these papers explore how a specific form of bias changes delegation. In Alex Frankel (2021), the principal and the agent value hard and soft information about a job candidate differently. In Che, Dessein, and Kartik (2013), the agent values the outside option differently than the principal; there is a sort of status quo misalignment. In our model, the recruiter prefers low-variance candidates more than the firm does, because these candidates are more likely to exceed a minimal level of competence and thus remain at the firm.

More broadly, our work is motivated by a desire to understand the "matching function" which is an important primitive in labor search and matching models. Following Shimer and Smith (2000) and Postel–Vinay and Robin (2002), our paper, like many other papers examining labor market sorting, considers workers with ex-ante productivity heterogeneity. Inspired by the individual specific and match specific productivity components in these models, we can think of the productivity expectation as individual ability, and the productivity variance as generated by match-specific effects which are not known until hire.

3 Model

Players and Actions: There is a single risk neutral firm that wants to fill a single job opening. To do this, she hires a recruiter to search for an ideal candidate. The recruiter is risk neutral and operates a sequential search technology over workers. We assume that workers are not players, and are either fired or quit exogenously when $a < 0$. For simplicity, we assume that the firm proposes the contract, and therefore extracts all surplus.

Search Technology, Information and Workers: The recruiter searches sequentially for a worker by paying a cost c per search. After each search, the recruiter observes a pair of attributes (μ, σ) describing the drawn worker's productivity. Specifically, productivity (a) conditional on the two attributes (μ, σ) is given by:

$$a|(\mu, \sigma) = \mu + \sigma \cdot \varepsilon \tag{1}$$

The random variable ε represents the remaining uncertainty about worker productivity. We make the semi-parametric assumption that it has a symmetric distribution, a zero mean, and a variance normalized to 1.[3] We denote probability density function and cumulative density function of ε by f and F, respectively.[4] Because $E[a|\mu, \sigma] = \mu$ and $Var(a|\mu, \sigma) = \sigma^2$, we refer to μ as productivity expectation, and to σ^2 as productivity variance. These attributes are distributed in the population according to a joint CDF G, and each search is an independent draw from G. After observing (μ, σ), the recruiter can offer the current applicant to the firm or continue their search.[5]

Contracts: The firm is restricted to contracts of the form: $t(a) = \alpha - \beta \mathbb{I}\{a < 0\}$. We call these *refund contracts* because α is the recruiter's payment if the search is successful and β is the refund when the employee is fired or quits for any reason.

Payoffs: If the recruiter rejects the contract, his outside option is assumed to be 0. The firm's ex-post profit is realized productivity less any payments to the recruiter: $\pi(a) = a - t(a)$. The recruiter's ex-post utility consists of payments from the firm less total search cost, which is unknown ex-ante, but is c times the number of searches (N) ex-post: $u(a) = t(a) - N \cdot c$.

3. An example of such distribution could be $\varepsilon \sim N(0, 1)$. Then $a|(\mu, \sigma)$ would be distributed as $N(\mu, \sigma^2)$.
4. For technical reasons, we also assume that ε has a continuously differentiable positive PDF on $\mathbb{R}$.
5. As is well-known, it is without loss to ignore recall of previously searched workers.

We restrict attention to cases where some search is optimal when the firm operates the search technology directly (in the first-best).[6]

3.1 Model Comments

Our framework is general, and is a reduced-form of several more structured models. Consider a model where the recruiter infers a worker's productivity based on observable characteristics, updating a prior over a. We do not specify the information structure, and instead focus on G, the joint distribution of posterior expectations and variances. We require only that posterior productivity can be written as in Equation 1. One such information structure is when all workers are part of a publicly observable group. Prior beliefs about each group are normal with potentially different variances and expectations. During search, the recruiter observes both group membership and a normal signal of ability, with precision that can be different for each group. Then signal precisions, prior variances and group proportions jointly determine the marginal distribution of σ. Likewise, these components and the prior expectation determine the marginal distribution of μ. μ, σ will not be independent in this case, because they both depend on group membership. We illustrate such an example in our application section and use it to explore how delegation influences statistical discrimination.

Our framework also nests other contexts. Consider a model where productivity has a firm, worker, and match-specific component. If the firm component is public knowledge and the recruiter can uncover the individual component through search, we can think of the match-specific component as a form of residual. Then, μ is the expected productivity of a firm and worker type, and σ^2 is the productivity variance of a firm and worker type. We describe how to map our results to an additive specification of such a matching model in the Appendix.

The firm payoff contains a rather than $\max\{a, 0\}$ because employment is an experience good: the firm must hire the employee in order to learn underlying productivity. One might also wonder why the firm does not receive a discounted sum of future profits when the employee is not fired. That is, why does the firm not receive a payoff of the form $a + \sum_t \delta^t \max\{a, 0\}$? This can be rationalized through symmetric learning and downwards rigid wages. Suppose a is productivity net of wages in an initial trial period. After this period, a is public. If $a > 0$, then the employee can negotiate higher wages so that future net productivity is 0. If $a < 0$, then the employee separates from the employer (due to downward rigid wages). Either way, the firm's profit is a less payments to the recruiter.

6. We define some search as the expected number of searches is strictly greater than one. When this condition does not hold, the firm can create a degenerate contract with $\beta = 0$, and the problem becomes uninteresting.

It should be noted that our qualitative results extend to these alternative specifications. To see this, note that both $\max\{a, 0\}$ and $a + \sum_t \delta^t \max\{a, 0\}$ are convex functions of a. Thus, when comparing workers with similar productivity expectations but different variances, the firm will tend to prefer higher variance candidates due to their option value. This increases misalignment between the firm and recruiter.

4 Analysis

In this section, we analyze the first-best benchmark and the actual equilibrium without imposing additional assumptions on G.

4.1 First-Best Benchmark

As a first-best benchmark, we consider the case when the firm can operate the search technology directly.[7] The firm is risk neutral, so it seeks to maximize expected profit. After searching an applicant, expected a is: $E[a|\mu, \sigma] = \mu$. As a result, the firm cares only about productivity expectation (μ). The first-best problem is thus a standard sequential search problem in the style of McCall (1970). The solution is a reservation rule, and the set of acceptable workers is given by an acceptance region of the form $\mathcal{D}_F = \{\mu, \sigma | \mu \geq \mu^*\}$, where μ^* solves:

$$c = \int_{\mu \geq \mu^*} (1 - G_\mu(\mu)) d\mu \tag{2}$$

These results are standard in the sequential search literature; in the interest of completeness, however, we provide derivations in the Appendix. To compare the first-best and equilibrium, it is informative to rewrite Equation 2 in the following way:

$$(\mathbb{E}[\mu|\mu > \mu^*] - \mu^*) \cdot \Pr(\mu > \mu^*) = c \tag{3}$$

From this, we see that μ^* is equal to expected profit from search:

$$\mu^* = \mathbb{E}[\mu|\mu \geq \mu^*] - \frac{c}{Pr(\mu \geq \mu^*)} \tag{4}$$

Since we assume that first-best recruitment is profitable, the right-hand-side is positive and μ^* must also be positive.

7. Equivalently, when there is no contract restriction. The firm would then optimally "sell-the-firm" by taking a fee from the recruiter and allowing the recruiter to be the residual claimant.

4.2 Delegation Equilibrium

We now consider a Perfect Bayesian Equilibrium where the firm must delegate search to the recruiter. The firm does not observe the search strategy of the recruiter. The contract space is also restricted to refund contracts. Such contracts consist of an upfront payment (α) and a refund (β) which is returned to the firm if the candidate terminates. We require that any contract be both individually rational and incentive compatible.

Incentive compatibility requires that the search strategy the firm requests must be the recruiter's optimal sequential search strategy after the contract is accepted. The recruiter is concerned solely with avoiding a refund. Upon searching for a worker, expected utility from selecting that worker given μ, σ is:

$$\alpha - \beta\mathbb{E}[\mathbb{I}\{a < 0\}|(\mu, \sigma)] = \alpha - \beta\left(1 - F\left(\frac{\mu}{\sigma}\right)\right) = (\alpha - \beta) + \beta F\left(\frac{\mu}{\sigma}\right) \tag{5}$$

where F denotes the standard error ε cumulative density function (CDF). Equation 5 shows that the ratio μ/σ is a sufficient statistic for the recruiter. We will call this ratio *standardized productivity* throughout this paper, and denote it $\tilde{\mu}$. Intuitively, it indexes how likely the worker's productivity is to be satisfactory (above 0). The recruiter searches through workers, evaluating them based on Equation 5. Ignoring the constant part of her utility, the decision to terminate search and hire the current worker is given by the following Bellman equation.

$$U = -c + \int \max\{\beta F(\tilde{\mu}), U\} d\tilde{G}(\tilde{\mu}) \tag{6}$$

U is the continuation value and $\tilde{G}$ is the CDF of $\tilde{\mu}$ derived from the joint CDF of (μ, σ). Equation 6 emphasizes that the recruiter's problem again reduces to standard sequential search, after we note that $\tilde{\mu}$ is a sufficient statistic and the function $\beta F(x)$ is strictly increasing for positive β. Similar to the first-best benchmark, we know from well-known properties of sequential search that the solution is a reservation rule in $\tilde{\mu}$. We formalize this in the following lemma.

Lemma 1 *In any incentive compatible contract, the recruiter's acceptance region takes the form:*

$$\mathcal{D}_R = \{\tilde{\mu} | \tilde{\mu} \geq \tilde{\mu}^*\}$$

where $\tilde{\mu}^*$ *solves:*[8]

$$c = \int_{\tilde{\mu} \geq \tilde{\mu}^*} \beta F(\tilde{\mu}) - \beta F(\tilde{\mu}^*) d\tilde{G}(\tilde{\mu}) \tag{IC}$$

The proof of Lemma 1 is in the Appendix. The economic intuition is as follows. The refund contract encourages the recruiter to care about standardized productivity, rather than expected productivity. There will be a fundamental misalignment between the firm and recruiter, which can be visualized by graphing the firm's isoprofit curves and the recruiter's indifference curves over the space of worker types. We focus on the case when $\beta > 0$, which we will later show is the most relevant case.

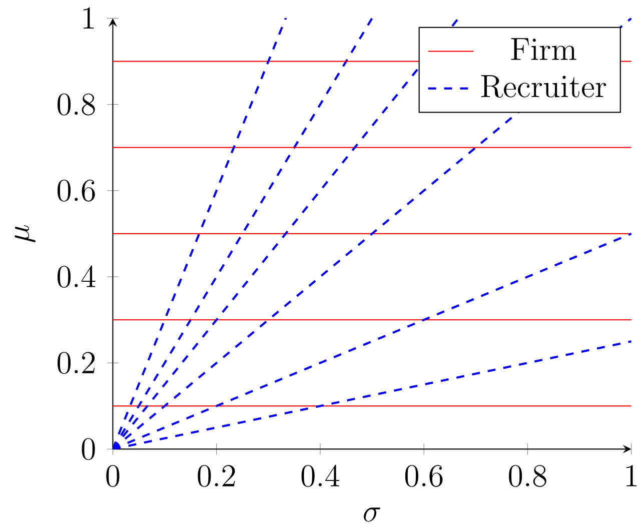


Figure 1: Indifference and Isoprofit Curves Over Worker Types

The recruiter's indifference curves all emanate from the origin, with higher indifference curves being more steeply sloped. In any incentive compatible contract, the recruiter will attempt to minimize productivity variance more than the firm would like, in order to climb to a higher indifference curve. An implication of Figure 1 and Lemma 1 is that the recruiter's acceptance region will be triangular[9], whereas the firm's will be rectangular. Figure 2 illustrates this.

8. This formulation is true for continuously distributed $\tilde{\mu}$ and a proper interior solution (non-degenerate search), but can be easily generalized to a system of inequalities otherwise.

9. or trapezoidal if the support of expected ability does not include 0.

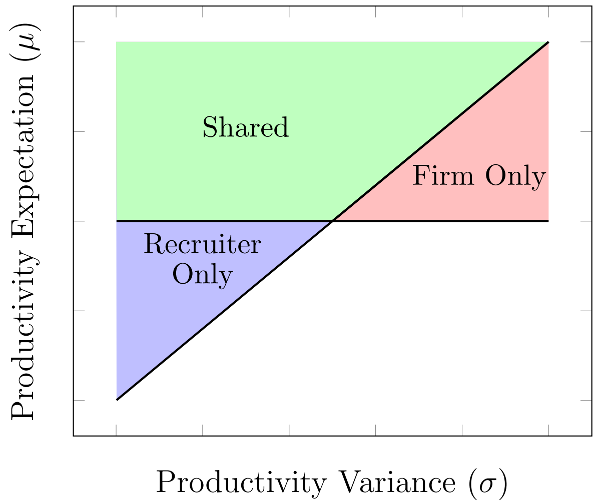


Figure 2: Recruiter vs. Firm Acceptance Regions Over Applicant Types

The figure displays three important partitions of the worker type space. In the recruiter only region are the low-variance, low-expectation workers the recruiter hires but the firm would prefer excluded. We refer to these as "safe bets," and they are inefficiently hired in equilibrium. In the firm only region are high-variance and high-expectation workers that the firm would like to hire, but the recruiter excludes. We call these workers "diamonds in the rough," and they are inefficiently excluded in equilibrium. In the Appendix, we show that under general conditions the acceptance regions will not be subsets of each other. A direct consequence of this is that there will always be a positive measure of diamonds in the rough and safe bets.

Figure 2 also provides intuition about how the equilibrium contract is determined. The firm effectively chooses the slope of the diagonal line that defines the three regions. Examining Equation IC reveals β controls the slope.[10] More powerful incentives (higher β) increase the slope. A steeper slope increases the share of workers that are inefficiently excluded, but decreases the share that are inefficiently included.

We now begin to characterize equilibrium. We have just shown that $\tilde{\mu}$ is a sufficient statistic for the recruiter. In order to focus on non-trivial cases, we introduce a weak assumption about the relationship between standardized productivity and expected productivity.

Assumption 1 $\mathbb{E}[\mu|\tilde{\mu} = x]$ *is weakly increasing in* x.[11]

For the rest of the paper, it is assumed to be satisfied. Intuitively, Assumption 1 means

10. Applying the implicit function theorem reveals that $\tilde{\mu}^*$ is increasing in β.

11. This condition is often referred to in the statistics literature as positive quadrant dependence in expectation, which is slightly weaker than positive quadrant dependence, and much weaker than positive affiliation.

that larger standardized productivity implies larger expected productivity of the candidate. This assumption is quite natural, given that $\tilde{\mu} = \mu/\sigma$.[12] If the conditional expectation is flat, then the problem becomes uninteresting. There is no way to encourage the recruiter to search strategically, and thus the firm will either not hire or offer a degenerate contract, where $\beta = 0$ and the recruiter returns the first applicant they searched.

We are now ready to characterize the delegation equilibrium. An equilibrium contract consists of an upfront payment α, contingent refund β and acceptance region $\mathcal{D}_R$, such that:

1. The firm maximizes profit.

2. The recruiter accepts the contract (individual rationality).

$$\mathbb{E}[t(a)|D_R] - \frac{c}{Pr((\mu, \sigma) \in D_R)} \geq 0 \tag{IR}$$

3. The acceptance region $\mathcal{D}_R$ is the optimal sequential search strategy of the recruiter, given the contract details (incentive compatibility).

This defines a quite general problem, with two dimensional sequential search and moral hazard. However, our prior discussion of the recruiter's problem, particularly Lemma 1, allows us to characterize the solution in a simple way. Lemma 1 shows that the acceptance region in the delegated search problem is defined by a reservation rule $\tilde{\mu}^*$. Moreover, any threshold $\tilde{\mu}^*$ can be induced by an incentive compatible and individually rational contract. As all parties are risk neutral and the firm pays the initial transfer α only to keep the recruiter indifferent between accepting and rejecting the contract (IR binds), then the firm extracts all social surplus and only cares about the acceptance region induced by the contract.

Theorem 1 *The delegated search equilibrium is given by the solution to a standard sequential search problem over* $\mathbb{E}[\mu|\tilde{\mu}]$. *The solution is determined by a reservation rule* $\tilde{\mu}^*$, *which solves:*

$$(\mathbb{E}[\mu|\tilde{\mu} \geq \tilde{\mu}^*] - \mathbb{E}[\mu|\tilde{\mu} = \tilde{\mu}^*]) \cdot \Pr(\tilde{\mu} \geq \tilde{\mu}^*) = c \tag{7}$$

and by (IC) and (IR) refund contract which induces the recruiter to select workers with $\tilde{\mu} \geq \tilde{\mu}^*$

Corollary 1.1 *The firm's profit under delegation is positive and equal to* $\mathbb{E}[\mu|\tilde{\mu} = \tilde{\mu}^*]$. *As a result, the optimal threshold* $\tilde{\mu}^*$ *is also positive.*

12. To break this assumption, the association between μ, σ needs to be so strong that the expectation grows faster than linearly. For example, $\sigma = \gamma\mu^2$ will cause the conditional expectation to be decreasing.

Theorem 1 has practical significance, proving that the general contracting problem is characterized by the solution to a much simpler problem. Indeed, the entire search strategy under moral hazard is defined by a threshold rule, which is uniquely pinned down by a single first-order condition. The entire problem essentially collapses into standard one-dimensional sequential search over a transformed distribution.

Theorem 1 also holds deeper economic insight. Comparing equations 7 and 3, we see that the equation characterizing equilibrium is identical to the one characterizing the first-best if we replace μ with $\tilde{\mu}$. Imposing the binary refund contract is equivalent to allowing the firm to search itself over $E[\mu|\tilde{\mu}]$ rather than μ. Thus, moral hazard in this setting has the effect of making the search technology more blunt or less accurate. The contract restriction requires the firm to resort to using the imperfect signal $\tilde{\mu}$ as a proxy for μ. In this way, search is noisy under delegation. This realization forms the foundation for the rest of our results.

5 Results

In the prior section, we provide a general characterization of the delegation equilibrium. In this section, we present first-best comparisons and comparative statics under minimal restrictions on the joint distribution of μ, σ. As shown in Section 4, first-best search is over μ directly while delegated search is over $\mathbb{E}[\mu|\tilde{\mu}]$. The latter has a distribution which is a mean preserving contraction of the former one as it is based on observing noisy "signal" $\tilde{\mu} = \mu/\sigma$ instead of observing μ explicitly.

In some special cases, the two distributions could be identical (for instance, if σ has a degenerate distribution). Thus, we begin by showing exactly when first-best profit and search strategy are not achieved under delegation. Throughout, we refer to strictly lower profit and a different search strategy under delegation as "not achieving the first-best."

Definition 1 *Let two random variables with CDFs F and G have compact supports S_F and S_G. F is a* ***strict mean preserving spread*** *of G if their expectations are the same and $\forall x \in int\ S_F$*

$$\int_x^{\bar{x}} (1 - F(s))ds > \int_x^{\bar{x}} (1 - G(s))ds$$

This definition is a slightly stronger notion of a mean-preserving spread. For the distributions of μ and $\mathbb{E}[\mu|\tilde{\mu}]$, the difference is as follows. If for some x, y it is true that thresholds on μ and $\tilde{\mu}$ select the same set of workers $\{(\mu,\sigma)|\mu \geq x\} = \{(\mu,\sigma)|\tilde{\mu} \geq y\} \notin \{\emptyset, \{(\mu,\sigma)\}\}$, then the inequality in the definition is violated (it becomes an equality) and for a specific search cost c, the first-best and the second-best acceptance regions are the same. This, however, only happens for very specific distributions of (μ, σ). This result is formalized in the next proposition.

Proposition 1 *First-best is not achieved in equilibrium for any search cost if and only if μ is a strict mean-preserving spread of $E[\mu|\tilde{\mu}]$.*

The formal proof consists of analyzing the firm's profits in two cases and the corresponding first-order conditions. The first order conditions equalize the excess wealth order – integrals from the definition – to the search cost c. As they are never equal to each other, the optimal thresholds can never be the same in the two cases. Moreover, the threshold must be strictly lower for the distribution of $\mathbb{E}[\mu|\tilde{\mu}]$, and thus, the profit must also be lower than it is in the first-best.

Expected productivity (μ) is always a mean-preserving spread of $E[\mu|\tilde{\mu}]$ using the standard definition. It will also be a strict mean-preserving spread, except for in a few knife-edge cases. Therefore, the proposition implies that the first-best is generally not achieved. We call the special cases when first-best is achieved knife-edge because they require a specific type of degeneracy in the joint distribution of (μ, σ). Our intuition suggests the first-best should not be achievable, because incentive compatibility requires the acceptance regions to be fundamentally different shapes. This is visualized in Figure 2. The proposition simply spells out when our intuition is violated.

As we showed earlier, the contract restriction makes the delegation problem equivalent to direct search with a less accurate search technology. Productivity variance is what garbles the search technology: when productivity variance is degenerate, standardized productivity $(\tilde{\mu})$ becomes a perfect signal of expected productivity, and the first-best is achieved.

Thinking about $\tilde{\mu}$ as a signal for μ allows us to view Proposition 1 through the lens of Blackwell (1953). First-best is not achievable whenever μ is statistically sufficient for $\tilde{\mu}$ in a strict sense, or put another way when μ dominates $\tilde{\mu}$ in the Blackwell order in a strict sense. Incentive compatibility and the refund contract convolute μ and σ in a specific way, such that the joint distribution of μ, σ will impact the equilibrium outcome. We will explore the mechanics of this relationship in the next few results.

With this established, we next wish to understand how the characteristics of accepted workers compare under delegation and the first-best. Particularly, we would like to see how the distribution of the workers' productivity expectation and variance differ in the delegated search and the first-best search benchmarks. We know that the firm searching directly cares more about the candidate's productivity expectation, and does not care about productivity variance at all. However, the recruiter prefers lower productivity variance. Intuitively, we should anticipate higher productivity expectation as well as higher productivity variance in the first-best than under delegation.

We begin with the productivity variance of accepted workers. When variance and expectation are independent, the distribution of variance among workers accepted in the first-best should be equal to the population distribution. The firm wishes to ignore productivity variance. However, the recruiter cares about productivity variance, and even when the two attributes are independent, the acceptable workers under delegation will tend to be lower variance than the general population. This is formalized in the following proposition.

Proposition 2 *If μ and σ are independent, then the productivity variance of accepted workers in the first-best is higher (in a first-order stochastic dominance sense) than under delegation.*

One way to understand this result is that refund contracts induce a special type of moral hazard, which biases the recruiter in favor of low-variance candidates, even when both the firm and the recruiter are risk neutral. A crucial question here is whether this bias in favor of low-variance candidates comes at the expense of high-expectation candidates. We turn to this next.

It follows directly from Proposition 1 that, because social surplus is higher in the first-best, expected productivity is higher on average than under delegation. However, under a broad set of circumstances, we can say that the entire distribution of productivity expectations of selected workers is higher in the first-best than under delegation, in a first-order stochastic dominance sense. A sufficient condition for this relies on search effort. We will define **search effort** as the percentage of candidates that are unacceptable, and denote it q. This object maps one-to-one with the expected number of searches, which is $1/(1-q)$. Thus search effort is itself an equilibrium outcome worth understanding. We anticipate less search effort in the delegated problem: if the search technology is less accurate, then the expected benefit of another search should be lower under delegation. Later in the section, we characterize necessary and sufficient conditions when this anticipated result holds. We first illustrate the connection between search effort and the distribution of productivity expectations in Lemma 2.

Lemma 2 *If search effort is lower under delegation than the first-best, then the productivity expectation of accepted workers in the first-best is higher (in a first-order stochastic dominance sense) than under delegation.*

To analyze search effort, we first observe that the strict mean-preserving spread condition from Proposition 1 does not generate a clear comparison of search effort under first-best and delegation. We need a stronger concept. As Chateauneuf, Cohen, and Meilijson (2004) and Zhou (2020) prove, one such concept is excess wealth order. We introduce this concept next, and prove that it is necessary and sufficient for generating clear comparative statics for all search costs.

Definition 2 *Let random variable X have a smooth CDF F. The excess wealth at threshold x^* is the expected benefit of one additional search if we already have x^*. It can be expressed as:*

$$EW_X(x^*) = E[(X - x^*)^+] = \int_{x^*}^{\infty} (1 - F(x))dx$$

To derive the necessary and sufficient condition for comparative statics in search effort, recall that the first-order condition of a sequential search problem over a random variable X with CDF F is:

$$(\mathbb{E}[X|X \geq x^*] - x^*)Pr(X^* \geq x) = EW_X(x^*) = c$$

Rewrite the equation in terms of quantiles of X:

$$EW_X(F^{-1}(q)) = c$$

The left-hand side is decreasing in q, implying that for a variable with greater excess wealth, the q that solves the equation will be also higher. To formalize this result, we define excess wealth order and state Theorem 2.

Definition 3 *A variable X_1 with CDF F_1 dominates a variable X_2 with CDF F_2 in the excess wealth order if:*

$$EW_{X_1}(F_1^{-1}(q)) \geq EW_{X_2}(F_2^{-1}(q)) \ \forall \ q \in (0, 1)$$

Theorem 2 *For any cost c search effort is greater for X_1 than for X_2 if and only if X_1 excess wealth order dominates X_2.*

Excess wealth order is a well-known variability order, and is discussed at length in Shaked and Shanthikumar (2007). It is sometimes referred to as the right-spread order. Excess wealth order implies that X_1 is a mean-preserving spread of X_2 if they have the same expectations, but the converse is not true. In this sense, it is stronger than the concept of a mean-preserving spread.[13] With this in hand, we present the following result.

13. However, it is weaker than the more widely-used dispersive order which, in past work, has been used to derive comparative statics in search intensity or duration.

Corollary 2.1 *If μ excess wealth order dominates $\mathbb{E}[\mu|\tilde{\mu}]$, then the search effort is greater in the first-best than in the delegated search benchmark. Therefore, the productivity expectation of accepted workers in the first-best is higher (in a first-order stochastic dominance sense) than under delegation.*

In the parametric examples section, we illustrate that for all joint lognormal or Pareto distributed (μ, σ), μ dominates $\mathbb{E}[\mu|\tilde{\mu}]$ in the excess wealth order. This is striking in the lognormal case: for arbitrarily high positive correlation between (μ, σ) dominance is maintained. In this sense, excess wealth order is not overly restrictive, and we have indeed shown that it is the least restrictive ordering needed for comparing equilibrium and first-best search effort.

We now reconsider the example discussed in the introduction. Candidate A, with a traditional resume, can be thought of as a safe-bet: low-variance, low-expectation. Candidate B, with a non-traditional resume, can be thought of as a diamond in the rough: high-variance, high-expectation. If the recruiter shared the firm's preferences, then he would select solely based on expected productivity and choose B. However, the refund contract introduces misalignment. The recruiter does not care about expected productivity, but rather how likely each candidate is to be above the firing threshold. This induces the recruiter to care about productivity variance, which will therefore cause the recruiter to inefficiently hire A over B.

Corollary 2.1 allows us to formalize the intuition from the example. While we already know that first-best social surplus is not achieved in equilibrium, we can say more: part of the lost social surplus is because the recruiter focuses search effort on finding low-variance candidates. The corollary proves that search is less efficient precisely because the recruiter is wasting his search effort minimizing variance instead of maximizing expectations. This manifests in a pool of acceptable workers which contains too few high-expectation candidates and too many low-variance candidates. As we discuss in the application section, this proposition implies that intermediation results in more variance-based statistical discrimination. The refund contract restriction results in an inefficient bias against high-variance candidates, despite risk neutrality of all actors.

Finally, we investigate how search effort, social surplus, and equilibrium profit change as the underlying distribution of workers in the labor market changes.

Proposition 3 *Consider a parameter θ of the joint distribution of μ, σ such that:*

1. *The marginal distribution of μ does not depend on θ*
2. *$\mathbb{E}_{\theta_1}[\mu|\tilde{\mu}]$ dominates $\mathbb{E}_{\theta_2}[\mu|\tilde{\mu}]$ in the excess wealth order for all $\theta_1 > \theta_2$.*
3. *$\mathbb{E}[\mu|\tilde{\mu}]$ converges to μ in distribution as θ converges to some value $\bar{\theta}$*

Then social surplus, equilibrium profit, and search effort are increasing in θ, and converge to first-best values as $\theta \to \bar{\theta}$

Restated verbally, the distortions and inefficiencies caused by delegation depend on the level of heterogeneity in the labor market. The proposition also illuminates a natural connection between our model and canonical multitasking models (Holmstrom and Milgrom 1991). To see this connection, suppose we define two "tasks": search along the expectation (μ) dimension and search along the variance (σ) dimension. Like in many multitasking models, the firm cannot provide incentives for each task individually, and can only encourage search over an aggregate metric. In our case, this metric is standardized productivity ($\tilde{\mu}$). If we take logs we have:

$$log(\tilde{\mu}) = log(\mu) - log(\sigma)$$

How do we interpret this expression? Well, it implies that the firm can only "buy" an increase in μ if it is willing to also "buy" a reduction in σ. Thus, we are in a situation where total search effort is rewarded, but there is a wasteful task which cannot be properly distinguished from the productive task. In equilibrium, this manifests in workers that are first-order stochastically dominated in terms of productivity expectation, but also first-order stochastically dominated in terms of variance compared to the first-best. In the proposition, raising θ effectively removes heterogeneity in productivity variance, the wasteful dimension of search. Removing this heterogeneity makes the recruiter focus on maximizing productivity expectation. In the extreme case when almost all heterogeneity in σ is removed, and $\mathbb{E}[\mu|\tilde{\mu}]$ converges to μ, first-best is achieved.

6 Parametric Examples

In the last section, we illustrate the main economic forces qualitatively, without imposing parametric assumptions on G, the distribution of productivity variance and expectation. In this section, we show more explicit results using specific parametric joint distributions.

6.1 Lognormal

Assumption 2 ***(Lognormal Productivity)*** *μ, σ are distributed joint lognormal. That is:*

$$\begin{pmatrix} log(\mu) \\ log(\sigma) \end{pmatrix} \sim N\left(\begin{bmatrix} m_\mu \\ m_\mu \end{bmatrix}, \begin{bmatrix} s_\mu^2 & s_{\sigma,\mu} \\ s_{\sigma,\mu} & s_\sigma^2 \end{bmatrix} \right)$$

Lognormal distributions are a common way to model positive-valued economic objects. The ability to easily incorporate correlation between μ, σ makes the parameterization more flexible. This also makes it a more credible way to model the posteriors generated by a proper information structure as posterior means and variances will generally not be independent.

In the Appendix, we provide and prove a series of lemmas that allow us to link our general results to the lognormal family of distributions. Two aspects of lognormal random variables make the analysis tractable. First, the multiplication of two lognormal random variables is, again, lognormal. This, taken together with the normal projection formula, implies that $E[\mu|\tilde{\mu}]$ will remain lognormal. Thus, first-best and equilibrium search will be over distributions within the same family, making the resulting problems easier to compare. Second, we show that two lognormal random variables with the same shift parameter m can be ranked in the excess wealth order by their shape parameters, s. Because $E[\mu|\tilde{\mu}]$ is a mean-preserving contraction of μ and both are lognormal, this makes comparison in the excess wealth order straightforward.

The last step needed to apply our general results is to observe that the shape parameter s_σ satisfies the conditions for θ laid out in Proposition 3. In particular, because s_σ is exactly the variance of $log(\sigma)$, it is clear to see how degeneracy of σ is achieved as it approaches 0.

Proposition 4 *If productivity variance and expectation are distributed according to Assumption 2 and further, there is not perfect positive dependence, then:*

1. *The first-best profit and acceptance region are not achieved.*

2. *Search effort is strictly greater in the first-best than under delegated search.*

3. *The productivity expectation of hired workers is higher in the first-best than delegated search.*

4. *Profit and search effort increase as s_σ decreases, and converge to first-best levels as $s_\sigma \to s^2_{\sigma,\mu}/s^2_\mu$.*

If additionally $s_{\sigma,\mu} = 0$, the productivity variance of hired workers is higher in the first-best than delegated search.

A nice feature of the result is that it is true for virtually all lognormal joint distributions, with arbitrarily negative or positive correlations. This illustrates that the necessary and sufficient excess wealth order condition is a quite weak requirement for some distributional families, further implying that our results regarding the under-provision of search effort and failure to achieve first-best are robust.

The fourth part of the proposition has a useful interpretation in terms of multiplicative noise. Consider a zero mean lognormal random variable Z that is independent of productivity variance and expectation. Multiplying productivity variance by Z is equal in distribution to increasing s_σ while holding all other parameters fixed. Therefore, we can use the proposition to conclude that additional multiplicative noise, or additional multiplicative heterogeneity in the variance dimension, makes search less efficient and reduces search effort.

6.2 Pareto

In this section, we assume the productivity attributes follow Pareto distributions. This assumption is convenient because it yields closed-form solutions for the threshold rules. In turn, these closed-form solutions allow straightforward comparative statics.

Assumption 3 ***(Pareto Productivity)*** *μ, σ are distributed independently with marginal Pareto distributions. That is, their joint probability density function is given by:*

$$g(\mu, \sigma) = \frac{\theta_\mu x_\mu^{\theta_\mu}}{\mu^{\theta_\mu+1}} \frac{\theta_\sigma x_\sigma^{\theta_\sigma}}{\sigma^{\theta_\sigma+1}} \mathbb{I}\{\mu \geq x_\mu\} \mathbb{I}\{\sigma \geq x_\sigma\}$$

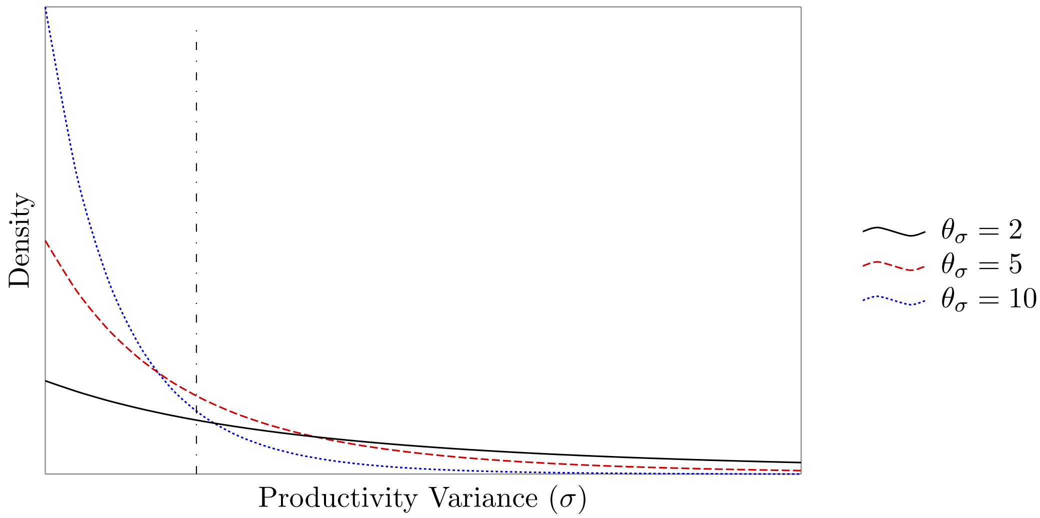
where both variables have finite expectations ($\theta_\mu > 1, \theta_\sigma > 1$).

Under a Pareto assumption, the random variable $\mathbb{E}[\mu|\tilde{\mu}]$ is almost Pareto: it has an atom at the beginning of the distribution, and is Pareto conditional on being above the atom. This is convenient for analysis because it implies that the equilibrium reservation rule will have a closed-form solution. We provide these closed form solutions in the Appendix. Additionally, we focus on the case where delegated search has an interior solution.[14]

The most important parameter for our analysis is the shape parameter θ_σ. In terms of interpretation, θ_σ represents the level of heterogeneity with respect to productivity variance in the worker pool. We show this visually in Figure 3, which plots three different Pareto densities with different values of θ_σ and $x_\sigma = 1$. As the notation implies, θ_σ satisfies all of the conditions laid out in Proposition 3. In particular, as θ rises, $E[\mu|\tilde{\mu}]$ rises in the excess wealth order.[15] As $\theta_\sigma \to \infty$ we approach a perfectly homogeneous population with respect to productivity variance.

14. In the Appendix we show an interior solution is guaranteed if $\frac{x_\mu \theta_\sigma}{(\theta_\mu+\theta_\sigma-1)(\theta_\mu-1)} \geq c$.
15. See the Appendix for a full proof.

Figure 3: Densities of σ for Different Values of θ_σ



Under this distributional assumption, we can apply our nonparameter results to show the following result.

Proposition 5 *If productivity variance and expectation are distributed according to Assumption 3, then:*

1. *The first-best profit and acceptance region are not achieved.*

2. *Search effort is greater in the first-best than under delegated search.*

3. *The productivity variance and productivity expectation of hired workers is higher in the first-best than delegated search.*

4. *Profit and search effort increase as θ_σ increases, and converge to first-best levels as $\theta_\sigma \to \infty$.*

The proof is provided in the Appendix, and is a direct consequence of the results presented in the general analysis section. The Pareto parameterization makes clear the connection between Proposition 3 and labor force heterogeneity. Increasing the θ_σ parameter reduces heterogeneity in the variance dimension. Since σ is a wasteful dimension of search in terms of social surplus, reducing heterogeneity in that dimension lowers the returns to searching along that dimension, and thus raises efficiency.

7 Applications

7.1 The Choice to Delegate

In the first-best benchmark, the firm operates the search technology directly. Therefore, the difference between first-best profit and profit under delegation is exactly the agency loss incurred by the firm. We can use this observation and Proposition 3 to understand the decision to delegate.

A firm must perform a cost-benefit analysis when deciding whether to delegate the recruiting function. The benefit of delegation is the comparative advantage of the recruiter. The cost is agency loss: as we have shown, there are fundamental differences between how the recruiter and the firm order potential workers. We can model this by supposing that, prior to designing the contract with the recruiter, the firm must decide whether it will delegate at all. If it chooses not to delegate, it can perform search directly at a different search cost c_F, which is strictly larger than the recruiter's search cost c.

The firm will decide to delegate when the comparative advantage of the recruiter outweighs the agency loss of delegation. We can apply Proposition 3 to understand how changes in the labor force impact the decision to delegate.

Proposition 6 *As heterogeneity in productivity variance decreases, the firm is more likely to delegate. When workers are homogeneous with respect to productivity variance, the firm will always delegate.*[16]

When the recruiter and the firm face the same search cost, we know from prior results that direct search is always more profitable than delegated search. However, when the recruiter has a lower search cost, there is the possibility that delegated search is more profitable. As heterogeneity in the wasteful dimension vanishes, profit under delegation rises, while profit from direct search stays constant. Eventually, when all agency loss is gone, only the comparative advantage effect remains, and delegated search is optimal.

If we think about different occupations as having different labor market pools, recruiter utilization in each occupation will depend on the amount of heterogeneity in productivity variance across workers. Occupations where workers have similar productivity variance will feature higher recruiter utilization, and occupations with more heterogeneous productivity variance will feature a larger share of firms recruiting directly.

One crucial distinction to make here is that all results in this paper are concerned with differences in variance, not the overall level of variance. Some occupations may have higher

16. In our model, a decrease in heterogeneity corresponds with a decrease in θ, in the same sense as Proposition 3.

average levels of productivity variance. This does not, however, imply that those occupations will have higher recruiter usage. What matter are the differences in variances among workers in the same occupation. As differences in variance become larger within an occupation, recruiter utilization becomes less likely.

7.2 Statistical Discrimination

In this section, we show delegation can amplify statistical discrimination. Along the way, we illustrate how to map an explicit information structure with a common prior to our primitives (μ, σ). We believe this section may be of interest to researchers who want to apply our framework to data.

Consider an economy in which all workers can be divided into two groups. The recruiter and firm have a common prior over productivity which is normal and the same for both groups: $a \sim N(\mu_0, \sigma_0^2)$. After searching a worker, the recruiter observes the group membership of the worker and a signal about productivity. The recruiter does not observe the group prior to the draw (there is no directed search). The signal takes a normal form: $Y = a + \xi_i^n$, $\xi_i^n \sim N(0, \tau_i^{-2})$, where i indexes the groups A and B. Notice that the only difference between the two groups is that they have different signal precision or information quality. Without loss, suppose that Group A has better signal precision: $\tau_A > \tau_B$. We can interpret this as the recruiter *better understanding*, the work histories of individuals from Group A.

First, we map this situation to our primitives. The recruiter Bayesian updates their normal prior with the normal signal, generating a normal posterior belief about productivity given the signal:

$$a|\xi_i = x \sim N\left(\frac{\tau_i^2}{1/\sigma_0^2}x + (1 - \frac{\tau_i^2}{1/\sigma_0^2})\mu_0, \frac{\sigma_0^2\tau_i^{-2}}{\sigma_0^2 + \tau_i^{-2}}\right)$$

From this expression, we can see that σ will take two values with equal probability. Because the precision of the signal for Group A is higher ($\tau_A > \tau_B$) we have that $\sigma_B > \sigma_A$. When the recruiter is faced with two candidates with the same expected productivity but from different groups, it will prefer the candidate from Group A. Additionally, $\mu_i = E[a|\xi_i]$ (that is expected productivity unconditional on the signal realization but conditional on group) will be normal.

We say there is statistical discrimination if the probability a hired worker is from Group A is higher than the probability that a hired worker is from Group B. Notice that even the first-best can feature statistical discrimination. Indeed, this is precisely what was pointed out in Heckman (1998): even if a decision maker uses the same expected productivity threshold for hiring all groups (meaning there is no taste-based discrimination) differences in productivity variance can cause arbitrary statistical discrimination in either direction. Whether the first-best features statistical discrimination against the high-variance group depends on whether the optimal threshold lies above or below the prior mean, $\mu^* > \mu_0$.

However, we wish to understand how delegation impacts statistical discrimination. Does delegation tend to reduce or increase statistical discrimination against group B? More formally, is the probability a hired worker is from Group B higher or lower than the first-best? If it is higher, Group B workers will have more labor market success under delegation. If it is lower Group B workers will more success under direct search.

Proposition 7 *The probability the hired worker is from Group B is lower under delegation than the first-best. Therefore, variance-based statistical discrimination is greater under delegation than the first-best.*

The proof is provided in the Appendix. Prior to this result we established that individuals with high productivity variance face a disadvantage under delegation. The proposition makes a stronger claim. If we think about the probability a hired worker is from Group B as the employment rate of Group B (this would be true under monopsony), then the result implies Group B will have a lower employment rate when search is delegated to a recruiter than when search is conducted directly. Put another way, Group B is unambiguously worse off under delegation and Group A is unambiguously better off. Recall that this occurs despite the fact that the two groups have identical underlying productivity distributions.

The proposition also highlights the potential power of homophily. One reason $\tau_A > \tau_B$ might be because the recruiter is a member of Group A. They better understand resumes or work histories from people who are similar to them. Without delegation, it is unclear which group fairs better. It will depend on the level of the hiring threshold μ^*. But when there is informational homophily delegation reduces the labor market success of Group B (the out group).

Suppose Group A workers are members of a racial majority, while Group B workers are members of a racial minority. Our model suggests that delegation can lead to discrimination against a minority group even in the absence of explicit bias. It is sufficient that people understand those of the same racial group better. This might explain racial hiring gaps persists in many occupations despite explicit verbal commitments by leaders to improve diversity. For example, black representation in computer and math occupations fell between 2002 and 2016 (Muro, Berube, and Whiton 2018). Overall, this section illustrates how delegation can form a microfoundation for variance-based statistical discrimination.

7.3 A Vicious Cycle

Under our model, we have established that recruiter utilization should be higher in occupations where the labor market is more homogeneous with respect to productivity variance. On the other hand, we have also established that delegation exacerbates statistical discrimination against groups or individuals with higher productivity variance. These two mechanisms dynamically interact in an interesting way.

Consider an occupation with an initially heterogeneous labor pool. Suppose firms decide whether to use a recruiter. In period 1, only the firms with the highest opportunity cost will choose to use a recruiter, meaning that the majority of hiring will take place directly. However, applicants with high productivity variance will still face a slight disadvantage, and be hired at slightly lower rates. Suppose workers become discouraged after lack of hiring success. They leave the occupation before the next period. In period 2, the workers who leave are replaced by workers with the same characteristics as the initial labor force. Even though new workers enter, the labor market will be more homogeneous with respect to productivity variance due to the disproportionate outflow of high-variance workers. This will increase the share of firms using recruiters. In turn, this increases the outflow of discouraged high-variance workers. As time progresses, a vicious cycle unfolds: recruiter utilization rises and the labor market becomes more homogeneous.

If we think productivity variance is connected to factors like socioeconomic status, this vicious cycle will tend to work against efforts to make the occupation more diverse. Crucially, this argument relies on discouraged workers leaving the occupation. It is not clear how powerful a force this cycle represents, and how it would be impacted by a fully specified search and matching model with congestion effects and other aspects. However, it is an observable implication of our model and one worth mentioning for future work.

8 Discussion

8.1 Beyond Recruiters

In our framework, the recruiter is induced to act risk averse. The specific mechanism we provide that generates this behavior is the refund contract, which is prolific among recruiters that are hired externally. However, there are many reasons why employees of the firm may also exhibit induced risk aversion. For example, an article from a leading human resource association mentions that many internal recruiters and human resource staff have bonuses based on the cost of hiring and the time to fill a position (Hirshman 2018). Such cost-based metrics ignore productivity, and can bias an agent in favor of candidates who do not have outside offers. The same article also suggests turnover as a possible metric for gauging recruiter and human resource performance. This is consistent with our interviews: the internal recruiter we interviewed, who also holds a dual human resources role, said that turnover can be an important part of human resource performance evaluation. Since a bonus based on turnover is similar in spirit to a refund contract, it could generate the same sort of misalignment that we discuss in this paper.

More generally, human resource employees are responsible for reducing legal risks arising from people. As Dufrane et al. (2021) put it:

> "There are numerous laws and regulations governing the employment relationship that HR professionals must understand and navigate in order to help ensure their organizations avoid costly fines and other penalties, including the potential harm to the organization's reputation."

Behaviors from employees that can give rise to "costly fines and penalties" include serious crimes committed by employees such as sexual harassment and fraud. Thus, human resource departments internalize the downside risks of hiring a malicious or negligent employee, but do not internalize the upside benefits of hiring a star.

These arguments highlight the possibility that our results extend to human resource employees. This greatly broadens the impact of our analysis, because while only some firms use recruiters, a much larger share have some form of human resource department. As of 2018, it is estimated that there are 671,140 human resource workers (Labor Statistics 2019). In 2018, the U.S. Census Bureau estimated that there were 664,757 businesses with 20 or more employees (Bureau 2021). Assuming that most businesses with less than 20 employees do not hire a human resource specialist, this therefore implies that there is more than one human resource worker for every business.

8.2 Heterogeneity in Productivity Variance

Many of our results examine how heterogeneity in the variance dimension of beliefs impacts search behavior. We have also shown that delegation results in differential treatment of two workers with the same productivity expectation but different productivity variance. These results beg the question: what generates differences in productivity variance across workers?

There are several sources worth discussing. First, job experience can generate variation in information quality, and thus generate differences in productivity variance for people of different ages. Many theoretical papers, starting with Jovanovic (1979), are based on the idea that work experience provides information about productivity. This idea is supported by empirical work. Fredriksson, Hensvik, and Skans (2018) show that match quality appears to be better among older workers. As a result, age and work experience can generate differences in productivity variance.

Continuing along this same vein, credit constraints can make it such that differences in parental income generate heterogeneity in productivity variance. High quality signals of productivity are expensive. For example, the cost of data science boot camps can be around $2,000-$17,000 for just a small period of instruction (Williams 2020). Prestigious universities are usually either extremely expensive (a year's tuition can be in excess of U.S. median annual earnings) or extremely selective. Even with financial aid, individuals from disadvantaged backgrounds often do not have the resources to invest in the preparatory work needed to be admitted.[17]

Unequal access to information can also contribute to inequality of opportunity. For example, currently, only 71% of eligible college applicants file the Free Application for Federal Student Aid (*How America Pays for College 2020*). As a result, job seekers will often need to pay for productivity signals using family support. Children of wealthy parents will tend to have lower productivity variance, meaning that if we compared two workers with the same expected ability but different parental wealth, we would expect the child of wealthier parents to be approached more by recruiters, even if recruiters have no intrinsic bias towards wealthy workers. This will tend to reinforce existing socioeconomic inequality.

Finally, a recruiter might be better able to interpret the resume or life experience of a worker from the same socioeconomic group. Factors like religion, nationality, language, and cultural background play large roles in processing signals of productivity. For example, Bencharit et al. (2019) find that 86 percent of European Americans want to convey excitement in a job interview, compared to 72 percent of Asian Americans and 48 percent of Hong Kong Chinese. In the same study, it was found European Americans rated their ideal job candidate as excited, while Hong Kong Chinese rated their ideal candidate as calm. This reflects a form of homophily which, in the language of our model, would make a recruiter of European descent have a higher productivity variance about a candidate who was of Asian descent than a candidate of European descent.

Taken together, these examples and our results suggest a form of two-way causality. On the one hand, our application section suggests that delegation is *more likely* in industries that are socioeconomically homogeneous. On the other hand, Proposition 2 and our other results suggest that delegation *helps homogenize* industries.

17. SAT preparation classes, tutoring, college admissions counseling, AP testing, etc.

9 Conclusion

We develop a theoretical model of delegated recruitment with uncertain productivity, and show that the general, nonparametric version of the model can be reduced to classic sequential search over modified objects. Our characterization reveals insights into how contract restrictions and delegation translate into moral hazard with a multitasking flavor: a risk neutral recruiter will over-select low-variance candidates at the expense of high-expectation candidates. Using this insight, we prove that the efficiency of delegation critically relies on the level of heterogeneity in the population with respect to productivity variance.

Our framework provides mechanisms for why firms outsource recruitment, why firms might statistically discriminate in the absence of bias, and why occupations can become homogeneous over time. Additionally, we argue that our results extend more broadly to many agents who are involved in hiring processes. We also discuss how socioeconomic inequality can be a source of productivity variance heterogeneity.

We believe that our approach is generally useful for analyzing economic situations in which search over objects of uncertain quality is delegated to a third party (recruiters, mortgage brokers, venture capitalists). The refund contracts is common in many settings; however we want to emphasize that our general approach is valid generally. Any type of contract shapes a recruiter's indifference curves over the distribution of posteriors. Then, solving the delegated search problem is equivalent to solving a standard sequential search problem for the distribution of expected productivity conditional on those indifference curves. Moreover, comparing the results to the first-best benchmark completely relies on the comparison of this conditional expectation distribution and unconditional expected productivity distribution. In other words, our procedure can be used for more complicated contracts and families of posterior distributions.

One limitation of our analysis also presents an opportunity for further work. Although we have a theory of delegated recruitment, it is unclear whether our theory and the testable implications we derive are consistent with recruiter and firm behavior. This is mainly because data on recruiter behavior is scant, and when it does exist, it is scattered across many platforms. Further empirical work will require a combination of innovative data sources, such as text analysis of recruiter LinkedIn messages and careful field experiments.[18] A particular challenge when evaluating our theory (and indeed, any theory of recruiter behavior) is mapping observed characteristics of workers to beliefs. Despite these challenges, we believe that this research frontier is worth pushing forward. Delegation is now a common practice across

18. Field experiments that elicit beliefs will require new techniques, like the two-sided audit. For an excellent example see Cowgill and Perkowski (2020).

many industries, and our paper offers theoretical evidence that it can generate distortions in the hiring process. Verifying the existence and the magnitude of these distortions will require taking ideas to data.

References

Armstrong, Mark, and John Vickers. 2010. "A model of delegated project choice." *Econometrica* 78 (1): 213–244.

Belzunce, Felix, Carolina Martinez-Riquelme, Jose M Ruiz, and Miguel A Sordo. 2016. "On sufficient conditions for the comparison in the excess wealth order and spacings." *Journal of Applied Probability* 53 (1): 33–46.

Bencharit, Lucy Zhang, Yuen Wan Ho, Helene H Fung, Dannii Y Yeung, Nicole M Stephens, Rainer Romero-Canyas, and Jeanne L Tsai. 2019. "Should job applicants be excited or calm? The role of culture and ideal affect in employment settings." *Emotion* 19 (3): 377.

Black, Ines, Sharique Hasan, and Rembrand Koning. 2020. "Hunting for talent: Firm-driven labor market search in America." *Available at SSRN.*

Blackwell, David. 1953. "Equivalent comparisons of experiments." *The annals of mathematical statistics,* 265–272.

Bureau, US Census. 2021. *2018 SUSB Annual Data Tables by Establishment Industry,* May. https://www.census.gov/data/tables/2018/econ/susb/2018-susb-annual.html.

Carlsson, Magnus, Luca Fumarco, and Dan-Olof Rooth. 2014. "Does the design of correspondence studies influence the measurement of discrimination?" *IZA Journal of Migration* 3 (1): 1–17.

Chateauneuf, Alain, Michele Cohen, and Isaac Meilijson. 2004. "Four notions of mean-preserving increase in risk, risk attitudes and applications to the rank-dependent expected utility model." *Journal of Mathematical Economics* 40 (5): 547–571.

Che, Yeon-Koo, Wouter Dessein, and Navin Kartik. 2013. "Pandering to persuade." *American Economic Review* 103 (1): 47–79.

Cowgill, Bo, and Patryk Perkowski. 2020. "Delegation in Hiring: Evidence from a Two-Sided Audit." *Columbia Business School Research Paper,* no. 898.

Dufrane, Amy, Michael Bonarti, Jim DeLoach, and Rebecca DeCook. 2021. *The HR Function's Compliance Role,* June. https://www.corporatecomplianceinsights.com/hr-function-compliance-role/.

Frankel, Alex. 2021. "Selecting Applicants." *Econometrica* 89 (2): 615–645.

Frankel, Alexander. 2014. "Aligned delegation." *American Economic Review* 104 (1): 66–83.

Fredriksson, Peter, Lena Hensvik, and Oskar Nordström Skans. 2018. "Mismatch of talent: Evidence on match quality, entry wages, and job mobility." *American Economic Review* 108 (11): 3303–38.

Gentzkow, Matthew, and Emir Kamenica. 2016. "A Rothschild-Stiglitz approach to Bayesian persuasion." *American Economic Review* 106 (5): 597–601.

Heckman, James J. 1998. "Detecting discrimination." *Journal of economic perspectives* 12 (2): 101–116.

Hirshman, Carolyn. 2018. *Incentives for Recruiters?,* April. https://www.shrm.org/hr-today/news/hr-magazine/pages/1103hirschman.aspx.

Holmstrom, Bengt, and Paul Milgrom. 1991. "Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design." *JL Econ. & Org.* 7:24.

Jovanovic, Boyan. 1979. "Job matching and the theory of turnover." *Journal of political economy* 87 (5, Part 1): 972–990.

Labor Statistics, U.S. Bureau of. 2019. *May 2018 national Occupational employment and wage estimates,* April. https://www.bls.gov/oes/2018/may/oes_nat.htm.

McCall, John Joseph. 1970. "Economics of information and job search." *The Quarterly Journal of Economics,* 113–126.

Muro, Mark, Alan Berube, and Jacob Whiton. 2018. *Black and Hispanic underrepresentation in tech: It's time to change the equation,* April. https://www.brookings.edu/research/black-and-hispanic-underrepresentation-in-tech-its-time-to-change-the-equation/.

Postel–Vinay, Fabien, and Jean–Marc Robin. 2002. "Equilibrium wage dispersion with worker and employer heterogeneity." *Econometrica* 70 (6): 2295–2350.

How America Pays for College 2020. https://www.salliemae.com/about/leading-research/how-america-pays-for-college/.

Shaked, Moshe, and J George Shanthikumar. 2007. *Stochastic orders.* Springer Science & Business Media.

Shimer, Robert, and Lones Smith. 2000. "Assortative matching and search." *Econometrica* 68 (2): 343–369.

Ulbricht, Robert. 2016. "Optimal delegated search with adverse selection and moral hazard." *Theoretical Economics* 11 (1): 253–278.

Williams, Alex. 2020. *Our Ultimate Guide to the Best Data Science Bootcamps,* April. https://www.coursereport.com/blog/best-data-science-bootcamps-the-complete-guide.

Zhou, Jidong. 2020. "Improved Information in Search Markets."

10 Appendix

10.1 Match-Specific Productivity

We have interpreted productivity expectation and variance (μ, σ) as characterizing the beliefs of the recruiter about a worker prior to hire. However, our results extend naturally to models where productivity of worker i at firm j can be decomposed into a worker effect, a firm effect, and a match-specific effect, like so:

$$a_{i,j} = \gamma_i + \kappa_j + \epsilon_{i,j}$$

For simplicity, assume that all three components are independent. It is reasonable to assume that the firm effect, κ_j, is known. Indeed, since a recruiter is searching for the same firm, it is constant across all workers, and thus only shifts the mean of the productivity distribution. We can think of the match-specific effect as being mean 0 with known variance σ_ϵ^2 and unpredictable prior to hire. The expertise of the recruiter lies in their ability to predict γ_i. After searching a worker, reading their resume, and conducting a preliminary interview, the recruiter forms an estimate of expected γ_i, denoted $\bar{\gamma}_i$, with a corresponding variance estimate σ_i.

Assuming normality of all components, the recruiter views $a_{i,j}$ after an interview as normal, with expectation $\kappa_j + \bar{\gamma}_i$ and variance $\sigma_\epsilon^2 + \sigma_i^2$. These two components correspond to our productivity expectation and variance, (μ, σ). If the recruiter can perfectly predict γ_i, $\sigma_i^2 = 0$ for all i and the productivity variance distribution is degenerate. First-best is achieved, and there are no distortions from delegation.

Consider the simple case when the recruiter can only distinguish between two groups of workers. The groups have the same expected worker effect $\bar{\gamma}$ but different worker effect variances σ_i. In this situation, our model predicts that the recruiter will always prefer the lower variance group, because this group has a higher standardized productivity. In any incentive compatible contract, the lower-variance group is always hired. However, the firm is indifferent between the groups, and will hire the first worker searched.

10.2 Proof of First-Best Search

The proof of the optimal sequential search strategy (without delegation) is well-known, but we include it for completeness. Denote V as the value function of the firm. Denote the marginal distribution of μ as F. The dynamic programming problem of the firm is given by:

$$V = -c + \int \max\{E[a|\mu = u], V\} dF(\mu)$$

Note that if there were recall (so that the highest previously viewed μ could be carried as a state variable), then the firm would never exercise the option.

$$V = -c + \int \max\{\mu, V\} dF(\mu)$$

Re-writing, this yields:

$$0 = -c + \int \max\{\mu - V, 0\} dF(\mu)$$

So, the optimal strategy is a reservation rule characterized by μ^*, where $V = \mu^*$. Thus:

$$c = \int \max\{\mu - V, 0\} dF(\mu) \leftrightarrow c = \int_{\mu > \mu^*} \mu - \mu^* dF(\mu)$$

Integration by parts gives:

$$c = -[(1 - F(\mu))(\mu - \mu^*)]_{\mu^*}^{\bar{\mu}} + \int_{\mu^*}^{\bar{\mu}} (1 - F(\mu)) d\mu$$

Since the first term is 0, this simplifies to:

$$c = \int_{\mu^*}^{\bar{\mu}} (1 - F(\mu)) d\mu$$

As an aside, note that we can re-arrange the intermediate equation this way:

$$c = \int_{\mu > \mu^*} \mu dF(\mu) - (1 - F(\mu^*))\mu^* \leftrightarrow \mu^* = \frac{1}{1 - F(\mu^*)}\left(\int_{\mu > \mu^*} \mu dF(\mu) - c\right)$$

which can be compactly re-written as:

$$\mu^* = E[\mu | \mu \geq \mu^*] - \frac{c}{Pr(\mu \geq \mu^*)}$$

10.3 Proof of Lemma 1

The dynamic programming problem of the recruiter is given by:

$$U = -c + \int \max\{-\beta E_a[\mathbb{I}\{a \leq 0\}|(u, s)], U\} dG(\mu, \sigma)$$

$$U = -c + \int \max\{-\beta F(-\mu/\sigma), U\} dG(\mu, \sigma)$$

Re-writing, this yields:

$$0 = -c + \int \max\{\beta F(\mu/\sigma) - U, 0\} dG(\mu, \sigma)$$

Observe that utility only depends on μ/σ, so we can reduce the problem to one-dimensional search. As long as β is negative, utility will be increasing in μ/σ. The firm will always set $\beta \leq 0$ when $E[a|\tilde{\mu} = x]$ is increasing in x (which is what we assumed). Thus, we will have a reservation rule strategy in the ratio μ/σ. Denote this reservation rule x^*. Returning to the recruiter's problem, we can re-write using standardized productivity ($\tilde{\mu}$):

$$c = \int \max\{\beta F(\tilde{\mu}), U\} d\tilde{G}(\tilde{\mu}) \leftrightarrow c = \int_{\tilde{\mu} \geq x^*} \beta F(\tilde{\mu}) - \beta F(x^*) d\tilde{G}(\tilde{\mu})$$

10.4 Proof of Theorem 1

The next lemma is implicitly used in the Proof of Theorem 1 while showing that the search over $\tilde{\mu}$ is equivalent to search over $\mathbb{E}[\mu|\tilde{\mu}]$, regardless of whether or not the first one contains more information about the worker.

Lemma 3 ***No-atom optimal search.*** *Let one search over a pool of uniformly distributed x with a payoff $f(x)$, ($f'(x) \geq 0$)) and a cost $c > 0$ per search. Let $x^* \in (0, 1)$ be a unique optimal search threshold. Then, $\forall \varepsilon > 0 : f(x^* - \varepsilon) < f(x^* + \varepsilon)$.*

Proof. The intuition of the statement is that one being able to set a threshold on the CDF of the search variable (rather than the variable itself) would never strictly prefer to set it within an atom than anywhere else. The problem described in the lemma can be stated as

$$\max_{x'}\{\mathbb{E}[f(x)|x \geq x'] - \frac{c}{1-x'}\}$$

The derivative with respect to x' is

$$(\mathbb{E}[f(x)|x \geq x'] - f(x')) * (1-x') - c = (*)$$

Let us suppose that x^* is the unique maximizer, and that $\exists\ \varepsilon > 0:$ s.t. $f(x)$ is flat on $(x^*-\varepsilon; x^*+\varepsilon)$. Let $\bar{x} = x^* + \varepsilon$. Locally for $x' \in (x^*-\varepsilon; x^*+\varepsilon)$

$$\mathbb{E}[f(x)|x \geq x'] = \frac{(1-\bar{x}) * \mathbb{E}[f(x)|x \geq \bar{x}] + (\bar{x}-x') * f(x^*)}{1-x'}$$

Then, simplifying the derivative of the outcome with respect to x' gives

$$(*) = (1-\bar{x}) * \mathbb{E}[f(x)|x \geq \bar{x}] + (\bar{x}-x') * f(x^*) - f(x^*) * (1-x') - c$$

$$= (1-\bar{x}) * (\mathbb{E}[f(x)|x \geq \bar{x}] - f(x^*)) - c$$

which apparently does not depend on x' and is constant for $x' \in (x^*-\varepsilon; x^*+\varepsilon)$. Then, x^* cannot be a unique maximizer, since – depending on the sign of the derivative – one should either increase or decrease the threshold, or is indifferent in some small neighborhood around x^*. ■

We apply Theorem 1 to the firm's problem, which is given by Equations OBJ, IR, IC, and VAL:

$$\max_{\alpha,\beta,\mathcal{D}_R} E[a - \beta\mathbb{I}\{a > 0\}|(\mu,\sigma) \in D_R] - \alpha$$

s.t.

$$\alpha + u^* \geq 0 \tag{IR}$$

$$c = \int_{u \geq u^*} (1 - M(u))du \tag{IC}$$

$$\mathcal{D}_R = \{\mu, \sigma | \mu/\sigma \geq F^{-1}\left(\frac{u^*}{\beta}\right)\} \tag{REGION}$$

First, we prove that the IR constraint must bind. Suppose it does not. Then, the firm could lower α by ϵ and increase maximized profit without violating any other constraints.

This contradicts optimality. Thus, IR binds at the optimum. From the end of the proof of Lemma 1, we have that:

$$u^* = E[u|u \geq u^*] - \frac{c}{Pr(u \geq u^*)}$$

Plugging this into binding IR and solving for α:

$$\alpha = -E[u|u \geq u^*] + \frac{c}{Pr(u \geq u^*)}$$

Substituting the result into the objective obtains:

$$\max_{\beta, \mathcal{D}_R} E[a|(\mu, \sigma) \in D_R] - \frac{c}{Pr((\mu, \sigma) \in D_R}$$

which is the desired form of the objective. Using Lemma 2, the modified problem becomes:

$$\max_{\beta, u^*} E[a|\mu/\sigma \geq F^{-1}(u^*/\beta)] - \frac{c}{Pr(\mu/\sigma \geq F^{-1}(u^*/\beta))}$$

$$c = \int_{u \geq u^*} (1 - M(u))du \tag{IC}$$

This makes apparent the fact that the objective is no longer constrained by the constraints (since we have an extra degree of freedom), and in fact only depends on $x := F^{-1}(u^*/\beta)$.

The firm's choice of the contract creates the incentives over $\tilde{\mu}$ in the recruiter's optimal stopping problem. This, along with the binding IR constraint in the delegated problem, means that the firm implicitly searches over $\tilde{\mu}$. Given the monotonicity Assumption 1 and Lemma 3, this is equivalent to searching over $\mathbb{E}[\mu|\tilde{\mu}]$, which is the firm's outcome. This optimal search is characterized by the first-order condition stated in the theorem.

Therefore, we can maximize the objective without constraints to derive x, then use the definition of x and the IC constraint to derive β, u^*. Finally, α can be retrieved from the binding IR constraint, and thus, the problem reduces in the way stated in the proposition.■

10.5 Proof of Proposition 2

Proof. Note that under independence, $\sigma|\mathcal{D}_F$ is the same as the unconditional distribution of σ. Then:

$$\begin{aligned}
Pr(\sigma \leq y|(\mu,\sigma) \in \mathcal{D}_R) &= Pr(\mu \leq y\tilde{\mu}^*|(\mu,\sigma) \in \mathcal{D}_R) * Pr(\sigma \leq y|\mu \leq y\tilde{\mu}^* \ \&\ (\mu,\sigma) \in \mathcal{D}_R) \\
&+ Pr(\mu > y\tilde{\mu}^*|(\mu,\sigma) \in \mathcal{D}_R) * Pr(\sigma \leq y|\mu > y\tilde{\mu}^* \ \&\ (\mu,\sigma) \in \mathcal{D}_R) \\
&= Pr(\mu \leq y\tilde{\mu}^*|(\mu,\sigma) \in \mathcal{D}_R) * 1 + (1 - Pr(\mu \leq y\tilde{\mu}^*|(\mu,\sigma) \in \mathcal{D}_R)) * G_\sigma(y) \\
&> G_\sigma(y) = Pr(\sigma \leq y|(\mu,\sigma) \in \mathcal{D}_F)
\end{aligned}$$

Notice that the first quantity is the conditional CDF in the recruiter acceptance region. The second-to-last line shows that the this CDF is essentially a weighted average of 1 and $G_\sigma(y)$, which is always weakly greater than $G_\sigma(y)$. This proves first-order stochastic dominance of σ by $\sigma|\mathcal{D}_F$. ∎

10.6 Proof of Lemma 2

We will use the notations from Figure 2.

$$p = \frac{Pr(A)}{Pr(A) + \Pr(B)}$$

$$q = \frac{Pr(C)}{Pr(C) + \Pr(B)}$$

We can conclude that $p > q$ since search effort is lower in the second-best ($Pr(\text{FB}) < Pr(\text{SB})$).

$$\mu|\text{SB} \sim (\mu|A)p(\mu|B)$$

$$\mu|\text{FB} \sim (\mu|C)q(\mu|B)$$

where notations on the RHS are used for mixture distribution. In other words, one could write each of them as a three-component mixture:

$$\mu|\text{SB} \sim (\mu|A)(\text{w/p } q) + (\mu|A)(\text{w/p } p - q) + (\mu|B)(\text{w/p } 1 - p)$$

$$\mu|\text{FB} \sim (\mu|C)(\text{w/p } q) + (\mu|B)(\text{w/p } p - q) + (\mu|B)(\text{w/p } 1 - p)$$

Given the support of $\mu|A$, $\mu|B$, $\mu|C$, it is trivial to conclude that $\mu|B \succ_{\text{FOSD}} \mu|A$ and $\mu|C \succ_{\text{FOSD}} \mu|A$. Thus, each of the components in the first-best μ mixture first order stochastically dominates the components in the second-best μ mixture. Given that the mixture probabilities are identical, this implies that the whole FB mixture dominates the SB mixture

$$\mu|\text{FB} \succ_{\text{FOSD}} \mu|\text{SB}$$

(this simply follows from the formula of a mixture CDF).

10.7 Proof of Proposition 3

The firm extracts all surplus from the recruiter, so profit and social surplus are the same. Consider any two values of the parameter $\theta' > \theta$. Then $E_{\theta'}[\mu|\tilde{\mu}]$ dominates $E_{\theta}[\mu|\tilde{\mu}]$ in the excess wealth order. Thus we have from Theorem 2 that search effort increases in θ. Excess wealth order dominance implies $E_{\theta'}[\mu|\tilde{\mu}]$ is a mean-preserving spread of $E_{\theta}[\mu|\tilde{\mu}]$. By well-known properties of sequential search, this implies profit is higher under θ' so profit and social surplus are increasing in θ. Finally, because the marginal distribution of μ does not depend on θ first-best profit, social surplus and search effort are fixed as θ rises. When $E_{\theta}[\mu|\tilde{\mu}] \stackrel{d}{=} \mu$, it must be that first-best values are archieved. Because $E_{\theta}[\mu|\tilde{\mu}]$ converges to μ in distribution as $\theta \to \bar{\theta}$, we then have that profit, social surplus and search effort converge to their first-best values from below.

10.8 Pareto Productivity Distribution

With a Pareto productivity distribution we can characterize μ^* this way:

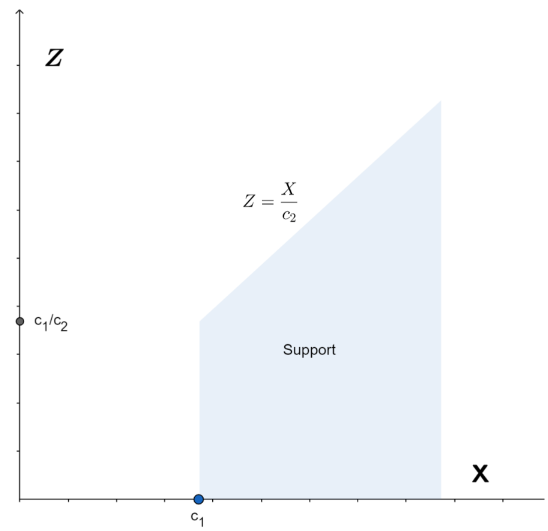
$$c = \int_{\mu^*}^{\infty} 1 - G_{\mu}(x)dx = \int_{\mu^*}^{\infty} \left(\frac{x_{\mu}}{x}\right)^{\theta_{\mu}} dx$$

Integration and solving for μ^* yields the optimal first-best threshold. Note that the first-best solution does not depend on the distribution of σ. We now derive several useful aspects of the joint distribution. Suppose that μ, σ is jointly distributed according to the density g from Assumption 3. We now derive the joint density of $\mu, \tilde{\mu} := \mu/\sigma$, which we denote f. By the transformation theorem, this is given by:

$$f(\mu, \tilde{\mu}) = g(\mu, \mu/\tilde{\mu}) \cdot \frac{\mu}{\tilde{\mu}^2}$$

$$f(\mu, \tilde{\mu}) = \frac{\theta_{\mu}\theta_{\sigma}x_{\mu}^{\theta_{\mu}}x_{\sigma}^{\theta_{\sigma}}}{x^{\theta_{\mu}+\theta_{\sigma}+1}} z^{-1+\theta_{\sigma}} \mathbb{I}\{\mu \geq x_{\mu}\}\mathbb{I}\{\mu/\tilde{\mu} \geq x_{\sigma}\}$$

Figure 4: Support for (X=μ,Z=$\tilde{\mu}$)



Now, we derive the marginal distribution of $\tilde{\mu}$. Consider first when $z \leq x_\mu/x_\sigma$. Then, the first indicator implies that the second is satisfied, and we can get the marginal:

$$f_{\tilde{\mu}}(\tilde{\mu}) = \int_{x_\mu}^{\infty} g(x,z)dx = \frac{\theta_\mu \theta_\sigma}{(\theta_\sigma + \theta_\mu)} z^{-1+\theta_\sigma} \left(\frac{x_\sigma}{x_\mu}\right)^{\theta_\sigma}$$

In the other case, the second indicator implies the first, so:

$$f_{\tilde{\mu}}(\tilde{\mu}) = \int_{x_\sigma \tilde{\mu}}^{\infty} f(\mu, \tilde{\mu})d\mu = \frac{\theta_\mu \theta_\sigma}{\theta_\mu + \theta_\sigma} z^{-1-\theta_\mu} \left(\frac{x_\mu}{x_\sigma}\right)^{\theta_\mu}$$

Now, we get the marginal CDF by cases:

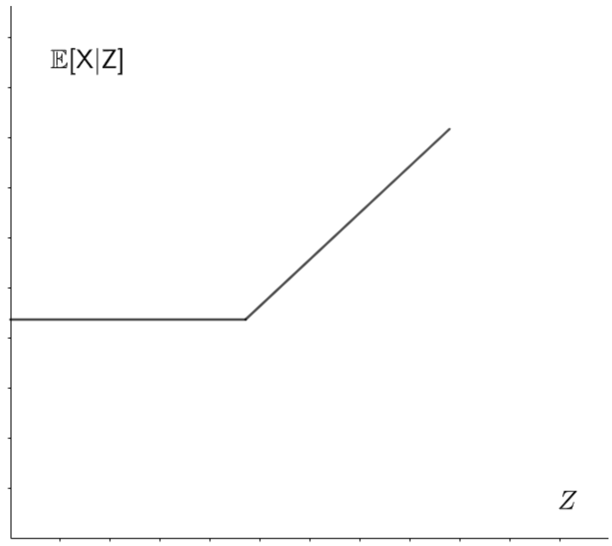
$$F(\tilde{\mu}) = \begin{cases} \frac{\theta_\mu}{\theta_\mu + \theta_\sigma} \left(\frac{x_\sigma}{x_\mu}\right)^{\theta_\sigma} \tilde{\mu}^{\theta_\sigma} \text{ if } \tilde{\mu} \leq x_\mu/x_\sigma \\ 1 - \frac{\theta_\sigma}{\theta_\mu + \theta_\sigma} \tilde{\mu}^{-\theta_\mu} \left(\frac{x_\mu}{x_\sigma}\right)^{\theta_\mu} \text{ else} \end{cases}$$

The conditional distribution is then:

$$f(\mu|\tilde{\mu}) = \frac{f(\mu, \tilde{\mu})}{f_{\tilde{\mu}}(\tilde{\mu})} = \begin{cases} \frac{x_\mu^{\theta_\mu + \theta_\sigma}(\theta_\mu + \theta_\sigma)}{\mu^{\theta_\mu + \theta_\sigma + 1}} \mathbb{I}\{\mu \geq x_\mu\} \text{ if } \tilde{\mu} \leq x_\mu/x_\sigma \\ \frac{(x_\sigma \tilde{\mu})^{\theta_\mu + \theta_\sigma}(\theta_\mu + \theta_\sigma)}{\mu^{\theta_\mu + \theta_\sigma + 1}} \mathbb{I}\{\mu \geq x_\sigma \tilde{\mu}\} \text{ else} \end{cases}$$

$$E[\mu|\tilde{\mu} = z] = \begin{cases} \frac{(\theta_\mu + \theta_\sigma)}{(\theta_\mu + \theta_\sigma - 1)} x_\mu \text{ if } z \leq x_\mu/x_\sigma \\ \frac{(\theta_\mu + \theta_\sigma)}{(\theta_\mu + \theta_\sigma - 1)} x_\sigma z \text{ else} \end{cases}$$

Figure 5: Conditional Expectation Function



For $z > x_\mu / x_\sigma$

$$\mathbb{E}[X|Z > z] = \frac{\theta_\mu + \theta_\sigma}{\theta_\mu + \theta_\sigma - 1} \cdot \frac{\theta_\mu}{\theta_\mu - 1} x_\sigma z$$

$$\mathbb{E}[X|Z > z] - \mathbb{E}[X|Z = z] = \frac{\theta_\mu + \theta_\sigma}{\theta_\mu + \theta_\sigma - 1} \cdot \frac{1}{\theta_\mu - 1} x_\sigma z$$

Thus, the First Order Condition determining SB search threshold z^* – $([\mu|\tilde{\mu} > z^*] - [\mu|\tilde{\mu} = z^*]) * \Pr(\tilde{\mu} > z^*) = c$ – for independently Pareto distributed μ and σ with parameters (x_μ, θ_μ) and (x_μ, θ_μ) can be re-written as

$$\frac{\theta_\mu + \theta_\sigma}{\theta_\mu + \theta_\sigma - 1} \cdot \frac{1}{\theta_\mu - 1} x_\sigma z^* \cdot \frac{\theta_\sigma}{\theta_\mu + \theta_\sigma} z^{*^{-\theta_\mu}} \left(\frac{x_\mu}{x_\sigma}\right)^{\theta_\mu} = c$$

or

$$\frac{\theta_\sigma}{(\theta_\mu + \theta_\sigma - 1)(\theta_\mu - 1)} \cdot \frac{x_\mu^{\theta_\mu}}{x_\sigma^{\theta_\mu - 1}} \cdot \frac{1}{c} = z^{*^{\theta_\mu - 1}}$$

$$z^* = \left(\frac{\theta_\sigma}{(\theta_\mu + \theta_\sigma - 1)(\theta_\mu - 1)}\right)^{\dfrac{1}{\theta_\mu - 1}} \cdot \frac{(x_\mu^{\theta_\mu}/c)^{\dfrac{1}{\theta_\mu - 1}}}{x_\sigma}$$

Re-arrange:

$$z^* = \left(\frac{x_\mu^{\theta_\mu} \theta_\sigma}{c(\theta_\mu + \theta_\sigma - 1)(\theta_\mu - 1)}\right)^{\dfrac{1}{\theta_\mu - 1}} \cdot \frac{1}{x_\sigma}$$

is increasing in x_μ, θ_σ and decreasing in x_σ, c (and probably increasing in θ_μ - not clear).

Note that the firm will select $z \geq x_\mu/x_\sigma$ if, and only if:

$$\frac{x_\mu \theta_\sigma}{(\theta_\mu + \theta_\sigma - 1)(\theta_\mu - 1)} \geq c$$

That is, so long as costs are not too large. If we plug in $z = x_\mu/x_\sigma$ (the knife-edge case), we have that:

$$\frac{\theta_\mu + \theta_\sigma}{\theta_\mu + \theta_\sigma - 1} \cdot \frac{1}{\theta_\mu - 1} \cdot \frac{\theta_\sigma}{\theta_\mu + \theta_\sigma} < \frac{c}{\theta_\mu x_\mu/(\theta_\mu - 1)} * \frac{\theta_\mu}{\theta_\mu - 1}$$

$$\frac{\theta_\sigma}{(\theta_\mu + \theta_\sigma - 1)\theta_\mu} < \frac{c}{\mathbb{E}[\mu]} < 1$$

It is possible that this is satisfied. We restrict attention to when it is not, which generates the assumption:

$$\frac{\theta_\sigma}{(\theta_\mu + \theta_\sigma - 1)\theta_\mu} < \frac{c}{\mathbb{E}[\mu]} \leftrightarrow \frac{\theta_\sigma x_\mu}{(\theta_\mu + \theta_\sigma - 1)(\theta_\mu - 1)} \geq c$$

For completeness, we consider the other FOC (when $z < x_\mu/x_\sigma$):

$$((p * \mathbb{E}[\mu|\tilde{\mu} = x_\mu/x_\sigma] + (1-p) * \mathbb{E}[\mu|\tilde{\mu} > x_\mu/x_\sigma]) - \mathbb{E}[\mu|\tilde{\mu} = x_\mu/x_\sigma]) * Pr(\tilde{\mu} > z^*) = c$$

Where $p = Pr(\tilde{\mu} < x_\mu/x_\sigma|\tilde{\mu} > z^*) \;\Rightarrow\; (1-p) * Pr(\tilde{\mu} < x_\mu/x_\sigma|\tilde{\mu} > z^*) = \dfrac{\theta_\sigma}{\theta_\mu + \theta_\sigma}$. Thus, the FOC is equivalent to

$$(1-p)Pr(\tilde{\mu} < x_\mu/x_\sigma|\tilde{\mu} > z^*)(\mathbb{E}[\mu|\tilde{\mu} > \frac{x_\mu}{x_\sigma}] - \mathbb{E}[\mu|\tilde{\mu} = \frac{x_\mu}{x_\sigma}]) = \frac{\theta_\sigma}{\theta_\mu + \theta_\sigma} \frac{\theta_\mu + \theta_\sigma}{\theta_\mu + \theta_\sigma - 1} \frac{x_\mu}{\theta_\mu - 1} = c$$

$$\frac{x_\mu \theta_\sigma}{(\theta_\mu + \theta_\sigma - 1)(\theta_\mu - 1)} = c$$

Thus, the closed form solutions assuming an interior solution are:

$$\tilde{\mu}^* = \frac{1}{x_\sigma}\left(\frac{x_\mu^{\theta_\mu}\theta_\sigma}{c(\theta_\mu + \theta_\sigma - 1)(\theta_\mu - 1)}\right)^{\frac{1}{\theta_\mu - 1}} \qquad \mu^* := \left(\frac{x_\mu^{\theta_\mu}}{c(\theta_\mu - 1)}\right)^{\frac{1}{\theta_\mu - 1}}$$

$$\frac{Pr_2}{Pr_1} = \frac{\theta_\sigma}{\theta_\mu + \theta_\sigma} \cdot \left(\frac{\theta_\mu + \theta_\sigma - 1}{\theta_\sigma}\right)^{\frac{\theta_\mu}{\theta_\mu - 1}}$$

$$\frac{\partial \log(Pr_2/Pr_1)}{\partial \theta_\sigma} = -\frac{\theta_\mu}{(\theta_\mu + \theta_\sigma - 1)(\theta_\mu + \theta_\sigma)\theta_\sigma} < 0$$

$$\lim_{\theta_\sigma \to \infty} \frac{Pr_2}{Pr_1} = 1$$

10.8.1 Proof of Proposition 5

All we need to apply the general results to this case is to establish an excess wealth order between μ and $\mathbb{E}[\mu|\tilde{\mu}]$ and how the latter one changes in θ_σ.

$$\mu \sim \text{Pareto}(x_\mu, \theta_\mu)$$

$$\mathbb{E}[\mu|\tilde{\mu}] \sim \begin{cases} \dfrac{\theta_\mu + \theta_\sigma}{\theta_\mu + \theta_\sigma - 1} * x_\mu, & \text{w/p } \dfrac{\theta_\mu}{\theta_\mu + \theta_\sigma}, \\ \text{Pareto}\left(\dfrac{\theta_\mu + \theta_\sigma}{\theta_\mu + \theta_\sigma - 1} * x_\mu, \theta_\mu\right), & \text{w/p } \dfrac{\theta_\sigma}{\theta_\mu + \theta_\sigma} \end{cases}$$

One could notice that the latter one is in fact a mean preserving contraction of the former one.

We can first derive excess wealth of μ for the quantile q threshold:

$$EW_\mu(q) = \int_{F_\mu^{-1}(1-q)}^{+\infty} (1 - F_\mu(x))dx = \frac{1}{\theta_\mu - 1} * x_\mu^{\theta_\mu} * \mu_q^{-(\theta_\mu - 1)} = \frac{1}{\theta_\mu - 1} * x_\mu * (1-q)^{\frac{\theta_\mu - 1}{\theta_\mu}}$$

Now we can derive it for the delegated search. Let us denote $p \equiv \frac{\theta_\sigma}{\theta_\mu + \theta_\sigma}$ and $A \equiv \frac{\theta_\mu + \theta_\sigma}{\theta_\mu + \theta_\sigma - 1}$.

$$EW_{\mathbb{E}[\mu|\tilde{\mu}]} = \int_{F^{-1}_{\mathbb{E}[\mu|\tilde{\mu}]}(1-q)}^{+\infty} (1 - F_{\mathbb{E}[\mu|\tilde{\mu}]}(x))dx = \frac{1}{\theta_\mu - 1} * p(x_\mu * A)^{\theta_\mu} (\mu_q')^{-(\theta_\mu - 1)}$$

where

$$\mu_q' = Ax_\mu * \left(\frac{p}{1-q}\right)^{\frac{1}{\theta_\mu}}$$

Therefore

$$EW_{\mathbb{E}[\mu|\tilde{\mu}]} = EW_\mu(q) * A * p^{\frac{1}{\theta_\mu}} = EW_\mu(q) * \frac{\theta_\mu + \theta_\sigma}{\theta_\mu + \theta_\sigma - 1} * \left(\frac{\theta_\sigma}{\theta_\mu + \theta_\sigma}\right)^{\frac{1}{\theta_\mu}} = EW_\mu(q) * \left(\frac{\text{Pr}_2}{\text{Pr}_1}\right)^{-\frac{\theta_\mu - 1}{\theta_\mu}}$$

Given the proof before we know that $EW_{\mathbb{E}[\mu|\tilde{\mu}]}$ is increasing in θ_σ and converges to 1 if $\theta_\sigma \to +\infty$. Thus μ EW order dominates $\mathbb{E}[\mu|\tilde{\mu}]$. Also, $\mathbb{E}[\mu|\tilde{\mu}]$ is EW increasing in θ_σ and converges to μ as $\theta_\sigma \to +\infty$. Theorem 2 and Proposition 3 conditions are satisfied which finishes the proof of Proposition 5.

10.9 Lognormal Productivity Distribution

Lemma 4 *Given two lognormal random variables $X_1 \sim LogNormal(u_1, s_1^2)$ and $X_2 \sim LogNormal(u_2, s_2^2)$ with the same mean, $X_1 \succ_{EW} X_2$ if $s_1 > s_2$.*

Proof. As shown by Belzunce et al. (2016) Corollary 2.1, a sufficient condition for $X_1 \succ_{EW} X_2$ is (1) $\lim_{p\to 0^+} F_{X_1}^{-1}(p) - F_{X_2}^{-1}(p)) \leq E[X_1] - E[X_2]$ and (2) that there exists a value p_0 such that:

1. $F_{X_1}^{-1}(p) - F_{X_2}^{-1}(p)$ is increasing on $[p_0, 1)$.
2. $F_{X_1}^{-1}(p) - F_{X_2}^{-1}(p)$ is decreasing on $(0, p_0)$.

Condition 1 is easy to check because in our case, X_1, X_2 have the same mean so the condition amounts to the $\lim_{p\to 0^+} F_{X_2}^{-1}(p) - F_{X_1}^{-1}(p)) \leq 0$. Since lognormal distribution are bounded below by 0, this condition holds because the quantile function evaluated at 0 for any lognormal is 0. To check the second condition, we write out the difference in quantile functions explicitly:

$$F_{X_1}^{-1}(p) - F_{X_2}^{-1}(p) = exp\Bigg(u_1 + s_1(\sqrt{2}erf^{-1}(2p-1))\Bigg) - exp\Bigg(u_2 + s_2(\sqrt{2}erf^{-1}(2p-1))\Bigg)$$

Taking the derivative wrt p gives:

$$s_1\sqrt{2}\sqrt{\pi}exp(-erf^{-1}(2p-1)^2)F_{X_1}^{-1}(p) - s_2\sqrt{2}\sqrt{\pi}exp(-erf^{-1}(2p-1)^2)F_{X_2}^{-1}(p)$$

We can factor out portions because they are positive:

$$\sqrt{2}\sqrt{\pi}exp(-erf^{-1}(2p-1)^2)\Bigg(s_1F_{X_1}^{-1}(p) - s_2F_{X_2}^{-1}(p)\Bigg)$$

Since the outside is positive we can focus just on the inside:

$$s_1F_{X_1}^{-1}(p) - s_2F_{X_2}^{-1}(p) = s_1exp\Bigg(u_1 + s_1(\sqrt{2}erf^{-1}(2p-1))\Bigg) - s_2exp\Bigg(u_2 + s_2(\sqrt{2}erf^{-1}(2p-1))\Bigg)$$

This is positive when:

$$log(s_1/s_2) + u_1 + s_1(\sqrt{2}erf^{-1}(2p-1))) \geq u_2 + s_2(\sqrt{2}erf^{-1}(2p-1)$$

Since $s_1 > s_2$, we can re-write again as:

$$\frac{log(s_1/s_2) + u_1 - u_2}{\sqrt{2}(s_2 - s_1)} \leq erf^{-1}(2p-1)$$

The left is constant in p. The right however is increasing in p, and as p approaches 0 the right approaches negative infinity meaning that the original function $F_{X_1}^{-1}(p) - F_{X_2}^{-1}(p)$ is decreasing over an initial range. As p approaches 1 the right approaches positive infinity, meaning that eventually the function is increasing. Thus there exists a value p_0 before which the function is decreasing and after which it is increasing if $s_2 < s_1$. Therefore we can apply the corollary from Belzunce et al. (2016) and $X_1 \succ_{EW} X_2$ if $s_1 > s_2$. The lemma implies that we can rank lognormal random variables with the same expectation using the shape parameter s. Higher s implies a higher excess wealth order.

Lemma 5 *Suppose two random variables X_1, X_2 are distributed according to Assumption 2. Then $V := E[X_1|X_1X_2^{-1}]$ is lognormal. Also $s_V < s_1$ as long as they are not perfectly positively dependent.*[19]

Proof.

$$(X_1, X_2) = \exp(Z_1, Z_2)$$

$$(Z_1, Z_2) \sim N(\cdot)$$

The distribution of $Z_1|Z_1 - Z_2 = y$ is:

$$N\left(u_1 + \frac{s_1^2 - s_{1,2}}{s_1^2 + s_2^2 - 2s_{1,2}}\left[y - u_1 + u_2\right], (1 - \frac{(s_1^2 + s_{1,2})^2}{(s_1^2 + s_2^2 - 2s_{1,2})(s_1^2 - s_{1,2})})s_1^2\right)$$

Thus $W := exp(Z_1)|Z_1 - Z_2 = y$ is log-normal with u, s parameters corresponding to the mean and variance of the underlying normal random variable $Z_1|Z_1 - Z_2 = y$. Then $E[exp(Z_1)|Z_1 - Z_2 = y]$ is $exp(u_W + s_W^2/2)$ which is:

$$exp\left(u_1 + \frac{s_1^2 - s_{1,2}}{s_1^2 + s_2^2 - 2s_{1,2}}\left[y - u_1 + u_2\right] + (1 - \frac{(s_1^2 + s_{1,2})^2}{(s_1^2 + s_2^2 - 2s_{1,2})(s_1^2 - s_{1,2})})s_1^2/2\right)$$

19. As long as the underlying normal random variables have a correlation coefficient that is not 1.

Notice that if we denote:

$$\gamma = exp\left(u_1 + \frac{s_1^2 - s_{1,2}}{s_1^2 + s_2^2 - 2s_{1,2}}\left[-u_1 + u_2 \right] + (1 - \frac{(s_1^2 + s_{1,2})^2}{(s_1^2 + s_2^2 - 2s_{1,2})(s_1^2 - s_{1,2})})s_1^2/2 \right)$$

$$\zeta = \frac{s_1^2 - s_{1,2}}{s_1^2 + s_2^2 - 2s_{1,2}}$$

we can write:

$$E[exp(Z_1)|Z_1 - Z_2 = y] = \gamma exp(\zeta y)$$

This is the conditional expectation function. To get the distirbution of $E[exp(Z_1)|Z_1 - Z_2]$ we apply this function to the random variable $Z_1 - Z_2$. This variable is equivalent to the log of X_1/X_2 so: $V = E[X_1|X_1X_2^{-1}] = E[exp(Z_1)|exp(Z_1 - Z_2)] \stackrel{d}{=} \gamma exp(\zeta log(X_1/X_2)) = \gamma(X_1/X_2)^{\zeta}$

By known properties of lognormal random variables, we also have that the power and multiplication of a lognormal remains lognormal, so:

$$\gamma(X_1/X_2)^{\zeta} \sim LogNormal(\zeta u_{X_1/X_2} + log(\gamma), \zeta^2 s^2_{X_1/X_2})$$

We now employ the last lemma to show that this distribution is excess wealth order dominated by the original distribution X_1. Because $E[X_1|X_1X_2^{-1}]$ is a mean-preserving contraction of X_1, we know the mean condition is satisfied. We then need only show the shape paramater is less than the original shape parameter of X_1, that is we wish to know if:

$$s_V^2 = \zeta^2 s^2_{X_1/X_2} < s_1^2$$

Plugging in $s^2_{X_1/X_2}$ and ζ:

$$\frac{(s_1^2 - s_{1,2})^2}{(s_1^2 + s_2^2 - 2s_{1,2})^2}(s_1^2 + s_2^2 - 2s_{1,2}^2) = \frac{(s_1^2 - s_{1,2})^2}{s_1^2 + s_2^2 - 2s_{1,2}} < s_1^2$$

$$\implies \frac{s_{1,2}}{s_1 s_2} = corr(X_1, X_2) \leq 1$$

Thus $s_V < s_1$ as long as there is no perfect positive dependence.

Lemma 6 *Suppose two random variables X_1, X_2 are distributed according to Assumption 2. Suppose X_1', X_2' are generated with all the same parameter values except with a higher value for $s_\sigma' > s_\sigma$. Then $s_V' < s_V$.*

Proof. In the last lemma we showed that $V := E[X_1|X_1X_2^{-1}]$ is lognormal when X_1, X_2

are lognormal. Further, the shape parameter s_V can be expressed as:

$$\frac{(s_1^2 - s_{1,2})^2}{s_1^2 + s_2^2 - 2s_{1,2}}$$

This expression is decreasing in s_2, therefore since σ corresponds to X_2, $s'_\sigma > s_\sigma$ implies $s'_V < s_V$. Because the expectation is invariant to changes in this parameter, we have that increases in this parameter result in decreases in the excess wealth order.

Lemma 7 *Suppose $Z \sim Lognormal(0, s_z^2)$ and is independent of X_1, X_2. $V := E[X_1|X_1X_2^{-1} \cdot Z^{-1}]$ is equal in distribution to $V' := E[X'_1|X'_1X_2'^{-1}]$ for some $s'_V < s_V$.*

Proof. Multiplying X_2 by an independent lognormal random variable with shift parameter of 0 is equivalent to a lognormal random variable with location parameter u_2 and shape parameter $s_z^2 + s_2^2$. The unlderying normal random variable has the same covariance with the normal random variable underlying X_1, thus $s_{1,2}$ is unchanged. Therefore it is equivalent to increasing s_2^2 directly. This implies that it is equivalent to adding mean-zero independent normal noise to the underlying normal random variables.

10.10 Proof of Proposition 4

Under the lognormal assumption without positive dependence, we can apply Lemma 5 with $X_1 = \mu, X_2 = \sigma$ and conclude that $E[\mu|\tilde{\mu}]$ is dominated by μ in the excess wealth order in a strict sense. Since excess wealth order implies mean preserving spread we have by Proposition 1 that first-best is not achieved. We also have directly by corollary 2.1 that search effort is lower under delegation as is the productivity expectation of hired workers in an FOSD sense. Finally, s_σ satisfies all requirements for θ so we can apply Proposition 3 to conclude profit and search effort increase and converge as s_σ decreases to 0.

If additional $s_{\sigma,\mu} = 0$ we have independence of μ, σ and we can apply Proposition 2 and say that productivity variance of hired workers is higher in the first-best.

10.11 Proof of Proposition 6

By the envelope theorem, the derivative of profit in the second-best with respect to search cost is:

$$\frac{\partial \pi^{SB}}{\partial c} = \frac{1}{Pr(\tilde{\mu} \geq \tilde{\mu}^*)} \geq 0$$

And in the first-best it looks similarly:

$$\frac{\partial \pi^{FB}}{\partial c} = \frac{1}{Pr(\mu \geq \mu^*)} \geq 0$$

Thus in both cases optimal profit is monotone decreasing in search cost. First-best profit is always weakly larger than delegated equilibrium profit for the same search cost. Thus there exists a value c^* such that all for all direct search costs above this value the firm delegates, and for all less it searches directly. As heterogeneity in σ decreases in the sense of Proposition 3, profit under delegation rises and it converges to first-best when the search costs are the same. Thus as heterogeneity decreases c^* falls and the firm delegates under a wider range of direct search costs. When σ is degenerate and workers are homogeneous in productivity variance, first-best profit is achieved. Then for any direct search cost that is weakly greater than the delegated search cost the firm delegates. Since we always assume there is some comparative advantage ($c_F > c$) the firm always delegates.

10.12 Inefficiency results

Proposition 8 *As soon as the first-best and delegated accepted regions are not identical (almost surely), then neither of them is a subset of the other one.*

Corollary 2.2 *In the delegated equilibrium, there is a set of excluded efficient candidates and a set of included inefficient candidates compared to the First Best.*

Proof.

Lemma 8 *There is a search over objects x in a set A with outcome $h(x)$. Let there be a total order on a finite set of possible search acceptance regions $\{A_k\}_{k=1}^n$:*

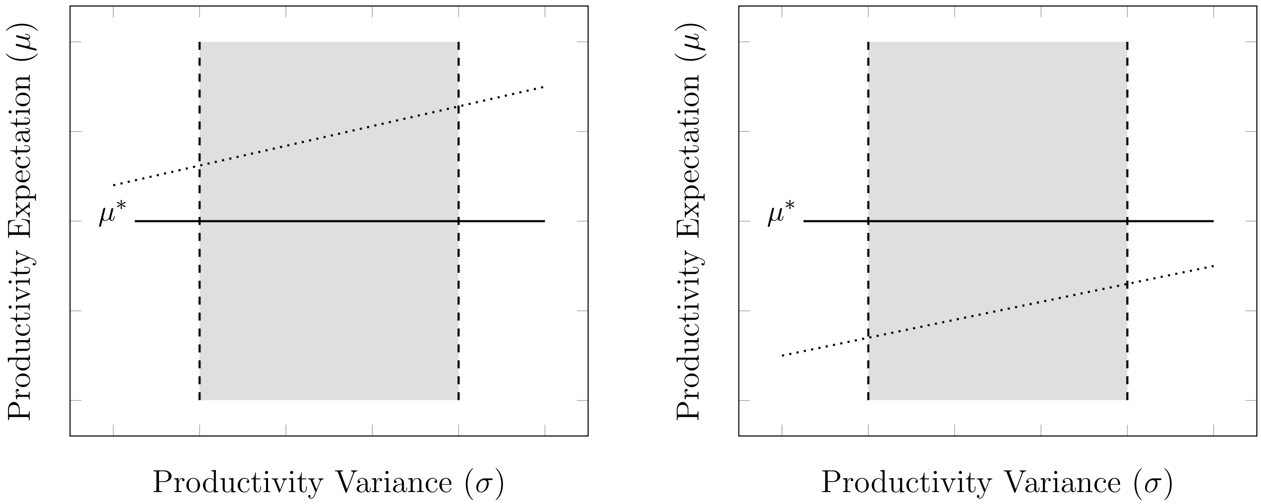
$$A \supset A_1 \supset A_2 \; ... \; \supset A_n$$

such that $\mathbb{E}[h(x)|A_k \backslash A_{k+1}]$ is increasing in k. The excess wealth is defined as

$$EW(A_k) = \Pr(A_{k+1}) * (\mathbb{E}[A_{k+1}] - \mathbb{E}[A_k \backslash A_{k+1}])$$

Then $EW(A_k)$ is decreasing in k.

Proposition 8 is obviously true in the case of full infinite support $S_{\mu,\sigma} = (-\infty, +\infty) \times (0, +\infty)$. In this case, the first-best and the delegated acceptance regions non-trivially intersect and define excluded efficient and included inefficient regions of positive measures. The cases of bounded or discreet support violating Proposition 8 statement are stylized in the two graphs below.



(a) Lower Boundary of Delegation Fully Above μ^* (b) Lower Boundary of Delegation Fully Below μ^*

Figure 6: Acceptance Regions with Bounded Support

Graph (a) can never be the case as the expected ability on the delegated acceptance line is higher than in the first best, which also means that the delegated firm's profit is higher than the first best which cannot be the case.

Using Lemma 8, we want to show that excess wealth for the first-best and delegated acceptance regions are not the same, which would violate one of the sequential search FOCs. All we need to do to apply Lemma 8 is to construct an ordered set of possible acceptance regions incorporating the first-best and the delegated excess wealth.

$$\begin{cases} A_1 = \{(\mu, \sigma) | \tilde{\mu} \geq \tilde{\mu}^*\} \\ A_2 = \{(\mu, \sigma) | \tilde{\mu} > \tilde{\mu}^*\} \\ A_3 = \{(\mu, \sigma) | \mu \geq \mu^*\} \\ A_4 = \{(\mu, \sigma) | \mu > \mu^*\} \end{cases}$$

The only technical condition to verify for applying Lemma 8 is that

$$\mathbb{E}[\mu | A_1 \backslash A_2] \leq \mathbb{E}[\mu | A_2 \backslash A_3] \leq \mathbb{E}[\mu | A_3 \backslash A_4]$$

which follows from staring at the graph and Assumption 1.

Now A_1 is the delegated acceptance region and A_3 is the first-best one and Lemma 8 together with search FOCs yields a contradiction

$$c = EW(A_1) > EW(A_3) = c$$

Therefore A_3 cannot be a subset of A_1. These two must non-trivially intersect creating non-empty excluded efficient and included inefficient regions. ■

10.13 Proof of Proposition 7

We can directly apply Proposition 8 to prove the statistical discrimination result as soon as we verify that the conditional expectation $E[\mu|\tilde{\mu} = x]$ is increasing, continuously differentiable and goes from negative to positive infinity. To do this we can write the conditional expectation as:

$$E[\mu|\tilde{\mu} = x] = pE[\mu|\tilde{\mu} = x, \sigma = \sigma_A] + (1 - p)E[\mu|\tilde{\mu} = x, \sigma = \sigma_B]$$

which given the definition of standardized productivity evaluates to:

$$E[\mu|\tilde{\mu} = x] = p\sigma_A x + (1 - p)x\sigma_B$$

This expression is clearly increasing because all coefficients are positive. It is also linear, meaning it spans the full real line for values of x that span the whole real line. Therefore the conditions for Proposition 8 are satisfied and we have that the delegated and first-best regions are not subsets. Because of the shapes of the regions, this implies that the proportion of hired workers from Group B must be smaller under delegation.