Estimating labor market power from job applications*

Pinjas Albagli†

September 24, 2025

[Latest version here]

Abstract

A growing literature on monopsony in labor markets emphasizes idiosyncratic preferences over differentiated jobs as a key source of market power, borrowing tools from industrial organization to estimate firm-level labor supply elasticities. While promising, this discrete choice approach —when applied to job applications data— typically assumes that each job seeker applies to only one job. This assumption is at odds with observed behavior and overlooks how a wage increase affects the supply of applications not only through substitution across jobs, but also through its impact on the number of applications submitted. This paper relaxes that assumption by extending the standard framework to allow for multiple applications in a simultaneous search environment, where uncertainty about job offers induces multiple-application behavior.

1 Introduction

This is not really an introduction. I just moved some incomplete paragraphs from the model section to the introduction because I think these are things that should be discussed here.

Mirroring the case of product markets, the measurement and estimation of labor market power is usually based on the identification of wage markdowns measuring the wedge between wages and the marginal revenue product of labor that results from imperfect competition in the labor market. While a part of the literature follows a direct approach to the estimation of wage markdowns

*[Acknowledgements]

†London School of Economics. Email: p.m.albagli@lse.ac.uk .

leveraging production function estimation techniques and revenue data (e.g., Brooks et al., 2021; Yeh et al., 2022; Mertens and Mottironi, 2023), this paper follows the tradition of estimating firm-level labor supply elasticities (e.g., Dal Bó et al., 2013; Azar et al., 2022; Roussille and Scuderi, 2025).[1] As discussed by Manning (2003, 2021), the key idea behind monopsony is that the labor supply curve to an individual employer is less than perfectly elastic. The markdown at the firm level is a function —typically the reciprocal— of this elasticity in a broad class of models of monopsony power (Azar and Marinescu, 2024)...

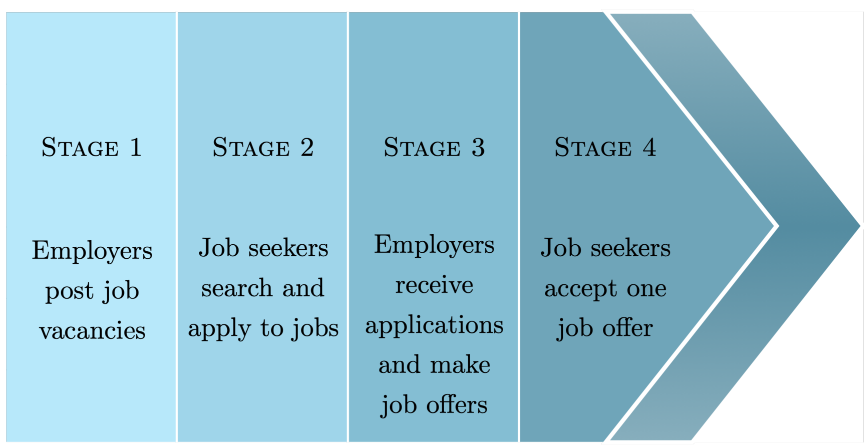
Card et al. (2018) and, more recently, Card (2022) advocate for the adoption of the industrial organization tradition of estimating discrete choice models of demand for differentiated products...

incomplete paragraph

As discussed by Azar and Marinescu (2024), the setups of different articles estimating the firm-level labour supply elasticity focus on different parts of the process that determines the firm's level of employment. While it is important to note how this implies the need for a further transformation of the estimates into labour supply elasticities, of equal importance is to note that models that appropriately describe one part of the process might be inappropriate for another. INCLUDE DIAGRAM! CITE AZAR, BERRY, MARINESCU (2022)! CITE HIRSCH ET AL (2022)!...

incomplete paragraph

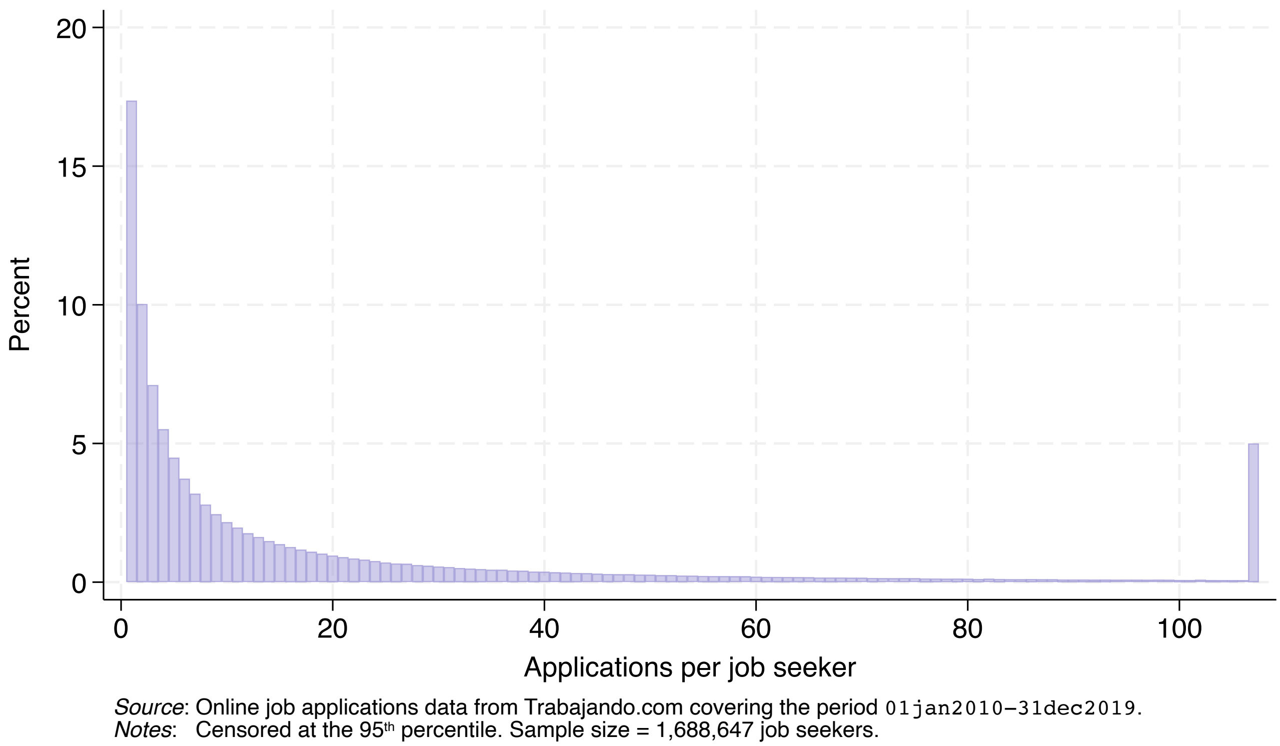
Figure 1: Timing of the recruitment process



Notes: Timeline of the recruitment process in a stylized labor market. Traditional discrete choice models that allow at most one job to be chosen are well-suited for stage 4, where workers decide among final job offers (see, e.g., Hirsch et al., 2022). However, when applied to stage 3, as in Azar et al. (2022), these models miss a key feature of the economic environment: under job-offer uncertainty and costly applications, job seekers optimally apply to multiple vacancies. Our framework captures this behavior.

[1]See Manning (2021) and Azar and Marinescu (2024) for an overview of both strands of the literature.

Figure 2: Histogram of the number of applications per job seeker



Notes: Histogram of the number of applications per job seeker on job board Trabajando.com over the period January 1, 2010 to December 31, 2019. Censored at the 95^{th} percentile. Sample size = 1,688,647 job seekers. DISCUSS IDENTIFICATION/DEFINITION OF SEARCH SPELLS! Multiple applications are pervasive, motivating our portfolio-choice framework and distinguishing our approach from single-application models

Need to re-do this graph, incorporating Alessandro's feedback: It needs to be clear that it is not a mechanical effect of multiple spells per job seeker/a timing issue. Maybe use the same sampling period as the empirical application (a shorter period would alleviate the concern).
Also need to improve the visuals. For example, remove the Stata note and change the font.

Even if a one-application model recovers mean utilities under revealed preference, its wage elasticities need not equal those from a portfolio model. In the portfolio model, a wage increase affects (i) the probability a firm enters an applicant's set holding the set size fixed, and (ii) the size of the set itself via the thresholds determined by $(\alpha_i, \gamma - i)$. The one-application model sets (ii) to zero by construction and forces all adjustments to be replacement within a singleton choice. Unless the data are such that $n_i = \in \{0, 1\}$ almost surely, this restriction induces a different elasticity mapping from $(\boldsymbol{\delta}, \beta)$ to applications. Thus, even with correctly estimated $\boldsymbol{\delta}$, one-application elasticities are generally misspecified when applicants send multiple applications.

Have a paragraph in the spirit of the above. Mentioning yet-undefined parameters might be too much, but the argument should appear here in the introduction and also in Section 2.3 as an implication. This may require discussing estimation and identification before I had planned.

2 A job differentiation model of labor supply

Consider a labor market where a finite set of firms $f \in \mathcal{F}$ each post a finite set $\mathcal{J}^f$ of job vacancies. A finite set $\mathcal{I}$ of job seekers, with size $I \equiv |\mathcal{I}|$ decide where to apply among the $J \equiv |\mathcal{J}|$ vacancies in the common choice set $\mathcal{J} \equiv \bigcup_{f \in \mathcal{F}} \mathcal{J}^f$. Each vacancy $j \in \mathcal{J}$ is fully characterized by an offered wage $w_j > 0$, a (column) vector of job characteristics $\boldsymbol{x}_j \in \mathbb{R}^K$ that we will assume observed by the econometrician when discussing estimation in Sections 2.3 and 4.2, and a scalar index ξ_j capturing other job characteristics that we will assume unobserved. Job characteristics $(\boldsymbol{x}_j', \xi_j)'$ are fixed at this stage of the recruitment process, and firms compete in wages to attract workers. Since our primary object of interest is the wage elasticity of the supply of job applications to the firm, we abstract away as much as possible from modeling the demand side of the market and market equilibrium.

Need to add more intuitive introduction and a bit more detail

2.1 Risky discrete choice and the job application portfolio problem

Consider the simultaneous search setting studied by Chade and Smith (2006), where each decision maker solves a static portfolio choice problem. Job seeker i faces a finite set $\mathcal{J}$ consisting of $J \equiv |\mathcal{J}|$ job vacancy advertisements and chooses a subset $A_i \subseteq \mathcal{J}$ of vacancies to apply to. The cost of applications, $c_i(n_i)$, depends only on the number of applications, $n_i \equiv |A_i|$, where $c_i : \mathbb{N} \to \mathbb{R}_+$ is increasing and convex with $c_i(0) = 0$. Conditional on applying, the job seeker gets an offer from job j with probability $\alpha_{ij} \in (0, 1]$. Recruitment decisions are independent in the sense that the events $\{j \text{ makes an offer to } i \mid i \text{ applied to } j\}$ and $\{\ell \text{ makes an offer to } i \mid i \text{ applied to } \ell\}$ are independent for $j, \ell \in \mathcal{J}$, $j \neq \ell$. The job seeker can accept at most one offer.

In this setting, each job vacancy represents a risky option, and at most one option will be exercised. Let $j = 0$ represent the outside option, corresponding to either unemployment or the current job if employed. The ex post payoff of exercising option j is represented by Bernoulli utility function $u_i : \mathcal{J} \cup \{0\} \to \mathbb{R}$, with shorthand notation $u_{ij} = u_i(j)$. We rule out weakly dominated (by the outside option) jobs by assuming $u_{ij} \geq u_{i0}$ for al $j \in \mathcal{J}$, implying the job seeker accepts at least one offer, if any.[2] Thus, the outside option is exercised only when either every application in

[2] Chade and Smith (2006) impose the stronger assumption that $\alpha_{ij} u_{ij} - c_i(1) > u_{i0}$ for all $j \in \mathcal{J}$, which further implies that at least one application is made. In contrast, we allow job seekers to make no applications by choosing $A_i = \emptyset$.

A_i is rejected or no applications are made ($A_i = \emptyset$). Realization of any option in the application portfolio depends on receiving an offer from that job and being rejected by every preferred job application.

Let $r_i : \mathcal{P}(\mathcal{J}) \times \{1, \ldots, J\} \to \mathcal{J}$ identify the k–th most preferred job within portfolio $A \subseteq \mathcal{J}$ by $r_i(A, k) \in A$, with shorthand notation r^A_{ik}. Here, $k \in \{1, \ldots, |A|\}$ and $\mathcal{P}(S)$ denotes the power set of set S. We assume that preferences are strict, meaning $r_i(\cdot, \cdot)$ is indeed a function (as opposed to a correspondence) and $u_{ir_i(\mathcal{J},1)} > \cdots > u_{ir_i(\mathcal{J},J)}$.[3] Each application portfolio $A \subseteq \mathcal{J}$ gives rise to a lottery over state space $\mathcal{J} \cup \{0\}$, where outcomes $j \in \mathcal{J}$ represent exercising option j —i.e., getting the job— and outcome $j = 0$ corresponds to exercising the outside option. The lottery assigns positive probability only to jobs in the application portfolio, $j \in A$, and to the outside option, $j = 0$. As discussed above, option $j \in A$ is exercised if and only if the job seeker (i) receives an offer from job j, and (ii) is rejected by every job in the portfolio that is (ex post) preferred to j. Therefore, if j is ranked in the k–th position among $m \in A$, then the probability of exercising this option is given by

$$p_i(A, j) = \alpha_{ij} \prod_{\ell=1}^{k-1} \left(1 - \alpha_{ir_i(A,\ell)}\right). \tag{1}$$

Similarly, the probability of exercising the outside option is[4]

$$p_i(A, 0) = \prod_{m \in A} (1 - \alpha_{im}). \tag{2}$$

Let $U_i : \mathcal{P}(\mathcal{J}) \to \mathbb{R}$ represent the (ex ante) von Neumann–Morgenstern expected utility of the lottery induced by portfolio $A \subseteq \mathcal{J}$ and, without loss of generality, normalize $u_{i0} = 0$. Then, considering the cost of applications —which is incurred in any event—,

$$U_i(A) = \sum_{k=1}^{n} u_{ir_i(A,k)}\, \alpha_{ir_i(A,k)} \prod_{\ell=1}^{k-1} \left(1 - \alpha_{ir_i(A,\ell)}\right) - c_i(n), \tag{3}$$

where $n = |A|$ is the size of portfolio A. The resulting utility maximization problem,

$$\max_{A \subseteq \mathcal{J}} U_i(A), \tag{4}$$

is a complex combinatorial optimization problem. In principle, it involves computation and comparison of the expected utilities from the $|\mathcal{P}(\mathcal{J})| = 2^J$ feasible application portfolios that can be chosen from $\mathcal{J}$ (including the empty set $A = \emptyset$). However, Chade and Smith (2006) exploit the downward-recursive structure of this class of portfolio choice problem to show that their marginal

[3] Moreover, for fixed $A \subseteq \mathcal{J}$, $r_i(A, \cdot)$ is a bijection from $\{1, \ldots, |A|\}$ to A. This implies the existence of an inverse $r_i^{-1}(A, j)$ that returns the position of alternative $j \in A$ in the ranking of the alternatives in A.

[4] Equation (2) is a special case of Equation (1) since every inside option is preferred to the outside option and the outside option is not risky ($\alpha_{i0} \equiv 1$). It can be verified that these probabilities sum to one over $A \cup \{0\}$.

improvement algorithm (MIA) efficiently finds the optimal portfolio in $J(J+1)/2 = O(J^2)$ steps. The MIA is a greedy algorithm that starts by identifying the best singleton portfolio, then finds the best alternative to add to the best singleton portfolio to form the best portfolio of size two, and so on until the next best portfolio addition decreases expected utility (see Appendix A for details).

The discrete choice methods typically used in the estimation of demand for differentiated products rely on revealed (or sometimes stated) preference in the sense that the (actual or hypothetical) ex ante choice of alternative j over alternative ℓ truthfully reveals that the decision maker prefers j to ℓ ex post. This is not generally true in our simultaneous search setting. In particular, $j \in A_i$ and $\ell \notin A_i \;\not\Longrightarrow\; u_{ij} > u_{i\ell}$. In a special case of this model, however, a revealed-preference structure emerges by imposing the following simplifying assumptions, which are maintained throughout the paper unless stated otherwise.

Assumption 1. Homogeneous admission probabilities: $\alpha_{ij} = \alpha_i \in (0,1), \forall j \in \mathcal{J}$.

Assumption 2. Constant marginal cost of applications: $c_i(|A|) = \gamma_i\,|A|\,, \forall A \subseteq \mathcal{J}$, where $\gamma_i > 0$.

Under Assumption 1, the model retains a sufficient degree of uncertainty to induce job seekers to make multiple applications, while the mechanism preventing preference revelation disappears as the order of the (ex ante) expected values of the risky options coincides with the ex-post preference order: $\alpha_i u_{ij} > \alpha_i u_{i\ell} \iff u_{ij} > u_{i\ell}$. Therefore, for any currently available pair j, ℓ such that $u_{ij} > u_{i\ell}$, the MIA will choose j over ℓ for the next optimal portfolio addition in any given iteration, giving portfolio choice the revealed-preference property $j \in A_i$ and $\ell \notin A_i \implies u_{ij} > u_{i\ell}$. This intuitive result follows as a corollary to Lemma 2 of Chade and Smith (2006). Further imposing Assumption 2 yields a stopping rule that determines the size of the optimal portfolio as a function of preferences and the parameters α_i and γ_i. This stopping rule follows directly from the MIA stopping rule. Proposition 1 below formalizes these insights. The resulting choice rule can be combined with an additive random utility model for the ex post utilities $\{u_{ij}\}_{j\in\mathcal{J}}$ to produce a tractable econometric model of portfolio choice.

Proposition 1. *Under Assumptions 1 and 2, the portfolio choice model* (3)–(4) *reduces to a two-stage choice rule comprising:*

(i) Stopping rule: Determine optimal portfolio size n_i following the rule

$$n_i = \max\left\{\left\{n \in \{1,\ldots,J\} : u_{i,r_i(\mathcal{J},n)} \geq \frac{\gamma_i}{\alpha_i\,(1-\alpha_i)^{n-1}}\right\} \cup \{0\}\right\}. \tag{5}$$

(ii) Choice of best ex post alternatives: Conditional on optimal portfolio size n_i, choose the optimal portfolio A_i of size n_i by including the alternatives with the n_i highest ex post utilities such that

$$A_i = \left\{r_i(\mathcal{J},1),\ldots,r_i(\mathcal{J},n_i)\right\}. \tag{6}$$

Proof. See Appendix B. □

It is easy to see that the two-step choice rule described in Proposition 1 can be equivalently —and more compactly— represented as a one-step rule of the form

$$j \in A_i \iff u_{ij} \geq \frac{\gamma_i}{\alpha_i(1-\alpha_i)^{r_i^{-1}(\mathcal{J},j)-1}}. \tag{7}$$

However, the sequential representation will prove useful in estimation after specifying an additive random utility model. We proceed to discuss this in more detail in Section 2.2 and **??** below.

2.2 An additive random utility model for ex-post job preferences

We complete our model of the supply of applications to the firm by specifying a random utility model (ARUM hereafter) for the Bernoulli utilities $\{u_{ij}\}_{j\in\mathcal{J}}$ representing job seekers' ex-post preferences over the available vacancies. We impose a simple logit preference structure in order to keep the model tractable while cleanly illustrating the mechanisms introduced by uncertainty and the application portfolio problem discussed in Section 2.1. This approach has the advantage of yielding closed-form solutions for the relevant choice probabilities —which are generalizations of the well-known choice probability in the one-application setting—, but at the cost of imposing restrictive assumptions on preference heterogeneity and substitution patterns as a consequence of the independence of irrelevant alternatives (IIA) property. We further discuss these limitations and compare the model to single-application benchmarks in Section 2.3.

To connect the portfolio-choice problem to observable data on job characteristics and wages, we assume that each job seeker's ex-post utilities are additively separable in a systematic component and an idiosyncratic shock. This assumption yields a tractable job-differentiation structure while capturing the key economic trade-off between hedging against job-offer uncertainty and costly applications. Workers apply to jobs based on their mean utilities, but randomness in tastes and outcomes still drives variation in portfolios. Formally, each job seeker $i \in \mathcal{I}$ faces the portfolio choice problem described by Equations (3) and (4). The ex-post utility that the job seeker derives from working in job $j \in \mathcal{J}$ takes the additively separable form

$$u_{ij} = \delta_j + \varepsilon_{ij}, \tag{8}$$

where $\delta_j \in \mathbb{R}$ is the deterministic component, or mean utility, and ε_{ij} is a random taste shock representing the idiosyncratic component of ex-post utility. Mean utility is linear in log-wage and job characteristics:

$$\delta_j = \beta \ln(w_j) + \boldsymbol{x_j'\theta} + \xi_j. \tag{9}$$

Equations (8)–(9), together with Assumption 3 below, comprise the core of our logit ARUM structure.[5]

Assumption 3. The idiosyncratic taste shocks ε_{ij} are independent and identically distributed draws from a standard type-1 extreme value distribution, with cumulative distribution function $F_\varepsilon(x) = \exp(-\exp(-x))$ for $x \in \mathbb{R}$.

Given this logit structure, we can derive closed-form expressions —up to integrating out job-seeker heterogeneity in the uncertainty and cost parameters— for the supply of job applications at the vacancy and firm levels. To simplify notation, let $\mathcal{J} = \{1, \dots, J\}$ so we can use vector notation for quantities such as $\boldsymbol{\delta} = (\delta_1, \dots, \delta_J)' \in \mathbb{R}^J$.[6] The expected number of applications to job j in our model is given by

$$q_j(\boldsymbol{\delta}) = I \sum_{n=1}^{J} \mathscr{s}_{j|n}(\boldsymbol{\delta}) \mathscr{s}_n(\boldsymbol{\delta}), \tag{10}$$

where

$$\mathscr{s}_{j|n}(\boldsymbol{\delta}) = \sum_{k=J-n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} \frac{\exp(\delta_j)}{\sum_{\ell \in \{j\} \cup A} \exp(\delta_\ell)} + \sum_{s=1}^{|\mathcal{B}_j \setminus A|} (-1)^s \sum_{B \in \mathcal{R}_s(\mathcal{B}_j \setminus A)} \frac{\exp(\delta_j)}{\sum_{\ell \in \{j\} \cup A \cup B} \exp(\delta_\ell)} \tag{11}$$

is the probability that job j belongs to the application portfolio conditional on portfolio size, $\mathbb{P}(j \in A_i \mid n_i = n)$ for $n \in \{1, \dots, J\}$, $\mathcal{B}_j \equiv \mathcal{J} \setminus \{j\}$ is the set of jobs excluding j, and $\mathcal{R}_k(S) = \{\sigma \subseteq S : |S| = k\}$ is the set of all size-k subsets of set S.[7] The conditional probability mass function (pmf) of portfolio size n_i —i.e., the number of applications— given admission probability α_i and marginal cost of application γ_i, $\mathbb{P}(n_i = n \mid \alpha_i, \gamma_i)$, is

$$\begin{aligned}\mathscr{s}_{n|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i) = \sum_{s=0}^{J} \sum_{A \in \mathcal{R}_s(\mathcal{J})} \prod_{\ell \in A} \left[F_\varepsilon\left(\psi_i^{n+1}\right)^{\exp(\delta_\ell)} - F_\varepsilon\left(\psi_i^{n}\right)^{\exp(\delta_\ell)} \right] \\ \sum_{t \in \tau_n^s} \sum_{B \in \mathcal{R}_t(\mathcal{J} \setminus A)} F_\varepsilon\left(\psi_i^{n}\right)^{\sum_{p \in B} \exp(\delta_p)} \prod_{q \in \mathcal{J} \setminus (A \cup B)} \left[1 - F_\varepsilon\left(\psi_i^{n+1}\right)^{\exp(\delta_q)} \right]\end{aligned} \tag{12}$$

[5]It is possible, at the risk of reduced tractability, to derive richer, more flexible models by combining a more general structure for Equation (8) with different distributional assumptions in place of Assumption 3. Such generalizations are out of the scope of this paper and are thus left for future research. See, for example, Section 3 of Berry and Haile (2021), Chapters 2–6 of Train (2009), or Chapter 2 of Aguirregabiria (2021) for detailed discussions in the setting where only one alternative is selected.

[6]Alternatively, fix a bijection $j : \mathcal{J} \to \{1, \dots, J\}$ such that $\boldsymbol{\delta} = (\delta_{j^{-1}(1)}, \dots, \delta_{j^{-1}(J)})'$ is simply the permutation of $\{\delta_j\}_{j \in \mathcal{J}}$ induced by $j(\cdot)$. So far, we have left the nature of job identities $\mathcal{J}$ unspecified for clarity when defining mappings from jobs to rankings of jobs. It will be useful to work with vectors in what follows, so it is convenient to fix an ordering of $\mathcal{J}$.

[7]Note that (i) $\mathscr{s}_{j|J}(\boldsymbol{\delta}) = 1$, consistent with the trivial fact that $\mathbb{P}(j \in A_i \mid n_i = J) = 1$; (ii) $\mathscr{s}_{j|1}(\boldsymbol{\delta}) = \exp(\delta_j) / \sum_{\ell \in \mathcal{J}} \exp(\delta_\ell)$ coincides with the well-known choice probability in the multinomial logit model; (iii) $\sum_{j \in \mathcal{J}} \mathscr{s}_{j|n}(\boldsymbol{\delta}) = n$, consistent with the conditioning event that n alternatives are chosen; and (iv) $\mathscr{s}_{j|n}(\boldsymbol{\delta})$ increases monotonically with n, consistent with the fact that, for any job seeker, the n most preferred alternatives include the $n-1$ most preferred alternatives.

for $n \in \{1, \ldots, J-1\}$, where $\tau_n^s = \{\max(J-n-s, 0), \ldots, \min(J-n, J-s)\}$ is a set of consecutive natural numbers, and

$$\psi_i^n = \frac{\gamma_i}{\alpha_i(1-\alpha_i)^{n-1}}, \quad n \in \{1, \ldots, J\} \tag{13}$$

is shorthand for the thresholds in part (i) of Proposition 1. For the extreme cases $n = 0$ and $n = J$, the conditional pmf is

$$\mathscr{s}_{0|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i) = F_\varepsilon\left(\psi_i^1\right)^{\sum_{\ell\in\mathcal{J}} \exp(\delta_\ell)} \tag{14}$$

and

$$\mathscr{s}_{J|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i) = 1 - \sum_{k=1}^{J} \sum_{A\in\mathcal{R}_k(\mathcal{J})} F_\varepsilon\left(\psi_i^J\right)^{\sum_{\ell\in A} \exp(\delta_\ell)} \prod_{m\in\mathcal{J}\setminus A} \left[1 - F_\varepsilon\left(\psi_i^J\right)^{\exp(\delta_m)}\right], \tag{15}$$

respectively. The corresponding unconditional pmf, $\mathbb{P}(n_i = n)$ for $n \in \{0, \ldots, J\}$, is

$$\mathscr{s}_n(\boldsymbol{\delta}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathscr{s}_{n|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i)\mathrm{d}F_\alpha(\alpha_i)\mathrm{d}F_\gamma(\gamma_i). \tag{16}$$

See Appendix C for a full derivation.

The elasticity of the supply of applications to job $j \in \mathcal{J}$ with respect to the wage of vacancy $\ell \in \mathcal{J}$ answers the question "*If the wage offered by job ℓ increases by one percent, what is the percent increase in the number of applications to job j?*" and is given by

$$\eta_{q_j,w_\ell} = \frac{1}{q_j(\boldsymbol{\delta})}\left[I\sum_{n=1}^{J} \frac{\partial \mathscr{s}_{j|n}(\boldsymbol{\delta})}{\partial \delta_\ell}\mathscr{s}_n(\boldsymbol{\delta}) + \mathscr{s}_{j|n}(\boldsymbol{\delta})\frac{\partial \mathscr{s}_n(\boldsymbol{\delta})}{\partial \delta_\ell}\right]\beta. \tag{17}$$

Our object of interest is the own-wage elasticity of the supply of applications to the firm. This quantity answers the question "*If the firm raises the wages it offers for all its vacancies by one percent, what is the percent increase in the total number of applications it receives?*". Since the total number of applications to firm $f \in \mathcal{F}$ posting job vacancies $\mathcal{J}^f$,

$$q^f(\boldsymbol{\delta}) = \sum_{j\in\mathcal{J}^f} q_j(\boldsymbol{\delta}), \tag{18}$$

is simply the sum of the supply of applications to each of its posted vacancies, its elasticity is a weighted average of the corresponding vacancy-level elasticities:

$$\eta_{q^f,\boldsymbol{w}^f} = \frac{1}{q^f(\boldsymbol{\delta})} \sum_{\ell\in\mathcal{J}^f}\sum_{j\in\mathcal{J}^f} q_j(\boldsymbol{\delta})\eta_{q_j,w_\ell}. \tag{19}$$

See Appendix D.1 for a derivation of the vacancy- and firm-level elasticities, and Appendix D.2 for closed-form expressions for the partial derivatives of $\mathscr{s}_{j|n}(\cdot)$ and $\mathscr{s}_n(\cdot)$ —up to integration over $F_\alpha(\cdot) \times F_\gamma(\cdot)$.

2.3 Implications and limitations

Differences with single-application models. Having developed the model and derived the implied supply of applications and own-wage elasticity, we now compare our framework to single-application benchmarks. These comparisons clarify the mechanisms driving differences in application behavior and wage elasticities.

The textbook multinomial logit (MNL) model assigns an idiosyncratic taste shock to the outside option. In contrast, our framework treats the outside option deterministically and assumes that all considered vacancies are at least as attractive as the status quo. Our treatment of the outside option is more natural in the context of job applications: rational job seekers would never consider applying to vacancies that are worse than their current position, be it a job or unemployment. To illustrate the implications of allowing multiple applications, we benchmark our model against two natural alternatives: (i) an MNL model with a deterministic, ex-post dominated outside option, and (ii) a restricted version of our model in which job seekers can submit at most one application. Comparing these models highlights how portfolio choice affects both the expected number of applications per vacancy and the implied wage elasticities.

Baseline model. To facilitate cleaner comparisons with single-application benchmarks, we focus on a simplified version of our model in which we abstract away from job-seeker heterogeneity by setting $\alpha_i = \alpha \in (0, 1)$ and $\gamma_i = \gamma > 0$ for all $i \in \mathcal{I}$. Under these degenerate distributions for the uncertainty and cost parameters, the unconditional pmf of the number of applications per job seeker in Equation (16) coincides with the conditional pmf in Equations (12), (14) and (15). In this case, the thresholds in Equation (13) simplify to

$$\psi^n = \frac{\gamma}{\alpha(1-\alpha)^{n-1}}.$$

The expected number of applications received by each job vacancy and the conditional application shares are then given by Equations (10) and (11), respectively. We use this baseline model as the point of comparison for the deterministic-outside option MNL benchmark and the restricted single-application version of our framework introduced below.

Deterministic-outside MNL benchmark. As a first benchmark, we derive an MNL model with a deterministic, ex-post dominated outside option. Intuitively, this corresponds to a setting where job seekers face no job-offer uncertainty ($\alpha = 1$) and therefore never apply to more than one job. The resulting model preserves the deterministic treatment of the outside option but shuts down the portfolio-choice mechanism entirely. Formally, we establish in Lemma 1 below that setting $\alpha = 1$ in the baseline model produces an MNL model where the outside option, with a deterministic utility normalized to 0, is chosen only when application costs are too high. Conditional on applying, the choice among the inside options is standard MNL.[8]

[8] A model resembling the textbook MNL with random outside utility can be obtained by setting $\alpha_i = 1$ for

Lemma 1. *When $\alpha_i = 1$ and $c_i(|A|) = \gamma\,|A| > 0$ for all $i \in \mathcal{I}$, the additive random utility model of portfolio choice in Equations (3), (4) and (8) with extreme value type 1 independent and identically distributed random taste shocks $\{\varepsilon_{ij}\}_{j\in\mathcal{J}}$ collapses to a model where:*

(i) The optimal portfolio A_i is either a singleton or the empty set:

$$n_i \equiv |A_i| \in \{0,1\}\,.$$

(ii) Job seekers choose not to apply only when applications are too costly, with probability

$$\mathcal{s}_0^{(i)}(\boldsymbol{\delta}) = F_\varepsilon\Bigg(\gamma - \ln\Bigg(\sum_{\ell\in\mathcal{J}} \exp\left(\delta_\ell\right)\Bigg)\Bigg) = \exp\Bigg(-\exp\left(-\gamma\right)\sum_{\ell\in\mathcal{J}} \exp\left(\delta_\ell\right)\Bigg).$$

(iii) Conditional on application —i.e., $n_i = 1$—, the expected share of applications to job $j \in \mathcal{J}$ takes the standard logit form

$$\mathcal{s}_{j|1}^{(i)}(\boldsymbol{\delta}) = \frac{\exp\left(\delta_j\right)}{\sum_{\ell\in\mathcal{J}}\exp\left(\delta_\ell\right)}.$$

Proof. See Appendix D.3. $\square$

The expected number of applications to job j in this benchmark model is given by

Need to complete this benchmarking exercise!

$$q_j^{(i)}(\boldsymbol{\delta}) = I\,\mathcal{s}_{j|1}^{(i)}(\boldsymbol{\delta})\left[1 - \mathcal{s}_0^{(i)}(\boldsymbol{\delta})\right] \tag{20}$$

Single-application benchmark.

[Second benchmark model here]

The roles of uncertainty and application costs.

[Discussion of the roles of α and γ here]

Discuss how the unconditional pmf (and its truncated counterpart) vary with typical (α,γ) and with their heterogeneity. Links to identification in MSM (discussed later). Add simulation plots to illustrate.

Limitations.

---

all $i \in \mathcal{I}$, taking $F_\gamma(\cdot) = F_\varepsilon(\cdot)$, and letting $\varepsilon_{i0} \equiv \gamma_i \overset{iid}{\sim} \mathrm{EV}_1$. Since the EV_1 distribution has support $\mathbb{R}$, this requires relaxing the convexity assumption on application costs and, under a linear cost $c_i(|A|)$ as in Assumption 2, some job seekers draw negative costs. For those with $\gamma_i > 0$, applications are costly and the model collapses to a standard MNL in which each job seeker applies to at most one job: $A_i = \{j_i^*\}$, where $j_i^* = \arg\max_{j\in\mathcal{J}\cup\{0\}}\ u_{ij} - \mathbb{1}\{j \in \mathcal{J}\}\,\gamma_i = \arg\max_{j\in\mathcal{J}\cup\{0\}}\ \widetilde{u}_{ij}$, $\widetilde{u}_{ij} = \delta_j + \varepsilon_{ij}$, and $\delta_0 = 0$. However, for job seekers with $\gamma_i \leq 0$, applications are (weakly) subsidized and the optimal choice is $A_i = \mathcal{J}$, meaning they apply to all vacancies even when only one will be exercised. Thus, the textbook MNL emerges as a special case for job seekers with positive application costs, but the equivalence is only partial due to the behavior of those with $\gamma_i \leq 0$.

```
[Discussion of limitations here]
```

Discuss:

- IIA and trstrictive substitution patterns
- Restricted heterogeneity in systematic preferences (only vertically differentiated; all horizontal differentiation comes from the idiosyncratic taste shocks). Important (i, j)-characteristics such as distance to place of work or skill-requirements/occupational matching left out.
- Labour supply $\neq$ supply of applications. Need to discuss how my object of interest maps to monopsony power, and under what assumptions it provides a reasonable proxy for the wage elasticity of labour supply .
- Exogeneity and homogeneity of the job-offer probabilities $\alpha_{ij} = \alpha_i$. It is not clear that the homogeneity assumption is compatible with an equilibrium setting with endogenous selection where employers optimally choose among the applicant pool. Or maybe it is in a setting where firms screen on a job seeker-specific scalar "quality" index with firm-specific thresholds that do not depend on utility-relevant job characteristics. Need to think more carefully about this.
- We abstract away from directed/competitive search considerations, which have been shown to be important in labor markets. This connects to the consideration set/unobserved choice set heterogeneity problem and potentially to costly information acquisition (for example, the partially revealed wage information may induce some info acquisition process
- Static model does not capture important dynamic aspects of labor markets such as entry and exit of job vacancies and job seekers (search spell initiation and duration). These are typically absent in standard school/college choice problems with synchronized application/recruitment cycles.

3 Econometrics: Taking the model to data

```
[Discussion of econometric setting here]
```

Discuss:

- Data as a partition of the choice set.
- Separation of $\boldsymbol{\delta}$ and $\boldsymbol{\nu}$ estimation.

- Observed and unobserved job characteristics.
- BLP instruments.
- Numerical estimation of elasticities through simulation given parameter estimates.

3.1 Identification

`[Discussion of identification here]`

3.2 Partially rank-ordered logit

`[Discussion of PROL maximum likelihood estimation here]`

Full derivation of minorize-maximize algorithm in Appendix E

3.3 Method of simulated moments

Technical optimization details in Appendix F

Let

$$\boldsymbol{m}_i(n_i) = \begin{pmatrix} m_1(n_i) \\ \vdots \\ m_M(n_i) \end{pmatrix}, \tag{21}$$

where $n_i = |A_i|$ is the number of applications by job seeker $i \in \mathcal{I}$ choosing optimal portfolio $A_i \subseteq \mathcal{J}$, and $M \geq L = \dim(\boldsymbol{\nu})$ is the number of moments used in estimation. Note that the population moment

$$\mathbb{E}\left[\boldsymbol{m}_i \mid n_i > 0\right] \equiv \boldsymbol{m}(\boldsymbol{\nu}_0 \mid \boldsymbol{\delta}_0),$$

is a function only of true parameters $\boldsymbol{\delta}_0$ and $\boldsymbol{\nu}_0$.[9] Similarly, the empirical moment

$$\overline{\boldsymbol{m}}_I(\boldsymbol{n}) = \frac{\sum_{i \in \mathcal{I}} \boldsymbol{m}_i(n_i)}{\sum_{i \in \mathcal{I}} \mathbb{1}\left\{n_i > 0\right\}}, \tag{22}$$

depends only on $\boldsymbol{n} = (n_1, \ldots, n_I)$, the data vector of observed individual applications per job seeker. Given knowledge of true mean utilities $\boldsymbol{\delta}_0$, a method of moments estimator would minimize the

[9] We condition on $n_i > 0$ because we only observe data for actual applicants —i.e., job seekers with at least one application. The functions $\{\mathscr{s}_n(\boldsymbol{\delta})\}_{n=0}^{J}$ are implicit functions of $\boldsymbol{\nu}$ because of integration over $F_\alpha \times F_\gamma$ on the right-hand side of (16). Here we use the notation $\boldsymbol{m}(\boldsymbol{\nu} \mid \boldsymbol{\delta})$ instead of the more standard $\boldsymbol{m}(\boldsymbol{\delta}, \boldsymbol{\nu})$ to emphasize our interest in estimating $\boldsymbol{\nu}_0$ given knowledge (or a consistent estimate) of $\boldsymbol{\delta}_{\mathbf{0}}$.

weighted distance between the implied population moment at candidate parameter value $\boldsymbol{\nu}$ and its sample counterpart,

$$Q_I(\boldsymbol{\nu} \mid \boldsymbol{\delta}_0, \boldsymbol{n}) = \Big(\boldsymbol{m}(\boldsymbol{\nu} \mid \boldsymbol{\delta}_0) - \overline{\boldsymbol{m}}_I(\boldsymbol{n})\Big)' W \Big(\boldsymbol{m}(\boldsymbol{\nu} \mid \boldsymbol{\delta}_0) - \overline{\boldsymbol{m}}_I(\boldsymbol{n})\Big), \tag{23}$$

where W is a positive-definite weight matrix.

While the closed-form solutions derived in Appendix C yield an analytic expression for the objective function, the combinatorics involved in its computation render the problem computationally infeasible for large J. The method of simulated moments (MSM) provides a convenient alternative (McFadden, 1989; Pakes and Pollard, 1989) where, instead of computing $\boldsymbol{m}(\boldsymbol{\nu} \mid \boldsymbol{\delta})$ analytically, we construct a frequency simulator $\widehat{\boldsymbol{m}}_S(\boldsymbol{\nu} \mid \boldsymbol{\delta})$. We draw S independent samples of $I \times J$ taste-shock matrix $\boldsymbol{\varepsilon}^s$ from $F_\varepsilon(\cdot)$, $I \times 1$ vector $\boldsymbol{\alpha}^s$ from $F_\alpha(\cdot \mid \boldsymbol{\nu}_\alpha)$, and $I \times 1$ vector $\boldsymbol{\gamma}^s$ from $F_\gamma(\cdot \mid \boldsymbol{\nu}_\gamma)$. We solve the model for each $s \in \{1, \ldots, S\}$, obtaining a simulated vector of individual application frequencies $\boldsymbol{n}_s = (n_1^s, \ldots, n_I^s)$ that allows us to compute the sample moment in the simulated data and average over simulations:

$$\widehat{\boldsymbol{m}}_S(\boldsymbol{\nu} \mid \boldsymbol{\delta}) = \frac{1}{S} \sum_{s=1}^{S} \frac{\sum_{i \in \mathcal{I}} \boldsymbol{m}_i(n_i)}{\sum_{i \in \mathcal{I}} \mathbb{1}\{n_i > 0\}}. \tag{24}$$

At mean utilities $\boldsymbol{\delta}$ and given sample $\boldsymbol{n}$, our MSM estimator is the minimizer of sample criterion function

$$Q_{I,S}(\boldsymbol{\nu} \mid \boldsymbol{\delta}, \boldsymbol{n}) = \Big(\widehat{\boldsymbol{m}}_S(\boldsymbol{\nu} \mid \boldsymbol{\delta}) - \overline{\boldsymbol{m}}_I(\boldsymbol{n})\Big)' \widehat{W} \Big(\widehat{\boldsymbol{m}}_S(\boldsymbol{\nu} \mid \boldsymbol{\delta}) - \overline{\boldsymbol{m}}_I(\boldsymbol{n})\Big), \tag{25}$$

where $\widehat{W}$ is a positive-semidefinite, consistent estimate of positive-definite weight matrix W. The optimal weight matrix is

$$W^* = \mathbb{E}\left[\Big(\widehat{\boldsymbol{m}}_S(\boldsymbol{\nu} \mid \boldsymbol{\delta}) - \overline{\boldsymbol{m}}_I(\boldsymbol{n})\Big)\Big(\widehat{\boldsymbol{m}}_S(\boldsymbol{\nu} \mid \boldsymbol{\delta}) - \overline{\boldsymbol{m}}_I(\boldsymbol{n})\Big)'\right], \tag{26}$$

and we estimate it non-parametrically via bootstrap (see Appendix F for details).

Our MSM estimator of $\boldsymbol{\nu}$ is infeasible in the sense that it requires knowledge of the mean-utility vector $\boldsymbol{\delta}_0$. In practice, we replace $\boldsymbol{\delta}_0$ with its MLE $\widehat{\boldsymbol{\delta}}$, obtained from the MM algorithm for the partially rank-ordered logit discussed in Appendix E. By standard extremum-estimator arguments, combined with the results on simulation estimators of McFadden (1989) and Pakes and Pollard (1989), the feasible MSM estimator $\widehat{\boldsymbol{\nu}}(\widehat{\boldsymbol{\delta}})$ converges in probability to the same limit as the infeasible version $\widehat{\boldsymbol{\nu}}(\boldsymbol{\delta}_0)$ as the sample size $I \to \infty$ and either the number of simulations S is large and fixed or $S \to \infty$ at a suitable rate.[10] Intuitively, consistency of $\widehat{\boldsymbol{\delta}}$, together with continuity of the simulated moment criterion in $\boldsymbol{\delta}$ and a uniform law of large numbers imply, via the continuous mapping theorem, that replacing $\boldsymbol{\delta}_0$ with $\widehat{\boldsymbol{\delta}}$ leaves the probability limit of the MSM estimator unchanged, and, because the finite-S discontinuities in $\boldsymbol{\nu}$ vanish in probability,

[10] See, e.g., Newey and McFadden (1994) or Ch. 5 of van der Vaart (1998).

the population criterion is continuous and well-behaved. We suppress the dependence on $\boldsymbol{\delta}$, $\boldsymbol{n}$, I and S from notation once the context is clear.

The feasible MSM estimator remains asymptotically normal under standard conditions. Its asymptotic variance, however, differs: it is inflated relative to the infeasible estimator because it incorporates the sampling error from the first-stage estimation of $\boldsymbol{\delta}_0$. This distinction matters for inference, but not for consistency or identification. Standard errors can, in principle, be obtained by bootstrapping the entire two-step procedure (estimating $\boldsymbol{\delta}$ and then $\boldsymbol{\nu}$). This, however, is computationally intensive, and is not pursued in our empirical application in Section 4 given its illustrative purpose.

3.4 From structural parameters to elasticities

[Discussion of numerical estimation of elasticities through simulation here]

4 An empirical application: Online job applications

In this section, we present an empirical application of our model to online job applications using microdata from a prominent Chilean job board. The main purpose here is illustrative, given our restrictive assumptions on preference heterogeneity, substitution patterns, and the selectivity of recruitment. We describe the data and institutional setting in Section 4.1. Details of the estimation strategy are provided in Section 4.2, while results are presented and discussed in Sections 4.3 and 4.4, respectively.

4.1 Chilean job board data

Most of this comes directly from my MRes paper. Needs heavy reframing and rewriting. Some of the detail here may need to go to an appendix, and other important discussions are missing.

We estimate our model with data from online job board Trabajando.com, comprising information on job advertisements, applicants, and their applications between January 1, 2018, and December 31, 2018. The board operates in several, mostly Spanish-speaking countries, including Argentina, Brazil, Colombia, Chile, Mexico, Peru, Portugal, Puerto Rico, Spain, Uruguay, and Venezuela. We focus on the Chilean platform.[11]

[11] See Banfi and Villena-Roldán (2019), Banfi et al. (2022), Choi et al. (2025), and Banfi et al. (2025) for other papers using job board data from Trabajando.com.

Registration is free for job seekers, while firms either purchase a pack of ads ranging from one "standard" ad to three "standard" plus two "featured" ads or pay for a subscription plan among a menu of four alternatives.[12] Applicants and firms both fill out information forms designed for job seekers and employers, respectively. The job seeker form requires information on the applicant's expected salary, with the option of hiding it from potential employers. Similarly, firms are also required to report in the employer form the expected wage for the vacancy being posted, with the option of hiding it from potential applicants.

Employers' ability to hide wages from potential applicants may raise some concerns. For example, if job seekers cannot see offered wages, then the extent to which their application choices respond to wage differences might be severely reduced. However, job seekers can filter job ads by narrow ranges of offered wage, and ads with hidden wages are listed in the corresponding results.[13] Therefore, applicants do observe a noisy but relatively accurate signal of the offered wage when the firm decides to hide it. Moreover, and as described below, an indicator for hidden wage is included in the dataset, allowing us to control for this in the analysis. A second concern is that hidden wages could be misreported. Banfi and Villena-Roldán (2019) show evidence that implicit or hidden wages are reliable, consistent with the incentives posed by the ability of job seekers to filter by wage ranges: reporting and hiding nonsensical wages would be detrimental for the firm posting the ad.

The information is contained in four main datasets. The first contains information on each posted job advertisement; the second comprises information on the set of firms posting on the platform; the third includes information on users, that is, job seekers; and the fourth is the applications dataset, linking applicants and the ads they applied to. Unfortunately, information on application outcomes is not recorded so this empirical exercise cannot say much about the final hiring stage.

The `job ads` dataset contains information on <u>1,137,965 job ads posted during the sample</u> update number
period. The variables in the dataset are the identity of the firm posting the ad, publication and expiry dates, number of vacancies, a dummy indicating that the wage is publicly posted, required experience (in years), posted (or unposted) wage, work arrangement, type of contract, required education level, area of the vacancy (e.g., administrative), required professional sector (if any), required level of computer skills, and job title. The dataset also contains four categorical variables used by Banfi and Villena-Roldán (2019), each corresponding to a list of words that are repeated more than 100 times as one of the first four meaningful words of the job title. Their approach is similar to that of Marinescu and Wolthoff (2020), and produces a list of 137 categories such as analyst, chief, manager, or assistant for the first meaningful word; 274 categories for the second

[12] See `https://gestion.trabajando.cl/companies/planSelection` for details on advertisement packs, and `https://empresas.trabajando.cl/planes-corporativos/` for subscription plans.

[13] See section II of Banfi and Villena-Roldán (2019) for details.

meaningful word; and 211 and 68 for the third and fourth meaningful words, repsectively.[14]

The `employers` dataset comprises information on 39,780 firms posting ads on the website during the sample period. Observed variables include industry, region, and size. Firm size is measured as the number of employees and is reported as a categorical variable with categories 1–10, 11–50, 51–150, 151–300, 301–500, 501–1,000, 1,001–5,000, and >5,000. Banfi and Villena-Roldán (2019) report that many ads are posted by recruiting firms on behalf of the actual employer. This is not directly observable in the dataset, so we follow their strategy to deal with this issue. They identify a recruiting firm as one posting a number of vacancies exceeding half of the upper limit of its reported size interval in a given month. We control for a recruiting firm dummy in our analysis.

update number

The `users` dataset contains observations on 3,618,392 individuals initially submitting their CVs between years 1999 and 2019. Nearly half of them (46.67%) applied to at least one vacancy during the sample period. Observed individual characteristics include date of birth, sex, nationality, region, city, and municipality of residence, marital status, educational attainment and profession, employment, beginning and ending dates and wage of latest work experience, availability to work, wage expectation and an indicator for it being observable to employers, and the dates of CV registration and latest modification. This information on job seekers' characteristics is used for sampling purposes since our model abstracts away from modelling applicant heterogeneity.

update number

Finally, the `applications` dataset includes applicant, firm, and job ad identifiers and the date of application for 39,480,855 applications during the sample period.

update number

4.2 Estimation strategy

Choice of moments: Our first $M-1$ moments are binned probabilities of the form

$$\mathbb{P}\left(\kappa_\ell - 1 \leq n_i \leq \kappa_\ell \mid n_i > 0\right) = \mathbb{E}\left[m_\ell(n_i) \mid n_i > 0\right],$$

where $k_\ell = 2\ell + 1$ and $m_\ell(n) = \mathbb{1}\left\{\kappa_\ell - 1 \leq n \leq \kappa_\ell\right\}$ for $\ell \in \{1, \ldots, M-1\}$. The M–th moment is the survivor function at κ_M,

$$\mathbb{P}\left(n_i \geq \kappa_M \mid n_i > 0\right),$$

where $\kappa_M = 2M$ has positive probability mass in the right tail of the empirical (truncated) distribution of n_i.

4.3 Empirical results

`[Results here]`

[14] See Section II.D of Banfi and Villena-Roldán (2019) for details. The numbers of categories reported here differ somewhat because our dataset covers a different period.

4.4 Discussion

[Discussion of limitations and implications here]

5 Conclusion

[Conclusion here]

References

Aguirregabiria, V. (2021): "Empirical Industrial Organization: Models, Methods, and Applications," Unpublished manuscript.

Azar, J., S. Berry, and I. Marinescu (2022): "Estimating Labor Market Power," Working Paper 30365, National Bureau of Economic Research.

Azar, J. and I. Marinescu (2024): "Chapter 10 - Monopsony power in the labor market," in *Handbook of Labor Economics*, ed. by C. Dustmann and T. Lemieux, Elsevier, vol. 5, 761–827.

Banfi, S., S. Choi, and B. Villena-Roldán (2022): "Sorting on-line and on-time," *European Economic Review*, 146, 104128.

——— (2025): "Deconstructing Job Search Behavior," Working paper.

Banfi, S. and B. Villena-Roldán (2019): "Do High-Wage Jobs Attract More Applicants? Directed Search Evidence from the Online Labor Market," *Journal of Labor Economics*, 37, 715–746.

Berry, S. T. and P. A. Haile (2021): "Chapter 1 - Foundations of demand estimation," in *Handbook of Industrial Organization*, ed. by K. Ho, A. Hortaçsu, and A. Lizzeri, Elsevier, vol. 4, 1–62.

Brooks, W. J., J. P. Kaboski, Y. A. Li, and W. Qian (2021): "Exploitation of labor? Classical monopsony power and labor's share," *Journal of Development Economics*, 150, 102627.

Card, D. (2022): "Who Set Your Wage?" *American Economic Review*, 112, 1075–90.

Card, D., A. R. Cardoso, J. Heining, and P. Kline (2018): "Firms and Labor Market Inequality: Evidence and Some Theory," *Journal of Labor Economics*, 36, S13–S70.

Chade, H. and L. Smith (2006): "Simultaneous Search," *Econometrica*, 74, 1293–1307.

Choi, S., B. Villena-Roldán, and N. Figueroa (2025): "Posted Wage Cyclicality: Evidence from High-Quality Vacancy Data," Working paper.

Dal Bó, E., F. Finan, and M. A. Rossi (2013): "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service," *The Quarterly Journal of Economics*, 128, 1169–1218.

Hirsch, B., E. J. Jahn, A. Manning, and M. Oberfichtner (2022): "The Wage Elasticity of Recruitment," Discussion Paper 15675, IZA - Institute of Labor Economics.

Manning, A. (2003): *Monopsony in Motion: Imperfect Competition in Labor Markets*, Princeton University Press.

——— (2021): "Monopsony in Labor Markets: A Review," *ILR Review*, 74, 3–26.

Marinescu, I. and R. Wolthoff (2020): "Opening the Black Box of the Matching Function: The Power of Words," *Journal of Labor Economics*, 38, 535–568.

McFadden, D. (1989): "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," *Econometrica*, 57, 995–1026.

Mertens, M. and B. Mottironi (2023): "Do larger firms exert more market power? Markups and markdowns along the size distribution," Discussion Paper No. 1945, Centre for Economic Performance.

Newey, W. K. and D. McFadden (1994): "Chapter 36 Large sample estimation and hypothesis testing," in *Handbook of Econometrics*, ed. by R. F. Engle and D. L. McFadden, Elsevier, vol. 4, 2111–2245.

Pakes, A. and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057.

Roussille, N. and B. Scuderi (2025): "Bidding for Talent: A Test of Conduct in a High-Wage Labor Market," Working Paper 33848, National Bureau of Economic Research.

Train, K. E. (2009): *Discrete Choice Methods with Simulation. Second Edition.*, Cambridge University Press.

van der Vaart, A. W. (1998): *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

Yeh, C., C. Macaluso, and B. Hershbein (2022): "Monopsony in the US Labor Market," *American Economic Review*, 112, 2099–2138.

Appendix

A Marginal improvement algorithm

This appendix describes the Chade and Smith (2006) marginal improvement algorithm (MIA) within the context and notation of Section 2.1. The MIA is a greedy algorithm in the sense that it makes a locally optimal choice in each iteration. Despite its greedy nature, it converges to the global optimum, as shown by Chade and Smith (2006).

Consider the portfolio choice problem described by Equations (3) and (4). The MIA follows the following iterative procedure to find the optimal portfolio $A_i = \arg\max_{A \in \mathcal{P}(\mathcal{J})} U_i(A)$. Let $\Lambda_0 = \emptyset$. At iteration $t \in \{1, \ldots, J\}$:

- Step 1: Choose any $j_t \in \underset{j \in \mathcal{J} \setminus \Lambda_{t-1}}{\arg\max}\, U_i(\Lambda_{t-1} \cup \{j\})$.
- Step 2: Stop if $U_i(\Lambda_{t-1} \cup \{j_t\}) - U_i(\Lambda_{t-1}) < 0$.
- Step 3: Set $\Lambda_t = \Lambda_{t-1} \cup \{j_t\}$ and go to step 1 for the next iteration.

The algorithm will stop at iteration $t = \min(n_i + 1, J)$, where $n_i \equiv |A_i| \leq J$, identifying A_i.

B Proof of Proposition 1

Proof. Consider the portfolio choice problem (3)–(4). Let us start by showing that Assumption 1 implies that, conditional on $|A_i| = n$ —where $A_i = \arg\max_{A \in \mathcal{P}(\mathcal{J})} U_i(A)$—, A_i consists of the n (ex-post) best alternatives. This can be established by induction.

Consider iteration $t = 1$ of the marginal improvement algorithm (MIA) described in Appendix A. The best singleton portfolio must be the best ex post alternative since the order of expected values $\{\alpha_i u_{ij}\}_{j \in \mathcal{J}}$ coincides with the order of ex post utilities $\{u_{ij}\}_{j \in \mathcal{J}}$. Formally,

$$\begin{aligned}
\underset{j \in \mathcal{J} \setminus \Lambda_0}{\arg\max}\, U_i(\Lambda_0 \cup \{j\}) &= \underset{j \in \mathcal{J}}{\arg\max}\, U_i(\{j\}) \\
&= \underset{j \in \mathcal{J}}{\arg\max}\, \alpha_i u_{ij} - c_i(1) \\
&= \underset{j \in \mathcal{J}}{\arg\max}\, u_{ij} \\
&= \Big\{ r_i(\mathcal{J}, 1) \Big\},
\end{aligned}$$

where the first equality follows from $\Lambda_0 = \emptyset$, the second equality follows by direct evaluation of (3) at $A = \{j\}$, the third equality follows because quantities $\alpha_i > 0$ and $c_i(1)$ do not vary with j, and the last equality follows from the definition of the ranking function $r_i(\cdot, \cdot)$.

Next, consider iteration $t > 1$ and suppose that

$$\Lambda_{t-1} = \Big\{ r_i(\mathcal{J}, 1), \ldots, r_i(\mathcal{J}, t-1) \Big\}, \tag{B.1}$$

i.e., the MIA-optimal portfolio of size $t-1$ consists of the $t-1$ (ex-post) best alternatives. The induction hypothesis (B.1) implies that any alternative still available for selection by the MIA must be ranked higher —i.e., worse— than all the alternatives the MIA has already selected in previous iterations. That is, for all $j \in \mathcal{J} \setminus \Lambda_{t-1}$ and $\ell \in \Lambda_{t-1}$,[15]

$$r_i^{-1}(\mathcal{J}, j) > r_i^{-1}(\mathcal{J}, \ell). \tag{B.2}$$

Moreover, the ranking order over Λ_{t-1} must obviously coincide with the first $t-1$ positions of the ranking order over $\mathcal{J}$, i.e.,

$$r_i(\Lambda_{t-1}, k) = r_i(\mathcal{J}, k) \tag{B.3}$$

for all $k \in \{1, \ldots, t-1\}$. It follows that the MIA-optimal addition to Λ_{t-1} in iteration t must be $r_i(\mathcal{J}, t)$ since

$$\begin{aligned}
\underset{j \in \mathcal{J} \setminus \Lambda_{t-1}}{\arg\max}\, U_i(\Lambda_{t-1} \cup \{j\}) &= \underset{j \in \mathcal{J} \setminus \{r_i(\mathcal{J},k)\}_{k=1}^{t-1}}{\arg\max}\, \alpha_i \left[\sum_{k=1}^{t-1} (1-\alpha_i)^{k-1} u_{i r_i(\mathcal{J},k)} + (1-\alpha_i)^{t-1} u_{ij} \right] - c_i(t) \\
&= \underset{j \in \mathcal{J} \setminus \{r_i(\mathcal{J},k)\}_{k=1}^{t-1}}{\arg\max}\, u_{ij} \\
&= \Big\{ r_i(\mathcal{J}, t) \Big\},
\end{aligned}$$

where the first equality follows from (B.2)–(B.3) and direct evaluation of (3) at $\Lambda_{t-1} \cup \{j\}$ under Assumption 1, the second equality follows by discarding all (non-negative when appropriate) quantities that do not vary with j, and the last equality follows from the definition of the ranking function. Since $t > 1$ is arbitrary and we have proved the induction hypothesis holds for $t = 1$, the principle of mathematical induction establishes part (ii) of Proposition 1.

Part (i) of Proposition 1 follows directly from the stopping rule in step 2 of the MIA under Assumptions 1 and 2 by noting that, by part (ii) of the proposition, the optimal portfolio size n_i is also the position in the ranking over $\mathcal{J}$ of the last chosen alternative. This means $r_i(\mathcal{J}, n_i)$ is the last alternative the MIA picks up. Hence, the optimal portfolio contains n_i alternatives if and only if (a) the MIA does not stop in step 2 of iteration n_i, and (b) either $n_i = J$ or the MIA stops in step 2 of iteration $n_i + 1$.

[15] An equivalent statement to (B.2) is $u_{ij} < u_{i\ell}$, but the former expression highlights the order of alternatives that determines the relevant lottery whose expected utility the MIA maximizes in iteration t.

From (a), we obtain

$$
\begin{aligned}
0 &\leq U_i(\Lambda_{n_i-1} \cup \{r_i(\mathcal{J}, n_i)\}) - U_i(\Lambda_{n_i-1}) \\
&= \alpha_i \sum_{k=1}^{n_i-1} (1-\alpha_i)^{k-1} u_{ir_i(\mathcal{J},k)} + \alpha_i (1-\alpha_i)^{n_i-1} u_{ir_i(\mathcal{J},n_i)} - \gamma_i n_i \\
&\qquad - \left[\alpha_i \sum_{k=1}^{n_i-1} (1-\alpha_i)^{k-1} u_{ir_i(\mathcal{J},k)} - \gamma_i (n_i - 1) \right] \\
&= \alpha_i (1-\alpha_i)^{n_i-1} u_{ir_i(\mathcal{J},n_i)} - \gamma_i,
\end{aligned}
$$

which holds if and only if

$$
u_{ir_i(\mathcal{J},n_i)} \geq \frac{\gamma_i}{\alpha_i (1-\alpha_i)^{n_i-1}}. \tag{B.4}
$$

Similarly from (b), either $n_i = J$ or

$$
\begin{aligned}
0 &> U_i(\Lambda_{n_i} \cup \{r_i(\mathcal{J}, n_i+1)\}) - U_i(\Lambda_{n_i}) \\
&= \alpha_i \sum_{k=1}^{n_i} (1-\alpha_i)^{k-1} u_{ir_i(\mathcal{J},k)} + \alpha_i (1-\alpha_i)^{n_i} u_{ir_i(\mathcal{J},n_i+1)} - \gamma_i (n_i+1) \\
&\qquad - \left[\alpha_i \sum_{k=1}^{n_i} (1-\alpha_i)^{k-1} u_{ir_i(\mathcal{J},k)} - \gamma_i n_i \right] \\
&= \alpha_i (1-\alpha_i)^{n_i} u_{ir_i(\mathcal{J},n_i+1)} - \gamma_i,
\end{aligned}
$$

which holds if and only if

$$
u_{ir_i(\mathcal{J},n_i+1)} < \frac{\gamma_i}{\alpha_i (1-\alpha_i)^{n_i}}. \tag{B.5}
$$

Finally, note that the following monotonicity properties must hold.

$$
u_{ir_i(\mathcal{J},k)} \geq \frac{\gamma_i}{\alpha_i (1-\alpha_i)^{k-1}}, \forall k \in \{1, \ldots, n_i - 1\}, \tag{B.6}
$$

$$
u_{ir_i(\mathcal{J},k)} < \frac{\gamma_i}{\alpha_i (1-\alpha_i)^{k-1}}, \forall k \in \{n_i + 2, \ldots, J\}, \tag{B.7}
$$

where $\{n_i + 2, \ldots, J\} \equiv \emptyset$ for $n_i \geq J - 1$. Suppose (B.6) does not hold, so $u_{ir_i(\mathcal{J},k)} < \gamma_i \alpha_i^{-1} (1 - \alpha_i)^{-(k-1)}$ for some $k \in \{1, \ldots, n_i - 1\}$. Then, we get the contradiction

$$
u_{ir^{\mathcal{J}}_{ik}} > u_{ir^{\mathcal{J}}_{in_i}} \geq \frac{\gamma_i}{\alpha_i (1-\alpha_i)^{n_i-1}} = (1-\alpha_i)^{k-n_i} \frac{\gamma_i}{\alpha_i (1-\alpha_i)^{k-1}} > (1-\alpha_i)^{k-n_i} u_{ir^{\mathcal{J}}_{ik}} > u_{ir^{\mathcal{J}}_{ik}}
$$

since $k < n_i$ and $\alpha_i \in (0, 1) \implies (1 - \alpha_i)^{k-n_i} > 1$. Similarly, suppose (B.7) does not hold, so $u_{ir_i(\mathcal{J},k)} \geq \gamma_i \alpha_i^{-1} (1 - \alpha_i)^{-(k-1)}$ for some $k \in \{n_i + 2, \ldots, J\}$. Then, we get the contradiction

$$
u_{ir^{\mathcal{J}}_{ik}} < u_{ir^{\mathcal{J}}_{in_i+1}} < \frac{\gamma_i}{\alpha_i (1-\alpha_i)^{n_i}} = (1-\alpha_i)^{k-(n_i+1)} \frac{\gamma_i}{\alpha_i (1-\alpha_i)^{k-1}} \leq (1-\alpha_i)^{k-(n_i+1)} u_{ir^{\mathcal{J}}_{ik}} < u_{ir^{\mathcal{J}}_{ik}}
$$

since $k > n_i + 1$ and $\alpha_i \in (0, 1) \implies (1 - \alpha_i)^{k-(n_i+1)} < 1$. Together, (B.4)–(B.7) establish part (i) of Proposition 1. $\square$

C Derivation of the job applications supply function

This appendix provides a full derivation of the applications supply function, the conditional applications share function, and the probability mass function (pmf) of the number of applications in Equations (10), (11) and (16), respectively. Given a finite set of job seekers, $\mathcal{I}$ with $|\mathcal{I}| \equiv I$, facing the portfolio choice problem (3)–(4) over applications to a finite set of jobs, $\mathcal{J}$ with $|\mathcal{J}| \equiv J$, the expected number of applications to job $j \in \mathcal{J}$ is

$$\begin{aligned}
\mathbb{E}\left[q_j\right] &= \mathbb{E}\left[\sum_{i \in \mathcal{I}} \mathbb{1}\left\{j \in A_i\right\}\right] \\
&= \sum_{i \in \mathcal{I}} \mathbb{E}\big[\mathbb{1}\left\{j \in A_i\right\}\big] \\
&= I\, \mathbb{P}\left(j \in A_i\right) \\
&= I \sum_{n=1}^{J} \mathbb{P}\left(j \in A_i \mid n_i = n\right) \mathbb{P}\left(n_i = n\right). \qquad \text{(C.1)}
\end{aligned}$$

Our model defines (i) a mapping $\mathcal{s}_{j|n}(\boldsymbol{\delta})$ from $\boldsymbol{\delta}$ to $\mathbb{P}\left(j \in A_i \mid n_i = n\right)$, and (ii) a mapping $\mathcal{s}_n(\boldsymbol{\delta})$ from $\boldsymbol{\delta}$ and the joint distribution of parameters (α_i, γ_i) to $\mathbb{P}\left(n_i = n\right)$. These mappings follow directly from parts (ii) and (i) of Proposition 1, respectively.

C.1 Conditional applications share function

Consider first the conditional (expected) applications share function $\mathcal{s}_{j|n}(\boldsymbol{\delta})$. The probability that j belongs to the application portfolio conditional on the job seeker applying to every job is trivially $\mathbb{P}\left(j \in A_i \mid n_i = J\right) = 1$. For $n \in \{1, \ldots, J-1\}$, the probability that job j belongs to the application portfolio conditional on the job seeker applying to n jobs is the probability that the ex post utility of job seeker i from job j is larger than the ex post utility from their $(n+1)$–th most preferred alternative, i.e., $\mathbb{P}\left(j \in A_i \mid n_i = n\right) = \mathbb{P}\left(u_{ij} > u_{ir_i(\mathcal{J},n+1)}\right)$. This is true since job seeker i applies to job j if and only if j is among i's $n_i = n$ most preferred alternatives. We can derive the expression for $\mathbb{P}\left(u_{ij} > u_{ir_i(\mathcal{J},n+1)}\right)$ as a function of $\boldsymbol{\delta}$ —and, obviously, of j and n, which we indicate by the subscript in $\mathcal{s}_{j|n}(\boldsymbol{\delta})$— defined by our model by applying a well-established result from the literature on order statistics.

Let $\{u_{i(n)}\}_{n=1}^{J}$ represent the order statistics of $\{u_{ij}\}_{j \in \mathcal{J}}$ such that $u_{i(1)} < \cdots < u_{i(J)}$, and note that

$$u_{ir_i(\mathcal{J},n+1)} = u_{i(J-n)} \qquad \text{(C.2)}$$

for all $n \in \{1, \ldots, J-1\}$. Similarly, let $\mathcal{B}_j \equiv \mathcal{J} \setminus \{j\}$ represent the leave-out set of available jobs excluding j, and $\{u^j_{i(n)}\}_{n=1}^{J-1}$ the order statistics of $\{u_{i\ell}\}_{\ell \in \mathcal{B}_j}$ such that $u^j_{i(1)} < \cdots < u^j_{i(J-1)}$.

Notice that "j `is among the best` n `jobs in` $\mathcal{J}$" if and only if "j `is better than the` $J-n$ `worse jobs in` $\mathcal{J}$" if and only if "j `is better than the` $J-n$ `worse jobs in` $\mathcal{B}_j$" for any $n \in \{1, \ldots, J-1\}$. The mutual independence of $\{u_{i\ell}\}_{\ell \in \mathcal{J}}$ implies that $u^j_{i(n)}$ is independent of u_{ij} for all $n \in \{1, \ldots, J-1\}$.

The *iid* assumption on $\{\varepsilon_{ij}\}_{j\in\mathcal{J}}$ implies that the ex post utilities $\{u_{ij}\}_{j\in\mathcal{J}}$ are independently but non-identically distributed with cumulative distribution function (cdf)

$$
\begin{aligned}
F_{u_j}(x) &\equiv \mathbb{P}\left(u_{ij} \leq x\right) \\
&= \mathbb{P}\left(\varepsilon_{ij} \leq x - \delta_j\right) \\
&= F_\varepsilon(x - \delta_j),
\end{aligned} \tag{C.3}
$$

where $F_\varepsilon(\cdot)$ is the marginal cdf of ε_{ij}. The cdf of the n–th order statistic $u^j_{i(n)}$ is then given by (see, e.g., David and Nagaraja, 2003, p. 96)

$$
\begin{aligned}
F_{u^j_{(n)}}(x) &= \sum_{k=n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} \prod_{\ell \in A} F_{u_\ell}(x) \prod_{m \in \mathcal{B}_j \setminus A} \left[1 - F_{u_m}(x)\right] \\
&= \sum_{k=n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} \prod_{\ell \in A} F_\varepsilon(x - \delta_\ell) \prod_{m \in \mathcal{B}_j \setminus A} \left[1 - F_\varepsilon(x - \delta_m)\right],
\end{aligned} \tag{C.4}
$$

where $\mathcal{R}_k(S) \equiv \{\sigma \subseteq S : |\sigma| = k\}$ is the set of all size-k subsets of set S —that is, all the k-combinations of S. Combining these results and leveraging the properties of the EV_1 distribution, $F_\varepsilon(x) = \exp\left(-\exp\left(-x\right)\right)$, we obtain

$$
\begin{aligned}
s_{j|n}(\boldsymbol{\delta}) &= \mathbb{P}\left(u_{ij} > u_{ir_i(\mathcal{J}, n+1)}\right) \\
&= \mathbb{P}\left(u_{ij} > u_{i(J-n)}\right) \\
&= \mathbb{P}\left(u_{ij} > u^j_{i(J-n)}\right) \\
&= \int_{-\infty}^{\infty} \mathbb{P}\left(x > u^j_{i(J-n)}\right) \mathrm{d}F_{u_j}(x) \\
&= \int_{-\infty}^{\infty} F_{u^j_{(J-n)}}(x) \mathrm{d}F_\varepsilon(x - \delta_j) \\
&= \int_{-\infty}^{\infty} \sum_{k=J-n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} \prod_{\ell \in A} F_\varepsilon(x - \delta_\ell) \prod_{m \in \mathcal{B}_j \setminus A} \left[1 - F_\varepsilon(x - \delta_m)\right] \mathrm{d}F_\varepsilon(x - \delta_j) \\
&= \int_{-\infty}^{\infty} \sum_{k=J-n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} \prod_{\ell \in A} F_\varepsilon(x - \delta_j)^{\frac{\exp(\delta_\ell)}{\exp(\delta_j)}} \prod_{m \in \mathcal{B}_j \setminus A} \left[1 - F_\varepsilon(x - \delta_j)^{\frac{\exp(\delta_m)}{\exp(\delta_j)}}\right] \mathrm{d}F_\varepsilon(x - \delta_j)
\end{aligned}
$$

$$= \int_0^1 \sum_{k=J-n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} \prod_{\ell \in A} u^{\frac{\exp(\delta_\ell)}{\exp(\delta_j)}} \prod_{m \in \mathcal{B}_j \setminus A} \left[1 - u^{\frac{\exp(\delta_m)}{\exp(\delta_j)}}\right] \mathrm{d}u$$

$$= \int_0^1 \sum_{k=J-n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} u^{\frac{\sum_{\ell \in A} \exp(\delta_\ell)}{\exp(\delta_j)}} \prod_{m \in \mathcal{B}_j \setminus A} \left[1 - u^{\frac{\exp(\delta_m)}{\exp(\delta_j)}}\right] \mathrm{d}u. \tag{C.5}$$

The second equality follows from (C.2). The third equality follows from equivalence of the events as discussed above. The fourth equality follows by integrating over the marginal distribution of u_{ij}. The fifth equality follows from (C.3) and the definition of the cdf of $u^j_{i(J-n)}$. The sixth equality follows from (C.4). The seventh equality follows from the fact that $F_\varepsilon(x - \ln(a)) = F_\varepsilon(x - \ln(b))^{a/b}$ for $a, b > 0$. The eigth equality follows by the change of variable $u = F_\varepsilon(x - \delta_j)$, and the last equality follows from the algebraic rules of exponentiation.

Equation (C.5) defining the conditional expected applications share function, $\mathscr{s}_{j|n}(\boldsymbol{\delta})$, is a generalization of the well-known choice probabilities of the multinomial logit model. We can easily verify that we get the standard choice probability for $n = 1$:

$$\mathscr{s}_{j|1}(\boldsymbol{\delta}) = \int_0^1 u^{\frac{\sum_{\ell \in \mathcal{J} \setminus \{j\}} \exp(\delta_\ell)}{\exp(\delta_j)}} \mathrm{d}u$$

$$= \frac{\exp(\delta_j)}{\sum_{\ell \in \mathcal{J}} \exp(\delta_\ell)},$$

since $\mathcal{R}_{J-1}(\mathcal{B}_j) = \mathcal{R}_{|\mathcal{B}_j|}(\mathcal{B}_j) = \{\mathcal{B}_j\}$ and $\mathcal{B}_j \setminus \mathcal{B}_j = \emptyset$. Note that the conditional expected shares $\mathscr{s}_{j|n}(\boldsymbol{\delta})$ satisfy the following recursive relation. We can rewrite (C.5) as

$$\mathscr{s}_{j|n}(\boldsymbol{\delta}) = \int_0^1 f_{j|n}(u, \boldsymbol{\delta})\, \mathrm{d}u,$$

where

$$f_{j|n}(u, \boldsymbol{\delta}) = \sum_{k=J-n}^{J-1} f_j(u, \boldsymbol{\delta}, k)$$

and

$$f_j(u, \boldsymbol{\delta}, k) = \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} u^{\frac{\sum_{\ell \in A} \exp(\delta_\ell)}{\exp(\delta_j)}} \prod_{m \in \mathcal{B}_j \setminus A} \left[1 - u^{\frac{\exp(\delta_m)}{\exp(\delta_j)}}\right].$$

For $n \in \{2, \ldots, J\}$, we can recursively decompose

$$f_{j|n}(u, \boldsymbol{\delta}) = \sum_{k=J-(n-1)}^{J-1} f_j(u, \boldsymbol{\delta}, k) + \sum_{k=J-n}^{J-n} f_j(u, \boldsymbol{\delta}, k)$$

$$= f_{j|n-1}(u, \boldsymbol{\delta}) + f_j(u, \boldsymbol{\delta}, J - n)$$

$$\vdots$$

$$= f_{j|1}(u, \boldsymbol{\delta}) + \sum_{k=J-n}^{J-2} f_j(u, \boldsymbol{\delta}, k),$$

implying the recursive relations

$$\mathscr{s}_{j|n}(\boldsymbol{\delta}) = \mathscr{s}_{j|n-1}(\boldsymbol{\delta}) + \int_0^1 f_j(u, \boldsymbol{\delta}, J-n)\,\mathrm{d}u, \tag{C.6}$$

$$\mathscr{s}_{j|n}(\boldsymbol{\delta}) = \mathscr{s}_{j|1}(\boldsymbol{\delta}) + \int_0^1 \sum_{k=J-n}^{J-2} f_j(u, \boldsymbol{\delta}, k)\,\mathrm{d}u. \tag{C.7}$$

Furthermore, since $f_j(u, \boldsymbol{\delta}, J-n) \geq 0$ for $u \in [0,1]$, Equation (C.6) establishes that the conditional share $\mathscr{s}_{j|n}(\boldsymbol{\delta})$ increases monotonically with the number of applications n. Finally, while our derivation of $\mathscr{s}_{j|n}(\boldsymbol{\delta})$ assumed $n \in \{1, \ldots, J-1\}$, it is possible to show that the resulting expression is also valid for $n = J$, integrating to $\mathscr{s}_{j|J}(\boldsymbol{\delta}) = 1$, and that $\sum_{j \in \mathcal{J}} \mathscr{s}_{j|n}(\boldsymbol{\delta}) = n$ for all $n \in \{1, \ldots, J\}$.

Given parameters $\boldsymbol{\delta}$, the integral on the right-hand side of Equation (C.5) can be accurately approximated by numerical quadrature for any $n \in \{1, \ldots, J\}$. Alternatively, we can obtain a closed-form solution by noting that

$$\prod_{m \in \mathcal{B}_j \setminus A} \left[1 - u^{\frac{\exp(\delta_m)}{\exp(\delta_j)}}\right] = 1 + \sum_{s=1}^{|\mathcal{B}_j \setminus A|} (-1)^s \sum_{B \in \mathcal{R}_s(\mathcal{B}_j \setminus A)} u^{\frac{\sum_{m \in B} \exp(\delta_m)}{\exp(\delta_j)}},$$

by standard combinatorics —e.g., by a straightforward generalization of the binomial theorem—, so (C.5) simplifies to

$$\begin{aligned}
\mathscr{s}_{j|n}(\boldsymbol{\delta}) &= \sum_{k=J-n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} \int_0^1 u^{\frac{\sum_{\ell \in A} \exp(\delta_\ell)}{\exp(\delta_j)}} \mathrm{d}u + \sum_{s=1}^{J-1-k} (-1)^s \sum_{B \in \mathcal{R}_s(\mathcal{B}_j \setminus A)} \int_0^1 u^{\frac{\sum_{\ell \in A \cup B} \exp(\delta_\ell)}{\exp(\delta_j)}} \mathrm{d}u \\
&= \sum_{k=J-n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} \frac{\exp(\delta_j)}{\sum_{\ell \in \{j\} \cup A} \exp(\delta_\ell)} + \sum_{s=1}^{J-1-k} (-1)^s \sum_{B \in \mathcal{R}_s(\mathcal{B}_j \setminus A)} \frac{\exp(\delta_j)}{\sum_{\ell \in \{j\} \cup A \cup B} \exp(\delta_\ell)},
\end{aligned} \tag{C.8}$$

where $\sum_{s=1}^{0}(\cdot) \equiv 0$ for notational consistency. Given parameter estimates $\widehat{\boldsymbol{\delta}}$, the computational burden in estimating these generalized conditional choice probabilities, either numerically or analytically, grows quickly with the number of alternatives due to the combinatorics involved.

C.2 Probability mass function of the number of applications

Consider now the conditional pmf of the number of applications conditional on the admission probability α_i and the cost of applications γ_i, $\mathscr{s}_{n|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i)$. We can recover the unconditional pmf, $\mathscr{s}_n(\boldsymbol{\delta})$, by integrating the conditional pmf over the joint distribution of parameters (α_i, γ_i), which we assume to be statistically independent. We start by obtaining the conditional pmf at

$n = 0$ despite the conditioning event $n_i = 0$ not appearing explicitly Equation (C.1).[16] The job seeker does not apply to any jobs when the expected utility of the singleton portfolio comprising the best ex post alternative is negative, i.e.,

$$n_i = 0 \iff U_i(\{r_i(\mathcal{J}, 1)\}) < 0 \iff u_{i(J)} < \psi_i^1,$$

where the thresholds $\{\psi_i^n\}_{n=1}^J$ are defined as functions of (α_i, γ_i) in Equation (13). Conditional on (α_i, γ_i), the probability of this event is

$$\begin{aligned} s_{0|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i) &= \mathbb{P}\left(n_i = 0 \mid \alpha_i, \gamma_i\right) \\ &= F_{u_{(J)}}\left(\psi_i^1\right) \\ &= \prod_{\ell \in \mathcal{J}} F_\varepsilon\left(\psi_i^1\right)^{\exp(\delta_\ell)} \\ &= F_\varepsilon\left(\psi_i^1\right)^{\sum_{\ell \in \mathcal{J}} \exp(\delta_\ell)}. \end{aligned} \tag{C.9}$$

Similarly, for the case $n = J$, the job seeker applies to every job when even the marginal gain in expected utility from expanding the locally-optimal size $J - 1$ portfolio to include their least preferred job is non-negative, i.e.,

$$n_i = J \iff U_i(\{r_{i1}^{\mathcal{J}}, \ldots, r_{iJ}^{\mathcal{J}}\}) - U_i(\{r_{i1}^{\mathcal{J}}, \ldots, r_{iJ-1}^{\mathcal{J}}\}) \geq 0 \iff u_{i(1)} \geq \psi_i^J.$$

The conditional probability is given by

$$\begin{aligned} s_{J|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i) &= \mathbb{P}\left(n_i = J \mid \alpha_i, \gamma_i\right) \\ &= 1 - F_{u_{(1)}}\left(\psi_i^J\right) \\ &= 1 - \sum_{k=1}^{J} \sum_{A \in \mathcal{R}_k(\mathcal{J})} \prod_{\ell \in A} F_\varepsilon\left(\psi_i^J - \delta_\ell\right) \prod_{m \in \mathcal{J} \setminus A} \left[1 - F_\varepsilon\left(\psi_i^J - \delta_m\right)\right] \\ &= 1 - \sum_{k=1}^{J} \sum_{A \in \mathcal{R}_k(\mathcal{J})} F_\varepsilon\left(\psi_i^J\right)^{\sum_{\ell \in A} \exp(\delta_\ell)} \prod_{m \in \mathcal{J} \setminus A} \left[1 - F_\varepsilon\left(\psi_i^J\right)^{\exp(\delta_m)}\right]. \end{aligned} \tag{C.10}$$

Finally, for the interior events $n_i = n \in \{1, \ldots, J-1\}$, the stopping rule in part (i) of Proposition 1 implies that

$$n_i = n \iff u_{i(J-n+1)} \geq \psi_i^n \text{ and } u_{i(J-n)} < \psi_i^{n+1},$$

[16] The application of the law of total probability in Equation (C.1) actually requires consideration of the case $n_i = 0$, but the conditional probability $\mathbb{P}(j \in A_i \mid n_i = 0)$ is obviously zero. We include the event $n_i = 0$ for completeness, but also because it illustrates the reasoning behind the derivations for $n_i > 0$ in the simplest possible scenario.

where $\psi_i^{n+1} > \psi_i^n$. That is, the event that the job seeker applies to n jobs depends on the realization of two consecutive order statistics. Instead of explicitly integrating over the joint distribution of the order statistics of ex post utilities, we can directly derive an expression for the probability that $u_{i(J-n+1)} \geq \psi_i^n$ and $u_{i(J-n)} < \psi_i^{n+1}$ by considering the following combinatorial arguments.

To find the probability measure of the set of all realizations of the expost utilities of a job seeker such that the $(J-n)$–th and $(J-n+1)$–th order statistics satisfy $u_{i(J-n+1)} \geq \psi_i^n$ and $u_{i(J-n)} < \psi_i^{n+1}$, we can partition this set according to how many realizations lie in the interval $[\psi_i^n, \psi_i^{n+1})$. Since the resulting subsets are disjoint events, we need simply compute the sum of the probabilities of each event in the partition. Figure C.1 below depicts the configurations of the order statistics that obtain for different sets in this partition.

Let s be the number of realizations in $[\psi_i^n, \psi_i^{n+1})$. As can be seen in Panel (a), the event $s = 0$ in our partition only includes realizations of random vector $\boldsymbol{u}_i$ such that exactly $J-n$ elements lie below ψ_i^n and the remaining n elements lie above ψ_i^{n+1}. The probability of this subset can be obtained by considering all possible combinations of $J-n$ alternatives and computing the probability that the utilities of these alternatives are less than ψ_i^n and the utilities of the remaining alternatives are larger than ψ_i^{n+1}, i.e.,

$$\sum_{B \in \mathcal{R}_{J-n}(\mathcal{J})} \prod_{p \in B} F_{u_p}(\psi_i^n) \prod_{q \in \mathcal{J} \setminus B} \left[1 - F_{u_q}(\psi_i^{n+1})\right].$$

Panel (b) of Figure C.1 illustrates the element of the partition where $s = J$. This case includes all realizations of $\boldsymbol{u}_i$ such that every element lies in $[\psi_i^n, \psi_i^{n+1})$. The probability of this subset is simply the probability that the utility of every alternative lies in $[\psi_i^n, \psi_i^{n+1})$ since there is only one combination of size J from $\mathcal{J}$ —i.e., $\mathcal{R}_J(\mathcal{J}) = \mathcal{R}_{|\mathcal{J}|}(\mathcal{J}) = \{\mathcal{J}\}$. The corresponding expression is

$$\prod_{\ell \in \mathcal{J}} \left[F_{u_\ell}(\psi_i^{n+1}) - F_{u_\ell}(\psi_i^n)\right].$$

Finally, for the cases $s \in \{1, \ldots, J-1\}$ depicted in panel (c), let $u_{i(r)}$ represent the smallest order statistic that lies in $[\psi_i^n, \psi_i^{n+1})$. Note that $u_{i(J-n)} < \psi_i^{n+1}$ implies the largest order statistic in $[\psi_i^n, \psi_i^{n+1})$ is at least the $(J-n)$–th, while $u_{i(J-n+1)} \geq \psi_i^n$ implies the smallest order statistic in $[\psi_i^n, \psi_i^{n+1})$ is at most the $(J-n+1)$–th. Therefore, r must satisfy $r + s - 1 \geq J - n$ and $r \leq J - n + 1$. Since the number of elements of $\boldsymbol{u}_i$ that lie in $(-\infty, \psi_i^n)$ is $r - 1$ and there are only $J - s$ elements that lie outside $[\psi_i^n, \psi_i^{n+1})$, the probability of the s–th subset in the partition can be obtained by (i) considering all combinations of size s of the J alternatives, $A \in \mathcal{R}_s(\mathcal{J})$, (ii) considering all the combinations of size $t \in \{\max(J-n-s, 0), \ldots, \min(J-n, J-s)\}$ of the remaining $J-s$ alternatives, $B \in \mathcal{R}_t(\mathcal{J} \setminus A)$, and (iii) computing the probability that the utilities of the alternatives in A lie between ψ_i^n and ψ_i^{n+1}, the utilities of the alternatives in B lie below ψ_i^n, and the remaining alternatives in $\mathcal{J} \setminus (A \cup B)$ have utilities larger than ψ_i^{n+1}. The corresponding expression is

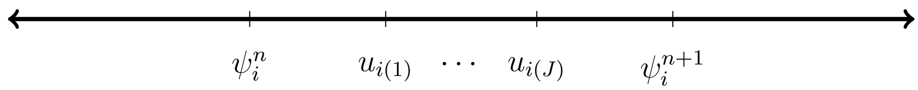
$$\sum_{A \in \mathcal{R}_s(\mathcal{J})} \prod_{\ell \in A} \left[F_{u_\ell}(\psi_i^{n+1}) - F_{u_\ell}(\psi_i^n)\right] \sum_{t \in \tau_n^s} \sum_{B \in \mathcal{R}_t(\mathcal{J} \setminus A)} \prod_{p \in B} F_{u_p}(\psi_i^n) \prod_{q \in \mathcal{J} \setminus (A \cup B)} \left[1 - F_{u_q}(\psi_i^{n+1})\right],$$

where $\tau_n^s \equiv \{\max(J-n-s, 0), \ldots, \min(J-n, J-s)\}$. Summing over all values of s, we obtain

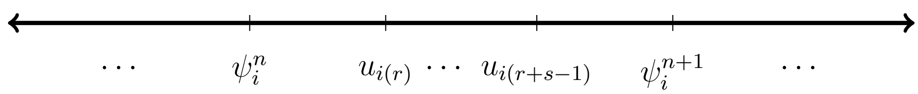
$$\begin{aligned} s_{n|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i) &= \mathbb{P}\left(n_i = n \mid \alpha_i, \gamma_i\right) \\ &= \sum_{s=0}^{J} \sum_{A \in \mathcal{R}_s(\mathcal{J})} \prod_{\ell \in A} \left[F_{u_\ell}\big(\psi_i^{n+1}\big) - F_{u_\ell}\big(\psi_i^n\big)\right] \\ &\qquad \sum_{t \in \tau_n^s} \sum_{B \in \mathcal{R}_t(\mathcal{J} \setminus A)} \prod_{p \in B} F_{u_p}\big(\psi_i^n\big) \prod_{q \in \mathcal{J} \setminus (A \cup B)} \Big[1 - F_{u_q}\big(\psi_i^{n+1}\big)\Big] \\ &= \sum_{s=0}^{J} \sum_{A \in \mathcal{R}_s(\mathcal{J})} \prod_{\ell \in A} \Big[F_\varepsilon\big(\psi_i^{n+1}\big)^{\exp(\delta_\ell)} - F_\varepsilon\big(\psi_i^n\big)^{\exp(\delta_\ell)}\Big] \\ &\qquad \sum_{t \in \tau_n^s} \sum_{B \in \mathcal{R}_t(\mathcal{J} \setminus A)} F_\varepsilon\big(\psi_i^n\big)^{\sum_{p \in B} \exp(\delta_p)} \prod_{q \in \mathcal{J} \setminus (A \cup B)} \Big[1 - F_\varepsilon\big(\psi_i^{n+1}\big)^{\exp(\delta_q)}\Big]. \end{aligned} \tag{C.11}$$

Figure C.1: Realizations of the order statistics consistent with n applications

(a) No realizations of u_{ij} lie between ψ_i^n and ψ_i^{n+1}



(b) All realizations of u_{ij} lie between ψ_i^n and ψ_i^{n+1}



(c) $s \in \{1, \ldots, J-1\}$ realizations of u_{ij} lie between ψ_i^n and ψ_i^{n+1}

Notes: This figure depicts the realizations of the order statistics of utilities u_{ij} that are consistent with the job seeker applying to n jobs according to the stopping rule in part (i) of Proposition 1. That is, $u_{i(J-n)} < \psi_i^{n+1}$ and $u_{i(J-n+1)} \geq \psi_i^n$ for $n \in \{1, \ldots, J-1\}$. The thresholds ψ_i^n and ψ_i^{n+1} are defined in Equation (13). Cases are indexed by the number of realizations of u_{ij} in the interval $\left[\psi_i^n, \psi_i^{n+1}\right)$, $s \in \{0, \ldots, J\}$. The case $s = 0$ in Panel (a) is equivalent to exactly $J - n$ realizations of u_{ij} below ψ_i^n and exactly n above ψ_i^{n+1}. The case $s = J$ in Panel (b) is equivalent to exactly J realizations of u_{ij} between ψ_i^n and ψ_i^{n+1}. For the cases $s \in \{1, \ldots, J-1\}$ in Panel (c), r must satisfy $r \leq J - n + 1$ so that $u_{i(J-n+1)} \geq \psi_i^n$, and $r + s - 1 \geq J - n$ so $u_{i(J-n)} < \psi_i^{n+1}$, where $u_{i(r)}$ is the smallest order statistic that lies between ψ_i^n and ψ_i^{n+1}. Then, we have at least $\max(J - n - s, 0)$ and at most $\min(J - n, J - s)$ realizations below ψ_i^n, with the remaining realizations above ψ_i^{n+1}.

D Other proofs and derivations

D.1 The wage elasticity of the job applications supply

The elasticity of the applications supply to job $j \in \mathcal{J}$ with respect to the wage of job $\ell \in \mathcal{J}$ is

$$\begin{aligned}
\eta_{q_j,w_\ell} &= \frac{\partial \ln\left(q_j(\boldsymbol{\delta})\right)}{\partial \ln\left(w_\ell\right)} \\
&= \frac{1}{q_j(\boldsymbol{\delta})}\frac{\partial q_j(\boldsymbol{\delta})}{\partial \ln\left(w_\ell\right)} \\
&= \frac{1}{q_j(\boldsymbol{\delta})}\frac{\partial q_j(\boldsymbol{\delta})}{\partial \delta_\ell}\frac{\partial \delta_\ell}{\partial \ln\left(w_\ell\right)} \\
&= \frac{1}{q_j(\boldsymbol{\delta})}\left[I\sum_{n=1}^{J}\frac{\partial \mathscr{s}_{j|n}(\boldsymbol{\delta})}{\partial \delta_\ell}\mathscr{s}_n(\boldsymbol{\delta}) + \mathscr{s}_{j|n}(\boldsymbol{\delta})\frac{\partial \mathscr{s}_n(\boldsymbol{\delta})}{\partial \delta_\ell}\right]\beta, \qquad \text{(D.1)}
\end{aligned}$$

where the last equality follows from partially differentiating Equation (10) with respect to δ_ℓ and Equation (8) —for job ℓ— with respect to $\ln(w_\ell)$.

The elasticity of the aggregate supply of applications at the firm level is

$$\begin{aligned}
\eta_{q^f,w_\ell} &= \frac{\partial \ln\left(q^f(\boldsymbol{\delta})\right)}{\partial \ln\left(w_\ell\right)} \\
&= \frac{1}{q^f(\boldsymbol{\delta})}\frac{\partial q^f(\boldsymbol{\delta})}{\partial \ln\left(w_\ell\right)} \\
&= \frac{1}{q^f(\boldsymbol{\delta})}\sum_{j\in\mathcal{J}^f}\frac{\partial q_j(\boldsymbol{\delta})}{\partial \ln\left(w_\ell\right)} \\
&= \frac{1}{q^f(\boldsymbol{\delta})}\sum_{j\in\mathcal{J}^f} q_j(\boldsymbol{\delta})\,\eta_{q_j,w_\ell}, \qquad \text{(D.2)}
\end{aligned}$$

where the third equality follows from (18), and the last equality follows from the definition of the vacancy-level elasticity.

Finally, the elasticity of the firm-level supply of applications with respect to a simultaneous increase of the wages the firm offers for all its vacancies, $\boldsymbol{w}^f = \{w_\ell\}_{\ell\in\mathcal{J}^f}$, is given by

$$\eta_{q^f,\boldsymbol{w}^f} = \frac{1}{q^f(\boldsymbol{\delta})}\sum_{\ell\in\mathcal{J}^f}\sum_{j\in\mathcal{J}^f} q_j(\boldsymbol{\delta})\,\eta_{q_j,w_\ell}. \qquad \text{(D.3)}$$

D.2 Closed-form derivatives

We can obtain closed-form solutions for the partial derivatives of the conditional share $\mathcal{s}_{j|n}(\boldsymbol{\delta})$ and the conditional pmf $\mathcal{s}_{n|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i)$ with respect to δ_ℓ for $n \in \{1, \ldots, J\}$. The partial derivative of the unconditional pmf, $\mathcal{s}_n(\boldsymbol{\delta})$, with respect to δ_ℓ is then obtained by integrating the partial of the conditional pmf over $F_\alpha(\cdot) \times F_\gamma(\cdot)$:

$$
\begin{aligned}
\frac{\partial \mathcal{s}_n(\boldsymbol{\delta})}{\partial \delta_\ell} &= \frac{\partial}{\partial \delta_\ell} \int \int \mathcal{s}_{n|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i) \mathrm{d}F_\alpha(\alpha_i) \mathrm{d}F_\gamma(\gamma_i) \\
&= \int \int \frac{\partial \mathcal{s}_{n|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i)}{\partial \delta_\ell} \mathrm{d}F_\alpha(\alpha_i) \mathrm{d}F_\gamma(\gamma_i), \qquad \text{(D.4)}
\end{aligned}
$$

where the second equality follows since $\mathcal{s}_{n|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i)$ is continuously differentiable in δ_ℓ and the supports of F_α and F_γ do not depend on δ_ℓ.[17]

To find the partial derivatives of $\mathcal{s}_{j|n}(\boldsymbol{\delta})$, fix $(j, n) \in \mathcal{J} \times \{1, \ldots, J\}$ and let

$$
E_\ell(S) = \frac{\exp(\delta_\ell)}{\sum_{k \in \{j\} \cup S} \exp(\delta_k)}, \qquad \text{(D.5)}
$$

for $\ell \in \mathcal{J}$ and $S \subseteq \mathcal{B}_j$. Note that the expression on the right-hand side of Equation (11) is a finite sum of terms —some with a negative sign— of the form $E_j(S)$ for different subsets S of the choice set that do not contain j. Each such term has partial derivative with respect to δ_ℓ

$$
\frac{\partial E_j(S)}{\partial \delta_\ell} = \begin{cases} E_j(S)\big[1 - E_j(S)\big] & \text{if } \ell = j \\[2ex] -\mathbb{1}\{\ell \in S\} E_j(S) E_\ell(S) & \text{otherwise} \end{cases}. \qquad \text{(D.6)}
$$

Thus,

$$
\begin{aligned}
\frac{\partial \mathcal{s}_{j|n}(\boldsymbol{\delta})}{\partial \delta_j} &= \sum_{k=J-n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} \frac{\partial E_j(A)}{\partial \delta_j} + \sum_{s=1}^{|\mathcal{B}_j \setminus A|} (-1)^s \sum_{B \in \mathcal{R}_s(\mathcal{B}_j \setminus A)} \frac{\partial E_j(A \cup B)}{\partial \delta_j} \\
&= \sum_{k=J-n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} E_j(A)\big[1 - E_j(A)\big] + \sum_{s=1}^{|\mathcal{B}_j \setminus A|} (-1)^s \sum_{B \in \mathcal{R}_s(\mathcal{B}_j \setminus A)} E_j(A \cup B)\big[1 - E_j(A \cup B)\big]
\end{aligned}
\qquad \text{(D.7)}
$$

and, similarly for $\ell \neq j$,

$$
\frac{\partial \mathcal{s}_{j|n}(\boldsymbol{\delta})}{\partial \delta_\ell} = \sum_{k=J-n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} \frac{\partial E_j(A)}{\partial \delta_\ell} + \sum_{s=1}^{|\mathcal{B}_j \setminus A|} (-1)^s \sum_{B \in \mathcal{R}_s(\mathcal{B}_j \setminus A)} \frac{\partial E_j(A \cup B)}{\partial \delta_\ell}
$$

[17] The function $\mathcal{s}_{n|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i)$ is a finite sum of products of terms of the form $F_\varepsilon(\psi_i^{n+1})^{\exp(\delta_k)} - F_\varepsilon(\psi_i^n)^{\exp(\delta_k)}$, $F_\varepsilon(\psi_i^n)^{\exp(\delta_k)}$, or $\big[1 - F_\varepsilon(\psi_i^{n+1})^{\exp(\delta_k)}\big]$. Each of these factors is uniformly bounded since $F_\varepsilon(x) \in [0, 1]$ for all $x \in \mathbb{R}$ and the thresholds $\{\psi_i^n\}_{n=1}^J$ do not depend on $\boldsymbol{\delta}$.

$$= -\sum_{k=J-n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} \Bigg[\mathbb{1}\left\{\ell \in A\right\} E_j(A) E_\ell(A) + \sum_{s=1}^{|\mathcal{B}_j \setminus A|} (-1)^s \sum_{B \in \mathcal{R}_s(\mathcal{B}_j \setminus A)} \mathbb{1}\left\{\ell \in A \cup B\right\} E_j(A \cup B) E_\ell(A \cup B) \Bigg]. \tag{D.8}$$

To find the partial derivative of $\mathscr{s}_{n|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i)$ with respect to δ_ℓ for $n \in \{1, \dots, J-1\}$, rewrite Equation (12) as

Need to do also for $n = J$. Maybe also mention that $n = 0$ is irrelevant.

$$\mathscr{s}_{n|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i) = \sum_{s=0}^{J} \sum_{A \in \mathcal{R}_s(\mathcal{J})} \sum_{t \in \tau_n^s} \sum_{B \in \mathcal{R}_t(\mathcal{J} \setminus A)} \left[\left(\prod_{k \in A} a_k \right) \left(\prod_{p \in B} b_p \right) \left(\prod_{q \in \mathcal{J} \setminus (A \cup B)} c_q \right) \right], \tag{D.9}$$

where $a_k = (F_{n+1})^{\exp(\delta_k)} - (F_n)^{\exp(\delta_k)}$, $b_k = (F_n)^{\exp(\delta_k)}$, $c_k = 1 - (F_{n+1})^{\exp(\delta_k)}$, and $F_k = F_\varepsilon(\psi_i^k)$ for $k \in \{1, \dots, J\}$. Note that the expression on the right-hand side of (D.9) is a finite sum of products of terms of the form a_k, b_p, or c_q for k, p, and q in different, mutually exclusive subsets of $\mathcal{J}$. Since ℓ belongs to only one of these subsets, the chain rule yields

$$\begin{aligned}
\frac{\partial \mathscr{s}_{n|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i)}{\partial \delta_\ell} = \sum_{s=0}^{J} \sum_{A \in \mathcal{R}_s(\mathcal{J})} \sum_{t \in \tau_n^s} \sum_{B \in \mathcal{R}_t(\mathcal{J} \setminus A)} \Bigg[& \mathbb{1}\left\{\ell \in A\right\} \frac{\partial a_\ell}{\partial \delta_\ell} \left(\prod_{k \in A \setminus \{\ell\}} a_k \right) \left(\prod_{p \in B} b_p \right) \left(\prod_{q \in \mathcal{J} \setminus (A \cup B)} c_q \right) \\
& + \mathbb{1}\left\{\ell \in B\right\} \frac{\partial b_\ell}{\partial \delta_\ell} \left(\prod_{k \in A} a_k \right) \left(\prod_{p \in B \setminus \{\ell\}} b_p \right) \left(\prod_{q \in \mathcal{J} \setminus (A \cup B)} c_q \right) \\
& + \mathbb{1}\left\{\ell \notin A \cup B\right\} \frac{\partial c_\ell}{\partial \delta_\ell} \left(\prod_{k \in A} a_k \right) \left(\prod_{p \in B} b_p \right) \left(\prod_{q \in \mathcal{J} \setminus (A \cup B \cup \{\ell\})} c_q \right) \Bigg] \\
= \sum_{s=0}^{J} \sum_{A \in \mathcal{R}_s(\mathcal{J})} \sum_{t \in \tau_n^s} \sum_{B \in \mathcal{R}_t(\mathcal{J} \setminus A)} \Bigg[& \left(\prod_{k \in A} a_k \right) \left(\prod_{p \in B} b_p \right) \left(\prod_{q \in \mathcal{J} \setminus (A \cup B)} c_q \right) \\
& \left(\mathbb{1}\left\{\ell \in A\right\} \frac{1}{a_\ell} \frac{\partial a_\ell}{\partial \delta_\ell} + \mathbb{1}\left\{\ell \in B\right\} \frac{1}{b_\ell} \frac{\partial b_\ell}{\partial \delta_\ell} + \mathbb{1}\left\{\ell \notin A \cup B\right\} \frac{1}{c_\ell} \frac{\partial c_\ell}{\partial \delta_\ell} \right) \Bigg],
\end{aligned} \tag{D.10}$$

where $\frac{\partial a_\ell}{\partial \delta_\ell} = \exp(\delta_\ell) \left[(F_{n+1})^{\exp(\delta_\ell)} \ln(F_{n+1}) - (F_n)^{\exp(\delta_\ell)} \ln(F_n) \right]$, $\frac{\partial b_\ell}{\partial \delta_\ell} = \exp(\delta_\ell) (F_n)^{\exp(\delta_\ell)} \ln(F_n)$, and $\frac{\partial c_\ell}{\partial \delta_\ell} = -\exp(\delta_\ell) (F_{n+1})^{\exp(\delta_\ell)} \ln(F_{n+1})$.

D.3 Proof of Lemma 1

Remark. The following proof makes use of the properties of the EV_1 distribution and the ARUM structure discussed in Appendix C, which we omit here to avoid repetition.

Proof. Start by noting how Equation (3) changes when $\alpha_i = 1$. In this case, the job seeker faces no uncertainty regarding her ability to exercise any option in the application portfolio —i.e., getting the job—, but the constraint that only one can be exercised binds. Given any nonempty application portfolio $A \neq \emptyset$, only the most ex-post preferred option in the portfolio, $r_i(A,1)$, will be exercised. Thus, the von Neumann–Morgenstern utility from nonempty portfolio $A \subseteq \mathcal{J}$ is

$$U_i(A) = u_{ir_i(A,1)} - c_i(|A|). \tag{D.11}$$

For an empty portfolio, expected utility simply coincides with the ex-post Bernoulli utility of the outside option:

$$U_i(\emptyset) = -c_i(0) = 0 = u_{i0}. \tag{D.12}$$

Now, let $\gamma > 0$, set $c_i(|A|) = \gamma\,|A|$, and note that

$$\begin{aligned} U_i(A) &= u_{ir_i(A,1)} - \gamma\,|A| \\ &\leq u_{ir_i(A,1)} - \gamma \\ &= U_i\big(\,\{r_i(A,1)\}\,\big) \end{aligned}$$

for any nonempty $A \subseteq \mathcal{J}$ since $|A| \in \{1,\dots,J\}$. Therefore, conditional on applying, the optimal portfolio is a singleton. Accounting for the case $A_i = \emptyset$, we conclude $A_i \in \{0,1\}$, establishing part (i) of Lemma 1.

Next, to prove part (ii), consider the non-application margin. Notice that, conditional on applying, the optimal portfolio is the singleton containing the best ex-post alternative:

$$\underset{A \in \{\sigma \subseteq \mathcal{J} : |\sigma| > 0\}}{\arg\max}\; U_i(A) = \{r_i(\mathcal{J},1)\}\,.$$

Not applying —i.e., choosing the outside option— is optimal if and only if the marginal cost of applications exceeds the highest ex-post utility among the inside alternatives:

$$A_i = \emptyset \iff U_i\big(\,\{r_i(\mathcal{J},1)\}\,\big) < U_i(\emptyset) \iff u_{ir_i(\mathcal{J},1)} - \gamma < 0.$$

This event has probability

$$\begin{aligned} \mathbb{P}\left(\max_{\ell \in \mathcal{J}} u_{i\ell} < \gamma\right) &= F_\varepsilon\left(\gamma - \ln\left(\sum_{\ell \in \mathcal{J}} \exp\left(\delta_\ell\right)\right)\right) \\ &= \exp\left(-\exp\left(-\gamma\right)\sum_{\ell \in \mathcal{J}} \exp\left(\delta_\ell\right)\right), \end{aligned}$$

establishing part (ii) of Lemma 1.

Finally, for any $j \in \mathcal{J}$, note that

$$\mathbb{P}\left(A_i = \{j\} \mid A_i \neq \emptyset\right) = \mathbb{P}\left(\max_{\ell \in \mathcal{J}} u_{i\ell} \leq u_{ij}\right)$$

$$= \frac{\exp\left(\delta_j\right)}{\sum_{\ell \in \mathcal{J}} \exp\left(\delta_\ell\right)},$$

which establishes part (iii) of Lemma 1. $\square$

E Minorize-maximize algorithm

This appendix closely follows Appendix D of Roussille and Scuderi (2025). The likelihood contribution of job seeker i can be written as

$$\begin{aligned}
f_i\left(\boldsymbol{\delta} \mid \mathcal{A}_i\right) &= \mathbb{P}\left(\bigcap_{j \in A_i, \ell \in \overline{A}_i} \left\{\delta_j + \varepsilon_{ij} > \delta_\ell + \varepsilon_{i\ell}\right\}\right) \\
&= \mathbb{P}\left(\bigcap_{j \in A_i} \left\{\delta_j + \varepsilon_{ij} > \max_{\ell \in \overline{A}_i} \delta_\ell + \varepsilon_{i\ell}\right\}\right) \\
&= \int_{-\infty}^{\infty} \left(\prod_{j \in A_i} 1 - F_\varepsilon\left(x - \delta_j\right)\right) \mathrm{d}F_\varepsilon\left(x - \ln\left(\sum_{\ell \in \overline{A}_i} \exp\left(\delta_\ell\right)\right)\right) \\
&= \int_{-\infty}^{\infty} \left(\prod_{j \in A_i} 1 - F_\varepsilon\left(x - \ln\left(\sum_{\ell \in \overline{A}_i} \exp\left(\delta_\ell\right)\right)\right)^{\frac{\exp\left(\delta_j\right)}{\sum_{\ell \in \overline{A}_i} \exp\left(\delta_\ell\right)}}\right) \mathrm{d}F_\varepsilon\left(x - \ln\left(\sum_{\ell \in \overline{A}_i} \exp\left(\delta_\ell\right)\right)\right) \\
&= \int_0^1 \left(\prod_{j \in A_i} 1 - u^{\frac{\exp\left(\delta_j\right)}{\sum_{\ell \in \overline{A}_i} \exp\left(\delta_\ell\right)}}\right) \mathrm{d}u \\
&= \int_0^1 \left(\prod_{j \in A_i} 1 - z^{\exp\left(\delta_j\right)}\right) \left(\sum_{\ell \in \overline{A}_i} \exp\left(\delta_\ell\right)\right) z^{\sum_{\ell \in \overline{A}_i} \exp\left(\delta_\ell\right) - 1} \mathrm{d}z, \qquad \text{(E.1)}
\end{aligned}$$

where $\mathcal{A}_i = \left\{A_i, \overline{A}_i\right\}$ is job seeker i's partition of the choice set into chosen and unchosen alternatives and $F_\varepsilon(x) = \exp\left(-\exp\left(-x\right)\right)$ is the cdf of the EV_1 distribution. The second equality follows from the equivalence of the corresponding events, the third equality follows from the assumption of independent observations and the fact that $\left\{\varepsilon_{ij}\right\}_{j \in \mathcal{J}} \overset{iid}{\sim} \mathrm{EV}_1 \implies \mathbb{P}\left(\max_{\ell \in \overline{A}_i} \delta_\ell + \varepsilon_{i\ell} \leq x\right) = F_\varepsilon\left(x - \ln\left(\sum_{\ell \in \overline{A}_i} \exp\left(\delta_\ell\right)\right)\right)$, the fourth equality uses the fact that $F_\varepsilon\left(x - \ln\left(a\right)\right) = F_\varepsilon\left(x - \ln\left(b\right)\right)^{a/b}$ for $a, b > 0$, the fifth equality applies the change of variable $u = F_\varepsilon\left(x - \ln\left(\sum_{\ell \in \overline{A}_i} \exp\left(\delta_\ell\right)\right)\right)$, and the last equality makes the change of variable $z = u^{1/\sum_{\ell \in \overline{A}_i} \exp\left(\delta_\ell\right)}$. Numerical evaluation of the resulting integral allows us to avoid iterating over all the permutations of the application portfolio A_i to break ties, which becomes an increasingly demanding computational task as the number of alternatives grows.

Given the *iid* assumption, the log-likelihood function takes the form

$$\ell\left(\boldsymbol{\delta} \mid \{\mathcal{A}_i\}_{i\in\mathcal{I}}\right) = \sum_{i\in\mathcal{I}} \ln\Big(f_i\left(\boldsymbol{\delta} \mid \mathcal{A}_i\right)\Big),$$

which could be directly maximized using the expression in Equation (E.1).[18] Instead, we gain some computational speed by implementing a minorize-maximize (MM) algorithm based on monotonically increasing a suitable surrogate function satisfying an ascent property that guarantees monotonic increases of the objective function.[19]

Let $\boldsymbol{\delta}^{(n)}$ represent the current iterate in our MM algorithm. A *minorizing function* of the real-valued function $\ell\left(\boldsymbol{\delta}\right)$ at the point $\boldsymbol{\delta}^{(n)}$ is any function $g\left(\boldsymbol{\delta} \mid \boldsymbol{\delta}^{(n)}\right)$ satisfying

$$g\left(\boldsymbol{\delta} \mid \boldsymbol{\delta}^{(n)}\right) \leq \ell\left(\boldsymbol{\delta}\right),\ \forall\,\boldsymbol{\delta}$$

$$g\left(\boldsymbol{\delta}^{(n)} \mid \boldsymbol{\delta}^{(n)}\right) = \ell\left(\boldsymbol{\delta}^{(n)}\right).$$

Note that if our iterative procedure is such that $g\left(\boldsymbol{\delta}^{(n+1)} \mid \boldsymbol{\delta}^{(n)}\right) \geq g\left(\boldsymbol{\delta}^{(n)} \mid \boldsymbol{\delta}^{(n)}\right)$ —i.e., each iteration (weakly) increases the corresponding surrogate minorizing function—, then

$$\begin{aligned}
\ell\left(\boldsymbol{\delta}^{(n+1)}\right) &\geq g\left(\boldsymbol{\delta}^{(n+1)} \mid \boldsymbol{\delta}^{(n)}\right) \\
&\geq g\left(\boldsymbol{\delta}^{(n)} \mid \boldsymbol{\delta}^{(n)}\right) \\
&= \ell\left(\boldsymbol{\delta}^{(n)}\right),
\end{aligned}$$

where the first inequality follows from the definition of $g\left(\cdot \mid \boldsymbol{\delta}^{(n)}\right)$ as a minorizing function of $\ell\left(\cdot\right)$ at $\boldsymbol{\delta}^{(n)}$, the second inequality is our assumption, and the equality follows again from the definition of a minorizing function. This ascent property of minorizing functions guarantees that MM algorithms force the objective function uphill.

MM algorithms typically construct a suitable surrogate minorizing function at the current iterate and then maximize it to obtain the next iterate, i.e.,

$$\boldsymbol{\delta}^{(n+1)} = \arg\max_{\boldsymbol{\delta}}\ g\left(\boldsymbol{\delta} \mid \boldsymbol{\delta}^{(n)}\right),$$

leading to significant computational efficiency gains when the surrogate is easy to maximize. However, the ascent property only requires *increasing* the surrogate function, as shown above. Consequently, we follow Roussille and Scuderi (2025) in replacing full maximization in the 'maximization' step with a single gradient ascent update.

[18] For notational simplicity, we hereafter suppress the dependence of the likelihood function on the data $\{\mathcal{A}_i\}_{i\in\mathcal{I}}$.

[19] See Wu and Lange (2010) for an introduction to MM algorithms.

To construct our minorizing surrogate of the log-likelihood function at $\boldsymbol{\delta}^{(n)}$, we start by defining

$$\rho_i\left(\delta_j \mid \boldsymbol{\delta}^{(n)}\right) = \frac{\exp\left(\delta_j\right)}{\sum_{\ell \in \overline{A}_i} \exp\left(\delta_\ell^{(n)}\right)},$$

$$\varphi_i\left(\boldsymbol{\delta}, z \mid \boldsymbol{\delta}^{(n)}\right) = \left(\prod_{j \in A_i} 1 - z^{\rho_i\left(\delta_j \mid \boldsymbol{\delta}^{(n)}\right)}\right) \sum_{\ell \in \overline{A}_i} \rho_i\left(\delta_\ell \mid \boldsymbol{\delta}^{(n)}\right) z^{\sum_{\ell \in \overline{A}_i} \rho_i\left(\delta_\ell \mid \boldsymbol{\delta}^{(n)}\right) - 1},$$

$$\pi_i\left(z \mid \boldsymbol{\delta}^{(n)}\right) = \frac{\varphi_i\left(\boldsymbol{\delta}^{(n)}, z \mid \boldsymbol{\delta}^{(n)}\right)}{\int_0^1 \varphi_i\left(\boldsymbol{\delta}^{(n)}, x \mid \boldsymbol{\delta}^{(n)}\right) \mathrm{d}x},$$

and noting that

$$\frac{f_i\left(\boldsymbol{\delta}\right)}{f_i\left(\boldsymbol{\delta}^{(n)}\right)} = \int_0^1 \frac{\varphi_i\left(\boldsymbol{\delta}, z \mid \boldsymbol{\delta}^{(n)}\right)}{\varphi_i\left(\boldsymbol{\delta}^{(n)}, z \mid \boldsymbol{\delta}^{(n)}\right)} \pi_i\left(z \mid \boldsymbol{\delta}^{(n)}\right) \mathrm{d}z,$$

which follows from the fact that $f_i\left(\boldsymbol{\delta} + \alpha \boldsymbol{\iota}\right) = f_i\left(\boldsymbol{\delta}\right) \; \forall \, \alpha \in \mathbb{R}$ and choosing $\alpha = -\ln\left(\sum_{\ell \in \overline{A}_i} \exp\left(\delta_\ell^{(n)}\right)\right)$, where $\boldsymbol{\iota}$ is a vector of ones. Since $\pi_i\left(z \mid \boldsymbol{\delta}^{(n)}\right) \geq 0$ and $\int_0^1 \pi_i\left(z \mid \boldsymbol{\delta}^{(n)}\right) \mathrm{d}z = 1$, applying Jensen's inequality yields

$$\ln\left(\int_0^1 \frac{\varphi_i\left(\boldsymbol{\delta}, z \mid \boldsymbol{\delta}^{(n)}\right)}{\varphi_i\left(\boldsymbol{\delta}^{(n)}, z \mid \boldsymbol{\delta}^{(n)}\right)} \pi_i\left(z \mid \boldsymbol{\delta}^{(n)}\right) \mathrm{d}z\right) \geq \int_0^1 \ln\left(\frac{\varphi_i\left(\boldsymbol{\delta}, z \mid \boldsymbol{\delta}^{(n)}\right)}{\varphi_i\left(\boldsymbol{\delta}^{(n)}, z \mid \boldsymbol{\delta}^{(n)}\right)}\right) \pi_i\left(z \mid \boldsymbol{\delta}^{(n)}\right) \mathrm{d}z$$

$$\iff \ell_i\left(\boldsymbol{\delta}\right) \geq \ell_i\left(\boldsymbol{\delta}^{(n)}\right) + \int_0^1 \ln\left(\frac{\varphi_i\left(\boldsymbol{\delta}, z \mid \boldsymbol{\delta}^{(n)}\right)}{\varphi_i\left(\boldsymbol{\delta}^{(n)}, z \mid \boldsymbol{\delta}^{(n)}\right)}\right) \pi_i\left(z \mid \boldsymbol{\delta}^{(n)}\right) \mathrm{d}z, \tag{E.2}$$

where $\ell_i\left(\boldsymbol{\delta}\right) = \ln\left(f_i\left(\boldsymbol{\delta}\right)\right)$ is the log-likelihood contribution of observation i. We obtain our first minorization of this log-likelihood contribution by defining

$$H_{\pi i}^{(n)} = -\int_0^1 \ln\left(\pi_i\left(z \mid \boldsymbol{\delta}^{(n)}\right)\right) \pi_i\left(z \mid \boldsymbol{\delta}^{(n)}\right) \mathrm{d}z$$

and rewriting (E.2) as

$$\ell_i\left(\boldsymbol{\delta}\right) \geq H_{\pi i}^{(n)} + \int_0^1 \ln\left(\varphi_i\left(\boldsymbol{\delta}, z \mid \boldsymbol{\delta}\right)^{(n)}\right) \pi_i\left(z \mid \boldsymbol{\delta}^{(n)}\right) \mathrm{d}z, \tag{E.3}$$

which holds with equality at $\boldsymbol{\delta} = \boldsymbol{\delta}^{(n)}$. We can improve on this minorization to obtain a surrogate function that is separable in $\boldsymbol{\delta}$ by noting that

$$\ln\left(\varphi_i\left(\boldsymbol{\delta}, z \mid \boldsymbol{\delta}\right)^{(n)}\right) = \sum_{j \in A_i} \ln\left(1 - z^{\rho_i\left(\delta_j \mid \boldsymbol{\delta}^{(n)}\right)}\right) + \ln\left(\sum_{\ell \in \overline{A}_i} \rho_i\left(\delta_\ell \mid \boldsymbol{\delta}^{(n)}\right)\right) + \left(\sum_{\ell \in \overline{A}_i} \rho_i\left(\delta_\ell \mid \boldsymbol{\delta}^{(n)}\right) - 1\right) \ln\left(z\right) \tag{E.4}$$

and

$$
\begin{aligned}
\ln\left(\sum_{\ell\in\overline{A}_i}\rho_i\left(\delta_\ell\mid\boldsymbol{\delta}^{(n)}\right)\right) &= \ln\left(\sum_{\ell\in\overline{A}_i}\frac{\exp\left(\delta_\ell\right)}{\exp\left(\delta_\ell^{(n)}\right)}\rho_i\left(\delta_\ell^{(n)}\mid\boldsymbol{\delta}^{(n)}\right)\right)\\
&\geq \sum_{\ell\in\overline{A}_i}\ln\left(\frac{\exp\left(\delta_\ell\right)}{\exp\left(\delta_\ell^{(n)}\right)}\right)\rho_i\left(\delta_\ell^{(n)}\mid\boldsymbol{\delta}^{(n)}\right)
\end{aligned}
$$

$$
\iff \ln\left(\sum_{\ell\in\overline{A}_i}\exp\left(\delta_\ell\right)\right)\geq\sum_{\ell\in\overline{A}_i}\delta_\ell\,\rho\left(\delta_\ell^{(n)}\mid\boldsymbol{\delta}^{(n)}\right)+H_{\rho i}^{(n)}, \tag{E.5}
$$

where $H_{\rho i}^{(n)}=-\sum_{\ell\in\overline{A}_i}\rho_i\left(\delta_\ell^{(n)}\mid\boldsymbol{\delta}^{(n)}\right)\ln\left(\rho_i\left(\delta_\ell^{(n)}\mid\boldsymbol{\delta}^{(n)}\right)\right)$ and the inequality follows from Jensen's inequality since $\rho_i\left(\delta_\ell^{(n)}\mid\boldsymbol{\delta}^{(n)}\right)\geq 0$ and $\sum_{\ell\in\overline{A}_i}\rho_i\left(\delta_\ell^{(n)}\mid\boldsymbol{\delta}^{(n)}\right)=1$. Notice that (E.5) holds with equality at $\boldsymbol{\delta}=\boldsymbol{\delta}^{(n)}$. Finally, combining with (E.3) and (E.4) yields

$$
\ell_i\left(\boldsymbol{\delta}\right)\geq H_i^{(n)}+g_i\left(\boldsymbol{\delta}\mid\boldsymbol{\delta}^{(n)}\right), \tag{E.6}
$$

where

$$
\begin{aligned}
g_i\left(\boldsymbol{\delta}\mid\boldsymbol{\delta}^{(n)}\right) &= \int_0^1\sum_{j\in A_i}\ln\left(1-z^{\rho_i\left(\delta_j\mid\boldsymbol{\delta}^{(n)}\right)}\right)\pi_i\left(z\mid\boldsymbol{\delta}^{(n)}\right)\mathrm{d}z+\sum_{\ell\in\overline{A}_i}\delta_\ell\,\rho_i\left(\delta_\ell^{(n)}\mid\boldsymbol{\delta}^{(n)}\right)\\
&\quad+\sum_{\ell\in\overline{A}_i}\rho_i\left(\delta_\ell\mid\boldsymbol{\delta}^{(n)}\right)\int_0^1\ln\left(z\right)\pi_i\left(z\mid\boldsymbol{\delta}^{(n)}\right)\mathrm{d}z,
\end{aligned}
$$

$$
H_i^{(n)}=H_{\pi i}^{(n)}+H_{\rho i}^{(n)}-\int_0^1\ln\left(z\right)\pi_i\left(z\mid\boldsymbol{\delta}^{(n)}\right)\mathrm{d}z,
$$

and (E.6) holds with equality at $\boldsymbol{\delta}=\boldsymbol{\delta}^{(n)}$. Thus, the log-likelihood function $\ell\left(\boldsymbol{\delta}\right)=\sum_{i\in\mathcal{I}}\ell_i\left(\boldsymbol{\delta}\right)$ is minorized at $\boldsymbol{\delta}^{(n)}$ by the surrogate function

$$
g\left(\boldsymbol{\delta}\mid\boldsymbol{\delta}^{(n)}\right)=H^{(n)}+\sum_{i\in\mathcal{I}}g_i\left(\boldsymbol{\delta}\mid\boldsymbol{\delta}^{(n)}\right), \tag{E.7}
$$

where $H^{(n)}=\sum_{i\in\mathcal{I}}H_i^{(n)}$.

In its n^{th} iteration, our MM algorithm looks for $\boldsymbol{\delta}^{(n+1)}$ such that $g\left(\boldsymbol{\delta}^{(n+1)}\mid\boldsymbol{\delta}^{(n)}\right)\geq g\left(\boldsymbol{\delta}^{(n)}\mid\boldsymbol{\delta}^{(n)}\right)$, producing an increase in the log-likelihood function by the ascent property. Notice that increasing $\sum_{i\in\mathcal{I}}g_i\left(\boldsymbol{\delta}\mid\boldsymbol{\delta}^{(n)}\right)$ is sufficient to obtain an increase in $g\left(\boldsymbol{\delta}\mid\boldsymbol{\delta}^{(n)}\right)$ since $H^{(n)}$ is constant in $\boldsymbol{\delta}$. The Newton-Raphson update for maximization of $\sum_{i\in\mathcal{I}}g_i\left(\boldsymbol{\delta}\mid\boldsymbol{\delta}^{(n)}\right)$ is given by

$$
\boldsymbol{\delta}^{(n+1)}=\boldsymbol{\delta}^{(n)}+\left(-\sum_{i\in\mathcal{I}}\frac{\partial^2 g_i\left(\boldsymbol{\delta}\mid\boldsymbol{\delta}^{(n)}\right)}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'}\right)^{-1}\left(\sum_{i\in\mathcal{I}}\frac{\partial g_i\left(\boldsymbol{\delta}\mid\boldsymbol{\delta}^{(n)}\right)}{\partial\boldsymbol{\delta}}\right)\Bigg|_{\boldsymbol{\delta}=\boldsymbol{\delta}^{(n)}}
$$

and, as mentioned above, we use only one such gradient ascent update in each iteration to obtain an increase in the objective function. The fact that $g_i\left(\boldsymbol{\delta} \mid \boldsymbol{\delta}^{(n)}\right)$ has a diagonal Hessian greatly simplifies computation of this update. The j^{th} entry of its gradient and the j^{th} diagonal element of its Hessian are respectively given by

$$\left.\frac{\partial g_i\left(\boldsymbol{\delta} \mid \boldsymbol{\delta}^{(n)}\right)}{\partial \delta_j}\right|_{\boldsymbol{\delta}=\boldsymbol{\delta}^{(n)}} = \mathbb{1}\left\{j \in A_i\right\}\left.\left(-\rho_i\left(\delta_j \mid \boldsymbol{\delta}^{(n)}\right)\int_0^1 \frac{z^{\rho_i\left(\delta_j\mid\boldsymbol{\delta}^{(n)}\right)}}{1-z^{\rho_i\left(\delta_j\mid\boldsymbol{\delta}^{(n)}\right)}}\ln(z)\pi_i\left(z \mid \boldsymbol{\delta}^{(n)}\right)\mathrm{d}z\right)\right|_{\boldsymbol{\delta}=\boldsymbol{\delta}^{(n)}}$$
$$+\mathbb{1}\left\{j \in \overline{A}_i\right\}\left.\left(\rho_i\left(\delta_j^{(n)} \mid \boldsymbol{\delta}^{(n)}\right)+\rho_i\left(\delta_j \mid \boldsymbol{\delta}^{(n)}\right)\int_0^1 \ln(z)\pi_i\left(z \mid \boldsymbol{\delta}^{(n)}\right)\mathrm{d}z\right)\right|_{\boldsymbol{\delta}=\boldsymbol{\delta}^{(n)}},$$

$$\left.\frac{\partial^2 g_i\left(\boldsymbol{\delta} \mid \boldsymbol{\delta}^{(n)}\right)}{\partial \delta_j^2}\right|_{\boldsymbol{\delta}=\boldsymbol{\delta}^{(n)}} = \mathbb{1}\left\{j \in A_i\right\}\left(-\rho_i\left(\delta_j \mid \boldsymbol{\delta}^{(n)}\right)\int_0^1 \frac{z^{\rho_i\left(\delta_j\mid\boldsymbol{\delta}^{(n)}\right)}}{1-z^{\rho_i\left(\delta_j\mid\boldsymbol{\delta}^{(n)}\right)}}\ln(z)\pi_i\left(z \mid \boldsymbol{\delta}^{(n)}\right)\mathrm{d}z\right.$$
$$\left.\left.-\rho_i\left(\delta_j \mid \boldsymbol{\delta}^{(n)}\right)^2\int_0^1 \frac{z^{\rho_i\left(\delta_j\mid\boldsymbol{\delta}^{(n)}\right)}}{\left[1-z^{\rho_i\left(\delta_j\mid\boldsymbol{\delta}^{(n)}\right)}\right]^2}\ln(z)^2\pi_i\left(z \mid \boldsymbol{\delta}^{(n)}\right)\mathrm{d}z\right)\right|_{\boldsymbol{\delta}=\boldsymbol{\delta}^{(n)}}$$
$$+\mathbb{1}\left\{j \in \overline{A}_i\right\}\left.\left(\rho_i\left(\delta_j \mid \boldsymbol{\delta}^{(n)}\right)\int_0^1 \ln(z)\pi_i\left(z \mid \boldsymbol{\delta}^{(n)}\right)\mathrm{d}z\right)\right|_{\boldsymbol{\delta}=\boldsymbol{\delta}^{(n)}}.$$

Therefore, since the Hessian is diagonal, the gradient ascent update for the j^{th} component of $\boldsymbol{\delta}^{(n)}$ takes the form

$$\delta_j^{(n+1)} = \delta_j^{(n)} + \frac{\sum_{i\in\mathcal{I}} \rho_{ij}^{(n)}\kappa_{ij}^{(n)}}{\sum_{i\in\mathcal{I}} \rho_{ij}^{(n)}\lambda_{ij}^{(n)}}, \tag{E.8}$$

where

$$\kappa_{ij}^{(n)} = \begin{cases} -\int_0^1 \frac{z^{\rho_{ij}^{(n)}}}{1-z^{\rho_{ij}^{(n)}}}\ln(z)\pi_i^{(n)}(z)\,\mathrm{d}z & \text{if } j \in A_i \\ \\ 1+\int_0^1 \ln(z)\pi_i^{(n)}(z)\,\mathrm{d}z & \text{if } j \in \overline{A}_i \end{cases},$$

$$\lambda_{ij}^{(n)} = \begin{cases} \int_0^1 \frac{z^{\rho_{ij}^{(n)}}}{1-z^{\rho_{ij}^{(n)}}}\ln(z)\pi_i^{(n)}(z)\,\mathrm{d}z + \rho_{ij}^{(n)}\int_0^1 \frac{z^{\rho_{ij}^{(n)}}}{\left(1-z^{\rho_{ij}^{(n)}}\right)^2}\ln(z)^2\pi_i^{(n)}(z)\,\mathrm{d}z & \text{if } j \in A_i \\ \\ -\int_0^1 \ln(z)\pi_i^{(n)}(z)\,\mathrm{d}z & \text{if } j \in \overline{A}_i \end{cases},$$

$\rho_{ij}^{(n)} = \rho_i\left(\delta_j^{(n)} \mid \boldsymbol{\delta}^{(n)}\right)$, $\pi_i^{(n)}(z) = \pi_i\left(z \mid \boldsymbol{\delta}^{(n)}\right)$, and all the integrals involved can be approximated by numerical quadrature. Finally, since the level of $\boldsymbol{\delta}$ is not identified, we impose the normalizations $\left|\left|\boldsymbol{\delta}^{(0)}\right|\right| = 1$ and $\sum_{j\in\mathcal{J}}\exp\left(\delta_j^{(N+1)}\right) = 1$ for the initial ($n = 0$) and terminal ($n = N+1$) values, respectively.

F Method of simulated moments estimation

This appendix details our simulated method of moments approach to the estimation of $\boldsymbol{\nu} = (\boldsymbol{\nu}_\alpha, \boldsymbol{\nu}_\gamma)$, the parameters of the distributions $F_\alpha()$ and $F_\gamma()$ of the job-offer uncertainty and marginal cost of applications parameters.

Bootstrap estimate of optimal weightmatrix:

Reparameterization: Let $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, \zeta_3)' \in \mathbb{R}^3$ and define

$$s_1 = \text{invlogit}(\zeta_1) \in [0,1]$$

$$s_2 = \text{invlogit}(\zeta_2) \in [0,1]$$

$$\overline{\gamma} = \exp(\zeta_3) > 0,$$

where $\text{invlogit}(x) = \exp(x)/(1+\exp(x))$. Map $(s_1, s_2) \in [0,1]^2$ to an ordered pair $(\underline{\alpha}, \overline{\alpha}) \in [0,1]^2$ by allocating a unit length stick such that

$$\underline{\alpha} = \underbrace{s_1(1-s_2)}_{\text{left piece}}$$

$$\overline{\alpha} = \underbrace{s_1(1-s_2)}_{\underline{\alpha}} + \underbrace{s_2(1-s_1)}_{\text{middle piece}}.$$

This guarantees $0 \leq \underline{\alpha} < \overline{\alpha} \leq 1$ for all $(\zeta_1, zeta_2) \in \mathbb{R}^2$. Intuition: e split the unit interval into three positive pieces, where the first two accumulate to $\overline{\alpha}$. Alternatively, consider the "center + half-width" reparameterization

$$c = \underbrace{\text{invlogit}(\zeta_1)}_{\text{center}} \in [0,1]$$

$$h = \underbrace{\text{invlogit}(\zeta_2)\min\{c, 1-c\}}_{\text{half-width}} \in [0, \min\{c, 1-c\}]$$

$$\overline{\gamma} = \exp(\zeta_3) > 0.$$

Then, the mapping

$$\underline{\alpha} = c - h$$

$$\overline{\alpha} = c + h$$

also enforces $0 \leq \underline{\alpha} < \overline{\alpha} \leq 1$.

Constant random numbers (CRNs): Explain (i) S draws of $U^s = \widehat{\boldsymbol{\delta}} \otimes \boldsymbol{\iota}_I' - \boldsymbol{\varepsilon}_i^{s\prime}$, $\boldsymbol{V}_\alpha^s \overset{iid}{\sim}$ Uniform$(0,1)_{1\times I}$, and $\boldsymbol{V}_\gamma^s \overset{iid}{\sim}$ Uniform$(0,1)_{1\times I}$ are kept fixed across the whole optimization routine; (ii) $\boldsymbol{\alpha}^{s,(k)} = \underline{\alpha}^{(k)} + (\overline{\alpha}^{(k)} - \underline{\alpha}^{(k)})\, \boldsymbol{V}_\alpha^s$ and $\boldsymbol{\gamma}^{s,(k)} = \underline{\gamma}^{(k)} + (\overline{\gamma}^{(k)} - \underline{\gamma}^{(k)})\, \boldsymbol{V}_\gamma^s$ are generated with the same common underlying $1 \times I$ Uniform$(0,1)$ draws in each optimizer iteration $k \in \mathbb{N}$; (iii) a subset $s \in \{1, \ldots, \widetilde{S}\}$ of the CRN draws, with $\widetilde{S} < S$, is used in some stages of the optimization routine to reduce computational costs; and (iv) how this minimizes simulation noise and makes the MSM criterion smoother.

Grid search for initial values: Discuss (i) Initial $\boldsymbol{\nu}$-space (structural parameters) bounding for starting values; (ii) $\boldsymbol{\zeta}$-space (reparameterization for numerical optimization) Latin hypercube sampling, polishing, trimming, and greedy maxmin distance refinement; (iii) tunning parameters (c, h) ensure the mapping $\boldsymbol{\zeta} \mapsto \boldsymbol{\nu}$ satisfies constraint $0 \leq \underline{\alpha} < \overline{\alpha} \leq 1$; (iv) cheap grid search with $\widetilde{S} = 1$ simulation draw over the trimmed and refined Latin hypercube $\rightarrow$ matrix of top K starting values $\boldsymbol{\nu}_k^{(0)}$ for $k \in \{1, \ldots, K\}$.

Initial Nelder-Mead refinement of initial values: Discuss (i) derivative-free optimization due to non-smoothness caused by underlying threshold jumping structure **(make sure this argument is correct!)** (ii) Nelder-Mead algorithm with relatively large/aggresive simplex in $\boldsymbol{\zeta}$-space; (ii) cheap evaluation of the objective function in each iteration for each parameter candidate since we use a subset of $1 \leq \widetilde{S} < S$ of the CRNs (i.e., nested CRNs); (iii) how this minimizes computational costs.

Final Nelder-Mead refinement of the top candidate: Discuss (i) Nelder-Mead algorithm with tighter simplex in $\boldsymbol{\zeta}$-space; (ii) more precise (but costly) evaluation of the objective function with the S CRN draws in each optimizer iteration; and (potentially) any further refinement I may add.

(Potentially) Asymptotic "sandwich" standard errors: Discuss (i) bootstrapping the entire estimation routine, including the MM algorithm for the partially rank-ordered logit, would be ideal but possibly too costly or computationally infeasible for a reasonable number of bootstrap replications; (ii) asymptotic delta-method standard errors can in principle be computed; but (iii) I need to make the numerical derivatives work to estimate the Jacobian matrix of the residual moments function (which I'm having some difficulty with).

G Data description

This appendix. . .

This is old and informal stuff... will need rewriting.

G.1 Job advertisements

Number of vacancies: A small number of job ads —6 out of 1,137,965, mapping to 54 of 39,480,855 applications— report zero vacancies being offered in the raw data. We treat the number of vacancies as missing for these observations.

Hours of work: Combined the two part-time categories into one (the distinction was only nominal).

Ad availability (dates and duration): Some of the reported publication and expiry dates and durations of job ads are nonsensical, leading to many application dates falling out of the reported ad availability period. We redefine job ad availability periods as spells of clustered applications according to the following procedure.

1. Define the maximum length (in days) of an application cluster, $\overline{\tau}$. If two consecutive applications to a given ad are more than $\overline{\tau}$ days apart, they belong to different application clusters. We set $\overline{\tau} = 120$ days.

2. For each job ad j identified by the unique ad ID in the raw data, let $\boldsymbol{d}_j^1 = (d_{j1}^1, \ldots, d_{jT_j}^1)'$ be a column vector containing the T_j numerical dates in which ad j received at least one application in ascending order, where numerical values are assigned to dates following, e.g., Stata's convention.

3. Assign all dates d_{jt}^1 such that $d_{jt}^1 - d_{j1}^1 + 1 \leq \overline{\tau}$ to the first application-cluster spell.

4. If all application dates fall within $\overline{\tau}$ dates of the first application date, stop —and job ad j has only one application-cluster spell. Otherwise, if the first application date more than $\overline{\tau}$ days apart from the first application date is the $\bar{t}_1^{\text{th}}$ one —i.e., $d_{jt}^1 - d_{j_1}^1 + 1 \leq \overline{\tau}\ \forall t \leq \bar{t}_1 - 1$ and $d_{jt}^1 - d_{j_1}^1 + 1 > \overline{\tau}\ \forall t \geq \bar{t}_1$—, repeat the previous steps for Let $\boldsymbol{d}_j^2 = (d_{j1}^2, \ldots, d_{j(T_j - \bar{t}_1 + 1)}^2)' = (d_{j\bar{t}_1}^1, \ldots, d_{jT_j}^1)'$.

The algorithm stops at a n^{th} iteration when $\dim(\boldsymbol{d}_j^n) > 0$ and $\dim(\boldsymbol{d}_j^{n+1}) = 0$, producing n distinct application-cluster spells with durations of at most $\overline{\tau}$ days. We define a new ad ID that maps to unique combinations of the original ad ID (j) and application-cluster spell (s).

Finally, we exploit the information contained in the original publication and expiry dates reported in the raw data by imputing the corresponding ad-availability spell publication and expiry dates as folows. Let $\boldsymbol{d}_{js} = (d_{js1}, \ldots, d_{jsT_{js}})'$ be a column vector containing the applications dates corresponding to application-cluster spell s of ad j in ascending order. We impute the publication date of ad j's s^{th} availability spell, $\underline{t}_{js}$, as ad j's originally reported publication date, $\underline{d}_j$, if the first application occurred at most Δ days after. Otherwise, we use the first application date of the application-cluster spell. That is, we define

$$\underline{t}_{js} = \begin{cases} \underline{d}_j & \text{if} \quad d_{js1} \in \left[\underline{d}_j,\, \underline{d}_j + \Delta\right] \\ \\ d_{js1} & \text{otherwise} \end{cases}.$$

Similarly, if the corresponding application-cluster spell contains more than one application date, we impute the expiry date of availability spell (j, s), $\overline{t}_{js}$, as the originally reported expiry date of ad j in the raw data, $\overline{d}_j$, if the last application occurred at most Δ days before, and the last application date otherwise. In the special case of unit duration —i.e., $T_{js} = 1$—

Appendix references

Chade, H. and L. Smith (2006): "Simultaneous Search," *Econometrica*, 74, 1293–1307.

David, H. A. and H. N. Nagaraja (2003): *Order Statistics*, John Wiley & Sons, third ed.

Roussille, N. and B. Scuderi (2025): "Bidding for Talent: A Test of Conduct in a High-Wage Labor Market," Working Paper 33848, National Bureau of Economic Research.

Wu, T. T. and K. Lange (2010): "The MM Alternative to EM," *Statistical Science*, 25, 492 – 505.