Who Gets the Job: A Model of Delegated Recruitment with Multidimensional Applicants

Jacob Kohlhepp and Stepan Aleksanko

September 8, 2020

Abstract

In this paper, we construct a theoretical model of delegated recruitment. In the model, applicants are multidimensional, differing in both expected ability and what we term "ability uncertainty." When the space of contracts is limited to the bonus contract commonly observed in the recruitment industry, the pool of applicants exhibits lower hiring risk then the first-best benchmark, even when recruiters are risk neutral. This implies recruiters pass-over high risk high reward applicants in favor of low risk, low-reward applicants. In a parametric implementation of the model, search efficiency and search intensity increase as the negative correlation between expected ability and hiring risk increases. Although the model is applied to recruiters, it extends well to any situations where search is delegated, the searched item has two dimensions, and there is an initial trial period.

1 Introduction

How people are allocated across firms is a central economic question, key to understanding how individuals escape poverty and governments reduce unemployment. It is also related to a question which captures the attention of every person who needs a paycheck: *how do employers decide who gets the job?* I suggest a novel perspective on this question by taking a step back. Employers only partially decide who gets the job. They delegate the first stage of the process to recruiters and other intermediaries, who build a pool of applicants from which the firm chooses. In this paper, I build a theoretical framework for delegated recruitment and show the typical recruiter contract has profound consequences for the types of applicants which are hired.

My approach is motivated by two primary observations. First, prior to hire, a recruiter or a firm forms beliefs over how well a prospective employee will perform once hired. These beliefs are based, initially, on a paper application, a LinkedIn profile or a resume. In an abstract sense, each possible item on a resume is a noisy signal of ability, and different items vary in their noisiness and in the way they cause a recruiter to update. Thus the whole resume is a package of signals, and variation in what each person has on their resume results in a distribution of possible posterior beliefs. To make this tractable, I assume beliefs can be characterized by a best estimate of the person's ability and an assessment of the uncertainty surrounding this estimate. To put it another way, the belief can be thought of as a forecast and a standard error of the forecast.

To illustrate this framework, consider two applicants applying for a data science position. Applicant A is traditional: they went to a four year college, got a degree in statistics, and interned at a large corporation. Applicant B is nontraditional: they only have a high school degree and are self-taught, but they won a Kaggle competition. Kaggle competitions often feature thousands of competitors, and winning requires a blend of ingenuity and luck. A recruiter might believe Applicant B has a higher expected ability. However, winning a Kaggle competition is a noisy signal, and applicant B is self-taught, so it is hard to say with certainty how B will do in a structured work environment. In comparison, A has a traditional background, and the recruiter has placed this type of candidate before. Thus the uncertainty surrounding A's forecast is lower.

Now for the second observation. There is a common form of recruiter-firm contracts. I interviewed three recruiters, and all reported the typical structure of recruiter pay is a fixed percentage of salary conditional on the employee being hired. They also stated that most recruiters offer some sort of guarantee: if the employee leaves for any reason prior to some agreed upon number of days, the recruiter will either refund the fee or place a new candidate for free. Among the three I interview, two specifically mentioned the period was 90 days. This is consistent with a survey conducted by Top Echelon[1], which found 96% of recruiters offered a guarantee, 86% had a guarantee period between 30 and 90 days. Additionally, 61% provided a replacement but not money back if the candidate failed to stay, while 26% offered a partial or full refund.

In the contracting literature, this payment scheme is called a bonus contract. The recruiters I interviewed also revealed typically commission rates are between 20 and 30 percent, meaning each placement represents a large fraction of a recruiter's total annual income. Since the bonus contract seems to be the norm, the question we ask is whether this contract induces misalignment between the types of applicants (in terms of expected ability and uncertainty) selected by the recruiter and the applicants the firm would ideally like to select.

Returning to Applicants A and B, one can imagine situations where a firm would be willing to take a risk on B. But because the recruiter is the one building the applicant pool, and the contract warps incentives, B may never make it to a hiring manager's desk. This is the sort of situation we are interested in understanding. The way recruiters handle noisy signals has implications for inequality and accessibility. High quality signals

[1] Top Echelon is a company that makes software for recruiters.

of ability are expensive. The cost of data science boot camps is often on the order of $2,000-$17,000 just for a small window 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.[2] They may not even be aware that they qualify: currently only 71% of college applicants file (*How America Pays for College 2020* n.d.).

In this paper, I build out a theoretical model of delegated sequential search, where the firm is restricted to bonus contracts. The model exhibits traditional moral hazard in that the firm cannot observe the search effort of the recruiter. It also exhibits moral hazard over the selection regions. The firm cannot observe what types of applicants the recruiter passes over for the job. Because search is over multidimensional objects, it is possible that two very different acceptance regions can require the same level of search effort, meaning the first form of moral hazard does not nest the second.

The model delivers two main theoretical predictions. First, if the firm is required to use only bonus contracts, the recruiter will over-select on applicant uncertainty compared to the first-best benchmark. In a sense, the recruiter will behave in a risk averse fashion and expend too much search effort finding applicants which are safe-bets. Second, efficiency, search intensity, and the size of bonus payments hinge critically on the correlation between applicant expected ability and hiring risk (uncertainty). Higher levels of correlation between the two induces agreement between the recruiter and the firm over which applicants are acceptable.

The implications of these results for human capital formation, income mobility, and inequality are large. To the extent precise signals of ability are costly, the findings in this paper suggest current contracts may be causing recruiters to distribute opportunity unevenly, by being unwilling to take risks on less traditional applicants. Jobs form the main method through which individuals escape poverty and form human capital, so the human costs of this uneven opportunity can add up quickly.

The results are consistent with the empirical observation that recruiters and hiring managers prefer longer resumes, as noted in Blackburn-Brockman and Belanger (2001). They also have interesting implications for proposals like "ban the box" or anonymous resume laws, as these programs generally increase uncertainty in the applicant pool. The model also provides a pathway for why informal references are so valuable: they greatly increase the precision of information presented on a resume, thus reducing the downside risk of a potential applicant.

Although I build the delegated search model with recruiters in mind, the results and the modeling environment can easily be applied to other contexts. For example, real estate agents search for houses for their clients. A real estate agent probably considers both the expected value of a house to a specific client as

[2] SAT preparation classes, tutoring, college admissions counseling, AP testing, etc.

well as the noisiness of the value, due to different amenities. Once the house is sold, there is usually an escrow period, during which the buyer may call off the sale. Managers of a venture capital firm search for startup companies on behalf of their investors. Some of these firms focus on taking a startup public, which involves preparing for and having an initial public offering. In both of these examples, search is delegated, the object of interest has several dimensions, and there is a waiting period before the searcher gets the payout.

The paper is organized as follows. In the next section I review other work and explain how this research fits within the broader literature. In the model section I explain the primitives of the model and define the restricted class of contracts. In the analysis section I provide some general results without restricting the distributions of expected ability, and then provide comparative statics assuming an explicit distribution. I conclude with a discussion of the policy implications and possible extensions.

2 Literature

Although our model is driven primarily by observations in the recruiting industry, it draws deeply from three strands of the theoretical literature. The first strand is rather recent and niche, and I term it the "delegated search" literature. Ulbricht (2016) derives results for the form of optimal contracts under both moral hazard and adverse selection. The strongest result, and one very relevant for this paper, is that the sequential search problem is such that, even with risk aversion, the first-best outcome can be achieved with moral hazard. This is in stark contract to the typical moral hazard problem, where effort is under-supplied. Ulbricht further proves that it is only under joint moral hazard and adverse selection where the first-best fails to be achieved. Our model features moral hazard, in the sense that the firm cannot observe the number of search or the features of the passed over applicants. It also features the delegation of a sequential search problem. However, I differ in that I restrain the contract shape to be of a bonus form, and I allow for two dimensions of ex-ante heterogeneity. Lewis (2012) explores delegated search with a focus on how contracting impacts agency loss. Crucially, the paper differs from ours because contracting is dynamic, and the objects of search differ only in their profitability. Where Lewis (2012) asks what is possible when all contracts are possible, we ask what is possible when contracts are constrained in a particular way.

The second strand is the more general delegated choice literature. Papers in this literature are characterized by models where an agent is hired to select among a set of options, and the principle has to decide how much discretion to allow. These models are generally not dealing with sequential search, but they work with the idea of permission sets, which inspired our definition of acceptance regions. Armstrong and Vickers (2010) is motivated by a regulator approving mergers, and it considers optimal permission sets that maximize welfare. One of their extensions involves a situation where the agent may invest effort to discover additional options. This extension is quite similar in spirit to our model, except that we consider how restrictions on contract form influences outcomes, while they allow for arbitrary delegation sets. The last closely related paper in this strand is Frankel (2014). In Frankel's model, there is uncertainty over the preferences of the agent, and the agent has private information relevant to the decision. Frankel's idea of alignment, which is when the agent plays as if they shared the principal's preferences, motivates our definition of alignment. The crucial difference between Frankel (2014) and our paper is that the former considers situations where agents differ in their bias, the principal wishes to minimize

The third strand is the multidimensional or multi-product search literature. In our model, applicants are themselves multidimensional, with two attributes which could potentially matter to the searcher. The paper from this literature which is most related to ours is Zhou (2014). Zhou (2014) constructs a consumer search model where consumers are trying to buy two different products and must search across firms. Consumers draw utility pairs for the goods after each search. Like in our case, search strategies are characterized by two-dimensional acceptance sets. Unlike our case, the distribution of the two attributes are assumed to be drawn independently and utility is additive. As a result the analysis presented does not extend to our case. The paper does develop an equilibrium model which is a helpful starting point for considering how our partial equilibrium model could extend.

In addition to the two theoretical strands, our model is motivated by the large literature on labor search and matching models. Early search and matching models incorporated homogeneous workers, but eventually models transitioned to include ex-ante heterogeneity along one dimension. Two papers which represent this stage of the literature well are Shimer and Smith (2000) and Postel–Vinay and Robin (2002). They interpret this ex-ante heterogeneity as "productivity" and "professional ability." More recently, the literature has expanded to consider models with multiple dimensions of heterogeneity. One such example is found in Lindenlaub (2017), where workers differ in both manual and cognitive skills. The idea that people who apply to jobs differ in multiple dimensions is a main idea of this paper, and it was in no small part inspired by the search and matching literature. Although our model does not claim to be general equilibrium, I believe it contributes to the search and matching literature by considering more closely the role of delegation in the matching process. A potential extension of our work, which I discuss in the conclusion, is the incorporation of a representative recruiter in a general equilibrium model of job search and matching.

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. The game is static.

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.[3] a is realized at the end of the game, and the firm exogenously fires the worker if $a < 0$. a is not observed by either the firm or the recruiter prior to hire. Instead, when the recruiter samples an applicant, it views a resume of the applicant, with two characteristics: (μ, σ) which are distributed in the population with joint cdf $G(\mu, \sigma)$ and a continuous joint pdf $g(\mu, \sigma)$. Conditional on observing these two characteristics, the distribution of a for that applicant is $N(\mu, \sigma^2)$.[4] The firm never observes μ, σ, even if the applicant is hired.

Search Technology: The recruiter can take i.i.d. draws of applicants at unit cost c. There 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. I assume there is no recall (this is without loss). Search takes place in a single period, so there is no discounting. The search process is fully private, so the firm does not observe any of the draws of the recruiter or the number of searches.

Payoffs and Contracts: The firm is restricted to contracts of the form: $\alpha + \beta\mathbb{I}\{a > 0\}$. That is, a bonus contract where α is the base fee and β is paid if the employee remains. The firm receives the realized net output a less the payments to the recruiter. The firm receives 0 if an applicant is not hired. The recruiter receives $\alpha + \beta\mathbb{I}\{a > 0\}$ less the search costs. I restrict attention to cases where recruitment is profitable in the first-best. That is, $E[a] = E[\mu] > c$.

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.
- 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, and if $a \leq 0$ the applicant is fired. The contract realizes.

3.1 Model Comments

In the above model, firing is not endogenous, and the firm receives a regardless of the fire decision. To justify this choice, it is best to interpret a as the benefit of hiring an employee less the market wage during an *initial*

[3]Another interpretation is that a is ability relative to some break-even type of worker, where break-even is a worker who produces exactly the market wage.

[4]Another interpretation: the recruiter receives one of many possible noisy signals of ability. Depending on which type of signal they receive, the prior is updated. More noisy signals result in higher variance posteriors.

employment period. During this period, the firm has to bare the cost (or benefit) beyond the market wage, which is exactly a. Once the period ends, if the employee is a net loss during the first period ($a < 0$) the firm either terminates or the employee leaves because the adjusted wage the firm offers is less than some outside option. If $a > 0$, the firm keeps the employee, and the wage is adjusted based on the publicly realized information so that wage and ability are equal. In either case, the additional profit is 0. In summary, our specification of payoffs reflects a situation where there is no commitment on either side.

The two applicant dimensions, μ and σ, can be thought of as quantifying the point estimate of applicant ability 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 and won a competition. It is possible a recruiter sees these two as having the same expected performance, but the second might be perceived as a greater risk. In this sense, σ represents how much information is present in a resume/application. If I assume that recruiter beliefs match reality, then characteristics that drive up μ can be viewed as true productivity boosters, while things that drive down σ can be thought of as signaling boosters. Many things, like university education, likely do both: they raise μ and reduce σ.

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 can also be connected to the literature on discrimination in hiring: recruiters may better understand the background of similar race applicants, or they may simply have a taste-based bias for or against certain applicants. All of these factors are captured by the μ, σ a recruiter assigns an applicant.

The main economic force I wish to capture with this model is the misalignment of the firm and the recruiter's selection strategy along the two dimensions. Misalignment is driven entirely by the restrictions on the contract form.

4 Non-Parametric Analysis

In this section, I analyze the first-best benchmark and the actual equilibrium without imposing additional assumptions on G. Our main object of interest is the acceptance region:

Definition 1 *(Acceptance Region) An acceptance region, denoted $\mathcal{D}_i$, is the set of applicant types (μ, σ) which entity i would select if they operated the search technology.*

This object describes the search strategy of a player, and is similar in spirit to the reservation wage or reservation utility in a uni-dimensional search model. I will also be interested in the following summary statistic.

Definition 2 *(Average Hire Risk) Denote the average hire risk of selection region i as r_i. Define it as:*

$$r_i := E[\sigma|(\mu, \sigma) \in \mathcal{D}_i]$$

Average hire risk quantifies the level of applicant pool risk chosen by entity i. If $r_i > r_j$ entity i selected applicants with on average higher uncertainty (σ). A main question of this paper is the relationship between recruiter average hire risk and the first-best average hire risk.

4.1 First-Best

For the first-best benchmark, I consider the case when the firm can operate the search technology directly.[5] 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:

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

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 care about the uncertainty (σ) dimension of an applicant. It is not worried about the downside risk of an employee because this downside risk is perfectly matched with the option value of picking a "diamond in the rough." Lemma 1 also shows that the optimal threshold μ^* is positive since it is equal to the firms 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).

[5]Equivalently, when there is no contract restriction.

4.2 Firm-Optimal Contract

I 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.

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

$$\alpha + E[U|(\mu,\sigma) \in \mathcal{D}_R] \geq 0 \tag{IR}$$

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

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

The firm is thus solving a traditional contracting problem, where it must choose a search strategy for the recruiter and a compensation scheme. Because search is not observed, the search strategy must be individually rational and incentive compatible given the contract form. Throughout the rest of this paper, we denote Φ to be the standard normal CDF. We can simplify utility of the recruiter after α is sunk as:

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

This simplification shows that the recruiter cares only about maximizing the ratio μ/σ – a candidate's parameter which we will call candidate's standardized ability through this paper. Comparing this to the first-best, 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 $a > 0$ to the firm. This is shown graphically in Figure 1.

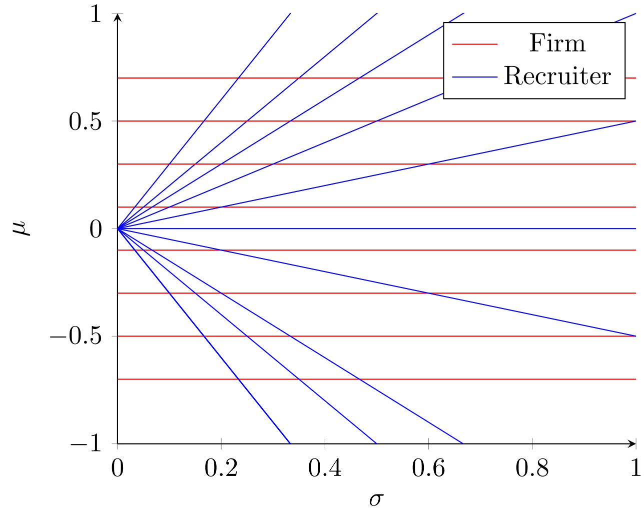

Figure 1: Indifference Curves

Definition 3 *(Standardized ability) Standardized ability of a candidate is the ratio of her expected ability over her ability uncertainty (in other words, how many standard deviations candidate's expected ability is away from zero)*

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

Standardized ability defines the strength of the signal $N(\mu, \sigma)$ about candidate's ability, which is the only thing the recruiter cares about. Larger μ means that the expected ability is higher, although accompanied with a higher σ, this may just decrease the probability of the candidate being a lemon. However, higher standardized ability $\tilde{\mu}$ definitely increases chances of the candidate being accepted to the job, which makes it a parameter of interest for the recruiter.

Further in the paper, we will mostly focus on $(\mu, \tilde{\mu})$ parameters of a candidate since one of them is of interest for the firm and the other is for the recruiter. It is also clear that there is a bijection between the set of candidates defined by (μ, σ) and by $(\mu, \tilde{\mu})$, hence, defining the distribution of workers ability through the joint distribution of (μ, σ) or $(\mu, \tilde{\mu})$ are equivalent. (Also, if the joint pdf of (μ, σ) is continuous, then the acquired pdf of $(\mu, \tilde{\mu})$ is also continuous, and vice versa.)

As we focus on the joint distribution of $(\mu, \tilde{\mu})$, we want to introduce a rather weak assumption about this joint distribution, which is expressed as follows.

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

Intuitively, Assumption 1 means that the larger standardized ability also means the larger expected ability of the candidate. The assumption is quite natural given that $\tilde{\mu} = \mu/\sigma$. To break this relation, the joint distribution of (μ, σ) must have an extremely strong positive association between expected ability and variance of the signal (where with larger μ, σ grows faster than linearly). For the rest of the analysis, this assumption is presumed to be satisfied.

From examining the shape of the recruiter's indifference curves we can see the shape of the acceptance regions. If μ has strictly positive support, 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 scenario.

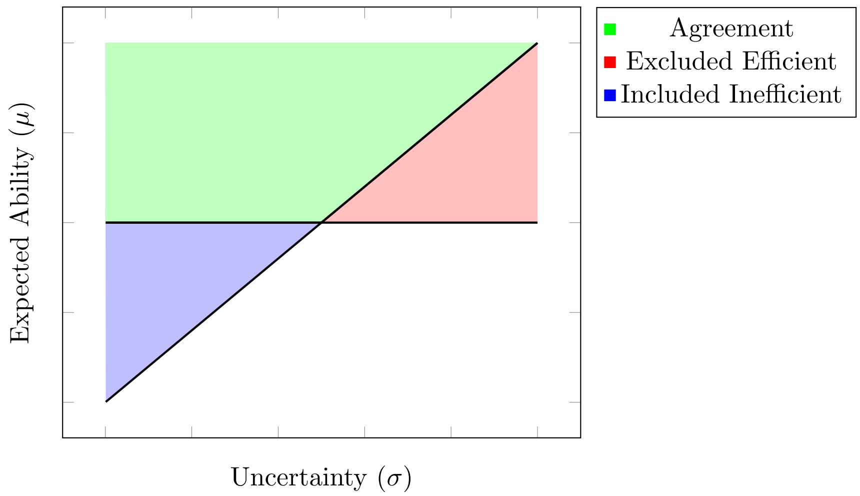

Figure 2: Recruiter vs. Firm Acceptance Regions Over Applicant Types

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 optimally wants to hire, but that are not selected by the recruiter. The blue region represents applicants which the firm would prefer not to hire, but the recruiter selects them. The firm uses β to choose the slope of the diagonal line, trading-off the two regions. We name these regions using the following formal definitions.

Definition 4 *The agreement region is the set of applicants with $\mu, \sigma \in \mathcal{D}_R \cap \mathcal{D}_F$.*

Definition 5 *The excluded efficient region is the set of applicants with $\mu, \sigma \in \mathcal{D}_F - \mathcal{D}_R$.*

Definition 6 *The included inefficient region is the set of applicants with $\mu, \sigma \in \mathcal{D}_R - \mathcal{D}_F$.*

Remark 4.1 *When the support of expected ability 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 "high-potential risks." 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:*

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

where u^* solves:

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

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^*. We next simplify the firm's problem using the previous lemmas.

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

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

F.O.C. below has a unique solution and it defines the firm-optimal contract

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

Then β can be obtained from the recruiter's IC constraint:

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

and α from the recruiter's IR constraint:

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

Corollary 1.1 *Under the firm-optimal contract, the firm's profit is positive and equal to $\mathbb{E}[\mu | \tilde{\mu} = x^*]$, which then also must be positive. Then x^* must be positive too.*

Theorem 1 proves that the problem can be solved by unconstrained, single variable maximization. The objective also makes clear that the firm is maximizing surplus within the constraints imposed by the contract form. The contract determines the slope of the acceptance region and also redistributes all of the rents of search. The first term in the objective is the value of the employee to the firm, and the second term is the expected search cost to find an acceptable employee. Before imposing distributional assumptions, I present several results that hold for broad conditions on the joint distribution G. The first result formalizes the idea that even if the two characteristics are independently distributed across the population, the form of the contract still induces the recruiter to under-select risky applicants relative to the first-best benchmark.

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

Proof. $Pr(\sigma \leq y | \sigma \in D_F) = G_\sigma(y)$ by independence. Then:

$$\begin{aligned} Pr(\sigma \leq y | \sigma \in \mathcal{D}_R) &= \int \mathbb{I}\{x^*\mu \leq y\} + \mathbb{I}\{x^*\mu \geq y\} G_\sigma(y) dG_\mu(\mu) \\ &= G_\mu(y/x^*) + (1 - G_\mu(y/x^*)) G_\sigma(y) \end{aligned}$$

This final term shows that the CDF conditional on the recruiter's acceptance region is a weighted average of 1 and $G_\sigma(y)$ which is always weakly greater than $G_\sigma(y)$. This is equivalent to first-order stochastic dominance of $\sigma | \sigma \in D_R$ by $\sigma || \sigma \in D_F$.$\blacksquare$

This proposition demonstrates that when there is no connection (dependence) between the two applicant attributes, 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, $\sigma, \mu/\sigma$ will often be negatively correlated. To see this, consider the following *example*, which shows that independent attributes can still induce negative affiliation.

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 clearly submodular: if the indicator is 0 for at least one of two pairs of values, it will also be 0 for the pairwise minimum. This is an example where independent μ, σ still results in negative affiliation, a much stronger condition than negative correlation.

Another interesting case is when there is some form of positive dependence between μ, σ. As the example hints at, it is not enough that μ, σ exhibit positive dependence: it must be that this positive dependence is "strong enough" where $\sigma, \mu/\sigma$ exhibit positive dependence. That is, the positive dependence must, in a sense, dominate the negative dependence between $\sigma, 1/\sigma$. This motivates the next proposition.

Proposition 2 *Suppose μ, σ are positively affiliated and $\mu, \mu/\sigma$ are negatively affiliated. Then the recruiter acceptance region is the full support of μ, σ. Also, the firm's acceptance region first-order stochastically dominates the distribution of precision within the recruiter's acceptance region.* [6]

Proof. If $\mu, \mu/\sigma$ are negatively affiliated this implies $E[\mu|\mu/\sigma \geq x]$ is decreasing in x. This implies the entire objective in Lemma 3 is decreasing in x. Thus optimal x will be set to $infmin\{\mu/\sigma|G(\mu,\sigma) > 0\}$. When the support of μ include 0, $x = 0$. Thus $\mathcal{D}_R$ includes the entire support of (μ, σ). To establish that the distribution of σ in the recruiter's acceptance region is first-order stochastically dominated by the distribution in the firm's acceptance region, first note that $\sigma|\sigma \in \mathcal{D}_R$ is equivalent to the unconditional distribution of σ. Next, note that $\sigma|\mu \geq \mu^*$ first-order stochastically dominates the unconditional distribution because μ, σ are positively affiliated.■

Propositions 1 and 2 show two conditions under which the recruiter selects less risky applicants. However, both conditions are rather unnatural. It is unlikely that μ, σ are independent or positively correlated: in general, applicants which appear to be more skilled also present more information or accomplishments in their resumes. Thus in the real world, μ, σ are likely to be negatively associated: more able candidates will tend to have lower uncertainty. Put simply, negative correlation captures the idea that most signals which boost expected ability also reduce uncertainty.

Remark 4.2 *When the two attributes are* ***negatively affiliated****, $\mu, \mu/\sigma$ are* ***positively affiliated****. This implies the objective in Lemma 3 has the potential to be increasing over an initial range, leading to a more interesting problem with an interior solution. However, this also implies that both $\sigma|\sigma \in \mathcal{D}_F$ and $\sigma|\sigma \in \mathcal{D}_R$ will generally first order stochastically dominate the unconditional distribution of σ. Whether one dominates the other is unclear, as it will depend on x^* that solves the problem in Propositions 1 and μ^* which solves the first-best problem in Lemma 1.*

What these propositions and the remark should make clear is that the relationship between $\mu, \mu/\sigma$ greatly impacts the acceptance regions. This idea is why the next section specifies the joint distribution of $\mu, \mu/\sigma$ rather than μ, σ.

5 Parametric Analysis

In order to compare the first and second-best regions, and perform comparative statics, I impose a parametric assumption. For this problem, it is much simpler to specify a joint distribution for $\mu, \tilde{\mu}$ rather than μ, σ directly. This avoids needing to apply transformation formulas when deriving results, and it highlights that it

[6] Note that it is necessary for μ, σ to not be negatively affiliated for the condition to hold.

is the relationship between $\mu, \tilde{\mu}$ that is most important. I then assume throughout that the below assumption holds.

Assumption 2 *(DIST) The joint cumulative distribution function of* $(\mu, \tilde{\mu})$ *is given by:*

$$H_{\mu,\tilde{\mu}}(\mu, \tilde{\mu}) = (1 - e^{-\lambda_1 \mu})(1 - e^{-\lambda_2 \tilde{\mu}})[1 + \rho e^{-\lambda_1 \mu - \lambda_2 \tilde{\mu}}]$$

The distribution in Assumption 1 is often referred to as a bivariate Gumbel-exponential distribution. It was first introduced in Gumbel (1960). For my purposes it has several desirable properties. First, it has positive support. Second, positive ρ induces a positive correlation between $\mu, \mu/\sigma$, negative ρ allows for negative correlation and $\rho = 0$ implies independence. Third, the marginal distributions of $(\mu, \mu/\sigma)$ are $exponential(\lambda_1), exponential(\lambda_2)$ respectively. Thus, changing λ_1, λ_2 allows us to easily introduce first-order stochastic shifts in the distributions. Different λ_1 values can be used to represent different industries or companies with different distributions of talent. Lower λ_1 implies a higher expected ability from a randomly drawn applicant. Different λ_2 values can represent situations with different types of applicant uncertainty. Lower λ_2 implies a higher expected standardized applicant ability. Holding the distribution of ability constant, this captures less aggregate applicant uncertainty.

Remark 5.1 *This assumption is about the joint distribution of* $\mu, \tilde{\mu}$. *To see how this connects to the joint distribution of* $\mu, 1/\sigma$, *note that:*

$$E[\sigma^{-1}|\mu] = E[\tilde{\mu}/\mu|\mu] = \mu^{-1}E[\tilde{\mu}|\mu] = \frac{1 + \rho(1/2 - e^{-\lambda_1 \mu})}{\mu \lambda_2}$$

$1/\sigma$ *is the precision of ability signal. Next note that when* $e^{-\lambda_1 \mu} = 1/2$, μ *is the median. For any* μ *greater than the median* μ, *higher* ρ *increases the conditional expectation of precision. For any* μ *less than the median, higher* ρ *reduces the conditional expectation of precision. Intuitively, higher* ρ *makes high ability individuals have more precise signals. Thus* ρ *also measures the association between* $\mu, 1/\sigma$.

Assumption 1 yields closed-form solutions to both the first-best and equilibrium problems. These are presented in the following proposition and theorem.

Proposition 3 *Under Assumption 1, the first-best acceptance region is given by:*

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

where has the closed-form solution:

$$\mu^* := -\frac{log(c\lambda_1)}{\lambda_1}$$

Proof. From Lemma 1, we know the general form of the acceptance region, what remains is to find μ^*. Next, note that under Assumption 1, the marginal distribution of μ is exponential, so the equation characterizing μ^* from Lemma 1 can be re-written as:

$$c = \int_{\mu^*}^{\infty} e^{-\lambda_1 \mu} d\mu = \frac{e^{-\lambda_1 \mu^*}}{\lambda_1}$$

Re-arrangement yields:

$$\mu^* = -\frac{log(c\lambda_1)}{\lambda_1}$$

which is the result. Note that the first-best solution does not depend on ρ.■

The Proposition shows that the size of the acceptance region, first-best profit, and first-best search intensity depend only on search cost (c) and the mean of applicant expected ability. When search costs are higher, the acceptance region increases in probability, indicating search intensity falls. When average expected ability increases the acceptance region rises in probability, even as μ^* rises. In this sense, improvement in applicant quality outpaces the rising standard for hire. I now turn to the equilibrium solution.

Theorem 2 *Under Assumption 1, the firm's problem has a unique solution with the following characteristics.*

1. ***Acceptance Region:***

$$\mathcal{D}_R = \{\mu, \sigma | \mu/\sigma \geq x^*\}$$

2. ***Fixed Payment*** α***:***

$$\alpha = -c\left\{\frac{e^{\lambda_2^2/2}[1-\Phi(x^*+\lambda_2)]}{\Phi(x^*)}\right\}^{-1}$$

3. ***Bonus Payment*** β***:***

$$\beta = c\left\{e^{\lambda_2^2/2}[1-\Phi(x^*+\lambda_2)]\right\}^{-1}$$

where

$$x^* := \begin{cases} \frac{1}{2\lambda_2} log\left(\frac{\rho}{2\lambda_1 c}\right) \; if \; \frac{\rho\lambda_2}{2\lambda_1 c} > 1 \\ 0 \; else \end{cases}$$

The Theorem can be proved by using Proposition 1 applied to Assumption 1. The full proof is provided in the Appendix because the steps are mainly algebraic with little economic insight. Recall that x^* is the slope of the lowest indifference curve within the recruiter's acceptance region. As ρ rises, x^* clearly rises, meaning the slope is increasing-expected utility from search after α is sunk is increasing. These closed-form solutions allow for several interesting comparative statics. Before proceeding, I introduce a useful measure of agency loss.

Definition 7 *Agency loss is defined as the difference between firm profit in the first-best benchmark and firm profit in equilibrium.*

Notice that because profit is equal to social surplus, this is simply the difference in social surplus under the two environments. First, I consider how agency loss due to delegation changes with ρ, the correlation between $\mu, \mu/\sigma$.

Theorem 3 *Under Assumption 1, agency loss declines with ρ. Additionally, incentives become stronger as ρ rises, in the sense that:*

1. *The equilibrium expected number of searches increases.*

2. *Equilibrium bonus payment β increases.*

Proof. Recall that agency loss is profit in the first-best less profit in equilibrium. ρ does not appear in first-best profit (because ρ impacts the joint distribution but not the marginal distributions), so it is only necessary to understand how equilibrium profit changes with ρ. Equilibrium profit is:

$$\Pi^*(\rho, x^*) = E[\mu|\tilde{\mu} \geq x^*] - \frac{c}{Pr(\tilde{\mu} \geq x^*)}$$

Under Assumption 1, the proof of Theorem 1 in the Appendix derives the conditional density and conditional expectation. Thus profit can be written explicitly as:

$$\Pi^*(\rho, x^*) = \lambda_1^{-1}(1 + \rho/2 - \rho/2e^{-\lambda_2 x^*}) - ce^{\lambda_2 x^*}$$

Using the Envelope Theorem, we have that:

$$\frac{d\Pi(x^*, \rho)}{d\rho} = \frac{1}{2}(1 - e^{-\lambda_2 x^*}) \geq 0$$

thus equilibrium profit is increasing in ρ implying agency loss is decreasing in ρ. Next recall that the expected number of searches is just $1/Pr(\tilde{\mu} \geq x^*)$. This can be expressed as $\left(\frac{2\lambda_1 c}{\rho\lambda_2}\right)^{-1/2}$ which is clearly increasing in ρ. Finally, from Theorem 1 β is increasing in ρ if x^* is increasing in ρ, which can be seen from inspection. ■

Because the firm extracts all surplus, Theorem 2 has a further implication: search becomes more efficient as ρ increases. Thus the model predicts delegated recruitment is more attractive in occupations where the candidates with the highest expected ability also tend to have the lowest uncertainty. This appears quite realistic: for entry-level positions there tends to be more equality in the level of uncertainty across candidates. Put another way, uncertainty is more independent of expected ability. However, for higher-level positions, better applicants tend to be more "tried and true"; that is uncertainty tends to fall more with expected ability. In this situation the model predicts firms will opt to delegate to recruiters more frequently for higher-level positions.

With this explicit solution we can understand how the sizes of the regions depicted in Figure 2 change with ρ.

Theorem 4 *Under Assumption 1, when ρ increases:*

1. *The* ***Recruiter Acceptance Region*** *decreases in probability.*
2. *The* ***Agreement Region*** *decreases in probability.*
3. *The* ***Included Inefficient Region*** *decreases in probability.*
4. *The* ***Excluded Efficient Region*** *increases in probability.*

The proof is provided in the Appendix. When ρ is negative, the conditions for Proposition 2 are met, and the solution is to set x^* to 0, which makes the acceptance region the entire support of (μ, σ). Once ρ is positive, ρ rising results in better alignment between the firm and recruiter, corresponding to an increase in the agreement region. The acceptance region shrinks in probability (the recruiter becomes more selective), which means that the inflow of excluded efficient applicant types is dominated by the outflow of included inefficient types. Efficiency improvements come more from not hiring the wrong people rather than from hiring more of the right people. If we interpret ρ as measuring industry correlation between $\mu, \tilde{\mu}$, the model predicts industries with higher correlation will have more selective recruiters. They will also favor applicants with

6 Discussion

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

7 Conclusion

In the course of this paper, I outline a new theoretical framework of delegated recruitment. Importantly, I explore the misalignment that occurs as a result of multidimensional job applicants and a constrained contract space. I characterize the problem in the general case, and provide some results without specifying a distribution. I then impose parametric assumptions in order to compute comparative statics. Under the parametric assumptions, I show how bonus payments respond to changes in search cost, correlation, and applicant characteristic distributions. I also highlight that it is the correlation between expected ability and expected ability divided by uncertainty which controls much of the dynamics. Specifically, when there is a high level of negative correlation between ability and uncertainty, this induces positive correlation between expected ability and expected ability divided by uncertainty. This generates agreement over which types of job applicants are acceptable, which in turn raises search intensity and efficiency.

The subject of intermediaries in the labor market has received little attention in both the theoretical and applied literature, and this paper is an attempt to build a cogent theory that captures the trade-offs of delegation in this environment. 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. This paper makes clear that job training programs must focus on more than just providing skills. They must be designed to minimize the hiring risk surrounding perspective job seekers, otherwise intermediary effects may undercut effectiveness.

Add section about extending to a general equilibrium model.

References

Armstrong, Mark and John Vickers (2010). "A model of delegated project choice". In: *Econometrica* 78.1, pp. 213–244.

Blackburn-Brockman, Elizabeth and Kelly Belanger (2001). "One page or two?: A national study of CPA recruiters' preferences for resume length". In: *The Journal of Business Communication (1973)* 38.1, pp. 29–45.

Frankel, Alexander (2014). "Aligned delegation". In: *American Economic Review* 104.1, pp. 66–83.

Gumbel, Emil J (1960). "Bivariate exponential distributions". In: *Journal of the American Statistical Association* 55.292, pp. 698–707.

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

Lewis, Tracy R (2012). "A theory of delegated search for the best alternative". In: *The RAND Journal of Economics* 43.3, pp. 391–416.

Lindenlaub, Ilse (2017). "Sorting multidimensional types: Theory and application". In: *The Review of Economic Studies* 84.2, pp. 718–789.

Postel–Vinay, Fabien and Jean–Marc Robin (2002). "Equilibrium wage dispersion with worker and employer heterogeneity". In: *Econometrica* 70.6, pp. 2295–2350.

Shimer, Robert and Lones Smith (2000). "Assortative matching and search". In: *Econometrica* 68.2, pp. 343–369.

Ulbricht, Robert (2016). "Optimal delegated search with adverse selection and moral hazard". In: *Theoretical Economics* 11.1, pp. 253–278.

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

Zhou, Jidong (2014). "Multiproduct search and the joint search effect". In: *American Economic Review* 104.9, pp. 2918–39.

8 Appendix

8.1 Proof of Lemma 1

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 firm is given by:

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

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

Re-writing, this yields:

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

Observe that utility only depends on μ/σ, so we can reduce the problem to one-dimensional search. Define the random variable $u = \beta \Phi(\mu/\sigma)$, and denote its CDF as M. Denote the new value function of the ratio as $\tilde{U}$. Then the above becomes:

$$0 = -c + \int \max\{u - \tilde{U}, 0\} dM(u)$$

The optimal strategy is a reservation rule (but in the ratio μ/σ). Denote reservation ratio as u^*, then by definition $\tilde{U} = u^*$, so:

$$c = \int \max\{u - \tilde{U}, 0\} dM(q) \leftrightarrow c = \int_{\mu > \mu^*} \mu - \mu^* dF(\mu)$$

Integration by parts gives:

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

Since the first term is 0, this simplifies to:

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

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

$$c = \int_{u > u^*} udM(u) - (1 - M(u^*))u^* \leftrightarrow u^* = \frac{1}{1 - M(u^*)} \left(\int_{u > u^*} udM(u) - c \right)$$

which can compactly be re-written as:

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

8.3 Proof of Proposition 1

We apply Lemma 2 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 $x := \Phi^{-1}(u^*/\beta)$. 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 Theorem 1

Given two variables X, Y distributed according to the distribution given in Assumption 1, we can derive the PDF:

$$h(x, y) = \lambda_1 \lambda_2 e^{-\lambda_1 x - \lambda_2 y}[1 + \rho(2e^{-\lambda_1 x} - 1)(2e^{-\lambda_2 y} - 1)]$$

And then the conditional density:

$$h(x|y) = \lambda_1 e^{-\lambda_1 x}[1 + \rho(2e^{-\lambda_1 x} - 1)(2e^{-\lambda_2 y} - 1)]$$

$$h(x|y) = \lambda_1 e^{-\lambda_1 x} + \lambda_1 \rho(2e^{-2\lambda_1 x} - e^{-\lambda_1 x})(2e^{-\lambda_2 y} - 1)]$$

$$h(x|y) = \lambda_1 \Big(e^{-\lambda_1 x}(1+\rho-2\rho e^{-\lambda_2 y}) - 2\rho e^{-2\lambda_1 x}(1-2e^{-\lambda_2 y}) \Big)$$

Then integrate this density to get the conditional expectation:

$$E[X|Y] = \lambda_1^{-1}(1+\rho/2-\rho e^{-\lambda_2 y})$$

To derive the conditional expectation that is relevant to the objective, apply Law of Iterated Expectations to $E[X|Y]$:

$$E[X|Y>y] = E[E[X|Y]|Y>y] = e^{\lambda_2 y}\int_0^\infty (\lambda_1^{-1}(1+\rho/2-\rho e^{-\lambda_2 y})\lambda_2 e^{-\lambda_2 y} dy = \lambda_1^{-1}\Big(1+\rho/2-\rho/2e^{-\lambda_2 y}\Big)$$

Also note that Y has a marginal exponential distribution, thus:

$$Pr(Y>y) = e^{-\lambda_2 y}$$

Applying these results to $(\mu, \mu/\sigma)$ we have that:

$$Pr(\mu/\sigma > x) = e^{-\lambda_2 x}$$

$$E[\mu|\mu/\sigma > x] = \lambda_1^{-1}\Big(1+\rho/2-\rho/2e^{-\lambda_2 x}\Big)$$

From Lemma 3, the problem is characterized in terms of a single choice variable, x. We can now make the problem explicit:

$$\max_x \lambda_1^{-1}\Big(1+\rho/2-\rho/2e^{-\lambda_2 x}\Big) - ce^{\lambda_2 x}$$

The first derivative is:

$$\frac{\rho\lambda_2}{2\lambda_1}e^{-\lambda_2 x} - c\lambda_2 e^{\lambda_2 x}$$

For $\rho/(2\lambda_1) < 1$ the derivative is negative for all positive x, meaning that the objective is strictly decreasing, and thus the solution is to set $x^* = 0$. Otherwise the function is positive then negative, implying the objective has a global maximum. Taking the derivative again yields:

$$-\frac{\rho\lambda_2^2}{2\lambda_1}e^{-\lambda_2 x} - c\lambda_2^2 e^{\lambda_2 x}$$

For $\rho \geq 0$, this is negative, and the function is concave, thus we can use the first-order condition to derive x^* when $\rho > 0$:

$$x^*(\rho, c, \lambda_1, \lambda_2) = \frac{1}{2\lambda_2} log\left(\frac{\rho}{2\lambda_1 c}\right)$$

To derive the other choice variables from x^*, we must first find u^*, reservation utility from search. This can be had from the following two questions:

$$c = \int_{x^*}^{\infty} (\beta\Phi(y) - u^*)\lambda_2 exp(-\lambda_2 y)dy \tag{1}$$

$$\beta\Phi(x^*) = u^* \tag{2}$$

Solving for β from 2 and substituting this into 1, we perform the following manipulations:

$$\begin{aligned}
c &= \int_{x^*}^{\infty} (u^* \frac{\Phi(y)}{\Phi(x^*)} - u^*)\lambda_2 exp(-\lambda_2 y)dy \\
&= -u^* e^{-\lambda_2 x^*} + \frac{u^*}{\Phi(x^*)} \int_{x^*}^{\infty} \Phi(y)\lambda_2 exp(-\lambda_2 y)dy \\
&= -u^* e^{-\lambda_2 x^*} + \frac{u^*}{\Phi(x^*)} \left\{ (-\Phi(y)e^{-\lambda_2 y}|_{x^*}^{\infty} + \int_{x^*}^{\infty} \phi(y)e^{-\lambda_2 y}dy \right\} \\
&= -u^* e^{-\lambda_2 x^*} + \frac{u^*}{\Phi(x^*)} \left\{ \Phi(x^*)e^{-\lambda_2 x^*} + \int_{x^*}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2 - \lambda_2 y}dy \right\} \\
&= -u^* e^{-\lambda_2 x^*} + \frac{u^*}{\Phi(x^*)} \left\{ \Phi(x^*)e^{-\lambda_2 x^*} + e^{\lambda_2^2/2} \int_{x^*}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(y+\lambda_2)^2/2}dy \right\} \\
&= -u^* e^{-\lambda_2 x^*} + \frac{u^*}{\Phi(x^*)} \left\{ \Phi(x^*)e^{-\lambda_2 x^*} + e^{\lambda_2^2/2}[1 - \Phi(x^* + \lambda_2)] \right\}
\end{aligned}$$

$$u^* = c\left\{ \frac{e^{\lambda_2^2/2}[1 - \Phi(x^* + \lambda_2)]}{\Phi(x^*)} \right\}^{-1}$$

The binding IR constraint implies $\alpha = -u^*$ (by the argument presented in the proof of Lemma 3). Thus:

$$\alpha = -c\left\{ \frac{e^{\lambda_2^2/2}[1 - \Phi(x^* + \lambda_2)]}{\Phi(x^*)} \right\}^{-1}$$

Finally, using equation 2:

$$\beta = c\left\{ e^{\lambda_2^2/2}[1 - \Phi(x^* + \lambda_2)] \right\}^{-1}$$

8.5 Proof of Corollary 4.2

The results simply require differentiating the expressions in Theorem 1:

$$\frac{\partial \beta}{\partial x^*} = \frac{c}{e^{\lambda_2^2/2}} \frac{\phi(x^* + \lambda_2}{(1 - \Phi(x^* + \lambda_2))^2}$$

$$\frac{\partial x^*}{\partial \rho} = \frac{1}{2\lambda_2 \rho}$$

$$\frac{\partial \beta}{\partial \rho} = \frac{\partial \beta}{\partial x^*} \frac{\partial x^*}{\partial \rho} = \frac{ce^{\lambda_2^2/2}}{(1 - \Phi(x^* + \lambda_2))^2} \phi(x^* + \lambda_2) \frac{1}{2\lambda_2 \rho}$$

This is clearly positive.

$$\frac{\partial x^*}{\partial c} = \frac{-1}{2\lambda_2 c}$$

$$\begin{aligned}
\frac{d\beta(x^*, c)}{dc} &= \frac{\partial \beta(x^*, c)}{\partial c} + \frac{\partial \beta}{\partial x^*} \frac{\partial x^*}{\partial c} \\
&= \{e^{\lambda_2^2/2}[1 - \Phi(x^* + \lambda_2)]\}^{-1} - \frac{c}{e^{\lambda_2^2/2}} \frac{\phi(x^* + \lambda_2)}{(1 - \Phi(x^* + \lambda_2))^2} \frac{1}{2\lambda_2 c} \\
&= \{e^{\lambda_2^2/2}[1 - \Phi(x^* + \lambda_2)]\}^{-1} \left(1 - \frac{\phi(x^* + \lambda_2)}{2\lambda_2(1 - \Phi(x^* + \lambda_2))}\right)
\end{aligned}$$

The outer term is positive, so this derivative's sign matches the sign of the term in the large parentheses. This inner term is a function of the normal hazard ratio, which is in turn a function of x^*. x^* is decreasing in c, and the normal hazard ratio is an increasing function over positive numbers. As a result, the whole inner term will become positive for high enough c.

$$\frac{\partial x^*}{\partial \lambda_1} = -\frac{1}{2\lambda_2 \lambda_1}$$

$$\frac{\partial \beta}{\partial \lambda_1} = \frac{\partial \beta}{\partial x^*} \frac{\partial x^*}{\partial \lambda_1} = -\frac{1}{2\lambda_2 \lambda_1} \frac{c}{e^{\lambda_2^2/2}} \frac{\phi(x^* + \lambda_2)}{(1 - \Phi(x^* + \lambda_2))^2}$$

This is always negative. ■

8.6 Proof of Theorem 3

Evaluating the joint CDF:

$$H(\mu^*, x^*) = (1 - e^{-\lambda_1 \mu^*})(1 - e^{-1/2log(\frac{\rho\lambda_2}{2\lambda_1 c})}(1 + \rho e^{-\lambda_1 \mu^*} e^{-1/2log(\frac{\rho\lambda_2}{2\lambda_1 c})})$$

which simplifies to:

$$H(\mu^*, x^*) = (1 - e^{-\lambda_1\mu^*})\left(1 - \left(\frac{\rho\lambda_2}{2\lambda_1 c}\right)^{-1/2}\right)\left[1 + \rho^{1/2}e^{-\lambda_1\mu^*}\left(\frac{\lambda_2}{2\lambda_1 c}\right)^{-1/2}\right]$$

We use these in the below expansions. Notice that the intersection of the acceptance regions is given by:

$$\mathcal{D}_R \cap \mathcal{D}_F = \{\mu, \sigma | \mu \geq \mu^* \& \mu/\sigma \geq x^*\}$$

The probability of the intersection can be re-written as:

$$\begin{aligned}
Pr(\mu \geq \mu^* \& \mu/\sigma \geq x^*) &= 1 - \{Pr(\mu \leq \mu^*) + Pr(\mu/\sigma \leq x^*) - Pr(\mu \leq \mu^* \& \mu/\sigma \leq x^*)\} \\
&= 1 - \{1 - c\lambda_1 + 1 - \left(\frac{2\lambda_1 c}{\rho}\right)^{1/2} - H(\mu^*, x^*)\} \\
&= c\lambda_1 - \left(1 - \left(\frac{2\lambda_1 c}{\rho}\right)^{1/2}\right)\left\{1 - (1 - e^{-\lambda_1\mu^*})\left[1 + \rho^{1/2}e^{-\lambda_1\mu^*}\left(\frac{1}{2\lambda_1 c}\right)^{-1/2}\right]\right\} \\
&= c\lambda_1 - \left(1 - \left(\frac{2\lambda_1 c}{\rho}\right)^{1/2}\right)\left\{c\lambda_1 - (1 - c\lambda_1)\left[\rho^{1/2}\left(\frac{1}{2\lambda_1^3 c^3}\right)^{-1/2}\right]\right\} \\
&= \left(\frac{2\lambda_1^3 c^3}{\rho}\right)^{1/2} + \left(1 - \left(\frac{2\lambda_1 c}{\rho}\right)^{1/2}\right)\left\{c\lambda_1 - (1 - c\lambda_1)\left[\rho^{1/2}\left(\frac{1}{2\lambda_1^3 c^3}\right)^{-1/2}\right]\right\} \\
&= c\lambda_1 - \left(1 - \left(\frac{2\lambda_1 c}{\rho}\right)^{1/2}\right)(1 - c\lambda_1)\left[\rho^{1/2}\left(\frac{1}{2\lambda_1^3 c^3}\right)^{-1/2}\right] \\
&= c\lambda_1 - (1 - c\lambda_1)\left[\rho^{1/2}\left(\frac{1}{2\lambda_1^3 c^3}\right)^{-1/2}\right] - 2\lambda_1^2 c^2(1 - c\lambda_1)
\end{aligned}$$

This function is decreasing in ρ. Relative agreement.

$$Pr(\mu \geq \mu^* \& \mu/\sigma \geq x^*)/Pr(\mu/\sigma \geq x^*) = \left(\frac{c\lambda_1\rho}{2}\right)^{1/2} - (1 - c\lambda_1)\frac{2\rho\lambda_1 c}{\lambda_2^{1/2}} - \rho^{1/2}2^{1/2}\lambda_1^{3/2}c^{3/2}(1 - c\lambda_1)$$

Next, we can compute the measure of applicant types which are in the first-best acceptance region but not the recruiter's. That is:

$$Pr(\mu \geq \mu^* \& \mu/\sigma \leq x^*) = Pr(\mu \geq \mu^*) - Pr(\mu \geq \mu^* \& \mu/\sigma \geq x^*)$$

The first term does not depend on ρ. The second term we showed was decreasing previously. As a result, the exclude efficient region is increasing in ρ. Finally we compute the measure of applicants which are in the recruiter region but not in the first-best region.

$$Pr(\mu/\sigma \geq x^* \& \mu \leq \mu^*) = \left(\frac{2\lambda_1 c}{\rho}\right)^{1/2} - c\lambda_1 + (1 - c\lambda_1)\left[\rho^{1/2}\left(\frac{1}{2\lambda_1^3 c^3}\right)^{-1/2}\right] + 2\lambda_1^2 c^2 (1 - c\lambda_1)$$

Taking derivatives:

$$\begin{aligned}\frac{\partial Pr(\mu/\sigma \geq x^* \& \mu \leq \mu^*)}{\partial \rho} &= -\left(\frac{\lambda_1 c}{2\rho^3}\right)^{1/2} + (1 - c\lambda_1)\left(\frac{\lambda_1^3 c^3}{2\rho}\right)^{1/2} \\ &= \left(\frac{\lambda_1 c}{2\rho}\right)^{1/2}\left[(1 - c\lambda_1)\lambda_1 c - \frac{1}{\rho}\right]\end{aligned}$$

Note that the first term inside the square brackets is the multiplication of two probabilities, and is thus less than 1. Meanwhile the second term is weakly greater than 1. Thus The derivative is negative.

9 Side Note

$$\begin{aligned}Pr(\mu \geq \mu^* \& \mu/\sigma \geq x^*) &= 1 - \{Pr(\mu \leq \mu^*) + Pr(\mu/\sigma \leq x^*) - Pr(\mu \leq \mu^* \& \mu/\sigma \leq x^*)\} \\ &= 1 - (G_\mu(\mu^*) + G_{\tilde{\mu}}(x^*) - G(\mu.\tilde{\mu})) \\ &= 1 - G_\mu(\mu^*) - G_{\tilde{\mu}}(x^*) + G_\mu(\mu^* | \tilde{\mu} \leq \mu^*) G_{\tilde{\mu}}(x^*)\end{aligned}$$