Delegated Recruitment and Hiring Distortions*

Jacob Kohlhepp† Stepan Aleksenko‡

August 12, 2021

Abstract

A firm delegates search for a worker to a recruiter. Productivity is uncertain prior to hire with recruiter beliefs characterized by an expectation and variance. Delegation occurs using a refund contract which is common in the industry. We analyze how delegation in this setting shapes search behavior and the composition of hires. We demonstrate that delegation is theoretically equivalent to making the search technology less accurate. This generates inefficiency: search effort and social surplus are lower under delegation than in the first-best benchmark. We show this inefficiency is driven by moral hazard with a multitasking flavor. The recruiter wastes search effort finding low variance workers at the expense of high expectation workers. As a result, as workers become more homogeneous with respect to productivity variance, delegation becomes more efficient. Our model provides a microfoundation for 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 and participants in the UCLA Theory Proseminar.

†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 with a similar message: people are everything. Yet, the search for talented people is frequently delegated. Between 2010 and 2018, the fraction of job postings for recruiting roles more than doubled. As of 2020, 18 percent of employed American workers 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 is this phenomenon economically meaningful? Put another way, are recruiters well-aligned agents of the firm, or does delegation distort search behavior and the types of workers hired?

To understand this question, we propose a model of delegated recruitment. In our model, a 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 throughout the paper we will assume is characterized by a variance and expectation. Contracts take a refund form, where the recruiter is paid an amount upfront but must refund a portion of the payment if the hired worker is fired.

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 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. Also, large firms are generally required by anti-discrimination law to treat employees fairly or equally (Carlsson, Fumarco, and Rooth 2014). This makes

1. We interviewed three recruiters and they reported such a contract was common, and the trial period is typically around 90 days. A survey by Top Echelon found 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.

it unlikely a firm can change its termination rules based on whether an employee was hired directly or through a recruiter.

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 mathematically equivalent to making the search technology less accurate. It is as if the firm faces a canonical multitasking problem, where the task she 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. We show that when the delegation-induced garbling of the search technology satisfies a weak condition, there is less search effort exerted under delegation. We further show this implies that the expected productivity distribution among selected workers is 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 strategy increase and converge to the first-best benchmark as heterogeneity in productivity variance decreases. This implies that agency loss in this setting depends crucially on the magnitude of differences in productivity uncertainty across candidates. When some workers exhibit less uncertainty (higher variance after search) than others, the recruiter will waste search effort hunting down such workers at the expense of expected productivity.

Finally, we show how to apply our results when productivity variance and expectation follow a Pareto or a log-normal distribution in the population. When our primitives follow a Pareto distribution, search strategies under delegation admit a closed-form and yield stark comparative statics. When our primitives follow a log-normal distribution, delegation always results in less search effort regardless of the correlation between productivity expectation and variance.

We explore two testable implications of the model. First, the specific way 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 show 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 micro-foundation for variance-based statistical discrimination that does not rely on risk attitudes. Second, our model predicts 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, than 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, the labor market in the future becomes more homogeneous with respect to productivity variance. This reduction in heterogeneity makes delegation more appealing, which increases the share of firms utilizing recruiters, thus increasing 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, in particular 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 that is popular in the discrimination literature, which 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 in this way. First, we describe how our work contributes to the existing literature. Second, we introduce the model. Third, we characterize the general solution to the delegation and first-best problems. Fourth, we derive results under semi-parametric assumptions on the joint distribution of productivity variance and expectation. Next, we apply our results to specific parametric examples and economic situations. Finally, we discuss the implications and conclude.

2 Literature

Our paper is most closely 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 contract shape to be of a refund form, and 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 (usually the agent has better information) with the bias of the agent. 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. We show that contract restrictions can generate misalignment like that 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 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 a prominent feature of many papers on labor search and matching. 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 the match specific productivity which is not known until hire.

3 Model

Players and Actions: There is a single risk neutral firm which wants to fill a single job opening. To fill the opening, it hires a recruiter to search. The recruiter is risk neutral and operates a sequential search technology for applicants. We assume that workers are not players, and are either fired or quit exogenously when $a < 0$. For simplicity we assume 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 fixed 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 (μ, σ) is given by:

$$a|(\mu, \sigma) = \mu + \sigma \cdot \varepsilon$$

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.[2] We denote PDF and CDF of ε by f and F respectively.[3] Because $E[a|\mu, \sigma] = \mu$ and $Var(a|\mu, \sigma) = \sigma^2$ we refer to μ as productivity expectation and we refer 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 search.[4]

Contracts: The firm is restricted to contracts of the form: $t(a) = \alpha - \beta\mathbb{I}\{a < 0\}$. We call these contract types *refund contracts* where α is the recruiter's payment if the search is successful and β is the refund when the employee fires 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: $u(a) = t(a) - N \cdot c$. We restrict attention to cases where some search would be optimal if the firm could operate the search technology directly.[5]

3.1 Model Comments

Our framework is general, and can be used to understand several mechanisms. One is an information story. 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, which is the joint distribution of posterior expectations and variances. Taking the distribution of posterior expectations and variances as our primitives makes our results quite general.

To see why, suppose a researcher has an information structure where the posterior belief about a can be written as in equation (reference here). One such information structure is

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

a normal-normal model, where the prior and the signal are normal. Then for that information structure the researcher can compute the resulting distribution of posterior means and variances. Our results can then be applied by checking whether the relevant conditions are satisfied. We demonstrate this process to understand how delegation impacts statistical discrimination in our application section.

Our framework can also be interpreted as capturing a matching process. Consider a model where productivity has a firm, individual, and match-specific component. If the firm component is public knowledge and the recruiter through search can uncover the individual component, 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. The joint distribution (μ, σ) is the resulting productivity distribution resulting from the matching of different types. Our results are then more general in the sense that we do not specify the types of firms and workers and the complementarity between them. 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 $a > 0$ the employee can negotiate higher wages so that future net productivity is 0. If $a < 0$ the employee separates from the employer (due to downward rigid wages). Either way, the firm's profit is a less payments to the recruiter.

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

For the first-best benchmark, we consider the case when the firm can operate the search technology directly.[6] 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:

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

These results are standard in the sequential search literature, however we provide derivations in Appendix Section 1 for completeness. To compare the first-best and equilibrium it is informative to rewrite Equation 1 in the following way:

$$(\mathbb{E}[\mu | \mu > \mu^*] - \mu^*) \cdot \Pr(\mu > \mu^*) = c \tag{2}$$

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

$$\mu^* = E[\mu | \mu \geq \mu^*] - \frac{c}{Pr(\mu \geq \mu^*)} \tag{3}$$

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

4.2 Delegation Equilibrium

We now consider the Bayesian Nash 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 a worker, expected utility from

6. 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.

selecting that worker given μ, σ is:

$$\alpha - \beta E[\mathbb{I}\{a < 0\}|(\mu,\sigma)] = \alpha - \beta\left(1 - F\left(\frac{\mu}{\sigma}\right)\right) = (\alpha - \beta) + \beta F\left(\frac{\mu}{\sigma}\right) \tag{4}$$

where F denotes the standard error ε cumulative density function. Equation 4 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 satisfactory (above 0). The recruiter searches over workers, evaluating them based on Equation 4. Ignoring the constant part of her utility, the decision to terminate search and hire the current worker is given by the following Bellman equation.

$$U = -c + \int \max\{\beta F(\tilde{\mu}), U\} d\tilde{G}(\tilde{\mu}) \tag{5}$$

U is the continuation value and $\tilde{G}$ is the CDF of $\tilde{\mu}$ derived from the joint CDF of (μ, σ). Equation 5 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:[7]

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

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 show later is the most relevant case.

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

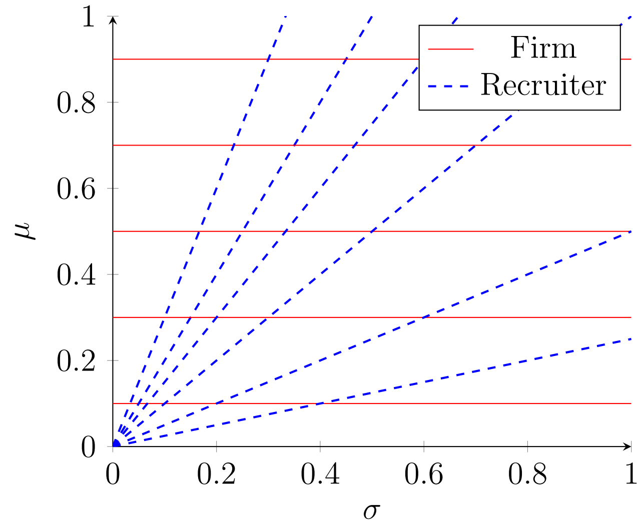


Figure 1: Indifference and Isoprofit Curves Over Worker Types

The recruiter's indifference curves all emanate from the origin, with higher indifference curves 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[8] while the firm's will be rectangular. Figure 2 illustrates this.

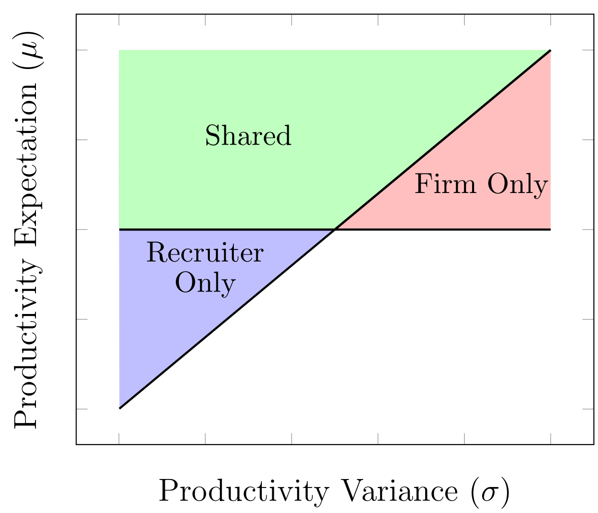


Figure 2: Recruiter vs. Firm Acceptance Regions Over Applicant Types

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*

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

equilibrium. In the firm only region are high variance and high expectation workers 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.* We use this diagram later to understand how incentive compatibility and the contract restriction impact equilibrium social surplus.

Figure 2 also provides intuition about the determination of equilibrium contract. The firm effectively chooses the slope of the diagonal line that defines the three regions. Examining Equation IC reveals β controls the slope.[9] More powerful incentives (higher β) increases the slope. A steeper slope increases the share of workers which are inefficiently excluded but decreases the share which 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 delegation, 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.[10]

For the rest of the paper it is assumed to be satisfied. Intuitively, Assumption 1 means larger standardized productivity implies larger expected productivity of the candidate. The assumption is quite natural given that $\tilde{\mu} = \mu/\sigma$.[11] If the conditional expectation is flat, the problem becomes uninteresting. There is no way to encourage the recruiter to search strategically, and so the firm will either not hire or offer a degenerate contract, where $\beta = 0$ and the recruiter returns the first applicant searched.

We are now ready to characterize the delegated search 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).

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

10. 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.

11. 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.

This defines a quite general problem, with two dimensional sequential search and moral hazard. However, our prior discussion of the recruiter's problem, in particular 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 the reservation rule $\tilde{\mu}$. Moreover, any threshold $\tilde{\mu}$ can be induced by an incentive compatible and individually rational mechanism. 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 cares only about it under 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:*

$$(\mathbb{E}[\mu|\tilde{\mu} \geq \tilde{\mu}^*] - \mathbb{E}[\mu|\tilde{\mu} = \tilde{\mu}^*]) \cdot \Pr(\tilde{\mu} \geq \tilde{\mu}^*) = c \tag{6}$$

and by (IC) and (IR) refund contract that 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.*

Theorem 1 has practical significance. It proves 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 6 and 2, 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, because we can always write $\mu = E[\mu|\tilde{\mu}] + \epsilon, E[\epsilon] = 0$. 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 nonparametric restrictions on the joint distribution of μ, σ. As it was 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 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 happens for only very specific non-generic 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 first order conditions characterizing them. 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 $\mathbb{E}[\mu|\tilde{\mu}]$, thus the profit must also be lower than in the first best.

Because a μ is always a mean-preserving spread of $E[\mu|\tilde{\mu}]$ and is a strict mean-preserving spread except a few knife edge cases, the proposition implies 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). Intuitively, we expect first-best to not be generally achievable because incentive compatibility requires the acceptance regions to be fundamentally different shapes. This is visualized in Figure 2. The proposition 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.

With this established, we 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, 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 the search effort, which is an important characteristic of search itself. We will define **search effort** as the percentage of candidates which are unacceptable and denote it q. This object maps one-to-one with the expected number of searches, which is $1/(1-q)$. We anticipate less search effort in the delegated problem as it happens in a standard moral hazard problem and characterize necessary and sufficient conditions for it later in this section. We first illustrate its connection to the productivity expectation 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 the mean-preserving spread condition from the last proposition does not generate clear comparative statics in terms of search effort. 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 to generate 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:*

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

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

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

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 and following the literature we define excess wealth order, and 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:*

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

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

Excess wealth order is a well-known variability order. It is discussed at length in Shaked and Shanthikumar (2007), and is sometimes called 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 second-order stochastic dominance.[12] With this in hand we present the following result.

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 $E[\mu|\tilde{\mu}]$ in the excess wealth order. In the lognormal case this is striking: for arbitrarily high positive correlation between between (μ, σ) dominance is maintained. This illustrates that for at least two parametric families the excess wealth order condition is fairly weak.

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, they would select solely based on expected productivity and choose B. However, the refund contract introduces misalignment. The recruiter cares not about expected productivity but rather how likely each candidate will be above the firing threshold. This induces the recruiter to care about productivity variance, which will cause the recruiter to inefficiently hire A over B.

We now formalize the intuition from the example. We know already that first-best social surplus is not achieved in equilibrium. However, we can say more: part of the lost social surplus is because the recruiter focuses search effort on finding low variance candidates. This focus is reflected in the composition of worker types in the equilibrium acceptance region.

The proposition highlights that the refund contract makes search less efficient. Additionally, search is less efficient precisely because the recruiter is wasting search effort minimizing variance instead of maximizing expectations. This manifests in a pool of acceptable workers that 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.

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

When does μ dominate $E[\mu|\tilde{\mu}]$ in the excess wealth order? Recalling that $\tilde{\mu} = \mu \cdot \sigma^{-1}$, we show in the parametric section that when μ, σ are log-normally distributed μ dominates $E[\mu|\tilde{\mu}]$ as long as there is not perfect positive correlation ($\rho < 1$). In this sense excess wealth order is not overly restrictive, and indeed, we have shown it is the least restrictive ordering needed to compare equilibrium and first-best search effort.

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. *$E_{\theta_1}[\mu|\tilde{\mu}]$ dominates $E_{\theta_2}[\mu|\tilde{\mu}]$ in the excess wealth order for all $\theta_1 > \theta_2$.*
3. *$E[\mu|\tilde{\mu}]$ converges to μ in distribution as θ converges to some value $\bar{\theta}$*

Then we have that social surplus, equilibrium profit, and expected number of searches are increasing in θ and converge to the first-best values as $\theta \to \bar{\theta}$

The proposition provides 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 a separate measure. In our case, this separate measure is standardized productivity ($\tilde{\mu}$). What is interesting is that if we take logs of the expression, we have:

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

How do we interpret the expression? Well, it implies 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 that cannot be properly distinguished from the productive task. In equilibrium, this manifests in workers which are first-order stochastically dominated in terms of productivity expectation but first-order stochastically dominant 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 $E[\mu|\tilde{\mu}]$ 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 joint 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. It also makes it a more credible way to model the posterior generated by a proper information structure. Generally posterior means and variances will not be independent.

In the Appendix we provide and prove a series of lemmas which 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 with the normal projection formula imply that $E[\mu|\mu \cdot \sigma^{-1}]$ 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, it can be shown that two lognormal random variables with the same shift parameter m can be ranked in the excess wealth order by their shape parameters, s.

The last step needed to apply our general results is to observe that s_σ satisfies the conditions for θ laid out in Proposition 4. 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. This further implies that our results regarding the under-provision of search effort and failure to achieve first-best will hold under a wide variety of conditions.

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 distributionally equivalent 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. This result can be connected back to our comparison to multitasking models. As we increase heterogeneity along the wasteful dimension through multiplicative noise.

6.2 Pareto

In this section we assume the productivity attributes are independently Pareto. The 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:*

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

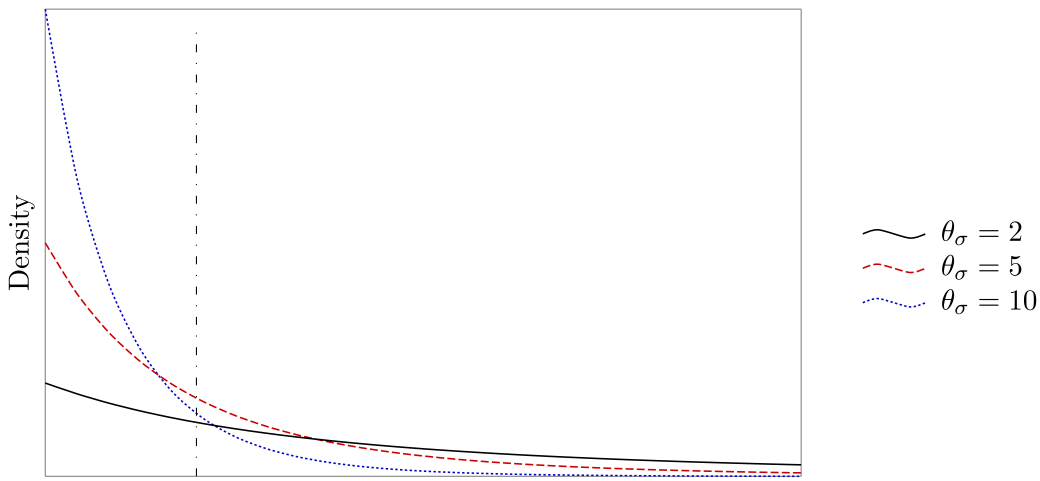
where both variables have finite expectations ($\theta_\mu > 1, \theta_\sigma > 1$*).*

Under this parametric distribution, the random variable $E[\mu|\tilde{\mu}]$ is almost Pareto: it has an atom at the beginning of the distribution and is Pareto conditional on being above the atom. This property is convenient for analysis because it implies the equilibrium reservation rule will have a closed-form solution.[13] We provide these closed form solutions in the Appendix.

13. It is well known that sequential search over Pareto distributed objects has an analytical solution.

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.[14] As $\theta_\sigma \to \infty$ we approach a perfectly homogeneous population with respect to productivity variance.

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 $\theta_\sigma \to \infty$.*

14. See the Appendix for a full proof.

The proof is provided in the Appendix, and is a direct application 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 essentially 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 weigh the cost and benefit when deciding whether to outsource, or delegate, the recruiting function. The benefit of delegation is the comparative advantage of the recruiter. The cost is agency loss: as we have shown through this paper, 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 to 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 effect 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.*[15]

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 delegated search is more profitable. As heterogeneity in the wasteful dimension vanishes, profit under delegation rises while profit from search directly stays constant. Eventually, when all agency loss is gone, only the comparative advantage effect remains, and delegated search is optimal.

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

If we think about different occupations as having different labor market pools, recruiter utilization in each occupations will depend on the amount of heterogeneity in productivity variance across workers. Occupations where workers have similar productivity variance will feature higher recruiter utilization. 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 average levels of productivity variance. This does not imply those occupations will have higher recruiter usage. What matters are the differences in variances among workers in the same occupation. As differences in variance become larger with an occupation, recruiter utilization becomes less likely.

7.2 Statistical Discrimination - To Be Finished

In this section, we use our model to understand how delegation can amplify statistical discrimination. To do so, we introduce a simple information structure. This has the added benefit of illustrating how (μ, σ), taken as primitives in our model, can be thought of as reduced-forms resulting from an information structure and a common prior.

Consider an economy where all workers can be divided into two groups. A share p belong to group A and a share $1-p$ to Group B. Index groups by i. Suppose the prior over productivity for both groups is given by $a_i \sim N(\bar{a}_i, \sigma_i^2)$. For simplicity assume that the prior mean is the same across the groups: $\bar{a}_A = \bar{a}_B$, but Group B has higher variance $\sigma_B > \sigma_A$. After searching a worker, the recruiter observes the group membership of the worker and a signal about productivity. The signal takes a normal form: $Y = a + \epsilon$, $\epsilon \sim N(0, s^2)$.

First we map this situation to our primitives. The recruiter Bayesian updates their normal prior with the normal signal. This generates a posterior mean and variance:

$$E[a_i|Y=y] = \bar{a} + \frac{\sigma_i^2}{\sigma_i^2 + s^2}(y - \bar{a}) \qquad Var(a_i|Y=y) = \frac{s^2\sigma_i^2}{\sigma_i^2 + s^2}$$

After integrating out the signal values y this yields a posterior mean distribution that is a a mixture of the following two normal distributions:

$$\mu|A \sim N(\bar{a}, \frac{\sigma_A^4}{\sigma_A^2 + s^2}) \qquad \mu|B \sim N(\bar{a}, \frac{\sigma_B^4}{\sigma_B^2 + s^2})$$

with mixture weights p and $1-p$. The posterior variance is a discrete random variable with two values:

$$\sigma|A = \frac{s^2\sigma_A^2}{\sigma_A^2 + s^2} \qquad \sigma|B = \frac{s^2\sigma_B^2}{\sigma_B^2 + s^2}$$

again with probabilities $p, 1-p$ respectively. Focusing first on the distribution of posterior expectations, $\sigma_B > \sigma_A$ implies that $\mu|A$ has a lower variance than $\mu|B$. Likewise the posterior variance for Group B is higher by the same argument.

We wish to understand how the fraction of Group B applicants hired changes between the first-best and equilibrium. That is, how does delegation impact who is hired? We show in the next proposition that delegation reduces the share of Group B applicants hired. Delegation induces variance-based statistical discrimination. This is because even if two applicants have the same productivity expectation, the recruiter will prefer applicant A because their posterior variance is lower.

8 Discussion

8.1 Beyond Recruiters

In our model, the recruiter is induced to act risk averse. The specific mechanism we provide which generates this behavior is the refund contract, which is prolific among recruiters that are hired externally. However, there are many reasons that 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 which do not have outside offers. The same article also suggests turnover as a possible metric to gauge 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 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 penalities" include serious crimes committed by employees, like sexual harassment and fraud. Thus human resource departments internalize many of the downside risks of hiring an incompetent or negligent employee, but do not internalize many of the upside benefits to hiring a high-performer.

These arguments support the idea that our main result, that delegation induces a form of risk aversion, is not limited to delegation *by external recruiters*. It is likely true for internal recruiters and human resource employees. This greatly broadens the impact of our analysis, because while only some firms use recruiters, a much larger share use some form of human resource department. It is estimated that as of 2018 there are 671,140 human resource workers (Labor Statistics 2019). In 2018, the U.S. Census Bureau estimated there were 664,757 businesses with 20 or more employees (Bureau 2021). Assuming most businesses with less than 20 employees do not hire a human resource specialist, this implies 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?

Several sources are 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. The cost of data science boot camps is often on the order of $2,000-$17,000 just for 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.[16]

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

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. This means 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 that 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 our context would likely make a European American recruiter have a higher productivity variance about a non-European American.

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

References

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

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.*

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.

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.

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.

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."

9 Appendix

9.1 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 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 corresponds 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 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.

9.2 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 was recall (so that the highest previously viewed μ could be carried as a state variable) the firm would never exercise the option. Because costs are already sunk, if it was previously optimal to search again it will still be optimal to search again the next period if the drawn μ is elss than the last

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

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 compactly be re-written as:

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

9.3 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 rewrite using standardized productivity ($\tilde{\mu}$):

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

9.4 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 the search over $\mathbb{E}[\mu|\tilde{\mu}]$ regardless of whether the first one contains more information about the worker or not.

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 $x' \in (x^* - \varepsilon; x^* + \varepsilon)$

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

Than 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 the threshold or decrease it 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 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 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 and the binding IR constraint in the delegated problem mean that the firm implicitly searches over $\tilde{\mu}$. Given the monotonicity Assumption 1 and Lemma 3, that is equivalent to searching over $\mathbb{E}[\mu|\tilde{\mu}]$, which is the firm's outcome. This optimal search is characterized by the FOC stated in the theorem.

Thus 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. Thus the problem reduces in the way stated in the proposition.■

9.5 Derivations for Pareto Productivity

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

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

Notice 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 $\sigma|\mathcal{D}_F$. ■

9.6 Proof of Proposition ??

From Lemma **??**, we know the acceptance region can be characterized by μ^* which solves:

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

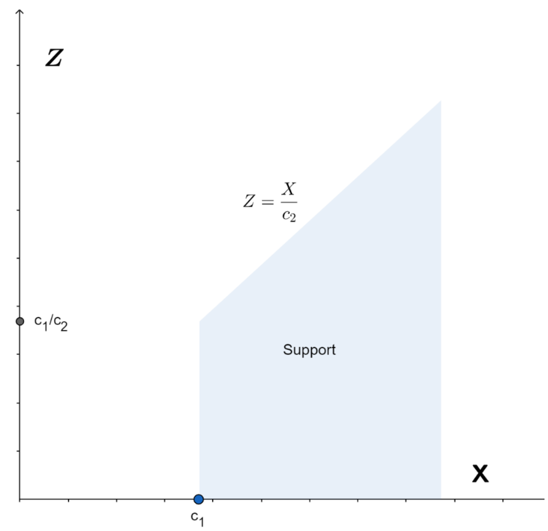
Integration and solving for μ^* yields the result. Note the first-best solution does not depend on the distribution of σ.

Suppose μ, σ 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 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:

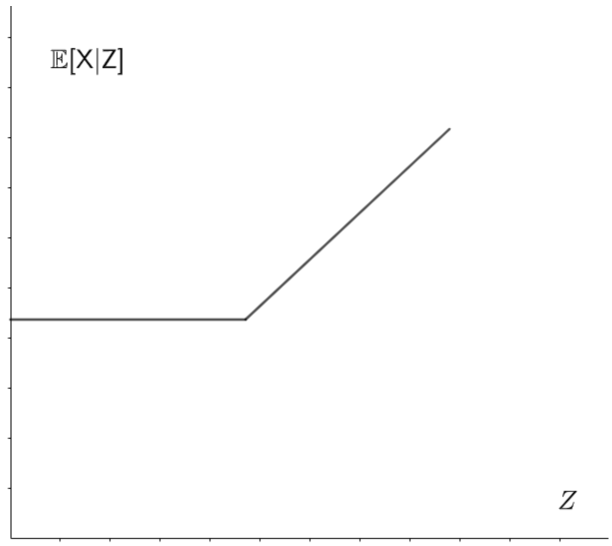
$$F(\tilde{\mu}) = \begin{cases} \frac{\theta_\mu}{\theta_\mu+\theta_\sigma} \left(\frac{x_\sigma}{x_\mu}\right)^{\theta_\sigma} \tilde{\mu}^{\theta_\sigma} \text{ if } \tilde{\mu} \leq x_\mu/x_\sigma \\ 1 - \frac{\theta_\sigma}{\theta_\mu+\theta_\sigma} \tilde{\mu}^{-\theta_\mu} \left(\frac{x_\mu}{x_\sigma}\right)^{\theta_\mu}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

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 rewritten 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, as 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 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:

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

9.7 Proof of Agency Loss Under Pareto

RAL is 1 minus the ratio of profits. We work with the ratio of profits for the proof:

$$\frac{\Pi_{SB}}{\Pi_{FB}} = (\theta_\mu + \theta_\sigma)\left(\frac{\theta_\sigma}{(\theta_\mu + \theta_\sigma - 1)^{\theta_\mu}}\right)^{\dfrac{1}{\theta_\mu - 1}}$$

This does not depend on c, x_μ, x_σ, proving part 1. For part 2:

$$log\left(\frac{\Pi_{SB}}{\Pi_{FB}}\right) = log(\theta_\mu + \theta_\sigma) + \frac{1}{\theta_\mu - 1}\left(log(\theta_\sigma) - \theta_\mu log(\theta_\mu + \theta_\sigma - 1)\right)$$

$$\frac{\partial}{\partial \theta_\sigma} log\left(\frac{\Pi_{SB}}{\Pi_{FB}}\right) = \frac{1}{\theta_\sigma + \theta_\mu} + \frac{1}{\theta_\mu - 1}\frac{1}{\theta_\sigma} - \frac{1}{\theta_\mu - 1}\frac{\theta_\mu}{\theta_\mu + \theta_\sigma - 1}$$

We want to know the sign of this derivative. After combining like terms it is positive if:

$$\frac{\theta_\mu^2 + \theta_\mu\theta_\sigma - \theta_\mu + \theta_\mu\theta_\sigma + \theta_\sigma^2 - \theta_\sigma + \theta_\mu^2\theta_\sigma + \theta_\mu\theta_\sigma^2 - \theta_\mu\theta_\sigma - \theta_\mu\theta_\sigma - \theta_\sigma^2 + \theta_\sigma - \theta_\mu^2\theta_\sigma - \theta_\mu\theta_\sigma^2}{(\theta_\mu + \theta_\sigma)\theta_\sigma(\theta_\mu + \theta_\sigma - 1)} \geq 0$$

which reduces to:

$$\frac{\theta_\mu(\theta_\mu - 1)}{(\theta_\mu + \theta_\sigma)\theta_\sigma(\theta_\mu + \theta_\sigma - 1)} \geq 0$$

Because we assumed both distributions had finite first moment: $\theta_\mu > 1$. So this is always true, and we have that RAL is decreasing in θ_σ. Now for part 3. Note we can re-write the limit:

$$\lim_{\theta_\sigma \to \infty} \tilde{\mu}^*(\theta_\sigma) = \frac{1}{x_\sigma}\mu^* \lim_{\theta_\sigma \to \infty}\left(\frac{\theta_\sigma}{\theta_\mu + \theta_\sigma - 1}\right)^{\frac{1}{\theta_\mu - 1}} = \frac{1}{x_\sigma}\mu^*\left(\lim_{\theta_\sigma \to \infty}\frac{\theta_\sigma}{\theta_\mu + \theta_\sigma - 1}\right)^{\frac{1}{\theta_\mu - 1}} = \frac{1}{x_\sigma}\mu^*$$

Since:

$$\Pi_{SB} = E[a|\tilde{\mu} = \tilde{\mu}^*] = \frac{\theta_\mu + \theta_\sigma}{\theta_\mu + \theta_\sigma - 1} x_\sigma \tilde{\mu}^*$$

Then:

$$\lim_{\theta_\sigma \to \infty} \Pi_{SB} = \lim_{\theta_\sigma \to \infty} \frac{\theta_\mu + \theta_\sigma}{\theta_\mu + \theta_\sigma - 1}\mu^* = \mu^*$$

Thus in the limit RAL goes to 0.

9.8 Proof of Proposition ??

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

$$Pr_1 \equiv Pr(\mu \geq \mu^*) = \frac{x_\mu^{\theta_\mu}}{\mu^{*\theta_\mu}}$$

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

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

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

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

Pr_1 – the first best acceptance region probability does not depend on the distribution of σ. The ratio Pr_2/Pr_1 is decreasing in θ_σ. Also

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

Thus we can also conclude that Pr_2 is always larger than Pr_1 (which ends the proof of proposition **??**).

9.9 Proof of Proposition ??

We will use the notations from Figure 2.

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

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

From proposition **??**, we can conclude that $p > q$ since $Pr(\text{FB}) < Pr(\text{SB})$.

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

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

where that notations on the RHS are used for mixture distribution. In other words, one could right 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, that 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).

9.10 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 and only if $s_1 > s_2$.*

The lemma implies that lognormals are easily ranked in terms of the 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 $s_V < s_\mu$ as long as they are not perfectly positively dependent.*[17]

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

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 $V' := E[X_1'|X_1'X_2'^{-1}]$ for some $s_V' < s_V$.*

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