

# Estimating labor market power from job applications

Pinjas Albagli\*

November 19, 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 application 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. Results show the model is rich enough to successfully replicate observed multi-application patterns and capture significant heterogeneity in firm-level labor supply elasticities that single-application models ignore.

---

\*London School of Economics and Centre for Economic Performance. Email: [p.m.albagli@lse.ac.uk](mailto:p.m.albagli@lse.ac.uk). I am deeply grateful to my supervisor, Alan Manning, for his continued advice, support, and kindness, as well as to my secondary adviser, Alessandro Gavazza, for his generous advice and encouragement. Special thanks to Nicolás Figueroa, whose early advice influenced the modeling approach taken in this paper, and to Benjamín Villena-Roldán and Sekyu Choi, without whose advice and generosity the empirical application would not have been possible. I also thank Ihsaan Bassier, Eduard Boehm, Rui Costa, Andrés Fajardo Ramirez, Lin Fan, Shadi Farahzadi, Cristián Figueroa, Duarte Gonçalves, Jeremias Klaeui, Roberto Maura-Rivero, Guy Michaels, Sidharth Moktan, Thomas Monk, Claudio Mora-García, Steve Pischke, Christopher Sandmann, Mark Schankerman, Lan Zou, and seminar participants at LSE for insightful comments and discussion.

# 1 Introduction

While the concept of monopsony dates back at least to [Robinson \(1933\)](#), and the perspective that labor markets are imperfectly competitive reconciles many otherwise puzzling stylized facts ([Manning, 2003](#)), this view did not gain much traction historically among labor economists. That is, until recently.<sup>1</sup> A falling labor share of national income and increasing inequality ([Autor et al., 2020](#); [ILO, 2019](#); [Karabarbounis and Neiman, 2013](#)) have revived interest in labor market power. As [Langella and Manning \(2021\)](#) put it, the line of enquiry is shifting from whether employers have any monopsony power to how much, how it varies, and what should be done about it.<sup>2</sup>

A growing literature follows [Card et al. \(2018\)](#) in borrowing tools from the discrete choice frameworks typically used in industrial organization to model the supply of labor to differentiated jobs in what [Manning \(2021\)](#) calls the “new classical monopsony” approach (e.g., [Roussille and Scuderi, 2025](#); [Azar et al., 2022](#); [Hirsch et al., 2022](#); [Lamadon et al., 2022](#)).<sup>3</sup> The standard discrete choice framework on which this literature relies is built on the assumption that each decision maker selects only one alternative.<sup>4</sup> When applied to job applications, as in [Azar et al. \(2022\)](#), the framework mechanically implies that each applicant submits only one application, an assumption that is starkly inconsistent with observed behavior and overlooks a fundamental feature of the economic environment, as we discuss below.

This paper relaxes the single-application assumption by extending the standard discrete choice framework to a simultaneous search setting in which uncertainty about job offers induces multi-application behavior. We show that in any model where job seekers apply to multiple vacancies in a wage-sensitive way, the wage elasticity of the supply of applications to the vacancy—and thus to the firm—decomposes into two structurally distinct components: a substitution effect across jobs holding the size of the worker’s application portfolio constant, and a portfolio-size effect capturing how the worker’s optimal number of applications changes. Importantly, this decomposition holds generally for any multi-application model, independent of the specific

---

<sup>1</sup>This is not to say there has been absolutely no work on imperfectly competitive labor markets since the 1930s. See, for example, the discussion of antecedents in Chapter 1 of [Manning \(2003\)](#) or the historical account in [Card \(2022\)](#). But labor economics has been particularly slow to recognize that many employers have wage-setting power, in contrast to other fields, such as industrial organization, where the existence of market power has been widely acknowledged and fully incorporated.

<sup>2</sup>See [Azar and Marinescu \(2024\)](#), [Manning \(2021\)](#), and [Sokolova and Sorensen \(2021\)](#) for recent reviews of the monopsony literature, and [Manning \(2011\)](#) for an earlier review.

<sup>3</sup>In general, market power is typically modeled as emerging from either frictions or idiosyncratic tastes. In the context of employer market power in labor markets, [Manning \(2021\)](#) distinguishes the “modern monopsony” approach based on search frictions (e.g., [Burdett and Mortensen, 1998](#)) from the “new classical monopsony” approach based on heterogeneity in workers’ preferences over differentiated jobs (e.g., [Card et al., 2018](#)).

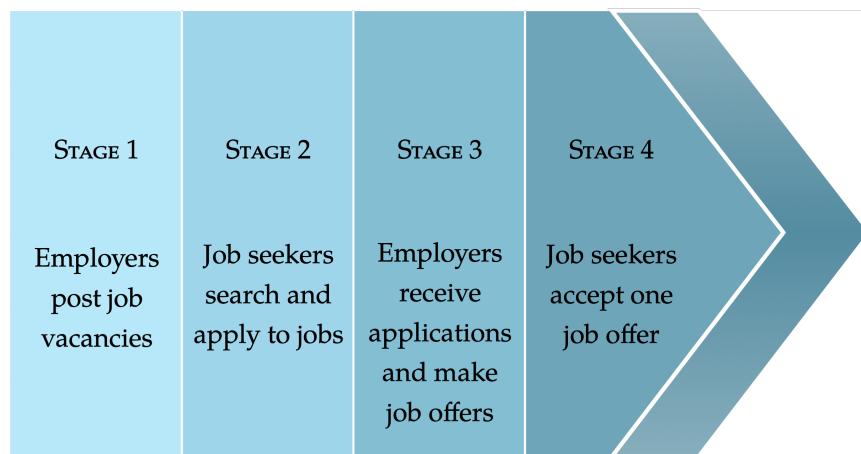
<sup>4</sup>Except for [Roussille and Scuderi \(2025\)](#). The discrete choice part of their model is based on observed partial rankings, and our estimation approach directly borrows from theirs in this regard.

behavioral and parametric structure. It is not an artifact of the particular assumptions we make to yield the first tractable model that formalizes the distinction between these two mechanisms. In an illustrative empirical application to job board data, we demonstrate how single-application models ignore economically relevant heterogeneity in monopsony power.

Mirroring the Lerner index (Lerner, 1934) in product markets, measurement of labor market power is typically based on wage markdowns capturing the wedge between wages and the marginal revenue product of labor that results from imperfect competition. While a strand of the literature follows a direct approach to markdown estimation, leveraging production function estimation techniques and revenue data (e.g., Mertens and Mottironi, 2025; Yeh et al., 2022; Brooks et al., 2021), this paper follows the tradition of estimating firm-level labor supply elasticities (e.g., Caldwell and Oehlsen, 2023; Dube et al., 2020; Dal Bó et al., 2013).

The key idea behind monopsony is that the labor supply curve to an individual employer is less than perfectly elastic (Manning, 2003, 2021). The markdown at the firm level is a function—typically the reciprocal—of this elasticity in a broad class of monopsony models (Azar and Marinescu, 2024). Articles estimating the firm-level labor supply elasticity focus on different parts of the process that determines the firm's level of employment. While it has been recognized how this implies the need for a further translation of the estimates into labor supply elasticities, it is equally important to note that models that appropriately describe one part of the process might be less suitable for another.

**Figure 1.** Timing of the recruitment process



*Notes:* Timeline of the recruitment process in a stylized labor market. Traditional discrete choice models that allow at most one job to be chosen are well-suited for Stage 4, where workers decide among final job offers (see, e.g., Hirsch et al., 2022). However, when applied to Stage 2, 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.

Figure 1 above depicts a stylized timeline of the recruitment process where (1) vacancies are posted, (2) job seekers apply, (3) employers make job offers, and (4) applicants accept their top

offer.<sup>5</sup> The single-choice assumption is naturally well-suited to the context of Stage 4, where candidates choose their most preferred job among their available offers (e.g., Hirsch et al., 2022). In contrast, when applied to job application data from Stage 2, as in Azar et al. (2022), this assumption is unrealistic and unnecessarily restrictive, as it mechanically shuts down the portfolio-size mechanism.

It is a well-documented empirical fact that job seekers typically submit multiple applications in any given search spell. See, for example, Coles et al. (2010) for the economics academic job market context, Belot et al. (2019) for a broader “field-in-the-lab” experimental context, or Azar et al. (2022) for an observational online job board context. Figure 2 below illustrates this point using application data from the Chilean job board Trabajando.com, where multiple applications are systematically prevalent.<sup>6</sup> It depicts the distribution of application counts at different levels of time aggregation to address the concern that the observed multi-application behavior could be an artifact of temporal clustering. Across all specifications, whether aggregating over the entire period, by week, or within search spells, the data consistently show substantial mass above one application per applicant.<sup>7</sup>

The question is then “*why do applicants submit more than one application?*” A quick introspection reveals a natural answer: uncertainty. Job seekers face uncertainty about which applications will result in offers, and this uncertainty induces them to submit multiple applications as a form of insurance. But then, “*why not apply to everything?*” Application costs discipline this behavior, leading to nondegenerate application counts. Yet single-application models cannot capture either mechanism by construction.

We explicitly model job-offer uncertainty and costly applications by layering an additive random utility model on top of the Chade and Smith (2006) portfolio choice framework. Workers have ex-post preferences over differentiated jobs, but face an ex-ante risky choice among available application portfolios: they must balance the benefit of diversifying across multiple applications against the cost of submitting each additional application, recognizing that at most one job offer will ultimately be accepted.

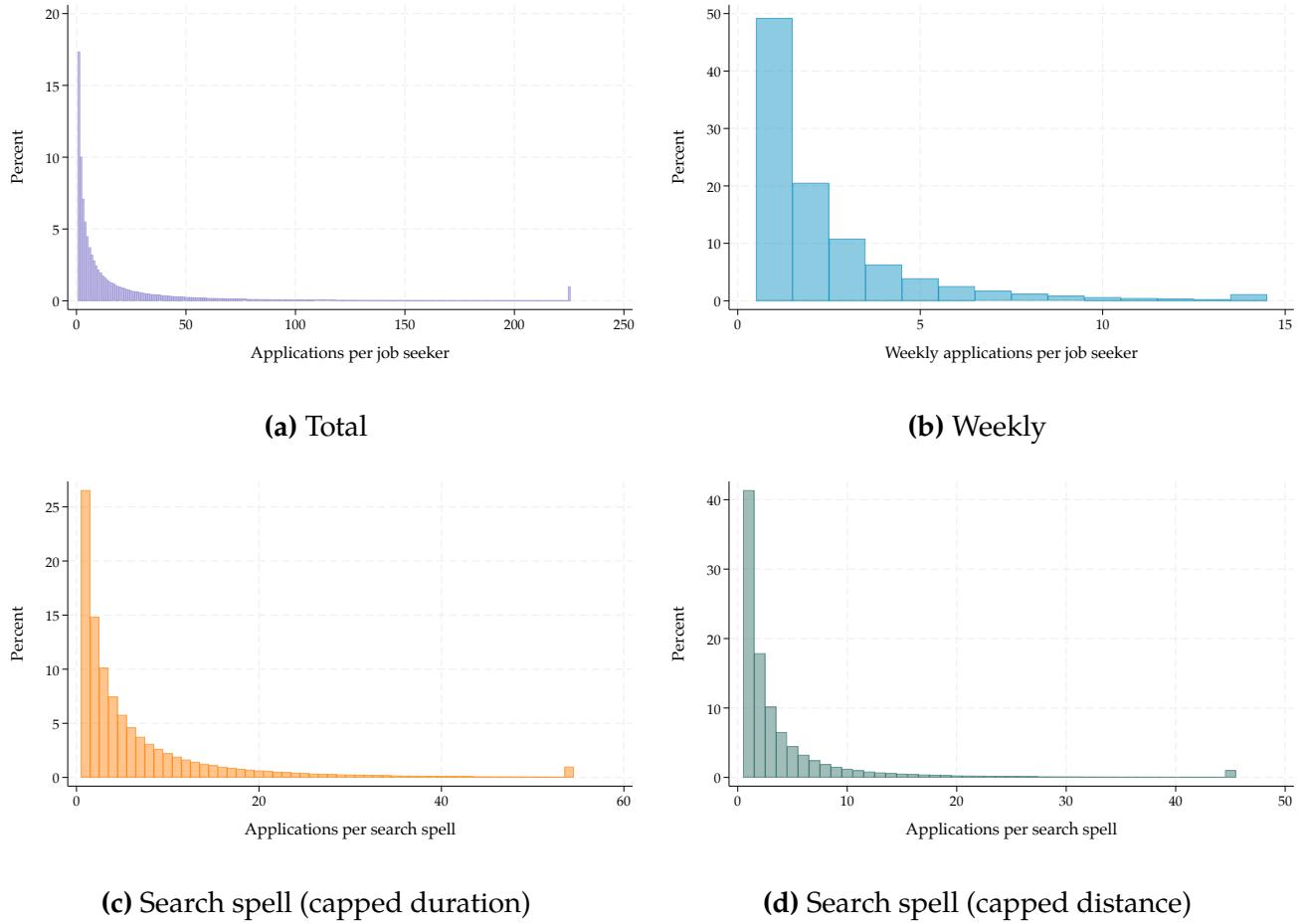
---

<sup>5</sup>Separations are another determinant of the level of employment; however, we abstract from dynamic considerations, on-the-job search, quits, and firings. The setting in Roussille and Scuderi (2025) does not align with this timeline, as they focus on a particular platform where employers bid for candidates who then choose a subset of employers to interview with. An alternative version of Stage 3, where job seekers receive employers’ bids and make “interview offers”, would accommodate their setup.

<sup>6</sup>Job application data from the same source have been used in other studies (e.g., Banfi and Villena-Roldán, 2019) and form the basis of our empirical application described below. See Section 4.1 for details on the data.

<sup>7</sup>See the figure notes for details, including search-spell definitions, and the [Online Appendix](#) for complementary figures with different frequencies and alternative search-spell definitions.

**Figure 2.** Distribution of the number of applications per job seeker



*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.

Each application portfolio induces a lottery over a state space in which each state represents getting a particular job —i.e., accepting an offer from that job. Only jobs in the application portfolio are realized with positive probability: applicants cannot get an offer from a job they did not apply to. Realization of each job in the portfolio is contingent on (i) receiving an offer from that job, and (ii) being rejected by all preferred jobs in the portfolio. Thus, the induced lottery depends on ex-post preferences. Job seekers optimally choose the portfolio that maximizes the expected utility from these lotteries, which can be the empty set representing the

non-application outside option.

A nontrivial challenge in this generalized setting is that ex-ante choice of application portfolios does not generally reveal ex-post preferences over jobs. That is, when we observe the applicant applying to a given job and not applying to another, we cannot infer that they prefer the former to the latter since the application decision is based on expected utility maximization. Importantly, this challenge also implicitly affects standard single-application discrete choice models since the events underlying the construction of choice probabilities are derived from revealed preference. The single application restriction can be thought of as assuming that every job yields an offer with certainty.

We obtain a tractable model of discrete choice under uncertainty by imposing two simplifying assumptions. First, job-offer probabilities vary across workers but not across jobs, making them a worker characteristic rather than a match-specific feature. Second, the marginal application costs are constant but vary across individuals. Under these restrictions, the model retains a revealed-preference structure that allows us to construct choice probabilities from ex-post utility comparisons, while generating endogenous multi-application behavior and rich heterogeneity patterns.

The optimal decision can be separated into two stages. Optimal portfolio size,  $n$ , is determined by a stopping rule comparing rank-ordered ex-post utilities to rank-specific thresholds that depend on the job-offer probability and the marginal cost of applications. Conditional on optimal portfolio size  $n$ , the optimal portfolio is formed by applying to the  $n$  top-ranked jobs. We treat the job-offer probabilities and marginal application costs as individual random effects with a parametric joint distribution and assume an i.i.d. extreme value type 1 (EV<sub>1</sub> henceforth) distribution for the idiosyncratic taste shocks. This yields closed-form expressions for choice probabilities that generalize the standard multinomial logit to the portfolio-choice setting.

We define the conditional inclusion probability as the probability that job  $j$  is included in worker  $i$ 's optimal portfolio conditional on submitting  $n$  applications. Application supply is the expectation of the number of applications to job  $j$  against the distribution defined by the composition of the conditional inclusion probabilities and the portfolio-size probability mass function (pmf hereafter). The model yields a decomposition of the wage elasticity of job applications into two components: a substitution effect, operating through the elasticities of the conditional inclusion probabilities (changes in which jobs enter the portfolio, holding portfolio size fixed), and a portfolio-size effect, operating through the elasticities of the portfolio-size pmf over its support (changes in how many applications each worker submits).

The elasticity 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 decom-

position explicit, we clarify which components of application behavior existing new classical monopsony models can and cannot capture. The single-application restriction is not merely an innocuous simplification: it mechanically shuts down the portfolio-size channel, rendering the estimated elasticities structurally misspecified even in the ideal setting where revealed preference holds and mean utilities are consistently estimated. Single-application models identify substitution effects but miss portfolio-size adjustments entirely.

This omission may have important implications for cross-firm heterogeneity in market power. Single-application models mechanically attribute all variation in elasticities to the substitution margin, since portfolio sizes are fixed at one by assumption. In contrast, our framework allows for richer patterns of variation through the interaction of substitution and portfolio-size effects. The empirical application discussed below suggests that this additional source of variation is quantitatively important: estimated elasticities exhibit substantially greater dispersion in the multi-application model than in a single-application benchmark. This is consistent with the portfolio-size margin varying in importance across firms, with systematically more attractive employers less exposed to this channel.

Our theoretical framework is tractable enough to take to observational job application data. We develop a multi-stage econometric method that exploits the model's separable structure for identification. Mean utilities are identified from conditional inclusion probabilities via a generalization of logit inversion, and the data structure yields what we call a "*partially rank-ordered logit*" model. Maximum likelihood estimation, however, involves combinatorially complex sums over portfolio subsets that quickly become computationally infeasible with large choice sets. We address this computational challenge using a minorize–maximize algorithm, originally developed by [Roussille and Scuderi \(2025\)](#), that makes estimation feasible with thousands of jobs.

We flexibly model the joint distribution of the random effects through a copula representation. The parameters governing this joint distribution individually control the mean and dispersion of the marginal distributions and the degree of comonotone or countermonotone dependence through the copula parameters. Since the marginal-cost random effect enters the stopping-rule thresholds linearly while the job-offer probability enters nonlinearly, changes in the respective mean and dispersion of the marginal distributions, as well as in copula parameters, produce distinct effects on the pmf of optimal portfolio sizes. Therefore, the random-effect distributional parameters are separately identified from the empirical distribution of portfolio sizes once mean utilities are known or consistently estimated. Rather than dealing with computationally intractable analytical moments, we employ a simulated method of moments estimator, bootstrapping the optimal weight matrix.

Application elasticities also depend on the coefficient on log-wage in the regression of mean

utilities on log-wages and job characteristics. This coefficient is identified through the exclusion of rival jobs' attributes from the equation defining the ex-post utility of a vacancy. Hence, we estimate it through BLP-style (Berry et al., 1995) instrumental variable regression using sums of rival-firm jobs' characteristics as excluded instruments. Finally, we compute firm-level elasticities via simulation and numerical differentiation, using directional derivatives to bypass expensive vacancy-level calculations.

An illustrative application to microdata from Trabajando.com, a prominent Chilean job board, demonstrates the method's feasibility and economic content. Using data from the first quarter of 2018 covering thousands of job seekers and job postings, we estimate the model and compute firm-level application wage elasticities. We model the joint distribution of job-offer probabilities and application costs through a Gaussian copula structure with a scalar correlation parameter, comparing results across progressively simpler specifications: independent marginals, degenerate marginals, and homogeneous parameters. The mean utility estimates are stable across both our multi-application framework and two single-application benchmarks (multinomial and nested logit), consistent with the three models relying on the same source of identification.

The estimates reveal substantial heterogeneity in both job-offer probabilities and application costs across unemployed job seekers. Our preferred copula specification indicates a mean job-offer probability of 54 percent with meaningful dispersion, suggesting workers face considerable uncertainty about whether applications will yield offers. Application costs exhibit even greater variation. Notably, the estimated Gaussian copula correlation parameter is strongly positive (0.85), indicating strong comonotone dependence: workers with better job prospects also tend to face higher application costs and vice versa. The model successfully replicates the observed distribution of portfolio sizes, with the copula specification achieving substantially better fit than alternatives imposing independence or homogeneity.

Comparing firm-level application elasticities between the multi-application model and the nested logit single-application benchmark reveals a striking difference in the implied distribution of market power. While both approaches yield similar median elasticities (around 0.4), the multi-application model produces substantially greater dispersion. The interquartile range is approximately six times as large in the multi-application specification (0.17 versus 0.03 in the nested logit benchmark), and the range extends from 0.06 to 3.40 compared to 0.29 to 0.88. Moreover, estimated elasticities in the multi-application model exhibit a negative relationship with mean utilities (correlation -0.25), while this pattern is weaker in the single-application benchmark (correlation -0.09). Most of the range variation between the multi-app and single-app empirical distributions stems from systematically unattractive firms (those with the smallest average mean utilities). This is consistent with the portfolio-size mechanism varying in im-

portance across firms: employers with systematically more attractive vacancies face less elastic application supply and are less exposed to the portfolio-size effect as their jobs are more likely to be included in small portfolios, while less attractive firms are more exposed to both substitution and portfolio-size adjustments.

These results suggest that single-application models may systematically underestimate the heterogeneity of market power across employers. We view our results as an open invitation to start thinking seriously about how uncertainty and search frictions shape job application behavior and translate into monopsony power. This paper establishes that accounting for multiple applications is both feasible and consequential. As a first step in this direction, our framework necessarily abstracts from several important features of labor markets. Relaxing the strong assumption that job-offer probabilities do not vary with jobs, endogenizing these probabilities in an equilibrium setting, and explicitly modeling the demand side (especially the selection dimension of recruitment) remain exciting directions for future research.

The remainder of the paper is organized as follows. Section 2 develops the theoretical framework, deriving the elasticity decomposition, comparing the model to single-application benchmarks, and exploring the comparative statics of the portfolio-size distribution with respect to the random-effect distributional parameters. Section 3 deals with econometrics, establishing identification and developing the multi-stage estimation strategy. Section 4 presents the empirical application, providing details on specification choices and discussing the results. Finally, Section 5 concludes. Formal proofs and derivations are relegated to the Appendix.

## 2 Theory: 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  $x_j \in \mathbb{R}^K$  that we will assume observed by the econometrician when discussing estimation in Sections 3.2 and 4.2, and a scalar index  $\xi_j$  capturing other job characteristics that we will assume unobserved. Job characteristics  $(x'_j, \xi_j)'$  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  $c_i : \mathbb{N} \rightarrow \mathbb{R}_+$  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  $u_i : \mathcal{J} \cup \{0\} \rightarrow \mathbb{R}$ , 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 all  $j \in \mathcal{J}$ , implying the job seeker accepts at least one offer, if any.<sup>8</sup> 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  $r_i : \mathcal{P}(\mathcal{J}) \times \{1, \dots, J\} \rightarrow \mathcal{J}$  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,  $k \in \{1, \dots, |A|\}$  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  $u_{ir_i(\mathcal{J}, 1)} > \dots > u_{ir_i(\mathcal{J}, J)}$ .<sup>9</sup> Each application portfolio  $A \subseteq \mathcal{J}$  gives rise to a lottery over state space  $\mathcal{J} \cup \{0\}$ , where outcomes  $j \in \mathcal{J}$  represent exercising option  $j$ —i.e., getting the job—and outcome  $j = 0$  corresponds to exercising the outside option. The lottery assigns positive probability only to jobs in the application portfolio,  $j \in A$ , and to the outside option,  $j = 0$ . As discussed above, option  $j \in A$  is exercised if and only if the job seeker (i) receives an offer from job  $j$ , and (ii) is rejected by every job in the portfolio that is (ex post) preferred to  $j$ . Therefore, if  $j$  is ranked in the  $k$ -th position among  $m \in A$ , then the probability of exercising this option is given by

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

---

<sup>8</sup>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  $A_i = \emptyset$ .

<sup>9</sup>Moreover, for fixed  $A \subseteq \mathcal{J}$ ,  $r_i(A, \cdot)$  is a bijection from  $\{1, \dots, |A|\}$  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  $A$ .

Similarly, the probability of exercising the outside option is<sup>10</sup>

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

Let  $U_i : \mathcal{P}(\mathcal{J}) \rightarrow \mathbb{R}$  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—,

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

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

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

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  $\ell$  truthfully reveals that the decision maker prefers  $j$  to  $\ell$  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  $j$ :

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

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

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

---

<sup>10</sup>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  $A \cup \{0\}$ .

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  $j, \ell$  such that  $u_{ij} > u_{i\ell}$ , the MIA will choose  $j$  over  $\ell$  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  $\alpha_i$  and  $\gamma_i$ . 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*

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

(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*

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

*Proof.* See Appendix B.1. □

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

$$u_{ij} = \delta_j + \varepsilon_{ij}, \quad (7)$$

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

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

Equations (7)–(8), together with Assumption 3 below, comprise the core of our logit ARUM structure.<sup>11</sup> We complete the model with a parametric distributional assumption on the individual uncertainty and cost parameters  $\alpha_i$  and  $\gamma_i$ , treating them as random effects, in Assumption 4.

**Assumption 3 (Logit ARUM).** The idiosyncratic taste shocks  $\varepsilon_{ij}$  are independent and identically distributed draws from a standard type-1 extreme value distribution, with cumulative distribution function  $F_\varepsilon(x) = \exp(-\exp(-x))$  for  $x \in \mathbb{R}$ , and independent of the random effects  $(\alpha_i, \gamma_i)$ .

**Assumption 4 (Uncertainty and cost random effects).** The random effects  $(\alpha_i, \gamma_i)$ , where  $\alpha_i$  is the individual job-offer probability and  $\gamma_i$  is the individual marginal cost of applications, are

---

<sup>11</sup>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.

independent and identically distributed draws from parametric joint cumulative distribution function  $F_{\alpha,\gamma}(x, y | \boldsymbol{\theta})$  for  $(x, y) \in (0, 1) \times (0, \infty)$ , with parameters  $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^L$ .

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  $\mathcal{J} = \{1, \dots, J\}$  so we can use vector notation for quantities such as  $\boldsymbol{\delta} = (\delta_1, \dots, \delta_J)' \in \mathbb{R}^J$ .<sup>12</sup> The expected number of applications to job  $j$  in our model is given by

$$q_j(\boldsymbol{\delta}, \boldsymbol{\theta}) = I \sum_{n=1}^J s_{j|n}(\boldsymbol{\delta}) s_n(\boldsymbol{\delta}, \boldsymbol{\theta}), \quad (9)$$

where

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

is the probability that job  $j$  belongs to the optimal application portfolio conditional on its size,  $\mathbb{P}(j \in A_i | n_i = n)$  for  $n \in \{1, \dots, J\}$ ,  $\mathcal{B}_j \equiv \mathcal{J} \setminus \{j\}$  is the set of jobs excluding  $j$ , and  $\mathcal{R}_k(S) = \{\sigma \subseteq S : |\sigma| = k\}$  is the set of all size- $k$  subsets of set  $S$ .<sup>13</sup> The conditional probability mass function (pmf) of portfolio size  $n_i$ —i.e., the number of applications—given admission probability  $\alpha_i$  and marginal cost of application  $\gamma_i$ ,  $\mathbb{P}(n_i = n | \alpha_i, \gamma_i)$ , is

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

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

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

---

<sup>12</sup>Alternatively, fix a bijection  $g : \mathcal{J} \rightarrow \{1, \dots, J\}$  such that  $\boldsymbol{\delta} = (\delta_{g^{-1}(1)}, \dots, \delta_{g^{-1}(J)})'$  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  $\mathcal{J}$ .

<sup>13</sup>Note that (i)  $s_{j|J}(\boldsymbol{\delta}) = 1$ , consistent with the trivial fact that  $\mathbb{P}(j \in A_i | n_i = J) = 1$ ; (ii)  $s_{j|1}(\boldsymbol{\delta}) = \exp(\delta_j) / \sum_{\ell \in \mathcal{J}} \exp(\delta_\ell)$  coincides with the well-known choice probability in the multinomial logit model; (iii)  $\sum_{j \in \mathcal{J}} s_{j|n}(\boldsymbol{\delta}) = n$ , consistent with the conditioning event that  $n$  alternatives are chosen; and (iv)  $s_{j|n}(\boldsymbol{\delta})$  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.

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

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

and

$$\vartheta_{J|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i) = 1 - \sum_{k=1}^J \sum_{A \in \mathcal{R}_k(\mathcal{J})} F_\varepsilon(\psi_i^J)^{\sum_{\ell \in A} \exp(\delta_\ell)} \prod_{m \in \mathcal{J} \setminus A} \left[ 1 - F_\varepsilon(\psi_i^J)^{\exp(\delta_m)} \right], \quad (14)$$

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

$$\vartheta_n(\boldsymbol{\delta}, \boldsymbol{\theta}) = \int \vartheta_{n|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i) dF_{\alpha,\gamma}(\alpha_i, \gamma_i | \boldsymbol{\theta}), \quad (15)$$

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  $\ell$  increases by one percent, what is the percent increase in the number of applications to job  $j$ ?*” and is given by

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

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:

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

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

$$\omega_{n|j}(\boldsymbol{\delta}, \boldsymbol{\theta}) = \frac{I \vartheta_{j|n}(\boldsymbol{\delta}) \vartheta_n(\boldsymbol{\delta}, \boldsymbol{\theta})}{q_j(\boldsymbol{\delta}, \boldsymbol{\theta})}$$

sum to one over  $n \in \{1, \dots, J\}$  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  $\mathcal{J}^f$ ,

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

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:

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

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 B.2 for a derivation of the vacancy- and firm-level elasticities, and the decomposition into the substitution and portfolio-size effects. See Appendix B.3 for closed-form expressions of the partial derivatives of  $s_{j|n}(\boldsymbol{\delta})$  and  $s_n(\boldsymbol{\delta}, \boldsymbol{\theta})$  with respect to  $\delta_\ell$  —up to integration over  $F_{\alpha, \gamma}(\cdot, \cdot | \boldsymbol{\theta})$ .

## 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  $n$ ,  $\mathbb{P}(j \in A_i | n_i = n)$  (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  $\mathbb{P}(j \in A_i | n_i = n)$  are well defined.

What the restrictive structure imposed by Assumptions 1 to 4 gives us is a set of injective mappings  $\{\boldsymbol{\delta} \mapsto \mathbb{P}(j \in A_i | n_i = n)\}_{j \in \mathcal{J}, n \in \{1, \dots, J-1\}}$  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  $\mathbb{P}(j \in A_i | n_i = n)$  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  $\mathbb{P}(j \in A_i \mid n_i = n) = \sigma_{j|n}(\mathbf{w}, \mathbf{X}, \Xi, \alpha, \gamma)$  and  $\mathbb{P}(n_i = n) = \sigma_n(\mathbf{w}, \mathbf{X}, \Xi, \alpha, \gamma)$  for  $j \in \mathcal{J}$  and  $n \in \{0, \dots, J\}$ , where the arguments are vectors and matrices suitably stacking  $\{(w_j, \mathbf{x}'_j, \xi_j)\}_{j \in \mathcal{J}}$  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  $(\mathbf{w}^0, \cdot)$ , so that the partial derivatives

$$\frac{\partial \sigma_{j|n}(\mathbf{w}, \mathbf{X}, \Xi, \alpha, \gamma)}{\partial \mathbf{w}} \Bigg|_{\mathbf{w}=\mathbf{w}^0}, \quad \frac{\partial \sigma_n(\mathbf{w}, \mathbf{X}, \Xi, \alpha, \gamma)}{\partial \mathbf{w}} \Bigg|_{\mathbf{w}=\mathbf{w}^0}$$

exist and are finite, then the semi-elasticities

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

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

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

follows through by differentiating the identity

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

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

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

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  $\theta$  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  $\theta = (\alpha, \gamma)$ . In this case, the thresholds in Equation (12) simplify to

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

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)

$$s_0(\delta, \alpha, \gamma) = F_\varepsilon\left(\frac{\gamma}{\alpha}\right)^{\sum_{\ell \in \mathcal{J}} \exp(\delta_\ell)}. \quad (21)$$

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  $\theta = (1, \gamma)$ , violating the support restriction on  $\alpha_i$  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.<sup>14</sup>

**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:*

$$n_i \equiv |A_i| \in \{0, 1\}. \quad (22)$$

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

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

(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*

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

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

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

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

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

*Proof.* See Appendix B.4. □

---

<sup>14</sup>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  $\varepsilon_{i0} \equiv \gamma_i \stackrel{\text{iid}}{\sim} \text{EV}_1$ . Since the  $\text{EV}_1$  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  $j_i^* = \arg \max_{j \in \mathcal{J} \cup \{0\}} u_{ij} - \mathbf{1}\{j \in \mathcal{J}\} \gamma_i = \arg \max_{j \in \mathcal{J} \cup \{0\}} \tilde{u}_{ij}$ ,  $\tilde{u}_{ij} = \delta_j + \varepsilon_{ij}$ , 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

$$\tilde{q}_j(\boldsymbol{\delta}, \gamma) = I\bar{s}_j(\boldsymbol{\delta}) [1 - \bar{s}_0(\boldsymbol{\delta}, \gamma)]. \quad (27)$$

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  $\gamma/\alpha$ , 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  $n_i = \mathbb{1}\{u_{ir_i(\mathcal{J},1)} \geq \gamma/\alpha\}$ . Alternatively, follow through the steps of the proof of Proposition 1 in Appendix B.1, 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:

$$\begin{aligned} \bar{q}_j(\boldsymbol{\delta}, \alpha, \gamma) &= I\bar{s}_j(\boldsymbol{\delta}) [1 - \bar{s}_0(\boldsymbol{\delta}, \alpha, \gamma)] \\ &= I \sum_{n=1}^J s_{j|1}(\boldsymbol{\delta}) s_n(\boldsymbol{\delta}, \alpha, \gamma) \\ &< I \sum_{n=1}^J s_{j|n}(\boldsymbol{\delta}) s_n(\boldsymbol{\delta}, \alpha, \gamma) \end{aligned}$$

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

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  $\{\varsigma_{j|n}(\boldsymbol{\delta})\}_{n=1}^J$  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,  $\alpha_i$ , and application costs,  $\gamma_i$ : they only enter the optimal stopping-rule that determines the optimal portfolio size,  $n_i$ . That is, together with the idiosyncratic taste shocks  $\varepsilon_{ij}$ , 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  $F_{\alpha, \gamma}(\alpha_i, \gamma_i | \boldsymbol{\theta})$  induce different portfolio-size distributions  $\varsigma_n(\boldsymbol{\delta}, \boldsymbol{\theta})$ . 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  $\gamma$  enters the stopping-rule thresholds (20) multiplicatively with a positive derivative that increases monotonically with  $n \in \{1, \dots, J\}$ ,

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

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  $\gamma$ . Therefore, since probabilities sum to one over the support,  $\mathbb{P}(n_i > 0)$  must fall with an increase in  $\gamma$ , 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  $\alpha$  does not shift all thresholds in the same

direction since the derivative

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

is negative for small  $n$  and positive for large  $n$ . In particular, there is a kink at  $n^* = 1/\alpha$ , so  $\mathbb{P}(n_i = n)$  increases with  $\alpha$  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

$$\frac{\psi^n - \psi^{n-1}}{\psi^{n-1}} = \frac{\alpha}{1 - \alpha}, \quad n \in \{2, \dots, J\} \quad (30)$$

is constant and depends only on  $\alpha$ . Therefore,  $\alpha$  determines the rate of decay of the right tail of the portfolio-size distribution. Higher  $\alpha$  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.<sup>15</sup> Both parameters reduce the frequency of large application portfolios, but operate through different mechanisms. Higher  $\gamma$  raises all thresholds proportionally, uniformly discouraging applications and compressing the distribution around small  $n$  and non-participation. In contrast, higher  $\alpha$  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,  $\gamma$  shifts the level of application intensity, while  $\alpha$  determines curvature.

Panel (a) of Figure 3 illustrates these comparative statics for  $\alpha$  on the right-hand plot and for  $\gamma$  on the left-hand plot. We simulate  $I = 100,000$  job seekers choosing among  $J = 1,000$  jobs for different values of  $\theta = (\alpha, \gamma)$  holding constant the mean utilities  $\delta$  and the  $EV_1$  taste shock draws across simulations. Note that as we vary  $\alpha$  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 G.1 in Appendix G illustrates cases with higher  $\alpha$  —and thus lower kinks— where the compression of the right tail is observed, but with zero mass at lower

---

<sup>15</sup>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  $\gamma$  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  $n$ .

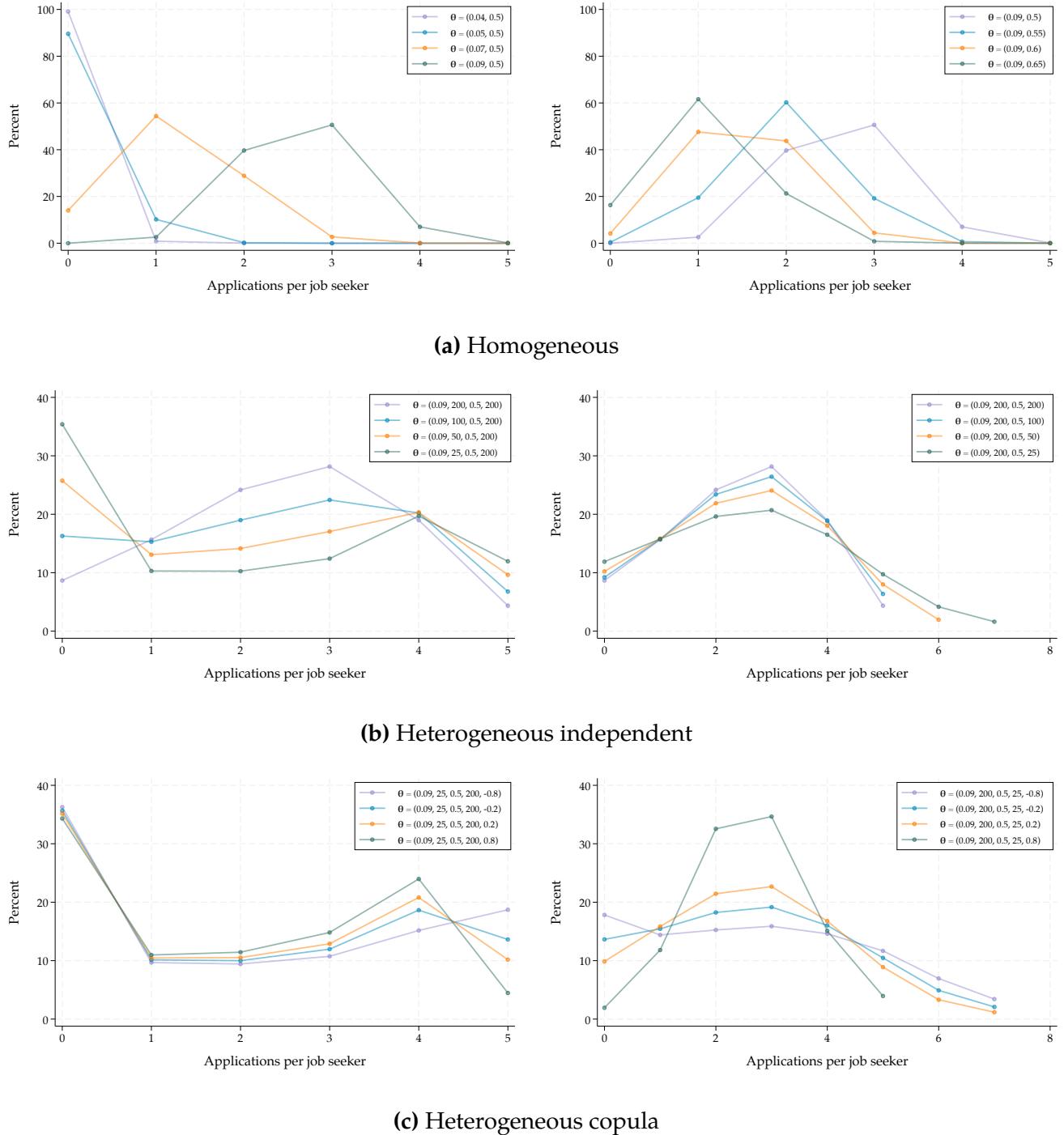
In the mutually-independent heterogeneous case, when the joint distribution of  $(\alpha_i, \gamma_i)$  is nondegenerate and factors into the product of the marginals, analogous arguments hold conditional on  $i$  since the functional form of the individual thresholds  $\psi_i^n$  in (20) remains the same. The observed pmf of  $n_i$  is a mixture of these baseline shapes where  $\gamma_i$  scales the level of all individual thresholds and  $\alpha_i$  controls their growth rate. Higher dispersion in  $\gamma_i$  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  $\gamma_i$ . Dispersion in  $\alpha_i$  introduces curvature effects: low- $\alpha_i$  individuals flatten the stopping sequence (slower threshold growth rate) and apply to more jobs, while higher- $\alpha_i$  individuals truncate their portfolios earlier (faster growing thresholds). As a result, variation in  $\alpha_i$  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  $\alpha_i$  is necessary, which will depend both on the mean and dispersion of  $\alpha_i$ . Hence, unlike  $\gamma_i$ -heterogeneity, which rescales thresholds uniformly,  $\alpha_i$ -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  $\alpha_i$  keeping its mean and the marginal distribution of  $\gamma_i$  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

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

setting the means at  $\mathbb{E}[\alpha_i] = \theta_1 = 0.09$  and  $\mathbb{E}[\gamma_i] = \theta_3 = 0.5$  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  $(\alpha_i, \gamma_i)$  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  $\text{Var}(\gamma_i) = (\theta_3)^2 / \theta_4$  over  $\{0.00125, 0.0025, 0.005, 0.01\}$  by respectively varying  $\theta_4$  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  $\text{Var}(\alpha_i) = \theta_1(1 - \theta_1) / (1 + \theta_2)$  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.

**Figure 3. Comparative statics in  $\theta$**



*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 ( $\alpha_i, \gamma_i$ ) are assumed homogeneous in Panel (a), heterogeneous, mutually independent with marginal distributions  $\alpha_i \stackrel{\text{iid}}{\sim} \text{Beta}(\theta_1\theta_2, (1-\theta_1)\theta_2)$  and  $\gamma_i \stackrel{\text{iid}}{\sim} \text{Gamma}(\theta_4, \theta_3/\theta_4)$  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  $\theta_5$ . We keep  $\delta$  and the  $\varepsilon_{i,j}$  EV<sub>1</sub> draws constant across simulations, and vary (i) the level in Panel (a) or (ii) the dispersion in Panel (b) of  $\alpha_i$  keeping that of  $\gamma_i$  constant, and vice versa. In Panel (c) we vary the Gaussian copula correlation parameter keeping the marginal distributions of  $\alpha_i$  and  $\gamma_i$  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  $(\alpha_i, \gamma_i)$  for different values of  $\theta$  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  $\theta_2$  for  $\alpha_i$  and  $\theta_4$  for  $\gamma_i$ . Portfolio size is discrete; probability mass points are connected by colored lines purely for visual continuity.

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  $\alpha_i$  and  $\gamma_i$ .<sup>16</sup> This correlation parameter captures the dependence between  $\alpha_i$  and  $\gamma_i$  and determines how much probability mass lies near the diagonal of the  $(\alpha, \gamma)$ -type space, where low  $\alpha_i$  is paired with low  $\gamma_i$  and high  $\alpha_i$  with high  $\gamma_i$ .<sup>17</sup> 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

$$s_n(\boldsymbol{\delta}, \boldsymbol{\theta}) = \int_{[0,1]^2} s_{n|\alpha,\gamma}(\boldsymbol{\delta}, F_\alpha^{-1}(u \mid \boldsymbol{\theta}_\alpha), F_\gamma^{-1}(v \mid \boldsymbol{\theta}_\gamma)) dC(u, v \mid \theta_C), \quad (31)$$

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

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

and how they intersect the the comonotone-rank path

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

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  $\{s_n(\boldsymbol{\delta}, \boldsymbol{\theta})\}_{n=0}^J$  shifts mass from the extremes towards the middle of the support as  $\theta_C$  increases. If instead the intersection of  $(\mathcal{S}_n^k(\boldsymbol{\delta}))_{k \in (0,1)}$  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  $\theta_C$  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  $\theta_C$  changes the simulated portfolio-size distribution in our calibrated setup from Panels (a) and

<sup>16</sup>See, e.g., Trivedi and Zimmer (2007) for details on copulas and, in particular, the Gaussian (or normal) copula.

<sup>17</sup>Here, formally, by diagonal we mean the comonotone diagonal in rank space:  $\{(U, V) \in (0, 1)^2 : U = V\}$ , where  $U = F_\alpha(\alpha_i)$  and  $V = F_\gamma(\gamma_i)$  are the marginal ranks.

(b). In particular, we hold fixed a pair of Beta–Gamma marginal distributions from Panel (b) such that dispersion in  $\alpha_i$  is relatively high while  $\gamma_i$ —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 G.1 for an alternative calibration where the second case is observed.

#### 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  $\delta_k$  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 1. 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 equilibrium 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 \(1998\)](#).<sup>18</sup> 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

---

<sup>18</sup>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.

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.

### 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  $\delta_0$ ,  $\theta_0$ , and  $\beta_{w0}$  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  $\eta_{q^f, w^f}(\delta_0, \theta_0, \beta_{w0})$  in Section 3.3.

#### 3.1 Identification

The conditional application share functions  $s_{j|n}(\delta)$  map  $\delta$  to the conditional inclusion probabilities  $\mathbb{P}(j \in A_i | n_i = n)$  and are directly derived from the EV<sub>1</sub> assumption on the idiosyncratic taste shocks  $\{\varepsilon_{ij}\}_{i \in \mathcal{I}, j \in \mathcal{J}}$ . They do not depend on the random effects  $(\alpha_i, \gamma_i)$  or their distribution  $F_{\alpha, \gamma}(\cdot, \cdot | \theta)$ . The latter determine the distribution of  $n_i$  across job seekers.<sup>19</sup> Let  $s_{j|n}$  and  $s_n$  denote the observed conditional shares and the empirical distribution of portfolio sizes, respectively. The equations

$$s_{j|n}(\delta_0) = \mathbb{E}[s_{j|n}], \quad j \in \mathcal{J}, \quad n \in \{1, \dots, J-1\} \quad (32)$$

cleanly pin down the underlying mean utilities  $\delta_0$  completely independently of  $(\alpha_i, \gamma_i)$  as a result of the revealed-preference structure induced by Assumption 1. Then, knowing  $\delta_0$ , the equations

$$s_n(\delta_0, \theta_0) = \mathbb{E}[s_n] \quad (33)$$

pin down  $\theta_0$ . This (conditional) separation underlies our two-step identification argument, which requires the following assumption.

---

<sup>19</sup>Technically, the optimal stopping rule in part (i) of Proposition 1 implies that  $\delta$  and  $\theta$  jointly determine the pmf of  $n_i$  through the distributions of the taste shocks  $\varepsilon_{ij}$  and the random effects  $(\alpha_i, \gamma_i)$ .

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

- (a) The pmf of portfolio sizes  $s_n(\delta_0, \theta_0)$  assigns positive probability to at least one interior  $n$ :

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

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

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

Part (a) of Assumption 5 guarantees that there is at least one informative portfolio size  $n$  in the support such that conditional inclusion probabilities  $s_{j|n}(\delta)$  identify  $\delta_0$ . The stronger condition in part (b) is necessary to ensure enough variation to identify the parameters of the random-effects distribution,  $\theta_0$ , from the shape of the pmf of  $n_i$ ,  $s_n(\delta_0, \theta_0)$ , given knowledge of mean utilities  $\delta_0$ . Identification of  $\delta_0$  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  $\delta_j$  is identified as

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

We can generalize this invertibility property to all interior  $n \in \{1, \dots, J-1\}$ .<sup>20</sup> Proposition 2 below formalizes this result.

**Proposition 2** (Identification of mean utilities). *Suppose Assumptions 1 to 5 hold. Fix an interior portfolio size  $n \in \{1, \dots, J-1\}$  and a base alternative  $b \in \mathcal{J}$ . Impose the location normalization  $\delta_b = 0$  and let  $\delta^b \in \mathbb{R}^{J-1}$  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  $s_{\cdot|n}^b : \mathbb{R}^{J-1} \rightarrow (0, 1)^{J-1}$ , stacking the conditional inclusion probabilities  $s_{j|n}^b(\delta^b) \equiv s_{j|n}(\delta - \delta_b \boldsymbol{\iota})$  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  $\mathbb{P}(n_i = n) > 0$  at size  $n$ , then  $\mathbb{E}[s_{j|n}]$  is observable at the population level. The restricted mapping  $s_{\cdot|n}^b(\cdot)$  is globally injective on  $\mathbb{R}^{J-1}$ , and thus a bijection from  $\mathbb{R}^{J-1}$  onto the image  $s_{\cdot|n}^b(\mathbb{R}^{J-1}) \subseteq (0, 1)^{J-1}$ . Therefore, the true normalized mean utility vector  $\delta_0^b \in \mathbb{R}^{J-1}$  is identified by inverting the  $(J-1) \times (J-1)$  system of equations*

$$s_{j|n}^b(\delta_0^b) = \mathbb{E}[s_{j|n}], \quad j \in \mathcal{B}_b, \tag{34}$$

*so the true mean utility vector  $\delta_0$  is identified up to a location normalization from the conditional inclusion probabilities  $\{s_{j|n}(\delta_0)\}_{j \in \mathcal{J}} = \{\mathbb{E}[s_{j|n}]\}_{j \in \mathcal{J}}$ .*

---

<sup>20</sup>The boundary case  $n = J$  is uninformative since  $\mathbb{P}(j \in A_i \mid n_i = J) \equiv 1$  holds identically for all  $(\delta, \theta)$ .

*Proof.* See Appendix B.5.  $\square$

Having established identification of  $\delta_0$ , we now consider identification of  $\theta_0$  given knowledge of  $\delta_0$ . To build intuition, we start with the homogenous baseline model studied in Sections 2.4.2 and 2.4.3. In this case,  $F_{\alpha,\gamma}(\cdot, \cdot | \theta_0)$  is degenerate at  $(\alpha_i, \gamma_i) = \theta_0 = (\alpha_0, \gamma_0) \in (0, 1) \times \mathbb{R}_{++}$ . We have already established in Section 2.4.3 that  $\gamma_0$  enters the homogeneous stopping-rule thresholds (20) linearly, scaling all simultaneously, while  $\alpha_0$  enters nonlinearly, shaping the growth rate of the threshold sequence. Moreover, given  $\delta_0$ , 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

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

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

**Proposition 3** (Identification of  $\theta_0$  in the homogeneous model). *Suppose Assumptions 1 to 5 hold,  $J \geq 2$ ,  $F_{\alpha,\gamma}(\cdot, \cdot | \theta_0)$  is degenerate at  $(\alpha_i, \gamma_i) = \theta_0 = (\alpha_0, \gamma_0) \in (0, 1) \times \mathbb{R}_{++}$ , and mean utilities  $\delta_0$  are known. Then,  $\theta_0$  is point-identified by the distribution of optimal portfolio sizes  $\{s_n(\delta_0, \theta_0)\}_{n=0}^J = \{\mathbb{E}[s_n]\}_{n=0}^J$ . In fact, it suffices to observe  $\mathbb{E}[s_0] = \mathbb{P}(n_i = 0)$  and one interior survivor probability  $\mathbb{E}\left[\sum_{k=n}^J s_k\right] = \mathbb{P}(n_i \geq n)$  for any  $n \in \{2, \dots, J\}$  such that  $\mathbb{P}(n_i \geq n) \in (0, 1)$ .*

*Proof.* See Appendix B.6.  $\square$

Proposition 3 emphasizes the role of the extensive margin in identifying  $\theta_0$  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  $\mathbb{E}[s_0]$  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  $\theta_0$  at least locally. This result is formalized in Lemma 2 below. Note that once  $(\alpha_0, \gamma_0)$  is known, the unobserved  $\mathbb{E}[s_0]$  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  $n \in \{1, \dots, J\}$ , let*

$$\mathcal{S}_n(\boldsymbol{\theta}) = \sum_{k=n}^J s_k(\delta_0, \boldsymbol{\theta}), \quad s_n^{>0}(\boldsymbol{\theta}) = \frac{s_n(\delta_0, \boldsymbol{\theta})}{1 - s_0(\delta_0, \boldsymbol{\theta})}, \quad \mathcal{S}_n^{>0}(\boldsymbol{\theta}) = \sum_{k=n}^J s_k^{>0}(\boldsymbol{\theta}) \quad (35)$$

*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  $\mathbb{P}(n_i \geq n)$ , the zero-truncated pmf of portfolio size  $\mathbb{P}(n_i = n | n_i > 0)$ , and*

the corresponding zero-truncated survivor function  $\mathbb{P}(n_i \geq n \mid n_i > 0)$ , respectively, at  $\boldsymbol{\delta}_0$ . At the true  $\boldsymbol{\theta}_0$ , the moment conditions

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

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

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

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

*Proof.* See Appendix B.7. □

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  $(\alpha_i, \gamma_i)$  is allowed. Our comparative-statics discussion in Section 2.4.3 already hints at this: since  $\gamma_i$  enters the stopping-rule thresholds (12) linearly, shifting overall scale, while  $\alpha_i$  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  $\alpha_i$  and  $\gamma_i$ , 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  $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^L$ , 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  $n \in \{0, \dots, J\}$ , it yields at most  $J$  linearly independent moments that can be used to identify  $\boldsymbol{\theta}_0$  as long as  $L = \dim(\boldsymbol{\theta}) \leq J$ .<sup>21</sup> 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  $(\alpha_i, \gamma_i)$  via copula theory.<sup>22</sup> By Sklar's theorem (Sklar, 1973), any  $m$ -dimensional joint cdf  $F(x_1, \dots, x_m)$  with marginals  $\{F_\ell(x_\ell)\}_{\ell=1}^m$  admits a representation of the form

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

where  $C : [0, 1]^m \rightarrow [0, 1]$  is a copula. Moreover, if the marginals are continuous, then this representation is unique in the sense that there is a unique copula  $C(u_1, \dots, u_m)$  that satisfies

---

<sup>21</sup>Proposition 4 imposes the stronger requirement that  $J \geq L + 1$  to deal with zero-truncation as in Lemma 2.

<sup>22</sup>See, e.g., Trivedi and Zimmer (2007) or Fan and Patton (2014) for accessible introductions to copula modeling.

(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  $F_{\alpha,\gamma}(\alpha_i, \gamma_i | \boldsymbol{\theta})$  admits a parametric copula representation

$$F_{\alpha,\gamma}(\alpha_i, \gamma_i | \boldsymbol{\theta}) = C(F_\alpha(\alpha_i | \boldsymbol{\theta}_\alpha), F_\gamma(\gamma_i | \boldsymbol{\theta}_\gamma) | \boldsymbol{\theta}_C), \quad (37)$$

where  $F_\alpha$  and  $F_\gamma$  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

$$s_n(\boldsymbol{\delta}, \boldsymbol{\theta}) = \int_{[0,1]^2} s_{n|\alpha,\gamma}(\boldsymbol{\delta}, F_\alpha^{-1}(u | \boldsymbol{\theta}_\alpha), F_\gamma^{-1}(v | \boldsymbol{\theta}_\gamma)) dC(u, v | \boldsymbol{\theta}_C), \quad (38)$$

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

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

- (a)  $F_{\alpha,\gamma}(\alpha_i, \gamma_i | \boldsymbol{\theta})$  admits the parametric copula representation (37), where the copula function  $C(F_\alpha(\alpha_i | \boldsymbol{\theta}_\alpha), F_\gamma(\gamma_i | \boldsymbol{\theta}_\gamma) | \boldsymbol{\theta}_C)$  has a continuously differentiable density  $c(u, v | \boldsymbol{\theta}_C)$  and its Fisher information matrix is nonsingular at the true  $\boldsymbol{\theta}_{C0}$ .
- (b) The mappings  $\boldsymbol{\theta}_\alpha \mapsto F_\alpha(\cdot | \boldsymbol{\theta}_\alpha)$  and  $\boldsymbol{\theta}_\gamma \mapsto F_\gamma(\cdot | \boldsymbol{\theta}_\gamma)$  are locally injective in a neighborhood of their respective true parameters  $\boldsymbol{\theta}_{\alpha0}$  and  $\boldsymbol{\theta}_{\gamma0}$ . Equivalently, their Fisher information matrices are nonsingular at  $\boldsymbol{\theta}_{\alpha0}$  and  $\boldsymbol{\theta}_{\gamma0}$ , respectively.
- (c) The copula density  $c(u, v | \boldsymbol{\theta}_C)$  and the marginal quantile functions

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

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  $\boldsymbol{\theta}_{C0}$ ,  $\boldsymbol{\theta}_{\alpha0}$ , and  $\boldsymbol{\theta}_{\gamma0}$ , 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,  $J \geq L+1$ ,  $F_{\alpha,\gamma}(\cdot, \cdot | \boldsymbol{\theta}_0)$  admits the parametric copula representation (37) where the marginals  $F_\alpha(\cdot | \boldsymbol{\theta}_\alpha)$  and  $F_\gamma(\cdot | \boldsymbol{\theta}_\gamma)$  and the copula  $C(\cdot, \cdot | \boldsymbol{\theta}_C)$  are continuously differentiable in  $\boldsymbol{\theta}_\alpha$ ,  $\boldsymbol{\theta}_\gamma$ ,  $\boldsymbol{\theta}_C$ , respectively,*

and mean utilities  $\delta_0$  are known. Let  $\mathcal{S}_n(\boldsymbol{\theta})$ ,  $\mathcal{S}_n^{>0}(\boldsymbol{\theta})$ , and  $\mathcal{S}_n^{>0}(\boldsymbol{\theta})$  denote the mappings from  $\boldsymbol{\theta} = (\boldsymbol{\theta}'_\alpha, \boldsymbol{\theta}'_\gamma, \boldsymbol{\theta}'_C)' \in \Theta \subseteq \mathbb{R}^L$  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  $1 \leq n_0 < n_1 < \dots < n_L \leq J$  such that  $\mathcal{S}_{n_\ell}^{>0}(\boldsymbol{\theta}_0) \in (0, 1)$  for all  $\ell \in \{0, \dots, L\}$ . Then, the  $L$  moment equations

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

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

*Proof sketch.* See Appendix B.8.  $\square$

Finally, we cover identification of  $\beta_{w0}$  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  $\mathbf{x}_j$  are observed by the econometrician, but there exist additional job attributes that are unobserved by the researcher with an effect on mean utilities  $\delta_j$  that is fully captured by the unobserved scalar index  $\xi_j$ . Following Berry (1994), we can treat Equation (8) as a linear estimating equation, where  $\xi_j$  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  $\mathbf{x}_j$  absorbs a potential nonzero mean of  $\xi_j$ . We further assume that job characteristics are predetermined at this stage of the recruitment process and (strongly) exogenous in the sense that

$$\mathbb{E}[\xi_j | \{\mathbf{x}_\ell\}_{\ell \in \mathcal{J}}] = 0, \quad \forall j \in \mathcal{J}. \quad (39)$$

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  $\mathbf{x}_\ell$  are excluded from Equation (8) defining  $\delta_j$  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 (square-integrable, measurable) function  $\mathbf{z}(\mathbf{X}_{-j})$ , where  $\mathbf{X}_{-j}$  stacks  $\{\mathbf{x}_\ell\}_{\ell \in \mathcal{B}_j}$ , provides a valid set of instruments (over)identifying  $(\beta_{w0}, \boldsymbol{\beta}_x)'$  through the set of moment equations

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

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

## 3.2 Estimation

Our three-stage estimation strategy follows our sequential identification argument from the previous section, where we show that  $\delta_0$  is identified from conditional inclusion probabilities alone, while, conditional on knowledge of  $\delta_0$ ,  $\theta_0$  and  $\beta_{w0}$  are independently identified from the portfolio-size distribution and exclusion of rival's characteristics from a job's ex-post utility, respectively. Consequently, we start by discussing maximum likelihood estimation of  $\theta_0$  in Section 3.2.1. We then discuss estimation of  $\beta_{w0}$  by linear instrumental variable regression of the first-stage consistent estimates  $\hat{\delta}$  on log-wages and job characteristics in Section 3.2.2. Finally, we discuss a simulated method of moments estimator for  $\theta_0$  that also takes  $\hat{\delta}$  as input in Section 3.2.3. Throughout, we focus on computational feasibility with large samples, especially large choice sets.

### 3.2.1 Estimating $\delta_0$ : Partially rank-ordered logit

Mean utilities  $\delta_0$  can be consistently estimated by maximum likelihood from data mapping a set of applicant unique identifiers into the subset of unique job identifiers of the vacancies they applied to.<sup>23</sup> We can formally think of such data as providing, for each applicant  $i \in \mathcal{I}_1 \equiv \{k \in \mathcal{I} : n_k \in \{1, \dots, J\}\}$ , a partition  $\mathcal{A}_i = \{A_i, \bar{A}_i\}$  of the choice set such that for every  $j \in \mathcal{J}$ ,  $j \in A_i$  if  $i$  applied to  $j$  and  $j \in \bar{A}_i$  otherwise. The underlying (ex-post) linear preference order over  $\mathcal{J}$  represented by the ex-post Bernoulli utilities  $\{u_{ij}\}_{j \in \mathcal{J}}$ ,  $\succ_i$ , induces a partial order  $\succ_i$  on  $\mathcal{A}_i$  defined by  $j \succ_i \ell \iff j \in A_i$  and  $\ell \in \bar{A}_i$ , where  $j, \ell \in \mathcal{J}$ . Naturally,  $\succ_i$  only partially reveals preferences in the sense that we can compare  $j \in A_i$  and  $\ell \in \bar{A}_i$  and infer  $j \succ_i \ell \implies j \succ_i \ell$ , but incompleteness prevents the revelation of preferences among  $j, \ell \in A_i$  or  $j, \ell \in \bar{A}_i$ . Intuitively, we only know that everything that was chosen is preferred to everything that was unchosen, nothing more and nothing less.

Given the partial revelation of preferences from observed choices, the likelihood function we can feasibly work with is the likelihood that a particular partial ordering is observed for each  $i$ . This likelihood is a function of the unknown parameters  $\delta$  given the observed data  $\{\mathcal{A}_i\}_{i \in \mathcal{I}_1}$ . The i.i.d. assumption on  $\{\varepsilon_i\}_{i \in \mathcal{I}_1}$  —where  $\varepsilon_i \in \mathbb{R}^J$  suitably stacks  $\{\varepsilon_{ij}\}_{j \in \mathcal{J}}$ — allows us to write the corresponding log-likelihood function as the sum of the individual log-likelihood contributions over applicants  $i \in \mathcal{I}_1$ ,

$$\mathcal{L}(\delta | \{\mathcal{A}_i\}_{i \in \mathcal{I}_1}) = \sum_{i \in \mathcal{I}_1} \ln \left( f_i(\delta | \mathcal{A}_i) \right), \quad (41)$$

where  $f_i(\delta | \mathcal{A}_i)$  is the likelihood contribution of applicant  $i$  with observed data  $\mathcal{A}_i$ . Hence, find-

---

<sup>23</sup>Here we use the term applicant instead of job seeker since this analysis is conditional on the observed optimal portfolios, and even if non-applicants are observed, they contribute no identifying variation to the estimation of  $\delta_0$ . That is, analysis here is necessarily conditional on  $n_i > 0$ .

ing an expression for the objective function in the maximum likelihood optimization problem boils down to finding an expression for the individual likelihood contributions  $f_i(\delta | \mathcal{A}_i)$ .

The maximum likelihood estimator for this econometric model, which we call the “*partially rank-ordered logit*” (PROL hereafter) model, has been studied elsewhere in the literature under different names. The individual likelihood contribution can be equivalently written as

$$f_i(\delta | \mathcal{A}_i) = \sum_{p \in \mathcal{Q}(A_i)} \prod_{j=1}^{|A_i|} \frac{\exp \delta_{p_j}}{\sum_{\ell \in \mathcal{J} \setminus \{p_k\}_{k=1}^{j-1}} \exp (\delta_\ell)} \quad (42)$$

$$= 1 + \sum_{k=1}^{|A_i|} (-1)^k \sum_{\sigma \in \mathcal{R}_k(A_i)} \frac{\sum_{\ell \in \bar{A}_i} \exp (\delta_\ell)}{\sum_{j \in \sigma \cup \bar{A}_i} \exp (\delta_j)} \quad (43)$$

$$= \int_0^1 \prod_{j \in A_i} 1 - z^{\exp (\delta_j) / \sum_{\ell \in \bar{A}_i} \exp (\delta_\ell)} dz, \quad (44)$$

where  $\mathcal{Q}(S)$  is the set of all the vectors in  $\mathbb{R}^{|S|}$  that can be formed from permutations of the elements of finite set  $S$ , and  $\mathcal{R}_k(S)$  is the set of all size- $k$  combinations from  $S$  already defined in Section 2.2.<sup>24</sup>

The expression on the right-hand side of (42) is a special case of the likelihood contribution derived by [Allison and Christakis \(1994\)](#) in their treatment of ties in the rank-ordered logit model of [Beggs et al. \(1981\)](#) and [Hausman and Ruud \(1987\)](#), corresponding to two tied ranks: one for  $j \in A_i$  and another for  $\ell \in \bar{A}_i$ .<sup>25</sup> The expression in (43) corresponds to the individual likelihood contribution in the multichoice logit model of [van Ophem et al. \(1999\)](#). [Albagli \(2025\)](#) shows that the multichoice logit model is a special case of the tied rank-ordered logit model in the sense that (42) and (43) are, unsurprisingly, mathematically equivalent. This provides a useful shortcut to implement the multichoice logit, since maximum likelihood routines for the general tied rank-ordered logit are readily available in statistical packages such as Stata. However, the combinatorial sums —over permutations in (42) and over combinations in (43)— can become very computationally intensive, even infeasible, when  $J$  is large.

The expression in Equation (44) is the partial-order likelihood derived by [Roussille and Scuderi \(2025\)](#).<sup>26</sup> It provides a convenient shortcut since the univariate integral can be accurately

<sup>24</sup>We define  $\{p_k\}_{k=1}^0 \equiv \emptyset$  for notational consistency.

<sup>25</sup>This expression is also equivalent to the degenerate case with only one tied rank corresponding to  $j \in A_i$ , where  $\ell \in \bar{A}_i$  are treated as unranked. This provides a useful computational trick by specifying only one tied rank for the chosen alternatives and setting the unchosen alternatives as unranked when using pre-programmed routines for the [Allison and Christakis \(1994\)](#) tied rank-ordered logit model. This avoids summing the rank-ordered logit likelihood contribution terms for the lowest rank, which identically sum to one and would require looping over all permutations of the unchosen set. See [Albagli \(2025\)](#) for details.

<sup>26</sup>We provide a full derivation of an equivalent integral representation in Appendix D. To see that our Equa-

approximated by numerical quadrature methods and avoids the cumbersome combinatorics in the alternative closed-form representations (42) and (43). To further reduce the computational burden and ensure tractability with large  $J$ , we maximize (41) indirectly through a minorize-maximize (MM henceforth) algorithm based on a minorization of (41) evaluated at (44) that was originally proposed by [Roussille and Scuderi \(2025\)](#) and has a diagonal Hessian. A detailed derivation of the MM algorithm that closely follows the original derivation in [Roussille and Scuderi \(2025\)](#) and includes a brief introduction to the concept of minorization is presented in Appendix D.

### 3.2.2 Estimating $\beta_{w0}$ : BLP-IV regression

We estimate  $\beta_{w0}$  by linear instrumental variable (IV) regression of the PROL mean utility estimates on log wages and job characteristics. From our brief discussion on the identification of  $\beta_{w0}$  in Section 3.1, it becomes clear that (i) log-wage is likely an endogenous regressor in Equation (8), and (ii) provided the necessary regularity and rank conditions are satisfied, the exclusion restriction on rival jobs' characteristics provides an infinite-dimensional class of valid instrumental variables at the population level: any square-integrable measurable function of rival jobs' characteristics.<sup>27</sup> It is well-known that there exist optimal instruments  $z^*(\mathbf{X}_{-j})$  that minimize the asymptotic variance of the generalized method of moments (GMM hereafter) estimator based on the unconditional moment restrictions (40) ([Hansen, 1985](#); [Chamberlain, 1987](#)). Optimal instruments, however, are hard to compute, so standard practice in the differentiated-product demand estimation literature typically follows [Berry et al. \(1995\)](#) in estimating them with sums of the characteristics of other products (jobs in our context) sold (posted) by (i) the same firm and (ii) other firms. A more recent literature highlights the importance of the quality of the approximation to optimal instruments, particularly in the context of mixed (or random coefficients) logit models of demand with aggregate data ([Armstrong, 2016](#); [Reynaert and Verboven, 2014](#)).

When multiple markets are observed, Hausman instruments ([Hausman, 1997](#); [Hausman et al., 1994](#)) provide an alternative source of identifying variation. These instruments are based on the idea that if firms' marginal costs (marginal product of labor in our context) are correlated across markets but demand (application supply) shocks are not, then average prices of products sold (wages offered) by the same firm in other markets can be used as instruments. See [Azar et al. \(2022\)](#) for an application of Hausman instruments in the context of job applications, where

---

<sup>27</sup>Of course, the number of valid, linearly independent instruments we can construct in a finite sample with fixed  $\{\mathbf{X}_{-j}\}_{j \in \mathcal{J}}$  is finite for any choice of  $z(\cdot)$ .

an argument in favor of this type of instrument is that firms may put weight on reducing within-firm, cross-market wage inequality when setting wages. We only use standard BLP instruments in our empirical application in Section 4 since we observe a single market.

### 3.2.3 Estimating $\theta_0$ : Method of simulated moments

Finally, our identification argument for  $\theta_0$  conditional on knowledge of  $\delta_0$ , based on the observed distribution of portfolio sizes, naturally suggests a GMM approach to estimation. The combinatorial nature of the closed-form expression for the portfolio-size pmf, however, complicates matters when  $J$  is large. We therefore follow a simulated method of moments approach, where an unbiased simulator is used instead of the closed-form expressions for the identifying moments. We focus on moments of the zero-truncated distribution of portfolio sizes in what follows to cover the truncated case, but the same arguments follow through with moments of the full, untruncated distribution. Optimization algorithm details are provided in Appendix E.

Formally, let

$$\mathbf{m}(n_i) = \begin{pmatrix} m_1(n_i) & \cdots & m_M(n_i) \end{pmatrix}', \quad (45)$$

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

$$\mu(\boldsymbol{\theta}_0 | \boldsymbol{\delta}_0) \equiv \mathbb{E} [\mathbf{m}(n_i) | n_i > 0],$$

is a function only of true parameters  $\boldsymbol{\delta}_0$  and  $\boldsymbol{\theta}_0$ .<sup>28</sup> Similarly, the empirical moment

$$\bar{\mathbf{m}}_I(\mathbf{n}) = \frac{\sum_{i \in \mathcal{I}} \mathbb{1}\{n_i > 0\} \mathbf{m}(n_i)}{\sum_{i \in \mathcal{I}} \mathbb{1}\{n_i > 0\}}, \quad (46)$$

depends only on  $\mathbf{n}$ , the data vector of observed portfolio sizes. Given knowledge of true mean utilities  $\boldsymbol{\delta}_0$ , a GMM estimator would minimize the weighted distance between the implied population moment at candidate parameter value  $\boldsymbol{\theta}$  and its sample counterpart,

$$Q_I(\boldsymbol{\theta} | \boldsymbol{\delta}_0, \mathbf{n}) = (\bar{\mathbf{m}}_I(\mathbf{n}) - \mu(\boldsymbol{\theta} | \boldsymbol{\delta}_0))' \mathbf{W} (\bar{\mathbf{m}}_I(\mathbf{n}) - \mu(\boldsymbol{\theta} | \boldsymbol{\delta}_0)), \quad (47)$$

where  $\mathbf{W}$  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 hereafter; [McFadden, 1989](#); [Pakes and Pollard, 1989](#)) provides a convenient alternative where, instead of computing  $\mu(\boldsymbol{\theta} | \boldsymbol{\delta})$  analytically, we construct a frequency simulator  $\hat{\mu}_S(\boldsymbol{\theta} | \boldsymbol{\delta})$  as follows. Given parameter values  $\boldsymbol{\delta}$  and  $\boldsymbol{\theta}$ , we start by taking  $S$  independent draws of (i)  $I \times J$  taste-shock matrix  $\boldsymbol{\varepsilon}^s$

---

<sup>28</sup>Here we use the notation  $\mu(\boldsymbol{\theta} | \boldsymbol{\delta})$  instead of the more standard  $\mu(\boldsymbol{\delta}, \boldsymbol{\theta})$  to emphasize our interest in estimating  $\boldsymbol{\theta}_0$  given knowledge (or a consistent estimate) of  $\boldsymbol{\delta}_0$ .

from  $F_\varepsilon(\cdot)$  and (ii)  $I \times 2$  random-effect matrix  $\begin{bmatrix} \alpha^s & \gamma^s \end{bmatrix}$  from  $F_{\alpha,\gamma}(\cdot, \cdot \mid \theta)$ . We then solve the model for each  $s \in \{1, \dots, S\}$ , obtaining a simulated vector of individual application frequencies  $\mathbf{n}^s = (n_1^s, \dots, n_I^s)'$  that allows us to compute the sample moment in the simulated data and average over simulations:

$$\hat{\mu}_S(\boldsymbol{\theta} \mid \boldsymbol{\delta}) = \frac{1}{S} \sum_{s=1}^S \frac{\sum_{i \in \mathcal{I}} \mathbb{1}_{\{n_i^s > 0\}} \mathbf{m}(n_i^s)}{\sum_{i \in \mathcal{I}} \mathbb{1}_{\{n_i^s > 0\}}}.$$
 (48)

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

$$Q_{I,S}(\boldsymbol{\theta} \mid \boldsymbol{\delta}_0, \mathbf{n}) = (\bar{\mathbf{m}}_I(\mathbf{n}) - \hat{\mu}_S(\boldsymbol{\theta} \mid \boldsymbol{\delta}_0))' \widehat{\mathbf{W}} (\bar{\mathbf{m}}_I(\mathbf{n}) - \hat{\mu}_S(\boldsymbol{\theta} \mid \boldsymbol{\delta}_0)),$$
 (49)

where  $\widehat{\mathbf{W}}$  is a positive-semidefinite, consistent estimate of positive-definite optimal weight matrix (Hansen, 1982)

$$\mathbf{W}^* = \mathbb{E} \left[ (\mathbf{m}(n_i) - \boldsymbol{\mu}(\boldsymbol{\theta}_0 \mid \boldsymbol{\delta}_0)) (\mathbf{m}(n_i) - \boldsymbol{\mu}(\boldsymbol{\theta}_0 \mid \boldsymbol{\delta}_0))' \mid n_i > 0 \right]^{-1},$$
 (50)

and we estimate it non-parametrically via bootstrap (see Appendix E 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  $\hat{\boldsymbol{\delta}}$ , obtained from the MM algorithm for the partially rank-ordered logit discussed in Section 3.2.1 and Appendix D. By standard extremum-estimator arguments, combined with the results on simulation estimators of McFadden (1989) and Pakes and Pollard (1989), the feasible MSM estimator  $\hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\delta}})$  converges in probability to the same limit as the infeasible version  $\hat{\boldsymbol{\theta}}(\boldsymbol{\delta}_0)$  as the sample size  $I \rightarrow \infty$  and either the number of simulations  $S$  is large and fixed or  $S \rightarrow \infty$  at a suitable rate.<sup>29</sup> Intuitively, consistency of  $\hat{\boldsymbol{\delta}}$ , 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  $\hat{\boldsymbol{\delta}}$  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  $\boldsymbol{\delta}$ ,  $\mathbf{n}$ ,  $I$ , and  $S$  from notation now 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 in point-estimation. 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, so it is not pursued in our empirical application in Section 4 given its illustrative purpose.

---

<sup>29</sup>See, e.g., Newey and McFadden (1994) or Ch. 5 of van der Vaart (1998).

### 3.3 From structural parameters to elasticities

With consistent estimates of  $(\delta_0, \theta_0, \beta_{w0})$  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 plug the closed-form solutions for  $s_{j|n}(\delta)$  and  $s_n(\delta, \theta)$ —together with the closed-form derivatives obtained in Appendix B.3—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— $q^s, q^{fs}, s_{j|n}^s, s_n^s$ —are averaged over simulation draws  $s \in \{1, \dots, S\}$ , and finally the resulting simulation estimates are numerically differentiated with respect to  $\delta$  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  $\mathbf{d}^f \in \mathbb{R}^J$  be a column vector with  $\ell$ -th element  $d_\ell^f = \mathbb{1}_{\{\ell \in \mathcal{J}^f\}}$ .<sup>30</sup> Consider the directional derivative of  $q^f(\delta, \theta)$  with respect to  $\delta$  in the direction of  $\mathbf{d}^f$ ,<sup>31</sup>

$$\left. \frac{dq^f(\delta + c \mathbf{d}^f, \theta)}{dc} \right|_{c=0} = \sum_{\ell \in \mathcal{J}^f} \frac{\partial q^f(\delta, \theta)}{\partial \delta_\ell}. \quad (51)$$

The fact that it is equivalent to the sum of the elements of the gradient of  $q^f(\delta, \theta)$  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

$$\eta_{q^f, w^f}(\delta, \theta, \beta_w) = \frac{1}{q^f(\delta, \theta)} \left( \left. \frac{dq^f(\delta + c \mathbf{d}^f, \theta)}{dc} \right|_{c=0} \right) \beta_w. \quad (52)$$

See Appendix B.9 for a derivation of these results.

---

<sup>30</sup>Recall we set  $\mathcal{J} = \{1, \dots, J\}$  for notational convenience in Section 2.2. Alternatively, suitably modify the notation to work with a bijection from  $\{1, \dots, J\}$  to a fixed permutation of  $\mathcal{J}$  as in Footnote 12.

<sup>31</sup>Here we define the directional derivative of  $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$  at  $\mathbf{x}^0 \in U$  in the direction of vector  $\mathbf{d} \in \mathbb{R}^n$  as

$$\nabla_{\mathbf{d}} f(\mathbf{x}^0) \equiv \lim_{h \rightarrow 0} \frac{f(\mathbf{x}^0 + h \mathbf{d}) - f(\mathbf{x}^0)}{h} = \lim_{h \rightarrow 0} \left. \frac{f(\mathbf{x}^0 + (c+h) \mathbf{d}) - f(\mathbf{x}^0 + c \mathbf{d})}{h} \right|_{c=0} \equiv \left. \frac{df(\mathbf{x}^0 + c \mathbf{d})}{dc} \right|_{c=0}.$$

We do not follow the standard convention of normalizing by the norm of direction vector  $\mathbf{d}$  since we are interested in a fixed-scale direction corresponding to an increase of  $\beta_w$  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  $\{w_\ell\}_{\ell \in \mathcal{J}^f}$ .

Equation (52) 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 the firm-level elasticities for the full sample by plugging the simulated quantities into (52), 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 (52) as a sanity check.<sup>32</sup>

## 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, (conditional) 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 Section 4.3. Finally, some empirical limitations qualifying the interpretation of results are discussed in Section 4.4.

### 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.<sup>33</sup> 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 has operated under stable institutional rules throughout the sample period: access is free for job seekers, while firms purchase ad packs or subscription

---

<sup>32</sup>Empirical implementation of the vacancy-level elasticity decomposition is currently work in progress.

<sup>33</sup>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.

plans to post job advertisements.<sup>34</sup> 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 choose whether to display this wage to job seekers, the information is always recorded internally in the dataset. Similarly, applicants must enter an expected wage when 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 the likelihood of high-quality 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.

**Available information.** 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, linking job seekers to ads and recording the date of each application. 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 aged 25 to 55, residing in Chile, and declaring salary expectations between CLP \$150,000 and CLP \$5,000,000. The sampling window is defined as follows. Start with job seekers whose most recent CV update was submitted between 01 January, 2017 and 31 December, 2018. Consider all their applications up to 365 before their last CV update. Define the first search spell of each individual as starting the first date they are observed applying in this sample. Segment spells when two consecutive applications are more than 90 days apart. Finally, retain only spells fully contained in the first quarter of calendar year 2018.

---

<sup>34</sup>See [advertisement packs](#) and [corporate subscription plans](#) for details on current options.

As explained in more detail in Section 4.2 below, we also estimate a single-application nested logit model in line with [Azar et al. \(2022\)](#)—although very simplified—for benchmarking purposes. Only for this benchmark nested logit, we segment the job board into markets—defined by occupation–region bins; see Appendix F for details on the occupational classification—and weeks. We keep applications in markets that have at least two posted vacancies every week. The resulting estimation sample contains 17,357 job seekers, 8,827 jobs, and 1,168 firms across 55 occupation–region groups.

**Data cleaning.** Firm identifiers can be fragmented when the same employer registers under slightly different names in the sense that each name gets a different employer ID in the raw data. In addition, a subset of ads are posted by recruiting agencies on behalf of client firms, and others are reposted by universities or public employment offices, which obfuscates the identity of the ultimate employer. We address these issues by flagging likely recruiting agencies following the heuristic of [Banfi and Villena-Roldán \(2019\)](#), and partially merging firm identifiers with administrative records from the Chilean tax authority. While these procedures mitigate the most salient inconsistencies, some residual noise remains in employer identity. Further details are provided in Appendix F, while complete variable descriptions are available in the [Online Appendix](#).

## 4.2 Estimation details

In this section, we document our empirical-design choices, including market definition, specification of the joint distribution of the random effects, choice of targeted moments in MSM, choice of included and excluded instruments, and two benchmark single-application models.

**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 multi-application 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 the random effects joint distribution.** For this empirical application we assume  $(\alpha_i, \gamma_i)$  are i.i.d. across job seekers, but mutually dependent through a Gaussian copula with a scalar correlation parameter  $\theta_5 \in (-1, 1)$  and marginal distributions

$$\alpha_i \stackrel{\text{iid}}{\sim} \text{Beta}\left(\theta_1 \theta_2, (1 - \theta_1) \theta_2\right), \quad \gamma_i \stackrel{\text{iid}}{\sim} \text{Gamma}\left(\theta_4, \theta_3 / \theta_4\right),$$

and independently of the i.i.d. EV<sub>1</sub> taste shocks  $\varepsilon_i$ . Here,  $\theta_1 \in (0, 1)$  and  $\theta_2 > 0$  are the mean and concentration of  $\alpha_i$ —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  $\gamma_i$ —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 extremes of the portfolio-size distribution.

We also estimate alternative, increasingly more restrictive specifications where:

- (i) The marginal distributions of  $\alpha_i$  and  $\gamma_i$  are also Beta and Gamma, respectively, but they are mutually independent. Here,  $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)'$  retain the same interpretation as in the main specification, but there is no Gaussian-copula correlation parameter.
- (ii) Mutually independent

$$\alpha_i \stackrel{\text{iid}}{\sim} \text{Beta}\left(\theta_1 \theta_2, (1 - \theta_1) \theta_2\right), \quad \gamma_i \stackrel{\text{iid}}{\sim} \text{Exponential}\left(\theta_3\right),$$

where  $\theta_3$  is the rate parameter of the exponential distribution, so  $\mathbb{E}[\gamma_i] = 1/\theta_3$  and  $\text{Var}(\gamma_i) = 1/(\theta_3)^2$ . Here,  $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)'$ , and the first two components retain the same meaning as in the previous more flexible specifications. Note that this specification is nested within the previous one by setting  $\theta_4 = 1$  in the Gamma distribution and reparameterizing  $\theta_3$  to its reciprocal. This special case of the Gamma distribution is fairly more restrictive since it imposes a constant coefficient of variation equal to one.<sup>35</sup>

- (iii) Homogeneous  $\alpha_i = \alpha$  and  $\gamma_i = \gamma$ . Here,  $\boldsymbol{\theta} = (\theta_1, \theta_2)' = (\alpha, \gamma)'$  and, as noted in Section 2.4.3, the model is incapable of generating a portfolio size distribution with significant probability mass at low and high  $n$  at the same time.

---

<sup>35</sup>We also consider alternative mutually independent Beta–Lognormal and Uniform–Uniform heterogeneity specifications. Results are available upon request. The best fit is obtained under our preferred Beta–Gamma specification.

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

$$m_\ell^{\text{pmf}}(n_i) = \mathbb{1} \{n_i = \kappa_\ell^{\text{pmf}}\}, \quad m_\ell^{\text{bin}}(n_i) = \mathbb{1} \{\kappa_\ell^{\text{bin}} \leq n_i \leq \nu_\ell^{\text{bin}}\}, \quad m_\ell^{\text{surv}}(n_i) = \mathbb{1} \{n_i \geq \kappa_\ell^{\text{surv}}\},$$

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:

$$\mathbb{E}[m_\ell^{\text{pmf}}(n_i) | n_i > 0] = \mathbb{P}(n_i = \kappa_\ell^{\text{pmf}} | n_i > 0) \quad (53)$$

$$\mathbb{E}[m_\ell^{\text{bin}}(n_i) | n_i > 0] = \mathbb{P}(\kappa_\ell^{\text{bin}} \leq n_i \leq \nu_\ell^{\text{bin}} | n_i > 0) \quad (54)$$

$$\mathbb{E}[m_\ell^{\text{surv}}(n_i) | n_i > 0] = \mathbb{P}(n_i \geq \kappa_\ell^{\text{surv}} | n_i > 0). \quad (55)$$

Table 1 below summarizes our chosen moments. Alternative results under a different set of moments that also includes binned probabilities are presented in Appendix G. With six moments, even our most flexible specification with five parameters is overidentified.

**Table 1.** MSM targeted moments

$\ell$	$\kappa_\ell$	$\nu_\ell$	Type	$\mathbb{E}[m_\ell(n_i)   n_i > 0]$
1	1		pmf	$\mathbb{P}(n_i = 1   n_i > 0)$
2	3		pmf	$\mathbb{P}(n_i = 3   n_i > 0)$
3	5		pmf	$\mathbb{P}(n_i = 5   n_i > 0)$
4	7		pmf	$\mathbb{P}(n_i = 7   n_i > 0)$
5	10		survivor	$\mathbb{P}(n_i \geq 10   n_i > 0)$
6	14		survivor	$\mathbb{P}(n_i \geq 14   n_i > 0)$

*Notes:* Description of the  $M = 6$  targeted moments chosen to define the MSM criterion function from the set of moment candidates spanned by Equations (53) to (55).

**Choice of included job attributes.** The vector of job characteristics we include as regressors in our BLP-IV regressions of mean utilities on log-wages and job attributes comprises (1) the number of vacancies posted in the job ad (it is sometimes more than one); (2) required experience in years (can be zero); and indicators for (3) salary disclosure; (4) paid ad; (5) permanent contract (so the excluded category is any other, non-permanent type of contract); (6) full-time job (the work arrangement variable has 6 different categories: full-time, part-time, commission-based, shifts, apprenticeship, temporary replacement); (7) part-time job; (8) high education

required (university or higher); (9) low or no education required (high school or less, so the excluded education category corresponds to vocational education); (10) high computer skills requirement; (11) no computer skills required (so the base category is low computer skills); (12) large firm (more than 1,000 employees); (13) small firm (1–50 employees; the excluded category groups employers with 51–1,000 employees). These, together with the constant, act as included instruments and as the basis for constructing our instrumental variables from rival job characteristics. All regressions also control for occupation  $\times$  region and (first meaningful word of) job title fixed effects, which also act as included instruments but are not used in the construction of excluded instruments.

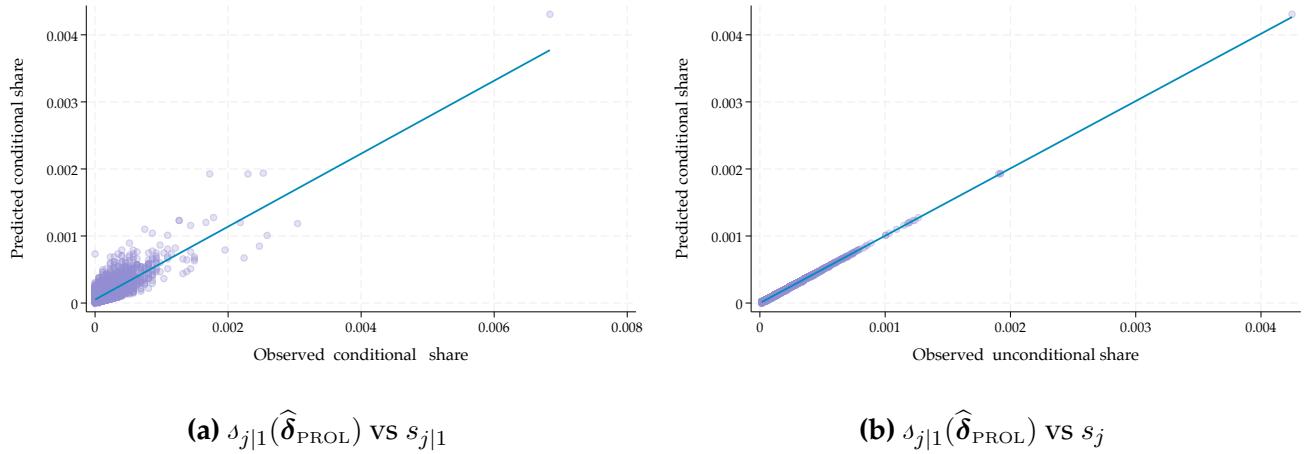
**Choice of excluded instruments.** Given our single-market definition, we only use standard BLP-style instruments. Moreover, the “leave-one-out, same-firm” and “rival-firm” instruments are highly collinear in the sample, possibly due to the high proportion of single-job employers (around 37%, with a median of 2 jobs per employer). Therefore, we only use the rival-firm BLP instruments, defined as the sum of job characteristics over jobs posted by other firms.

**Benchmark single-application model.** As briefly mentioned in the discussion of sample construction in Section 4.1, we also estimate a single-application nested logit (NL hereafter) model as a benchmark. See [Azar et al. \(2022\)](#) for details on the single-application NL model. Our NL specification is a simplification of [Azar et al. \(2022\)](#) in the sense that we include only one applicant-level variable in the top-level logit: a dummy indicating whether the applicant lives in the same region of the job ad. Only for this model, we segment markets based on occupation–region bins. We allow the nesting parameter to vary by market by interacting it with market fixed effects in the top-level logit. We force a single-application sample by treating each application by the same applicant as if submitted by different applicants. This is achieved by simply recoding the applicant ID into the observation number in the applicant–job applications dataset, and thus effectively treating the ranking within  $A_i$  as the sum over every ranking where each  $j \in A_i$  is ranked first and every other  $k \in A_i$  is (partially) ranked second together with every  $\ell \in \bar{A}_i$ . While the BLP framework would prescribe instruments defined at the region  $\times$  occupation  $\times$  week level to be consistent with the market segmentation in this NL model, we use exactly the same instruments as in our full, multi-application model —where the whole job board is treated as a single market— for comparability. Finally, we also estimate a standard multinomial logit model under the same remapping of applicant IDs, enforcing a single-application sample. We present this model’s BLP-IV results for comparison, but we do not pursue plug-in elasticity computation since it would require an estimate of application costs (the job-offer probability can be thought assumed equal to one here), which is not available without data on the extensive margin (zero versus one application).

### 4.3 Empirical results

This section proceeds in four parts. We first estimate mean utilities and assess the fit of the PROL model. We then estimate the impact of log-wages on 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 and compare the elasticities implied by our full, multi-application model to those implied by the single-application NL model.

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



*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  $\text{Corr}(s_{j|1}(\widehat{\delta}_{\text{PROL}}), s_{j|1}) = 0.8044$  in Panel (a) and  $\text{Corr}(s_{j|1}(\widehat{\delta}_{\text{PROL}}), s_j) = 0.9997$  in Panel (b). We predict only the application share conditional on a portfolio size of one because of its straightforward computation compared to higher-order inclusion probabilities that involve combinatorial sums.

Since the mean utility vector is very high-dimensional and  $\delta$  is identified up to location, it is impractical to present detailed estimates. Instead, we show two scatter plots in Figure 4 above. Panel (a) displays the scatter plot of the model-predicted estimates of each job's application share conditional on one application,  $s_{j|1}(\widehat{\delta}_{\text{PROL}})$ , against its sample counterpart  $s_{j|1}$ . We choose this moment since its computation is trivial: it does not involve combinatorics and is simply equal to the standard MNL share (without an outside option). The predicted and observed quantities display a reasonably high correlation of around 0.8, suggesting a good fit of the PROL maximum likelihood estimator (MLE from now on). Panel (b) is included only as a reassurance on convergence of the MM algorithm used to find the MLE. It displays the scatter plot of the same model-predicted conditional share against the unconditional application share  $s_j$ , with a nearly perfect fit (correlation equal to 0.9997). A poor correlation here would be a sign of non-convergence of the numerical optimization algorithm.

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

	PROL		MNL		NL	
	OLS (1)	IV (2)	OLS (3)	IV (4)	OLS (5)	IV (6)
Log monthly salary (CLP)	0.016 (0.026)	0.363*** (0.100)	0.002 (0.022)	0.302*** (0.088)	0.005 (0.016)	0.429*** (0.051)
Salary disclosure	-0.042 (0.037)	-0.031 (0.037)	-0.007 (0.025)	0.002 (0.025)	0.031* (0.019)	0.044** (0.019)
Number of vacancies	0.005*** (0.002)	0.006*** (0.002)	0.004*** (0.002)	0.005*** (0.002)	0.000 (0.001)	0.001 (0.001)
Required experience	-0.054*** (0.007)	-0.081*** (0.011)	-0.054*** (0.007)	-0.078*** (0.010)	-0.021*** (0.005)	-0.055*** (0.007)
Paid advertisement	0.264*** (0.047)	0.221*** (0.049)	0.286*** (0.044)	0.250*** (0.045)	0.034 (0.034)	-0.017 (0.036)
Permanent contract	0.072*** (0.021)	0.051** (0.022)	0.057*** (0.019)	0.039* (0.020)	-0.005 (0.014)	-0.030** (0.014)
Full-time job	0.067 (0.042)	0.057 (0.042)	0.051 (0.031)	0.042 (0.032)	0.039* (0.021)	0.027 (0.022)
Part-time job	0.063 (0.056)	0.214*** (0.072)	0.034 (0.050)	0.165*** (0.062)	0.082** (0.034)	0.268*** (0.040)
High education requirement ( $\geq$ university)	0.049* (0.029)	-0.086* (0.047)	0.043 (0.027)	-0.074* (0.043)	-0.000 (0.019)	-0.166*** (0.028)
Low education requirement ( $\leq$ high school)	0.060** (0.030)	0.094*** (0.031)	0.061** (0.026)	0.091*** (0.028)	0.062*** (0.018)	0.104*** (0.020)
High computer skills requirement	-0.078** (0.033)	-0.095*** (0.034)	-0.038 (0.023)	-0.053** (0.024)	-0.038** (0.016)	-0.058*** (0.017)
No computer skills required	0.028 (0.028)	0.066** (0.031)	0.008 (0.026)	0.041 (0.028)	-0.037** (0.018)	0.009 (0.020)
Large firm (>1,000 emp.)	0.150*** (0.024)	0.144*** (0.024)	0.151*** (0.023)	0.145*** (0.023)	0.104*** (0.016)	0.096*** (0.017)
Small firm (1–150 emp.)	-0.059** (0.027)	-0.075*** (0.028)	-0.034* (0.020)	-0.047** (0.020)	0.013 (0.014)	-0.007 (0.014)
Kleibergen-Paap			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 collinearity 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 2 above presents the results of our ordinary least squares (OLS) and two-stage least squares (2SLS) BLP-IV linear regressions of estimated mean utilities on  $\ln(w_j)$  and  $x_j$ , where the dependent variable is  $\widehat{\delta}_{\text{PROL}}$  in Columns (1) and (2),  $\widehat{\delta}_{\text{MNL}}$  in Columns (3) and (4), and  $\widehat{\delta}_{\text{NL}}$  in Columns (5) and (6). Eicker-Huber-White robust standard errors are reported for each regression, and the Kleibergen and Paap (2006) weak-IV statistic—an heteroskedasticity-robust version of the Cragg and Donald (1993) statistic used in the Stock and Yogo (2005) test for weak instruments—for the 2SLS regressions. Unsurprisingly, since they share the same first stage, the Kleibergen–Paap statistics coincide across the IV Columns (2), (4) and (6), with a value of 88.83 that is high enough to reject the standard weak-IV null hypotheses if compared against the Stock–Yogo critical values.

Results consistently show statistically insignificant (at conventional levels) OLS estimates of the effect of log-wage on location-normalized mean utility, while the IV estimates fluctuate in the 0.3–0.43 range and significantly differ from zero at the 1% level. Among the job attributes that display consistent patterns across all columns, we find that higher required experience and computer skills diminish applicants' systematic valuation of a job. Jobs requiring a high-school or lower educational level appear more attractive relative to those requiring vocational education. Similarly, part-time jobs appear more valuable compared to other less standard work arrangements, as do larger employers against medium-sized ones. Some attributes share sign and magnitude patterns between the PROL and MNL specifications but differ in the NL specification. These include a negative but statistically insignificant PROL/MNL effect of salary disclosure while positive and significant at 10–5% in NL; a positive PROL/MNL but null or negative NL impact of the number of vacancies, paid advertisements, and permanent contracts; and a negative PROL/MNL but null NL effect of small- relative to medium-sized employers.

Overall, the fact that we obtain similar estimates of  $\beta_w$  across multi- and single-application models is reassuring and consistent with our identification results: the three models—PROL, MNL, and NL—exploit roughly the same identifying variation thanks to the (partial) revelation of preferences. However, a key result of our theoretical discussion is that, even if mean utilities can be identified and consistently estimated in single-application models, the post-estimation mapping from structural-parameter estimates to application wage elasticities is misspecified since single-application models shut down the portfolio-size mechanism by construction. This is exactly what our MSM and elasticity estimates below suggest.

The MSM estimates in Panel (a) of Table 3 below reveal substantial heterogeneity in both job-offer probabilities and application costs across job seekers. Our preferred Beta–Gamma copula specification in Column (4) estimates a mean job-offer probability of 0.54, suggesting that the average applicant faces roughly even odds of receiving an offer conditional on applying. The implied variance of 0.025 indicates meaningful dispersion: some job seekers face considerably

better prospects than others, even when applying to the same jobs. The application cost estimates show even greater heterogeneity. The mean marginal cost of 9.97 (in units of mean utility) masks substantial variation, with an estimated variance of 159. This dispersion suggests that while some job seekers face negligible application costs, others encounter substantial barriers, whether through time constraints, information costs, or other sources of application costs.

**Table 3.** Estimates of the random-effects distribution parameters

Panel (a)	Independent				Copula
	Homogeneous (1)	Beta-Exponential		Beta-Gamma	Beta-Gamma (4)
		(2)	(3)		
<b>Estimates</b>					
$\theta_1$	0.7482		0.1865	0.1742	0.5383
$\theta_2$	0.0968		3.4717	5.1058	8.8311
$\theta_3$			0.0954	48.6479	9.9681
$\theta_4$				1.0290	0.6245
$\theta_5$					0.8536
Objective function	5,317.0888		21.2028	5.4828	1.0134
<b>Implied marginals</b>					
$\mathbb{E}[\alpha_i]$	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 (1)	Simulated (2)	Contrib. (3)	Simulated (4)	Contrib. (5)	Simulated (6)	Contrib. (7)	Simulated (8)	Contrib. (9)
<b>Moments</b>									
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  $\theta$  of the joint distribution  $F_{\alpha,\gamma}(\cdot, \cdot | \theta)$  of the random effects  $(\alpha_i, \gamma_i)$ . 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 estimated Gaussian copula correlation of 0.85 indicates strong comonotone dependence between job-offer probabilities and application costs. This pattern suggests that job seekers

who face higher application costs also tend to face better job-offer prospects, while those with lower application costs tend to have worse offer probabilities. This positive correlation may reflect underlying heterogeneity within the unemployed population. For instance, higher-skilled workers might face higher opportunity costs of time or higher psychic costs of applying to positions below their preferred level (high  $\gamma_i$ ), while simultaneously being able to attract more offers due to their qualifications (high  $\alpha_i$ ). Alternatively, workers with caregiving responsibilities or geographic constraints may face higher effective application costs while representing a positively selected group conditional on searching. Conversely, workers experiencing longer unemployment spells might have more available time for applications (low  $\gamma_i$ ) but face lower offer probabilities (low  $\alpha_i$ ) due to duration dependence or skill depreciation.

Comparing across specifications, the Beta–Gamma copula model achieves the lowest objective function value, substantially outperforming both the independent Beta-Gamma and the more restrictive independent Beta-Exponential specifications. Heuristically contrasting—since our estimator is based on simulation—the implied Hansen overidentification statistics in Panel (a) against the critical values of the  $\chi^2$  distribution with  $M - L$  degrees of freedom, we would (i) strongly reject in Column (1):  $Q = 5,317 > \chi^2_4(0.99) = 13.277$ ; (ii) reject in Column (2):  $Q = 21.203 > \chi^2_3(0.99) = 11.345$ ; (iii) not reject at 5% but reject at 10% in Column (3):  $\chi^2_2(0.90) = 6.251 < Q = 5.483 < \chi^2_2(0.95) = 7.815$ ; and (iv) not reject at conventional levels in Column (4):  $Q = 1.01 < \chi^2_1(0.90) = 2.706$ .<sup>36</sup>

Panel (b) demonstrates how this superior fit stems from better matching of the targeted moments, particularly at the extremes of the portfolio-size distribution. Column (1) shows the empirical moments. Columns (2), (4), (6), and (8) display the simulated moments averaged across the  $S = 50$  simulation draws used in estimation for each specification. Columns (3), (5), (7), and (9) report each moment's normalized contribution to the overall MSM objective function, measured as the (signed) proportion of the total weighted misfit attributed to that moment, taking into account the full covariance weighting from our bootstrap estimate of the optimal weight matrix.<sup>37</sup> As predicted by our theoretical discussion in previous sections, the homogeneous model is incapable of putting significant mass at low and high portfolio sizes simultaneously, as illustrated by the zero mass it puts on the moments associated with higher  $n$  while trying to match the empirical mass at the lower- $n$  moments. As we move from more

---

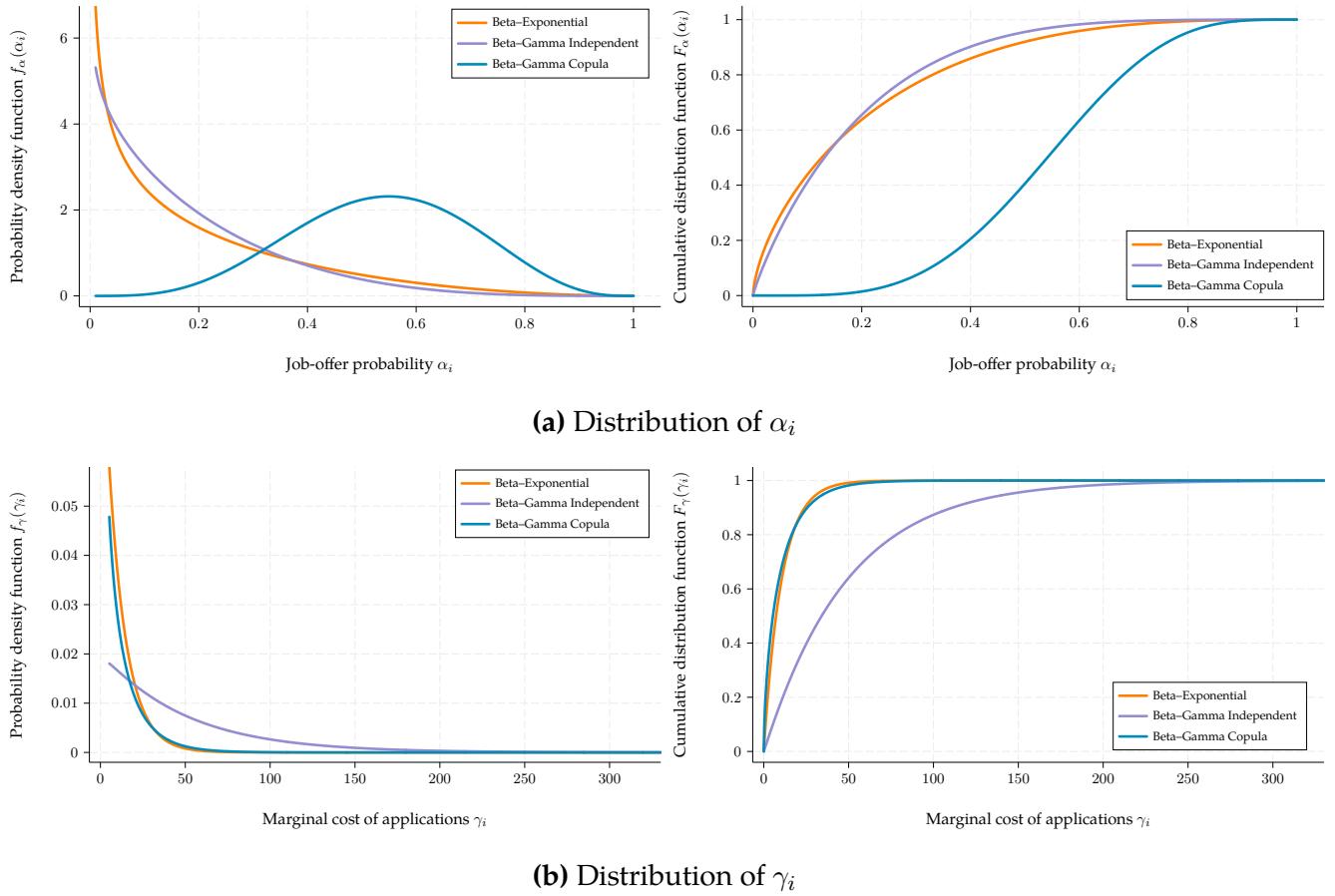
<sup>36</sup>Here, the overidentification statistic is the optimized objective function value without rescaling by sample size since (the inverse of) our bootstrap estimate of the optimal weight matrix actually estimates the *variance of the sample mean* of the individual moment functions, inversely scaling by the truncated sample size. See Appendix E for details.

<sup>37</sup>Specifically, the MSM criterion at the estimates is  $Q(\hat{\theta}) = \mathbf{d}' \widehat{\mathbf{W}} \mathbf{d}$ , where  $\mathbf{d} = \bar{\mathbf{m}}_I(n) - \widehat{\mathbf{m}}_S(\hat{\theta})$ . The signed proportional contribution of moment  $\ell \in \{1, \dots, M\}$  is computed as  $c_\ell = d_\ell(\widehat{\mathbf{W}} \mathbf{d})_\ell / Q(\hat{\theta})$ , where subscript  $\ell$  denotes the  $\ell$ -th component of the vector. Because  $\widehat{\mathbf{W}}$  is non-diagonal, some  $c_\ell$  may be negative even though the total  $Q(\hat{\theta})$  is positive.

restrictive towards more flexible heterogeneity specifications, the model's ability to match the empirical moments increases monotonically.

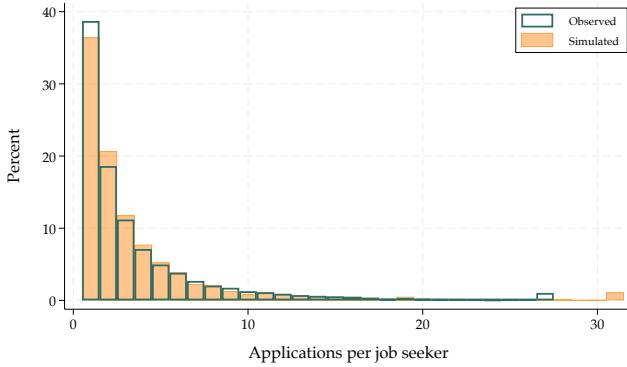
In particular, relaxing the constant (and equal to one) assumption on the coefficient of variation of  $\gamma_i$  when moving from the independent Beta–Exponential to the independent Beta–Gamma specification pushes the estimator towards a higher-mean and extremely more dispersed marginal distribution of  $\gamma_i$  in order to better match the high mass at very low  $n$  and the long right tail. Introducing dependence through the Gaussian copula specification allows the model to push mass towards the extremes without requiring huge dispersion in application costs by moderately increasing the mean and dispersion of job-offer probabilities. These distributional shifts are depicted in Figure 5 below, where we plot the pdf and cdf of  $\alpha_i$  and  $\gamma_i$  implied by each of our heterogeneous sets of estimates. It is clearly visible how the pdf of  $\alpha_i$  becomes more bell-shaped and concentrated around its mode, while the distribution of  $\gamma_i$  goes back closer to the Beta–Exponential case.

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

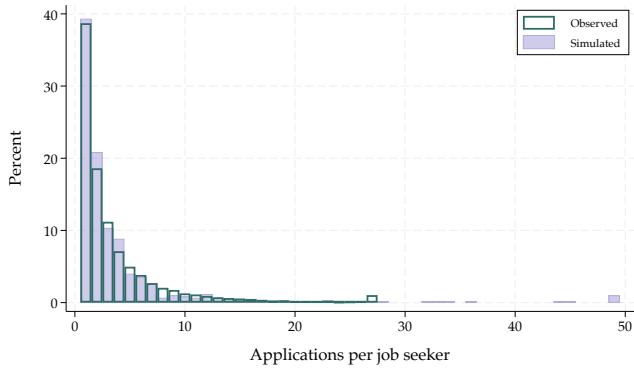


*Notes:* Marginal distributions of the random effects  $(\alpha_i, \gamma_i)$  implied by the MSM estimates of  $\theta$  in Table 3 for the nondegenerate specifications (i) independent Beta–Exponential, (ii) independent Beta–Gamma, and Beta–Gamma copula. Supports are visually truncated at  $\alpha_i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

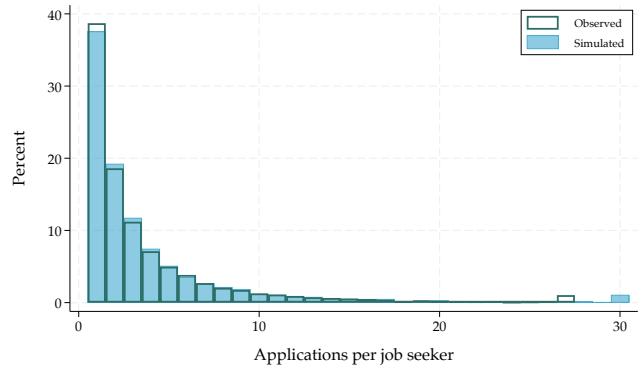
**Figure 6.** Assessing the fit of the structural parameter estimates



**(a) Beta-Exponential**



**(b) Beta-Gamma Independent**



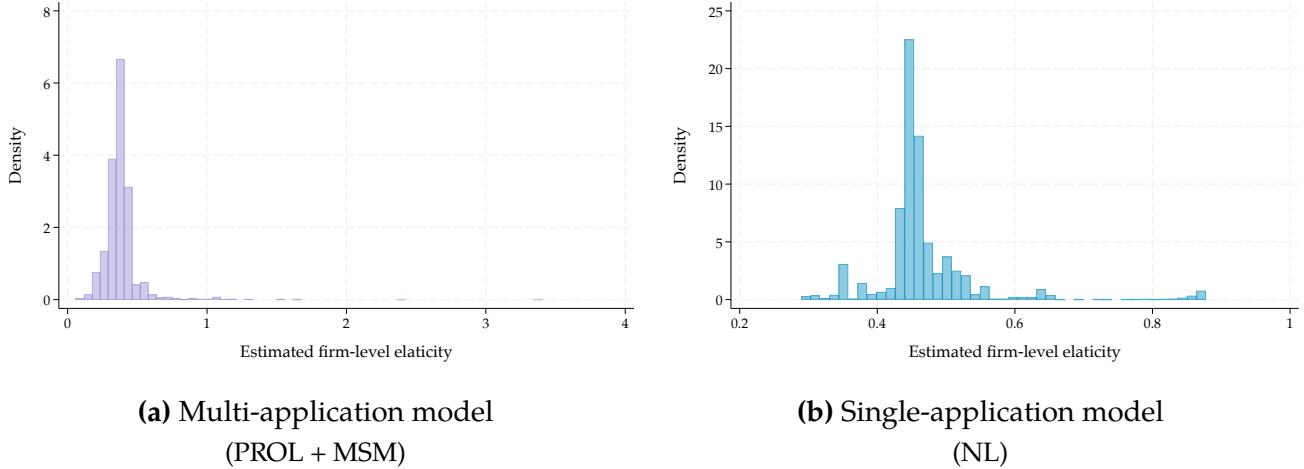
**(c) Beta-Gamma Copula**

*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  $\delta$  from Figure 4 and the MSM estimates of  $\theta$  from Table 3. Empirical supports are censored at the corresponding 99-th percentile. This figure, showing the full zero-truncated pmf of portfolio sizes, mixes targeted ununtargeted moments by our MSM estimator, partially serving as a model fit diagnostic.

To assess how well our  $\delta$  and  $\theta$  estimates jointly reproduce the empirical distribution of portfolio sizes, Figure 6 above compares the observed and simulated zero-truncated portfolio-size pmf for a single simulation draw. The model's ability to replicate the empirical portfolio-size distribution provides a simple heuristic validation device for our structural approach, particularly since the full (zero-truncated) pmf includes both moments that are targeted by our MSM estimator and other moments that are not. All three heterogeneous specifications successfully capture the key features of the data: the modal concentration at low application counts, the declining probability mass at intermediate values, and the presence of a right tail representing intensive searchers. The Beta-Gamma copula specification in Panel (c) achieves the closest match to the empirical distribution, particularly in replicating both the peak at small portfolio sizes and the gradual decline through intermediate values. Single-application

models completely ignore this rich source of variation in application elasticities. Our final set of results directly compares our model's distributional implications for firm-level elasticities against the single-application NL benchmark, illustrating how the behavioral heterogeneity we have documented thus far translates into heterogeneity in labor market power.

**Figure 7.** Distribution of the firm-level elasticity estimates



*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. \(2022\)](#).

Figure 7 above shows the distributions of the firm-level elasticity estimates from our multi-application model and the single-application NL benchmark model, while Table 4 summarizes the corresponding descriptive statistics. Both models yield median application elasticities significantly below the range of estimates in the literature. If we translate them into labor supply elasticities following the heuristic of [Azar et al. \(2022\)](#)—which consists of multiplying by two—the multi-application model implies a median firm-level labor supply elasticity of 0.73, while the single-application benchmark implies a median elasticity of 0.9, both implying over 100% median markdowns of wages below the marginal revenue product of labor.<sup>38</sup>

While such degree of monopsony power may seem too large to be realistic—and we discuss in Sections 2.4.4 and 4.4 several limitations that suggest our estimates should not be taken as a definitive measure of monopsony power—a point made by [Langella and Manning \(2021\)](#) in the context of estimating how separation elasticities vary with wages, the job offer arrival rate, and

<sup>38</sup>The wage markdown (or exploitation rate) is obtained from the Lerner condition-style ([Lerner, 1934](#)) equation  $(\text{MRPL} - w)/w = 1/\eta$ , which can be derived from the first order condition of a static partial equilibrium model of monopsony ([Manning, 2003](#), Ch. 2).

over time, similarly applies here. More than estimating the absolute level of monopsony power, the purpose of our empirical application lies in illustrating how single-application models miss a behavioral margin that our multi-application model captures. If the (source of) bias in elasticity levels is in a sense constant across models, then our conclusions about differences across single- and multi-application models remain valid.

A prominent feature of Figure 8 that is corroborated by the descriptive statistics in Table 4 is the markedly larger concentration of the distribution of the NL estimates. The interquartile range (IQR) of the multi-application estimates is about 6 times as large as the IQR of the single-application estimates. Similarly, the ratio of the standard deviations is approximately 3, the ratio of the coefficients of variation is  $0.5381/0.1632 \approx 3.3$ , and the range of the multi-application model's empirical distribution ( $3.4028 - 0.0599 = 3.3429$ ) is about 5.7 times as large as that of the single-application model ( $0.8775 - 0.2893 = 0.5882$ ). While this pattern may be partially driven by certain specification choices—e.g., market segmentation or applicant-specific covariates in the NL model—, these results suggest that accounting for endogenous portfolio choice may reveal firm-level heterogeneity in labor market power that is difficult to capture in single-application models.

**Table 4.** Descriptive statistics of firm-level elasticities

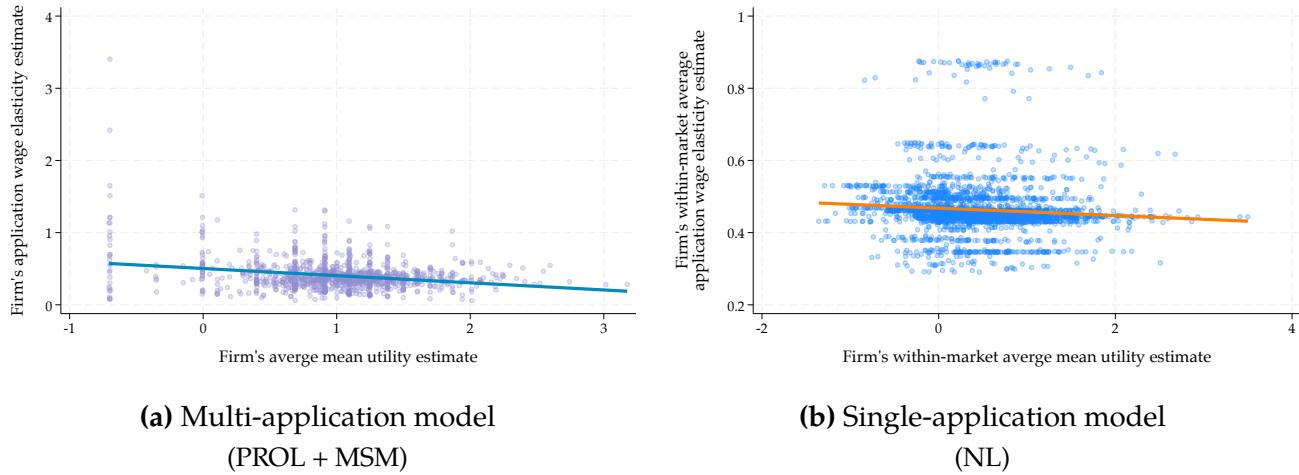
	Mean	Median	IQR	Std. dev.	Min.	Max.	Mode	Obs.
Multi-application model	0.4051	0.3655	0.1665	0.2180	0.0599	3.4028	0.5854	1,168
Single-application model	0.4645	0.4488	0.0277	0.0758	0.2893	0.8775	0.4476	20,950

*Notes:* Descriptive statistics of the estimated firm-level elasticities from the full multi-application model and the benchmark single-application nested logit model. The unit of observation is the firm the multi-application model, and firm by market-week in the nested logit model. The mean utility estimates underlying the computation of elasticities 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 the multi-application model correspond to those in Column (4) of Panel (a) of Table 3. Firm-level elasticities in the multi-application model are computed through the simulation-numerical differentiation procedure for the directional derivatives described in Section 3.3. Elasticities in the single-application model are based on the closed-form expressions in Azar et al. (2022).

This comparison illustrates our central methodological point: single-application models mechanically shut down the portfolio-size mechanism, understating the degree of heterogeneity in the wage-responsiveness of the application supply to the firm. When heterogeneous workers facing different chances of getting a job offer and marginal costs of applications can adjust both which jobs they apply to and how many applications they submit, wage increases generate larger variation in supply responses than models that restrict workers to a single application can predict. Firms with systematically more attractive jobs —i.e., higher mean utilities— will tend to rank near the top of job seekers' ex-post preferences and will thus be generally more likely to be included in applicants' optimal portfolios regardless of wage changes. As an extreme

example, imagine a dominant job with so high (but finite) a systematic utility that it is virtually always included in optimal portfolios. This stellar job is included with probability arbitrarily close to one in optimal portfolios of larger size (and very high even at  $n = 1$ ), so its application supply is largely unresponsive to the portfolio-size mechanism apart from the extensive margin (which shrinks as this job becomes more attractive). For finite but high  $\delta_j$ , wage changes affect applications primarily through substitution, with only a small residual extensive margin effect.<sup>39</sup> Wage changes only affect it through the substitution mechanism: if its wage falls so much that it moves down through the individual rankings of some applicants, it may be excluded from their optimal portfolios. Thus, we expect the application supply to more attractive jobs and the firms that post them to respond less and in a more homogeneous way to wage changes. That is, we expect more market power (lower elasticities) and lower variation in market power (more concentrated distribution slices at higher mean utilities) for more attractive jobs and firms.

**Figure 8.** Correlation of firm-level elasticities and average mean utilities



*Notes:* Scatter plots of the estimated firm-level elasticities against firm's average mean utilities from our full multi-application model in Panel (a), and from the benchmark single-application nested logit model in Panel (b). The unit of observation is the firm in Panel (a), and firm by market 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). The raw correlations are -0.2501 and -0.0869 for the multi- and single-application models, respectively; both statistically different from 0 at the 1% significance level. The coefficients from a regression of elasticities on average mean utilities are -0.0990 and -0.0103, for the multi- and single-application models, respectively; both statistically different from 0 at the 1% significance level using heteroskedasticity-robust standard errors. Panel (a) excludes two outlier firms with very low average mean utilities for visual clarity. Results in terms of raw and partial correlation (i.e., linear regression coefficient) remain virtually unchanged when including these two outliers and are available upon request.

Our model predicts that systematically more attractive employers face less wage-sensitive application supply because they are already near full inclusion. Moreover, the wage-responsiveness

<sup>39</sup> Both the substitution and portfolio-size effects shrink as  $\delta_j$  grows, but the portfolio-size effect becomes second-order for finite large  $\delta_j$ , while the substitution effect vanishes only in the limit as  $\delta_j \rightarrow \infty$ .

of supply among such firms will tend to be less varied since only (or mainly) the substitution effect operates. Figure 8 above explores this relationship by plotting the firm-level elasticity estimates against the firm-level average mean utility of the jobs posted by the firm. The unit of observation is the firm in Panel (a) and firm by market in Panel (b) —i.e., we aggregate over weeks in the case of the NL model. We find that elasticities do tend to concentrate more at lower values as average mean utility increases in the multi-application model. This relationship is not clearly visible, especially in terms of concentration, in the single-application model, although it also exhibits a (weaker) negative correlation between elasticities and systematic desirability. Specifically, the raw correlations are -0.2501 and -0.0869 and the coefficients from a regression of elasticities on average mean utilities are -0.0990 and -0.0103 for the multi- and single-application models, respectively, all statistically significant at the 1% level (using heteroskedasticity-robust standard errors for the regression coefficients).

## 4.4 A brief discussion of empirical limitations

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, and to illustrate our main methodological point that single-application models mechanically ignore the portfolio-size margin and the rich behavioral heterogeneity it introduces. 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, introducing 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 2.4.4, 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. Nonetheless, as in Section 4.1, we note that the evidence in [Choi et al. \(2025\)](#) suggests Trabajando.com's administrative records are reasonably representative of the Chilean labor market.

## 5 Conclusion

This paper addresses a fundamental gap in the recent empirical IO-inspired literature on labor market power: the assumption that job seekers submit at most one application. While this restriction simplifies estimation and fits naturally when modeling final job acceptance decisions, it is inconsistent with observed application behavior and overlooks how uncertainty shapes job search. When workers face uncertainty about which applications will yield offers, they optimally diversify by applying to multiple jobs. This portfolio choice introduces an additional margin through which wages affect the supply of applications, a mechanism that, by construction, single-application models entirely miss.

We show that in any model allowing job seekers to submit multiple applications, the wage elasticity of job applications decomposes into substitution and portfolio-size components. This decomposition follows from the law of total probability and holds generally, independent of the specific parametric structure used for estimation. This clarifies which aspects of application behavior standard discrete choice models of monopsony can and cannot capture: the single-application framework identifies substitution effects but mechanically shuts down portfolio-size adjustments, mismeasuring the responsiveness of labor supply to wages and understating its heterogeneity.

To operationalize this insight, we develop a tractable structural framework that models both margins jointly. Building on the simultaneous search setting of [Chade and Smith \(2006\)](#) and their marginal improvement algorithm, we specify an additive random utility model where workers with idiosyncratic preferences over differentiated jobs face a portfolio choice problem under job-offer uncertainty and costly applications. A nontrivial complication that also implicitly affects single-application discrete choice models is that ex-ante choice of application portfolios does not generally reveal ex-post preferences over jobs. Under simplifying assumptions —worker-specific only job-offer probabilities and constant marginal application costs—, however, the model retains a revealed-preference structure while generating endogenous multi-application behavior and allowing for rich heterogeneity patterns. We derive closed-form expressions for the relevant choice probabilities and develop a multi-stage estimation strategy that overcomes the severe computational challenges posed by the typically large choice sets in job application settings.

Our empirical application to Chilean job board data demonstrates that the framework is implementable at scale and rich enough to replicate observed application patterns. The estimates reveal substantial heterogeneity in both job-offer probabilities and application costs across unemployed job seekers, with strong positive dependence suggesting that better job prospects are associated with higher search frictions and vice versa. Most importantly, the

estimated firm-level application elasticities exhibit considerably greater dispersion than those from a single-application benchmark, consistent with the portfolio-size mechanism varying in strength across firms based on their degree of systematic attractiveness.

The framework developed here necessarily abstracts from several important features of labor markets. The logit structure imposes restrictive substitution patterns —although IIA holds only conditional on portfolio size—, the static model ignores learning and dynamic adjustment, and the analysis focuses on application flows rather than completed employment relationships. The strong assumptions on job-offer probabilities —exogenous and homogeneous across jobs— are difficult to reconcile with realistic firm screening behavior. Relaxing these assumptions, embedding the model in equilibrium, and incorporating employers' selection decisions remain important and exciting directions for future research. Despite the current limitations, the results establish that accounting for multi-application behavior is both feasible and consequential. The portfolio-size margin is not merely a technical complication but a substantive economic mechanism that shapes how labor supply responds to wages. Ignoring this margin risks mischaracterizing the distribution of market power across firms and misunderstanding the sources of monopsony.

We view our empirical application, and this paper in general, as an open invitation to start thinking seriously about how uncertainty and search frictions shape job application behavior and translate into monopsony power. As administrative and online platform data on job applications become increasingly available, frameworks that accommodate the multi-application nature and inherent uncertainty of job search will become essential for credibly estimating labor market power and informing policy.

## 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.
- ALBAGLI, P. (2025): "Estimating Preferences from Coarse Rankings: On the Equivalence of the Multichoice and Tied Rank-Ordered Logit Models," Working Paper.
- ALI, S. N. AND R. I. SHORRER (2025): "Hedging When Applying: Simultaneous Search with Correlation," *American Economic Review*, 115, 571–98.
- ALLISON, P. D. AND N. A. CHRISTAKIS (1994): "Logit Models for Sets of Ranked Items," *Sociological Methodology*, 24, 199–228.
- ARMSTRONG, T. B. (2016): "Large Market Asymptotics for Differentiated Product Demand Estimators With Economic Models of Supply," *Econometrica*, 84, 1961–1980.
- AUTOR, D., D. DORN, L. F. KATZ, C. PATTERSON, AND J. VAN REENEN (2020): "The Fall of the Labor Share and the Rise of Superstar Firms," *The Quarterly Journal of Economics*, 135, 645–709.
- 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.
- BEGGS, S., S. CARDELL, AND J. HAUSMAN (1981): "Assessing the Potential Demand for Electric Cars," *Journal of Econometrics*, 17, 1–19.
- BELOT, M., P. KIRCHER, AND P. MULLER (2019): "Providing Advice to Jobseekers at Low Cost: An Experimental Study on Online Advice," *The Review of Economic Studies*, 86, 1411–1447.

- 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.
- BROOKS, W. J., J. P. KABOSKI, Y. A. LI, AND W. QIAN (2021): "Exploitation of labor? Classical monopsony power and labor's share," *Journal of Development Economics*, 150, 102627.
- 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.
- CALDWELL, S. AND E. OEHLSSEN (2023): "Gender, Outside Options, and Labor Supply: Experimental Evidence from the Gig Economy," Working Paper.
- CARD, D. (2022): "Who Set Your Wage?" *American Economic Review*, 112, 1075–90.
- CARD, D., A. R. CARDOSO, J. HEINING, AND P. KLINE (2018): "Firms and Labor Market Inequality: Evidence and Some Theory," *Journal of Labor Economics*, 36, S13–S70.
- CHADE, H. AND L. SMITH (2006): "Simultaneous Search," *Econometrica*, 74, 1293–1307.
- CHAMBERLAIN, G. (1987): "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics*, 34, 305–334.
- CHOI, S., B. VILLENA-ROLDÁN, AND N. FIGUEROA (2025): "Posted Wage Cyclical: Evidence from High-Quality Vacancy Data," Working paper.
- COLES, P., J. CAWLEY, P. B. LEVINE, M. NIEDERLE, A. E. ROTH, AND J. J. SIEGFRIED (2010): "The Job Market for New Economists: A Market Design Perspective," *Journal of Economic Perspectives*, 24, 187–206.
- CRAGG, J. G. AND S. G. DONALD (1993): "Testing Identifiability and Specification in Instrumental Variable Models," *Econometric Theory*, 9, 222–240.

- 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.
- HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.
- (1985): "A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators," *Journal of Econometrics*, 30, 203–238.
- HAUSMAN, J. (1997): "Valuation of New Goods under Perfect and Imperfect Competition," in *The Economics of New Goods*, ed. by T. F. Bresnahan and R. J. Gordon, University of Chicago Press, 209–237.
- HAUSMAN, J., G. LEONARD, AND J. D. ZONA (1994): "Competitive Analysis with Differentiated Products," *Annales d'Économie et de Statistique*, 159–180.
- HAUSMAN, J. A. AND P. A. RUUD (1987): "Specifying and testing econometric models for rank-ordered data," *Journal of Econometrics*, 34, 83–104.
- HIRSCH, B., E. J. JAHN, A. MANNING, AND M. OBERFICHTNER (2022): "The Wage Elasticity of Recruitment," Discussion Paper 15675, IZA - Institute of Labor Economics.
- ILO (2019): "The Global Labour Income Share and Distribution: Key Findings," Tech. rep., Data Production and Analysis Unit, International Labour Organization Department of Statistics.
- KARABARBOUNIS, L. AND B. NEIMAN (2013): "The Global Decline of the Labor Share," *The Quarterly Journal of Economics*, 129, 61–103.
- KLEIBERGEN, F. AND R. PAAP (2006): "Generalized Reduced Rank Tests Using the Singular Value Decomposition," *Journal of Econometrics*, 133, 97–126.

- LAMADON, T., M. MOGSTAD, AND B. SETZLER (2022): "Imperfect Competition, Compensating Differentials, and Rent Sharing in the US Labor Market," *American Economic Review*, 112, 169–212.
- LANGELLA, M. AND A. MANNING (2021): "Marshall Lecture 2020 The Measure of Monopsony," *Journal of the European Economic Association*, 19, 2929–2957.
- LE BARBANCHON, T., R. RATHÉLOT, AND A. ROULET (2020): "Gender Differences in Job Search: Trading off Commute against Wage," *The Quarterly Journal of Economics*, 136, 381–426.
- LERNER, A. P. (1934): "The Concept of Monopoly and the Measurement of Monopoly Power," *The Review of Economic Studies*, 1, 157–175.
- MANNING, A. (2003): *Monopsony in Motion: Imperfect Competition in Labor Markets*, Princeton University Press.
- (2011): "Imperfect Competition in the Labor Market," in *Handbook of Labor Economics*, ed. by D. Card and O. Ashenfelter, Elsevier, vol. 4, 973–1041.
- (2021): "Monopsony in Labor Markets: A Review," *ILR Review*, 74, 3–26.
- MARINESCU, I. AND R. RATHÉLOT (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.
- MATĚJKÁ, 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.
- MERTENS, M. AND B. MOTTIROLI (2025): "Do larger firms exert more market power? Markups and markdowns along the size distribution," *Economics Letters*, 255, 112500.
- MORTENSEN, D. T. (2003): *Wage Dispersion: Why Are Similar Workers Paid Differently?*, The MIT Press.
- NEWHEY, 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.
- REYNAERT, M. AND F. VERBOVEN (2014): "Improving the performance of random coefficients demand models: The role of optimal instruments," *Journal of Econometrics*, 179, 83–98.
- ROBINSON, J. (1933): *The Economics of Imperfect Competition*, London: Macmillan.
- 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.
- SKLAR, A. (1973): "Random variables, joint distribution functions, and copulas," *Kybernetika*, 9, 449–460.
- SOKOLOVA, A. AND T. SORENSEN (2021): "Monopsony in Labor Markets: A Meta-Analysis," *ILR Review*, 74, 27–55.
- STOCK, J. H. AND M. YOGO (2005): "Testing for Weak Instruments in Linear IV Regression," in *Essays in Honor of Thomas Rothenberg*, ed. by D. W. K. Andrews and J. H. Stock, Cambridge University Press, 80–108.
- 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.
- VAN OPHEM, H., P. STAM, AND B. VAN PRAAG (1999): "Multichoice Logit: Modeling Incomplete Preference Rankings of Classical Concerts," *Journal of Business & Economic Statistics*, 17, 117–128.
- YEH, C., C. MACALUSO, AND B. HERSHBEIN (2022): "Monopsony in the US Labor Market," *American Economic Review*, 112, 2099–2138.

# Appendix

## A Marginal improvement algorithm

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

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

- Step 1: Choose any  $j_t \in \arg \max_{j \in \mathcal{J} \setminus \Lambda_{t-1}} U_i(\Lambda_{t-1} \cup \{j\})$ .
- 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 Proofs and derivations

### B.1 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,

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

$$= \{r_i(\mathcal{J}, 1)\},$$

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

$$\Lambda_{t-1} = \{r_i(\mathcal{J}, 1), \dots, r_i(\mathcal{J}, t-1)\}, \quad (\text{B.1})$$

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  $\ell \in \Lambda_{t-1}$ ,<sup>40</sup>

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

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

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

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

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

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  $r_i(\mathcal{J}, n_i)$

---

<sup>40</sup>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  $t$ .

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  $n_i + 1$ .

From (a), we obtain

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

which holds if and only if

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

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

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

which holds if and only if

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

Finally, note that the following monotonicity properties must hold.

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

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

where  $\{n_i + 2, \dots, J\} \equiv \emptyset$  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  $k \in \{1, \dots, n_i - 1\}$ . Then, we get the contradiction

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

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  $k \in \{n_i + 2, \dots, J\}$ . Then, we get the contradiction

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

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.  $\square$

## B.2 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

$$\begin{aligned} \eta_{q_j, w_\ell}(\boldsymbol{\delta}, \boldsymbol{\theta}, \beta_w) &= \frac{\partial \ln(q_j(\boldsymbol{\delta}, \boldsymbol{\theta}))}{\partial \ln(w_\ell)} \\ &= \frac{1}{q_j(\boldsymbol{\delta}, \boldsymbol{\theta})} \frac{\partial q_j(\boldsymbol{\delta}, \boldsymbol{\theta})}{\partial \ln(w_\ell)} \\ &= \frac{1}{q_j(\boldsymbol{\delta}, \boldsymbol{\theta})} \frac{\partial q_j(\boldsymbol{\delta}, \boldsymbol{\theta})}{\partial \delta_\ell} \frac{\partial \delta_\ell}{\partial \ln(w_\ell)} \\ &= \frac{1}{q_j(\boldsymbol{\delta}, \boldsymbol{\theta})} \left[ I \sum_{n=1}^J \frac{\partial s_{j|n}(\boldsymbol{\delta})}{\partial \delta_\ell} s_n(\boldsymbol{\delta}, \boldsymbol{\theta}) + s_{j|n}(\boldsymbol{\delta}) \frac{\partial s_n(\boldsymbol{\delta}, \boldsymbol{\theta})}{\partial \delta_\ell} \right] \beta_w, \end{aligned} \quad (\text{B.8})$$

where the last equality follows from partially differentiating Equation (9) with respect to  $\delta_\ell$  and Equation (7) —for job  $\ell$ — 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

$$\begin{aligned} \eta_{s_{j|n}, w_\ell}(\boldsymbol{\delta}, \beta_w) &= \frac{\partial \ln(s_{j|n}(\boldsymbol{\delta}))}{\partial \ln(w_\ell)} \\ &= \frac{1}{s_{j|n}(\boldsymbol{\delta})} \frac{\partial s_{j|n}(\boldsymbol{\delta})}{\partial \ln(w_\ell)} \\ &= \frac{1}{s_{j|n}(\boldsymbol{\delta})} \frac{\partial s_{j|n}(\boldsymbol{\delta})}{\partial \delta_\ell} \frac{\partial \delta_\ell}{\partial \ln(w_\ell)} \\ &= \frac{1}{s_{j|n}(\boldsymbol{\delta})} \frac{\partial s_{j|n}(\boldsymbol{\delta})}{\partial \delta_\ell} \beta_w \end{aligned} \quad (\text{B.9})$$

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

to  $w_\ell$  by

$$\begin{aligned}
\eta_{s_n, w_\ell}(\boldsymbol{\delta}, \boldsymbol{\theta}, \beta_w) &= \frac{\partial \ln(s_n(\boldsymbol{\delta}, \boldsymbol{\theta}))}{\partial \ln(w_\ell)} \\
&= \frac{1}{s_n(\boldsymbol{\delta}, \boldsymbol{\theta})} \frac{\partial s_n(\boldsymbol{\delta}, \boldsymbol{\theta})}{\partial \ln(w_\ell)} \\
&= \frac{1}{s_n(\boldsymbol{\delta}, \boldsymbol{\theta})} \frac{\partial s_n(\boldsymbol{\delta}, \boldsymbol{\theta})}{\partial \delta_\ell} \frac{\partial \delta_\ell}{\partial \ln(w_\ell)} \\
&= \frac{1}{s_n(\boldsymbol{\delta}, \boldsymbol{\theta})} \frac{\partial s_n(\boldsymbol{\delta}, \boldsymbol{\theta})}{\partial \delta_\ell} \beta_w.
\end{aligned} \tag{B.10}$$

Then, we can rewrite Equation (B.8) as

$$\begin{aligned}
\eta_{q_j, w_\ell}(\boldsymbol{\delta}, \boldsymbol{\theta}, \beta_w) &= \sum_{n=1}^J \frac{I s_{j|n}(\boldsymbol{\delta}) s_n(\boldsymbol{\delta}, \boldsymbol{\theta})}{q_j(\boldsymbol{\delta}, \boldsymbol{\theta})} \frac{1}{s_{j|n}(\boldsymbol{\delta})} \frac{\partial s_{j|n}(\boldsymbol{\delta})}{\partial \delta_\ell} \beta_w + \sum_{n=1}^J \frac{I s_{j|n}(\boldsymbol{\delta}) s_n(\boldsymbol{\delta}, \boldsymbol{\theta})}{q_j(\boldsymbol{\delta}, \boldsymbol{\theta})} \frac{1}{s_n(\boldsymbol{\delta}, \boldsymbol{\theta})} \frac{\partial s_n(\boldsymbol{\delta}, \boldsymbol{\theta})}{\partial \delta_\ell} \beta_w \\
&= \sum_{n=1}^J \omega_{n|j}(\boldsymbol{\delta}, \boldsymbol{\theta}) \eta_{s_{j|n}, w_\ell}(\boldsymbol{\delta}, \beta_w) + \sum_{n=1}^J \omega_{n|j}(\boldsymbol{\delta}, \boldsymbol{\theta}) \eta_{s_n, w_\ell}(\boldsymbol{\delta}, \boldsymbol{\theta}, \beta_w)
\end{aligned} \tag{B.11}$$

where the weights

$$\omega_{n|j}(\boldsymbol{\delta}, \boldsymbol{\theta}) = \frac{I s_{j|n}(\boldsymbol{\delta}) s_n(\boldsymbol{\delta}, \boldsymbol{\theta})}{q_j(\boldsymbol{\delta}, \boldsymbol{\theta})} \tag{B.12}$$

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

$$\sum_{n=1}^J \omega_{n|j}(\boldsymbol{\delta}, \boldsymbol{\theta}) = 1 \tag{B.13}$$

trivially by Equation (9). The first term on the right-hand side of Equation (B.11) 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  $\ell$  is

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

$$= \frac{1}{q^f(\boldsymbol{\delta}, \boldsymbol{\theta})} \sum_{j \in \mathcal{J}^f} q_j(\boldsymbol{\delta}, \boldsymbol{\theta}) \eta_{q_j, w_\ell}(\boldsymbol{\delta}, \boldsymbol{\theta}, \beta_w), \quad (\text{B.14})$$

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,  $\mathbf{w}^f = \{w_\ell\}_{\ell \in \mathcal{J}^f}$ , is given by

$$\eta_{q^f, \mathbf{w}^f}(\boldsymbol{\delta}, \boldsymbol{\theta}, \beta_w) = \frac{1}{q^f(\boldsymbol{\delta}, \boldsymbol{\theta})} \sum_{\ell \in \mathcal{J}^f} \sum_{j \in \mathcal{J}^f} q_j(\boldsymbol{\delta}, \boldsymbol{\theta}) \eta_{q_j, w_\ell}(\boldsymbol{\delta}, \boldsymbol{\theta}, \beta_w). \quad (\text{B.15})$$

### B.3 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  $\delta_\ell$  for  $n \in \{1, \dots, J\}$ . The partial derivative of the unconditional pmf,  $s_n(\boldsymbol{\delta})$ , with respect to  $\delta_\ell$  is then obtained by integrating the partial of the conditional pmf over  $F_{\alpha, \gamma}(\cdot, \cdot | \boldsymbol{\theta})$ :

$$\begin{aligned} \frac{\partial s_n(\boldsymbol{\delta}, \boldsymbol{\theta})}{\partial \delta_\ell} &= \frac{\partial}{\partial \delta_\ell} \int s_{n|\alpha, \gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i) dF_{\alpha, \gamma}(\alpha_i, \gamma_i | \boldsymbol{\theta}) \\ &= \int \frac{\partial s_{n|\alpha, \gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i)}{\partial \delta_\ell} dF_{\alpha, \gamma}(\alpha_i, \gamma_i | \boldsymbol{\theta}), \end{aligned} \quad (\text{B.16})$$

where the second equality follows since  $s_{n|\alpha, \gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i)$  is continuously differentiable in  $\delta_\ell$  and the support of  $F_{\alpha, \gamma}$  does not depend on  $\delta_\ell$ , allowing us to pass the derivative operator through the integral sign.<sup>41</sup>

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

$$E_\ell(S) = \frac{\exp(\delta_\ell)}{\sum_{k \in \{j\} \cup S} \exp(\delta_k)}, \quad (\text{B.17})$$

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  $\delta_\ell$

$$\frac{\partial E_j(S)}{\partial \delta_\ell} = \begin{cases} E_j(S)[1 - E_j(S)] & \text{if } \ell = j \\ -\mathbb{1}\{\ell \in S\} E_j(S) E_\ell(S) & \text{otherwise} \end{cases}. \quad (\text{B.18})$$

---

<sup>41</sup>The function  $s_{n|\alpha, \gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i)$  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  $[1 - F_\varepsilon(\psi_i^{n+1})^{\exp(\delta_k)}]$ . 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  $\boldsymbol{\delta}$ .

Thus,

$$\begin{aligned}
\frac{\partial s_{j|n}(\boldsymbol{\delta})}{\partial \delta_j} &= \sum_{k=J-n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} \frac{\partial E_j(A)}{\partial \delta_j} + \sum_{s=1}^{|\mathcal{B}_j \setminus A|} (-1)^s \sum_{B \in \mathcal{R}_s(\mathcal{B}_j \setminus A)} \frac{\partial E_j(A \cup B)}{\partial \delta_j} \\
&= \sum_{k=J-n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} E_j(A) [1 - E_j(A)] + \sum_{s=1}^{|\mathcal{B}_j \setminus A|} (-1)^s \sum_{B \in \mathcal{R}_s(\mathcal{B}_j \setminus A)} E_j(A \cup B) [1 - E_j(A \cup B)]
\end{aligned} \tag{B.19}$$

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

$$\begin{aligned}
\frac{\partial s_{j|n}(\boldsymbol{\delta})}{\partial \delta_\ell} &= \sum_{k=J-n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} \frac{\partial E_j(A)}{\partial \delta_\ell} + \sum_{s=1}^{|\mathcal{B}_j \setminus A|} (-1)^s \sum_{B \in \mathcal{R}_s(\mathcal{B}_j \setminus A)} \frac{\partial E_j(A \cup B)}{\partial \delta_\ell} \\
&= - \sum_{k=J-n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} \left[ \mathbb{1}_{\{\ell \in A\}} E_j(A) E_\ell(A) \right. \\
&\quad \left. + \sum_{s=1}^{|\mathcal{B}_j \setminus A|} (-1)^s \sum_{B \in \mathcal{R}_s(\mathcal{B}_j \setminus A)} \mathbb{1}_{\{\ell \in A \cup B\}} E_j(A \cup B) E_\ell(A \cup B) \right]. \tag{B.20}
\end{aligned}$$

To find the partial derivative of  $s_{n|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i)$  with respect to  $\delta_\ell$  for  $n \in \{1, \dots, J-1\}$ , rewrite Equation (11) as

$$s_{n|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i) = \sum_{s=0}^J \sum_{A \in \mathcal{R}_s(\mathcal{J})} \sum_{t \in \tau_n^s} \sum_{B \in \mathcal{R}_t(\mathcal{J} \setminus A)} \left[ \left( \prod_{k \in A} a_k \right) \left( \prod_{p \in B} b_p \right) \left( \prod_{q \in \mathcal{J} \setminus (A \cup B)} c_q \right) \right], \tag{B.21}$$

where  $a_k = (F_{n+1})^{\exp(\delta_k)} - (F_n)^{\exp(\delta_k)}$ ,  $b_k = (F_n)^{\exp(\delta_k)}$ ,  $c_k = 1 - (F_{n+1})^{\exp(\delta_k)}$ , and  $F_k = F_\varepsilon(\psi_i^k)$  for  $k \in \{1, \dots, J\}$ . Note that the expression on the right-hand side of (B.21) 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  $\ell$  belongs to only one of these subsets, the chain rule yields

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

$$\begin{aligned}
&= \sum_{s=0}^J \sum_{A \in \mathcal{R}_s(\mathcal{J})} \sum_{t \in \tau_n^s} \sum_{B \in \mathcal{R}_t(\mathcal{J} \setminus A)} \left[ \left( \prod_{k \in A} a_k \right) \left( \prod_{p \in B} b_p \right) \left( \prod_{q \in \mathcal{J} \setminus (A \cup B)} c_q \right) \right. \\
&\quad \left. \left( \mathbb{1}_{\{\ell \in A\}} \frac{1}{a_\ell} \frac{\partial a_\ell}{\partial \delta_\ell} + \mathbb{1}_{\{\ell \in B\}} \frac{1}{b_\ell} \frac{\partial b_\ell}{\partial \delta_\ell} + \mathbb{1}_{\{\ell \notin A \cup B\}} \frac{1}{c_\ell} \frac{\partial c_\ell}{\partial \delta_\ell} \right) \right], \tag{B.22}
\end{aligned}$$

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

## B.4 Proof of Lemma 1

*Remark.* The following proof makes use of the properties of the EV<sub>1</sub> 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

$$U_i(A) = u_{ir_i(A,1)} - c_i(|A|). \tag{B.23}$$

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

$$U_i(\emptyset) = -c_i(0) = 0 = u_{i0}. \tag{B.24}$$

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

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

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

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

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

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:

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

This event has probability

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

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  $j \in \mathcal{J}$ ,

$$\begin{aligned} \mathbb{P}\left(A_i = \{j\} \mid A_i \neq \emptyset\right) &= \mathbb{P}\left(\max_{\ell \in \mathcal{J}} u_{i\ell} \leq u_{ij}\right) \\ &= \frac{\exp(\delta_j)}{\sum_{\ell \in \mathcal{J}} \exp(\delta_\ell)}, \end{aligned}$$

which establishes part (iii) of Lemma 1.

Part (iv) of Lemma 1 follows directly from the properties of the EV<sub>1</sub> 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

$$\begin{aligned} \tilde{s}_0(\boldsymbol{\delta}, \gamma) &= F_\varepsilon(\gamma)^{\sum_{\ell \in \mathcal{J}} \exp(\delta_\ell)} \\ &< F_\varepsilon\left(\frac{\gamma}{\alpha}\right)^{\sum_{\ell \in \mathcal{J}} \exp(\delta_\ell)} \\ &= s_0(\boldsymbol{\delta}, \alpha, \gamma). \end{aligned}$$

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

$$\begin{aligned} \frac{\partial \tilde{s}_0(\boldsymbol{\delta}, \gamma)}{\partial \delta_\ell} &= \exp(\delta_\ell) F_\varepsilon(\gamma)^{\sum_{k \in \mathcal{J}} \exp(\delta_k)} \ln(F_\varepsilon(\gamma)) \\ &< \exp(\delta_\ell) F_\varepsilon\left(\frac{\gamma}{\alpha}\right)^{\sum_{k \in \mathcal{J}} \exp(\delta_k)} \ln\left(F_\varepsilon\left(\frac{\gamma}{\alpha}\right)\right) \\ &= \frac{\partial s_0(\boldsymbol{\delta}, \alpha, \gamma)}{\partial \delta_\ell}. \end{aligned}$$

□

## B.5 Proof of Proposition 2

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

$$\sum_{j \in \mathcal{J}} \sigma_{j|n}(\delta) = n$$

to obtain

$$\sum_{j \in \mathcal{J}} \frac{\partial \sigma_{j|n}(\delta)}{\partial \delta_\ell} = 0, \quad \forall \ell \in \mathcal{J}, \quad (\text{B.25})$$

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

$$\sigma_{j|n}(\delta + c \boldsymbol{\iota}) = \sigma_{j|n}(\delta), \quad \forall c \in \mathbb{R},$$

implies

$$\frac{d\sigma_{j|n}(\delta + c \boldsymbol{\iota})}{dc} = 0, \quad \forall c \in \mathbb{R}.$$

Evaluating at  $c = 0$ , we obtain

$$\begin{aligned} 0 &= \left. \frac{d\sigma_{j|n}(\delta + c \boldsymbol{\iota})}{dc} \right|_{c=0} \\ &= \left. \frac{\partial \sigma_{j|n}(\delta + c \boldsymbol{\iota})}{\partial (\delta + c \boldsymbol{\iota})'} \frac{d(\delta + c \boldsymbol{\iota})}{dc} \right|_{c=0} \\ &= \left. \frac{\partial \sigma_{j|n}(\delta)}{\partial \delta'} \boldsymbol{\iota} \right. \\ &= \sum_{\ell \in \mathcal{J}} \frac{\partial \sigma_{j|n}(\delta)}{\partial \delta_\ell}, \end{aligned} \quad (\text{B.26})$$

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,

$$\text{rank} \left( \frac{\partial \sigma_{\cdot|n}(\delta)}{\partial \delta} \right) \leq J - 1. \quad (\text{B.27})$$

Next, we note that the location normalization  $\delta_b = 0$  yields a full-rank, nonsingular Jacobian of the restricted mapping  $\sigma_{\cdot|n}^b(\delta^b)$  for every  $\delta^b \in \mathbb{R}^{J-1}$  by eliminating the degree of freedom from

location invariance and the counting identity. From Equations (B.19) and (B.20) in Appendix B.3, the unrestricted Jacobian follows the sign pattern

$$\frac{\partial s_{j|n}(\boldsymbol{\delta})}{\partial \delta_j} > 0, \quad \frac{\partial s_{j|n}(\boldsymbol{\delta})}{\partial \delta_\ell} < 0, \quad \ell \neq j, \quad (\text{B.28})$$

reflecting that an increase in  $\delta_j$  raises job  $j$ 's conditional inclusion probability while reducing that of competing jobs. From (B.26) we obtain

$$\begin{aligned} 0 &< -\frac{\partial s_{j|n}(\boldsymbol{\delta})}{\partial \delta_b} \\ &= \frac{\partial s_{j|n}(\boldsymbol{\delta})}{\partial \delta_j} + \sum_{\ell \in \mathcal{J} \setminus \{j, b\}} \frac{\partial s_{j|n}(\boldsymbol{\delta})}{\partial \delta_\ell} \\ &= \left| \frac{\partial s_{j|n}^b(\boldsymbol{\delta}^b)}{\partial \delta_j} \right| - \sum_{\ell \in \mathcal{J} \setminus \{j, b\}} \left| \frac{\partial s_{j|n}^b(\boldsymbol{\delta}^b)}{\partial \delta_\ell} \right| \end{aligned} \quad (\text{B.29})$$

for  $j \neq b$ , where the inequality follows from (B.28) and the last equality follows from the definition of the restricted mapping and the sign pattern in (B.28). 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,  $s_{\cdot|n}(\cdot)$  is globally injective on  $\mathbb{R}^{J-1}$ .<sup>42</sup> Finally, any injective function is a bijection from its domain onto its image.  $\square$

## B.6 Proof of Proposition 3

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

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

so we can solve for

$$\psi_0^1 = \frac{\gamma_0}{\alpha_0} = F_\varepsilon^{-1}\left(\mathbb{E}[s_0]^{\frac{1}{\sum_{\ell \in \mathcal{J}} \exp(\delta_{\ell 0})}}\right) \quad (\text{B.30})$$

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  $(\alpha_0, \gamma_0)$  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

---

<sup>42</sup>Each conditional inclusion probability  $s_{j|n}(\boldsymbol{\delta})$  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  $s_{\cdot|n}^b(\cdot)$  is continuously differentiable.

$\{u_{ir_i(\mathcal{J},n)} \geq \psi^n\}$ . 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

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

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  $\psi_0^n$ . Since the cdf of the  $(J-n+1)$ -th order statistic,

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

(see David and Nagaraja, 2003, p. 96), is continuous and strictly increasing, the corresponding survivor function  $1 - F_{u_{(J-n+1)}}(x)$  is invertible. Hence, we conclude that, for any  $n \in \{2, \dots, J\}$  with  $\mathbb{P}(n_i \geq n) \in (0, 1)$  —which exists by Assumption 5—,  $\psi_0^n$  is identified as

$$\psi_0^n = F_{u_{(J-n+1)}}^{-1} \left( 1 - \sum_{k=n}^J \mathbb{E}[s_k] \right). \quad (\text{B.31})$$

Finally, with  $\psi_0^1$  and  $\psi_0^n$  pinned down for some  $n \in \{2, \dots, J\}$  with nondegenerate survivor probability, we can use (20) and (B.30) to solve for  $\boldsymbol{\theta}_0$ :

$$\alpha_0 = 1 - \left( \frac{\psi_0^1}{\psi_0^n} \right)^{\frac{1}{n-1}}, \quad \gamma_0 = \psi_0^1 \left[ 1 - \left( \frac{\psi_0^1}{\psi_0^n} \right)^{\frac{1}{n-1}} \right]. \quad (\text{B.32})$$

Note that the mapping  $(\alpha_0, \gamma_0) \mapsto \{\mathbb{E}[s_n]\}_{n=0}^J$  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  $\{s_n(\boldsymbol{\delta}_0, \alpha^*, \gamma^*)\}_{n=0}^J = \{s_n(\boldsymbol{\delta}_0, \alpha_0, \gamma_0)\}_{n=0}^J$ , so both  $\boldsymbol{\theta}$  parameters map to the same distribution of optimal portfolio sizes  $\{\mathbb{E}[s_n]\}_{n=0}^J$ . Then,  $\psi^{1*} = \psi_0^1$  by (B.30) and  $\psi^{n*} = \psi_0^n$  by (B.31), so  $(\alpha^*, \gamma^*) = (\alpha_0, \gamma_0)$  by (B.32), a logical contradiction. Therefore,  $\{s_n(\boldsymbol{\delta}_0, \alpha^*, \gamma^*)\}_{n=0}^J = \{s_n(\boldsymbol{\delta}_0, \alpha_0, \gamma_0)\}_{n=0}^J \implies (\alpha^*, \gamma^*) = (\alpha_0, \gamma_0)$  and the mapping is injective.

□

## B.7 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  $\mathbf{g} : (0, 1) \times \mathbb{R}_{++} \rightarrow \mathbb{R}^2$  defined by

$$\mathbf{g}(\boldsymbol{\theta}) = \left( \ln \left( \frac{\mathcal{S}_{n_1}^{>0}(\boldsymbol{\theta})}{\mathcal{S}_{n_2}^{>0}(\boldsymbol{\theta})} \right) \quad \ln \left( \frac{\mathcal{S}_{n_3}^{>0}(\boldsymbol{\theta})}{\mathcal{S}_{n_2}^{>0}(\boldsymbol{\theta})} \right) \right)'$$

$$= \begin{pmatrix} \ln(\mathcal{S}_{n_1}(\boldsymbol{\theta})) - \ln(\mathcal{S}_{n_2}(\boldsymbol{\theta})) & \ln(\mathcal{S}_{n_3}(\boldsymbol{\theta})) - \ln(\mathcal{S}_{n_2}(\boldsymbol{\theta})) \end{pmatrix}',$$

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

$$\frac{\partial \mathbf{g}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} = \left( \frac{\partial \ln(\mathcal{S}_{n_1}(\boldsymbol{\theta}_0))}{\partial \boldsymbol{\theta}} - \frac{\partial \ln(\mathcal{S}_{n_2}(\boldsymbol{\theta}_0))}{\partial \boldsymbol{\theta}}, \frac{\partial \ln(\mathcal{S}_{n_3}(\boldsymbol{\theta}_0))}{\partial \boldsymbol{\theta}} - \frac{\partial \ln(\mathcal{S}_{n_2}(\boldsymbol{\theta}_0))}{\partial \boldsymbol{\theta}} \right)'. \quad (\text{B.33})$$

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

From the proof of Proposition 3 in Appendix B.6, we know that

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

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  $\psi_0^n$  is the  $n$ -th stopping-rule threshold from Proposition 1 specialized to the homogeneous  $(\alpha_i, \gamma_i)$  case and evaluated at  $\boldsymbol{\theta}_0$ . Thus, for any  $n \in \{1, \dots, J\}$  such that  $\mathcal{S}_n(\boldsymbol{\theta}_0) \in (0, 1)$ ,

$$\frac{\partial \ln(\mathcal{S}_n(\boldsymbol{\theta}_0))}{\partial \boldsymbol{\theta}} = -\frac{f_{u_{(J-n+1)}}(\psi_0^n)}{1 - F_{u_{(J-n+1)}}(\psi_0^n)} \frac{\partial \psi_0^n}{\partial \boldsymbol{\theta}} \quad (\text{B.34})$$

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  $\psi_0^n$  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 (B.34) takes the form

$$\frac{\partial \ln(\mathcal{S}_n(\boldsymbol{\theta}_0))}{\partial \boldsymbol{\theta}} = -c_n \mathbf{v}_n, \quad c_n = \frac{f_{u_{(J-n+1)}}(\psi_0^n)}{1 - F_{u_{(J-n+1)}}(\psi_0^n)} \psi_0^n, \quad \mathbf{v}_n = \begin{pmatrix} \frac{\alpha_0 n - 1}{\alpha_0(1 - \alpha_0)} & \frac{1}{\gamma_0} \end{pmatrix}'. \quad (\text{B.35})$$

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  $\mathbf{v}_n \in \mathbb{R}^2$  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  $\alpha_0 - v_n^1 = 0$  and  $\mathbf{v}_n$  remains a well-defined vertical vector on the  $(\alpha, \gamma)$ -plane. This means that, for any two distinct  $n_k \neq n_\ell \in \{1, \dots, J\}$ , the vectors  $\mathbf{v}_{n_k}$  and  $\mathbf{v}_{n_\ell}$  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

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

are pairwise linearly independent, so the Jacobian (B.33) has linearly independent rows and is thus nonsingular. In fact, it can be directly verified that the determinant of the Jacobian,

$$\det\left(\frac{\partial \mathbf{g}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right) = \frac{1}{\gamma_0} \left[ -c_{n_1} c_{n_3} (v_{n_3}^1 - v_{n_1}^1) + c_{n_1} c_{n_2} (v_{n_2}^1 - v_{n_1}^1) + c_{n_2} c_{n_3} (v_{n_3}^1 - v_{n_2}^1) \right], \quad (\text{B.36})$$

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 (B.36) 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,  $\mathbf{g}(\boldsymbol{\theta})$  is invertible in a neighborhood of  $\boldsymbol{\theta}_0$ , so  $\boldsymbol{\theta}_0$  is locally identified by inverting the system of equations

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

□

## B.8 Proof of Proposition 4 (sketch)

*Remark.* Throughout the proof sketch below, read

$$(\alpha_i, \gamma_i) = \left(Q_\alpha(u \mid \boldsymbol{\theta}_\alpha), Q_\gamma(v \mid \boldsymbol{\theta}_\gamma)\right), \quad (u, v) = \left(F_\alpha(\alpha_i \mid \boldsymbol{\theta}_\alpha), F_\gamma(\gamma_i \mid \boldsymbol{\theta}_\gamma)\right)$$

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  $1 \leq n_1 < \dots < n_L \leq J$  such that  $\mathcal{S}_{n_\ell}(\boldsymbol{\theta}_0) \in (0, 1)$  for  $\ell \in \{0, \dots, L\}$ . Consider the mapping  $\mathbf{g} : \Theta \rightarrow \mathbb{R}^L$  defined by

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

where the second equality follows since  $\mathcal{S}_n^{>0}(\boldsymbol{\theta}) = \mathcal{S}_n(\boldsymbol{\theta}) / (1 - \mathcal{S}_0(\boldsymbol{\delta}_0, \boldsymbol{\theta}))$  for all  $n \in \{1, \dots, J\}$  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  $k \geq n$ ,

$$\mathcal{S}_{n|\alpha,\gamma}(\alpha_i, \gamma_i) = \sum_{k=n}^J \mathcal{S}_{k|\alpha,\gamma}(\boldsymbol{\delta}_0, \alpha_i, \gamma_i)$$

$$= 1 - F_{u(J-n+1)}(\psi^n(\alpha_i, \gamma_i)),$$

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 B.6 since  $(\alpha_i, \gamma_i)$  are held fixed here. For the same reason, analogous arguments to those in Appendix B.7 establish that

$$\frac{\partial \mathcal{S}_{n|\alpha,\gamma}(\alpha_i, \gamma_i)}{\partial(\alpha_i, \gamma_i)} = -f_{u(J-n+1)}(\psi^n(\alpha_i, \gamma_i))\psi^n(\alpha_i, \gamma_i) \left( \frac{\alpha_i n-1}{\alpha_i(1-\alpha_i)} - \frac{1}{\gamma_i} \right)' . \quad (\text{B.37})$$

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

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

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

$$\frac{\partial \mathcal{S}_n(\boldsymbol{\theta})}{\partial \theta_\alpha^r} = \int_0^1 \kappa_n^\alpha(u | \boldsymbol{\theta}) \frac{\partial Q_\alpha(u | \boldsymbol{\theta}_\alpha)}{\partial \theta_\alpha^r} du,$$

where

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

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

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

where

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

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

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

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

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

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

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

$$\begin{aligned}\frac{\partial g_\ell(\boldsymbol{\theta}_0)}{\partial \theta_\alpha^p} &= \int_0^1 \left( \frac{\kappa_{n_\ell}^\alpha(u \mid \boldsymbol{\theta}_0)}{\mathcal{S}_{n_\ell}(\boldsymbol{\theta}_0)} - \frac{\kappa_{n_0}^\alpha(u \mid \boldsymbol{\theta}_0)}{\mathcal{S}_{n_0}(\boldsymbol{\theta}_0)} \right) \frac{\partial Q_\alpha(u \mid \boldsymbol{\theta}_{\alpha 0})}{\partial \theta_\alpha^p} du, \\ \frac{\partial g_\ell(\boldsymbol{\theta}_0)}{\partial \theta_\gamma^q} &= \int_0^1 \left( \frac{\kappa_{n_\ell}^\gamma(v \mid \boldsymbol{\theta}_0)}{\mathcal{S}_{n_\ell}(\boldsymbol{\theta}_0)} - \frac{\kappa_{n_0}^\gamma(v \mid \boldsymbol{\theta}_0)}{\mathcal{S}_{n_0}(\boldsymbol{\theta}_0)} \right) \frac{\partial Q_\gamma(v \mid \boldsymbol{\theta}_{\gamma 0})}{\partial \theta_\gamma^q} dv, \\ \frac{\partial g_\ell(\boldsymbol{\theta}_0)}{\partial \theta_C^r} &= \int_{[0,1]^2} \left( \frac{\mathcal{S}_{n_\ell|\alpha,\gamma}(Q_u^\alpha, Q_v^\gamma \mid \boldsymbol{\theta}_0^{\alpha,\gamma})}{\mathcal{S}_{n_\ell}(\boldsymbol{\theta}_0)} - \frac{\mathcal{S}_{n_0|\alpha,\gamma}(Q_u^\alpha, Q_v^\gamma \mid \boldsymbol{\theta}_0^{\alpha,\gamma})}{\mathcal{S}_{n_0}(\boldsymbol{\theta}_0)} \right) \frac{\partial \ln(c(u, v \mid \boldsymbol{\theta}_{C 0}))}{\partial \theta_C^r} dC(u, v \mid \boldsymbol{\theta}_{C 0})\end{aligned}$$

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

Just as in our discussion of comparative statics in the homogeneous model in Section 2.4.3, we can conclude from Equation (B.37) 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  $\gamma_i$  with stronger effect for larger  $n$ , while it is nonmonotonic in  $\alpha_i$ : increasing for  $\alpha_i < 1/n$ , and decreasing for  $\alpha_i > 1/n$ . Averaging over ranks  $(U, V) \sim C(u, v \mid \boldsymbol{\theta}_{C 0})$  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  $\boldsymbol{\theta}_\gamma \mapsto Q_\gamma(\cdot \mid \boldsymbol{\theta}_\gamma)$  imposed by Assumption 6 implies the  $\partial \mathbf{g}(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}'_\alpha$  and  $\partial \mathbf{g}(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}'_\gamma$  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  $\gamma_i$ ) can cancel a kinked pattern that necessarily flips signs (conditional survivor derivative in  $\alpha_i$ ).

Finally, the  $\partial \mathbf{g}(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}'_C$  block of the Jacobian is also full rank and linearly independent of the other two blocks. To see this, note that the copula scores  $\{\partial \ln(c(u, v \mid \boldsymbol{\theta}_{C 0}))/\partial \theta_C^r\}_{r=1}^{\dim(\boldsymbol{\theta}_C)}$  are linearly independent in  $L^2(C_0)$ , where  $C_0 \equiv C(\cdot, \cdot \mid \boldsymbol{\theta}_{C 0})$ , by the nonsingular Fisher information matrix regularity condition in Assumption 6. Since each  $\theta_C^r$ -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  $n_\ell$  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  $\mathbf{g}(\boldsymbol{\theta}_0)$  is locally injective at  $\boldsymbol{\theta}_0$  by the inverse function theorem, and the  $L$  moment

equations

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

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

□

## B.9 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  $\mathbf{d}^f$  with  $\ell$ -th entry  $d_\ell^f = \mathbb{1}\{\ell \in \mathcal{J}^f\}$ . 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,  $\{\delta_\ell\}_{\ell \in \mathcal{J}^f}$ :

$$\begin{aligned} \frac{dq^f(\boldsymbol{\delta} + c \mathbf{d}^f, \boldsymbol{\theta})}{dc} \Big|_{c=0} &= \left( \frac{d}{dc} \sum_{j \in \mathcal{J}^f} q_j(\boldsymbol{\delta} + c \mathbf{d}^f, \boldsymbol{\theta}) \right) \Big|_{c=0} \\ &= I \sum_{j \in \mathcal{J}^f} \sum_{n=1}^J \left( \frac{\partial}{\partial (\boldsymbol{\delta} + c \mathbf{d}^f)} s_{j|n}(\boldsymbol{\delta} + c \mathbf{d}^f) \frac{d}{dc} (\boldsymbol{\delta} + c \mathbf{d}^f) s_n(\boldsymbol{\delta} + c \mathbf{d}^f, \boldsymbol{\theta}) \right. \\ &\quad \left. + \frac{\partial}{\partial (\boldsymbol{\delta} + c \mathbf{d}^f)} s_n(\boldsymbol{\delta} + c \mathbf{d}^f, \boldsymbol{\theta}) \frac{d}{dc} (\boldsymbol{\delta} + c \mathbf{d}^f) s_{j|n}(\boldsymbol{\delta} + c \mathbf{d}^f) \right) \Big|_{c=0} \\ &= I \sum_{j \in \mathcal{J}^f} \sum_{n=1}^J \frac{\partial s_{j|n}(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}'} \mathbf{d}^f s_n(\boldsymbol{\delta}, \boldsymbol{\theta}) + \frac{\partial s_n(\boldsymbol{\delta}, \boldsymbol{\theta})}{\partial \boldsymbol{\delta}'} \mathbf{d}^f s_{j|n}(\boldsymbol{\delta}) \\ &= I \sum_{j \in \mathcal{J}^f} \sum_{n=1}^J \sum_{\ell \in \mathcal{J}^f} \frac{\partial s_{j|n}(\boldsymbol{\delta})}{\partial \delta_\ell} s_n(\boldsymbol{\delta}, \boldsymbol{\theta}) + \frac{\partial s_n(\boldsymbol{\delta}, \boldsymbol{\theta})}{\partial \delta_\ell} s_{j|n}(\boldsymbol{\delta}) \\ &= \sum_{\ell \in \mathcal{J}^f} \sum_{j \in \mathcal{J}^f} \frac{\partial q_j(\boldsymbol{\delta}, \boldsymbol{\theta})}{\partial \delta_\ell} \\ &= \sum_{\ell \in \mathcal{J}^f} \frac{\partial q^f(\boldsymbol{\delta}, \boldsymbol{\theta})}{\partial \delta_\ell}. \end{aligned} \tag{B.38}$$

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  $(\boldsymbol{\delta} + c \mathbf{d}^f, \boldsymbol{\theta})$  and differentiating with respect to  $c$ . The third equality follows by direct evaluation at  $c = 0$ .

The fourth equality follows from the property that  $\mathbf{x}'\mathbf{d}^f = \sum_{\ell \in \mathcal{J}^f} x_\ell$  for any  $\mathbf{x} \in \mathbb{R}^J$  by the definition of  $\mathbf{d}^f$  as a vector-valued 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  $\delta_\ell$ . 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 (B.38):

$$\begin{aligned}\eta_{q^f, w^f}(\boldsymbol{\delta}, \boldsymbol{\theta}, \beta_w) &= \sum_{\ell \in \mathcal{J}^f} \frac{\partial \ln(q^f(\boldsymbol{\delta}, \boldsymbol{\theta}))}{\partial \ln(w_\ell)} \\ &= \frac{1}{q^f(\boldsymbol{\delta}, \boldsymbol{\theta})} \sum_{\ell \in \mathcal{J}^f} \frac{\partial q^f(\boldsymbol{\delta}, \boldsymbol{\theta})}{\partial \delta_\ell} \frac{\partial \delta_\ell}{\partial \ln(w_\ell)} \\ &= \frac{1}{q^f(\boldsymbol{\delta}, \boldsymbol{\theta})} \left( \frac{dq^f(\boldsymbol{\delta} + c \mathbf{d}^f, \boldsymbol{\theta})}{dc} \right) \Big|_{c=0} \beta_w.\end{aligned}\tag{B.39}$$

The first equality is the definition of this elasticity. The second equality follows from the chain rule. The last equality follows from Equations (B.38) and (8).

## C 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

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

Our model defines (i) a mapping  $s_{j|n}(\delta)$  from  $\delta$  to  $\mathbb{P}(j \in A_i \mid n_i = n)$ , and (ii) a mapping  $s_n(\delta, \theta)$  from  $\delta$  and  $\theta$  to  $\mathbb{P}(n_i = n)$ . These mappings follow directly from parts (ii) and (i) of Proposition 1, respectively.

## C.1 The conditional application share function

Consider first the conditional (expected) application share function  $s_{j|n}(\delta)$ . The probability that  $j$  belongs to the application portfolio conditional on the job seeker applying to every job is trivially  $\mathbb{P}(j \in A_i \mid n_i = J) = 1$ . For  $n \in \{1, \dots, J-1\}$ , 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.,  $\mathbb{P}(j \in A_i \mid n_i = n) = \mathbb{P}(u_{ij} > u_{ir_i(\mathcal{J}, n+1)})$ . 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  $\mathbb{P}(u_{ij} > u_{ir_i(\mathcal{J}, n+1)})$  as a function of  $\delta$ —and, obviously, of  $j$  and  $n$ , which we indicate by the subscript in  $s_{j|n}(\delta)$ —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  $u_{i(1)} < \dots < u_{i(J)}$ , and note that

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

for all  $n \in \{1, \dots, J-1\}$ . Similarly, let  $\mathcal{B}_j \equiv \mathcal{J} \setminus \{j\}$  represent the leave-out set of available jobs excluding  $j$ , and  $\{u_{i(n)}^j\}_{n=1}^{J-1}$  the order statistics of  $\{u_{i\ell}\}_{\ell \in \mathcal{B}_j}$  such that  $u_{i(1)}^j < \dots < u_{i(J-1)}^j$ . 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  $n \in \{1, \dots, J-1\}$ . The mutual independence of  $\{u_{i\ell}\}_{\ell \in \mathcal{J}}$  implies that  $u_{i(n)}^j$  is independent of  $u_{ij}$  for all  $n \in \{1, \dots, J-1\}$ .

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)

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

where  $F_\varepsilon(\cdot)$  is the marginal cdf of  $\varepsilon_{ij}$ . The cdf of the  $n$ -th order statistic  $u_{i(n)}^j$  is then given by (see, e.g., [David and Nagaraja, 2003](#), p. 96)

$$F_{u_{i(n)}^j}(x) = \sum_{k=n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} \prod_{\ell \in A} F_{u_\ell}(x) \prod_{m \in \mathcal{B}_j \setminus A} [1 - F_{u_m}(x)]$$

$$= \sum_{k=n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} \prod_{\ell \in A} F_\varepsilon(x - \delta_\ell) \prod_{m \in \mathcal{B}_j \setminus A} [1 - F_\varepsilon(x - \delta_m)], \quad (\text{C.4})$$

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<sub>1</sub> distribution,  $F_\varepsilon(x) = \exp(-\exp(-x))$ , we obtain

$$\begin{aligned} s_{j|n}(\boldsymbol{\delta}) &= \mathbb{P}(u_{ij} > u_{ir_i(\mathcal{J}, n+1)}) \\ &= \mathbb{P}(u_{ij} > u_{i(J-n)}) \\ &= \mathbb{P}\left(u_{ij} > u_{i(J-n)}^j\right) \\ &= \int_{-\infty}^{\infty} \mathbb{P}\left(x > u_{i(J-n)}^j\right) dF_{u_j}(x) \\ &= \int_{-\infty}^{\infty} F_{u_{i(J-n)}^j}(x) dF_\varepsilon(x - \delta_j) \\ &= \int_{-\infty}^{\infty} \sum_{k=J-n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} \prod_{\ell \in A} F_\varepsilon(x - \delta_\ell) \prod_{m \in \mathcal{B}_j \setminus A} [1 - F_\varepsilon(x - \delta_m)] dF_\varepsilon(x - \delta_j) \\ &= \int_{-\infty}^{\infty} \sum_{k=J-n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} \prod_{\ell \in A} F_\varepsilon(x - \delta_j)^{\frac{\exp(\delta_\ell)}{\exp(\delta_j)}} \prod_{m \in \mathcal{B}_j \setminus A} \left[1 - F_\varepsilon(x - \delta_j)^{\frac{\exp(\delta_m)}{\exp(\delta_j)}}\right] dF_\varepsilon(x - \delta_j) \\ &= \int_0^1 \sum_{k=J-n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} \prod_{\ell \in A} u^{\frac{\exp(\delta_\ell)}{\exp(\delta_j)}} \prod_{m \in \mathcal{B}_j \setminus A} \left[1 - u^{\frac{\exp(\delta_m)}{\exp(\delta_j)}}\right] du \\ &= \int_0^1 \sum_{k=J-n}^{J-1} \sum_{A \in \mathcal{R}_k(\mathcal{B}_j)} u^{\frac{\sum_{\ell \in A} \exp(\delta_\ell)}{\exp(\delta_j)}} \prod_{m \in \mathcal{B}_j \setminus A} \left[1 - u^{\frac{\exp(\delta_m)}{\exp(\delta_j)}}\right] du. \end{aligned} \quad (\text{C.5})$$

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  $u_{i(J-n)}^j$ . 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 eighth 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.<sup>43</sup> We can easily verify that we get the

---

<sup>43</sup>Here, we use the word “inclusion” to emphasize that the underlying (conditional) event is  $\{j \in A_i \mid n_i = n\}$ , i.e.  $\{j \text{ is included in the portfolio} \mid n \text{ alternatives are included}\}$ , whereas in the standard MNL model the underlying event is  $\{j \text{ is chosen}\}$ .

standard choice probability for  $n = 1$ :

$$\begin{aligned}\vartheta_{j|1}(\boldsymbol{\delta}) &= \int_0^1 u^{\frac{\sum_{\ell \in \mathcal{J} \setminus \{j\}} \exp(\delta_\ell)}{\exp(\delta_j)}} du \\ &= \frac{\exp(\delta_j)}{\sum_{\ell \in \mathcal{J}} \exp(\delta_\ell)},\end{aligned}$$

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

$$\vartheta_{j|n}(\boldsymbol{\delta}) = \int_0^1 f_{j|n}(u, \boldsymbol{\delta}) du,$$

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

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

implying the recursive relations

$$\vartheta_{j|n}(\boldsymbol{\delta}) = \vartheta_{j|n-1}(\boldsymbol{\delta}) + \int_0^1 f_j(u, \boldsymbol{\delta}, J-n) du, \tag{C.6}$$

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

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

Given parameters  $\boldsymbol{\delta}$ , the integral on the right-hand side of Equation (C.5) can be accurately approximated by numerical quadrature for any  $n \in \{1, \dots, J\}$ . 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

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

where  $\sum_{s=1}^0(\cdot) \equiv 0$  for notational consistency. Given parameter estimates  $\hat{\boldsymbol{\delta}}$ , 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 The probability mass function of optimal portfolio size

Consider now the conditional pmf of the number of applications conditional on the admission probability  $\alpha_i$  and the cost of applications  $\gamma_i$ ,  $s_{n|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i)$ . We can recover the unconditional pmf,  $s_n(\boldsymbol{\delta}, \boldsymbol{\nu})$ , by integrating the conditional pmf over the joint distribution of parameters  $(\alpha_i, \gamma_i)$ , 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).<sup>44</sup> 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  $(\alpha_i, \gamma_i)$  in Equation (12). Conditional on  $(\alpha_i, \gamma_i)$ , the probability of this event is

$$s_{0|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i) = \mathbb{P}(n_i = 0 \mid \alpha_i, \gamma_i)$$

---

<sup>44</sup>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.

$$\begin{aligned}
&= F_{u_{(J)}}(\psi_i^1) \\
&= \prod_{\ell \in \mathcal{J}} F_\varepsilon(\psi_i^1)^{\exp(\delta_\ell)} \\
&= F_\varepsilon(\psi_i^1)^{\sum_{\ell \in \mathcal{J}} \exp(\delta_\ell)}. \tag{C.9}
\end{aligned}$$

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.,

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

The conditional probability is given by

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

Finally, for the interior events  $n_i = n \in \{1, \dots, J - 1\}$ , 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  $\mathbf{u}_i$  such that exactly  $J - n$  elements lie below  $\psi_i^n$  and the remaining  $n$  elements lie above  $\psi_i^{n+1}$ . 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  $\psi_i^n$  and the utilities of the remaining alternatives are larger than  $\psi_i^{n+1}$ , i.e.,

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

Panel (b) of Figure C.1 illustrates the element of the partition where  $s = J$ . This case includes all realizations of  $\mathbf{u}_i$  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

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

Finally, for the cases  $s \in \{1, \dots, J - 1\}$  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  $\mathbf{u}_i$  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  $t \in \{\max(J - n - s, 0), \dots, \min(J - n, J - s)\}$  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  $\psi_i^n$  and  $\psi_i^{n+1}$ , the utilities of the alternatives in  $B$  lie below  $\psi_i^n$ , and the remaining alternatives in  $\mathcal{J} \setminus (A \cup B)$  have utilities larger than  $\psi_i^{n+1}$ . The corresponding expression is

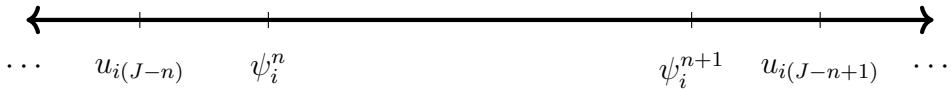
$$\sum_{A \in \mathcal{R}_s(\mathcal{J})} \prod_{\ell \in A} [F_{u_\ell}(\psi_i^{n+1}) - F_{u_\ell}(\psi_i^n)] \sum_{t \in \tau_n^s} \sum_{B \in \mathcal{R}_t(\mathcal{J} \setminus A)} \prod_{p \in B} F_{u_p}(\psi_i^n) \prod_{q \in \mathcal{J} \setminus (A \cup B)} [1 - F_{u_q}(\psi_i^{n+1})],$$

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

$$\begin{aligned} s_{n|\alpha,\gamma}(\boldsymbol{\delta}, \alpha_i, \gamma_i) &= \mathbb{P}(n_i = n \mid \alpha_i, \gamma_i) \\ &= \sum_{s=0}^J \sum_{A \in \mathcal{R}_s(\mathcal{J})} \prod_{\ell \in A} [F_{u_\ell}(\psi_i^{n+1}) - F_{u_\ell}(\psi_i^n)] \\ &\quad \sum_{t \in \tau_n^s} \sum_{B \in \mathcal{R}_t(\mathcal{J} \setminus A)} \prod_{p \in B} F_{u_p}(\psi_i^n) \prod_{q \in \mathcal{J} \setminus (A \cup B)} [1 - F_{u_q}(\psi_i^{n+1})] \end{aligned}$$

$$\begin{aligned}
&= \sum_{s=0}^J \sum_{A \in \mathcal{R}_s(\mathcal{J})} \prod_{\ell \in A} \left[ F_\varepsilon(\psi_i^{n+1})^{\exp(\delta_\ell)} - F_\varepsilon(\psi_i^n)^{\exp(\delta_\ell)} \right] \\
&\quad \sum_{t \in \tau_n^s} \sum_{B \in \mathcal{R}_t(\mathcal{J} \setminus A)} F_\varepsilon(\psi_i^n)^{\sum_{p \in B} \exp(\delta_p)} \prod_{q \in \mathcal{J} \setminus (A \cup B)} \left[ 1 - F_\varepsilon(\psi_i^{n+1})^{\exp(\delta_q)} \right]. \quad (\text{C.11})
\end{aligned}$$

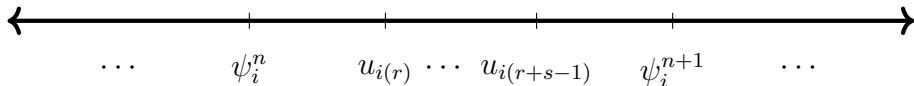
**Figure C.1.** Realizations of the order statistics consistent with  $n$  applications



**(a)** No realizations of  $u_{ij}$  lie between  $\psi_i^n$  and  $\psi_i^{n+1}$



**(b)** All realizations of  $u_{ij}$  lie between  $\psi_i^n$  and  $\psi_i^{n+1}$



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

*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  $n \in \{1, \dots, J-1\}$ . The thresholds  $\psi_i^n$  and  $\psi_i^{n+1}$  are defined in Equation (12). Cases are indexed by the number of realizations of  $u_{ij}$  in the interval  $[\psi_i^n, \psi_i^{n+1}]$ ,  $s \in \{0, \dots, J\}$ . The case  $s = 0$  in Panel (a) is equivalent to exactly  $J-n$  realizations of  $u_{ij}$  below  $\psi_i^n$  and exactly  $n$  above  $\psi_i^{n+1}$ . The case  $s = J$  in Panel (b) is equivalent to exactly  $J$  realizations of  $u_{ij}$  between  $\psi_i^n$  and  $\psi_i^{n+1}$ . For the cases  $s \in \{1, \dots, J-1\}$  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  $\psi_i^n$  and  $\psi_i^{n+1}$ . Then, we have at least  $\max(J-n-s, 0)$  and at most  $\min(J-n, J-s)$  realizations below  $\psi_i^n$ , with the remaining realizations above  $\psi_i^{n+1}$ .

## D Partially rank-ordered logit: Minorize–maximize algorithm

This appendix closely follows Appendix E of [Roussille and Scuderi \(2025\)](#). The likelihood contribution of applicant  $i$  can be written as

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

where  $\mathcal{A}_i = \{A_i, \bar{A}_i\}$  is applicant  $i$ '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  $\{\varepsilon_{ij}\}_{j \in \mathcal{J}} \stackrel{iid}{\sim} EV_1 \implies \mathbb{P}(\max_{\ell \in \bar{A}_i} \delta_\ell + \varepsilon_{i\ell} \leq x) = F_\varepsilon(x - \ln(\sum_{\ell \in \bar{A}_i} \exp(\delta_\ell)))$ , 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  $u = F_\varepsilon(x - \ln(\sum_{\ell \in \bar{A}_i} \exp(\delta_\ell)))$ , and the last equality makes the change of variable  $z = u^{1/\sum_{\ell \in \bar{A}_i} \exp(\delta_\ell)}$ . 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

$$\mathcal{L}(\boldsymbol{\delta} \mid \{\mathcal{A}_i\}_{i \in \mathcal{I}_1}) = \sum_{i \in \mathcal{I}_1} \ln \left( f_i(\boldsymbol{\delta} \mid \mathcal{A}_i) \right),$$

which could be directly maximized using the expression in Equation (D.1).<sup>45</sup> 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.<sup>46</sup>

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

$$g(\delta \mid \delta^{(n)}) \leq \mathcal{L}(\delta), \forall \delta$$

$$g(\delta^{(n)} \mid \delta^{(n)}) = \mathcal{L}(\delta^{(n)}).$$

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

$$\begin{aligned} \mathcal{L}(\delta^{(n+1)}) &\geq g(\delta^{(n+1)} \mid \delta^{(n)}) \\ &\geq g(\delta^{(n)} \mid \delta^{(n)}) \\ &= \mathcal{L}(\delta^{(n)}), \end{aligned}$$

where the first inequality follows from the definition of  $g(\cdot \mid \delta^{(n)})$  as a minorizing function of  $\mathcal{L}(\cdot)$  at  $\delta^{(n)}$ , 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.,

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

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  $\delta^{(n)}$ , we start by defining

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

---

<sup>45</sup>For notational simplicity, we hereafter suppress the dependence of the likelihood function on the data  $\{\mathcal{A}_i\}_{i \in \mathcal{I}_1}$ .

<sup>46</sup>See [Wu and Lange \(2010\)](#) for an introduction to MM algorithms.

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

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

and noting that

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

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

$$\begin{aligned} \ln \left( \int_0^1 \frac{\varphi_i(\boldsymbol{\delta}, z \mid \boldsymbol{\delta}^{(n)})}{\varphi_i(\boldsymbol{\delta}^{(n)}, z \mid \boldsymbol{\delta}^{(n)})} \pi_i(z \mid \boldsymbol{\delta}^{(n)}) dz \right) &\geq \int_0^1 \ln \left( \frac{\varphi_i(\boldsymbol{\delta}, z \mid \boldsymbol{\delta}^{(n)})}{\varphi_i(\boldsymbol{\delta}^{(n)}, z \mid \boldsymbol{\delta}^{(n)})} \right) \pi_i(z \mid \boldsymbol{\delta}^{(n)}) dz \\ \iff \mathcal{L}_i(\boldsymbol{\delta}) &\geq \mathcal{L}_i(\boldsymbol{\delta}^{(n)}) + \int_0^1 \ln \left( \frac{\varphi_i(\boldsymbol{\delta}, z \mid \boldsymbol{\delta}^{(n)})}{\varphi_i(\boldsymbol{\delta}^{(n)}, z \mid \boldsymbol{\delta}^{(n)})} \right) \pi_i(z \mid \boldsymbol{\delta}^{(n)}) dz, \end{aligned} \tag{D.2}$$

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

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

and rewriting (D.2) as

$$\mathcal{L}_i(\boldsymbol{\delta}) \geq H_{\pi i}^{(n)} + \int_0^1 \ln \left( \varphi_i(\boldsymbol{\delta}, z \mid \boldsymbol{\delta})^{(n)} \right) \pi_i(z \mid \boldsymbol{\delta}^{(n)}) dz, \tag{D.3}$$

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

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

and

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

$$\geq \sum_{\ell \in \bar{A}_i} \ln \left( \frac{\exp(\delta_\ell)}{\exp(\delta_\ell^{(n)})} \right) \rho_i \left( \delta_\ell^{(n)} \mid \boldsymbol{\delta}^{(n)} \right)$$

$$\iff \ln \left( \sum_{\ell \in \bar{A}_i} \exp(\delta_\ell) \right) \geq \sum_{\ell \in \bar{A}_i} \delta_\ell \rho_i \left( \delta_\ell^{(n)} \mid \boldsymbol{\delta}^{(n)} \right) + H_{\rho i}^{(n)}, \quad (\text{D.5})$$

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

$$\mathcal{L}_i(\boldsymbol{\delta}) \geq H_i^{(n)} + g_i \left( \boldsymbol{\delta} \mid \boldsymbol{\delta}^{(n)} \right), \quad (\text{D.6})$$

where

$$g_i \left( \boldsymbol{\delta} \mid \boldsymbol{\delta}^{(n)} \right) = \int_0^1 \sum_{j \in A_i} \ln \left( 1 - z^{\rho_i(\delta_j \mid \boldsymbol{\delta}^{(n)})} \right) \pi_i \left( z \mid \boldsymbol{\delta}^{(n)} \right) dz + \sum_{\ell \in \bar{A}_i} \delta_\ell \rho_i \left( \delta_\ell^{(n)} \mid \boldsymbol{\delta}^{(n)} \right)$$

$$+ \sum_{\ell \in \bar{A}_i} \rho_i \left( \delta_\ell \mid \boldsymbol{\delta}^{(n)} \right) \int_0^1 \ln(z) \pi_i \left( z \mid \boldsymbol{\delta}^{(n)} \right) dz,$$

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

and (D.6) holds with equality at  $\boldsymbol{\delta} = \boldsymbol{\delta}^{(n)}$ . Thus, the log-likelihood function  $\mathcal{L}(\boldsymbol{\delta}) = \sum_{i \in \mathcal{I}_1} \mathcal{L}_i(\boldsymbol{\delta})$  is minorized at  $\boldsymbol{\delta}^{(n)}$  by the surrogate function

$$g \left( \boldsymbol{\delta} \mid \boldsymbol{\delta}^{(n)} \right) = H^{(n)} + \sum_{i \in \mathcal{I}_1} g_i \left( \boldsymbol{\delta} \mid \boldsymbol{\delta}^{(n)} \right), \quad (\text{D.7})$$

where  $H^{(n)} = \sum_{i \in \mathcal{I}_1} H_i^{(n)}$ .

In its  $n^{\text{th}}$  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  $\sum_{i \in \mathcal{I}_1} g_i \left( \boldsymbol{\delta} \mid \boldsymbol{\delta}^{(n)} \right)$  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  $\sum_{i \in \mathcal{I}_1} g_i \left( \boldsymbol{\delta} \mid \boldsymbol{\delta}^{(n)} \right)$  is given by

$$\boldsymbol{\delta}^{(n+1)} = \boldsymbol{\delta}^{(n)} + \left( - \sum_{i \in \mathcal{I}_1} \frac{\partial^2 g_i \left( \boldsymbol{\delta} \mid \boldsymbol{\delta}^{(n)} \right)}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}'} \right)^{-1} \left( \sum_{i \in \mathcal{I}_1} \frac{\partial g_i \left( \boldsymbol{\delta} \mid \boldsymbol{\delta}^{(n)} \right)}{\partial \boldsymbol{\delta}} \right) \Big|_{\boldsymbol{\delta}=\boldsymbol{\delta}^{(n)}}$$

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  $j^{\text{th}}$  entry of its gradient and the  $j^{\text{th}}$  diagonal element of its Hessian are respectively given by

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

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

$$\delta_j^{(n+1)} = \delta_j^{(n)} + \frac{\sum_{i \in \mathcal{I}_1} \rho_{ij}^{(n)} \kappa_{ij}^{(n)}}{\sum_{i \in \mathcal{I}_1} \rho_{ij}^{(n)} \lambda_{ij}^{(n)}}, \quad (\text{D.8})$$

where

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

$\rho_{ij}^{(n)} = \rho_i(\delta_j^{(n)} | \boldsymbol{\delta}^{(n)})$ ,  $\pi_i^{(n)}(z) = \pi_i(z | \boldsymbol{\delta}^{(n)})$ , and all the integrals involved can be approximated by numerical quadrature.

We demean  $\boldsymbol{\delta}^{(n)}$  in each iteration to improve numerical stability. In the event that demeaning is not enough to prevent numerical overflow when computing the integrals, we rescale by normalizing  $\boldsymbol{\delta}^{(n)}$  to unit norm. While the likelihood function is not invariant to such rescaling, it typically allows the algorithm to continue in its ascent path. In the rare event that the  $n$ -th

iteration update produces a decrease in the objective function due to the true log-likelihood increase being smaller than machine precision, we start a step-halving loop until an increase is produced or the maximum number of iterations is reached. Finally, since the location of  $\delta$  is not identified, we impose the normalization  $\delta_b^{(N)} = 0$  for base alternative  $b$  at the terminal values.

## E Method of simulated moments: Optimization details

This appendix details our simulated method of moments approach to estimation of the parameters governing the joint distribution of the random effects  $(\alpha_i, \gamma_i)$ ,  $\boldsymbol{\theta}_0$ . To make matters more concrete, the discussion here is framed in terms of our preferred specification from our empirical application in Section 4.2. There, we partition  $\boldsymbol{\theta} = (\boldsymbol{\theta}'_\alpha, \boldsymbol{\theta}'_\gamma, \theta_C)' = ((\theta_1, \theta_2), (\theta_3, \theta_4), \theta_5)'$  and specify the random-effect joint distribution as represented by a Gaussian copula of the form

$$F_{\alpha, \gamma}(\alpha_i, \gamma_i | \boldsymbol{\theta}) = C(F_\alpha(\alpha_i | \boldsymbol{\theta}_\alpha), F_\gamma(\gamma_i | \boldsymbol{\theta}_\gamma) | \theta_C)$$

where

$$C(u, v | \theta_C) = \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v) | \theta_C),$$

$\Phi(\cdot)$  is the cdf of the standard normal distribution, and  $\Phi_2(\cdot, \cdot | \theta_C)$  is the cdf of the standard bivariate normal distribution with correlation parameter  $\theta_C \in (-1, 1)$ . The scalar Gaussian-copula correlation parameter  $\theta_C$  measures the degree of comonotone (if positive) or countermotone (if negative) dependence between  $(\alpha_i, \gamma_i)$ . The marginal distributions  $F_\alpha(\cdot | \boldsymbol{\theta}_\alpha)$  and  $F_\gamma(\cdot | \boldsymbol{\theta}_\gamma)$  are specified as

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

where  $\theta_1 \in (0, 1)$  and  $\theta_2 > 0$  are the mean and concentration of  $\alpha_i$ , while  $\theta_3 > 0$  and  $\theta_4 > 0$  are the mean and shape parameter of  $\gamma_i$ . Our optimization algorithm to find the MSM estimator proceeds as follows.

**Bootstrap estimate of the optimal weight matrix.** The optimal weight matrix is

$$\mathbf{W}^* = \text{Var}(\mathbf{m}(n_i) | n_i > 0)^{-1}, \tag{E.1}$$

We construct our nonparametric bootstrap estimator by taking  $B$  random samples with replacement,  $\mathbf{n}_1^b = (n_1^b, \dots, n_{I_1}^b)$ , from the restriction of the observed vector  $\mathbf{n}$  to  $n_i > 0$ ,  $n_1$ , where  $I_1 \equiv |\mathcal{I}_1|$  is the observed number of applicants.<sup>47</sup> For each  $b \in \{1, \dots, B\}$ , we compute the bootstrapped empirical moments  $\bar{\mathbf{m}}_{I_1}(\mathbf{n}_1^b)$  and then estimate their variance through the method of

---

<sup>47</sup>Note that  $\mathbf{n}_1 = \mathbf{n}$  in zero-truncated data.

moments estimator

$$\widehat{\mathbf{W}}^{-1} = \frac{1}{B-1} \sum_{b=1}^B \left( \overline{\mathbf{m}}_{I_1}(\mathbf{n}^b) - \frac{1}{B} \sum_{s=1}^B \overline{\mathbf{m}}_{I_1}(\mathbf{n}^s) \right) \left( \overline{\mathbf{m}}_{I_1}(\mathbf{n}^b) - \frac{1}{B} \sum_{s=1}^B \overline{\mathbf{m}}_{I_1}(\mathbf{n}^s) \right)' . \quad (\text{E.2})$$

Note that this produces a consistent estimate of the variance of the aggregate empirical moments as opposed to the individual moment functions. In particular, our variance estimate is scaled by  $1/I_1$  compared to the estimate that uses individual moment functions instead of averages across individuals. While  $I_1 \equiv |\{i \in \mathcal{I} : n_i > 0\}| = \dim(\mathbf{n}_1)$  is itself a random variable, it is deterministic in the bootstrap setting. Moreover, when constructing the sample criterion function, the random scaling factor  $I_1$  is simply an irrelevant positive constant that does not affect the optimizers. That is, the MSM estimator yields exactly the same estimates  $\widehat{\boldsymbol{\theta}}$  whether we use our alternative estimator or a direct, non-rescaled estimate of  $\mathbf{W}^*$ .

Compared to directly estimating (E.1), this simple procedure has the computational advantage of requiring the storage of smaller matrices in memory, while yielding the exact same MSM estimates. Furthermore, our value function  $Q(\widehat{\boldsymbol{\theta}})$  is already scaled by sample size, so it is immediately interpretable as Hansen's overidentification statistic—except for the detail that the population moments are simulated rather than computed analytically—with an asymptotic  $\chi^2$  distribution with  $M - L$  degrees of freedom.

Finally, we invert (E.2) through an adaptive diagonal-ridge procedure that deals with numerical singularity. If  $\widehat{\mathbf{W}}^{-1}$  is positive definite and its condition number does not exceed a threshold  $\kappa_{\max}$ , we directly invert this matrix. Otherwise, we iteratively add  $\tau_t \mathbf{d}$  to its diagonal, where  $\mathbf{d}$  represents the original diagonal of  $\widehat{\mathbf{W}}^{-1}$  with any non-positive entries replaced by the median positive entry, and the scalar  $\tau_t$  is adaptively determined as follows. Set initial values  $\tau_0$  and  $\widetilde{\mathbf{W}}_0^{-1} = \widehat{\mathbf{W}}^{-1}$ . In iteration  $t$ , compare the ratio of the largest to smallest eigen values of  $\widetilde{\mathbf{W}}_t^{-1} = \widetilde{\mathbf{W}}_{t-1}^{-1} + \tau_{t-1} \mathbf{d}$ ,  $\kappa_t = \lambda_t^{\max}/\lambda_t^{\min}$ , to  $\kappa_{\max}$  if  $\lambda_t^{\min} > 0$ . If  $\kappa_t \leq \kappa_{\max}$ , set the inverse of  $\widetilde{\mathbf{W}}_t^{-1}$  as our optimal weight matrix estimate. Otherwise, if either  $\lambda_t^{\min} \leq 0$  or  $\kappa_t > \kappa_{\max}$ , update  $\tau_t = g_\tau \tau_{t-1}$ , where  $g_\tau$  is a predefined growth factor, and continue to the next iteration. If at any iteration  $\tau_t$  exceeds a predefined cap  $\tau_{\max}$ , set the estimate as the inverse of  $\widehat{\mathbf{W}}^{-1} + \tau \mathbf{I}_M$  if it exists, where  $\tau$  is a predefined small constant and  $\mathbf{I}_M$  is the  $M \times M$  identity matrix, or declare failure otherwise.

In our empirical application in Section 4, we set  $B = 10,000$ ,  $\kappa_{\max} = 10^8$ ,  $\tau_0 = 10^{-8}$ ,  $\tau_{\max} = 10^4$ ,  $g_\tau = 5$ , and  $\tau = 10^{-8}$ .

**Constant random numbers (CRNs).** The use of constant random numbers in estimation is essential to minimize the simulation noise of the MSM estimator by ensuring the optimizer does not drift in uninformative directions chasing artificial criterion reductions produced by simulation error (see, e.g., [Cameron and Trivedi, 2005](#), Ch. 12.8). We hold all the underly-

ing pseudo-random draws constant across iterations of our optimization algorithm, drawing each vector or matrix (and keeping it in memory) only as it becomes necessary, to improve computational speed.

Specifically, we draw  $S$  independent samples of the following CRNs.

- (i)  $J \times I$  ex-post utility matrix  $\mathbf{U}^s = \widehat{\boldsymbol{\delta}} \otimes \boldsymbol{\iota}'_I - \boldsymbol{\varepsilon}^{s'}$ , where  $\widehat{\boldsymbol{\delta}}$  is our PROL MLE of mean utilities,  $\boldsymbol{\iota}_I$  is an  $I \times 1$  vector of ones, and  $\boldsymbol{\varepsilon}^s$  is an  $I \times J$  i.i.d. matrix draw from  $F_{\varepsilon}(\cdot)$ .
- (ii) Two  $1 \times I$  i.i.d. vector draws from the Uniform(0, 1) distribution,  $\mathbf{u}_{\alpha}^s$  and  $\mathbf{v}_{\gamma}^s$ , that we use to reconstruct the  $1 \times I$  random effect draws  $\boldsymbol{\alpha}^s(\boldsymbol{\theta}_{\alpha}, \mathbf{u}_{\alpha}^s)$  and  $\boldsymbol{\gamma}^s(\boldsymbol{\theta}_{\gamma}, \theta_C, \mathbf{u}_{\alpha}^s, \mathbf{v}_{\gamma}^s)$  for different values of  $\boldsymbol{\theta}$  as:

$$\alpha_i^s(\boldsymbol{\theta}_{\alpha}, u_{\alpha i}^s) = Q_{\alpha}(u_{\alpha i}^s | \boldsymbol{\theta}_{\alpha})$$

$$\gamma_i^s(\boldsymbol{\theta}_{\gamma}, \theta_C, u_{\alpha i}^s, v_{\gamma i}^s) = \frac{\theta_3}{\theta_4} Q_{\gamma}\left(\Phi\left(\theta_5 \Phi^{-1}(u_{\alpha i}^s) + \sqrt{1 - (\theta_5)^2} \Phi^{-1}(v_{\gamma i}^s)\right) | \boldsymbol{\theta}_{\gamma}\right),$$

where  $Q_{\alpha}(u | \boldsymbol{\theta}_{\alpha})$ ,  $Q_{\gamma}(v | \boldsymbol{\theta}_{\gamma})$ , and  $\Phi^{-1}(u)$  represent the quantile functions of the Beta( $\theta_1\theta_2, (1 - \theta_1)\theta_2$ ), Gamma( $\theta_4, \theta_3/\theta_4$ ), and Normal(0, 1) distributions, respectively.

We set  $S = 50$  in our empirical application.

**Optimization-space reparameterization.** Due to the finite-sample nonsmoothness of the MSM criterion caused by the underlying threshold-jumping structure, we base our estimation routine on the derivative-free, simplex-based [Nelder and Mead \(1965\)](#) numerical optimization algorithm. Given the constrained parameter domains  $\theta_1 \in (0, 1)$ ,  $\theta_2, \theta_3, \theta_4 \in \mathbb{R}_{++}$ , and  $\theta_5 \in (-1, 1)$ , we reparameterize  $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^L$  into an unconstrained parameter vector  $\boldsymbol{\zeta} \in \mathbb{R}^L$  only for optimization purposes. The bijective reparameterization mapping is defined by

$$\begin{aligned} \boldsymbol{\zeta}(\boldsymbol{\theta}) &= \left(\text{logit}(\theta_1), \ln(\theta_2), \ln(\theta_3), \ln(\theta_4), \text{arctanh}(\theta_5)\right)' \\ \boldsymbol{\theta}(\boldsymbol{\zeta}) &= \left(\text{invlogit}(\zeta_1), \exp(\zeta_2), \exp(\zeta_3), \exp(\zeta_4), \tanh(\zeta_5)\right)', \end{aligned}$$

where  $\text{logit}(x) = \ln(x/(1-x)) \in \mathbb{R}$  and  $\text{arctanh}(x) = 1/2 \ln((1+x)/(1-x)) \in \mathbb{R}$ ,  $\text{invlogit}(x) = \exp(x)/(1+\exp(x)) \in (0, 1)$ , and  $\tanh(x) = (\exp(x) - \exp(-x))/(\exp(x) + \exp(-x)) \in (-1, 1)$  are the logit, inverse hyperbolic tangent, inverse logit, and hyperbolic tangent functions, respectively. We map back from  $\boldsymbol{\zeta}$ - to  $\boldsymbol{\theta}$ -space before reporting estimation results.

**Grid search for initial values.** Our estimation algorithm starts with a coarse grid search for initial values. We start by specifying parameter bounds in  $\boldsymbol{\theta}$ -space,  $\boldsymbol{\theta}^{\min}$  and  $\boldsymbol{\theta}^{\max}$ . We then map

these bounds into  $\zeta$ -space and build an initial grid by Latin-hypercube sampling (McKay et al., 1979)  $N^{\text{grid}}$  points inside the  $L$ -dimensional box defined by  $\zeta(\boldsymbol{\theta}^{\min})$  and  $\zeta(\boldsymbol{\theta}^{\max})$ .<sup>48</sup> We choose our initial values by cheaply evaluating the MSM objective function with only one simulation draw. Therefore, we draw the CRNs corresponding to  $s = 1$  at this stage. Grid points are ranked according to the coarse, one-draw MSM criterion and the top  $K_1$  (lowest values) are retained as initial values. In our empirical application, we set  $N^{\text{grid}} = 500$  and  $K_1 = 10$ .

**Initial Nelder–Mead refinement with wide simplex.** The second stage of our MSM estimation routine comprises the refinement of the  $K_1$  selected initial values. We initiate the Nelder–Mead algorithm at each initial value and rank the minimizers according to the terminal values of the MSM criterion. The top  $K_2$  candidates with the lowest values are selected for a final refinement in the next stage. Here, we balance computational costs and accuracy by using only  $1 \leq S_2 < S$  simulation draws. Thus, we draw the CRNs corresponding to  $s \in \{2, \dots, S_2\}$  at this stage (no additional draws if  $S_2 = 1$ ). The initial simplex is defined as

$$\mathcal{N}(\boldsymbol{\theta}^{(0)}, \Delta) = \left\{ \boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(0)} + (\Delta_1, 0, \dots, 0)', \boldsymbol{\theta}^{(0)} + (0, \Delta_2, 0, \dots, 0)', \dots, \boldsymbol{\theta}^{(0)} + (0, 0, \dots, 0, \Delta_L)' \right\},$$

where  $\boldsymbol{\theta}^{(0)}$  represents the initial value. We set the simplex deltas  $\Delta = (\Delta_1, \dots, \Delta_L)'$  as a capped fraction of the initial  $\zeta$ -box span  $\zeta(\boldsymbol{\theta}^{\max}) - \zeta(\boldsymbol{\theta}^{\min})$ :

$$\Delta_\ell^{\text{wide}} = \min \left\{ p^{\text{wide}} \left[ \zeta(\theta_\ell^{\max}) - \zeta(\theta_\ell^{\min}) \right], \bar{\Delta}_\ell \right\}, \quad \ell \in \{1, \dots, L\}.$$

In our empirical application, we set  $K_2 = 1$ ,  $S_2 = 3$ ,  $p^{\text{wide}} = 0.2$ , and  $\bar{\Delta} = (0.8, 0.7, 0.5, 0.35, 0.4)'$ .

**Final Nelder–Mead refinement with tighter simplex.** Finally, we refine the top  $K_2$  candidates by feeding them to the Nelder–Mead algorithm as starting values. We tighten the simplex in this stage by setting the simplex deltas as a smaller fraction  $p^{\text{tight}} < p^{\text{wide}}$  of the initial box span, using the same caps  $\bar{\Delta}$ . We improve accuracy, although at a significant increase in computation costs, by using the full  $S$  simulation draws, meaning we draw the remaining CRNs corresponding to  $s \in \{S_2 + 1, \dots, S\}$  only at this stage. We rank the minimizers according to the terminal MSM criterion (only if  $K_2 > 1$ ), and choose the one with the lowest value as our MSM estimate. We set  $p^{\text{tight}} = 0.07$  in our empirical application.

**Standard errors.** We do not pursue standard error estimation in the current version of the draft since our main interest lies in comparing elasticity point estimates rather than on statistical

---

<sup>48</sup>In general, it is possible that not all sampled points in  $\zeta$ -space map back to points inside the  $\boldsymbol{\theta}$ -space box. This can be caused by nonlinearities in the  $\boldsymbol{\theta} \mapsto \zeta$  mapping, for example in the reparameterization of a Uniform–Uniform specification we do not report in the paper. In such cases, we trim the grid by dropping any points that map outside the  $\boldsymbol{\theta}$ -box. It is not generally the case for the specifications we use in our empirical application.

inference on  $\theta$ , although the null hypothesis  $\theta_C = 0$  might be of particular interest. It is possible to obtain delta-method asymptotic standard errors based on the standard efficient GMM sandwich form, but this requires numerically differentiating the residual moment function  $\bar{m}_I(n) - \hat{\mu}_S(\theta)$  with respect to  $\theta$ . A more accurate but significantly costlier approach is bootstrapping the entire two-stage procedure where  $\delta$  is estimated and then  $\theta$  is estimated using the  $\delta$  estimates as input. This would account for the variance introduced by PROL-MLE estimation of  $\delta$  in the first stage. Given the sample size and the long running times, it seems impractical to bootstrap standard errors with any reasonable number of bootstrap replications in our empirical application.

## F Data details

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

### F.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](#).

## F.2 Occupation and workplace classification

Occupations are classified using the Chilean National Statistics Institute (*Instituto Nacional de Estadísticas, 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.

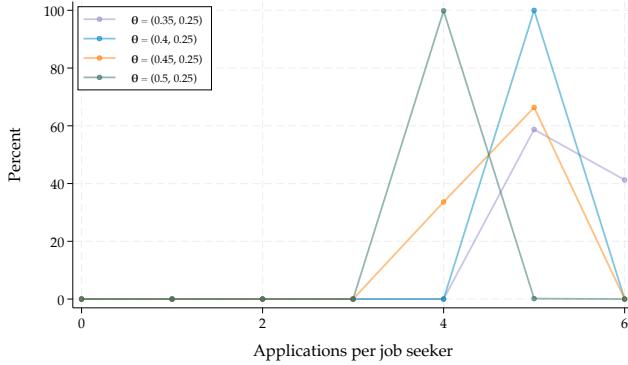
## F.3 Cleaning

**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.

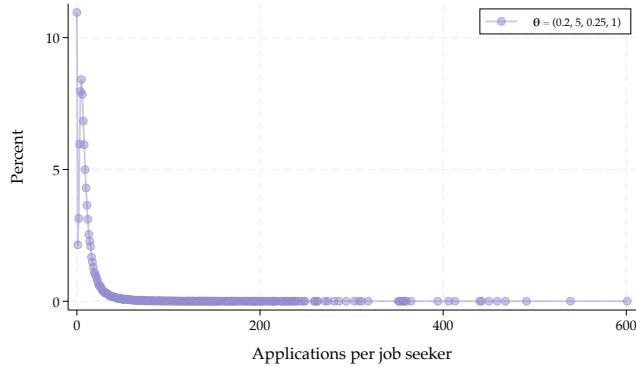
**Hours of work:** There are two part-time categories in the raw data. We combine them into one since the distinction is only nominal.

## G Further results

**Figure G.1.** Further comparative statics in  $\theta$



**(a) Homogeneous**



**Table G.1.** MSM alternative targeted moments sets

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

Notes: Description of the  $M = 6$  targeted moments chosen to define the MSM criterion function from the set of moment candidates spanned by Equations (53) to (55).

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

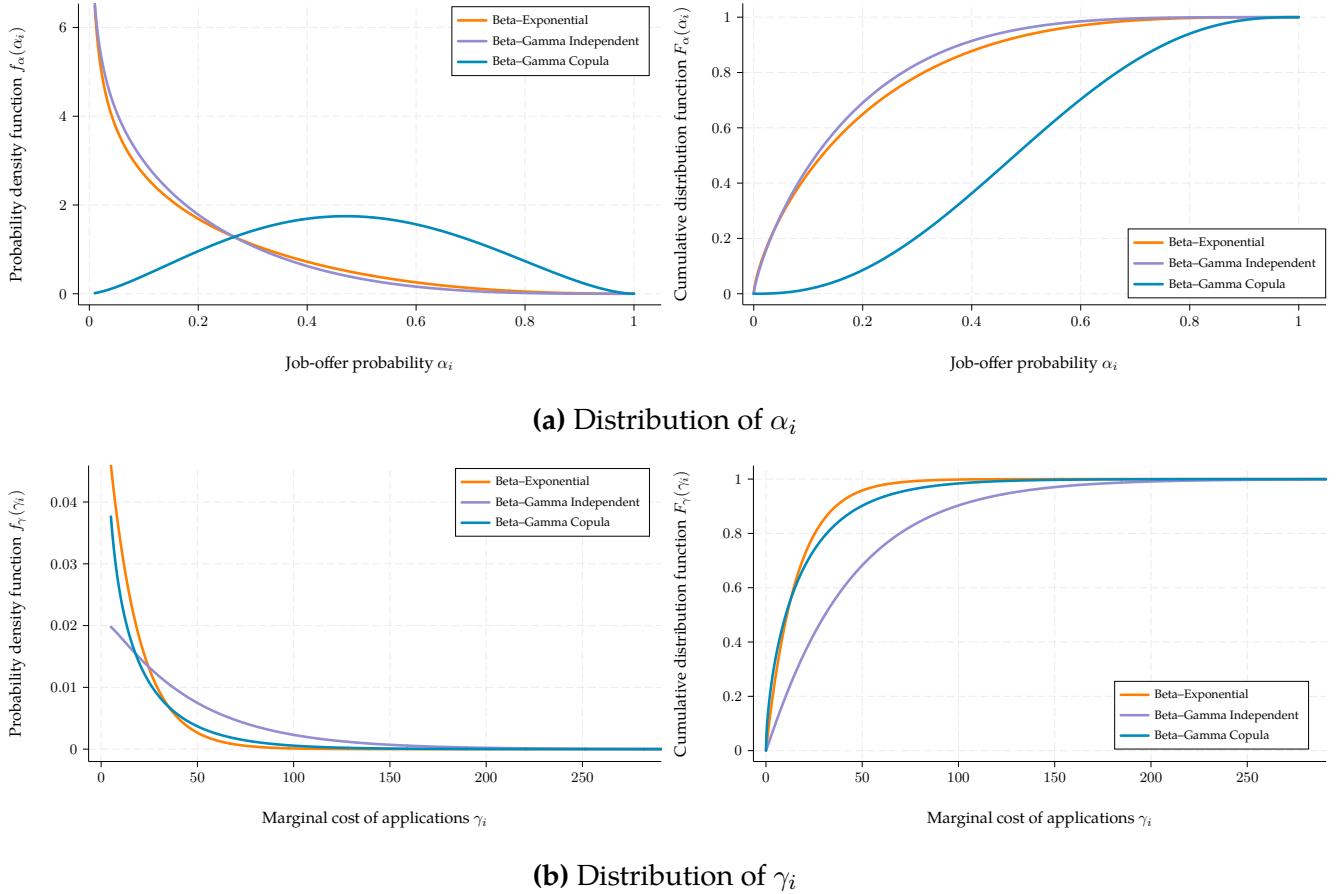
Panel (a)	Independent				Copula	
	Homogeneous (1)	Beta–Exponential (2)		Beta–Gamma (3)		Beta–Gamma (4)
<b>Estimates</b>						
$\theta_1$	1.3491e-5		0.1792		0.1601	0.4817
$\theta_2$	1.1513e-4		3.9738		4.9339	5.2574
$\theta_3$			0.0634		43.4031	19.1699
$\theta_4$					1.0500	0.6183
$\theta_5$						0.3567
Objective function	4,243.6432		22.1203		8.9051	7.7577
<b>Implied marginals</b>						
$\mathbb{E}[\alpha_i]$	1.3491e-5		0.1792		0.1601	0.4817
$\text{Var}(\alpha_i)$	0		0.0296		0.0227	0.0398
$\mathbb{E}[\gamma_i]$	1.1513e-4		15.7733		43.4031	19.1699
$\text{Var}(\gamma_i)$	0		248.7970		1,794.0500	594.3600

Panel (b)	Independent				Copula				
	Homogeneous	Beta–Exponential		Beta–Gamma		Beta–Gamma			
		Empirical (1)	Simulated (2)	Contrib. (3)	Simulated (4)	Contrib. (5)	Simulated (6)	Contrib. (7)	Simulated (8)
<b>Moments</b>									
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

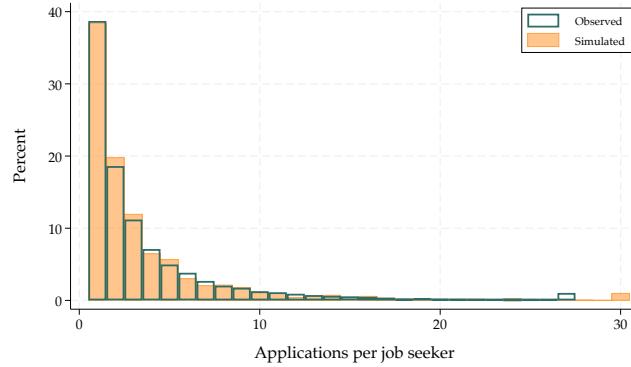
Notes: Method-of-simulated-moments estimates of the parameters  $\boldsymbol{\theta}$  of the joint distribution  $F_{\alpha,\gamma}(\cdot, \cdot | \boldsymbol{\theta})$  of the random effects  $(\alpha_i, \gamma_i)$ . The bootstrap estimate of the optimal weight matrix has full rank 6 and condition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

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

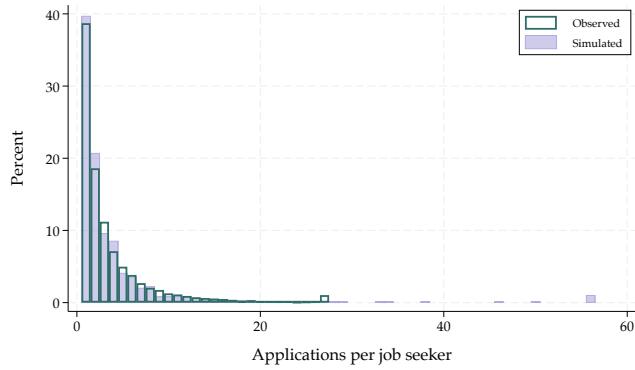


*Notes:* Marginal distributions of the random effects  $(\alpha_i, \gamma_i)$  implied by the alternative MSM estimates of  $\theta$  in Table G.2 for the nondegenerate specifications (i) independent Beta-Exponential, (ii) independent Beta-Gamma, and Beta-Gamma copula. Supports are visually truncated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

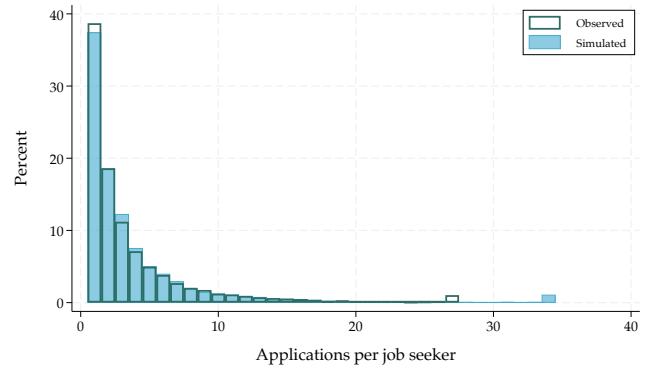
**Figure G.3.** Assessing the fit of the alternative structural parameter estimates



**(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  $\delta$  from Figure 4 and the MSM estimates of  $\theta$  from Table 3. Empirical supports are censored at the corresponding 99-th percentile. This figure, showing the full zero-truncated pmf of portfolio sizes, mixes targeted ununtargeted 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.
- CAMERON, A. C. AND P. K. TRIVEDI (2005): *Microeometrics: Methods and Applications*, Cambridge University Press.
- 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.
- McKAY, M. D., R. J. BECKMAN, AND W. J. CONOVER (1979): “Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code,” *Technometrics*, 21, 239–245.
- NELDER, J. A. AND R. MEAD (1965): “A Simplex Method for Function Minimization,” *The Computer Journal*, 7, 308–313.
- 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.