Estimating labor market power from job applications

Pinjas Albagli*

November 3, 2025

[Latest version here]

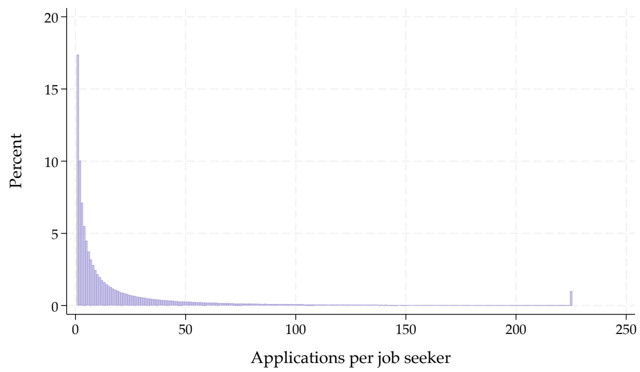
Abstract

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. Standard discrete choice models, however, are built on the assumption that each decision-maker selects only one alternative. When applied to job applications data, this assumption implies that each applicant submits only one application —an unrealistic restriction that is inconsistent with observed behavior and overlooks how wage increases affect the supply of applications not only through substitution across jobs, but also through their impact on the number of applications each worker submits. This paper relaxes the single-application assumption by extending the standard framework to a simultaneous-search setting in which uncertainty about job offers induces multi-application behavior. We derive a decomposition of the wage elasticity of job applications into substitution and portfolio-size components, and develop a multi-stage econometric method that overcomes the computational challenges of estimation with large choice sets. An illustrative application to job-board data demonstrates the method's feasibility and shows that the model is rich enough to successfully replicate observed multi-application patterns and yield meaningful firm-level labor supply elasticity estimates.

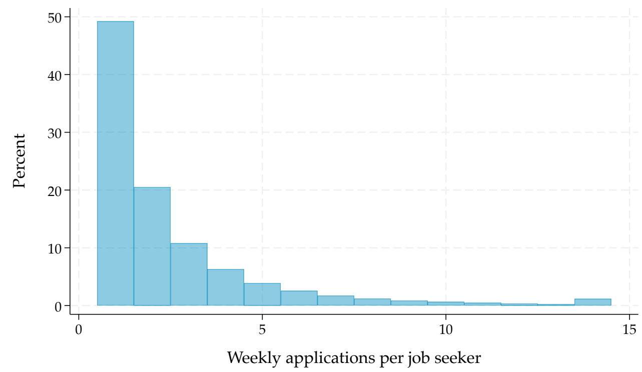
*London School of Economics and Centre for Economic Performance. Email: p.m.albagli@lse.ac.uk.
[Acknowledgements here].

1 Introduction

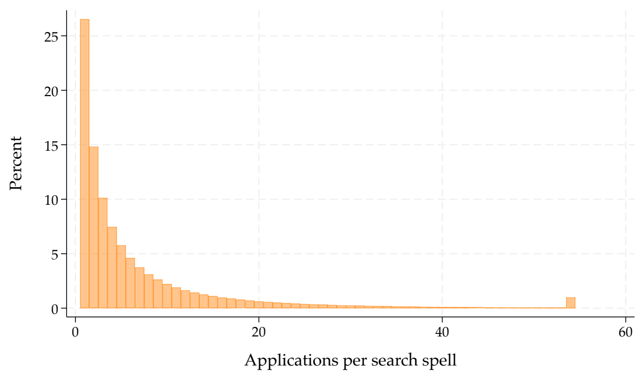
Figure 1. Distribution of the number of applications per job seeker



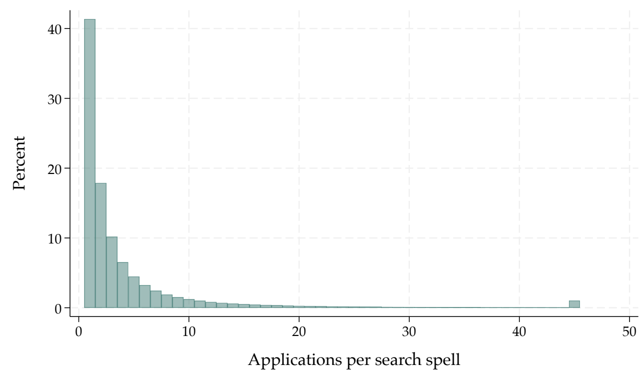
(a) Total



(b) Weekly



(c) Search spell (capped duration)



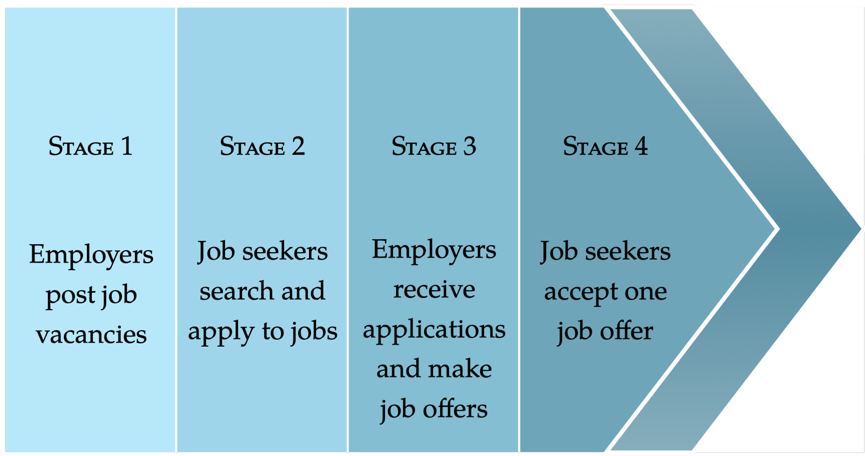
(d) Search spell (capped distance)

Notes: Histogram of the number of applications per job seeker on Chilean job board Trabajando.com. All plots censored at the 99–th percentile. The sample comprises 39,460,403 applications by 1,688,647 job seekers over the period January 1, 2010 to December 31, 2019. The number of applications is left-truncated at zero since the sample only includes applicants. See Section 4.1 for details on the dataset. Panel (a) focuses on the raw, total number of applications per applicant over the sampling period. To address the concern that the observed multi-application behavior at the aggregate level may be an artifact of time aggregation, Panel (b) focuses on weekly applications per job seeker, while Panels (c) and (d) focus on the number of applications per search spell for alternative spell definitions. Search spells are defined as clusters of consecutive application dates by the same individual, where (i) the distance between any application date and the start of the spell does not exceed 90 days in Panel (c), or (ii) the distance between any pair of consecutive application dates does not exceed 15 days in Panel (d). See the Online Appendix for histograms of application counts at different frequencies and under alternative search-spell definitions. The main message of this figure is that multiple applications are pervasive, motivating our portfolio-choice framework and distinguishing our approach from single-application models.

We show that in any model where job seekers apply to multiple vacancies, the wage elasticity of applications decomposes into two structurally distinct components: a substitution margin across jobs holding portfolio size constant, and a portfolio-size margin in the number of applications per worker. The decomposition itself is algebraically simple: it follows directly from the law of total probability. Yet this simple structure highlights a fundamental modeling gap: single-application models collapse two distinct behavioral margins into one. By making this decomposition explicit, we clarify which components of application behavior standard monopsony models can and cannot capture.

One-application models, such as standard discrete-choice, random-utility formulations, mechanically shut down the second channel, rendering their estimated labor-supply elasticities misspecified even under ideal identification and estimation conditions. We develop a structural framework that provides a tractable first step towards modeling both margins jointly, under explicit assumptions that ensure identification and estimation feasibility. This model offers a first parametric implementation of a general property of multi-application labor markets.

Figure 2. 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.

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 3.2 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.

2.1 Discrete choice under uncertainty: The 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.[1] Thus, the outside option is exercised only when either every application in 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_{ik}^A. 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 $u_{ir_i(\mathcal{J},1)} > \cdots > u_{ir_i(\mathcal{J},J)}$.[2] 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

[1] 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 $A_i = \emptyset$.

[2] 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.

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[3]

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

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 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 $\ell \notin A_i \;\not\Rightarrow\; u_{ij} > u_{i\ell}$. 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 job-offer probabilities)**.** The probability that job seeker $i \in \mathcal{I}$ gets an offer from job $j \in \mathcal{J}$ after applying does not depend on the identity of job j:

$$\alpha_{ij} = \alpha_i \in (0, 1), \quad \forall j \in \mathcal{J}.$$

[3] 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 $A \cup \{0\}$.

Assumption 2 (Constant marginal cost of applications)**.** Job seekers $i \in \mathcal{I}$ face a constant marginal cost of application:

$$c_i(|A|) = \gamma_i |A|, \quad \gamma_i > 0, \quad A \subseteq \mathcal{J}.$$

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 (Optimal portfolio)**.** *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

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

(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

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

Proof. See Appendix B. □

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

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

However, the sequential representation in Proposition 1 will prove useful for identification and estimation of the model's parameters after specifying an additive random utility model (see Section 3)

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 "conditional inclusion probabilities" —which are generalizations of the well-known choice probability in the one-application setting— and the probability mass function (pmf) of optimal portfolio sizes, but at the cost of imposing restrictive assumptions on preference heterogeneity and conditional substitution patterns. We further discuss these limitations in Section 2.4.4, after discussing our framework's implications in Sections 2.4.1 to 2.4.3.

To connect the portfolio choice problem to observable data on posted wages and job characteristics, we assume that job seekers' ex-post utilities are additively separable in a systematic component and an idiosyncratic random taste shock. This assumption yields a tractable job differentiation structure while capturing the central economic trade-off between hedging against job-offer uncertainty and costly applications. Workers apply to jobs considering their mean utilities, but randomness in tastes and application outcomes drives variation in optimal portfolios. Formally, each job seeker $i \in \mathcal{I}$ faces the portfolio choice problem defined in 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

$$u_{ij} = \delta_j + \varepsilon_{ij}, \tag{7}$$

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:

$$\delta_j = \beta_w \ln(w_j) + \boldsymbol{x}_j' \boldsymbol{\beta}_{\boldsymbol{x}} + \xi_j. \tag{8}$$

Equations (7)–(8), together with Assumption 3 below, comprise the core of our logit ARUM structure.[4] We complete the model with a parametric distributional assumption on the individual uncertainty and cost parameters α_i and γ_i, treating them as random effects, in Assumption 4.

Assumption 3 (Logit ARUM)**.** The idiosyncratic taste shocks ε_{ij} are independent and identically distributed draws from a standard type-1 extreme value distribution, with cumulative

[4] It is possible, at the risk of reduced tractability, to derive richer, more flexible models by generalizing the additive structure in (7)–(8) and/or substituting Assumption 3 with an alternative, more flexible distributional assumption. Such generalizations are outside 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.

distribution function $F_\varepsilon(x) = \exp(-\exp(-x))$ for $x \in \mathbb{R}$, and independent of the random effects (α_i, γ_i).

Assumption 4 (Uncertainty and cost random effects)**.** The random effects (α_i, γ_i), where α_i is the individual job-offer probability and γ_i is the individual marginal cost of applications, are independent and identically distributed draws from parametric joint cumulative distribution function $F_{\alpha,\gamma}(x, y \mid \boldsymbol{\theta})$ for $(x, y) \in (0,1) \times (0, \infty)$, with parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^L$.

Given this logit structure, we can derive closed-form expressions —up to integrating out the job-seeker heterogeneity embedded in the random effects— for the supply of job applications at the vacancy and firm levels. To simplify notation, let $\mathcal{J} = \{1, \ldots, J\}$ so we can use vector notation for quantities such as $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_J)' \in \mathbb{R}^J$.[5] The expected number of applications to job j in our model is given by

$$q_j(\boldsymbol{\delta}, \boldsymbol{\theta}) = I \sum_{n=1}^{J} \mathcal{s}_{j|n}(\boldsymbol{\delta}) \mathcal{s}_n(\boldsymbol{\delta}, \boldsymbol{\theta}), \tag{9}$$

where

$$\mathcal{s}_{j|n}(\boldsymbol{\delta}) = \sum_{k=J-n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} \frac{\exp(\delta_j)}{\sum\limits_{\ell \in \{j\} \cup A} \exp(\delta_\ell)} + \sum_{s=1}^{|\mathcal{B}_j \setminus A|} (-1)^s \sum_{B \in \mathcal{R}_s(\mathcal{B}_j \setminus A)} \frac{\exp(\delta_j)}{\sum\limits_{\ell \in \{j\} \cup A \cup B} \exp(\delta_\ell)} \tag{10}$$

is the probability that job j belongs to the optimal application portfolio conditional on its size, $\mathbb{P}(j \in A_i \mid n_i = n)$ for $n \in \{1, \ldots, J\}$, $\mathcal{B}_j \equiv \mathcal{J} \setminus \{j\}$ is the set of jobs excluding j, and $\mathcal{R}_k(S) = \{\sigma \subseteq S : |\sigma| = k\}$ is the set of all size–k subsets of set S.[6] 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

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

[5]Alternatively, fix a bijection $g : \mathcal{J} \to \{1, \ldots, J\}$ such that $\boldsymbol{\delta} = (\delta_{g^{-1}(1)}, \ldots, \delta_{g^{-1}(J)})'$ is simply the unique permutation of $\{\delta_j\}_{j \in \mathcal{J}}$ induced by $g(\cdot)$. 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 $\mathcal{J}$.

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

$$\psi_i^n = \frac{\gamma_i}{\alpha_i(1-\alpha_i)^{n-1}}, \quad n \in \{1, \ldots, J\} \tag{12}$$

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

$$\mathscr{s}_{0|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i) = F_\varepsilon\big(\psi_i^1\big)^{\sum_{\ell\in\mathcal{J}} \exp(\delta_\ell)} \tag{13}$$

and

$$\mathscr{s}_{J|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i) = 1 - \sum_{k=1}^{J} \sum_{A\in\mathcal{R}_k(\mathcal{J})} F_\varepsilon\big(\psi_i^J\big)^{\sum_{\ell\in A}\exp(\delta_\ell)} \prod_{m\in\mathcal{J}\setminus A} \Big[1 - F_\varepsilon\big(\psi_i^J\big)^{\exp(\delta_m)}\Big], \tag{14}$$

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

$$\mathscr{s}_n(\boldsymbol{\delta}, \boldsymbol{\theta}) = \int \mathscr{s}_{n|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i)\, \mathrm{d}F_{\alpha,\gamma}(\alpha_i, \gamma_i \mid \boldsymbol{\theta}), \tag{15}$$

where integration is over the support of the joint distribution $F_{\alpha,\gamma}$. See Appendix C for a full derivation.

2.3 The wage elasticity of the job application supply

Having characterized the individual application probabilities and the supply of applications implied by the model, we can now study how they respond to wage changes through their effects on mean utilities. 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

$$\eta_{q_j,w_\ell}(\boldsymbol{\delta}, \boldsymbol{\theta}, \beta_w) = \frac{1}{q_j(\boldsymbol{\delta}, \boldsymbol{\theta})} \left[I \sum_{n=1}^{J} \frac{\partial \mathscr{s}_{j|n}(\boldsymbol{\delta})}{\partial \delta_\ell} \mathscr{s}_n(\boldsymbol{\delta}, \boldsymbol{\theta}) + \mathscr{s}_{j|n}(\boldsymbol{\delta}) \frac{\partial \mathscr{s}_n(\boldsymbol{\delta}, \boldsymbol{\theta})}{\partial \delta_\ell} \right] \beta_w. \tag{16}$$

These vacancy-level elasticities can be decomposed into a substitution effect that operates through changes in the composition of the optimal portfolio for a given portfolio size, and an intensive-margin effect on the distribution of portfolio sizes:

$$\eta_{q_j,w_\ell}(\boldsymbol{\delta}, \boldsymbol{\theta}, \beta_w) = \underbrace{\sum_{n=1}^{J} \omega_{n|j}(\boldsymbol{\delta}, \boldsymbol{\theta})\, \eta_{s_{j|n},w_\ell}(\boldsymbol{\delta}, \beta_w)}_{\text{substitution} \mid n_i} + \underbrace{\sum_{n=1}^{J} \omega_{n|j}(\boldsymbol{\delta}, \boldsymbol{\theta})\, \eta_{s_n,w_\ell}(\boldsymbol{\delta}, \boldsymbol{\theta}, \beta_w)}_{\text{intensive margin: pmf of } n_i}, \tag{17}$$

where $\eta_{s_{j|n},w_\ell}(\boldsymbol{\delta}, \beta_w)$ is the elasticity of the conditional inclusion probability $\mathscr{s}_{j|n}(\boldsymbol{\delta})$ with respect to w_ℓ, $\eta_{s_n,w_\ell}(\boldsymbol{\delta}, \boldsymbol{\theta}, \beta_w)$ is the elasticity of the pmf at $n_i = n$, $\mathscr{s}_n(\boldsymbol{\delta}, \boldsymbol{\theta})$, with respect to w_ℓ, and the weights

$$\omega_{n|j}(\boldsymbol{\delta}, \boldsymbol{\theta}) = \frac{I \mathscr{s}_{j|n}(\boldsymbol{\delta}) \mathscr{s}_n(\boldsymbol{\delta}, \boldsymbol{\theta})}{q_j(\boldsymbol{\delta}, \boldsymbol{\theta})}$$

sum to one over $n \in \{1, \ldots, J\}$ and represent the proportion of applications to job j from size–n portfolios.

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 $\mathcal{J}^f$,

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

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:

$$\eta_{q^f, \boldsymbol{w}^f}(\boldsymbol{\delta}, \boldsymbol{\theta}, \beta_w) = \sum_{j \in \mathcal{J}^f} \omega_{j|f}(\boldsymbol{\delta}, \boldsymbol{\theta}) \sum_{\ell \in \mathcal{J}^f} \eta_{q_j, w_\ell}(\boldsymbol{\delta}, \boldsymbol{\theta}, \beta_w), \tag{19}$$

where the weights $\omega_{j|f}(\boldsymbol{\delta}, \boldsymbol{\theta}) = q_j(\boldsymbol{\delta}, \boldsymbol{\theta}) / q^f(\boldsymbol{\delta}, \boldsymbol{\theta})$ sum to one over $j \in \mathcal{J}^f$ and represent the proportion of applications to firm f that correspond to vacancy j. See Appendix D.1 for a derivation of the vacancy- and firm-level elasticities, and the decomposition into the substitution and portfolio-size effects. See Appendix D.2 for closed-form expressions of the partial derivatives of $\mathscr{s}_{j|n}(\boldsymbol{\delta})$ and $\mathscr{s}_n(\boldsymbol{\delta}, \boldsymbol{\theta})$ with respect to δ_ℓ —up to integration over $F_{\alpha,\gamma}(\cdot, \cdot \mid \boldsymbol{\theta})$.

2.4 Discussion: Implications and limitations

2.4.1 Generality of the elasticity decomposition

While our model imposes a heavy structure on the economic setting, our decomposition of the wage elasticity of the job application supply into substitution and portfolio-size margins is indeed more general than it may first appear. The starting point of our derivation of the job application supply function in Appendix C is a nonparametric decomposition of the expected number of applications to job j, $\mathbb{E}[q_j]$, into a sum of terms involving the pmf of portfolio sizes, $\mathbb{P}(n_i = n)$, and the inclusion probabilities of job j conditional on portfolio size n, $\mathbb{P}(j \in A_i \mid n_i = n)$ (see Equation (C.1)). This decomposition follows directly from the law of total probability under mild regularity assumptions that ensure that the expectation $\mathbb{E}[q_j]$ exists and is finite, and the probabilities $\mathbb{P}(n_i = n)$ and $\mathbb{P}(j \in A_i \mid n_i = n)$ are well defined.

What the restrictive structure imposed by Assumptions 1 to 4 gives us is a set of injective mappings $\{\boldsymbol{\delta} \mapsto \mathbb{P}(j \in A_i \mid n_i = n)\}_{j \in \mathcal{J}, n \in \{1, \ldots, J-1\}}$ and $\{(\boldsymbol{\delta}', \boldsymbol{\theta}')' \mapsto \mathbb{P}(n_i = n)\}_{n=1}^{J}$ from the model's structural parameters to the primitive behavioral probabilities, together with the convenient separability implied by the fact that the $\mathbb{P}(j \in A_i \mid n_i = n)$ do not depend on $\boldsymbol{\theta}$. This structure underlies our approach to estimation of $\boldsymbol{\delta}$ and $\boldsymbol{\theta}$, but it is not necessary to decompose the wage elasticity of the job application supply into its substitution and portfolio-size components. To see this, suppose that some other, more general structure, paired with the portfolio-choice problem from Section 2.1, induces the nonparametric mappings $\mathbb{P}(j \in A_i \mid n_i = n) = \sigma_{j|n}(\boldsymbol{w}, \boldsymbol{X}, \Xi, \boldsymbol{\alpha}, \boldsymbol{\gamma})$ and $\mathbb{P}(n_i = n) = \sigma_n(\boldsymbol{w}, \boldsymbol{X}, \Xi, \boldsymbol{\alpha}, \boldsymbol{\gamma})$ for $j \in \mathcal{J}$ and $n \in \{0, \ldots, J\}$, where the arguments are vectors and matrices suitably stacking $\left\{(w_j, \boldsymbol{x}_j', \xi_j)\right\}_{j\in\mathcal{J}}$, and $\{(\alpha_{ij})_{j\in\mathcal{J}}, \gamma_i\}_{i\in\mathcal{I}}$. If these functions are bounded away from zero for $n > 0$ and continuously differentiable in a neighborhood of the evaluation point $(\boldsymbol{w}^0, \cdot)$, so that the partial derivatives

$$\left.\frac{\partial \sigma_{j|n}(\boldsymbol{w}, \boldsymbol{X}, \Xi, \boldsymbol{\alpha}, \boldsymbol{\gamma})}{\partial \boldsymbol{w}}\right|_{\boldsymbol{w}=\boldsymbol{w}^0}, \qquad \left.\frac{\partial \sigma_{n}(\boldsymbol{w}, \boldsymbol{X}, \Xi, \boldsymbol{\alpha}, \boldsymbol{\gamma})}{\partial \boldsymbol{w}}\right|_{\boldsymbol{w}=\boldsymbol{w}^0}$$

exist and are finite, then the semi-elasticities

$$\upsilon^0_{q_j, w_\ell} = \frac{1}{\mathbb{E}[q_j]}\frac{\partial\, \mathbb{E}[q_j]}{\partial w^0_\ell}, \quad \upsilon^0_{s_{j|n}, w_\ell} = \frac{1}{\sigma_{j|n}(\boldsymbol{w}^0, \cdot)}\frac{\partial \sigma_{j|n}(\boldsymbol{w}^0, \cdot)}{\partial w^0_\ell}, \quad \upsilon^0_{s_n, w_\ell} = \frac{1}{\sigma_{n}(\boldsymbol{w}^0, \cdot)}\frac{\partial \sigma_{n}(\boldsymbol{w}^0, \cdot)}{\partial w^0_\ell},$$

where w_ℓ represents the ℓ–th wage coordinate in $\boldsymbol{w}$ and superscript 0 denotes evaluation at $\boldsymbol{w}^0$, are well defined for each $(j, \ell) \in \mathcal{J}^2$ and $n \in \{0, \ldots, J\}$ —for the latter two. The decomposition

$$\upsilon^0_{q_j, w_\ell} = \sum_{n=1}^{J} \varpi^0_{n|j}\left(\upsilon^0_{s_{j|n}, w_\ell} + \upsilon^0_{s_n, w_\ell}\right)$$

follows through by differentiating the identity

$$\mathbb{E}[q_j] = I \sum_{n=1}^{J} \sigma_{j|n}(\boldsymbol{w}, \cdot)\, \sigma_n(\boldsymbol{w}, \cdot)$$

at $\boldsymbol{w} = \boldsymbol{w}^0$, where the weights

$$\varpi^0_{n|j} = \frac{I\, \sigma_{j|n}(\boldsymbol{w}^0, \cdot)\, \sigma_n(\boldsymbol{w}^0, \cdot)}{\mathbb{E}[q_j]}$$

measure the share of expected applications to job j that come from size-n portfolios and sum to one.

This means that the misspecification in single-application models goes deeper than revealed preference and the identification of ex-post mean utilities: it is not an econometric issue, it is structural.

2.4.2 Contrast 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.

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 arising as job-offer uncertainty vanishes, and (ii) a restricted version of our model in which job seekers are constrained to submit at most one application. Comparing these models highlights how portfolio choice affects both the expected number of applications per vacancy.

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 $i \in \mathcal{I}$. This baseline, homogeneous setup will also serve as a basis for our investigation of comparative statics in the $\boldsymbol{\theta}$ parameters in Section 2.4.3. Under these degenerate distributions for the uncertainty and cost parameters, the unconditional pmf of the number of applications per job seeker in Equation (15) coincides with the conditional pmf in Equations (11), (13) and (14) with $\boldsymbol{\theta} = (\alpha, \gamma)$. In this case, the thresholds in Equation (12) simplify to

$$\psi^n = \frac{\gamma}{\alpha(1-\alpha)^{n-1}}. \tag{20}$$

The expected number of applications received by each job vacancy and the conditional application shares are then given by Equations (9) and (10), respectively. The extensive margin, capturing the decision to apply ($n_i > 0$) versus not applying ($n_i = 0$) is characterised by the outside-option share (over job seekers)

$$\mathscr{s}_0(\boldsymbol{\delta}, \alpha, \gamma) = F_\varepsilon\left(\frac{\gamma}{\alpha}\right)^{\sum_{\ell \in \mathcal{J}} \exp(\delta_\ell)}. \tag{21}$$

We use this baseline model as the point of comparison for the no-uncertainty constrained and portfolio-size single-application benchmark models introduced below.

No-uncertainty 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. For notational consistency, note that this restricted model can be written as a special case of our baseline model with $\boldsymbol{\theta} = (1, \gamma)$, violating the support restriction on α_i in Assumption 1. 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.[7]

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 (7) 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:

$$n_i \equiv |A_i| \in \{0, 1\}. \tag{22}$$

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

$$\widetilde{s}_0(\boldsymbol{\delta}, \gamma) = F_\varepsilon(\gamma)^{\sum_{\ell \in \mathcal{J}} \exp(\delta_\ell)} = s_0(\boldsymbol{\delta}, 1, \gamma). \tag{23}$$

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

$$\widetilde{s}_j(\boldsymbol{\delta}) = \frac{\exp(\delta_j)}{\sum_{\ell \in \mathcal{J}} \exp(\delta_\ell)} = s_{j|1}(\boldsymbol{\delta}). \tag{24}$$

(iv) The outside option is less attractive than in the baseline model:

$$\widetilde{s}_0(\boldsymbol{\delta}, \gamma) < s_0(\boldsymbol{\delta}, \alpha, \gamma), \quad \forall \boldsymbol{\delta} \in \mathbb{R}^J, \quad \forall \alpha \in (0, 1), \quad \forall \gamma > 0. \tag{25}$$

(v) The extensive non-participation margin is less sensitive to mean utilities than in the baseline model:

$$\frac{\partial \widetilde{s}_0(\boldsymbol{\delta}, \gamma)}{\partial \delta_j} < \frac{\partial s_0(\boldsymbol{\delta}, \alpha, \gamma)}{\partial \delta_j}, \quad \forall j \in \mathcal{J}, \quad \forall \boldsymbol{\delta} \in \mathbb{R}^J, \quad \forall \alpha \in (0, 1), \quad \forall \gamma > 0. \tag{26}$$

Proof. See Appendix D.3. ☐

[7] A model resembling the textbook MNL with random outside utility can be obtained by setting $\alpha_i = 1$ for all $i \in \mathcal{I}$, taking $F_\gamma(\cdot) = F_\varepsilon(\cdot)$ —where we assume mutually independent random effects: $F_{\alpha,\gamma}(x, y) = F_\alpha(x) F_\gamma(y)$— and letting $\varepsilon_{i0} \equiv \gamma_i \overset{\text{iid}}{\sim} \text{EV}_1$. 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 $\gamma_i \leq 0$.

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

$$\widetilde{q}_j(\boldsymbol{\delta},\gamma) = I\,\widetilde{\mathscr{s}}_j(\boldsymbol{\delta})\left[1-\widetilde{\mathscr{s}}_0(\boldsymbol{\delta},\gamma)\right]. \tag{27}$$

How it compares to the expected application supply to job j in the baseline model, $q_j(\boldsymbol{\delta},\alpha,\gamma)$, is not immediately obvious because of the result in part (iv) of Lemma 1. While job j only receives applications from job seekers submitting exactly one application, causing a fall in its expected number of applications since there are no incoming expected applications from larger portfolio sizes, the outside option is also less attractive without uncertainty, pushing more job seekers over the extensive margin from 0-size portfolios to singleton portfolios and thus offsetting the former effect. Which effect dominates cannot be determined a priori, so it is ultimately an empirical question.

Constrained portfolio size benchmark. We next compare the baseline model to a constrained version where job seekers are not allowed to submit more than one application, so the expected-utility maximization problem becomes

$$\max_{A\in\{\sigma\subseteq\mathcal{J}\,:\,|\sigma|\leq 1\}} U_i(A),$$

where the expected utility of the lottery induced by portfolio A remains defined by Equation (3) and the only difference is that the job seeker only compares it over singleton application portfolios and the outside option (the empty set). The solution to this problem is trivial: apply to the best ex-post alternative if the payoff exceeds the application threshold γ/α, otherwise do not apply. One way to see this formally is restricting the support to $n\in\{1\}$ in part (i) of Proposition 1, yielding $n_i = \mathbb{1}\left\{u_{ir_i(\mathcal{J},1)} \geq \gamma/\alpha\right\}$. Alternatively, follow through the steps of the proof of Proposition 1 in Appendix B, noting that the constraint implies the MIA must stop after the first iteration that compares the best singleton to the outside option. In contrast to the first, no-uncertainty benchmark model, here job seekers face job-offer uncertainty that incentivizes multiple applications, but they are constrained and can submit at most one. Conditional on application, we have standard MNL choice among $j\in\mathcal{J}$. What the singleton portfolio-size constraint does to the distribution of optimal n_i is redistributing all the probability mass from $n\in\{2,\dots,J\}$ to $n=1$, leaving the mass at $n=0$ unchanged. Job j receives fewer applications in this constrained benchmark:

$$\begin{aligned}
\overline{q}_j(\boldsymbol{\delta},\alpha,\gamma) &= I\overline{\mathscr{s}}_j(\boldsymbol{\delta})\left[1-\overline{\mathscr{s}}_0(\boldsymbol{\delta},\alpha,\gamma)\right] \\
&= I\sum_{n=1}^{J} \mathscr{s}_{j|1}(\boldsymbol{\delta})\mathscr{s}_n(\boldsymbol{\delta},\alpha,\gamma) \\
&< I\sum_{n=1}^{J} \mathscr{s}_{j|n}(\boldsymbol{\delta})\mathscr{s}_n(\boldsymbol{\delta},\alpha,\gamma)
\end{aligned}$$

$$= q_j(\boldsymbol{\delta}, \alpha, \gamma),$$

where the second equality follows from the MNL structure conditional on application in the benchmark model, the common extensive margin across models, and the fact that the pmf of n_i integrates (sums) to one. The inequality follows from the monotonicity of $\left\{s_{j|n}(\boldsymbol{\delta})\right\}_{n=1}^{J}$ in n.

No-uncertainty versus constrained portfolio size. Finally, by comparing the two benchmark models, we gain further insight into how job-offer uncertainty drives application behavior. In both models, job seekers either submit one application or do not apply at all, with standard MNL choice conditional on application. The only difference is that uncertainty makes singleton-application less attractive, increasing the proportion of job seekers who choose not to apply. Thus, uncertainty has a dual role by incentivizing multiple applications on the one hand, but making the outside option more attractive on the other.

2.4.3 Comparative statics: The roles of uncertainty and application costs

The two-stage representation of the optimal decision rule in Proposition 1 cleanly isolates the roles of job-offer uncertainty, α_i, and application costs, γ_i: they only enter the optimal stopping-rule that determines the optimal portfolio size, n_i. That is, together with the idiosyncratic taste shocks ε_{ij}, they shape the portfolio-size mechanism in our decomposition of the wage elasticity of job applications in Equation (17). For any given value of mean utilites $\boldsymbol{\delta} \in \mathbb{R}^J$, different random-effects distributions $F_{\alpha,\gamma}(\alpha_i, \gamma_i \mid \boldsymbol{\theta})$ induce different portfolio-size distributions $s_n(\boldsymbol{\delta}, \boldsymbol{\theta})$. To gain insight into this relationship, we first consider the homogeneous baseline model from Section 2.4.2, where $\boldsymbol{\theta} = (\alpha, \gamma) \in (0,1) \times \mathbb{R}_{++}$ and $F_{\alpha,\gamma}$ is degenerate at $\boldsymbol{\theta}$.

The homogeneous marginal cost of applications γ enters the stopping-rule thresholds (20) multiplicatively with a positive derivative that increases monotonically with $n \in \{1, \ldots, J\}$,

$$\frac{\partial \psi^n}{\partial \gamma} = \frac{1}{\alpha(1-\alpha)^{n-1}}. \tag{28}$$

Holding everything else constant, an increase in the marginal cost of applications generates proportional increases in all thresholds, with larger impact on those corresponding to higher n and thus shifting the probability mass of the portfolio-size distribution to the left. We can directly see from Equation (21) that the proportion of job seekers who choose the outside option increases with γ. Therefore, since probabilities sum to one over the support, $\mathbb{P}(n_i > 0)$ must fall with an increase in γ, and any increase in $\mathbb{P}(n_i = n)$ for some relatively small but positive n must be of a lower magnitude than the overall mass decrease at higher n.

In contrast, the homogeneous job-offer probability α does not shift all thresholds in the same direction since the derivative

$$\frac{\partial \psi^n}{\partial \alpha} = \gamma \, \frac{\alpha n - 1}{\alpha^2 (1-\alpha)^n} \tag{29}$$

is negative for small n and positive for large n. In particular, there is a kink at $n^* = 1/\alpha$, so $\mathbb{P}\left(n_i = n\right)$ increases with α for $n < n^*$ and falls for $n > n^*$. The job-offer probability parameter governs how the stopping-rule thresholds grow with n, since the growth rate

$$\frac{\psi^n - \psi^{n-1}}{\psi^{n-1}} = \frac{\alpha}{1-\alpha}, \qquad n \in \{2, \ldots, J\} \tag{30}$$

is constant and depends only on α. Therefore, α determines the rate of decay of the right tail of the portfolio-size distribution. Higher α induces faster growing thresholds and thus a shorter right tail, making extreme portfolio sizes less likely.

These results are fairly intuitive: higher application (marginal) costs discourage the overall application intensity, while a higher likelihood of getting an offer and landing a job disincentivizes diversification through very large application portfolios since applications are less risky.[8] Both parameters reduce the frequency of large application portfolios, but operate through different mechanisms. Higher γ raises all thresholds proportionally, uniformly discouraging applications and compressing the distribution around small n and non-participation. In contrast, higher α alters the risk–return tradeoff: greater job-offer chances reduce the need to apply broadly, thinning the right tail but redistributing probability mass toward intermediate portfolio sizes rather than zero. In other words, γ shifts the level of application intensity, while α determines curvature.

Panel (a) of Figure 3 illustrates these comparative statics for α on the right-hand plot and for γ on the left-hand plot. We simulate $I = 100,000$ job seekers choosing among $J = 1,000$ jobs for different values of $\boldsymbol{\theta} = (\alpha, \gamma)$ holding constant the mean utilities $\boldsymbol{\delta}$ and the EV_1 taste shock draws across simulations. Note that as we vary α over $\{0.04, 0.05, 0.07, 0.09\}$ for fixed $\gamma = 0.5$ on the right-hand panel, the supports of the simulated distributions are bounded away from the corresponding kink points $n^* = 1/\alpha \in \{25, 20, 14.29, 11.11\}$. Therefore, only the mass-increase effect for relatively-low n is observed. The homogeneous baseline model is incapable of generating a portfolio size distribution with significant probability mass at low and high n values at the same time. Figure H.1 in Appendix H illustrates cases with higher α —and thus lower kinks— where the compression of the right tail is observed, but with zero mass at lower

[8] By diversification, here we mean increasing the likelihood of landing a job by investing in a larger number of risky options with uncorrelated chances of success. It must not be conflated with hedging in the sense of applying to a job for its high offer probability —which in our model is constant across jobs— when not justified by its expected payoff. Regarding the latter, Chade and Smith (2006) show that a "'no safety school" property holds in the simultaneous search framework with independent admission events. See Ali and Shorrer (2025) for a more general simultaneous search setting where hedging emerges optimally.

n. As we increase γ through $\{0.5, 0.55, 0.6, 0.65\}$ for fixed $\alpha = 0.09$ on the right-hand panel, probability mass monotonically shifts to the left, towards zero/very-low n.

In the mutually-independent heterogeneous case, when the joint distribution of (α_i, γ_i) is nondegenerate and factors into the product of the marginals, analogous arguments hold conditional on i since the functional form of the individual thresholds ψ_i^n in (20) remains the same. The observed pmf of n_i is a mixture of these baseline shapes where γ_i scales the level of all individual thresholds and α_i controls their growth rate. Higher dispersion in γ_i shifts the overall level of application intensity, stretching the distribution of portfolio sizes as the proportion of job seekers facing very low application costs grows. Of course, the magnitude of this effect also depends on the mean of γ_i. Dispersion in α_i introduces curvature effects: low–α_i individuals flatten the stopping sequence (slower threshold growth rate) and apply to more jobs, while higher–α_i individuals truncate their portfolios earlier (faster growing thresholds). As a result, variation in α_i thickens the low-application region and the right tail of the distribution without necessarily extending the support (to the right). To push far into high n, persistently low α_i is necessary, which will depend both on the mean and dispersion of α_i. Hence, unlike γ_i–heterogeneity, which rescales thresholds uniformly, α_i–heterogeneity changes their relative spacing, thus reshaping the distribution rather than only stretching it.

These independent-heterogeneous comparative statics can be seen in Panel (b) of Figure 3, where we vary the dispersion of α_i keeping its mean and the marginal distribution of γ_i constant on the left-hand panel, and vice versa on the right-hand panel. For this calibration, we assume the random effects are mutually independent with marginal distributions

$$\alpha_i \overset{\text{iid}}{\sim} \mathrm{Beta}\big(\theta_1\,\theta_2,\ (1-\theta_1)\,\theta_2\big), \qquad \gamma_i \overset{\text{iid}}{\sim} \mathrm{Gamma}\big(\theta_4,\ \theta_3\,/\,\theta_4\big),$$

setting the means at $\mathbb{E}\left[\alpha_i\right] = \theta_1 = 0.09$ and $\mathbb{E}\left[\gamma_i\right] = \theta_3 = 0.5$ for comparability with the homogeneous simulations in Panel (a). Here, we censor the simulated portfolio-size distributions at their corresponding 99–th percentiles to improve readability. We add to the fixed simulation setting from Panel (a) a fixed draw of two i.i.d. standard uniform random variates for each job seeker, which we use to construct the (α_i, γ_i) draws for the different parameter values through the corresponding marginal quantile functions, while keeping the underlying pseudo-randomness of simulation fixed. On the right-hand panel we see that increasing $\mathrm{Var}\left(\gamma_i\right) = (\theta_3)^2/\theta_4$ over $\{0.00125, 0.0025, 0.005, 0.01\}$ by respectively varying θ_4 over $\{200, 100, 50, 25\}$ for fixed $\theta_3 = 0.5$, shifts probability mass out of the modal region towards the extremes and, in our calibration, extends the support of the portfolio-size distribution rightward. In contrast, increasing $\mathrm{Var}\left(\alpha_i\right) = \theta_1(1-\theta_1)/(1+\theta_2)$ over (approximately) $\{0.0004, 0.0008, 0.0016, 0.0032\}$ primarily thickens the low-n mass and the right tail without extending the support rightward for this set of calibrated parameters.

Our framework allows arbitrary dependence between job-offer uncertainty and application costs. A convenient way to model this dependence is, for example, through a Gaussian copula, which requires introducing only one additional correlation parameter for any specified parametric marginal distributions of α_i and γ_i.[9] This correlation parameter captures the dependence between α_i and γ_i and determines how much probability mass lies near the diagonal of the (α, γ)–type space, where low α_i is paired with low γ_i and high α_i with high γ_i.[10] Under positive correlation, the joint distribution concentrates near the diagonal. Higher positive dependence moves mass towards the diagonal (more low–low and high–high types) and away from the off-diagonal corners (fewer low–high and high–low types). A negative correlation parameter and increasing (in magnitude) negative dependence have the opposite effect.

To fix ideas, let $\boldsymbol{\theta}_\alpha$ and $\boldsymbol{\theta}_\gamma$ be the subvectors of $\boldsymbol{\theta}$ determining the marginal distributions $F_\alpha(\alpha_i \mid \boldsymbol{\theta}_\alpha)$ and $F_\gamma(\gamma_i \mid \boldsymbol{\theta}_\gamma)$, respectively, and $\theta_C \in (-1, 1)$ represent the Gaussian copula correlation parameter. Then, we can rewrite Equation (15) for the unconditional pmf of portfolio size as

$$\mathscr{s}_n(\boldsymbol{\delta}, \boldsymbol{\theta}) = \int_{[0,1]^2} \mathscr{s}_{n|\alpha,\gamma}\big(\boldsymbol{\delta}, F_\alpha^{-1}(u \mid \boldsymbol{\theta}_\alpha), F_\gamma^{-1}(v \mid \boldsymbol{\theta}_\gamma)\big)\, \mathrm{d}C\big(u, v \mid \theta_C\big), \tag{31}$$

where $C\big(u, v \mid \theta_C\big) = \Phi_2\big(\Phi^{-1}(u), \Phi^{-1}(v) \mid \theta_C\big)$, $\Phi(\cdot)$ is the cdf of the standard normal distribution, and $\Phi_2\big(\cdot, \cdot \mid \theta_C\big)$ is the cdf of the standard bivariate normal distribution with correlation parameter θ_C. To understand how the shape of the (unconditional) portfolio-size pmf is affected by changes in the Gaussian copula correlation parameter θ_C, it is useful to consider the k–level sets of the conditional pmf $\big\{\mathscr{s}_{n|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i)\big\}_{n=0}^J$ in (α, γ)–space for $k \in (0, 1)$ and fixed $\boldsymbol{\delta} \in \mathbb{R}^J$,

$$\mathcal{S}_n^k(\boldsymbol{\delta}) \equiv \Big\{(\alpha_i, \gamma_i) \in (0, 1) \times \mathbb{R}_{++} : \mathscr{s}_{n|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i) = k\Big\},$$

and how they intersect the the comonotone-rank path

$$\mathcal{D}(\boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_\gamma) \equiv \Big\{\big(F_\alpha^{-1}(u \mid \boldsymbol{\theta}_\alpha), F_\gamma^{-1}(u \mid \boldsymbol{\theta}_\gamma)\big) \in (0, 1) \times \mathbb{R}_{++} : u \in (0, 1)\Big\}$$

If $\boldsymbol{\theta}_\alpha$ and $\boldsymbol{\theta}_\gamma$ are such that along the diagonal $\mathcal{D}(\boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_\gamma)$ the conditional probability mass of n_i concentrates on moderate n, then the unconditional pmf $\{\mathscr{s}_n(\boldsymbol{\delta}, \boldsymbol{\theta})\}_{n=0}^J$ shifts mass from the extremes towards the middle of the support as θ_C increases. If instead the intersection of $\big(\mathcal{S}_n^k(\boldsymbol{\delta})\big)_{k\in(0,1)}$ and $\mathcal{D}(\boldsymbol{\theta}_\alpha, \boldsymbol{\theta}_\gamma)$ is such that low–low types concentrate probability around large n while high–high types concentrate around very low n (especially $n = 0$), then higher θ_C pushes mass from the middle towards the very-low n region and the right tail.

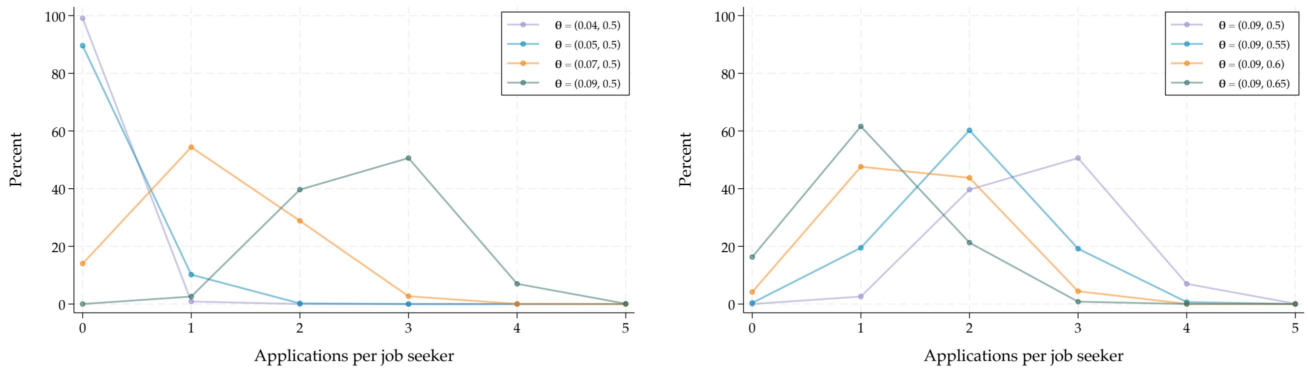
We conclude our comparative statics exercise in Panel (c) of Figure 3 by studying how varying θ_C changes the simulated portfolio-size distribution in our calibrated setup from Panels (a) and (b). In particular, we hold fixed a pair of Beta–Gamma marginal distributions from Panel (b)

[9]See, e.g., Trivedi and Zimmer (2007) for details on copulas and, in particular, the Gaussian (or normal) copula.

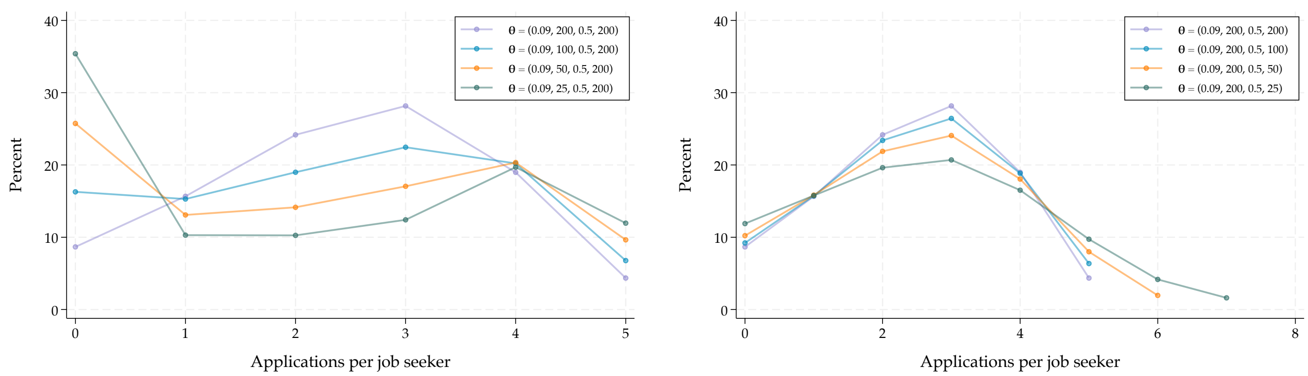
[10]Here, formally, by diagonal we mean the comonotone diagonal in rank space: $\big\{(U, V) \in (0, 1)^2 : U = V\big\}$, where $U = F_\alpha(\alpha_i)$ and $V = F_\gamma(\gamma_i)$ are the marginal ranks.

such that dispersion in α_i is relatively high while γ_i–dispersion relatively low on the left-hand panel, and vice versa on the right-hand panel. We vary $\theta_5 = \theta_C$ over $\{-0.8, -0.2, 0.2, 0.8\}$ for the two pairs of fixed marginal distributions —i.e., holding $(\theta_1, \theta_2, \theta_3, \theta_4)$ constant. We see that as positive dependence increases, the extremes of the distribution shrink while the middle expands, consistent with the first case described in the previous paragraph. See Panel (c) of Figure H.1 for an alternative calibration where the second case is observed.

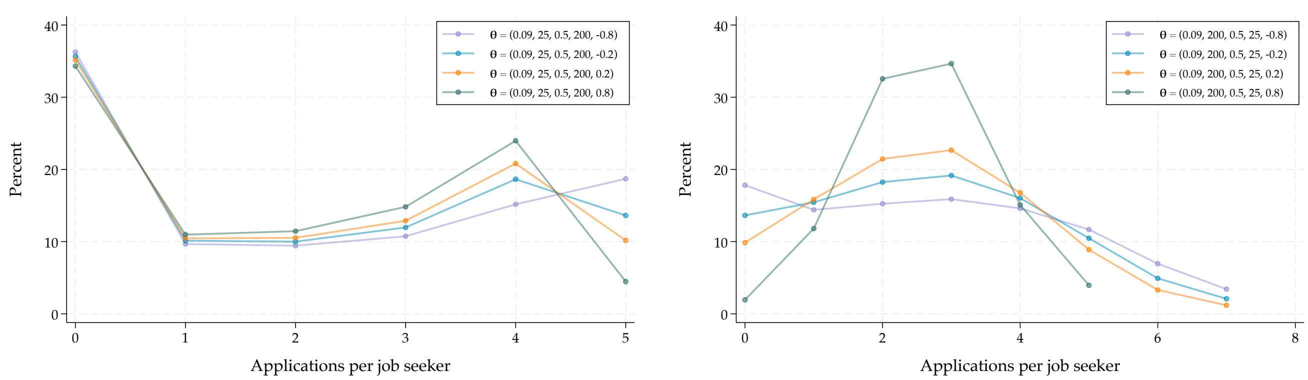
Figure 3. Comparative statics in $\boldsymbol{\theta}$



(a) Homogeneous



(b) Heterogeneous independent



(c) Heterogeneous copula

Notes: Histograms of portfolio size for different simulated samples comprising $I = 100,000$ job seekers choosing among $J = 1,000$ jobs. The individual job-offer probabilities and marginal costs of applications (α_i, γ_i) are assumed homogeneous in Panel (a), heterogeneous, mutually independent with marginal distributions $\alpha_i \overset{\text{iid}}{\sim} \text{Beta}(\theta_1\theta_2, (1-\theta_1)\theta_2)$ and $\gamma_i \overset{\text{iid}}{\sim} \text{Gamma}(\theta_4, \theta_3/\theta_4)$ in Panel (b), and heterogeneous and mutually dependent through a Gaussian copula structure in Panel (c) with the same marginal distributions from Panel (b) and correlation parameter θ_5. We keep $\boldsymbol{\delta}$ and the $\varepsilon_{i,j}$ EV_1 draws constant across simulations, and vary (i) the level in Panel (a) or (ii) the dispersion in Panel (b) of α_i keeping that of γ_i constant, and vice versa. In Panel (c) we vary the Gaussian copula correlation parameter keeping the marginal distributions of α_i and γ_i constant. We also draw $I \times 1$ vectors of i.i.d. standard uniform random variates that we keep fixed across simulation and use to sample (α_i, γ_i) for different values of $\boldsymbol{\theta}$ in Panels (b) and (c). Since $\mathbb{E}[\alpha_i] = \theta_1$, $\text{Var}(\alpha_i) = \theta_1(1-\theta_1)/(1+\theta_2)$, $\mathbb{E}[\gamma_i] = \theta_3$, and $\text{Var}(\gamma_i) = (\theta_3)^2/\theta_4$ in the heterogeneous specifications, we increase dispersion keeping expectations fixed by lowering θ_2 for α_i and θ_4 for γ_i. Portfolio size is discrete; probability mass points are connected by colored lines purely for visual continuity.

2.4.4 Limitations

Our framework necessarily abstracts from several important aspects of job search and matching. Here we briefly discuss some of these limitations and how they may be addressed in future work.

IIA and systematic preference heterogeneity. While the conditional logit structure within each portfolio-size level imposes the IIA property —i.e., conditional on a given number of applications, the relative inclusion probabilities between any two vacancies depend only on their ex-post utility differences (McFadden, 1974; Train, 2009)—, our model does not satisfy IIA unconditionally. The relative unconditional inclusion probabilities $\mathbb{P}(j \in A_i) / \mathbb{P}(\ell \in A_i)$ do depend on δ_k for $k \notin \{j, \ell\}$. The composition of conditional inclusion probabilities with the endogenous distribution of portfolio sizes, which itself depends on all mean utilities, induces richer cross-vacancy substitution patterns. In particular, the portfolio-size channel and heterogeneity in job-offer probabilities and application costs generate asymmetric cross-effects in the Jacobian of the application supply function, breaking the symmetry and column-constant cross-elasticities characteristic of the standard multinomial logit.

On the other hand, our ARUM specification in (7)–(8) implies all horizontal differentiation arises from the idiosyncratic extreme-value taste shocks, while systematic preferences are vertically differentiated. This leaves out salient (i, j)-specific characteristics, such as commuting distance or occupation-specific match quality, that are likely to influence application behavior. Indeed, empirical evidence suggests distance to the place of work is an important determinant of systematic preferences in the contexts of job applications and final job-offer acceptance (Marinescu and Rathelot, 2018; Le Barbanchon et al., 2020; Azar et al., 2022; Hirsch et al., 2022) —i.e., stages 3 and 4 in Figure 2. Relaxing conditional IIA through nested logit, mixed logit, or other generalized discrete-choice structures would allow richer substitution patterns that can accommodate these dimensions at the cost of increased computational costs in a setting where the simplest logit structure already generates choice probabilities of a complex combinatorial nature.

Labor supply versus applications supply. Our framework models the supply of applications rather than the supply of accepted employment relationships. A full model of monopsony power needs to track applications through subsequent stages: firms select among applicants, offers are made, and workers accept or reject them. Given our assumptions, acceptance is automatic once an offer is received, but the process of applicant selection is left unmodeled. As a result, our theoretical elasticities capture the responsiveness of application flows to wages, not necessarily the responsiveness of employment. This abstraction is deliberate, but it limits the model's ability to provide a complete mapping from firm wages to realized labor supply. This issue has already been recognized and to an extent dealt with in the literature (Azar and Marinescu, 2024). In the case of job-to-job moves, Manning (2011) shows that since the recruit of one firm is a quit from another, the labor supply elasticity is roughly twice the recruitment elasticity. Based on the available empirical evidence (Manning, 2011; Dal Bó et al., 2013; Marinescu and Wolthoff, 2020; Dube et al., 2020), Azar et al. (2022) conclude that it is reasonable to assume that the application elasticity is roughly equal to the hiring elasticity, and thus, for illustrative purposes, estimate the labor supply elasticity as twice the application elasticity.

Strong assumptions on job-offer probabilities. We assume that job-offer probabilities are exogenous and homogeneous across jobs, depending only on the worker ($\alpha_{ij} = \alpha_i$). This is a useful simplification, but it may be difficult to reconcile with a setting in which firms optimally select from their applicant pools. One way to rationalize this assumption is to imagine that firms screen on a one-dimensional worker "quality" index, with thresholds that vary across firms but do not interact with job characteristics relevant to workers' utility. A full model of firm screening, or relaxing the homogeneity restriction, would help connect our framework more directly to equilibrium hiring behavior, but we would lose the identifying power of revealed preference.

Equilibrium considerations. Our framework focuses squarely on the supply of applications and abstracts from the demand side of the market and equilbrium interactions. Job characteristics are treated as exogenous, and we do not model how firms would adjust them (or wages) in response to application behavior in market equilibrium. This abstraction is consistent with our primary object of interest—the wage elasticity of the supply of applications to the firm—but it also limits our ability to capture feedback between supply and demand. Embedding the framework into a richer environment with firm selectivity responses remains an important direction for future research.

Directed and competitive search. We abstract away from directed (competitive) search considerations that determine the shape of the equilibrium wage distribution, and where modeling multiple-application behavior has been found to increase the explanatory power of the theory. In particular, allowing multiple applications in directed search generates wage dispersion with a decreasing density of posted wages, even with homogeneous agents (Galenianos and Kircher, 2009). This contrasts with one-application directed search, which produces a degenerate posted-wage distribution (Burdett et al., 2001), and with random-search models, which produce non-degenerate wage distributions with increasing density (Burdett and Mortensen, 1998). Empirically, wage densities are typically unimodal, rising at the bottom and declining in the upper tail (Mortensen, 2003). The decreasing density in Galenianos and Kircher (2009) aligns better with the observed decline at higher wages than the monotonically increasing density implied by Burdett and Mortensen (1998).[11] Our assumption that $\alpha_{ij} = \alpha_i$ rules out any worker–firm specific heterogeneity in offer probabilities that could induce sorting and directed application behavior.

Information frictions and rational inattention. Our framework also abstracts away from costly information acquisition and the formation of consideration sets. There is a direct connection between rational inattention (RI) and (generalized) multinomial logit choice probabilities (Matějka and McKay, 2015). Recent empirical work applies endogenous information acquisition to health insurance plan choice (Brown and Jeon, 2024). Extending our model to allow job seekers to optimally allocate attention across vacancies—thereby endogenizing consideration sets and the intensity of directed search—would be a natural next step. Evidence from Banfi and Villena-Roldán (2019) suggests that even when wages are hidden, applicants respond to implicit signals of pay, consistent with costly information acquisition shaping consideration sets. The data in our empirical application in Section 4 are well-suited for exploring this extension: wage posting is voluntary, but job seekers can infer information from wage brackets and search categories, and from other non-wage job attributes. More speculatively, rational inattention could also provide microfoundations for directed search behavior: if workers optimally allocate scarce attention across vacancies, the resulting attention targeting may resemble the directed application choices emphasized in the competitive search literature. Exploring this connection remains an open avenue for future research.

Static model. Our simultaneous search framework is static. It does not capture important dynamic aspects of labor markets such as the entry and exit of job vacancies, the initiation and duration of search spells, or sequential updating by workers. These features are absent in many portfolio-choice contexts such as school choice with synchronized admission cycles, but are central to labor markets. Incorporating dynamics would allow richer analysis of learning, timing, and adjustment, albeit with substantial computational costs and data requirements. A sequential search framework may be more appropriate to capture these dynamic aspects of job search and employment determination.

[11] In Burdett et al. (2001), the symmetric equilibrium with one application per worker features a single posted wage, so the distribution degenerates at a point. In contrast, in the classical Burdett and Mortensen (1998) model of random search with on-the-job search, the equilibrium wage distribution is non-degenerate and has an increasing density under standard assumptions. The multiple-application directed search environment in Galenianos and Kircher (2009), based on the simultaneous search setting studied by Chade and Smith (2006), has the explanatory power to reconcile theory with the empirically observed decline in the upper tail of the wage distribution.

3 Econometrics: Taking the model to data

In this section, we consider estimation of our model with observational data on job applications consisting of (i) a set of job seekers, (ii) a set of job vacancies, (iii) posted wages and job characteristics, (iv) a set of employers, (v) a mapping from job seekers to the jobs they applied to, and (vi) a mapping from job vacancies to the firms that posted them. Here we follow the standard convention of denoting our estimands, the unknown population parameters, with subscript 0. We establish identification of the structural parameters $\boldsymbol{\delta}_0$, $\boldsymbol{\theta}_0$, and β_{w0} in Section 3.1, and develop a multi-stage estimation strategy grounded in the empirical IO tradition in Section 3.2. Finally, we discuss a plug-in, simulation-based estimator for the firm-level wage elasticities $\eta_{q^f,\boldsymbol{w}^f}(\boldsymbol{\delta}_0, \boldsymbol{\theta}_0, \beta_{w0})$ in Section 3.3.

3.1 Identification

The conditional application share functions $\mathscr{s}_{j|n}(\boldsymbol{\delta})$ map $\boldsymbol{\delta}$ to the conditional inclusion probabilities $\mathbb{P}\left(j \in A_i \mid n_i = n\right)$ and are directly derived from the EV_1 assumption on the idiosyncratic taste shocks $\{\varepsilon_{ij}\}_{i\in\mathcal{I},j\in\mathcal{J}}$. They do not depend on the random effects (α_i, γ_i) or their distribution $F_{\alpha,\gamma}(\cdot,\cdot \mid \boldsymbol{\theta})$. The latter determine the distribution of n_i across job seekers.[12] Let $s_{j|n}$ and s_n denote the observed conditional shares and the empirical distribution of portfolio sizes, respectively. The equations

$$\mathscr{s}_{j|n}(\boldsymbol{\delta}_0) = \mathbb{E}\left[s_{j|n}\right], \quad j \in \mathcal{J}, \quad n \in \{1, \ldots, J-1\} \tag{32}$$

cleanly pin down the underlying mean utilities $\boldsymbol{\delta}_0$ completely independently of (α_i, γ_i) as a result of the revealed-preference structure induced by Assumption 1. Then, knowing $\boldsymbol{\delta}_0$, the equations

$$\mathscr{s}_n(\boldsymbol{\delta}_0, \boldsymbol{\theta}_0) = \mathbb{E}\left[s_n\right] \tag{33}$$

pin down $\boldsymbol{\theta}_0$. This (conditional) separation underlies our two-step identification argument, which requires the following assumption.

Assumption 5 (Nondegenerate application behavior)**.** The distribution of optimal portfolio sizes satisfies the following conditions.

(a) The pmf of portfolio sizes $\mathscr{s}_n(\boldsymbol{\delta_0}, \boldsymbol{\theta}_0)$ assigns positive probability to at least one interior n:

$$\exists n \in \{1, \ldots, J-1\} : \mathbb{P}\left(n_i = n\right) > 0.$$

[12] Technically, the optimal stopping rule in part (i) of Proposition 1 implies that $\boldsymbol{\delta}$ and $\boldsymbol{\theta}$ jointly determine the pmf of n_i through the distributions of the taste shocks ε_{ij} and the random effects (α_i, γ_i).

(b) The distribution of n_i is not fully degenerate at any single n:

$$\mathbb{P}\left(n_i = n\right) < 1,\ \forall n \in \{0, \ldots, J\}.$$

Part (a) of Assumption 5 guarantees that there is at least one informative portfolio size n in the support such that conditional inclusion probabilities $\mathcal{s}_{j|n}(\boldsymbol{\delta})$ identify $\boldsymbol{\delta}_0$. The stronger condition in part (b) is necessary to ensure enough variation to identify the parameters of the random-effects distribution, $\boldsymbol{\theta}_0$, from the shape of the pmf of n_i, $\mathcal{s}_n(\boldsymbol{\delta}_0, \boldsymbol{\theta}_0)$, given knowledge of mean utilities $\boldsymbol{\delta}_0$. Identification of $\boldsymbol{\delta}_0$ from conditional inclusion probabilities is easier to see in the case $n = 1$ since it follows directly from the well-known MNL inversion result (see Berry, 1994). Set a base alternative $b \in \mathcal{J}$ and normalize location such that $\delta_b = 0$. Then, straightforward algebra implies that, for all $j \in \mathcal{B}_b$, mean utility δ_j is identified as

$$\delta_j = \ln\left(\mathbb{E}\left[s_{j|1}\right]\right) - \ln\left(\mathbb{E}\left[s_{b|1}\right]\right).$$

We can generalize this invertibility property to all interior $n \in \{1, \ldots, J-1\}$.[13] Proposition 2 below formalizes this result.

Proposition 2 (Identification of mean utilities)**.** *Suppose Assumptions 1 to 5 hold. Fix an interior portfolio size $n \in \{1, \ldots, J-1\}$ and a base alternative $b \in \mathcal{J}$. Impose the location normalization $\delta_b = 0$ and let $\boldsymbol{\delta}^b \in \mathbb{R}^{J-1}$ represent the mean utility vector restricted to the free alternatives $j \in \mathcal{B}_b \equiv \mathcal{J} \setminus \{b\}$. Define the restricted vector-valued mapping $\mathcal{s}^b_{\cdot|n} : \mathbb{R}^{J-1} \to (0,1)^{J-1}$, stacking the conditional inclusion probabilities $\mathcal{s}^b_{j|n}(\boldsymbol{\delta}^b) \equiv \mathcal{s}_{j|n}(\boldsymbol{\delta} - \delta_b \boldsymbol{\iota})$ for $j \in \mathcal{B}_b$, where $\boldsymbol{\iota}$ is a $J \times 1$ vector of ones.*

If the distribution of optimal portfolio sizes assigns non-zero mass $\mathbb{P}\left(n_i = n\right) > 0$ at size n, then $\mathbb{E}\left[s_{j|n}\right]$ is observable at the population level. The restricted mapping $\mathcal{s}^b_{\cdot|n}(\cdot)$ is globally injective on $\mathbb{R}^{J-1}$, and thus a bijection from $\mathbb{R}^{J-1}$ onto the image $\mathcal{s}^b_{\cdot|n}(\mathbb{R}^{J-1}) \subseteq (0,1)^{J-1}$. Therefore, the true normalized mean utility vector $\boldsymbol{\delta}^b_0 \in \mathbb{R}^{J-1}$ is identified by inverting the $(J-1) \times (J-1)$ system of equations

$$\mathcal{s}^b_{j|n}(\boldsymbol{\delta}^b_0) = \mathbb{E}\left[s_{j|n}\right], \qquad j \in \mathcal{B}_b, \tag{34}$$

so the true mean utility vector $\boldsymbol{\delta}_0$ is identified up to a location normalization from the conditional inclusion probabilities $\left\{\mathcal{s}_{j|n}(\boldsymbol{\delta}_0)\right\}_{j \in \mathcal{J}} = \left\{\mathbb{E}\left[s_{j|n}\right]\right\}_{j \in \mathcal{J}}$.

Proof. See Appendix D.4. $\square$

Having established identification of $\boldsymbol{\delta}_0$, we now consider identification of $\boldsymbol{\theta}_0$ given knowledge of $\boldsymbol{\delta}_0$. To build intuition, we start with the homogenous baseline model studied in Sections 2.4.2 and 2.4.3. In this case, $F_{\alpha,\gamma}(\cdot, \cdot \mid \boldsymbol{\theta}_0)$ is degenerate at $(\alpha_i, \gamma_i) = \boldsymbol{\theta}_0 = (\alpha_0, \gamma_0) \in (0,1) \times \mathbb{R}_{++}$. We have already established in Section 2.4.3 that γ_0 enters the homogeneous stopping-rule

[13] The boundary case $n = J$ is uninformative since $\mathbb{P}\left(j \in A_i \mid n_i = J\right) \equiv 1$ holds identically for all $(\boldsymbol{\delta}, \boldsymbol{\theta})$.

thresholds (20) linearly, scaling all simultaneously, while α_0 enters nonlinearly, shaping the growth rate of the threshold sequence. Moreover, given $\boldsymbol{\delta_0}$, the mass at $n = 0$ depends only on the entry threshold $\psi_0^1 = \gamma_0/\alpha_0$. Thus, intuitively, the non-application margin uniquely pins down this entry threshold, fixing the level of the threshold sequence, by inverting the equation

$$F_\varepsilon\left(\psi_0^1\right)^{\sum_{\ell \in \mathcal{J}} \exp(\delta_{\ell 0})} = \mathbb{E}\left[s_0\right],$$

while the survivor function $\mathbb{P}\left(n_i \geq n\right)$ for some $n \in \{2, \ldots, J\}$ pins down $\psi_0^n/\psi_0^1 = (1-\alpha_0)^{-(n-1)}$. Proposition 3 below formalizes this intuition.

Proposition 3 (Identification of $\boldsymbol{\theta}_0$ in the homogeneous model)**.** *Suppose Assumptions 1 to 5 hold, $J \geq 2$, $F_{\alpha,\gamma}(\cdot, \cdot \mid \boldsymbol{\theta}_0)$ is degenerate at $(\alpha_i, \gamma_i) = \boldsymbol{\theta}_0 = (\alpha_0, \gamma_0) \in (0,1) \times \mathbb{R}_{++}$, and mean utilities $\boldsymbol{\delta}_0$ are known. Then, $\boldsymbol{\theta}_0$ is point-identified by the distribution of optimal portfolio sizes $\left\{\mathscr{s}_n(\boldsymbol{\delta}_0, \boldsymbol{\theta}_0)\right\}_{n=0}^J = \left\{\mathbb{E}\left[s_n\right]\right\}_{n=0}^J$. In fact, it suffices to observe $\mathbb{E}\left[s_0\right] = \mathbb{P}\left(n_i = 0\right)$ and one interior survivor probability $\mathbb{E}\left[\sum_{k=n}^J s_k\right] = \mathbb{P}\left(n_i \geq n\right)$ for any $n \in \{2, \ldots, J\}$ such that $\mathbb{P}\left(n_i \geq n\right) \in (0,1)$.*

Proof. See Appendix D.5. $\square$

Proposition 3 emphasizes the role of the extensive margin in identifying $\boldsymbol{\theta_0}$ in the homogeneous model. In practice, it is reasonable to expect job applications data to cover applicants only, with no information on the proportion of job seekers who optimally choose not to apply. In this case $\mathbb{E}\left[s_0\right]$ can be thought as unobserved at the population level, so identification cannot be based on this moment. With $J \geq 3$ and sufficient variation in portfolio sizes at the population level, the zero-truncated pmf of n_i still identifies $\boldsymbol{\theta}_0$ at least locally. This result is formalized in Lemma 2 below. Note that once (α_0, γ_0) is known, the unobserved $\mathbb{E}\left[s_0\right]$ can be recovered from Equation (21).

Lemma 2 (Local identification in the zero-truncated homogeneous model)**.** *Suppose the same assumptions from Proposition 3 hold. For $n \in \{1, \ldots, J\}$, let*

$$\mathcal{S}_n(\boldsymbol{\theta}) = \sum_{k=n}^J \mathscr{s}_k(\boldsymbol{\delta_0}, \boldsymbol{\theta}), \qquad \mathscr{s}_n^{>0}(\boldsymbol{\theta}) = \frac{\mathscr{s}_n(\boldsymbol{\delta_0}, \boldsymbol{\theta})}{1 - \mathscr{s}_0(\boldsymbol{\delta_0}, \boldsymbol{\theta})}, \qquad \mathcal{S}_n^{>0}(\boldsymbol{\theta}) = \sum_{k=n}^J \mathscr{s}_k^{>0}(\boldsymbol{\theta}) \tag{35}$$

denote the mappings from $\boldsymbol{\theta} = (\alpha, \gamma) \in (0,1) \times \mathbb{R}_{++}$ to the model-implied survivor function of the full distribution of portfolio size $\mathbb{P}\left(n_i \geq n\right)$, the zero-truncated pmf of portfolio size $\mathbb{P}\left(n_i = n \mid n_i > 0\right)$, and the corresponding zero-truncated survivor function $\mathbb{P}\left(n_i \geq n \mid n_i > 0\right)$, respectively, at $\boldsymbol{\delta}_0$. At the true $\boldsymbol{\theta}_0$, the moment conditions

$$\mathcal{S}_n(\boldsymbol{\theta}_0) = \sum_{k=n}^J \mathbb{E}\left[s_k\right], \qquad \mathscr{s}_n^{>0}(\boldsymbol{\theta}_0) = \frac{\mathbb{E}\left[s_n\right]}{1 - \mathbb{E}\left[s_0\right]}, \qquad \mathcal{S}_n^{>0}(\boldsymbol{\theta}_0) = \frac{\sum_{k=n}^J \mathbb{E}\left[s_k\right]}{1 - \mathbb{E}\left[s_0\right]}$$

hold. Further assume that $J \geq 3$ and there exist three distinct portfolio sizes $1 \leq n_1 < n_2 < n_3 \leq J$ such that $\mathcal{S}_\ell^{>0}(\boldsymbol{\theta}_0) \in (0,1)$ for $\ell \in \{n_1, n_2, n_3\}$. Then, the moment equations

$$\ln\left(\frac{\mathcal{S}_\ell^{>0}(\boldsymbol{\theta}_0)}{\mathcal{S}_{n_2}^{>0}(\boldsymbol{\theta}_0)}\right) = \ln\left(\sum_{k=\ell}^{J} \mathbb{E}\left[s_k\right]\right) - \ln\left(\sum_{k=n_2}^{J} \mathbb{E}\left[s_k\right]\right), \qquad \ell \in \{n_1, n_3\}$$

locally identify $\boldsymbol{\theta}_0$.

Proof. See Appendix D.6. $\square$

What Proposition 3 and Lemma 2 show is that the distribution of optimal portfolio sizes contains enough information to identify $\boldsymbol{\theta}_0$ in the homogeneous model. Intuition naturally suggests that, under some regularity conditions, this result should carry to the full, heterogeneous model where arbitrary dependence of (α_i, γ_i) is allowed. Our comparative-statics discussion in Section 2.4.3 already hints at this: since γ_i enters the stopping-rule thresholds (12) linearly, shifting overall scale, while α_i enters nonlinearly, controlling the sequence's growth rate, the portfolio-size pmf responds in distinct ways to (i) changes in the mean or dispersion of the marginal distributions of α_i and γ_i, and (ii) changes in comonotone or countermonotone dependence between these random effects. What we need to establish local identification of $\boldsymbol{\theta}_0$ in this more general setting is local invertibility at $\boldsymbol{\theta}_0$ of the model-implied, vector-valued mapping from $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^L$, given $\boldsymbol{\delta}_0$, to the portfolio-size pmf. Proposition 4 below formalizes this logic. It is worth noting that, since the pmf sums to one over $n \in \{0, \ldots, J\}$, it yields at most J linearly independent moments that can be used to identify $\boldsymbol{\theta}_0$ as long as $L = \dim(\boldsymbol{\theta}) \leq J$.[14] It is fairly plausible in empirical settings —as in our empirical application in Section 4— that $L < J$, so $\boldsymbol{\theta}_0$ is overidentified.

Following our analysis of comparative statics in Section 2.4.3, we model dependence of the random effects (α_i, γ_i) via copula theory.[15] By Sklar's theorem (Sklar, 1973), any m–dimensional joint cdf $F(x_1, \ldots, x_m)$ with marginals $\{F_\ell(x_\ell)\}_{\ell=1}^m$ admits a representation of the form

$$F(x_1, \ldots, x_m) = C\big(F_1(x_1), \ldots, F_m(x_m)\big), \tag{36}$$

where $C : [0,1]^m \to [0,1]$ is a copula. Moreover, if the marginals are continuous, then this representation is unique in the sense that there is a unique copula $C(u_1, \ldots, u_m)$ that satisfies (36). Therefore, the use of copulas to model dependence is without loss of generality. To model dependence parametrically within our framework, however, we further impose the assumption that the joint cdf of the random effects $F_{\alpha,\gamma}(\alpha_i, \gamma_i \mid \boldsymbol{\theta})$ admits a parametric copula representation

$$F_{\alpha,\gamma}(\alpha_i, \gamma_i \mid \boldsymbol{\theta}) = C\big(F_\alpha(\alpha_i \mid \boldsymbol{\theta}_\alpha), F_\gamma(\gamma_i \mid \boldsymbol{\theta}_\gamma) \mid \boldsymbol{\theta}_C\big), \tag{37}$$

[14] Proposition 4 imposes the stronger requirement that $J \geq L + 1$ to deal with zero-truncation as in Lemma 2.

[15] See, e.g., Trivedi and Zimmer (2007) or Fan and Patton (2014) for accessible introductions to copula modeling.

where F_α and F_γ are the corresponding marginals and we partition $\boldsymbol{\theta} = (\boldsymbol{\theta}_\alpha', \boldsymbol{\theta}_\gamma', \boldsymbol{\theta}_C')'$. This decomposition separates the effects of the marginal distributions of job-offer uncertainty and application costs, captured through $\boldsymbol{\theta}_\alpha$ and $\boldsymbol{\theta}_\gamma$, from the effect of their dependence captured by the copula parameters $\boldsymbol{\theta}_C$. Specifically, it allows us to write the (model-implied, unconditional) portfolio-size pmf as

$$\mathscr{s}_n(\boldsymbol{\delta}, \boldsymbol{\theta}) = \int_{[0,1]^2} \mathscr{s}_{n|\alpha,\gamma}\big(\boldsymbol{\delta}, F_\alpha^{-1}(u \mid \boldsymbol{\theta}_\alpha), F_\gamma^{-1}(v \mid \boldsymbol{\theta}_\gamma)\big)\, \mathrm{d}C\big(u, v \mid \boldsymbol{\theta}_C\big), \tag{38}$$

which is a generalized version of Equation (31) where we imposed a Gaussian copula with a scalar correlation parameter θ_C. In order to establish local identification of $\boldsymbol{\theta}_0$ in Proposition 4, Assumption 6 imposes some regularity conditions on the joint distribution $F_{\alpha,\gamma}(\cdot, \cdot \mid \boldsymbol{\theta})$ and its associated copula and marginals $C\big(\cdot, \cdot \mid \boldsymbol{\theta}_C\big)$, $F_\alpha(\cdot \mid \boldsymbol{\theta}_\alpha)$, $F_\gamma(\cdot \mid \boldsymbol{\theta}_\gamma)$.

Assumption 6 (Random-effects joint distribution regularity conditions)**.** The joint distribution of the random effects (α_i, γ_i), $F_{\alpha,\gamma}(\alpha_i, \gamma_i \mid \boldsymbol{\theta})$, satisfies the following properties:

(a) $F_{\alpha,\gamma}(\alpha_i, \gamma_i \mid \boldsymbol{\theta})$ admits the parametric copula representation (37), where the copula function $C\big(F_\alpha(\alpha_i \mid \boldsymbol{\theta}_\alpha), F_\gamma(\gamma_i \mid \boldsymbol{\theta}_\gamma) \mid \boldsymbol{\theta}_C\big)$ has a continuously differentiable density $c\big(u, v \mid \boldsymbol{\theta}_C\big)$ and its Fisher information matrix is nonsingular at the true $\boldsymbol{\theta}_{C0}$.

(b) The mappings $\boldsymbol{\theta}_\alpha \mapsto F_\alpha(\cdot \mid \boldsymbol{\theta}_\alpha)$ and $\boldsymbol{\theta}_\gamma \mapsto F_\gamma(\cdot \mid \boldsymbol{\theta}_\gamma)$ are locally injective in a neighborhood of their respective true parameters $\boldsymbol{\theta}_{\alpha 0}$ and $\boldsymbol{\theta}_{\gamma 0}$. Equivalently, their Fisher information matrices are nonsingular at $\boldsymbol{\theta}_{\alpha 0}$ and $\boldsymbol{\theta}_{\gamma 0}$, respectively.

(c) The copula density $c\big(u, v \mid \boldsymbol{\theta}_C\big)$ and the marginal quantile functions

$$Q_\alpha(u \mid \boldsymbol{\theta}_\alpha) \equiv F_\alpha^{-1}(u \mid \boldsymbol{\theta}_\alpha), \qquad Q_\gamma(u \mid \boldsymbol{\theta}_\gamma) \equiv F_\gamma^{-1}(u \mid \boldsymbol{\theta}_\gamma)$$

are continuously differentiable in their parameters $\boldsymbol{\theta}_C$, $\boldsymbol{\theta}_\alpha$, and $\boldsymbol{\theta}_\gamma$, respectively, in a neighborhood of their respective true values $\boldsymbol{\theta}_{C0}$, $\boldsymbol{\theta}_{\alpha 0}$, and $\boldsymbol{\theta}_{\gamma 0}$, and are each bounded by an integrable envelope.

Proposition 4 (Local identification of $\boldsymbol{\theta}_0$ in the full model)**.** *Suppose Assumptions 1 to 6 hold, $J \geq L+1$, $F_{\alpha,\gamma}(\cdot, \cdot \mid \boldsymbol{\theta}_0)$ admits the parametric copula representation (37) where the marginals $F_\alpha(\cdot \mid \boldsymbol{\theta}_\alpha)$ and $F_\gamma(\cdot \mid \boldsymbol{\theta}_\gamma)$ and the copula $C(\cdot, \cdot \mid \boldsymbol{\theta}_C)$ are continuously differentiable in $\boldsymbol{\theta}_\alpha$, $\boldsymbol{\theta}_\gamma$, $\boldsymbol{\theta}_C$, respectively, and mean utilities $\boldsymbol{\delta}_0$ are known. Let $\mathcal{S}_n(\boldsymbol{\theta})$, $\mathscr{s}_n^{>0}(\boldsymbol{\theta})$, and $\mathcal{S}_n^{>0}(\boldsymbol{\theta})$ denote the mappings from $\boldsymbol{\theta} = \big(\boldsymbol{\theta}_\alpha', \boldsymbol{\theta}_\gamma', \boldsymbol{\theta}_C'\big)' \in \boldsymbol{\Theta} \subseteq \mathbb{R}^L$ to the model-implied survivor function, zero-truncated pmf, and zero-truncated survivor function defined in Equation (35). Further assume there exist at least $L+1$ distinct portfolio sizes $1 \leq n_0 < n_1 < \cdots < n_L \leq J$ such that $\mathcal{S}_{n_\ell}^{>0}(\boldsymbol{\theta}_0) \in (0,1)$ for all $\ell \in \{0, \ldots, L\}$. Then, the L moment equations*

$$\ln\left(\frac{\mathcal{S}_{n_\ell}^{>0}(\boldsymbol{\theta}_0)}{\mathcal{S}_{n_0}^{>0}(\boldsymbol{\theta}_0)}\right) = \ln\left(\sum_{k=n_\ell}^{J} \mathbb{E}\left[s_k\right]\right) - \ln\left(\sum_{k=n_0}^{J} \mathbb{E}\left[s_k\right]\right), \qquad \ell \in \{1, \ldots, L\}$$

locally identify $\boldsymbol{\theta}_0$.

Proof sketch. See Appendix D.7. $\square$

Finally, we cover identification of β_{w0} less formally since it is standard in the BLP (Berry et al., 1995) tradition estimating discrete choice models of demand for differentiated products in empirical IO. We assume that posted wages w_j and job characteristics $\boldsymbol{x}_j$ are observed by the econometrician, but there exist additional job attributes that are unobserved by the researcher with an effect on mean utilities δ_j that is fully captured by the unobserved scalar index ξ_j. Following Berry (1994), we can treat Equation (8) as a linear estimating equation, where ξ_j is the error term, since we have already established that $\boldsymbol{\delta}_0$ is identified —i.e., we can use a consistent estimate on the left-hand side. Without loss of generality, we assume $\mathbb{E}[\xi_j] = 0$ —including a constant in $\boldsymbol{x}_j$ absorbs a potential nonzero mean of ξ_j. We further assume that job characteristics are predetermined at this stage of the recruitment process and (strongly) exogenous in the sense that

$$\mathbb{E}\left[\xi_j \mid \{\boldsymbol{x}_\ell\}_{\ell \in \mathcal{J}}\right] = 0, \qquad \forall j \in \mathcal{J}. \tag{39}$$

Posted wages, on the other hand, are possibly correlated with both observed and unobserved job attributes through, e.g., wage-setting equilibrium (which is left unmodeled here). Since $\boldsymbol{x}_\ell$ are excluded from Equation (8) defining δ_j for $\ell \neq j$, they are also excluded from (7) defining ex-post utilities u_{ij}. Therefore, provided a nondegenerate first stage such that the standard rank condition is satisfied, the mean independence condition (39) implies —through the law of iterated expectations— that any function $\boldsymbol{z}(\boldsymbol{X}_{-j})$, where $\boldsymbol{X}_{-j}$ stacks $\{\boldsymbol{x}_\ell\}_{\ell \in \mathcal{B}_j}$, provides a valid set of instruments (over)identifying $(\beta_{w0}, \boldsymbol{\beta}_{\boldsymbol{x}}{}')'$ through the set of moment equations

$$\mathbb{E}\left[\begin{pmatrix}\boldsymbol{z}(\boldsymbol{X}_{-j})' & \boldsymbol{x}_j'\end{pmatrix}' \xi_j(\beta_{w0}, \boldsymbol{\beta}_{\boldsymbol{x}})\right] = \boldsymbol{0}, \tag{40}$$

where $\xi_j(\beta_{w0}) \equiv \delta_{j0} - \beta_{w0} \ln(w_j) - \boldsymbol{x}_j'\boldsymbol{\beta}_{\boldsymbol{x}}$.

3.2 Estimation

Based on our sequential identification arguments from the previous section, where we show that $\boldsymbol{\delta}_0$ is identified from conditional inclusion probabilities alone, while, conditional on knowledge of $\boldsymbol{\delta}_0$, $\boldsymbol{\theta}_0$ and β_{w0} are independently identified from the portfolio-size distribution and exclusion of rival's characteristics from a job's ex-post utility, respectively.

3.2.1 Estimating δ_0: Partially rank-ordered logit

3.2.2 Estimating β_{w0}: BLP-IV regression

3.2.3 Estimating θ_0: Method of simulated moments

Let

$$\boldsymbol{m}_i(n_i) = \begin{pmatrix} m_1(n_i) \\ \vdots \\ m_M(n_i) \end{pmatrix}, \tag{41}$$

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 $M \geq L = \dim(\boldsymbol{\theta})$ is the number of moments used in estimation. Note that the population moment

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

is a function only of true parameters $\boldsymbol{\delta}_0$ and $\boldsymbol{\theta}_0$.[16] Similarly, the empirical moment

$$\overline{\boldsymbol{m}}_I(\boldsymbol{n}) = \frac{\sum_{i\in\mathcal{I}} \boldsymbol{m}_i(n_i)}{\sum_{i\in\mathcal{I}} \mathbb{1}\left\{n_i > 0\right\}}, \tag{42}$$

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

$$Q_I(\boldsymbol{\theta} \mid \boldsymbol{\delta}_0, \boldsymbol{n}) = \Big(\boldsymbol{m}(\boldsymbol{\theta} \mid \boldsymbol{\delta}_0) - \overline{\boldsymbol{m}}_I(\boldsymbol{n})\Big)' W \Big(\boldsymbol{m}(\boldsymbol{\theta} \mid \boldsymbol{\delta}_0) - \overline{\boldsymbol{m}}_I(\boldsymbol{n})\Big), \tag{43}$$

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 $\boldsymbol{m}(\boldsymbol{\theta} \mid \boldsymbol{\delta})$ analytically, we construct a frequency simulator $\widehat{\boldsymbol{m}}_S(\boldsymbol{\theta} \mid \boldsymbol{\delta})$. 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 $F_\alpha(\cdot \mid \boldsymbol{\theta}_\alpha)$, and $I \times 1$ vector $\boldsymbol{\gamma}^s$ from $F_\gamma(\cdot \mid \boldsymbol{\theta}_\gamma)$. 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

[16] 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 $\{\mathscr{s}_n(\boldsymbol{\delta}, \boldsymbol{\theta})\}_{n=0}^J$ depend on $\boldsymbol{\theta}$ through the integration over $F_{\alpha,\gamma}(\cdot, \cdot \mid \boldsymbol{\theta})$ on the right-hand side of (15). Here we use the notation $\boldsymbol{m}(\boldsymbol{\theta} \mid \boldsymbol{\delta})$ instead of the more standard $\boldsymbol{m}(\boldsymbol{\delta}, \boldsymbol{\theta})$ to emphasize our interest in estimating $\boldsymbol{\theta}_0$ given knowledge (or a consistent estimate) of $\boldsymbol{\delta}_0$.

simulated data and average over simulations:

$$\widehat{\boldsymbol{m}}_S(\boldsymbol{\theta} \mid \boldsymbol{\delta}) = \frac{1}{S}\sum_{s=1}^{S} \frac{\sum_{i\in\mathcal{I}} \boldsymbol{m}_i(n_i)}{\sum_{i\in\mathcal{I}} \mathbb{1}\{n_i > 0\}}. \tag{44}$$

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

$$Q_{I,S}(\boldsymbol{\theta} \mid \boldsymbol{\delta}, \boldsymbol{n}) = \Big(\widehat{\boldsymbol{m}}_S(\boldsymbol{\theta} \mid \boldsymbol{\delta}) - \overline{\boldsymbol{m}}_I(\boldsymbol{n})\Big)' \widehat{W} \Big(\widehat{\boldsymbol{m}}_S(\boldsymbol{\theta} \mid \boldsymbol{\delta}) - \overline{\boldsymbol{m}}_I(\boldsymbol{n})\Big), \tag{45}$$

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

$$W^* = \mathbb{E}\left[\Big(\widehat{\boldsymbol{m}}_S(\boldsymbol{\theta} \mid \boldsymbol{\delta}) - \overline{\boldsymbol{m}}_I(\boldsymbol{n})\Big)\Big(\widehat{\boldsymbol{m}}_S(\boldsymbol{\theta} \mid \boldsymbol{\delta}) - \overline{\boldsymbol{m}}_I(\boldsymbol{n})\Big)'\right]^{-1}, \tag{46}$$

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

Our MSM estimator of $\boldsymbol{\theta}_0$ 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 $\widehat{\boldsymbol{\theta}}(\widehat{\boldsymbol{\delta}})$ converges in probability to the same limit as the infeasible version $\widehat{\boldsymbol{\theta}}(\boldsymbol{\delta}_0)$ 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.[17] 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 $\boldsymbol{\theta}$ vanish in probability, 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 $\boldsymbol{\delta}_0$ and then $\boldsymbol{\theta}_0$). This, however, is computationally intensive, and is not pursued in our empirical application in Section 4 given its illustrative purpose.

3.3 From structural parameters to elasticities

With consistent estimates of $(\boldsymbol{\delta}_0, \boldsymbol{\theta}_0, \beta_{w0})$ in hand, we are in position to compute the implied vacancy- and firm-level wage elasticities of the job application supply. In principle, we could

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

plug the closed-form solutions for $\mathcal{s}_{j|n}(\boldsymbol{\delta})$ and $\mathcal{s}_n(\boldsymbol{\delta}, \boldsymbol{\theta})$ —together with the closed-form derivatives obtained in Appendix D.2— into Equation (16) to compute the vacancy-level elasticities, and then aggregate to the firm level following Equation (19). The involved combinatorial nature of these closed-form expressions, however, renders this direct, plug-in approach impractical for even moderate J. To deal with the typically large J in observational job-application data, we propose a simulation approach where the model is simulated and solved S times, the relevant quantities —$q_j^s, q^{f\,s}, s_{j|n}^s, s_n^s$— are averaged over simulation draws $s \in \{1, \ldots, S\}$, and finally the resulting simulation estimates are numerically differentiated with respect to $\boldsymbol{\delta}$ and accordingly plugged into their corresponding equations.

While this approach avoids the nested combinatorial sums in the closed forms, it is still computationally intensive and may become impractical for large J. We therefore take advantage of an alternative representation of the firm-level elasticities based on directional derivatives that significantly reduces dimensionality by focusing directly on a $|\mathcal{F}| \times 1$ output vector, where each element is differentiated with respect to a single scalar quantity. For each firm $f \in \mathcal{F}$, let $\boldsymbol{d}^f \in \mathbb{R}^J$ be a column vector with ℓ–th element $d_\ell^f = \mathbb{1}\left\{\ell \in \mathcal{J}^f\right\}$.[18] Consider the directional derivative of $q^f(\boldsymbol{\delta}, \boldsymbol{\theta})$ with respect to $\boldsymbol{\delta}$ in the direction of $\boldsymbol{d}^f$,[19]

$$\left.\frac{\mathrm{d}q^f\left(\boldsymbol{\delta} + c\,\boldsymbol{d}^f, \boldsymbol{\theta}\right)}{\mathrm{d}c}\right|_{c=0} = \sum_{\ell \in \mathcal{J}^f} \frac{\partial q^f\left(\boldsymbol{\delta}, \boldsymbol{\theta}\right)}{\partial \delta_\ell}. \tag{47}$$

The fact that it is equivalent to the sum of the elements of the gradient of $q^f(\boldsymbol{\delta}, \boldsymbol{\theta})$ with respect to the mean utilities of the firm's own vacancies implies we can equivalently compute the firm-level elasticities with respect to a simultaneous wage increase in all the firm's vacancies in (19) as

$$\eta_{q^f, \boldsymbol{w}^f}\left(\boldsymbol{\delta}, \boldsymbol{\theta}, \beta_w\right) = \frac{1}{q^f\left(\boldsymbol{\delta}, \boldsymbol{\theta}\right)} \left.\left(\frac{\mathrm{d}q^f\left(\boldsymbol{\delta} + c\,\boldsymbol{d}^f, \boldsymbol{\theta}\right)}{\mathrm{d}c}\right)\right|_{c=0} \beta_w. \tag{48}$$

See Appendix D.8 for a derivation of these results.

Equation (48) provides a convenient shortcut to computing firm-level elasticities, but its aggregate nature does not admit a decomposition like (17). In order to provide insight into the relative importance of the substitution and portfolio-size mechanisms, we propose computing

[18] Recall we set $\mathcal{J} = \{1, \ldots, J\}$ for notational convenience in Section 2.2. Alternatively, suitably modify the notation to work with a bijection from $\{1, \ldots, J\}$ to a fixed permutation of $\mathcal{J}$ as in Footnote 5.

[19] Here we define the directional derivative of $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ at $\boldsymbol{x}^0 \in U$ in the direction of vector $\boldsymbol{d} \in \mathbb{R}^n$ as

$$\nabla_{\boldsymbol{d}}\, f(\boldsymbol{x}^0) \equiv \lim_{h \to 0} \frac{f\left(\boldsymbol{x}^0 + h\,\boldsymbol{d}\right) - f\left(\boldsymbol{x}^0\right)}{h} = \lim_{h \to 0} \left.\frac{f\left(\boldsymbol{x}^0 + (c+h)\,\boldsymbol{d}\right) - f\left(\boldsymbol{x}^0 + c\,\boldsymbol{d}\right)}{h}\right|_{c=0} \equiv \left.\frac{\mathrm{d}f\left(\boldsymbol{x}^0 + c\,\boldsymbol{d}\right)}{\mathrm{d}c}\right|_{c=0}.$$

We do not follow the standard convention of normalizing by the norm of direction vector $\boldsymbol{d}$ since we are interested in a fixed-scale direction corresponding to an increase of β_w in each $\{\delta_j\}_{j \in \mathcal{J}^f}$ caused by a simultaneous unit increase in each $\{\ln(w_\ell)\}_{\ell \in \mathcal{J}^f}$, i.e., a simultaneous 1% increase in $\{w_\ell\}_{\ell \in \mathcal{J}^f}$.

the firm-level elasticities for the full sample by plugging the simulated quantities into (48), and then randomly drawing a representative subsample of firms of manageable size to compute the decomposition terms in (17) at the vacancy level. Representativeness of the resulting subsample can be directly assessed in terms of the firm-level elasticities. Similarly, the vacancy-level elasticities for the subsample can be aggregated to the firm level using (19) to compare with the single-shift, directional-derivative elasticities from (48) as a sanity check.

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

Our empirical application uses microdata from Trabajando.com, one of the largest private job boards in Chile and with presence throughout Latin America.[20] The dataset contains millions of job applications over more than a decade, with detailed records of job seekers, employers, and job advertisements. For tractability and to keep the empirical exercise illustrative, we restrict attention to a one-quarter window spanning applications between January 1 and March 31, 2018. Administrative records from Trabajando.com covering different periods have been used in other empirical work that exploits the same institutional features discussed below (see, e.g., Banfi and Villena-Roldán, 2019; Banfi et al., 2022; Choi et al., 2025; Banfi et al., 2025).

Institutional setting. The site operates under stable institutional rules throughout the sample period: access is free for job seekers, while firms purchase ad packs or subscription plans to post job advertisements.[21] Each job ad requires the firm to enter a monthly salary offer (net of taxes and contributions), which we henceforth refer to as the wage. Although firms may choose whether to display this wage to job seekers, the information is always recorded internally in the researcher dataset. Similarly, applicants must enter an expected wage when

[20]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.

[21]See advertisement packs and corporate subscription plans for details on current options.

creating their CV, with the option of keeping it hidden from employers. While only a minority of ads display pay publicly, hidden wages remain informative: applicants can filter ads by salary ranges, meaning misreporting would harm employers by reducing relevant matches. Banfi and Villena-Roldán (2019) show that observable job characteristics strongly predict hidden wages, and that applications to hidden-wage ads are indeed wage responsive. This feature is central to our empirical application, since the dataset contains information on both posted and hidden wages.

Representativeness. Although online job boards do not cover the entire labor market, Trabajando.com accounts for a large share of mid- and high-skill vacancies in Chile. Comparisons with nationally representative surveys show that, once postings are weighted by the number of openings, the distribution of offered wages closely matches the wage distribution of newly hired workers in household survey data (Choi et al., 2025). Thus, while the data underrepresent informal and very low-skill employment, they provide meaningful variation in formal-sector labor demand and job seeker behavior.

Information available. The raw data comprise four linked datasets: (i) job advertisements, including wages (posted or hidden), requirements, and job details; (ii) employers, with firm identifiers, industries, and size categories; (iii) users, with rich job seeker information (demographics, education, work history, expected wage, and job search activity); and (iv) applications, which link seekers to ads and record the date of each application. Unlike administrative social security records, the data do not include callbacks or hiring outcomes, so the analysis focuses on application flows. From these raw datasets, we construct an estimation sample tailored to our model, as described below.

Sample construction. For the estimation sample, we restrict attention to unemployed job seekers residing in Chile with declared expected wages between CLP $150,000 and CLP $5,000,000. Search spells are defined using CV updates: a spell begins at the most recent CV modification date in 2017–2018, and we include applications up to 365 days before in it. Spells are segmented when two consecutive applications are more than 90 days apart, and we retain only spells fully contained in calendar year 2018. The resulting estimation sample contains 17,357 job seekers, 8,808 jobs, and 1,167 firms across 55 occupation–region groups.

Data cleaning. Reported ad publication and expiry dates are sometimes inconsistent with observed application dates, and firm identifiers can be duplicated or fragmented when the same employer registers under slightly different names. In addition, a subset of postings are placed by recruiting agencies on behalf of client firms, and others are reposted by universities or public employment offices, which can obscure the identity of the ultimate employer. We address these issues by redefining ad availability spells using application clusters, flagging likely recruiting agencies following the heuristic of Banfi and Villena-Roldán (2019), and partially merging firm identifiers with administrative sources. While these procedures mitigate the most salient inconsistencies, some residual noise remains in employer identity and ad timing. Further details and complete variable descriptions are provided in Appendix G.

Overall, the data combine scale, detail, and credible wage information, making them well suited for analyzing application behavior for estimating firm-level labor supply elasticities in our structural framework.

4.2 Estimation details

Market definition. We treat the whole board as a single market over the sampling period. This choice is grounded on network analysis that reveals the largest connected component of the graph connecting vacancies through job applicants contains more than 99% of the full set of ads. This result is consistent with the high documented connectivity (over 90%) of the graph connecting employers through workers' job-to-job moves in Chilean matched employer-employee administrative records from the unemployment insurance system (Cruz and Rau, 2022), and in line with evidence from average firm-level labor market flows constructed from rich, high-frequency administrative tax records, ranking Chile highest in terms of job reallocation in a sample of 25 OECD countries (Albagli et al., 2023). More granular market definitions, such as the occupation–location bins defining the top-level branches in the nested logit model of Azar et al. (2022), would be problematic in our setting since search is simultaneous and job seekers' decisions to include vacancies from different markets in the optimal portfolio cannot be easily separated.

Specification of random effects distributions. For this empirical application we assume (α_i, γ_i) are i.i.d. across job seekers with

$$\alpha_i \sim \text{Beta}\Big(\theta_1\,\theta_2,\ (1-\theta_1)\,\theta_2\Big), \qquad \gamma_i \sim \text{Gamma}\Big(\theta_4,\ \theta_3\,/\,\theta_4\Big),$$

independently of the i.i.d. EV_1 taste shocks $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iJ})'$. Here $\theta_1 \in (0,1)$ and $\theta_2 > 0$ are the mean and concentration of α_i —so $\mathbb{E}[\alpha_i] = \theta_1$ and $\text{Var}(\alpha_i) = \theta_1(1-\theta_1)/(1+\theta_2)$—, while $\theta_3 > 0$ and $\theta_4 > 0$ are the mean and shape parameter of γ_i —so $\mathbb{E}[\gamma_i] = \theta_3$ and $\text{Var}(\gamma_i) = (\theta_3)^2/\theta_4$. This choice respects the natural supports $\alpha_i \in (0,1)$ and $\gamma_i > 0$, while allowing intuitive interpretation of the parameters in terms of means and dispersion and flexibly accommodating different patterns of application behavior through the tails of the portfolio-size distribution.[22]

Choice of moments. For our MSM estimator, we combine three types of moments of the portfolio-size distribution based on functions

$$m_\ell^{\text{pmf}}(n_i) = \mathbb{1}\left\{n_i = \kappa_\ell^{\text{pmf}}\right\}, \qquad m_\ell^{\text{bin}}(n_i) = \mathbb{1}\left\{\kappa_\ell^{\text{bin}} \le n_i \le \nu_\ell^{\text{bin}}\right\}, \qquad m_\ell^{\text{surv}}(n_i) = \mathbb{1}\left\{n_i \ge \kappa_\ell^{\text{surv}}\right\},$$

each yielding a set of moments defined as the (i) probability mass at single points in the support, (ii) probability mass at nondegenerate bins in the support, and (iii) survivor probabilities at single points in the support, respectively:

$$\mathbb{E}\left[m_\ell^{\text{pmf}}(n_i) \mid n_i > 0\right] = \mathbb{P}\left(n_i = \kappa_\ell^{\text{pmf}} \mid n_i > 0\right) \tag{49}$$

$$\mathbb{E}\left[m_\ell^{\text{bin}}(n_i) \mid n_i > 0\right] = \mathbb{P}\left(\kappa_\ell^{\text{bin}} \le n_i \le \nu_\ell^{\text{bin}} \mid n_i > 0\right) \tag{50}$$

$$\mathbb{E}\left[m_\ell^{\text{surv}}(n_i) \mid n_i > 0\right] = \mathbb{P}\left(n_i \ge \kappa_\ell^{\text{surv}} \mid n_i > 0\right). \tag{51}$$

Table 1 below summarizes our chosen moments.

Table 1. MSM targeted moments

ℓ	κ_ℓ	ν_ℓ	Type	$\mathbb{E}\left[m_\ell(n_i) \mid n_i > 0\right]$
1	1		pmf	$\mathbb{P}\left(n_i = 1 \mid n_i > 0\right)$
2	3		pmf	$\mathbb{P}\left(n_i = 3 \mid n_i > 0\right)$
3	5		pmf	$\mathbb{P}\left(n_i = 5 \mid n_i > 0\right)$
4	7		pmf	$\mathbb{P}\left(n_i = 7 \mid n_i > 0\right)$
5	10		survivor	$\mathbb{P}\left(n_i \ge 10 \mid n_i > 0\right)$
6	14		survivor	$\mathbb{P}\left(n_i \ge 14 \mid n_i > 0\right)$

Notes: Description of the $M = 6$ targeted moments chosen to define the MSM criterion function from the set of moment candidates spanned by Equations (49) to (51).

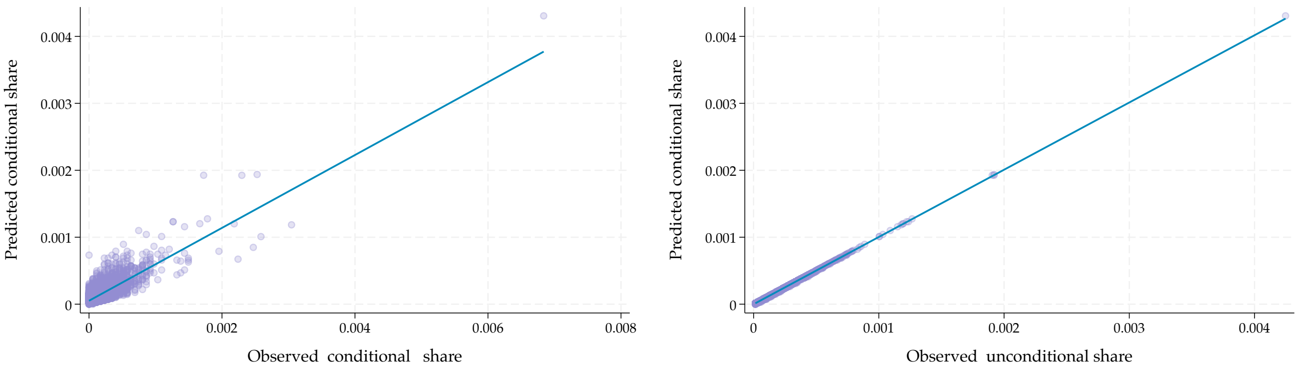
4.3 Empirical results

This section proceeds in four parts. We first estimate the mean utilities and assess the fit of the partially rank-ordered logit (PROL) model. We then examine how posted wages enter

[22]As a robustness check, we consider alternative Beta–Exponential and Uniform–Uniform heterogeneity specifications. The best fit is obtained under our preferred Beta–Gamma specification.

mean utilities, comparing the resulting estimates with those from benchmark multinomial and nested logit models. Next, we use these results to recover the parameters of the random-effects distribution and visualize the implied heterogeneity. Finally, we translate the structural estimates into firm-level wage elasticities.

Figure 4. Fit of the partially rank-ordered logit MLE



(a) $s_{j|1}(\widehat{\boldsymbol{\delta}}_{\mathrm{PROL}})$ vs $s_{j|1}$ **(b)** $s_{j|1}(\widehat{\boldsymbol{\delta}}_{\mathrm{PROL}})$ vs s_j

Notes: Scatter plots of the PROL predicted application share conditional on one application per job seeker against (a) its sample counterpart, the observed conditional share, and (b) the observed unconditional share. We include Panel (b) as a reassurance of the numerical convergence of our estimation routine. The correlations are Corr $\left(s_{j|1}(\widehat{\boldsymbol{\delta}}_{\mathrm{PROL}}), s_{j|1}\right) = 0.8044$ in Panel (a) and Corr $\left(s_{j|1}(\widehat{\boldsymbol{\delta}}_{\mathrm{PROL}}), s_j\right) = 0.9997$ in Panel (b). We predict only the application shares conditional on a portfolio size of one because of its straightforward combination. Higher order inclusion probabilities.

Table 2. Impact of log wage on the ex-post mean utility derived from a job

	PROL		MNL		NL	
	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Log monthly salary (CLP)	0.016	0.363***	0.002	0.302***	0.005	0.429***
	(0.026)	(0.100)	(0.022)	(0.088)	(0.016)	(0.051)
Salary disclosure	-0.042	-0.031	-0.007	0.002	0.031*	0.044**
	(0.037)	(0.037)	(0.025)	(0.025)	(0.019)	(0.019)
Number of vacancies	0.005***	0.006***	0.004***	0.005***	0.000	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Required experience	-0.054***	-0.081***	-0.054***	-0.078***	-0.021***	-0.055***
	(0.007)	(0.011)	(0.007)	(0.010)	(0.005)	(0.007)
Paid advertisement	0.264***	0.221***	0.286***	0.250***	0.034	-0.017
	(0.047)	(0.049)	(0.044)	(0.045)	(0.034)	(0.036)
Permanent contract	0.072***	0.051**	0.057***	0.039*	-0.005	-0.030**
	(0.021)	(0.022)	(0.019)	(0.020)	(0.014)	(0.014)
Full-time job	0.067	0.057	0.051	0.042	0.039*	0.027
	(0.042)	(0.042)	(0.031)	(0.032)	(0.021)	(0.022)
Part-time job	0.063	0.214***	0.034	0.165***	0.082**	0.268***
	(0.056)	(0.072)	(0.050)	(0.062)	(0.034)	(0.040)
High education requirement ($\geq$ university)	0.049*	-0.086*	0.043	-0.074*	-0.000	-0.166***
	(0.029)	(0.047)	(0.027)	(0.043)	(0.019)	(0.028)
High education requirement ($\leq$ high school)	0.060**	0.094***	0.061**	0.091***	0.062***	0.104***
	(0.030)	(0.031)	(0.026)	(0.028)	(0.018)	(0.020)
High computer skills requirement	-0.078**	-0.095***	-0.038	-0.053**	-0.038**	-0.058***
	(0.033)	(0.034)	(0.023)	(0.024)	(0.016)	(0.017)
No computer skills required	0.028	0.066**	0.008	0.041	-0.037**	0.009
	(0.028)	(0.031)	(0.026)	(0.028)	(0.018)	(0.020)
Large firm (>1,000 emp.)	0.150***	0.144***	0.151***	0.145***	0.104***	0.096***
	(0.024)	(0.024)	(0.023)	(0.023)	(0.016)	(0.017)
Small firm (1–150 emp.)	-0.059**	-0.075***	-0.034*	-0.047**	0.013	-0.007
	(0.027)	(0.028)	(0.020)	(0.020)	(0.014)	(0.014)
Kleibergen-Paap		88.83		88.83		88.83
Observations	8,800	8,800	8,800	8,800	8,800	8,800

Notes: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Columns (1), (3), and (5) report the estimated coefficients of ordinary least squares regressions (OLS) of mean utilities on log wages and job characteristics, while Columns (2), (4), and (6) report results from analogous two-stage least squares instrumental variables regressions (IV). The dependent variables are the maximum likelihood estimates of mean utilities from the partially rank-ordered logit model (PROL) in Columns (1) and (2), the standard multinomial logit model (MNL) in Columns (3) and (4), and the nested logit model (NL) in Columns (5) and (6). All regressions control for occupation $\times$ region and first word of job title fixed effects. BLP instruments are sums of the characteristics of jobs posted by rival firms. Other own-job characteristics are not used as excluded instruments because of colinearity with the rival-firm instruments. Even if the NL model assumes a more granular market definition, the market definition for instrumental variable construction treats the whole job board as a single market for consistency with the PROL model.

Table 3. Estimates of the random-effects distribution parameters

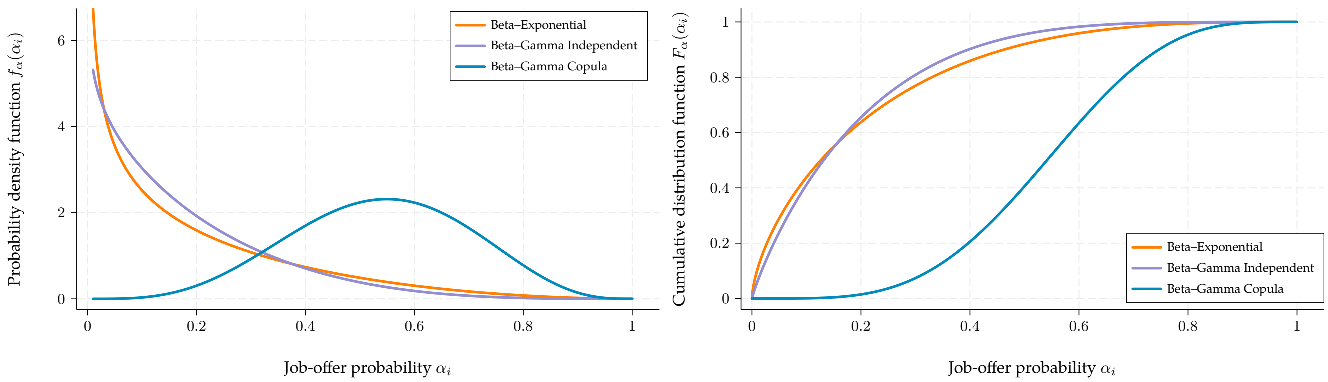
Panel (a)		Independent		Copula
	Homogeneous	Beta–Exponential	Beta–Gamma	Beta–Gamma
	(1)	(2)	(3)	(4)
Estimates				
θ_1	0.7482	0.1865	0.1742	0.5383
θ_2	0.0968	3.4717	5.1058	8.8311
θ_3		0.0954	48.6479	9.9681
θ_4			1.0290	0.6245
θ_5				0.8536
Objective function	5,317.0888	21.2028	5.4828	1.0134
Implied marginals				
$\mathbb{E}[\alpha_i]$	0.7482	0.1865	0.1742	0.5383
$\text{Var}(\alpha_i)$	0	0.0339	0.0236	0.0253
$\mathbb{E}[\gamma_i]$	0.0968	10.4799	48.6479	9.9681
$\text{Var}(\gamma_i)$	0	109.8275	2,299.9073	159.1107

Panel (b)				Independent				Copula	
		Homogeneous		Beta–Exponential		Beta–Gamma		Beta–Gamma	
	Empirical	Simulated	Contrib.	Simulated	Contrib.	Simulated	Contrib.	Simulated	Contrib.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Moments									
Moment 1	0.3871	0.4852	-0.0568	0.3807	0.2013	0.3841	0.0904	0.3866	0.0472
Moment 2	0.1120	0.0320	0.2945	0.1158	0.0410	0.1132	0.0311	0.1122	-0.0063
Moment 3	0.0498	0.0000	0.2284	0.0497	0.0015	0.0519	0.2891	0.0499	-0.0032
Moment 4	0.0270	0.0000	0.1227	0.0254	0.1126	0.0265	0.0331	0.0270	-0.0014
Moment 5	0.0904	0.0000	0.4163	0.0876	0.3387	0.0902	0.0384	0.0893	0.7287
Moment 6	0.0496	0.0000	-0.0052	0.0519	0.3049	0.0515	0.5179	0.0500	0.2349

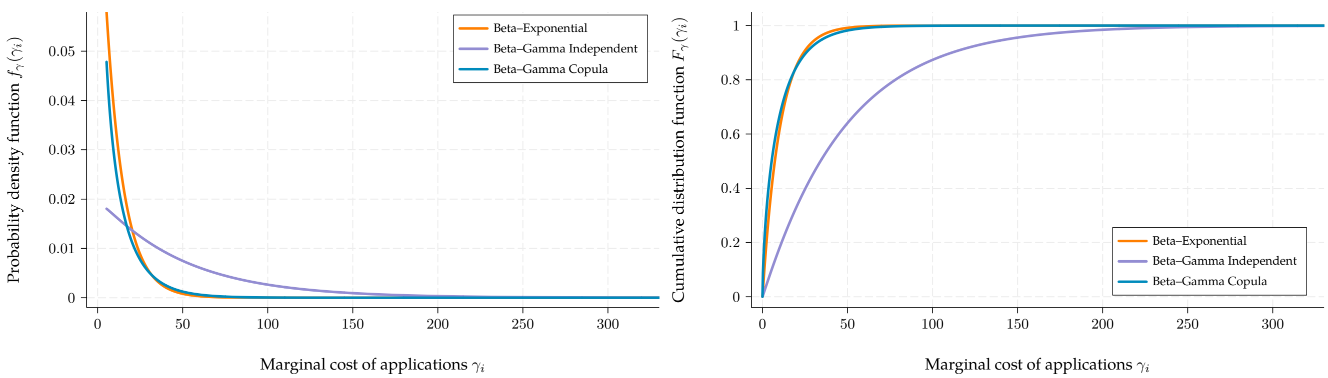
Notes: Method-of-simulated-moments estimates of the parameters $\boldsymbol{\theta}$ of the joint distribution $F_{\alpha,\gamma}(\cdot,\cdot \mid \boldsymbol{\theta})$ of the random effects (α_i, γ_i). The bootstrap estimate of the optimal weight matrix has full rank 6 and condition number 16.2688. Targeted moments correspond to those listed in Table 1.

The implied distribution of α_i indicates that roughly half of job seekers have offer probabilities below 0.5, while γ_i shows substantial mass near zero, consistent with heterogeneity in application costs. Figure 5 visualizes these implied distributions for the three random-effects specifications.

Figure 5. Marginal distributions of the random effects implied by the MSM estimates



(a) Distribution of α_i

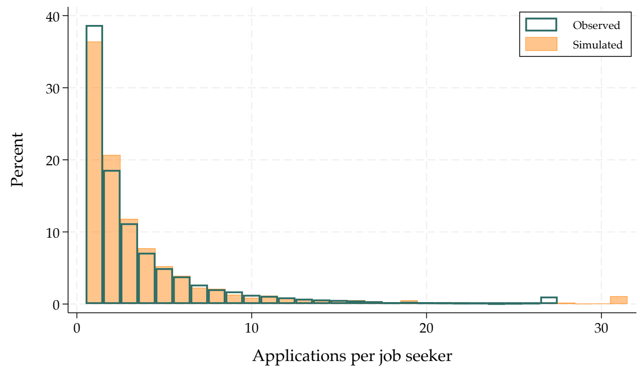


(b) Distribution of γ_i

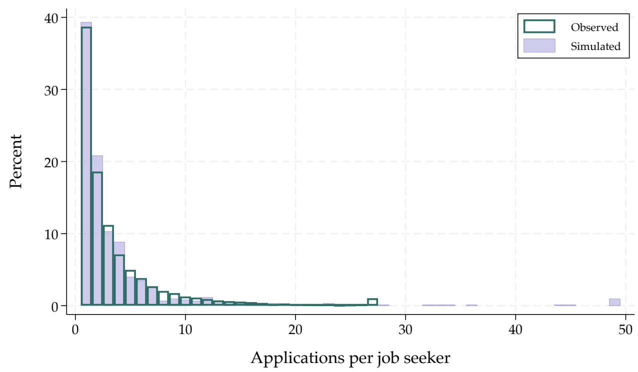
Notes: Marginal distributions of the random effects (α_i, γ_i) implied by the MSM estimates of $\boldsymbol{\theta}$ in Table 3 for the nondegenerate specifications (i) independent Beta–Exponential, (ii) independent Beta–Gamma, and Beta–Gamma copula. Supports are visually truncated at $\alpha_i = 0.01$ and $\gamma_i = 5$ in the pdf plots for readability.

To assess how well these estimated parameters reproduce the empirical distribution of portfolio sizes, Figure 6 compares the simulated and observed conditional pmfs of applications per job seeker.

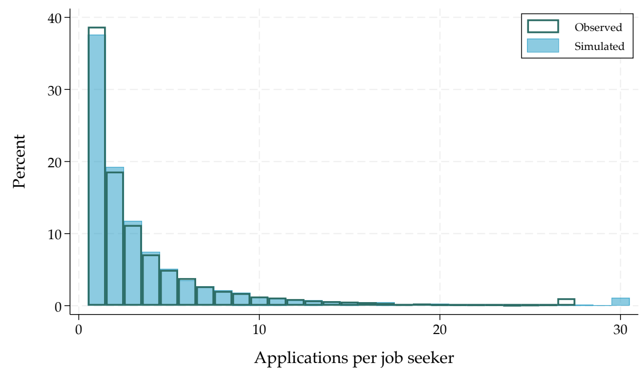
Figure 6. Full conditional pmf fit of the structural parameters



(a) Beta–Exponential



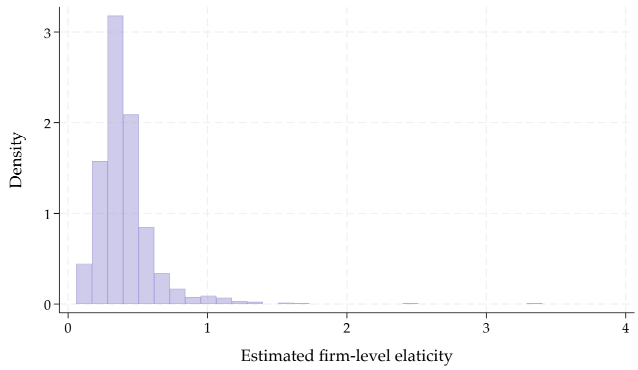
(b) Beta-Gamma Independent



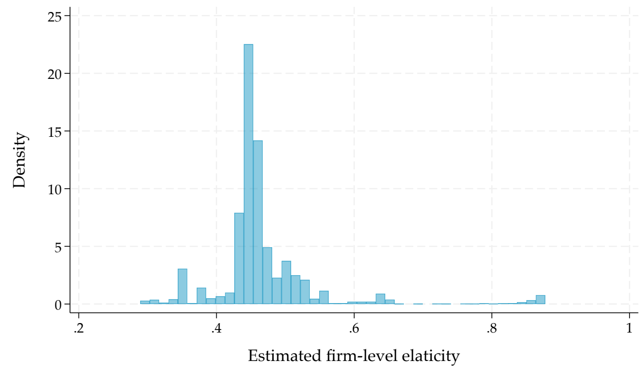
(c) Beta–Gamma Copula

Notes: Histograms of the simulated versus empirical portfolio size distributions, conditional on applying $n_i > 0$. Simulated conditional distributions are obtained from one simulation draw of 17,404 job seekers under the data-generating process defined by the PROL MLE of $\boldsymbol{\delta}$ from Figure 7 and the MSM estimates of $\boldsymbol{\theta}$ from Table 3. Empirical supports are censored at the corresponding 99–th percentile. This figure, showing the full pmf of portfolio sizes mixes targeted un untargeted moments by our MSM estimator, partially serving as a model fit diagnostic.

Figure 7. Distribution of the firm-level elasticity estimates



(a) Multi-application model
(PROL + MSM)



(b) Single-application model
(NL)

Notes: Histograms of the estimated firm-level elasticities from our full multi-application model in Panel (a), and from benchmark single-application nested logit model in Panel (b). The unit of observation is the firm in Panel (a), and firm by market-week in Panel (b). The mean utility estimates underlying the computation of elasticities in Panels (a) and (b) correspond to those used as the dependent variable in Columns (2) and (6) of Table 2, respectively. The MSM estimates underlying the elasticity computations in Panel (a) correspond to those in Column (4) of Panel (a) of Table 3. The firm-level elasticities in Panel (a) are computed through the simulation–numerical differentiation procedure for the directional derivatives described in Section 3.3. Elasticities in panel (b) are based on the closed-form expressions in Azar et al. (2022).

Our deliberately sparse nested logit specification yields substantially more concentrated elasticity estimates than our multi-application framework, with an interquartile range less than one-eighth as large. While differences in model specification contribute to this pattern, the results suggest that accounting for endogenous portfolio choices may reveal firm-level heterogeneity in labor market power that is difficult to capture in single-application models.

Firms that are systematically ranked near the top of applicants' preference distributions exhibit lower estimated labor-supply elasticities. In our framework, these firms are almost always included in applicants' portfolios regardless of wage changes, so the portfolio-size mechanism is effectively inactive. In contrast, less-preferred firms benefit from both substitution and portfolio-size adjustments, resulting in higher elasticities. This pattern is consistent with an endogenous form of monopsony power: firms with higher baseline attractiveness face less wage-sensitive application supply because they are already near full inclusion.

4.4 Discussion

Our empirical application is illustrative: It is not meant to provide definitive estimates of monopsony power in Chile, but to demonstrate how our basic framework can be applied in settings with large numbers of job seekers and job vacancies. We view our results from Section 4.3 as an open invitation to start giving serious consideration to the inherent uncertainty in the job application process. More flexible models may produce better estimates; here, we have taken a first step towards a better understanding of how this uncertainty drives multiple-application behavior, adding an additional mechanism through which the number of applications to the firm varies with posted wages. With this objective in mind, we proceed to discuss some limitations that, together with the theoretical limitations explored in Section 2.4, qualify the interpretation of our estimates.

Interpretation of application elasticities. Because the theoretical framework models applications rather than employment, the elasticities we estimate empirically should be interpreted as proxies for labor supply elasticities rather than direct measures. They are informative about monopsony power to the extent that applications are a necessary step toward hires, but they omit subsequent stages such as employers' selection among applicants. Thus, the empirical results should be understood as measuring the responsiveness of applicant inflows to wages, which under reasonable assumptions provides a lower-bound proxy for the wage elasticity of labor supply.

Voluntary wage posting. Posting wages is voluntary on the platform. Salary disclosure decisions are unlikely to be exogenous. As emphasized by Banfi and Villena-Roldán (2019), this generates a selection effect: firms that anticipate strong applicant inflows without revealing wages may withhold that information, while firms seeking to attract more applicants may disclose it. Their evidence shows, however, that hidden-wage ads attract more applications when the hidden wage (which we do observe in the data) is higher. This suggests that voluntary wage posting does not prevent job seekers from inferring which jobs offer higher wages.

Firms that post higher wages may also differ systematically in unobserved amenities, and the decision to reveal the wage may itself reflect underlying firm characteristics. To address this concern, we control for salary disclosure in the linear-regression stage of our estimation procedure. Our IV strategy further alleviates this concern to the extent that our BLP instruments satisfy the exclusion restriction when salary disclosure correlates with wages and unobserved job characteristics. As in all applications of this approach, instrument validity is an identifying assumption that cannot be directly tested.

Employer identification. Firms are not perfectly identified in the raw job board data in the sense that some employers appear under multiple IDs due to spelling differences in reported names. We have partially addressed this using external administrative records that include firm names, but our fuzzy matching admits refinement. In addition, some job ads are posted by recruiting agencies or reposted by higher-education and public institutions on behalf of different employers. While these issues complicate the mapping between vacancies and employers for elasticity aggregation, they do not affect the vacancy-level analysis at the heart of structural parameter estimation in this paper. Following Banfi and Villena-Roldán (2019), recruitment agency postings can be partially controlled for, and future work using text analysis could further refine employer identification in the data. These limitations reinforce our framing of the empirical exercise as illustrative rather than definitive.

External validity. The analysis relies on online job postings, which may not be representative of the full set of vacancies in the labor market. Jobs advertised online may differ in occupation, formality, or applicant pool composition compared to offline vacancies. For this reason, the estimates reported here should be interpreted as illustrative of the methodology rather than as definitive measures of monopsony power.

5 Conclusion

References

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

Albagli, E., A. Chovar, E. Luttini, C. Madeira, A. Naudon, and M. Tapia (2023): "Labor market flows: Evidence for Chile using microdata from administrative tax records," *Latin American Journal of Central Banking*, 4, 100102.

Ali, S. N. and R. I. Shorrer (2025): "Hedging When Applying: Simultaneous Search with Correlation," *American Economic Review*, 115, 571–98.

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. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *The RAND Journal of Economics*, 25, 242–262.

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.

Berry, S. T., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

Brown, Z. Y. and J. Jeon (2024): "Endogenous Information and Simplifying Insurance Choice," *Econometrica*, 92, 881–911.

Burdett, K. and D. T. Mortensen (1998): "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, 39, 257–273.

Burdett, K., S. Shi, and R. Wright (2001): "Pricing and Matching with Frictions," *Journal of Political Economy*, 109, 1060–1085.

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.

Cruz, G. and T. Rau (2022): "The effects of equal pay laws on firm pay premiums: Evidence from Chile," *Labour Economics*, 75, 102135.

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.

Dube, A., J. Jacobs, S. Naidu, and S. Suri (2020): "Monopsony in Online Labor Markets," *American Economic Review: Insights*, 2, 33–46.

Fan, Y. and A. J. Patton (2014): "Copulas in Econometrics," *Annual Review of Economics*, 6, 179–200.

Galenianos, M. and P. Kircher (2009): "Directed Search with Multiple Job Applications," *Journal of Economic Theory*, 144, 445–471.

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

Le Barbanchon, T., R. Rathelot, and A. Roulet (2020): "Gender Differences in Job Search: Trading off Commute against Wage*," *The Quarterly Journal of Economics*, 136, 381–426.

Manning, A. (2011): "Imperfect Competition in the Labor Market," in *Handbook of Labor Economics*, ed. by D. Card and O. Ashenfelter, Elsevier, vol. 4, 973–1041.

Marinescu, I. and R. Rathelot (2018): "Mismatch Unemployment and the Geography of Job Search," *American Economic Journal: Macroeconomics*, 10, 42–70.

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.

Matějka, F. and A. McKay (2015): "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *American Economic Review*, 105, 272–98.

McFadden, D. (1974): "Chapter Four: Conditional Logit Analysis of Qualitative Choice Behavior," Academic Press, Frontiers in Econometrics, 105–142.

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

MORTENSEN, D. T. (2003): *Wage Dispersion: Why Are Similar Workers Paid Differently?*, The MIT Press.

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.

SKLAR, A. (1973): "Random variables, joint distribution functions, and copulas," *Kybernetika*, 9, 449–460.

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

TRIVEDI, P. K. AND D. M. ZIMMER (2007): "Copula Modeling: An Introduction for Practitioners," *Foundations and Trends in Econometrics*, 1, 1–111.

VAN DER VAART, A. W. (1998): *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

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 $\ell \in \Lambda_{t-1}$,[23]

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

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

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

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 $n_i + 1$.

[23] 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 application supply function

This appendix provides a full derivation of the application supply function, the conditional application share function, and the probability mass function (pmf) of the number of applications in Equations (9), (10) and (15), 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 $s_{j|n}(\boldsymbol{\delta})$ from $\boldsymbol{\delta}$ to $\mathbb{P}\left(j \in A_i \mid n_i = n\right)$, and (ii) a mapping $s_n(\boldsymbol{\delta}, \boldsymbol{\theta})$ from $\boldsymbol{\delta}$ and $\boldsymbol{\theta}$ to $\mathbb{P}\left(n_i = n\right)$. These mappings follow directly from parts (ii) and (i) of Proposition 1, respectively.

C.1 Conditional application share function

Consider first the conditional (expected) application share function $s_{j|n}(\boldsymbol{\delta})$. 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 $s_{j|n}(\boldsymbol{\delta})$— 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 i.i.d. 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

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

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

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

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.

The "conditional inclusion probabilities" in Equation (C.5), $s_{j|n}(\boldsymbol{\delta})$, generalize the well-known choice probabilities in the multinomial logit model.[24] We can easily verify that we get the standard choice probability for $n = 1$:

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

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

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 $s_{j|n}(\boldsymbol{\delta})$ satisfy the following recursive relation. We can rewrite (C.5) as

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

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

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

[24] Here, we use the word "inclusion" to emphasize that the underlying (conditional) event is $\{j \in A_i \mid n_i = n\}$, i.e. $\{j\ \texttt{is included in the portfolio} \mid n\ \texttt{alternatives are included}\}$, whereas in the standard MNL model the underlying event is $\{j\ \texttt{is chosen}\}$.

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

$$\vdots$$

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

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 from 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, $s_n(\boldsymbol{\delta}, \boldsymbol{\nu})$, 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).[25] 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 (12). 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

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

[25] 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.

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

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

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

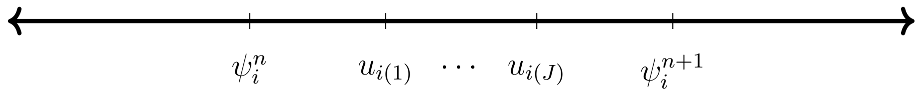
$$\sum_{A\in\mathcal{R}_s(\mathcal{J})} \prod_{\ell\in A} \left[F_{u_\ell}(\psi_i^{n+1}) - F_{u_\ell}(\psi_i^n)\right] \sum_{t\in\tau_n^s} \sum_{B\in\mathcal{R}_t(\mathcal{J}\setminus A)} \prod_{p\in B} F_{u_p}(\psi_i^n) \prod_{q\in\mathcal{J}\setminus(A\cup B)} \left[1 - F_{u_q}(\psi_i^{n+1})\right],$$

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

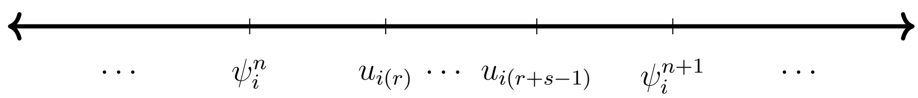
$$\begin{aligned}
\jmath_{n|\alpha,\gamma}(\boldsymbol{\delta},\alpha_i,\gamma_i) &= \mathbb{P}\left(n_i = n \mid \alpha_i,\gamma_i\right) \\
&= \sum_{s=0}^{J} \sum_{A\in\mathcal{R}_s(\mathcal{J})} \prod_{\ell\in A} \left[F_{u_\ell}\big(\psi_i^{n+1}\big) - F_{u_\ell}\big(\psi_i^n\big)\right] \\
&\qquad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

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 (12). 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 application 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

$$\eta_{q_j, w_\ell}(\boldsymbol{\delta}, \boldsymbol{\theta}, \beta_w) = \frac{\partial \ln\left(q_j(\boldsymbol{\delta}, \boldsymbol{\theta})\right)}{\partial \ln\left(w_\ell\right)}$$

$$= \frac{1}{q_j(\boldsymbol{\delta}, \boldsymbol{\theta})} \frac{\partial q_j(\boldsymbol{\delta}, \boldsymbol{\theta})}{\partial \ln(w_\ell)}$$

$$= \frac{1}{q_j(\boldsymbol{\delta}, \boldsymbol{\theta})} \frac{\partial q_j(\boldsymbol{\delta}, \boldsymbol{\theta})}{\partial \delta_\ell} \frac{\partial \delta_\ell}{\partial \ln(w_\ell)}$$

$$= \frac{1}{q_j(\boldsymbol{\delta}, \boldsymbol{\theta})} \left[I \sum_{n=1}^{J} \frac{\partial s_{j|n}(\boldsymbol{\delta})}{\partial \delta_\ell} s_n(\boldsymbol{\delta}, \boldsymbol{\theta}) + s_{j|n}(\boldsymbol{\delta}) \frac{\partial s_n(\boldsymbol{\delta}, \boldsymbol{\theta})}{\partial \delta_\ell} \right] \beta_w, \tag{D.1}$$

where the last equality follows from partially differentiating Equation (9) with respect to δ_ℓ and Equation (7) —for job ℓ— with respect to $\ln(w_\ell)$.

We can further decompose the vacancy-level elasticity into a substitution or composition component and an intensive margin effect on the distribution of portfolio sizes. Let

$$\eta_{s_{j|n}, w_\ell}(\boldsymbol{\delta}, \beta_w) = \frac{\partial \ln(s_{j|n}(\boldsymbol{\delta}))}{\partial \ln(w_\ell)}$$

$$= \frac{1}{s_{j|n}(\boldsymbol{\delta})} \frac{\partial s_{j|n}(\boldsymbol{\delta})}{\partial \ln(w_\ell)}$$

$$= \frac{1}{s_{j|n}(\boldsymbol{\delta})} \frac{\partial s_{j|n}(\boldsymbol{\delta})}{\partial \delta_\ell} \frac{\partial \delta_\ell}{\partial \ln(w_\ell)}$$

$$= \frac{1}{s_{j|n}(\boldsymbol{\delta})} \frac{\partial s_{j|n}(\boldsymbol{\delta})}{\partial \delta_\ell} \beta_w \tag{D.2}$$

represent the elasticity of the conditional inclusion probability of job j with respect to the wage of vacancy ℓ. Similarly, denote the elasticity of the pmf of portfolio sizes at $n_i = n$ with respect to w_ℓ by

$$\eta_{s_n, w_\ell}(\boldsymbol{\delta}, \boldsymbol{\theta}, \beta_w) = \frac{\partial \ln(s_n(\boldsymbol{\delta}, \boldsymbol{\theta}))}{\partial \ln(w_\ell)}$$

$$= \frac{1}{s_n(\boldsymbol{\delta}, \boldsymbol{\theta})} \frac{\partial s_n(\boldsymbol{\delta}, \boldsymbol{\theta})}{\partial \ln(w_\ell)}$$

$$= \frac{1}{s_n(\boldsymbol{\delta}, \boldsymbol{\theta})} \frac{\partial s_n(\boldsymbol{\delta}, \boldsymbol{\theta})}{\partial \delta_\ell} \frac{\partial \delta_\ell}{\partial \ln(w_\ell)}$$

$$= \frac{1}{s_n(\boldsymbol{\delta}, \boldsymbol{\theta})} \frac{\partial s_n(\boldsymbol{\delta}, \boldsymbol{\theta})}{\partial \delta_\ell} \beta_w. \tag{D.3}$$

Then, we can rewrite Equation (D.1) as

$$\eta_{q_j, w_\ell}(\boldsymbol{\delta}, \boldsymbol{\theta}, \beta_w) = \sum_{n=1}^{J} \frac{I s_{j|n}(\boldsymbol{\delta}) s_n(\boldsymbol{\delta}, \boldsymbol{\theta})}{q_j(\boldsymbol{\delta}, \boldsymbol{\theta})} \frac{1}{s_{j|n}(\boldsymbol{\delta})} \frac{\partial s_{j|n}(\boldsymbol{\delta})}{\partial \delta_\ell} \beta_w + \sum_{n=1}^{J} \frac{I s_{j|n}(\boldsymbol{\delta}) s_n(\boldsymbol{\delta}, \boldsymbol{\theta})}{q_j(\boldsymbol{\delta}, \boldsymbol{\theta})} \frac{1}{s_n(\boldsymbol{\delta}, \boldsymbol{\theta})} \frac{\partial s_n(\boldsymbol{\delta}, \boldsymbol{\theta})}{\partial \delta_\ell} \beta_w$$

$$= \sum_{n=1}^{J} \omega_{n|j}(\boldsymbol{\delta},\boldsymbol{\theta})\, \eta_{s_{j|n},w_\ell}(\boldsymbol{\delta},\beta_w) + \sum_{n=1}^{J} \omega_{n|j}(\boldsymbol{\delta},\boldsymbol{\theta})\, \eta_{s_n,w_\ell}(\boldsymbol{\delta},\boldsymbol{\theta},\beta_w) \tag{D.4}$$

where the weights

$$\omega_{n|j}(\boldsymbol{\delta},\boldsymbol{\theta}) = \frac{I s_{j|n}(\boldsymbol{\delta}) s_n(\boldsymbol{\delta},\boldsymbol{\theta})}{q_j(\boldsymbol{\delta},\boldsymbol{\theta})} \tag{D.5}$$

measure the proportion of applications to job j that (in expectation) come from applicants with a portfolio of size n. Note that

$$\sum_{n=1}^{J} \omega_{n|j}(\boldsymbol{\delta},\boldsymbol{\theta}) = 1 \tag{D.6}$$

trivially by Equation (9). The first term on the right-hand side of Equation (D.4) captures substitution across vacancies for a fixed number of applications per job seeker, while the second term captures adjustments in each individual's number of applications.

The elasticity of the aggregate supply of applications at the firm level with respect to the wage of job ℓ is

$$\begin{aligned}
\eta_{q^f,w_\ell}(\boldsymbol{\delta},\boldsymbol{\theta},\beta_w) &= \frac{\partial \ln\left(q^f(\boldsymbol{\delta},\boldsymbol{\theta})\right)}{\partial \ln(w_\ell)} \\
&= \frac{1}{q^f(\boldsymbol{\delta},\boldsymbol{\theta})} \frac{\partial q^f(\boldsymbol{\delta},\boldsymbol{\theta})}{\partial \ln(w_\ell)} \\
&= \frac{1}{q^f(\boldsymbol{\delta},\boldsymbol{\theta})} \sum_{j\in\mathcal{J}^f} \frac{\partial q_j(\boldsymbol{\delta},\boldsymbol{\theta})}{\partial \ln(w_\ell)} \\
&= \frac{1}{q^f(\boldsymbol{\delta},\boldsymbol{\theta})} \sum_{j\in\mathcal{J}^f} q_j(\boldsymbol{\delta},\boldsymbol{\theta})\, \eta_{q_j,w_\ell}(\boldsymbol{\delta},\boldsymbol{\theta},\beta_w),
\end{aligned} \tag{D.7}$$

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

$$\eta_{q^f,\boldsymbol{w}^f}(\boldsymbol{\delta},\boldsymbol{\theta},\beta_w) = \frac{1}{q^f(\boldsymbol{\delta},\boldsymbol{\theta})} \sum_{\ell\in\mathcal{J}^f} \sum_{j\in\mathcal{J}^f} q_j(\boldsymbol{\delta},\boldsymbol{\theta})\, \eta_{q_j,w_\ell}(\boldsymbol{\delta},\boldsymbol{\theta},\beta_w). \tag{D.8}$$

D.2 Closed-form derivatives

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

$$
\begin{aligned}
\frac{\partial \mathscr{s}_n(\boldsymbol{\delta},\boldsymbol{\theta})}{\partial \delta_\ell} &= \frac{\partial}{\partial \delta_\ell} \int \mathscr{s}_{n|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i) \, \mathrm{d}F_{\alpha,\gamma}(\alpha_i, \gamma_i \mid \boldsymbol{\theta}) \\
&= \int \frac{\partial \mathscr{s}_{n|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i)}{\partial \delta_\ell} \, \mathrm{d}F_{\alpha,\gamma}(\alpha_i, \gamma_i \mid \boldsymbol{\theta}) \, , \qquad \text{(D.9)}
\end{aligned}
$$

where the second equality follows since $\mathscr{s}_{n|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i)$ is continuously differentiable in δ_ℓ and the support of $F_{\alpha,\gamma}$ does not depend on δ_ℓ, allowing us to pass the derivative operator through the integral sign.[26]

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

$$
E_\ell(S) = \frac{\exp(\delta_\ell)}{\sum_{k \in \{j\} \cup S} \exp(\delta_k)}, \qquad \text{(D.10)}
$$

for $\ell \in \mathcal{J}$ and $S \subseteq \mathcal{B}_j$. Note that the expression on the right-hand side of Equation (10) 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

$$
\frac{\partial E_j(S)}{\partial \delta_\ell} = \begin{cases} E_j(S)\big[1 - E_j(S)\big] & \text{if } \ell = j \\ \\ -\mathbb{1}\{\ell \in S\} \, E_j(S) E_\ell(S) & \text{otherwise} \end{cases} . \qquad \text{(D.11)}
$$

Thus,

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

(D.12)

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

$$
\begin{aligned}
\frac{\partial \mathscr{s}_{j|n}(\boldsymbol{\delta})}{\partial \delta_\ell} &= \sum_{k=J-n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} \frac{\partial E_j(A)}{\partial \delta_\ell} + \sum_{s=1}^{|\mathcal{B}_j \setminus A|} (-1)^s \sum_{B \in \mathcal{R}_s(\mathcal{B}_j \setminus A)} \frac{\partial E_j(A \cup B)}{\partial \delta_\ell} \\
&= - \sum_{k=J-n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} \Big[\mathbb{1}\{\ell \in A\} \, E_j(A) E_\ell(A)
\end{aligned}
$$

[26] The function $\mathscr{s}_{n|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i)$ 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 $\boldsymbol{\delta}$.

$$+\sum_{s=1}^{|\mathcal{B}_j\setminus A|}(-1)^s\sum_{B\in\mathcal{R}_s(\mathcal{B}_j\setminus A)}\mathbb{1}\left\{\ell\in A\cup B\right\}E_j(A\cup B)E_\ell(A\cup B)\Bigg]. \tag{D.13}$$

To find the partial derivative of $\mathscr{s}_{n|\alpha,\gamma}(\boldsymbol{\delta},\alpha_i,\gamma_i)$ with respect to δ_ℓ for $n\in\{1,\ldots,J-1\}$, rewrite Equation (11) as

$$\mathscr{s}_{n|\alpha,\gamma}(\boldsymbol{\delta},\alpha_i,\gamma_i)=\sum_{s=0}^{J}\sum_{A\in\mathcal{R}_s(\mathcal{J})}\sum_{t\in\tau_n^s}\sum_{B\in\mathcal{R}_t(\mathcal{J}\setminus A)}\left[\left(\prod_{k\in A}a_k\right)\left(\prod_{p\in B}b_p\right)\left(\prod_{q\in\mathcal{J}\setminus(A\cup B)}c_q\right)\right], \tag{D.14}$$

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 $k\in\{1,\ldots,J\}$. Note that the expression on the right-hand side of (D.14) 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

$$\begin{aligned}\frac{\partial \mathscr{s}_{n|\alpha,\gamma}(\boldsymbol{\delta},\alpha_i,\gamma_i)}{\partial\delta_\ell}=&\sum_{s=0}^{J}\sum_{A\in\mathcal{R}_s(\mathcal{J})}\sum_{t\in\tau_n^s}\sum_{B\in\mathcal{R}_t(\mathcal{J}\setminus A)}\Bigg[\mathbb{1}\left\{\ell\in A\right\}\frac{\partial a_\ell}{\partial\delta_\ell}\left(\prod_{k\in A\setminus\{\ell\}}a_k\right)\left(\prod_{p\in B}b_p\right)\left(\prod_{q\in\mathcal{J}\setminus(A\cup B)}c_q\right)\\&+\mathbb{1}\left\{\ell\in B\right\}\frac{\partial b_\ell}{\partial\delta_\ell}\left(\prod_{k\in A}a_k\right)\left(\prod_{p\in B\setminus\{\ell\}}b_p\right)\left(\prod_{q\in\mathcal{J}\setminus(A\cup B)}c_q\right)\\&+\mathbb{1}\left\{\ell\notin A\cup B\right\}\frac{\partial c_\ell}{\partial\delta_\ell}\left(\prod_{k\in A}a_k\right)\left(\prod_{p\in B}b_p\right)\left(\prod_{q\in\mathcal{J}\setminus(A\cup B\cup\{\ell\})}c_q\right)\Bigg]\\=&\sum_{s=0}^{J}\sum_{A\in\mathcal{R}_s(\mathcal{J})}\sum_{t\in\tau_n^s}\sum_{B\in\mathcal{R}_t(\mathcal{J}\setminus A)}\Bigg[\left(\prod_{k\in A}a_k\right)\left(\prod_{p\in B}b_p\right)\left(\prod_{q\in\mathcal{J}\setminus(A\cup B)}c_q\right)\\&\left(\mathbb{1}\left\{\ell\in A\right\}\frac{1}{a_\ell}\frac{\partial a_\ell}{\partial\delta_\ell}+\mathbb{1}\left\{\ell\in B\right\}\frac{1}{b_\ell}\frac{\partial b_\ell}{\partial\delta_\ell}+\mathbb{1}\left\{\ell\notin A\cup B\right\}\frac{1}{c_\ell}\frac{\partial c_\ell}{\partial\delta_\ell}\right)\Bigg],\end{aligned} \tag{D.15}$$

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

$$U_i(A) = u_{ir_i(A,1)} - c_i(|A|). \tag{D.16}$$

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

$$U_i(\emptyset) = -c_i(0) = 0 = u_{i0}. \tag{D.17}$$

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 $|A| \in \{1, \ldots, J\}$. 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

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

which coincides with Equation (13) evaluated at $(\alpha_i, \gamma_i) = (1, \gamma)$, establishing part (ii) of Lemma 1.

To derive the inside-MNL choice probabilities, note that for any $j \in \mathcal{J}$,

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

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

which establishes part (iii) of Lemma 1.

Part (iv) of Lemma 1 follows directly from the properties of the EV_1 cdf $F_\varepsilon(\cdot)$ and the exponential function. In particular, $\exp(x) > 0$ and $F_\varepsilon(x) > 0$ for all $x \in \mathbb{R}$, and $F_\varepsilon(\cdot)$ is strictly increasing. Thus, since $\alpha \in (0,1)$, we have $\gamma < \gamma/\alpha$ and

$$\begin{aligned}
\widetilde{\mathscr{s}}_0(\boldsymbol{\delta}, \gamma) &= F_\varepsilon\big(\gamma\big)^{\sum_{\ell \in \mathcal{J}} \exp\left(\delta_\ell\right)} \\
&< F_\varepsilon\Big(\frac{\gamma}{\alpha}\Big)^{\sum_{\ell \in \mathcal{J}} \exp\left(\delta_\ell\right)} \\
&= \mathscr{s}_0(\boldsymbol{\delta}, \alpha, \gamma).
\end{aligned}$$

Finally, part (v) follows from the same properties and the fact that the natural logarithm is a strictly increasing function:

$$\begin{aligned}
\frac{\partial \widetilde{\mathscr{s}}_0(\boldsymbol{\delta}, \gamma)}{\partial \delta_\ell} &= \exp\left(\delta_\ell\right) F_\varepsilon\big(\gamma\big)^{\sum_{k \in \mathcal{J}} \exp\left(\delta_k\right)} \ln\Big(F_\varepsilon\big(\gamma\big)\Big) \\
&< \exp\left(\delta_\ell\right) F_\varepsilon\left(\frac{\gamma}{\alpha}\right)^{\sum_{k \in \mathcal{J}} \exp\left(\delta_k\right)} \ln\left(F_\varepsilon\left(\frac{\gamma}{\alpha}\right)\right) \\
&= \frac{\partial \mathscr{s}_0(\boldsymbol{\delta}, \alpha, \gamma)}{\partial \delta_\ell}.
\end{aligned}$$

□

D.4 Proof of Proposition 2

Proof. We start by noting that the Jacobian of the unrestricted conditional inclusion probability mapping $\mathscr{s}_{\cdot|n}\big(\boldsymbol{\delta}\big)$ has rank at most $J-1$ for all $\boldsymbol{\delta} \in \mathbb{R}^J$. To see this, fix $n \in \{1, \ldots, J-1\}$ and $\boldsymbol{\delta} \in \mathbb{R}^J$, and differentiate the counting identity

$$\sum_{j \in \mathcal{J}} \mathscr{s}_{j|n}\big(\boldsymbol{\delta}\big) = n$$

to obtain

$$\sum_{j \in \mathcal{J}} \frac{\partial \mathscr{s}_{j|n}\big(\boldsymbol{\delta}\big)}{\partial \delta_\ell} = 0, \qquad \forall \ell \in \mathcal{J}, \tag{D.18}$$

i.e., each column of the $J \times J$ unrestricted Jacobian sums to zero. Similarly, the location invariance of the conditional inclusion probabilities,

$$\mathscr{s}_{j|n}\big(\boldsymbol{\delta} + c\,\boldsymbol{\iota}\big) = \mathscr{s}_{j|n}\big(\boldsymbol{\delta}\big), \qquad \forall c \in \mathbb{R},$$

implies

$$\frac{\mathrm{d}\scriptstyle{\mathcal{s}}_{j|n}\big(\boldsymbol{\delta}+c\,\boldsymbol{\iota}\big)}{\mathrm{d}c}=0, \qquad \forall c\in\mathbb{R}.$$

Evaluating at $c=0$, we obtain

$$\begin{aligned}
0 &= \left.\frac{\mathrm{d}\mathcal{s}_{j|n}\big(\boldsymbol{\delta}+c\,\boldsymbol{\iota}\big)}{\mathrm{d}c}\right|_{c=0} \\
&= \left.\frac{\partial \mathcal{s}_{j|n}\big(\boldsymbol{\delta}+c\,\boldsymbol{\iota}\big)}{\partial\big(\boldsymbol{\delta}+c\,\boldsymbol{\iota}\big)'}\frac{\mathrm{d}\big(\boldsymbol{\delta}+c\,\boldsymbol{\iota}\big)}{\mathrm{d}c}\right|_{c=0} \\
&= \frac{\partial \mathcal{s}_{j|n}\big(\boldsymbol{\delta}\big)}{\partial\boldsymbol{\delta}'}\boldsymbol{\iota} \\
&= \sum_{\ell\in\mathcal{J}}\frac{\partial \mathcal{s}_{j|n}\big(\boldsymbol{\delta}\big)}{\partial\delta_\ell},
\end{aligned} \tag{D.19}$$

where the second equality follows from the chain rule, the third equality follows from evaluation at $c=0$, and the last equality expands the inner product. That is, each row of the unrestricted Jacobian sums to zero. Thus,

$$\operatorname{rank}\left(\frac{\partial \boldsymbol{\mathcal{s}}_{\cdot|n}\big(\boldsymbol{\delta}\big)}{\partial\boldsymbol{\delta}}\right)\le J-1. \tag{D.20}$$

Next, we note that the location normalization $\delta_b=0$ yields a full-rank, nonsingular Jacobian of the restricted mapping $\boldsymbol{\mathcal{s}}^{\boldsymbol{b}}_{\cdot|\boldsymbol{n}}\big(\boldsymbol{\delta}^b\big)$ for every $\boldsymbol{\delta}^b\in\mathbb{R}^{J-1}$ by eliminating the degree of freedom from location invariance and the counting identity. From Equations (D.12) and (D.13) in Appendix D.2, the unrestricted Jacobian follows the sign pattern

$$\frac{\partial \mathcal{s}_{j|n}\big(\boldsymbol{\delta}\big)}{\partial\delta_j}>0, \qquad \frac{\partial \mathcal{s}_{j|n}\big(\boldsymbol{\delta}\big)}{\partial\delta_\ell}<0, \qquad \ell\neq j, \tag{D.21}$$

reflecting that an increase in δ_j raises job j's conditional inclusion probability while reducing that of competing jobs. From (D.19) we obtain

$$\begin{aligned}
0 &< -\frac{\partial \mathcal{s}_{j|n}\big(\boldsymbol{\delta}\big)}{\partial\delta_b} \\
&= \frac{\partial \mathcal{s}_{j|n}\big(\boldsymbol{\delta}\big)}{\partial\delta_j}+\sum_{\ell\in\mathcal{J}\setminus\{j,b\}}\frac{\partial \mathcal{s}_{j|n}\big(\boldsymbol{\delta}\big)}{\partial\delta_\ell} \\
&= \left|\frac{\partial \mathcal{s}^b_{j|n}\big(\boldsymbol{\delta}^b\big)}{\partial\delta_j}\right|-\sum_{\ell\in\mathcal{J}\setminus\{j,b\}}\left|\frac{\partial \mathcal{s}^b_{j|n}\big(\boldsymbol{\delta}^b\big)}{\partial\delta_\ell}\right|
\end{aligned} \tag{D.22}$$

for $j\neq b$, where the inequality follows from (D.21) and the last equality follows from the definition of the restricted mapping and the sign pattern in (D.21). Thus, the restricted Jacobian is a P–matrix for all $\boldsymbol{\delta}^b \in \mathbb{R}^{J-1}$ since it has positive diagonal entries, negative off-diagonal entries, and a strictly dominant diagonal, and is therefore nonsingular (see Gale and Nikaido, 1965). Moreover, by the Gale–Nikaido global univalence theorem, $\mathcal{s}_{\cdot|n}(\cdot)$ is globally injective on $\mathbb{R}^{J-1}$.[27] Finally, any injective function is a bijection from its domain onto its image. $\square$

D.5 Proof of Proposition 3

Proof. Start by noting that, from Equation (21), the probability mass at $n = 0$ is

$$\mathbb{E}\left[s_0\right] = \mathbb{P}\left(n_i = 0\right) = F_\varepsilon\left(\frac{\gamma_0}{\alpha_0}\right)^{\sum_{\ell\in\mathcal{J}} \exp\left(\delta_{\ell 0}\right)},$$

so we can solve for

$$\psi_0^1 = \frac{\gamma_0}{\alpha_0} = F_\varepsilon^{-1}\left(\mathbb{E}\left[s_0\right]^{\frac{1}{\sum_{\ell\in\mathcal{J}} \exp\left(\delta_{\ell 0}\right)}}\right) \tag{D.23}$$

by the properties of $\exp(\cdot)$ and $F_\varepsilon(\cdot)$.

Next, evaluate the optimal stopping rule (5) in part (i) of Proposition 1 at homogeneous (α_0, γ_0) and note that the (decreasing) monotonicity of order statistics together with (increasing) monotonicity of the thresholds $\{\psi^n\}_{n=1}^J$ imply the equivalence of the events $\{n_i \geq n\}$ and $\left\{u_{ir_i(\mathcal{J},n)} \geq \psi^n\right\}$. Thus, given the mapping from rank-ordered ex-post utilities to the order statistics of $\{u_{ij}\}_{j\in\mathcal{J}}$ in Equation (C.2) (see Appendix C.1), the probability of this event is given by the survivor function

$$\mathbb{E}\left[\sum_{k=n}^J s_k\right] = \mathbb{P}\left(n_i \geq n\right) = 1 - F_{u_{(J-n+1)}}\left(\psi_0^n\right),$$

where the expression on the right-hand side is the (marginal) survivor function of the $(J-n+1)$–th order statistic of $\{u_{ij}\}_{j\in\mathcal{J}}$ evaluated at the true ψ_0^n. Since the cdf of the $(J-n+1)$–th order statistic,

$$F_{u_{(J-n+1)}}(x) = \sum_{k=n}^J \sum_{A\in\mathcal{R}_k(\mathcal{J})} \prod_{\ell\in A} F_\varepsilon\left(x - \delta_{\ell 0}\right) \prod_{m\in\mathcal{J}\setminus A} \left[1 - F_\varepsilon\left(x - \delta_{m0}\right)\right]$$

(see David and Nagaraja, 2003, p. 96), is continuous and strictly increasing, the corresponding survivor function $1 - F_{u_{i(J-n+1)}}(x)$ is invertible. Hence, we conclude that, for any $n \in \{2, \ldots, J\}$ with $\mathbb{P}\left(n_i \geq n\right) \in (0,1)$ —which exists by Assumption 5—, ψ_0^n is identified as

$$\psi_0^n = F_{u_{(J-n+1)}}^{-1}\left(1 - \sum_{k=n}^J \mathbb{E}\left[s_k\right]\right). \tag{D.24}$$

[27]Each conditional inclusion probability $\mathcal{s}_{j|n}(\boldsymbol{\delta})$ in Equation (10) is a smooth function of $\boldsymbol{\delta}$ because it is a finite sum of logit-type terms of the form $\exp(\delta_j)/\sum_{\ell\in S}\exp(\delta_\ell)$ for different subsets $S \subseteq \mathcal{J}$ and their inclusion–exclusion combinations. Hence $\mathcal{s}^b_{\cdot|n}(\cdot)$ is continuously differentiable.

Finally, with ψ_0^1 and ψ_0^n pinned down for some $n \in \{2, \dots, J\}$ with nondegenerate survivor probability, we can use (20) and (D.23) to solve for $\boldsymbol{\theta}_0$:

$$\alpha_0 = 1 - \left(\frac{\psi_0^1}{\psi_0^n}\right)^{\frac{1}{n-1}}, \qquad \gamma_0 = \psi_0^1 \left[1 - \left(\frac{\psi_0^1}{\psi_0^n}\right)^{\frac{1}{n-1}}\right]. \tag{D.25}$$

Note that the mapping $(\alpha_0, \gamma_0) \mapsto \left\{ \mathbb{E}\left[s_n\right] \right\}_{n=0}^{J}$ is injective. To see this, suppose not: there exists $(\alpha^*, \gamma^*) \in (0,1) \times \mathbb{R}_{++}$ such that $(\alpha^*, \gamma^*) \neq (\alpha_0, \gamma_0)$ but $\left\{\mathscr{s}_n(\boldsymbol{\delta}_0, \alpha^*, \gamma^*)\right\}_{n=0}^{J} = \left\{\mathscr{s}_n(\boldsymbol{\delta}_0, \alpha_0, \gamma_0)\right\}_{n=0}^{J}$, so both $\boldsymbol{\theta}$ parameters map to the same distribution of optimal portfolio sizes $\left\{ \mathbb{E}\left[s_n\right] \right\}_{n=0}^{J}$. Then, $\psi^{1*} = \psi_0^1$ by (D.23) and $\psi^{n*} = \psi_0^n$ by (D.24), so $(\alpha^*, \gamma^*) = (\alpha_0, \gamma_0)$ by (D.25), a logical contradiction. Therefore, $\left\{\mathscr{s}_n(\boldsymbol{\delta}_0, \alpha^*, \gamma^*)\right\}_{n=0}^{J} = \left\{\mathscr{s}_n(\boldsymbol{\delta}_0, \alpha_0, \gamma_0)\right\}_{n=0}^{J} \implies (\alpha^*, \gamma^*) = (\alpha_0, \gamma_0)$ and the mapping is injective.

□

D.6 Proof of Lemma 2

Proof. Fix three distinct portfolio sizes $1 \leq n_1 < n_2 < n_3 \leq J$ such that $\mathcal{S}_\ell(\boldsymbol{\theta}_0) \in (0,1)$ for $\ell \in \{n_1, n_2, n_3\}$. Consider the mapping $\boldsymbol{g} : (0,1) \times \mathbb{R}_{++} \to \mathbb{R}^2$ defined by

$$\begin{aligned}
\boldsymbol{g}(\boldsymbol{\theta}) &= \left(\ln\left(\frac{\mathcal{S}^{>0}_{n_1}(\boldsymbol{\theta})}{\mathcal{S}^{>0}_{n_2}(\boldsymbol{\theta})}\right) \quad \ln\left(\frac{\mathcal{S}^{>0}_{n_3}(\boldsymbol{\theta})}{\mathcal{S}^{>0}_{n_2}(\boldsymbol{\theta})}\right)\right)' \\
&= \left(\ln\left(\mathcal{S}_{n_1}(\boldsymbol{\theta})\right) - \ln\left(\mathcal{S}_{n_2}(\boldsymbol{\theta})\right) \quad \ln\left(\mathcal{S}_{n_3}(\boldsymbol{\theta})\right) - \ln\left(\mathcal{S}_{n_2}(\boldsymbol{\theta})\right)\right)',
\end{aligned}$$

where the second equality follows since $\mathcal{S}^{>0}_n(\boldsymbol{\theta}) = \mathcal{S}_n(\boldsymbol{\theta})/(1 - \mathscr{s}_0(\boldsymbol{\delta}_0, \boldsymbol{\theta}))$ for all $n \in \{1, \dots, J\}$ so the (nonzero) denominators cancel out. The Jacobian of $\boldsymbol{g}$ at the true parameters $\boldsymbol{\theta}_0 = (\alpha_0, \gamma_0)$ is therefore

$$\frac{\partial \boldsymbol{g}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} = \left(\frac{\partial \ln\left(\mathcal{S}_{n_1}(\boldsymbol{\theta}_0)\right)}{\partial \boldsymbol{\theta}} - \frac{\partial \ln\left(\mathcal{S}_{n_2}(\boldsymbol{\theta}_0)\right)}{\partial \boldsymbol{\theta}} \quad \frac{\partial \ln\left(\mathcal{S}_{n_3}(\boldsymbol{\theta}_0)\right)}{\partial \boldsymbol{\theta}} - \frac{\partial \ln\left(\mathcal{S}_{n_2}(\boldsymbol{\theta}_0)\right)}{\partial \boldsymbol{\theta}} \right)'. \tag{D.26}$$

We will prove that this Jacobian is nonsingular, so the mapping $\boldsymbol{g}$ is invertible in a neighborhood of $\boldsymbol{\theta}_0$.

From the proof of Proposition 3 in Appendix D.5, we know that

$$\mathcal{S}_n(\boldsymbol{\theta}_0) = 1 - F_{u_{(J-n+1)}}\left(\psi_0^n\right),$$

where $F_{u_{(J-n+1)}}(\cdot)$ is the strictly increasing, continuously differentiable cdf of the $(J-n+1)$–th order statistic of ex-post utilities $\{u_{ij}\}_{j \in \mathcal{J}}$ with strictly positive density $f_{u_{(J-n+1)}}(\cdot)$, and ψ_0^n is the n–th stopping-rule threshold from Proposition 1 specialized to the homogeneous (α_i, γ_i) case and evaluated at $\boldsymbol{\theta}_0$. Thus, for any $n \in \{1, \dots, J\}$ such that $\mathcal{S}_n(\boldsymbol{\theta}_0) \in (0,1)$,

$$\frac{\partial \ln\left(\mathcal{S}_n(\boldsymbol{\theta}_0)\right)}{\partial \boldsymbol{\theta}} = -\frac{f_{u_{(J-n+1)}}\left(\psi_0^n\right)}{1 - F_{u_{(J-n+1)}}\left(\psi_0^n\right)} \frac{\partial \psi_0^n}{\partial \boldsymbol{\theta}} \tag{D.27}$$

is a negative scalar multiple of the gradient of the n–th stopping-rule threshold with respect to the parameters at their true value $\boldsymbol{\theta}_0$. From Equations (28) and (29), we can write the gradient of ψ_0^n as

$$\frac{\partial \psi_0^n}{\partial \boldsymbol{\theta}} = \psi_0^n \begin{pmatrix} \frac{\alpha_0 n - 1}{\alpha_0(1-\alpha_0)} & \frac{1}{\gamma_0} \end{pmatrix}',$$

so the gradient of the log-survivor in (D.27) takes the form

$$\frac{\partial \ln\left(\mathcal{S}_n(\boldsymbol{\theta}_0)\right)}{\partial \boldsymbol{\theta}} = -c_n\, \boldsymbol{v}_n, \qquad c_n = \frac{f_{u_{(J-n+1)}}\left(\psi_0^n\right)}{1 - F_{u_{(J-n+1)}}\left(\psi_0^n\right)} \psi_0^n, \qquad \boldsymbol{v}_n = \begin{pmatrix} \frac{\alpha_0 n - 1}{\alpha_0(1-\alpha_0)} & \frac{1}{\gamma_0} \end{pmatrix}'. \tag{D.28}$$

The scalar c_n is strictly positive since the hazard rate of the $(J-n+1)$–th order statistic of u_{ij} and the n–th stopping-rule threshold are both strictly positive. The vector $\boldsymbol{v}_n \in \mathbb{R}^2$ has slope —given by the ratio of its components—

$$\frac{v_n^2}{v_n^1} = \frac{\alpha_0(1-\alpha_0)}{\gamma_0(\alpha_0 n - 1)},$$

which is a nonzero, injective function of n on $\mathbb{N} \setminus \{1/\alpha_0\}$. When $n = 1/\alpha_0$ —a distinction that matters only occasionally when $1/\alpha_0$ is an integer, a measure-zero subset of $(0,1)$, the domain of α_0— $v_n^1 = 0$ and $\boldsymbol{v}_n$ remains a well-defined vertical vector on the (α, γ)–plane. This means that, for any two distinct $n_k \neq n_\ell \in \{1, \ldots, J\}$, the vectors $\boldsymbol{v}_{n_k}$ and $\boldsymbol{v}_{n_\ell}$ are linearly independent since (i) $v_{n_k}^1 = v_{n_\ell}^1 \implies n_k = n_\ell$, and (ii) the vectors point in different directions whenever $v_{n_k}^1 \neq v_{n_\ell}^1$.

Finally, since $1 \leq n_1 < n_2 < n_3 \leq J$, the gradients

$$\frac{\partial \ln\left(\mathcal{S}_\ell(\boldsymbol{\theta}_0)\right)}{\partial \boldsymbol{\theta}} = -c_\ell \boldsymbol{v}_\ell, \qquad \ell \in \{n_1, n_2, n_3\}$$

are pairwise linearly independent, so the Jacobian (D.26) has linearly independent rows and is thus nonsingular. In fact, it can be directly verified that the determinant of the Jacobian,

$$\det\left(\frac{\partial \boldsymbol{g}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right) = \frac{1}{\gamma_0}\Big[- c_{n_1} c_{n_3}\big(v_{n_3}^1 - v_{n_1}^1\big) + c_{n_1} c_{n_2}\big(v_{n_2}^1 - v_{n_1}^1\big) + c_{n_2} c_{n_3}\big(v_{n_3}^1 - v_{n_2}^1\big)\Big], \tag{D.29}$$

is generically nonzero —except on a measure-zero subset of $(0,1) \times \mathbb{R}_{++}$ where $\boldsymbol{\theta}_0$ maps to constants $(c_{n_1}, c_{n_2}, c_{n_3})$ that align to make (D.29) vanish— since $n_1 < n_2 < n_3 \implies v_{n_1}^1 < v_{n_2}^1 < v_{n_3}^1$. By the inverse function theorem, $\boldsymbol{g}(\boldsymbol{\theta})$ is invertible in a neighborhood of $\boldsymbol{\theta}_0$, so $\boldsymbol{\theta}_0$ is locally identified by inverting the system of equations

$$\ln\left(\frac{\mathcal{S}_\ell^{>0}(\boldsymbol{\theta}_0)}{\mathcal{S}_{n_2}^{>0}(\boldsymbol{\theta}_0)}\right) = \ln\left(\sum_{k=\ell}^{J} \mathbb{E}\left[s_k\right]\right) - \ln\left(\sum_{k=n_2}^{J} \mathbb{E}\left[s_k\right]\right), \qquad \ell \in \{n_1, n_3\}.$$

$\square$

D.7 Proof of Proposition 4 (sketch)

Remark. Throughout the proof sketch below, read

$$(\alpha_i, \gamma_i) = \Big(Q_\alpha\big(u \mid \boldsymbol{\theta}_\alpha\big), Q_\gamma\big(v \mid \boldsymbol{\theta}_\gamma\big)\Big), \qquad (u, v) = \Big(F_\alpha\big(\alpha_i \mid \boldsymbol{\theta}_\alpha\big), F_\gamma\big(\gamma_i \mid \boldsymbol{\theta}_\gamma\big)\Big)$$

interchangeably, where the pair $(\alpha_i, \gamma_i) \in (0,1) \times \mathbb{R}_{++}$ represent job seeker i's random effects and $(u, v) \in [0,1]^2$ their ranks in the corresponding marginal distributions.

Proof sketch. Fix $L+1$ distinct portfolio sizes $1 \leq n_1 < \cdots < n_L \leq J$ such that $\mathcal{S}_{n_\ell}(\boldsymbol{\theta}_0) \in (0,1)$ for $\ell \in \{0, \ldots, L\}$. Consider the mapping $\boldsymbol{g} : \boldsymbol{\Theta} \to \mathbb{R}^L$ defined by

$$\begin{aligned}
\boldsymbol{g}(\boldsymbol{\theta}) &= \left(\ln\left(\frac{\mathcal{S}^{>0}_{n_1}(\boldsymbol{\theta})}{\mathcal{S}^{>0}_{n_0}(\boldsymbol{\theta})}\right) \quad \cdots \quad \ln\left(\frac{\mathcal{S}^{>0}_{n_L}(\boldsymbol{\theta})}{\mathcal{S}^{>0}_{n_0}(\boldsymbol{\theta})}\right)\right)' \\
&= \left(\ln\Big(\mathcal{S}_{n_1}(\boldsymbol{\theta})\Big) - \ln\Big(\mathcal{S}_{n_0}(\boldsymbol{\theta})\Big) \quad \cdots \quad \ln\Big(\mathcal{S}_{n_L}(\boldsymbol{\theta})\Big) - \ln\Big(\mathcal{S}_{n_0}(\boldsymbol{\theta})\Big)\right)',
\end{aligned}$$

where the second equality follows since $\mathcal{S}^{>0}_n(\boldsymbol{\theta}) = \mathcal{S}_n(\boldsymbol{\theta})/(1 - \mathcal{s}_0(\boldsymbol{\delta}_0, \boldsymbol{\theta}))$ for all $n \in \{1, \ldots, J\}$ so the (nonzero) denominators cancel out.

Let $\psi^n(\alpha_i, \gamma_i)$ represent the n–th stopping-rule threshold defined in Equation (12) and write the conditional survivor function at n as the sum of the conditional pmf over $k \geq n$,

$$\begin{aligned}
\mathcal{S}_{n|\alpha,\gamma}(\alpha_i, \gamma_i) &= \sum_{k=n}^{J} \mathcal{s}_{k|\alpha,\gamma}(\boldsymbol{\delta}_0, \alpha_i, \gamma_i) \\
&= 1 - F_{u(J-n+1)}\big(\psi^n(\alpha_i, \gamma_i)\big),
\end{aligned}$$

where $F_{u_{(J-n+1)}}(\,\cdot\,)$ is the continuously differentiable, strictly increasing cdf of the $(J-n+1)$–th order statistic of ex-post utilities $\{u_{ij}\}_{j \in \mathcal{J}}$. The second equality follows from the same arguments used in the proof of Proposition 3 in Appendix D.5 since (α_i, γ_i) are held fixed here. For the same reason, analogous arguments to those in Appendix D.6 establish that

$$\frac{\partial \mathcal{S}_{n|\alpha,\gamma}(\alpha_i, \gamma_i)}{\partial(\alpha_i, \gamma_i)} = -f_{u(J-n+1)}\big(\psi^n(\alpha_i, \gamma_i)\big)\psi^n(\alpha_i, \gamma_i)\left(\frac{\alpha_i n - 1}{\alpha_i(1-\alpha_i)} \quad \frac{1}{\gamma_i}\right)'. \tag{D.30}$$

Use the copula representation (37) in density form to write the unconditional survivor function as an integral over the unit square,

$$\mathcal{S}_n(\boldsymbol{\theta}) = \int_{[0,1]^2} \mathcal{S}_{n|\alpha,\gamma}\Big(Q_\alpha\big(u \mid \boldsymbol{\theta}_\alpha\big), Q_\gamma\big(v \mid \boldsymbol{\theta}_\gamma\big)\Big)\, c\big(u, v \mid \boldsymbol{\theta}_C\big)\, \mathrm{d}u\, \mathrm{d}v$$

Under the regularity conditions in Assumption 6, differentiation under the integral sign is valid by the dominated convergence theorem. For a component θ^r_α of $\boldsymbol{\theta}_\alpha$ with $r \in \{1, \ldots, \dim(\boldsymbol{\theta}_\alpha)\}$,

$$\frac{\partial \mathcal{S}_n(\boldsymbol{\theta})}{\partial \theta^r_\alpha} = \int_0^1 \kappa^\alpha_n(u \mid \boldsymbol{\theta})\, \frac{\partial Q_\alpha\big(u \mid \boldsymbol{\theta}_\alpha\big)}{\partial \theta^r_\alpha}\, \mathrm{d}u,$$

where

$$\kappa_n^\alpha(u \mid \boldsymbol{\theta}) = \int_0^1 \frac{\partial \mathcal{S}_{n|\alpha,\gamma}\Big(Q_\alpha\big(u \mid \boldsymbol{\theta}_\alpha\big), Q_\gamma\big(v \mid \boldsymbol{\theta}_\gamma\big)\Big)}{\partial \alpha_i} c\big(u, v \mid \boldsymbol{\theta}_C\big)\,\mathrm{d}v.$$

Similarly, for a component θ_γ^r of $\boldsymbol{\theta}_\gamma$ with $r \in \{1, \ldots, \dim(\boldsymbol{\theta}_\gamma)\}$,

$$\frac{\partial \mathcal{S}_n(\boldsymbol{\theta})}{\partial \theta_\gamma^r} = \int_0^1 \kappa_n^\gamma(v \mid \boldsymbol{\theta})\, \frac{\partial Q_\gamma\big(v \mid \boldsymbol{\theta}_\gamma\big)}{\partial \theta_\gamma^r}\,\mathrm{d}v,$$

where

$$\kappa_n^\gamma(v \mid \boldsymbol{\theta}) = \int_0^1 \frac{\partial \mathcal{S}_{n|\alpha,\gamma}\Big(Q_\alpha\big(u \mid \boldsymbol{\theta}_\alpha\big), Q_\gamma\big(v \mid \boldsymbol{\theta}_\gamma\big)\Big)}{\partial \gamma_i} c\big(u, v \mid \boldsymbol{\theta}_C\big)\,\mathrm{d}u.$$

In contrast, for θ_C^r a component of $\boldsymbol{\theta}_C$ with $r \in \{1, \ldots, \dim(\boldsymbol{\theta}_C)\}$,

$$\begin{aligned}\frac{\partial \mathcal{S}_n(\boldsymbol{\theta})}{\partial \theta_C^r} &= \int_{[0,1]^2} \mathcal{S}_{n|\alpha,\gamma}\Big(Q_\alpha\big(u \mid \boldsymbol{\theta}_\alpha\big), Q_\gamma\big(v \mid \boldsymbol{\theta}_\gamma\big)\Big)\, \frac{\partial c\big(u, v \mid \boldsymbol{\theta}_C\big)}{\partial \theta_C^r}\,\mathrm{d}u\,\mathrm{d}v \\ &= \int_{[0,1]^2} \mathcal{S}_{n|\alpha,\gamma}\Big(Q_\alpha\big(u \mid \boldsymbol{\theta}_\alpha\big), Q_\gamma\big(v \mid \boldsymbol{\theta}_\gamma\big)\Big)\, \frac{\partial \ln\Big(c\big(u, v \mid \boldsymbol{\theta}_C\big)\Big)}{\partial \theta_C^r}\,\mathrm{d}C\big(u, v \mid \boldsymbol{\theta}_C\big),\end{aligned}$$

where the second equality follows form the properties of the logarithm.

Now, partitioning the $L \times L$ Jacobian of $\boldsymbol{g}$ at $\boldsymbol{\theta}_0$, i.e., differentiating blockwise,

$$\frac{\partial \boldsymbol{g}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} = \left[\begin{bmatrix} \frac{\partial g_1(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'_\alpha} \\ \vdots \\ \frac{\partial g_L(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'_\alpha} \end{bmatrix} \quad \begin{bmatrix} \frac{\partial g_1(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'_\gamma} \\ \vdots \\ \frac{\partial g_L(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'_\gamma} \end{bmatrix} \quad \begin{bmatrix} \frac{\partial g_1(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'_C} \\ \vdots \\ \frac{\partial g_L(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'_C} \end{bmatrix}\right] = \begin{bmatrix} \frac{\partial \boldsymbol{g}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'_\alpha} & \frac{\partial \boldsymbol{g}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'_\gamma} & \frac{\partial \boldsymbol{g}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'_C} \end{bmatrix},$$

where, for each $\ell \in \{1, \ldots, L\}$, $p \in \{1, \ldots, \dim(\boldsymbol{\theta}_\alpha)\}$, $q \in \{1, \ldots, \dim(\boldsymbol{\theta}_\gamma)\}$, and $r \in \{1, \ldots, \dim(\boldsymbol{\theta}_C)\}$,

$$\frac{\partial g_\ell(\boldsymbol{\theta}_0)}{\partial \theta_\alpha^p} = \int_0^1 \left(\frac{\kappa_{n_\ell}^\alpha(u \mid \boldsymbol{\theta}_0)}{\mathcal{S}_{n_\ell}(\boldsymbol{\theta}_0)} - \frac{\kappa_{n_0}^\alpha(u \mid \boldsymbol{\theta}_0)}{\mathcal{S}_{n_0}(\boldsymbol{\theta}_0)}\right) \frac{\partial Q_\alpha\big(u \mid \boldsymbol{\theta}_{\alpha 0}\big)}{\partial \theta_\alpha^p}\,\mathrm{d}u,$$

$$\frac{\partial g_\ell(\boldsymbol{\theta}_0)}{\partial \theta_\gamma^q} = \int_0^1 \left(\frac{\kappa_{n_\ell}^\gamma(v \mid \boldsymbol{\theta}_0)}{\mathcal{S}_{n_\ell}(\boldsymbol{\theta}_0)} - \frac{\kappa_{n_0}^\gamma(v \mid \boldsymbol{\theta}_0)}{\mathcal{S}_{n_0}(\boldsymbol{\theta}_0)}\right) \frac{\partial Q_\gamma\big(v \mid \boldsymbol{\theta}_{\gamma 0}\big)}{\partial \theta_\gamma^q}\,\mathrm{d}v,$$

$$\frac{\partial g_\ell(\boldsymbol{\theta}_0)}{\partial \theta_C^r} = \int_{[0,1]^2} \left(\frac{\mathcal{S}_{n_\ell|\alpha,\gamma}\big(Q_u^\alpha, Q_v^\gamma \mid \boldsymbol{\theta}_0^{\alpha,\gamma}\big)}{\mathcal{S}_{n_\ell}(\boldsymbol{\theta}_0)} - \frac{\mathcal{S}_{n_0|\alpha,\gamma}\big(Q_u^\alpha, Q_v^\gamma \mid \boldsymbol{\theta}_0^{\alpha,\gamma}\big)}{\mathcal{S}_{n_0}(\boldsymbol{\theta}_0)}\right) \frac{\partial \ln\Big(c\big(u, v \mid \boldsymbol{\theta}_{C0}\big)\Big)}{\partial \theta_C^r}\,\mathrm{d}C\big(u, v \mid \boldsymbol{\theta}_{C0}\big)$$

and we write $\mathcal{S}_{n|\alpha,\gamma}\big(Q_u^\alpha, Q_v^\gamma \mid \boldsymbol{\theta}^{\alpha,\gamma}\big) \equiv \mathcal{S}_{n|\alpha,\gamma}\Big(Q_\alpha\big(u \mid \boldsymbol{\theta}_\alpha\big), Q_\gamma\big(v \mid \boldsymbol{\theta}_\gamma\big)\Big)$ for readability.

Just as in our discussion of comparative statics in the homogeneous model in Section 2.4.3, we can conclude from Equation (D.30) that, since the order-statistic pdf $f_{u(J-n+1)}(\cdot)$ and the sopping threshold $\psi^n(\alpha_i, \gamma_i)$ are both strictly positive, the conditional survivor $\mathcal{S}_{n|\alpha,\gamma}(\alpha_i, \gamma_i)$ is strictly decreasing in γ_i with stronger effect for larger n, while it is nonmonotonic in α_i: increasing for $\alpha_i < 1/n$, and decreasing for $\alpha_i > 1/n$. Averaging over ranks $(U, V) \sim C\big(u, v \mid \boldsymbol{\theta}_{C0}\big)$ preserves these patterns in the kernels $\kappa_n^{\gamma}(v \mid \boldsymbol{\theta}_0)$ (uniformly negative, strictly ordered in n) and $\kappa_n^{\alpha}(u \mid \boldsymbol{\theta}_0)$ (flipping side between two portfolio sizes on opposite sides of the kink $1/\alpha_i$). With $L+1$ distinct portfolio sizes, the local injectivity of the marginal quantile mappings $\boldsymbol{\theta}_\alpha \mapsto Q_\alpha(\cdot \mid \boldsymbol{\theta}_\alpha)$ and $\boldsymbol{\theta}_\gamma \mapsto Q_\gamma(\cdot \mid \boldsymbol{\theta}_\gamma)$ implied by Assumption 6, the $\partial \boldsymbol{g}(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}'_\alpha$ and $\partial \boldsymbol{g}(\boldsymbol{\gamma}_0)/\partial \boldsymbol{\theta}'_\alpha$ blocks of the Jacobian are each full rank. Moreover, they are and mutually independent since no linear combination of a uniform-sign pattern (conditional survivor derivative in γ_i) can cancel a kinked pattern that necessarily flips signs (conditional survivor derivative in α_i).

Finally, the $\partial \boldsymbol{g}(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}'_C$ block of the Jacobian is also full rank and linearly independent of the other two blocks. To see this, note that the copula scores $\left\{\partial \ln \big(c(u, v \mid \boldsymbol{\theta}_{C0})\big)/\partial \theta_C^r\right\}_{r=1}^{\dim(\boldsymbol{\theta}_C)}$ are linearly independent in $L^2(C_0)$, where $C_0 \equiv C(\cdot, \cdot \mid \boldsymbol{\theta}_{C0})$, by the nonsingular Fisher information matrix regularity condition in Assumption 6. Since each θ_C^r–column is an inner product in $L^2(C_0)$ between the n_0–differenced conditional log-survivor derivatives and the copula scores, and the former vary with n_ℓ on a set with positive C_0–measure, this Jacobian block has full rank. Moreover, since the copula parameters $\boldsymbol{\theta}_C$ do not change the marginals $F_\alpha(\cdot \mid \boldsymbol{\theta}_\alpha)$ and $F_\gamma(\cdot \mid \boldsymbol{\theta}_\gamma)$, each copula score is orthogonal to any u– or v–only function such as the kernels that enter the $\boldsymbol{\theta}_\alpha$ and $\boldsymbol{\theta}_\gamma$ Jacobian blocks. Therefore, the copula columns of the Jacobian cannot lie in the span of the marginals' columns, establishing the linear independence of this Jacobian block from the other two.

Since the three Jacobian blocks are full rank and mutually linearly independent, we conclude that the full Jacobian is nonsingular with full column rank L. Thus, the moment condition mapping $\boldsymbol{g}(\boldsymbol{\theta}_0)$ is locally injective at $\boldsymbol{\theta}_0$ by the inverse function theorem and the L moment equations

$$\ln\left(\frac{S_{n_\ell}^{>0}(\boldsymbol{\theta}_0)}{S_{n_0}^{>0}(\boldsymbol{\theta}_0)}\right) = \ln\left(\sum_{k=n_\ell}^{J} \mathbb{E}\left[s_k\right]\right) - \ln\left(\sum_{k=n_0}^{J} \mathbb{E}\left[s_k\right]\right), \qquad \ell \in \{1, \ldots, L\}$$

locally identify $\boldsymbol{\theta}_0$. Note that this system of equations deals with zero-truncation but identification equivalently holds in the non-truncated case.

□

D.8 Firm-level elasticities through single-shift directional derivatives

Consider the directional derivative of the supply of applications to firm $f \in \mathcal{F}$, $q^f(\boldsymbol{\delta}, \boldsymbol{\theta})$, with respect to mean utilities $\boldsymbol{\delta}$ in the direction of vector $\boldsymbol{d}^f$ with ℓ–th entry $d_\ell^f = \mathbb{1}\left\{\ell \in \mathcal{J}^f\right\}$. This scalar-valued derivative is equivalent to the sum of the elements of the gradient of $q^f(\boldsymbol{\delta}, \boldsymbol{\theta})$ with respect to the mean ex-post utilities of its vacancies, $\{\delta_\ell\}_{\ell\in\mathcal{J}^f}$:

$$\begin{aligned}
\left.\frac{\mathrm{d}q^f\left(\boldsymbol{\delta}+c\,\boldsymbol{d}^f,\boldsymbol{\theta}\right)}{\mathrm{d}c}\right|_{c=0} &= \left.\left(\frac{\mathrm{d}}{\mathrm{d}c}\sum_{j\in\mathcal{J}^f} q_j\left(\boldsymbol{\delta}+c\,\boldsymbol{d}^f,\boldsymbol{\theta}\right)\right)\right|_{c=0} \\
&= I\sum_{j\in\mathcal{J}^f}\sum_{n=1}^{J}\left(\frac{\partial}{\partial\left(\boldsymbol{\delta}+c\,\boldsymbol{d}^f\right)'}s_{j|n}\left(\boldsymbol{\delta}+c\,\boldsymbol{d}^f\right)\frac{\mathrm{d}}{\mathrm{d}c}\left(\boldsymbol{\delta}+c\,\boldsymbol{d}^f\right)s_n\left(\boldsymbol{\delta}+c\,\boldsymbol{d}^f,\boldsymbol{\theta}\right)\right. \\
&\qquad\left.\left.+\frac{\partial}{\partial\left(\boldsymbol{\delta}+c\,\boldsymbol{d}^f\right)'}s_n\left(\boldsymbol{\delta}+c\,\boldsymbol{d}^f,\boldsymbol{\theta}\right)\frac{\mathrm{d}}{\mathrm{d}c}\left(\boldsymbol{\delta}+c\,\boldsymbol{d}^f\right)s_{j|n}\left(\boldsymbol{\delta}+c\,\boldsymbol{d}^f\right)\right)\right|_{c=0} \\
&= I\sum_{j\in\mathcal{J}^f}\sum_{n=1}^{J}\frac{\partial s_{j|n}\left(\boldsymbol{\delta}\right)}{\partial\boldsymbol{\delta}'}\boldsymbol{d}^f s_n\left(\boldsymbol{\delta},\boldsymbol{\theta}\right)+\frac{\partial s_n\left(\boldsymbol{\delta},\boldsymbol{\theta}\right)}{\partial\boldsymbol{\delta}'}\boldsymbol{d}^f s_{j|n}\left(\boldsymbol{\delta}\right) \\
&= I\sum_{j\in\mathcal{J}^f}\sum_{n=1}^{J}\sum_{\ell\in\mathcal{J}^f}\frac{\partial s_{j|n}\left(\boldsymbol{\delta}\right)}{\partial\delta_\ell}s_n\left(\boldsymbol{\delta},\boldsymbol{\theta}\right)+\frac{\partial s_n\left(\boldsymbol{\delta},\boldsymbol{\theta}\right)}{\partial\delta_\ell}s_{j|n}\left(\boldsymbol{\delta}\right) \\
&= \sum_{\ell\in\mathcal{J}^f}\sum_{j\in\mathcal{J}^f}\frac{\partial q_j\left(\boldsymbol{\delta},\boldsymbol{\theta}\right)}{\partial\delta_\ell} \\
&= \sum_{\ell\in\mathcal{J}^f}\frac{\partial q^f\left(\boldsymbol{\delta},\boldsymbol{\theta}\right)}{\partial\delta_\ell}. \qquad (\text{D.31})
\end{aligned}$$

The first equality follows directly from the definition of the firm-level supply in Equation (18). The second equality follows from the linearity of differentiation, by evaluating (9) at $\left(\boldsymbol{\delta}+c\,\boldsymbol{d}^f,\boldsymbol{\theta}\right)$ and differentiating with respect to c. The third equality follows by direct evaluation at $c=0$. The fourth equality follows from the property that $\boldsymbol{x}'\boldsymbol{d}^f=\sum_{\ell\in\mathcal{J}^f}x_\ell$ for any $\boldsymbol{x}\in\mathbb{R}^J$ by the definition of $\boldsymbol{d}^f$ as a vector-valued an indicator function. The fifth equality follows by rearranging summation indices and noting that the inner summation —including the I term— is the partial derivative of the right-hand side of (9) with respect to δ_ℓ. The last equality follows, again, from the definition of the supply of applications to the firm in (18) and the linearity of the derivative operator.

Using this result, we can write the elasticity of the supply of applications to the firm with respect to a firm-level offered-pay increase in terms of the directional derivative on the left-hand side of (D.31):

$$\begin{aligned}
\eta_{q^f,\boldsymbol{w}^f}\left(\boldsymbol{\delta},\boldsymbol{\theta},\beta_w\right) &= \sum_{\ell\in\mathcal{J}^f}\frac{\partial\ln\left(q^f\left(\boldsymbol{\delta},\boldsymbol{\theta}\right)\right)}{\partial\ln\left(w_\ell\right)} \\
&= \frac{1}{q^f\left(\boldsymbol{\delta},\boldsymbol{\theta}\right)}\sum_{\ell\in\mathcal{J}^f}\frac{\partial q^f\left(\boldsymbol{\delta},\boldsymbol{\theta}\right)}{\partial\delta_\ell}\frac{\partial\delta_\ell}{\partial\ln\left(w_\ell\right)}
\end{aligned}$$

$$= \frac{1}{q^f(\boldsymbol{\delta}, \boldsymbol{\theta})} \left(\frac{\mathrm{d} q^f(\boldsymbol{\delta} + c\, \boldsymbol{d}^f, \boldsymbol{\theta})}{\mathrm{d}c} \right) \Bigg|_{c=0} \beta_w. \tag{D.32}$$

The first equality is the definition of this elasticity. The second equality follows from the chain rule. The last equality follows from Equations (D.31) and (8).

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

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

where $\mathcal{A}_i = \{A_i, \overline{A}_i\}$ is job seeker i's partition of the choice set into chosen and unchosen alternatives and $F_\varepsilon(x) = \exp(-\exp(-x))$ 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 $\{\varepsilon_{ij}\}_{j \in \mathcal{J}} \overset{\text{iid}}{\sim} \mathrm{EV}_1 \implies \mathbb{P}\left(\max_{\ell \in \overline{A}_i} \delta_\ell + \varepsilon_{i\ell} \leq x\right) = F_\varepsilon\left(x - \ln\left(\sum_{\ell \in \overline{A}_i} \exp(\delta_\ell)\right)\right)$, the fourth equality uses the fact that $F_\varepsilon(x - \ln(a)) = F_\varepsilon(x - \ln(b))^{a/b}$ for $a, b > 0$, the fifth equality applies the change of variable $u = F_\varepsilon\left(x - \ln\left(\sum_{\ell \in \overline{A}_i} \exp(\delta_\ell)\right)\right)$, and the last equality makes the change of variable $z =$

$u^{1/\sum_{\ell \in \overline{A}_i} \exp(\delta_\ell)}$. 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 i.i.d. 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).[28] 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.[29]

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

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

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

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 $\ell(\cdot)$ 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.,

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

[28]For notational simplicity, we hereafter suppress the dependence of the likelihood function on the data $\{\mathcal{A}_i\}_{i\in\mathcal{I}}$.
[29]See Wu and Lange (2010) for an introduction to MM algorithms.

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.

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

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

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

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

where

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

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

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

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

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

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: Optimization details

This appendix details our simulated method of moments approach to the estimation of $\boldsymbol{\theta} = (\theta_\alpha^1, \theta_\alpha^2, \theta_\gamma^1, \theta_\gamma^3)$, the parameters of the distribution $F_{\alpha,\gamma}(\cdot,\cdot \mid \boldsymbol{\theta}) = F_\alpha(\cdot \mid \boldsymbol{\theta}_\alpha)$ and $F_\gamma(\cdot \mid \boldsymbol{\theta}_\gamma)$, where we have assumed mutual independence of the job-offer uncertainty and marginal cost of applications random effects. In our Beta-Exponential specification, the parameters are $\boldsymbol{\theta}_\alpha = (\theta_\alpha^1, \theta_\alpha^2$, controlling the mean and dispersion of the Beta distribution of α_i, and $\boldsymbol{\theta}_\gamma = (\theta_\gamma^1, \theta_\gamma^2$, controlling the mean and shape of the Gamma distribution of γ_i.

Bootstrap estimate of optimal weightmatrix: Constant random numbers (CRNs): Explain (i) S draws of $U^s = \widehat{\boldsymbol{\delta}} \otimes \boldsymbol{\iota}_I' - \boldsymbol{\varepsilon}_i^{s\prime}$, $\boldsymbol{V}_\alpha^s \overset{\text{iid}}{\sim} \text{Uniform}(0,1)_{1\times I}$, and $\boldsymbol{V}_\gamma^s \overset{\text{iid}}{\sim} \text{Uniform}(0,1)_{1\times I}$ are kept fixed across the whole optimization routine; (ii) $\boldsymbol{\alpha}^{s,(k)} = \text{invibeta}\big(\theta_\alpha^1\,\theta_\alpha^2, (1-\theta_\alpha^1)\theta_\alpha^2, \boldsymbol{V}_\alpha^s\big)$ and $\boldsymbol{\gamma}^{s,(k)} = (\theta_\gamma^1/\theta_\gamma^2)\,\text{invgammap}\big(\theta_\gamma^2, \theta_\gamma^1/\theta_\gamma^2, \boldsymbol{V}_\gamma^s\big)$ are generated with the same common underlying $1 \times I$ $\text{Uniform}(0,1)$ 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 $\to$ 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; (iii) 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); (iv) 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 details

This appendix provides detail on the Trabajando.com dataset, including variable definitions, sample construction, and cleaning procedures.

G.1 Raw datasets

The data are organized into four linked datasets:

- `Job ads`: vacancy-level information including identifiers, publication and expiry dates, job title and field, employer information, required education and experience, contract type, working hours, wage (posted or hidden), number of openings, and whether the ad was paid. Additional requisites and location fields are recorded as unstructured text.
- `Employers`: firm-level data with anonymized identifiers, names (text strings), industry, region, city, and categorical size bins (1–10, 11–50, 51–150, 151–300, 301–500, 501–1,000, 1,001–5,000, and >5,000 employees). Recruiting firms are flagged following Banfi and Villena-Roldán (2019).
- `Users`: job-seeker records including demographics (sex, date of birth, marital status, nationality, residence), education (highest degree, up to three study programs with institution and status), labor market history (employment status, up to three prior jobs with start/end dates, titles, and wages), and job search activity (registration and update dates, availability to work, expected wage, and disclosure preference).

- `Applications`: applicant–ad links recording identifiers and daily application dates. Application outcomes are not observed.

Variable dictionaries are provided in the Online Appendix

G.2 Sample construction

The estimation sample is defined as follows:

- Job seekers: unemployed, residing in Chile, with declared expected wages between CLP $150,000 and CLP $5,000,000.
- Search spells: begin at the most recent CV update in 2017–2018, include applications up to 365 days prior, segmented if gaps exceed 90 days, and restricted to spells fully contained in calendar year 2018.
- Outliers: applications above the 99–th percentile of seeker-level counts are removed.

The final sample consists of 17,357 job seekers, 8,808 jobs, and 1,167 firms across 55 occupation–region bins.

G.3 Occupation and workplace classification

Occupations are classified using the INE automatic classifier, which maps job titles and ad text to ISCO08.CL two-digit codes. Workplace addresses are standardized using the Google Maps API to recover districts and regions. These two dimensions (occupation and region) define labor markets for the nested logit benchmark. Network analysis confirms that 99% of jobs belong to the largest connected component of the application graph, consistent with treating the full sample as a single market in the main portfolio-choice analysis.

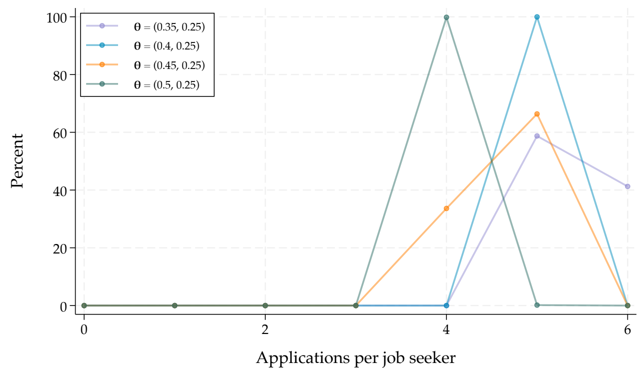
G.4 Cleaning

Number of vacancies: A tiny number of job ads report zero vacancies being offered in the raw data. We treat the number of vacancies as missing for these observations.

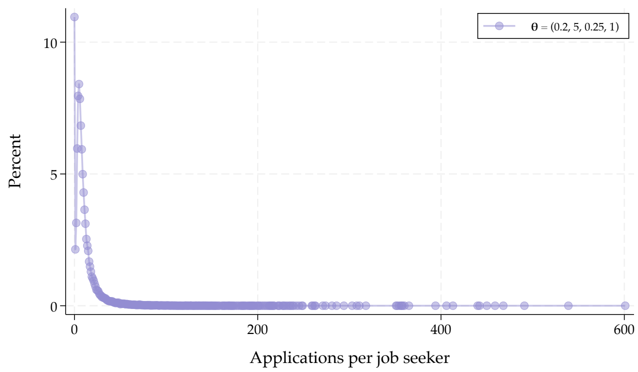
Hours of work: We Combine the two part-time categories into one since the distinction was only nominal.

H Further results

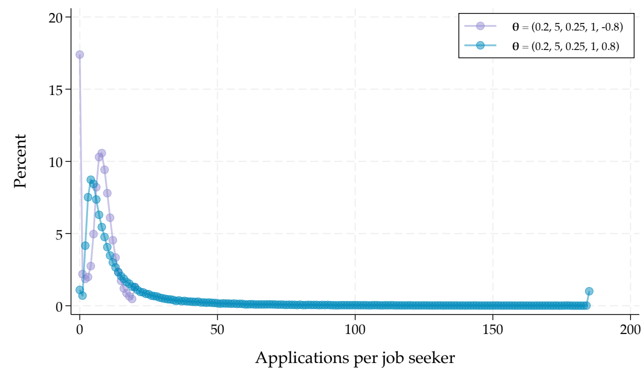
Figure H.1. Further comparative statics in $\boldsymbol{\theta}$



(a) Homogeneous



(b) Heterogeneous independent



(c) Heterogeneous copula

Notes: Additional portfolio size histograms complementing the calibrated simulations from Figure 3 in the main text, illustrating different effects for calibrated parameters in alternative regions. See the notes of Figure 3 for simulation details. In Panel (a), we show how the right tail contracts with higher α when the kink point $n^* = 1/\alpha$ lies in the support. In Panel (b), we show how the independent heterogeneous model can generate long right tails for certain parameter values. In Panel (c), we show that for relatively higher mean α_i compared to Panel (c) of Figure 3, a positive correlation in the Gaussian copula model expands the right tail and concentrates mass at low (but not zero) n relative to negatively correlated α_i and γ_i.

Table H.1. MSM alternative targeted moments sets

ℓ	κ_ℓ	ν_ℓ	Type	$\mathbb{E}\left[m_\ell(n_i) \mid n_i > 0\right]$
1	1		pmf	$\mathbb{P}\left(n_i = 1 \mid n_i > 0\right)$
2	2		pmf	$\mathbb{P}\left(n_i = 2 \mid n_i > 0\right)$
3	4		pmf	$\mathbb{P}\left(n_i = 4 \mid n_i > 0\right)$
4	6		pmf	$\mathbb{P}\left(n_i = 6 \mid n_i > 0\right)$
5	8	10	bin prob.	$\mathbb{P}\left(8 \leq n_i \leq 10 \mid n_i > 0\right)$
6	12		survivor	$\mathbb{P}\left(n_i \geq 12 \mid n_i > 0\right)$

Notes: Description of the $M = 6$ targeted moments chosen to define the MSM criterion function from the set of moment candidates spanned by Equations (49) to (51).

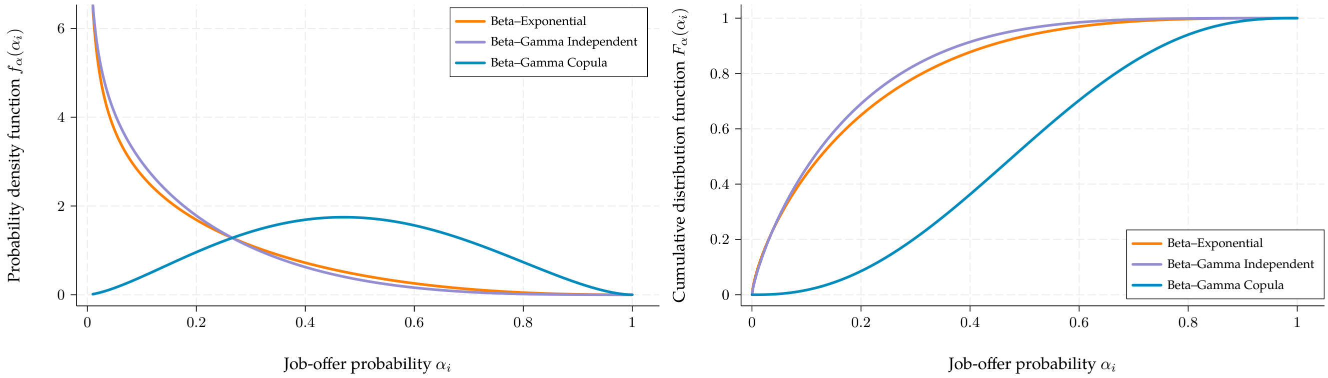
Table H.2. Alternative estimates of the random-effects distribution parameters

Panel (a)		Independent		Copula
	Homogeneous	Beta–Exponential	Beta–Gamma	Beta–Gamma
	(1)	(2)	(3)	(4)
Estimates				
θ_1	1.3491e-5	0.1792	0.1601	0.4817
θ_2	1.1513e-4	3.9738	4.9339	5.2574
θ_3		0.0634	43.4031	19.1699
θ_4			1.0500	0.6183
θ_5				0.3567
Objective function	4,243.6432	22.1203	8.9051	7.7577
Implied marginals				
$\mathbb{E}[\alpha_i]$	1.3491e-5	0.1792	0.1601	0.4817
$\mathrm{Var}(\alpha_i)$	0	0.0296	0.0227	0.0398
$\mathbb{E}[\gamma_i]$	1.1513e-4	15.7733	43.4031	19.1699
$\mathrm{Var}(\gamma_i)$	0	248.7970	1,794.0500	594.3600

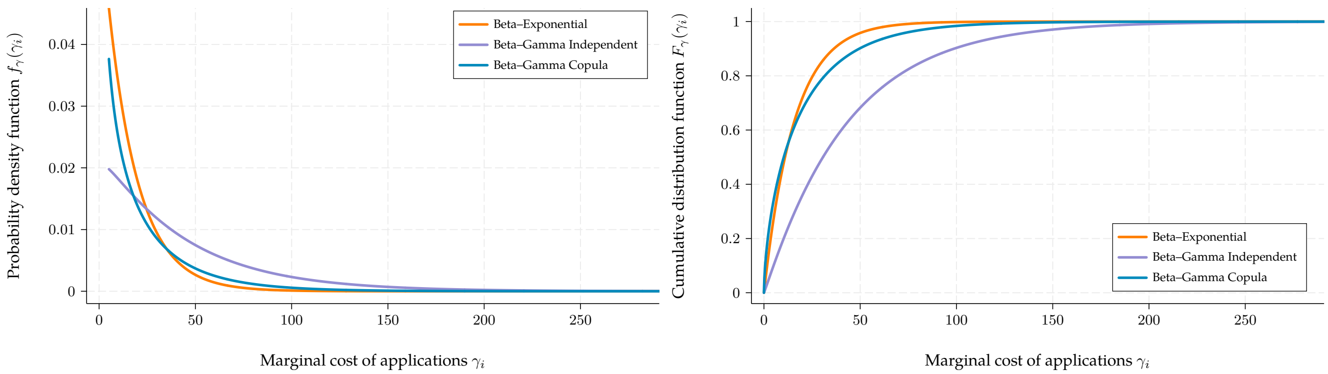
Panel (b)				Independent				Copula	
		Homogeneous		Beta–Exponential		Beta–Gamma		Beta–Gamma	
	Empirical	Simulated	Contrib.	Simulated	Contrib.	Simulated	Contrib.	Simulated	Contrib.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Moments									
Moment 1	0.3871	0.3720	0.0082	0.3812	0.0403	0.3839	0.0874	0.3832	0.1419
Moment 2	0.1861	0.3224	0.3575	0.1949	0.3795	0.1905	0.1546	0.1914	0.2668
Moment 3	0.0712	0.0804	0.0017	0.0743	0.1273	0.0706	0.0159	0.0722	0.0210
Moment 4	0.0383	0.0080	0.1167	0.0348	0.2403	0.0366	0.1710	0.0366	0.1982
Moment 5	0.0509	0.0005	0.2246	0.0476	0.1521	0.0479	0.3622	0.0481	0.3674
Moment 6	0.0663	0.0000	0.2913	0.0684	0.0604	0.0692	0.2090	0.0670	0.0046

Notes: Method-of-simulated-moments estimates of the parameters $\boldsymbol{\theta}$ of the joint distribution $F_{\alpha,\gamma}(\cdot,\cdot \mid \boldsymbol{\theta})$ of the random effects (α_i, γ_i). The bootstrap estimate of the optimal weight matrix has full rank 6 and condition number 12.9026. Targeted moments correspond to those listed in Table H.1.

Figure H.2. Marginal distributions of the random effects implied by the alternative MSM estimates



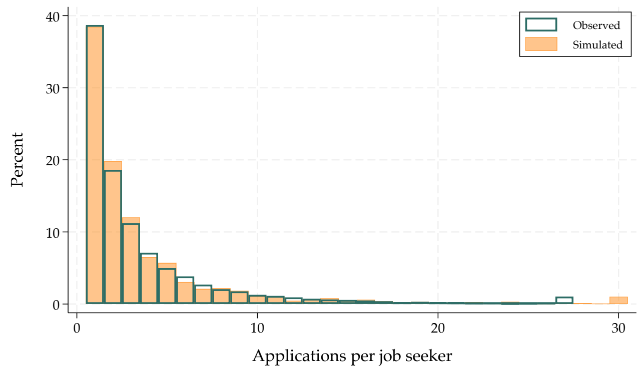
(a) Distribution of α_i



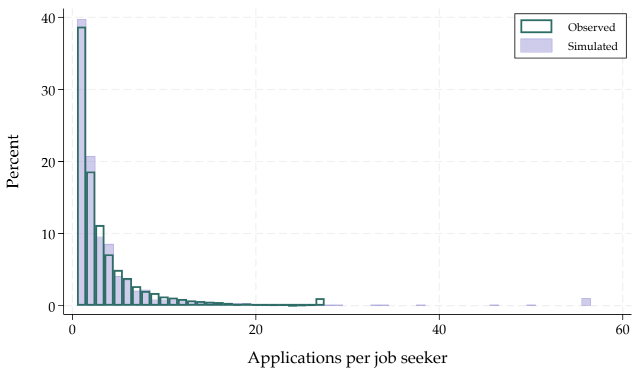
(b) Distribution of γ_i

Notes: Marginal distributions of the random effects (α_i, γ_i) implied by the alternative MSM estimates of $\boldsymbol{\theta}$ in Table H.2 for the nondegenerate specifications (i) independent Beta–Exponential, (ii) independent Beta–Gamma, and Beta–Gamma copula. Supports are visually truncated at $\alpha_i = 0.01$ and $\gamma_i = 5$ in the pdf plots for readability.

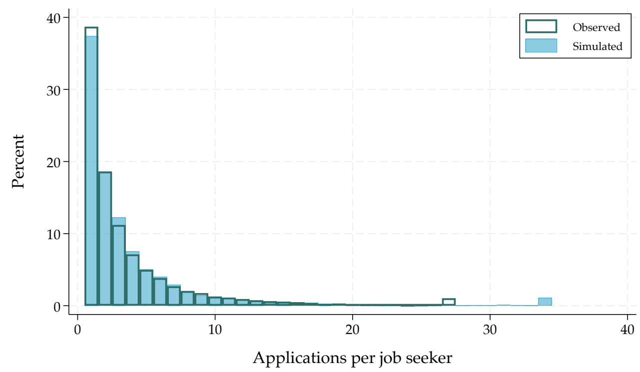
Figure H.3. Full conditional pmf fit of the structural parameters



(a) Beta–Exponential



(b) Beta-Gamma Independent



(c) Beta–Gamma Copula

Notes: Histograms of the simulated versus empirical portfolio size distributions, conditional on applying $n_i > 0$. Simulated conditional distributions are obtained from one simulation draw of 17,404 job seekers under the data-generating process defined by the PROL MLE of $\boldsymbol{\delta}$ from Figure 7 and the MSM estimates of $\boldsymbol{\theta}$ from Table 3. Empirical supports are censored at the corresponding 99–th percentile. This figure, showing the full conditional —i.e., trucated at zero— pmf of portfolio sizes mixes targeted un untargeted moments by our MSM estimator, partially serving as a model fit diagnostic.

Appendix references

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.

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.

Gale, D. and H. Nikaido (1965): "The Jacobian matrix and global univalence of mappings," *Mathematische Annalen*, 159, 81–93.

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.