Delegated Recruitment and Hiring Mismatch*

Jacob Kohlhepp† Stepan Aleksenko‡

April 16, 2021

Abstract

A firm hires a recruiter to find a worker. Search over candidates is noisy: the recruiter does not know productivity exactly but instead forms an estimate of a worker's productivity expectation and variance. We analyze how delegation under the commonly used refund contract between the firm and the recruiter affects search. We find the contract induces hiring mismatch, where the recruiter focuses on finding low variance instead of high expected productivity workers. More heterogeneity in worker productivity variance causes greater hiring mismatch and welfare loss. Our model predicts variance-based statistical discrimination in hiring like that suggested by the Heckman-Siegalman critique. We also analyze the relationship between direct and delegated search, establishing delegation is equivalent to adding noise to the search technology.

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.

1 Introduction

The allocation of talent across firms has always been an important part of the economy. This importance is reflected in the large literature on job search. In many of the models underlying this literature, *workers search for firms.* However, recent years have seen the rise of the reverse phenomenon: *firms searching for workers.* When firms search, they often delegate the process to recruiters or headhunters. As a result, it is estimated that 18 percent of employed American workers were hired by a recruiter (Black, Hasan, and Koning 2020). In this paper, we ask how delegation impacts the types of workers hired.

To answer this question, we build a model of delegated recruitment with uncertain worker productivity. We identify a form of hiring mismatch: the recruiter wastes search effort selecting workers with lower productivity variance than in the first-best benchmark. When workers are heterogeneous with respect to productivity variance, the recruiter focuses on finding low variance rather than high expected productivity candidates, making delegation less efficient compared to direct search. The implication is that firms will be more likely to hire a recruiter to search when workers are more similar in terms of productivity variance.

Our model predicts risk-neutral firms that outsource recruitment will statistically discriminate against groups with large unobserved productivity variance. Even if two workers have the same productivity expectation, the one with lower unobserved productivity variance will be hired. This provides a micro-foundation for variance-based statistical discrimination, essentially extending the Heckman-Siegalman critique to recruiters. Put another way, our results suggest delegation can generate labor market discrimination even in the absence of bias.

The model works in the following way. A firm can employ a recruiter to sequentially search for a worker. Search is noisy: the recruiter does not know productivity exactly but instead searches over an estimate of a worker's productivity expectation and variance. The search process is not observed: the recruiter and firm can only contract on the realized productivity of the hired worker.

To fix ideas, consider two candidates for a data science position at a risk-neutral firm. Candidate A is traditional: they went to a four year college, got a degree in statistics, and interned at a different well-known firm. Candidate B is nontraditional: they only have a high school degree and are self-taught, but they won a freelance machine-learning competition.[1] When comparing A and B prior to hire, B's productivity might have a higher expectation but also higher variance than A's.

We restrict attention to refund contracts. In such contracts, the recruiter receives an upfront payment and refunds some of the payment if the worker leaves for any reason during a trial period. Both anecdotal and survey evidence suggests this is the main contract form used by recruiters.[2][3] We assume the firm fires an employee if productivity is below an exogenous threshold. We think it is reasonable to focus on refund contracts because measuring productivity is hard but measuring whether someone is stilled employed is straight forward. The exogenous firing threshold is realistic in the context of refund contracts because firms are often required by anti-discrimination law to treat employees fairly. This prohibits most firms from shifting firing thresholds based on whether an employee was hired directly or through a recruiter.

With this in mind, we reconsider the example candidates A and B. 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, and as we show in the paper, this will cause the recruiter to inefficiently hire A over B.

Although the focus of this paper is recruitment, our general model can be used to think

1. E.g. a Kaggle competition. Such competitions often feature thousands of competitors, and winning requires a blend of ingenuity and luck.

2. We interviewed three recruiters and they reported such a contract was common, and the trial period is typically around 90 days.

3. 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.

about any economic context where a principal delegates search over uncertain objects to an agent using refund-type contracts. For example, some hedge fund managers are paid a fixed fee plus a bonus if an investor's portfolio achieves a certain return. Because a bonus is just a negative refund, our model applies to this setting.

When appropriately applied, our framework is convenient because it simplifies the original delegation search problem into a standard search problem without delegation. A brief intuition for this is as follows. Regardless of the contract parameters, the refund contract form preserves the recruiter's preference over candidates; therefore, the delegated search is equivalent to the firm searching directly over recruiter's indifference curves of workers. Equivalently, under delegated search the firm cannot distinguish between candidates that are equally good for the recruiter; it is as if there is additional "noise" in the search technology. This theoretical contribution provides a clean way to analyze delegated search over objects of uncertain utility, and it may be of independent interest for future work on other subjects.

The paper is structured in this way. In the next section we describe how our work contributes to and fits into the broader literature. Next, we introduce the model. Following that, we present non-parametric results assuming only independence of the productivity variance and productivity mean. We then derive additional comparative statics assuming a Pareto distribution for both productivity characteristics. Finally, we discuss interpretation of our results, as well as implications for policy and further research.

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.

We also consider our paper to be 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 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. In this way, our paper is similar in spirit to Frankel (2020). Frankel considers a hiring manager who selects applicants to hire for a job. The firm can discipline the manager's decisions using hard information like a test score. Similar to our paper, Frankel considers a specific context rather than an abstract delegation problem.

Our model is motivated by the large literature on labor search and matching. Early search and matching models incorporated homogeneous workers, but eventually transitioned to include heterogeneity in productivity prior to hire. Two papers which represent this stage of the literature well are . Another literature, starting with , highlights how worker-firm match quality is learned over the course of the employment relationship. Our paper considers how intermediation and the distribution of ex-ante heterogeneity interact to impact the matching process. By focusing on a previously overlooked part of the matching process (recruiters), we help build a more comprehensive picture of the aggregate labor market.

Finally, our paper is related and contributes to the large literature on labor search and matching. We study recruiters, a intermediary through which matching occurs. We draw from papers like Shimer and Smith (2000) and Postel–Vinay and Robin (2002) which feature applicants 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. Our paper is also related to learning about match quality. Just as in Jovanovic (1979), in our model, bad match quality leads to termination of employment relationships. We contribute to this literature by noting that termination is observable and thus easier to contract on then productivity in recruiter-firm contracts. As a result, learning about match quality will impact who gets hired *through recruiters.* We believe our paper provides a look into one part of the matching function, and may be helpful for future search and matching models which wish to examine the general equilibrium impact of intermediation on labor markets.

3 Model

Players and Actions: There is a single risk neutral firm which desires 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.

Applicants and Information: Applicants are fully described by a single attribute, a, which can be interpreted as output produced net of some fixed market wage.[4] a is realized at the end of the game and it is not observed by either the firm or the recruiter prior to hire. Instead, when the recruiter samples an applicant, it observes two characteristics: (μ, σ). These attributes are distributed in the population with joint CDF $G(\mu, \sigma)$. Conditional on observing these two characteristics, the distribution of a for the applicant is $N(\mu, \sigma^2)$. μ can be interpreted as the productivity expectation of worker i. σ^2 can be interpreted as the productivity variance. It is important to understand that these are μ, σ describe the worker after search but prior to hire. They are best thought about as describing the conditional distribution of a after observing some characteristics of the worker.

Search Technology: The recruiter can take i.i.d. draws of applicants at unit cost c. There

4. Another interpretation is that a is productivity relative to some break-even type of worker, where break-even is a worker who produces exactly the market wage.

is no limit to the number of searches. After drawing an applicant, the recruiter views their attributes (μ, σ) and then may either suggest the applicant for hire or continue search. We assume there is no recall (this is without loss). Search takes place in a single period, so there is no discounting. The firm does not observe any of the applicants the firm searches but does not select.

Payoffs and Contracts: The firm is restricted to contracts of the form: $t(a) = \alpha - \beta \mathbb{I}\{a < 0\}$. That is a refund contract where α is the recruiter's payment if the search is successful and β is the refund if the employee does not fit the job (the recruiter must return it to the firm if the realized productivity a is less than 0). If the recruiter rejects the contract, both the firm and the recruiter get 0. If the recruiter accepts the contract she runs a search over the applicant pool and offers one of them to the firm. The firm receives the realized net output a less the payments to the recruiter $t(a)$ and the recruiter receives the transfer $t(a)$ according to the contract less the search costs. We restrict attention to cases where recruitment is profitable. That is, $E[a] = E[\mu] > c$. [5]

Timing:

- First the firm commits to a contract consisting of (α, β).
- Then the recruiter accepts or rejects the contract. If she rejects, she receives her outside option 0 (the firm receives 0 too).
- If she accepts, the recruiter sequentially searches for an applicant by taking i.i.d. draws from G.
- The recruiter suggests one applicant to the firm, and the firm hires the applicant.
- Finally, a realizes, the firm receives a. The contract realizes.

5. There is always a contract inducing the recruiter to choose the first sampled applicant ($\alpha = c,\ \beta = 0$). Under the assumption $E[a] > c$, this contract is profitable for the firm, and therefore the outside option of not hiring anyone (which would yield profit of 0) is never a part of any equilibrium.

3.1 Model Comments

We can interpret a as the residual benefit of hiring an employee above the market wage per period. Then total profit is exactly equal to a because after the first period, a is realized. If $a > 0$, the employee can request a wage increase so that in subsequent periods $a' = 0$. If $a < 0$ and a is only observed by the firm and the employee, the firm cannot lower the wage because the employee can still obtain the market wage elsewhere. Thus the firm fires the employee.

Our modeling of productivity uncertainty is motivated by the idea that upon finding a worker, a recruiter forms a belief about productivity based on observable characteristics (experience, education, etc.). We do not formally specify the information structure or beliefs. Rather, we specify G which is the joint distribution of the posterior mean and variance. Intuitively, this joint distribution is a reduced-form object generated from a prior belief over a and some information structure.

In our model, firing is not endogenous: the firm terminates the employee if $a < 0$. We think this is reasonable for three reasons. First, there is a sense in which productivity is unobservable and not easily contractible. Indeed, this is why we limit contracts to be binary. Firms can probably prove a worker was not meeting some average standard, but it is unlikely they can prove an employee is below an arbitrary threshold. Second, firms can sometimes be sued for discrimination or wrongful termination if their firing decision is not based on "just cause." There is a sense in which it would be unfair if the firm had a different termination rule for recruited and directly hired applicants. Third, there is a sense in which the firm cannot commit to a firing rule ex-ante, due to the hard-to-observe nature of a.

The two applicant dimensions, μ and σ, can be thought of as quantifying the point estimate of applicant productivity and the uncertainty surrounding the estimate. As an example, consider again Applicants A and B from the introduction. Recall Applicant A is traditional, while Applicant B is nontraditional. It is possible a recruiter sees these two as having the same expected performance, but the second might be perceived as having greater variance, because there is a large chance they will either outperform or under-perform the market wage.

Importantly, two different recruiters could have different beliefs over the same applicant due to either true information asymmetries or biases. In this sense, σ also captures how familiar an applicant is to a recruiter: how frequently the recruiter has placed similar candidates. It also may be related to issues of homophily and statistical discrimination: recruiters may better understand the background of similar race applicants. In this way, the primitive $G(\mu, \sigma)$ can be thought of as a reduced form of the true distribution of productivity combined with an information structure.

3.2 Other Examples

The working example throughout will be the firm/recruiter relationship. But we believe our framework extends to many circumstances where search is over objects of uncertain quality. In the context of mortgage brokers, a would represent the return to the lender including the possibility of default. The broker then earns a fixed fee after closing the deal, and likely only holds liability for the return if the loan turns out to be unsuitable, which is determined by a set of legal requirements. In the case of venture capitalists, a represents the profit of the startup. When venture capital fund managers collect a fixed fee, they may be concerned with whether or not the return exceeds a specific threshold rather than with the exact realization of a. These examples highlight the key ingredients which make our model a good approximation:

1. Search is delegated.

2. The value of the searched object is uncertain prior to consumption.

3. The principal wishes to maximize expected value.

4. The agent wishes to maximize the probability the value exceeds a certain threshold.

5. The probability of exceeding a threshold is well approximated by the ratio of the mean and variance.[6]

6. The firm does not have much control of this threshold.

When a situations fits these requirements reasonably well, our model provides a parsimonious way to capture it. As we show in the next section, the value of our framework is that the originally complicated delegation problem can be reduced to single variable maximization problem.

4 Non-Parametric 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.[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 μ and the problem reduces to one-dimensional search.

Lemma 1 *In the first-best benchmark, where the firm operates the search technology directly, the acceptance region is given by:*

$$\mathcal{D}_F = \{\mu, \sigma | \mu \geq \mu^*\}$$

where μ^* solves:

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

6. This takes the form of the normal assumption on $a|\mu, \sigma$ in our model.
7. Equivalently, when there is no contract restriction.

Or equivalently

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

The proof is provided in the Appendix. This lemma formalizes the idea that the firm does not intrinsically care about the uncertainty (σ) dimension of an applicant. Symmetry of the distribution of a combined with risk-neutrality implies that the firm is indifferent between the upside potential and downside risk. Lemma 1 also shows that the optimal threshold μ^* is positive since it is equal to the firm's expected profit, which is also positive due to the assumption that $\mathbb{E}[a] > c$. The last expression emphasizes that the firm selects μ^* by equating the marginal gain of an additional search (the left side) with the cost of an additional search (the right side).

4.2 Firm-Optimal Contract

We now consider the firm-proposing optimal contract, where the firm must delegate search to the recruiter. The full problem is characterized below.

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

s.t.

$$\alpha - \beta + U \geq 0 \tag{IR}$$

$$\mathcal{D}_R = \{\mu, \sigma|\beta E_a[\mathbb{I}\{a > 0\}|(\mu,\sigma)] - U \geq 0\} \tag{IC}$$

where U is the value function of the recruiter (after α is sunk) during the sequential search problem, defined as:

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

Throughout this paper, we call solution to this problem the ”second-best” or the ”delegated search equilibrium.” In this problem, the firm must choose a search strategy for the recruiter and a compensation scheme. Because search is not observed, the search strategy must be incentive compatible given the contract form restriction. Throughout the rest of this paper, we denote Φ to be the standard normal CDF. With this notation, we can write the utility the recruiter obtains from a candidate with characteristics μ, σ as:

$$\beta E[\mathbb{I}\{a > 0\}|(\mu, \sigma)] = \beta \Phi\left(\frac{\mu}{\sigma}\right)$$

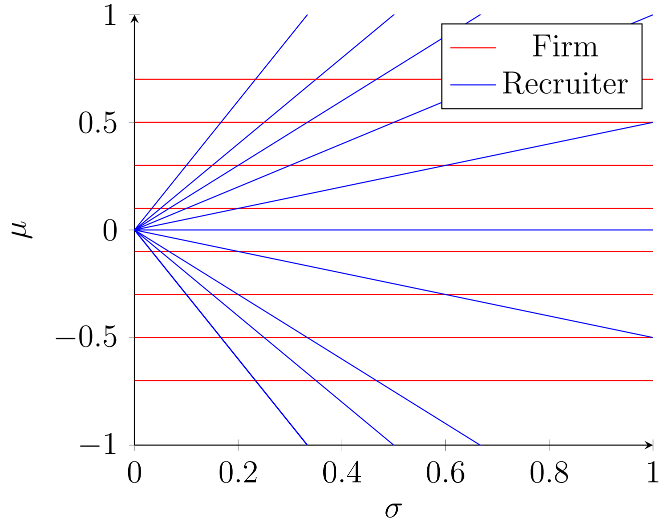
This simplification shows that the recruiter ranks candidates according to the ratio μ/σ. We will call this ratio **standardized productivity** throughout this paper.

Definition 1 *(Standardized productivity) Standardized productivity, denoted $\tilde{\mu}$, of a candidate is the ratio of her expected productivity over her productivity uncertainty (in other words, how many standard deviations candidate's expected productivity is away from zero)*

$$\tilde{\mu} = \frac{\mu}{\sigma}$$

Comparing delegated and direct search, the firm's indifference curves are horizontal lines while the recruiter's curves are sloped lines emanating from the origin. Higher indifference curves have steeper slopes. When μ is positive, the recruiter prefers lower applicant uncertainty. When μ is negative, she prefers higher σ. Intuitively, the recruiter gains or loses nothing from applicant upside or downside, and wants to pick someone with the highest probability of being acceptable ($a > 0$). This is shown graphically in Figure 1.

Figure 1: Indifference Curves



In order to focus on non-trivial cases of delegation, we introduce a weak assumption about the joint distribution G, which is expressed as follows.

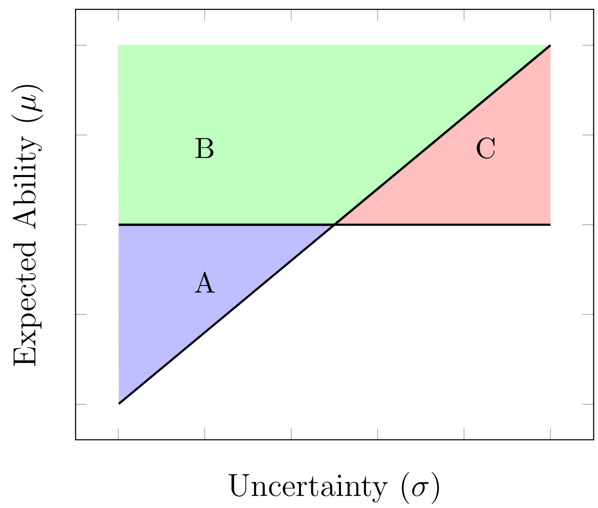
Assumption 1 $\mathbb{E}[\mu|\tilde{\mu} = x]$ *is weakly increasing in* x.

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. Intuitively, Assumption 1 means that larger standardized productivity implies larger expected productivity of the candidate. The assumption is quite natural given that $\tilde{\mu} = \mu/\sigma$. To break this assumption, the association between μ, σ most be positive and extremely strong. Consider the case of perfect dependence, where $\sigma = \gamma\mu$ for some constant γ. Even this is not sufficient, because this results in a flat conditional expectation. What is necessary is some form of dependence where the expectation grows faster than linearly or is not monotonic. For example, $\sigma = \gamma\mu^2$ will cause the conditional expectation to be decreasing.

For the rest of the paper this is assumed to be satisfied. This assumption allows there to be a clear relationship between the firm's and the recruiter's preferences over $\tilde{\mu}$. This is important for further analysis as it eliminates situations where there is degenerate optimal search due to the diametrically opposed preferences over applicants.

From examining the shape of the recruiter's indifference curves we can see the shape of the acceptance regions. The recruiter's acceptance region will be a triangle, the area above one of the upward sloping blue lines. The firm's acceptance region will be a rectangle. Figure 2 illustrates this.

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



The green region represents applicants which both the firm (in the first-best benchmark) and the recruiter hire. The red region represents applicants which the firm would hire if it conducted search directly, but that are not selected by the recruiter in equilibrium. The blue region represents applicants which the firm would prefer not to hire, but that are selected anyway by the recruiter in equilibrium. The firm uses β to choose the slope of the diagonal line, trading-off the applicants which are inefficiently included or excluded.

Remark 4.1 *When the support of expected productivity is positive, we can see from the graph that the recruiter is over-selecting "safe-bets." These low μ, low σ applicants are represented by the blue region. We call them "included inefficient" because they are selected but the firm would prefer they were excluded. The recruiter is also under-selecting "diamonds in the rough." These high μ, high σ applicants are represented by the red region. We call them "excluded efficient" because the firm would prefer they were included but they are excluded.*

We formalize our observations about the recruiter's problem in Lemma 2.

Lemma 2 *Given β, define $M(u)$ as the CDF of $u := \beta\Phi(\tilde{\mu})$. In any incentive compatible contract, the recruiter's acceptance region is given by:*

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

where u^* solves:[8]

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

Lemma 2 is proved in the Appendix, but intuitively the proof is the same as Lemma 1, with the small additional step of defining the random variable u. By reducing the two-dimensional problem into one dimension, we can use the well-known result that the optimal strategy is a reservation rule, u^*. Using Lemma 2 we can greatly simplify the second-best delegated search problem to one of single-variable maximization, where the firm maximizes surplus but searches over $\tilde{\mu}$.

Theorem 1 *The firm-optimal contract can be solved by solving the unconstrained problem:*

$$\max_x \mathbb{E}[\mu | \tilde{\mu} \geq x] - \frac{c}{Pr(\tilde{\mu} \geq x)}$$

which has the below F.O.C., uniquely defining $\tilde{\mu}^*$:

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

Then β and α are given by:

$$\beta = \frac{c}{(E[\Phi(\tilde{\mu}) | \tilde{\mu} \geq \tilde{\mu}^*] - E[\Phi(\tilde{\mu}) | \tilde{\mu} = \tilde{\mu}^*]) \cdot \Pr(\tilde{\mu} > \tilde{\mu}^*)}$$

8. 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.

$$\alpha = \beta - \left(\beta \cdot E[\Phi(\tilde{\mu})|\tilde{\mu} \geq \tilde{\mu}^*]) - \frac{c}{\Pr(\tilde{\mu} > \tilde{\mu}^*)}\right) = \beta \cdot (1 - E[\Phi(\tilde{\mu})|\tilde{\mu} = \tilde{\mu}^*])$$

Corollary 1.1 *Under the above assumption, the firm's profit is positive and equal to* $\mathbb{E}[\mu|\tilde{\mu} = \tilde{\mu}^*]$. *Then* $\tilde{\mu}^*$ *must positive too since the profit is positive under the assumption made in Section 3.*

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 again by a simple threshold rule, which is defined by a single first-order condition. The entire problem essentially collapses into standard sequential search over modified objects.

Theorem 1 also holds deeper economic insight. Imposing the binary refund contract is equivalent to allowing the firm to search directly but only allowing the firm to observe $\tilde{\mu}$ rather than μ. Moral hazard in this setting is similar to making the search technology less accurate. The firm now must search based on an object that is correlated with what it cares about, but not perfectly correlated.

Proposition 1 *If* μ *and* σ *are independent, the distribution of* σ *in the first-best acceptance region* $\mathcal{D}_F$ *first-order stochastically dominates the distribution of* σ *in the recruiter's acceptance region* $\mathcal{D}_R$.

Proposition 1 demonstrates the contract form induces the recruiter to select less risky (lower σ) applicants. Part of the reason for this is that even if μ, σ are independent, $\sigma, \tilde{\mu}$ will often be negatively affiliated. This in turn implies that $\sigma|\tilde{\mu} > x$ will stochastically dominate the unconditional distribution of σ. Consider the following parametric example.

Example: $\mu \sim U[0, 1], \sigma \sim exp(\lambda)$ where the two variables are independent. Then the joint pdf of $v := \mu/\sigma, \sigma$ is given by $f(v, \sigma) = \mathbb{I}\{v\sigma \leq 1\}\lambda exp(-\lambda\sigma)\sigma$. Two random variables are negatively affiliated if and only if their joint pdf is log submodular. Taking logs of the pdf gives: $log(f(v, \sigma)) = log(\mathbb{I}\{v\sigma \leq 1\}) - \lambda\sigma + log(\sigma)$. This function is submodular: if the indicator is 0 for at least one of two pairs of values, it will also be 0 for the pairwise minimum.

5 Parametric Analysis

In order to compare the first and second-best regions, and perform comparative statics, we impose a parametric assumption.

Assumption 2 ***Parametric Assumption.*** *μ, σ are distributed independently with marginal Pareto distributions. That is, their joint pdf 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 ($\theta_\mu > 1, \theta_\sigma > 1$).

In this parameterization, x_μ, x_σ are the shift parameters, which give the lower bounds of the support. $\theta_\mu, \theta_\sigma$ are the shape parameters which control the amount of mass near the beginning of the distribution. Assumption 2 yields closed-form solutions to both the first-best and equilibrium problems. In the Appendix, we derive various intermediate joint and marginal distributions, including the distribution of $\tilde{\mu}$. Of particular interest is the following conditional expectation:

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

Note that this is weakly increasing in z, as required by Assumption 1. However, it is flat over an initial range. We know from Lemma 3 the firm will never select $\tilde{\mu}^*$ to be on the flat portion: it will always set $\tilde{\mu}^*$ to be the minimum of the support or it will set it above the flat portion. The first case is rather strange. it means that the search technology is not even being used, because the first applicant drawn is immediately hired. Thus, if the firm had the outside option of drawing a random applicant, then it would never use the recruiter. For this reason, and to focus on an interior solution for comparative statics, we make one additional assumption.

Assumption 3 ***Non-degenerate Search.*** *Search cost c and the parameters of the joint Pareto distribution satisfy:*

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

Then we can characterize interior closed-form solutions for the first-best and the second-best problems for the given distribution G.

Proposition 2 ***Closed Form Solutions.*** *If Assumption 2 is satisfied, the first-best benchmark has a unique solution characterized by:*

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

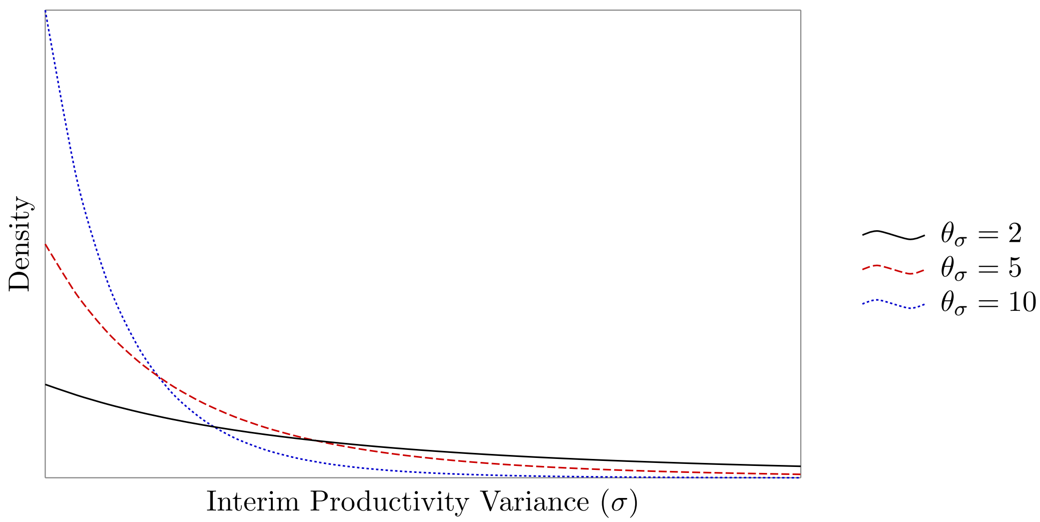
If Assumption 3 is also satisfied, the firm-optimal contract has a unique solution characterized by:

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

The proof is provided in the Appendix because the steps are mainly algebraic with little economic insight.

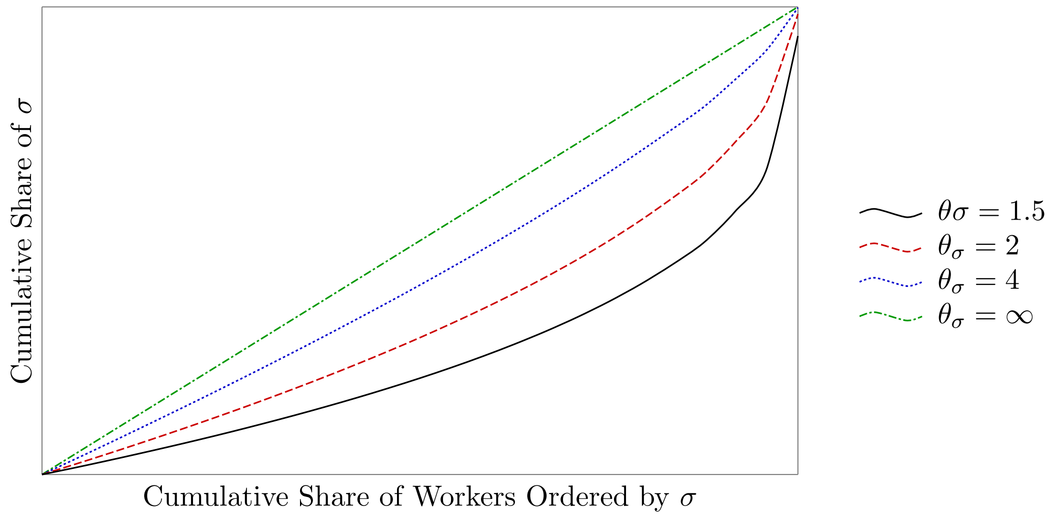
In the last set of results, we will focus on comparative statics in θ_σ. θ_σ represents the level of heterogeneity with respect to σ. As θ_σ rises, the population becomes more homogeneous. A greater number of workers have an productivity variance near the lower bound of the support given by x_σ. When it falls, the population becomes more heterogeneous. This can be visualized in Figure 3 which plots three different Pareto distributions with the same x_σ but different θ_σ.

Figure 3: Densities of σ for Different Values of θ_σ



θ_σ is also a measure of inequality in a distribution, which is why it is often called the Pareto Index. In our case, θ_σ measures the level of inequality of productivity variance. When θ_σ is around 1.16 we can say 20 percent of workers possess 80 percent of productivity variance. When it increases to 1.59 the distribution becomes more equitable: around 33 percent of workers possess 67 percent of productivity variance. When $\theta_\sigma \to \infty$ we achieve perfect equality, where all workers have the same productivity variance. Figure 4 visualizes the inequality interpretation by plotting Lorenz curves of σ for different θ_σ values. The 45 degree line represents perfect equality and it is the Lorenz curve we converge to when $\theta_\sigma \to \infty$.

Figure 4: Lorenz Curves for Different Values of θ_σ



An interesting question is whether there is more or less search in the first-best as opposed to the second-best. We measure search intensity as $Pr((/mu, \sigma) \in \mathcal{D}_i)^{-1}$, which is the expected number of searches given acceptance region $\mathcal{D}_i$.

Proposition 3 ***Search Intensity.***

1. *Search intensity is always higher in the first-best than the second-best, that is:*

$$\frac{1}{Pr((\mu, \sigma) \in \mathcal{D}_{FB})} \geq \frac{1}{Pr((\mu, \sigma) \in \mathcal{D}_{FB})}$$

2. *Second-best search intensity is increasing in θ_σ and is equal to the first best search intensity in the limit $\theta_\sigma \to \infty$ (the acceptance regions are identical in the limit as well).*

Proposition 3 illustrates there is always less search in the delegation equilibrium than in the first-best. This is because the marginal value of each additional search is lower when the firm has to delegate: the recruiter will sometimes pass over workers the firm wants and will sometimes select someone the firm does not want. As variance heterogeneity decreases

(θ_σ rises) the agency loss declines and the marginal value of additional search improves. Eventually, when heterogeneity vanishes, first-best search intensity is achieved.

This proposition means that the contract restriction in combination with variance heterogeneity induces a form of moral hazard, which is similar in flavor to canonical multitasking models (Holmstrom and Milgrom 1991). To see this connection, suppose we define two "tasks': search along the μ dimension and search along the 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 μ/σ. What is interesting is that if we take logs of the expression, we have:

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

What does this mean? Well, it implies the firm can only "buy" a 1 percent increase in μ if it is willing to also "buy" a 1 percent decrease in the maximum σ. Thus we are in a situation where total effort is rewarded, but there is a wasteful task that cannot be properly distinguished from the productive task.

Proposition 4 ***Comparison of First-Best and Second-Best Acceptance regions.***

1. *The distribution of σ in the first-best acceptance region first-order stochastically dominates the distribution of σ in the recruiter's acceptance region.*

2. *The distribution of μ in the first-best acceptance region first-order stochastically dominates the distribution of μ in the recruiter's acceptance region.*

Proposition 4 first restates the general finding from the nonparametric results section. Under delegated search, accepted workers have lower productivity variance than in the first-best. The second part of the proposition states something new. Under the joint Pareto assumption, first-best accepted workers have higher productivity expectation than second-best accepted workers in a first-order stochastic sense. Thus, in this particular parameterization, the recruiter is constructing a candidate pool which is safer but worse. There is a sense in which "safe bets" are being accepted at the expense of "diamonds in the rough," because the recruiter cannot be made to value upside potential.

Social welfare is another way to measure misalignment between the first and the second best solutions in this model. We wish to focus on the relative share of first-best social surplus which is lost due to delegation. To accomplish this, we introduce the concept of relative agency loss.

Definition 2 *Relative agency loss (RAL) is defined as the percentage of first-best surplus lost due to delegation. The social surplus is equal to the firm's profit in both FB and SB. Thus, we can compute relative agency loss in the following way.*

$$RAL = 1 - \frac{\Pi_{SB}}{\Pi_{FB}}$$

Comparative statics of relative agency loss are given in Theorem 2.

Theorem 2 *Relative agency loss has the following characteristics:*

1. *invariant to c, x_μ, x_σ*
2. *increasing in θ_σ*
3. $\lim_{\theta_\mu \to \infty} RAL(\theta_\mu) = 0$

Part one of the theorem states the overall level of variances and means in the population (x_σ, x_μ) do not impact efficiency. It is the amount of heterogeneity or inequality that matters. The shape parameter θ_σ captures the degree of inequality in productivity variance distribution. When θ_σ is near 1, the distribution exhibits near perfect inequality: a small number of applicants have near infinite variance while the rest have low variance. This makes it profitable for the recruiter to waste effort avoiding high variance applicants, which is reflected in the agency loss. When θ_σ becomes large, all applicants essentially have the same variance. The recruiter becomes less tempted to search along the variance dimension, less search effort is wasted, and we approach first-best social surplus.

The theorem has a nice interpretation when we think about estimating the productivity of a worker based on observable characteristics. Then μ is the prediction and σ is expected prediction error. When observable characteristics are informative about σ, the recruiter effectively wastes energy examining these characteristics. To be more concrete, suppose there is some observable characteristic that impacts variance but not the mean. The firm would actually be better off if this characteristic was not observable.

One application of our findings about relative agency loss has to do with the decision to outsource recruiting. When firms decide whether to outsource the recruiting function, they must trade-off the agency loss that comes with outsourcing with the savings in terms of opportunity cost. We analyze how productivity variance heterogeneity impacts this decision in one final proposition.

Proposition 5 ***The Choice To Delegate.*** *Suppose the firm can either outsource recruiting to a recruiter who can perform search at unit cost c_R or it can perform search directly at unit cost c_F. Let c^* be the highest ratio of search costs c_R/c_F at which the firm prefers to delegate. Then c^* is increasing in θ_σ and converges to 1 as $\theta_\sigma \to \infty$.*

Intuitively, the ratio of c_R/c_F measures the comparative advantage of a recruiter conducting search compared to someone like the CEO. A recruiter is probably more efficient at searching for candidates, and a CEO probably has a higher opportunity cost of time, so $c_R/C_F < 1$. [9] c^* is the maximum relative cost of search for the recruiter under which delegation is viable. If θ_σ increases, workers become more homogeneous with respect to productivity variance and delegation is preferred under a broader range of conditions. When workers are homogeneous ($\theta_\sigma \approx \infty$) delegation results in 0 agency loss and the firm will delegate for any recruiter search cost up to and including c_F. The proof is in the Appendix, but this result follows almost directly from Theorem 2. This is because relative agency loss

9. For example, a CEO could use the time to plan firm strategy.

has a direct mapping to the choice to delegate if the firm 's outside option is to conduct search on its own at higher cost.

The proposition also has some general equilibrium implications. Consider a competitive market of homogeneous recruiters facing firms of heterogeneous opportunity costs c_F. The proposition implies the share of firms using a recruiter will depend on θ_σ. If we think about different industries as having different labor market pools, recruiter utilization in each industry will depend on the amount of heterogeneity in productivity variance in each labor market. We expect to see more recruiters in industries where worker productivity exhibits similar variance.

6 Discussion

Our model has several testable predictions about recruiters and the labor market. First, industries with large amounts of heterogeneity in productivity variance will tend to not have recruiters.

Second, recruiters should suggest older experienced workers over younger inexperienced workers. Since Jovanovic (1979), studies have found empirical evidence of learning about match quality (lange and other one). In particular, Fredriksson, Hensvik, and Skans 2018 shows match quality is better among experienced workers than inexperienced. This is consistent with the existence of better information about experienced workers. In the language of our paper, older experienced workers should exhibit lower productivity variance, which in turn should make them preferred by recruiters. This might help explain the empirical observation that bad labor market conditions hurt younger workers more than older workers.

Third, recruiters should be more likely to work with candidates of similar socioeconomic background, even when they have no intrinsic biases or prejudice. This is because similarity of backgrounds should improve communication and thus reduce the level of perceived productivity variance.

Fourth, when recruiters can partially direct search they will search among more homogenous workers. To see this, recall the closed form solution for $\tilde{\mu}^*$ in Proposition 2. $\tilde{\mu}^*$ is the expected profit of the recruiter before search but after receiving the upfront payment. $\tilde{\mu}^*$ is increasing in θ_σ, meaning that the recruiter ranks worker pools by their level of variance heterogeneity. If given the choice, the recruiter will choose to search among the more homogeneous pool (all else equal).

All the analysis in this paper highlights one key insight: the common bonus contract causes recruiters to waste effort finding applicants with low hiring risk rather than applicants with high expected productivity, despite the fact that the firm wants the recruiter to focus solely on expected productivity. This misaligned search strategy results in distorted applicant pools. Economy-wide, this means some applicants are receiving less attention from recruiters than is socially optimal.

Our findings are important for socioeconomic mobility. 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 the median annual salary) 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.[10]

Without credit constraints it would not necessarily be an issue that high quality signals of productivity are expensive. But credit constraints and imperfect information are a real part of the labor market landscape. 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

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

if recruiters have no intrinsic bias towards wealthy candidates. This will tend to reinforce existing socioeconomic inequality.

6.1 Statistical Discrimination and the Heckman-Siegelman Critique

Our findings are important for the literature on discrimination in hiring. Most past work considers the traditional setting where workers search for a job and a firm rejects or accepts them. Our paper suggests statistical discrimination may also be present when firms search for a worker. In particular, contract limitations in the recruiter-firm relationship can result in statistical discrimination. To see this consider two groups which have the same productivity expectation but different variances. Our main theoretical result implies firms which delegate to a recruiter will hire a larger share of the low variance group *even when the firm is indifferent between the groups.*

For a policymaker wishing to reduce statistical discrimination, our paper highlights the importance of policies which equalize uncertainty about productivity across groups. One class of policies that accomplish this goal are those which make high quality signals less costly. Our paper also highlights the unintended consequences of policies like "ban-the-box." Such policies, which make it illegal to ask whether someone has a criminal background, can exacerbate differences in perceived variance. There is evidence such policies have negative impacts on direct job applicants from protected groups (Agan and Starr 2018). Our results suggest there will be similar negative consequences for firm-directed search.

Our paper is also tightly connected to the Heckman-Siegelman critique. Introduced in Heckman and Siegelman (1993) and expanded upon in Heckman (1998), a central part of the critique is that differences in unobserved productivity variance across two groups can confound inference about discrimination in either direction (false positives or false negatives). This is true even when the researcher can randomly assign applicants to firms using a correspondence study. Our results imply this concern will also be valid in the context of recruiters.

Correspondence studies are now quite popular. As researchers began adapting such methods to test for discrimination in recruiter behavior (or more generally, in firm-driven search behavior) it will be important to adjust for differences in productivity variance. Failing to do so will bias results in an unknown direction.

An interesting consequence of our theorem about relative agency loss is that industries with more recruiter activity should be less sensitive to the Heckman-Siegelman critique. To see why, note that the Heckman-Siegelman critique relies on heterogeneity in variances. If variances are homogeneous, the critique is not relevant. We showed that as variance becomes more heterogeneous firms are less likely to hire a recruiter. As variance becomes homogeneous firms are more likely to hire a recruiter. In the extreme case of constant variance the Heckman-Siegelman critique has no bite and hiring a recruiter is always optimal.

Since the 1990s, correspondence study methods have advanced. In particular Neumark (2012) proposed a convenient method for adjusting correspondence study estimates to account for differences in unobserved productivity variance.[11] The method involves estimating σ for each group using a heteroskedastic probit. This allows a researcher to decompose discrimination into a direct and variance-based effect. The literature thus far has mainly used this method as a robustness test: if coefficients remain significant after the adjustment then there is evidence of taste-based discrimination. But our paper gives the method new life: it can be used to estimate the full joint distribution of μ, σ. Given this joint distribution estimated for several industries, it would be possible to test our main theorems. It should be the case that measures of recruiter density should be positively correlated with the amount of heterogeneity in worker unobserved productivity variance.

6.2 The Contract Restriction

Restrictions on the space of contracts generate inefficiency in our model. If productivity is perfectly contractible a "sell-the-firm" contract achieves the first-best outcome and profit.

11. In any future studies of recruiter behavior we recommend a similar method be used.

Indeed, simply allowing the firm to commit to a firing rule for the employee upfront will allow the firm to achieve the first-best. Why then is the the binary refund contract (which motivated this entire paper) so common? Before engaging in this discussion, we want to highlight this is a similar question discussed in Holmstrom and Milgrom (1991) under the heading "Missing Incentive Clauses in Contracts." As the authors point out, much of contracting theory points to the importance of explicit incentives in achieving efficiency. Yet, real-world contracts often lack any explicit contingencies. Like many before us, we believe the lack of explicit incentives can be traced back to a desire for transparency and simplicity.

First, measuring productivity of an individual person is difficult if not impossible, especially in jobs with team-based work and productivity spillovers. Measuring whether someone still works somewhere is easy, and much less ambiguous. Thus the binary refund contract might be popular because it is less likely to be challenged and not costly to measure. Even when productivity can be measured, firms may be hesitant to disclose information needed to verify productivity in court. Such information might give away trade secrets. Secondly, a binary contract matches how people often think: in dichotomies. Is the engineering team better or worse since New Employee A joined? Is New Employee A performing adequately or not? These binary judgments are more straightforward and less difficult than assessing productivity on some form of continuous, absolute scale. Third, firms are bound by anti-discrimination laws to treat employees equally (Carlsson, Fumarco, and Rooth 2014). Since a company cannot set firing standards that are unique for each employee, this renders the threshold essentially exogenous and fixed.

7 Conclusion

In the course of this paper, we develop a novel theoretical model of delegated recruitment with uncertain productivity. We show that the general, nonparametric version of the model can be reduced to single variable unconstrained maximization. We show under an independence assumption that a risk-neutral recruiter will exhibit risk-averse behavior and over-select low-variance candidates. We then impose a parametric assumption in order to compute comparative statics. Under the parametric assumption, we show the agency loss from delegation depends on the amount of variance heterogeneity in the labor force. We expand upon this result and link it to the decision to outsource recruitment.

The subject of intermediaries in the labor market has received little attention in both the theoretical and applied literature. This paper builds a cogent framework that captures the impact of delegation on the types of workers which are hired. Moving up the job ladder is a process that frequently operates through recruiters, and it is key to improving socioeconomic outcomes. As a result, better understanding the incentives current contracts create for recruiters is more than just an academic exercise. It has wide-ranging implications for socioeconomic mobility, labor force composition, discrimination and government policy. To this end, an extension of our analysis would be empirically test our model through an experiment. In such an experiment, artificial candidate profiles would be posted to a job search website like LinkedIn. By varying observable characteristics, profiles could be constructed to have different productivity variance and expectations. The study would then measure the number of recruiters which messaged the profile and see how this varies with productivity variance and expectations. Our model predicts recruiters should focus attention on low-variance candidates.

Our model represents a first step towards understanding the trade-offs firms face when deciding how to recruit. We show that agency loss from outsourcing or delegating the recruiting function depends on the level of heterogeneity in productivity variance. Our model treats the composition of the labor market as a primitive, and we abstract from workers as strategic players. An important extension we leave for future work is how the presence of recruiters impacts labor market composition. In particular, it would be interesting to understand how signaling concerns endogenously generate the joint distribution of productivity variance and productivity expectations.

References

Agan, Amanda, and Sonja Starr. 2018. "Ban the box, criminal records, and racial discrimination: A field experiment." *The Quarterly Journal of Economics* 133 (1): 191–235.

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

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

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.

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

———. 2020. "Selecting Applicants."

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.

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

Heckman, James J, and Peter Siegelman. 1993. "The Urban Institute audit studies: Their methods and findings." *Clear and convincing evidence: Measurement of discrimination in America:* 187–258.

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.

Neumark, David. 2012. "Detecting discrimination in audit and correspondence studies." *Journal of Human Resources* 47 (4): 1128–1157.

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/`.

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`.

8 Appendix

8.1 Proof of Lemma 1

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:

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

8.2 Proof of Lemma 2

The dynamic programming problem of the recruiter is given by:

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

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

Re-writing, this yields:

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

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, that is we will select all μ, σ that satisfy:

$$u^* \leq \beta \Phi(-\mu/\sigma) \leftrightarrow \beta \Phi^{-1}(\frac{u^*}{\beta}) \leq \frac{\mu}{\sigma}$$

where u^* can be uniquely obtained in a similar argument to the last proof. Simply define M as the cdf of $u := \beta \Phi(-\mu/\sigma)$. Then:

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

Integration by parts gives:

$$c = -[(1 - M(u))(u - u^*)]_{u^*}^{\bar{u}} + \int_{u^*}^{\bar{u}} (1 - M(u)) du$$

Since the first term is 0, this simplifies to:

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

8.3 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.

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

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

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

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 2, 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:

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

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

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 $x := \Phi^{-1}(u^*/\beta)$.

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.■

8.4 Proof of Proposition 1

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

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

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 $\sigma|\mathcal{D}_F$. ■

8.5 Proof of Proposition 2

From Lemma 1, we know the acceptance region can be characterized by μ^* which solves:

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

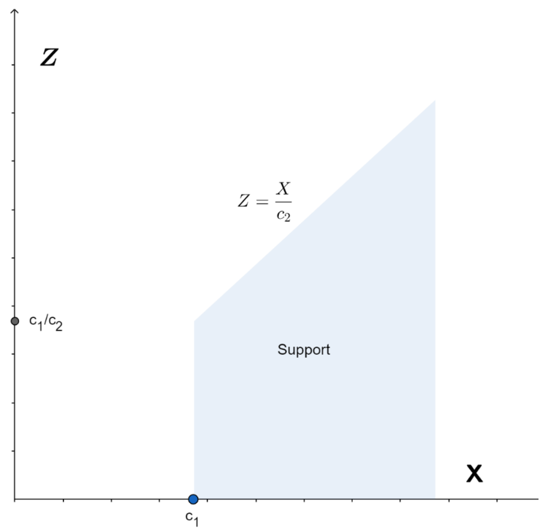
Integration and solving for μ^* yields the result. Note the first-best solution does not depend on the distribution of σ.

Suppose μ, σ jointly distributed according to the density g from Assumption 2. 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 5: 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:

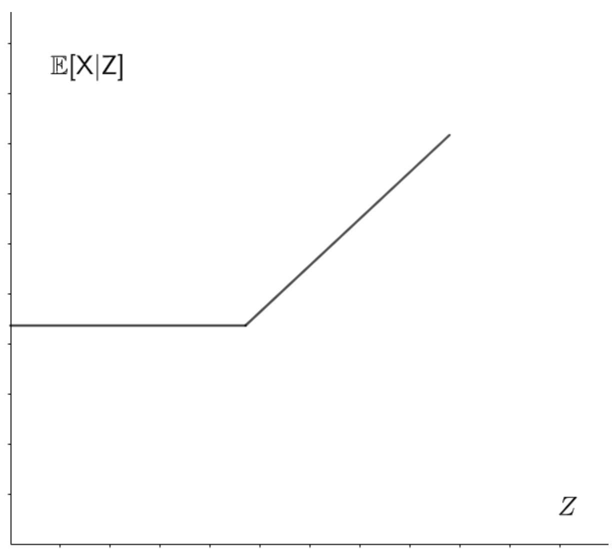
$$F(\tilde{\mu}) = \begin{cases} \frac{\theta_\mu}{\theta_\mu+\theta_\sigma}\left(\frac{x_\sigma}{x_\mu}\right)^{\theta_\sigma}\tilde{\mu}^{\theta_\sigma} \text{ if } \tilde{\mu} \leq x_\mu/x_\sigma \\ 1 - \frac{\theta_\sigma}{\theta_\mu+\theta_\sigma}\tilde{\mu}^{-\theta_\mu}\left(\frac{x_\mu}{x_\sigma}\right)^{\theta_\mu}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

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 6: 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 3:

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

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

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

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

8.6 Proof of Theorem 2

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:

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

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

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.

8.7 Proof of Proposition 5

We wish to find the ratio c_p/c_A where the principal is indifferent between delegation and direct search. That is, we wish to solve for search costs where profit in the first and second best is equalized. To do this, note that:

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

We wish to solve for the ratio of costs where this is equal to 1:

$$\left[\frac{c_P}{c_A}\right]^{\frac{1}{\theta_\mu - 1}} (\theta_\mu + \theta_\sigma) \left(\frac{\theta_\sigma}{(\theta_\mu + \theta_\sigma - 1)^{\theta_\mu}} \right)^{\dfrac{1}{\theta_\mu - 1}} = 1$$

$$\left[(\theta_\mu + \theta_\sigma) \left(\frac{\theta_\sigma}{(\theta_\mu + \theta_\sigma - 1)^{\theta_\mu}} \right)^{\dfrac{1}{\theta_\mu - 1}} \right]^{\theta_\mu - 1} = \frac{c_A}{c_P}$$

Call this ratio c^*. If $\frac{c_A}{c_P} < c^*$ then the principal will choose to delegate. If not, the principal will conduct search itself. Then we wish to compute comparative statics in c^*. It does not depend on x_μ, x_σ. To prove the proposition, we realize that c^* is equal to the ratio of profits when search costs is the same raised to a power:

$$c^* = \left[\frac{\Pi_{SB}(c)}{\Pi_{FB}(c)} \right]^{\theta_\mu - 1}$$

from the proof of Theorem 2 we have that the ratio of profits with the same search cost is increasing in θ_σ, thus an increasing function of the ratio is also increasing. For the limit, note that the function $f(x) = x^{\theta_\mu - 1}$ when $\theta_\mu > 1$ is continuous for all x, so we can pass the limit through the function and say that:

$$\lim_{\theta_\sigma \to \infty} \left[\frac{\Pi_{SB}(c)}{\Pi_{FB}(c)} \right]^{\theta_\mu - 1} = \left[\lim_{\theta_\sigma \to \infty} \frac{\Pi_{SB}(c)}{\Pi_{FB}(c)} \right]^{\theta_\mu - 1} = 1$$

8.8 Proof of Proposition 3

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

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

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

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

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

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

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 3).

8.9 Proof of Proposition 4

We will use the notations from Figure 2.

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

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

From proposition 3, we can conclude that $p > q$ since $Pr(\mathrm{FB}) < Pr(\mathrm{SB})$.

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

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

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).