# Employer Market Power with Endogenous Labor Market Boundaries*

Olalekan Bello
New York University

Claudia Macaluso
FRB Richmond

Abdoulaye Ndiaye
New York University

May 20, 2026

**Abstract**

Employer market power is typically measured and analyzed under the assumption of fixed labor market boundaries, often approximated by geographic units such as commuting zones or counties. We show that this practice misses an important margin: workers broaden the geographic scope of their job search when their home market is concentrated or cyclically weak. Using over 63 million online applications to U.S. hourly jobs, we document two stylized facts. First, the spatial breadth of search is heterogeneous across workers: about one in three searches in a market either smaller or larger than a typical commuting zone. Second, search breadth is also cyclical: when local labor markets deteriorate, an increasing share of workers searches farther from home. To interpret these facts, we propose a search-and-matching model that endogenizes labor market boundaries by combining granular firms with multi-market search. Relative to a fixed-market benchmark, the model has a set of distinct predictions: first, workers in concentrated markets reallocate search effort away from them; second, re-application to large firms is dampened; third, within-firm wages fall with distance to market boundaries, and, fourth, productivity pass-through is attenuated where cross-market search is more costly. Each prediction holds in the data. Effective labor market monopsony exposure is thus jointly determined by local market structure and workers' search-effort reallocation.

**Keywords:** monopsony; labor market boundaries; granular search; spatial search; concentration.
**JEL Codes:** E24, J64, J31, J42.

*We thank Katarina Borovickova, Pierre-Daniel Sarte, Chen Yeh and seminar participants at the Federal Reserve Bank of Richmond and Society for Economic Dynamics 2025 for helpful comments. Moxie Clifford provided excellent research assistance. Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the Federal Reserve Bank of Richmond, nor the Federal Reserve System.

# 1 Introduction

Employer market power is typically measured and analyzed under the assumption of fixed labor market boundaries, often approximated by geographic units such as commuting zones or counties. But how substitutable are employers within and across such units? Whether and to what extent workers consider employers across administrative lines is what ultimately determines which firms compete for any given worker. Yet direct evidence on this margin is scarce. This paper closes this gap with a two-pronged approach. We use a novel dataset of online job applications to characterize the geographic breadth of job seekers' application portfolios, both across workers and over the business cycle, and we interpret the patterns through a search-and-matching model in which monopsonistic firms compete for labor and workers direct search effort across spatially separated markets, at a distance cost.

Our analysis rests on a large dataset of online job applications and postings for hourly workers in the United States, spanning over 63 million applicant–job matched observations between 2020 and 2026 and covering more than 12 million workers. We observe full application portfolios with worker–firm linkages, ZIP-code geolocation on both sides of the market, and eight-digit occupational classification.

We use the data to document two stylized facts about the spatial breadth of job search. The first is cross-sectional. Search breadth is heterogeneous across workers: about 19% of U.S. workers systematically search in a labor market smaller than the median zipcode, while 12% apply to jobs located in an area larger than the median commuting zone. Roughly one in three workers therefore searches in a market that is smaller or larger than a typical commuting zone. This heterogeneity departs from the convention of treating markets as homogeneous and constant in their boundaries across workers.

The second stylized fact is cyclical. There is abundant evidence that the average worker searches within a narrow geographic range [Kennan and Walker, 2011, Manning and Petrongolo, 2017, Marinescu and Rathelot, 2018]. We complement this evidence by characterizing how workers' search range shifts over the cycle. In locations with accelerating unemployment, job seekers apply to farther jobs: in booming economies, 50% of workers search within roughly 6 miles of their residence, while in deteriorating labor markets this share falls to 40.6%. The "catching area" of any given job posting expands precisely when local labor markets weaken, with direct consequences for the labor supply employers face.

To interpret these facts, we propose a search-and-matching model that endogenizes labor market boundaries by combining the granular-search framework of Jarosch, Nimczik, and Sorkin [2024], in which a finite number of firms holding positive vacancy shares generates monopsony power, and the multi-market spatial search structure of Manning and Petrongolo [2017], in which workers direct effort across spatially distinct markets at a distance cost. In this environment, the wage markdown is a function of *both* the degree of within-market concentration *and* the ease of cross-market escape.

Granular firms affect workers' outside option, and hence their wages, through an embargo threat: if bargaining breaks down, the firm can withdraw its own vacancies from the worker's set of possible matches. Concentration sharpens the embargo, as fewer employers command a larger share of vacancies; easier cross-market access, by contrast, shortens the expected punishment duration. Effective monopsony exposure thus emerges from the interaction of market structure *and* worker behavior.

Relative to a fixed-market benchmark, the model delivers four testable predictions. First, workers in more concentrated home markets reallocate search effort away from them. Second, re-application to large firms is dampened. Third, within-firm wages fall with distance to market boundaries. Fourth, productivity pass-through into wages is attenuated where cross-market search is more costly. A key feature of these predictions is that they concern *applications*, rather than realized wages or employment: where workers send their job search effort, how often they re-apply to the same firm, how application portfolios respond to home-market concentration. Our data are uniquely suited to test them.

We bring each prediction to the data and find support for all four. First, job search varies systematically with home-market concentration: workers in more concentrated commuting zones send a larger share of their applications outside the home CZ. Second, re-applications to large firms are mechanically more frequent in concentrated markets, where these firms hold a larger share of vacancies, but the rate is dampened relative to a random-search benchmark, and the dampening is sharpest at the largest firms—a pattern consistent with a firm-level embargo whose grip tightens with vacancy share. Third, like much of the literature, we find a decreasing reduced-form relationship between wages and market concentration [Azar, Marinescu, and Steinbaum, 2022, Berger, Herkenhoff, and Mongey, 2022]; we further show that the within-firm wage gradient with respect to distance from the CZ boundary steepens with concentration, so that establishments further from the CZ boundary pay lower wages than establishments of the same firm located closer to the boundary. Fourth, productivity pass-through into wages is attenuated in concentrated CZs and more so where cross-market search is more costly, consistent with higher markdowns limiting the transmission of productivity gains into wages.

Taken together, our results imply that effective labor market monopsony exposure is jointly determined by local market structure and workers' search-effort reallocation. The standard practice of treating commuting zones or counties as homogeneous, constant labor markets misses therefore an important margin of adjustment.

**Related literature.** Our work connects to two strands of literature on the geography of labor markets. On the spatial-equilibrium side, we connect to work on the importance of cross-geography flows for a variety of outcomes: Monte, Redding, and Rossi-Hansberg [2018] show that the employment impact of local demand shocks depends on commuting connectivity, while Gaubert, Kline, Vergara, and Yagan [2025] show that optimal place-based transfers depend sharply on migration responsiveness; Yagan [2019] and Autor, Dorn, and Hanson [2013] document the persistence of local shocks. Relatedly, Bilal [2023] traces persistent regional disparities to the interaction of location choice and search frictions. Our work focuses on how workers allocate their job search efforts over space, and how costly cross-market applications interact with search frictions.

The spatial aspect of job search is a common theme in migration and labor literatures, too. Kennan and Walker [2011] is a seminal paper establishing that migration is costly but income-responsive. More focused on job search, Marinescu and Rathelot [2018] and Schmutz and Sidibé [2019] use online and administrative data to estimate a large distance elasticity of applications, while Manning and Petrongolo [2017] document that most workers search locally using detailed data from the U.K. Faberman and Kudlyak [2019] document heterogeneity in search intensity using the same data source on hourly workers as we do. Our paper advances this line of inquiry by documenting that the spatial breadth of search responds endogenously to local market structure and to aggregate conditions.

We also contribute to a growing body of work studying labor market monopsony in quantitative macroeconomic models. Empirically, Yeh, Macaluso, and Hershbein [2022] estimate substantial wage markdowns in U.S. manufacturing, with workers paid on average 65% of their marginal product. Berger, Herkenhoff, and Mongey [2022] embed nested-CES labor supply with finite firms in a general equilibrium framework and quantify the macroeconomic consequences of such markdowns. Subsequent work by Berger, Hasenzagl, Herkenhoff, Mongey, and Posner [2025] and Hurst, Kehoe, Pastorino, and Winberry [2025] extends this framework to study minimum-wage policy and labor market dynamics. Jarosch, Nimczik, and Sorkin [2024] provide an alternative framework to rationalize markdowns through a tractable granular-search model with finite firms in a single market and show that concentration depresses wages through a self-competition channel. Our contribution is to embed the Jarosch, Nimczik, and Sorkin [2024] embargo mechanism in a multi-market environment with distance frictions.[1]

**Roadmap.** The paper is organized as follows. Section 2 describes the data and illustrates cross-sectional and cyclical variation in the spatial breadth of job search. Section 3 develops the model. Section 4 presents a two-market quantitative illustration and outlines a road map for structural estimation. Section 5 develops the model's four core predictions and brings them to the data. Section 5.6 discusses implications for the measurement of monopsony power and for local labor-market policy. Section 6 concludes.

[1]It is worth mentioning that our model draws on a tractable formulation of the bargaining environment close in spirit to Blanchard and Diamond [1994] and Fernández-Villaverde, Mandelman, Yu, and Zanetti [2021]. The latter study a self-competition mechanism in which firms forgive past bargaining breakdowns with some exogenous probability, a feature we combine with multi-market search to obtain our cross-market escape channel.

## 2 The spatial breadth of job search

### 2.1 Data

Our primary data come from the largest U.S. job-search and application website specialized in hourly work, with a focus on low- and middle-wage occupations. The platform is free for job seekers and paid for by job-posting employers. We observe the universe of applications submitted on the platform between January 1, 2020 and April 20, 2026, a window that spans the COVID-19 recession, the subsequent tight-labor market period, and the cooling that followed. The raw data cover approximately 30 million job postings and about 12 million applicants, of whom roughly 70% are repeat users (defined as having submitted at least two applications). After dropping single-application applicants, we work with over 63 million applicant-job posting matched observations.

For each application we observe a precise location on both sides of the market—residential ZIP code for the applicant and work-premises ZIP code for the job—together with a detailed job title (eight-digit SOC).[2] The level of detail in ZIP-code geocoding, combined with the platform's coverage of hourly work, makes the data well-suited to studying the spatial breadth of search of low- and middle-wage workers, a population for which both limited employer mobility and higher wage markdowns are of special concern.

A brief description of the platform helps clarify what is captured by the data, and what is not. On the employer side, the platform offers a structured posting process. Firms move sequentially through company information, job details, and qualification specifications. On the job-seeker side, applicants create free profiles with personal information. They can search for jobs using keywords and location filters. The platform also offers algorithmic recommendations based on profile information, account activity, and local demand. Two application methods are available: external links (the applicant is redirected to the employer's site) or platform-internal profiles. We observe the universe of applications, regardless of method. On the processing side, the platform tracks applications for job seekers. The dataset available to us is the result of this tracking: application histories by individual linked to the jobs applied for, with longitudinal links for both the job seeker and the employers posting the jobs. Unfortunately, we do not observe the eventual hire decision, nor do we have access, at this time, to demographic or skill information. The final dataset comprises over 60 million applicant-job posting matched observations spanning the entire U.S.[3]

[2]Throughout, we trim the top 0.5% of individual application volumes and the top 0.5% of geographic-distance applications. These are often third-party applicants, for instance staffing agencies using automatic application bots. Robustness exercises use alternative thresholds and the specific choice is largely immaterial.

[3]We complement the platform data with administrative and survey sources. We use the Quarterly Census of Employment and Wages (QCEW) for industry employment shares at the CZ level, the Occupational Employment and Wages Statistics (OEWS) for occupational employment and wages, the BLS LAUS series for local unemployment rates, BLS JOLTS-derived figures for vacancies at the industry or state level, and national productivity growth data from BLS KLEMS. BLS JOLTS and QCEW are used chiefly for benchmarking the data, but also for our empirical exercises. We construct two Bartik shocks at the CZ level from QCEW and KLEMS, which we use in Section 5 to test for heterogeneity in productivity pass-through across markets.

### 2.1.1 Sample construction

The estimation sample retains applicants who satisfy four conditions: (i) at least two applications over the observation window, ensuring that we can construct a meaningful application portfolio per applicant; (ii) non-missing residential address; (iii) non-missing job location; and (iv) non-missing eight-digit SOC for the job (a crucial control variable).

The applicant population is broadly representative of U.S. low- and middle-wage hourly workers. Applicants are concentrated in California, Florida, and the Northeastern megalopolis—a pattern that mirrors U.S. population distribution—though the platform also has substantial coverage in less densely populated regions, including Alaska, Hawaii, and Puerto Rico.

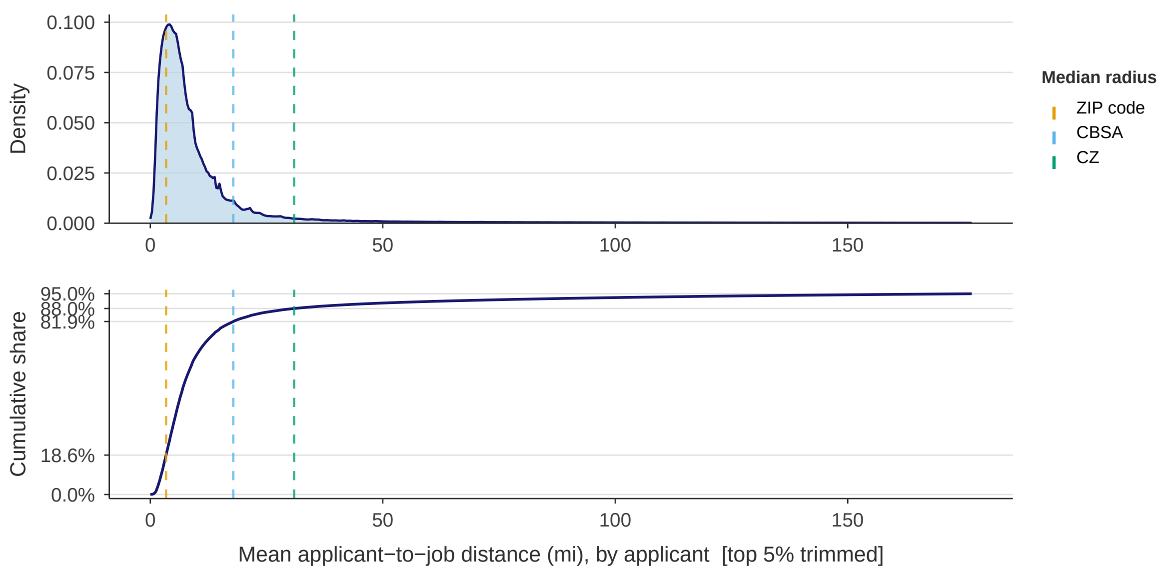


Figure 1: Empirical distribution of applicant-to-job distance per applicant. Empirical CDF (black) and PDF (blue). Top 5% trimmed for visibility. About 19% of individuals apply within an area smaller than the median ZIP code radius; about 12% to an area larger than the median commuting zone radius.

## 2.2 Variation in labor market "leakage"

The literature has long recognized that workers' applications are highly localized—a fact our data reproduce. The median worker's search radius in our data is 6.89 miles, comparable to a typical commute distance. However, the cross-sectional distribution is wide. Let an individual $p$'s search radius $\bar{d}_p$ be the average geographic distance, in miles, between the applicant's residential ZIP code and the work ZIP code of each job in her application set. Figure 1 displays the empirical CDF of $\bar{d}_p$, trimmed at the top 5% for visibility. There is significant mass at the tails. About 19% of individuals apply within an area *smaller* than the median ZIP-code radius, and roughly 12% apply to an area *larger* than the median commuting-zone radius. It follows that identifying the labor

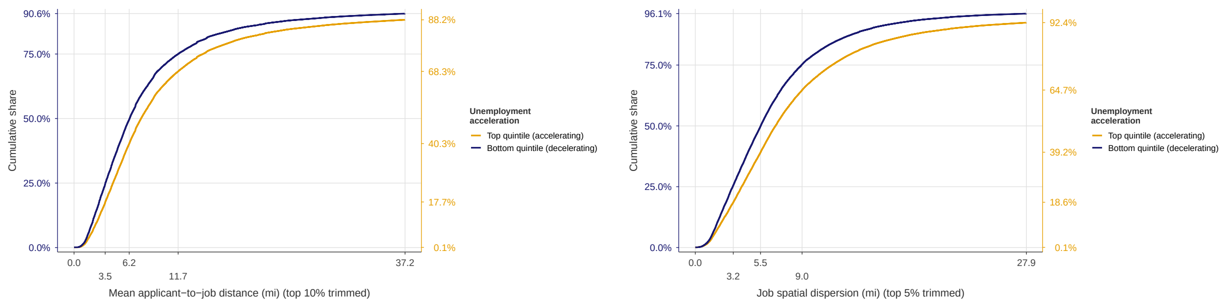


Figure 2: Empirical CDF of applicant-to-job distance for CZs with unemployment rates monthly changes below versus above the 75th percentile of the national distribution. In booming labor markets, 50% of workers search within 6.23 miles of their location; in deteriorating markets, only about 40.6% do.

market with the commuting zone, a modal choice in quantitative spatial macro models, leaves a non-trivial fraction of search outside (or well inside!) the market definition.

### 2.3 Leakage rises in recessions

Not only is the geographic breadth of search heterogeneous across workers, but it also varies systematically with aggregate conditions. We find it is countercyclical. We split commuting zones by their unemployment rate change relative to the 75th percentile of the national distribution. In CZs with a tightening labor market, 50% of workers search for jobs within roughly six miles of their residence. In deteriorating labor markets, however, where unemployment is rapidly increasing, the corresponding share falls to about 40.6%. More workers reach farther from home when the local market is weak. Figure 2 shows the empirical CDFs across the two groups.

These two facts—cross-sectional heterogeneity in search radius, and cyclical variation—motivate a theoretical framework in which the boundary of the labor market is itself an equilibrium object.

## 3 A model of granular firms and costly cross-market search

This section presents a search and bargaining model that builds on the granular-search framework of Jarosch, Nimczik, and Sorkin [2024], in which a finite number of firms, each holding a positive share of vacancies, generates monopsony power. Our model also incorporates the multi-market search structure of Manning and Petrongolo [2017], in which workers direct search across spatially separated labor markets. Concentration matters in this environment through an *embargo* mechanism. When bargaining with a firm breaks down, that firm may refuse to rehire the worker on later encounters, depressing her outside option. We characterize how market concentration shapes wages, search intensity, and the cross-market allocation of effort.

### 3.1 Environment

Time is discrete and the horizon is infinite. Workers and firms share the discount factor $\beta \in (0, 1)$. The economy has $M$ labor markets indexed by $m = 1, \ldots, M$, each populated by $N_m$ symmetric firms. A market with small $N_m$ is *concentrated*; cross-market variation in $N_m$ identifies the model's predictions in the data. Markets are otherwise connected through a single homogeneous good produced and consumed across all of them, so output in any market enters a common resource constraint.

Workers come in $M$ types indexed by $k \in \{1, \ldots, M\}$, with $L_k$ workers of type $k$. A worker's type is her *home market*, the market in which she lives and at which her search bears no distance disutility. Effort directed at non-home markets carries a distance disutility introduced below.

A filled match in market $m$ produces output

$$y_m \;=\; z \cdot x_m,$$

where $z > 0$ is aggregate productivity, common across markets, and $\{x_m\} > 0$ are market-specific productivity shifters so that productivity and concentration enter the analysis as separable drivers of cross-market wage variation. Matches dissolve exogenously at rate $\delta \in (0, 1)$ each period, returning the worker to unemployment, where she collects flow value $b$.

### 3.2 Search and matching

A type-$k$ unemployed worker chooses search effort $e_m^k \geq 0$ in market $m$, equal to the number of applications she sends there. Let $u_m = \sum_k e_m^k \, u_k L_k$ denote the effort-weighted mass of unemployed workers searching in $m$ (equivalently, the total application pool). Total vacancies are $v_m$, and market tightness is $\theta_m = v_m/u_m$. Each market $m$ uses an urn-ball matching technology. Each application is sent to a random vacancy in $m$ (this approximation, while stark, is not necessarily too limiting when it comes to application choices of low- to middle-wage workers). Workers thus direct search *across* markets through their effort allocation, but search within a market is undirected.

Because applications cannot be coordinated, the number received by any given vacancy is Poisson with mean $1/\theta_m$. The probability a vacancy receives zero applications is

$$q_m \;\equiv\; e^{-1/\theta_m},$$

and the per-vacancy match rate is $1 - q_m$. Equivalently, $q_m$ is the probability that a worker applying to a given vacancy is its sole applicant.

Following Jarosch, Nimczik, and Sorkin [2024], we focus on the symmetric vacancy-posting equilibrium, in which each firm posts $v_m/N_m$ vacancies. With non-strategic firms, every vacancy yields the same expected return regardless of which firm posts it, leaving the apportionment of $v_m$ across the $N_m$ firms indeterminate. Symmetric posting fixes the firm-level share at

$$f_m = \frac{1}{N_m},$$

the object that drives the embargo mechanism's bite described below.

If a vacancy receives multiple applications, the firm selects one applicant at random and Nash-bargains with her over the wage. Following Blanchard and Diamond [1994], a no-commitment assumption rules out bidding among applicants for the same vacancy, and a no-side-payments assumption rules out up-front transfers, so the bilateral Nash solution is well-defined ex post. While firms still select hires vacancy-by-vacancy, their memory of past bargaining breakdowns is shared across all of their vacancies: this is the embargo mechanism.

The per-application hire probability is

$$\pi_m \equiv \theta_m\bigl(1 - e^{-1/\theta_m}\bigr).$$

A worker sending a single application matches with probability $\pi_m$, recovering the standard urn-ball job-finding rate [Shimer, 2005]. Following Manning and Petrongolo [2017], we work in a regime where each worker matches in at most one market per period.[4] Workers take $\theta_m$ (and hence $\pi_m$) as given; the congestion externality of aggregate effort on market tightness is not internalized at the individual level. Under this regime, the per-period match probability for $e^k_m$ applications takes the closed form

$$\lambda_m(e^k_m, \theta_m) = 1 - \exp(-\pi_m e^k_m).$$

Because a worker can hold at most one job at a time, the gains from additional effort are concave: each marginal application is valuable only in the event that all prior ones failed.

Search effort enters the worker's problem at a convex cost, $\frac{\psi}{2}\bigl(\sum_m e^k_m\bigr)^2$. This form treats total search as a scarce resource: pushing more applications into one market raises the marginal cost of searching elsewhere, generating substitution across markets. It can be rationalized as limited attention span, for instance, so that distinguishing the $n$-th job posting becomes harder for workers the higher the number of jobs they have already sorted through. Effort directed at non-home markets carries a pairwise distance disutility $d(k, m)\, e^k_m$, with $d(k, k) = 0$, a reduced-form way to model the additional information costs involved in finding out work conditions away from one's home location. In our two-market quantitative illustration below ($M = 2$), the distance cost is symmetric: $d(k, m) = \kappa$ for $m \neq k$.

[4] In other words, $\sum_m \lambda^k_m \leq 1$ holds in equilibrium, and we verify numerically that the constraint binds with slack in our comparative statics.

### 3.3 Embargo mechanism

The embargo is the model's granular-search departure from anonymous random search and the channel through which concentration enters wages [Jarosch, Nimczik, and Sorkin, 2024]. Without it, a worker bargaining with a firm in market $m$ would treat future encounters with that firm's vacancies as part of her outside option: should this bargain fail, she could re-apply to the same firm at a later vacancy and recover much of her continuation value. With finitely many firms, the bargaining firm competes against itself, propping up the worker's threat point and squeezing its own surplus. A continuum of firms makes this immaterial—each firm has zero share in the outside option—but with finite $N_m$ the self-competition bites. The embargo lets a granular firm exclude itself from its own future competition.

Formally, when bargaining between a type-$k$ worker and a firm in market $m$ breaks down, with probability $\varphi \in [0, 1]$ the firm retains the worker's identity and refuses to rehire her on subsequent encounters. With the complementary probability the breakdown is forgotten, instead, and the worker returns to the normal unemployment state. The embargo binds firm-by-firm. An embargoed worker's applications to market $m$ split according to the firm-level vacancy share $f_m = 1/N_m$. With probability $f_m$ an application lands at the embargoing firm, where the firm enforces the embargo by selecting another applicant (unless the embargoed worker is the sole applicant). The embargo is therefore enforced with probability $1 - q_m$, where $q_m = e^{-1/\theta_m}$. With probability $1 - f_m$ the application reaches a non-embargoing firm and is treated under the standard urn-ball protocol at per-application rate $\pi_m$.

The per-application hire rate under embargo is

$$\pi_m^d \;=\; (1 - f_m)\,\pi_m + f_m\, q_m,$$

strictly below $\pi_m$ since $q_m < \pi_m$ for $\theta_m > 0$. The worker's embargoed contact rate in market $m$ is therefore $\lambda_m^{d,k} = 1 - \exp(-\pi_m^d e_m^k)$; her contact rates in all other markets are unchanged, as the embargo binds only at the firm that imposed it.

The parameter $\varphi$ is a reduced-form primitive capturing the strength of firm memory—equivalently, the firm's willingness to enforce the embargo on a worker it previously failed to hire. At $\varphi = 0$, firms have no memory: re-encounters are treated as fresh interactions, the embargo never binds, and the model collapses to standard anonymous random search with no monopsony from this channel. At $\varphi = 1$, firms always remember and always punish, recovering the full embargo limit of Jarosch, Nimczik, and Sorkin [2024]. Intermediate values represent imperfect memory, partial enforcement, or any reduced-form interpretation that maps to a probabilistic punishment. This formulation is closely related to the "forgiveness" parameter of Fernández-Villaverde, Mandelman, Yu, and Zanetti [2021], who study a similar self-competition mechanism.

Concentration sharpens the embargo. As $N_m$ falls, $f_m = 1/N_m$ rises: a larger share of an embargoed worker's applications lands at the embargoing firm and is subject to the sole-applicant rule, so $\pi_m^d$

falls and the embargoed contact rate drops further. The harder the embargo bites, the lower the worker's punishment-state value $U^k_{d,m}$, the lower her threat point in bargaining, and the lower the wage. Concentrated markets generate larger embargo wage penalties.

It is worth noting that, in this framework, firms take both market tightness $\theta_m$ and their own vacancy share $f_m$ as given. In particular, they do not internalize that increasing their share would sharpen the embargo's bite against workers they have previously failed to hire. Granularity operates entirely on the worker's side, therefore, through threat-point compression. In addition, the embargo is single-period only: any contact in any market releases the worker, and all subsequent encounters with the previously embargoing firm are treated as fresh [Jarosch, Nimczik, and Sorkin, 2024]. This tractability assumption bounds the worker's punishment history and prevents the state space from growing in the number of past breakdowns.

### 3.4 Value functions

This subsection lays out the value functions and derives the Nash-bargained wage that links them. We take the worker side first—employment, unemployment, and the embargoed state—before turning to the firm.

**Employment value.** A type-$k$ worker employed in market $m$ at the Nash-bargained wage $w^k_m$ collects $w^k_m$ each period until the match exogenously dissolves at rate $\delta$, returning her to normal unemployment with value $U^k$. Her value satisfies

$$W^k_m \;=\; w^k_m + \beta\big[(1-\delta)\,W^k_m + \delta\,U^k\big] \quad\Longrightarrow\quad W^k_m \;=\; \frac{w^k_m + \beta\delta\,U^k}{1-\beta(1-\delta)}.$$

Note that $W^k_m$ varies across markets only through the wage $w^k_m$: firms within a market are symmetric and pay the same Nash-bargained wage to a given type, and firm identity matters only for tracking which firm holds an embargo against a worker. Second, exogenous separation returns the worker to $U^k$ rather than to an embargoed state: the embargo triggers on bargaining breakdowns, not on involuntary dissolutions.

**Unemployment value.** An unemployed type-$k$ worker collects flow value $b$ and chooses search effort $\{e^k_m\}_m \geq 0$, paying convex cost $\frac{\psi}{2}\big(\sum_m e^k_m\big)^2$ and pairwise distance disutility $\sum_m d(k,m)\,e^k_m$. With probability $\lambda^k_m$ she contacts a vacancy in market $m$ and enters employment at value $W^k_m$; with the residual probability $1-\sum_m \lambda^k_m$ she remains unemployed. The Bellman is

$$U^k \;=\; \max_{\{e^k_m \geq 0\}} \; b - \tfrac{\psi}{2}\Big(\sum_m e^k_m\Big)^2 - \sum_m d(k,m)\,e^k_m + \beta\Big[\sum_m \lambda^k_m\,W^k_m + \big(1-\sum_m \lambda^k_m\big)U^k\Big]. \tag{1}$$

Match surplus is positive under Nash bargaining, so the worker accepts every offer.

**Embargoed unemployment value.** A type-$k$ worker enters the embargoed state $U^k_{d,m}$ following a bargaining breakdown when the firm enforces the embargo, which it does with probability $\varphi$. In this state, her contact rate in market $m$ falls from $\lambda^k_m$ to $\lambda^{d,k}_m = 1 - \exp(-\pi^d_m e^k_m)$ with $\pi^d_m < \pi_m$, while contact rates in other markets are unchanged at $\lambda^k_j$. Any contact, in $m$ or elsewhere, releases the embargo, and the worker enters employment at the normal-state value $W^k_j$.

Embargoed workers do not re-optimize effort. Because match surplus is positive, the embargo state is off-path and $U^k_{d,m}$ enters the model only as the worker's threat point. We therefore evaluate $U^k_{d,m}$ at the normal-state effort optimum. The Bellman is

$$U^k_{d,m} \;=\; b - \tfrac{\psi}{2}(E^k)^2 - \sum_j d(k,j)\, e^k_j + \beta\Big[\lambda^{d,k}_m\, W^k_m + \sum_{j\neq m} \lambda^k_j\, W^k_j + \big(1 - \lambda^{d,k}_m - \sum_{j\neq m} \lambda^k_j\big)\, U^k_{d,m}\Big],$$

where $E^k = \sum_j e^k_j$. The first continuation term is release through a contact in market $m$ at the reduced rate, the second is release through any other market at the normal rate, and the third is remaining embargoed for another period. Solving the linear equation,

$$U^k_{d,m} \;=\; \frac{b - \frac{\psi}{2}(E^k)^2 - \sum_j d(k,j)\, e^k_j + \beta\big[\lambda^{d,k}_m\, W^k_m + \sum_{j\neq m} \lambda^k_j\, W^k_j\big]}{1 - \beta\big(1 - \lambda^{d,k}_m - \sum_{j\neq m} \lambda^k_j\big)}.$$

**Lemma 1** (Embargo lowers value)**.** *At any interior match ($S^k_m > 0$) with $\varphi > 0$ and $\lambda^{d,k}_m < \lambda^k_m$, we have $U^k_{d,m} < U^k$.*

*Proof sketch.* Differencing the two Bellmans and applying the Nash relation reduces the gap $U^k - U^k_{d,m}$ to a positive multiple of $\beta(\lambda^k_m - \lambda^{d,k}_m)\,\eta S^k_m$. Both $\lambda^k_m - \lambda^{d,k}_m$ and $S^k_m$ are positive under the lemma's hypotheses, so the gap is strictly positive. See Appendix A for the full derivation. □

The gap $U^k - U^k_{d,m}$ gives the embargo bite at the bargaining stage: the worker's threat point falls below $U^k$ by $\varphi(U^k - U^k_{d,m})$, depressing the wage in proportion.

**Filled job value.** A firm matched with a type-$k$ worker in market $m$ collects flow profit $y_m - w^k_m$ each period until the match exogenously dissolves at rate $\delta$, after which the firm reverts to posting a vacancy with no flow profit. Its value is

$$J^k_m \;=\; (y_m - w^k_m) + \beta(1-\delta)\, J^k_m \quad\Longrightarrow\quad J^k_m \;=\; \frac{y_m - w^k_m}{1 - \beta(1-\delta)}.$$

There is free entry and the value of a vacancy is zero in equilibrium. The embargo selection rule is costless to implement: when the embargoed worker is the sole applicant, the firm's outside option is an unfilled vacancy at zero profit, so it strictly prefers to hire her despite the embargo. The embargo therefore requires no commitment power on the firm's part—only that the firm recognize which workers it has previously punished.

**Nash bargaining.** When a type-$k$ worker contacts a firm in market $m$, the wage is set by Nash bargaining with worker share $\eta$. The worker's threat point combines the normal- and embargo-state values weighted by the embargo probability:

$$\widetilde{U}_m^k \;=\; (1-\varphi)\,U^k + \varphi\,U_{d,m}^k.$$

With probability $\varphi$ a bargaining failure triggers an embargo; with probability $1-\varphi$ the worker returns to normal unemployment. The firm's threat is zero: an unfilled vacancy yields no flow profit. Total match surplus is

$$S_m^k \;=\; W_m^k + J_m^k - \widetilde{U}_m^k \;=\; \frac{y_m + \beta\delta U^k}{1-\beta(1-\delta)} - \widetilde{U}_m^k,$$

and the Nash split allocates fraction $\eta$ to the worker and $1-\eta$ to the firm:

$$W_m^k - \widetilde{U}_m^k \;=\; \eta\,S_m^k, \qquad J_m^k \;=\; (1-\eta)\,S_m^k.$$

Solving for the wage,

$$w_m^k \;=\; \eta\,y_m + (1-\eta)\big[(1-\beta(1-\delta))\,\widetilde{U}_m^k - \beta\delta\,U^k\big]. \tag{2}$$

The threat point $\widetilde{U}_m^k$ is the only object in the wage equation that responds to the embargo. The next two results characterize this response.

**Lemma 2** (Embargo depresses wages)**.** *Under the hypotheses of Lemma 1 and with $\varphi > 0$, the Nash wage $w_m^k$ falls below the no-embargo benchmark $w_m^{k,0} \equiv \eta\,y_m + (1-\eta)(1-\beta)\,U^k$ by*

$$w_m^k - w_m^{k,0} \;=\; (1-\eta)\,(1-\beta(1-\delta))\,\varphi\,\big(U_{d,m}^k - U^k\big) \;<\; 0.$$

*Proof.* At $\varphi = 0$, $\widetilde{U}_m^k = U^k$ and the wage equation collapses to $w_m^{k,0} = \eta y_m + (1-\eta)((1-\beta(1-\delta)) - \beta\delta)U^k = \eta y_m + (1-\eta)(1-\beta)U^k$. Differencing and substituting $\widetilde{U}_m^k - U^k = \varphi(U_{d,m}^k - U^k)$ yields $w_m^k - w_m^{k,0} = (1-\eta)\,(1-\beta(1-\delta))\,\varphi\,(U_{d,m}^k - U^k)$. Strict negativity follows from $U_{d,m}^k < U^k$ (Lemma 1). □

**Corollary 1** (Wage penalty rises with concentration)**.** *Under the hypotheses of Lemma 2 and holding $U^k$ fixed (partial equilibrium), the wage penalty magnitude $w_m^{k,0} - w_m^k$ is strictly decreasing in $N_m$ and strictly increasing in concentration $1/N_m$.*

*Proof.* By Lemma 2, $w_m^{k,0} - w_m^k = (1-\eta)\,(1-\beta(1-\delta))\,\varphi\,(U^k - U_{d,m}^k) > 0$. With $U^k$ held fixed, the sign of $\partial(w_m^{k,0} - w_m^k)/\partial N_m$ is the sign of $-\partial U_{d,m}^k/\partial N_m$. Appendix A establishes $\partial U_{d,m}^k/\partial N_m > 0$: a larger $N_m$ lowers the firm-level vacancy share $f_m = 1/N_m$, raises the embargoed contact rate $\pi_m^d$, and slackens the punishment. The penalty therefore decreases in $N_m$ and increases in $1/N_m$. □

The corollary identifies the model's central monopsony channel. In more concentrated markets, an embargoed worker is more likely to re-encounter the embargoing firm; her embargoed contact rate $\pi_m^d$ is lower, $U_{d,m}^k$ falls further below $U^k$, the threat point $\widetilde{U}_m^k$ compresses more, and through the wage equation the Nash wage drops more. Concentration enters wages here entirely through the embargoed value $U_{d,m}^k$; productivity $y_m$ and the normal-state value $U^k$ are held fixed across the comparative static.

### 3.5 The concentration parameter $\tau_m^k$

To summarize the wage markdown, define

$$\Delta_m \;\equiv\; \lambda_m^k - \lambda_m^{d,k}, \qquad \Lambda_d^{m,k} \;\equiv\; \sum_j \lambda_j^k - \Delta_m \;=\; \lambda_m^{d,k} + \sum_{j\neq m} \lambda_j^k.$$

Algebra (detailed in Appendix A) yields

$$U^k - U_{d,m}^k \;=\; \tau_m^k \cdot \big(W_m^k - U^k\big), \qquad \tau_m^k \;\equiv\; \frac{\beta\,\Delta_m}{1-\beta\big(1-\Lambda_d^{m,k}\big)}.$$

The sufficient statistic $\tau_m^k$ has two factors. The numerator $\beta\Delta_m$ captures the per-period contact-rate penalty from the embargo: how much harder it is to match in market $m$ once embargoed. This penalty rises with $f_m = 1/N_m$ and with the gap $\pi_m - q_m$. The denominator $1-\beta(1-\Lambda_d^{m,k})$ captures the expected punishment duration: the inverse total contact rate while embargoed. Better cross-market access (higher $\sum_{j\neq m}\lambda_j^k$) raises $\Lambda_d^{m,k}$, shortens the embargo, and lowers $\tau_m^k$. We call this the *cross-market escape* channel: it is the mechanism through which spatial frictions interact with within-market concentration.

The wage markdown then has a one-line representation:

$$w_m^k - w_m^{k,0} \;=\; -(1-\eta)\,(1-\beta(1-\delta))\,\varphi\,\tau_m^k\,(W_m^k - U^k) \;<\; 0. \tag{3}$$

Three predictions follow, in partial equilibrium and evaluated at the normal-state effort optimum (the convention used throughout Appendix A). First, higher concentration ($N_m \downarrow$) raises the firm-level vacancy share $f_m$, increases the contact-rate penalty $\Delta_m$ and the sufficient statistic $\tau_m^k$, and depresses wages. Second, holding $\tau_m^k$ fixed, higher punishment risk ($\varphi \uparrow$) scales the markdown linearly and amplifies it most where $\tau_m^k$ is already large. Third, primitives that ease cross-market access—a lower distance cost $\kappa$, higher competition elsewhere ($N_{j\neq m} \uparrow$), or higher tightness elsewhere ($\theta_{j\neq m} \uparrow$)—raise the embargoed worker's total contact rate $\Lambda_d^{m,k}$, shorten the expected punishment duration, lower $\tau_m^k$, and shrink the markdown.

### 3.6 Search effort: optimum and reallocation

Section 3.5 showed that concentration depresses wages. Workers respond by reallocating search effort across markets. We now characterize that response, first deriving the worker's first-order condition, then specializing to the two-market case to obtain transparent comparative statics, and finally stating the main proposition on the reallocation of effort.

**First-order condition.** The worker's effort choice trades the convex cost of additional applications against the embargo-compressed marginal benefit. Substituting the Nash share $W_m^k = \widetilde{U}_m^k + \eta S_m^k$ into the unemployment Bellman (1) and differentiating with respect to $e_m^k$ yields

$$\psi\, E + d(k,m) \;=\; G_m^k \,\exp(-\pi_m e_m^k), \qquad E \equiv \sum_{m'} e_{m'}^k, \tag{4}$$

where

$$G_m^k \;\equiv\; \beta\,\pi_m\left(W_m^k - U^k\right) \;=\; \underbrace{\beta\,\pi_m\,\eta\,S_m^k}_{\text{Nash surplus share}} \;+\; \underbrace{\beta\,\pi_m\,\varphi\,(U_{d,m}^k - U^k)}_{\text{embargo correction} \,\leq 0}$$

is the marginal benefit of an application in market $m$ at $e_m^k = 0$. The left-hand side is the marginal cost: a convex cost shared across markets through total effort $E$, plus a local distance disutility. The right-hand side is the congestion-adjusted marginal benefit, decreasing in $e_m^k$ because each additional application is more likely to overlap with one the worker has already sent.

The decomposition of $G_m^k$ is where concentration enters the effort decision. The first term is the standard return on a successful application. The second term is the embargo correction. Because the threat point in bargaining is $\widetilde{U}_m^k = (1-\varphi)U^k + \varphi U_{d,m}^k$ rather than $U^k$, and $U_{d,m}^k < U^k$ whenever $\varphi > 0$, the embargo compresses realized surplus and lowers $G_m^k$. Concentration thus enters the search margin through the realized surplus, not the contact technology: $\pi_m$ at the search margin is unchanged.

The interior FOC can be solved for effort in market $m$ as a function of total effort $E$:

$$e_m(E) \;=\; \max\!\left(0,\; -\frac{1}{\pi_m}\log\frac{\psi E + d(k,m)}{G_m^k}\right), \tag{5}$$

with the maximum handling corner cases ($G_m^k \leq d(k,m)$ implies $e_m = 0$). Conditional on $E$, the cross-market allocation depends only on the ratios $G_m^k/(\psi E + d(k,m))$: effort flows toward markets where the marginal benefit is high relative to the local cost. The level of $E$ is pinned down by the consistency requirement

$$\sum_m e_m(E) \;=\; E. \tag{6}$$

The $M$-dimensional optimization collapses to a one-dimensional fixed point in $E$.

### 3.7 Two-market specialization

To build intuition for the comparative statics, we specialize to $M = 2$. Index a type-$k$ worker's residence market by $h$ (home) and her other market by $a$ (away), with distance disutilities $d(k, h) = 0$ and $d(k, a) = \kappa > 0$. The FOCs become

$$\psi E \;=\; G^k_h \exp(-\pi_h e^k_h), \qquad \psi E + \kappa \;=\; G^k_a \exp(-\pi_a e^k_a).$$

At the optimum, the marginal benefit of an away application exceeds that of a home application by exactly $\kappa$. The home and away effort levels can be expressed in terms of $E$ alone:

$$e^k_h(E) \;=\; \max\Big(0,\; -\tfrac{1}{\pi_h}\log(\psi E/G^k_h)\Big), \qquad e^k_a(E) \;=\; \max\Big(0,\; -\tfrac{1}{\pi_a}\log((\psi E + \kappa)/G^k_a)\Big).$$

This is the simplest non-trivial setting in which the embargo–distance interaction can be traced analytically. The interior solution—both markets active—has an intuitive shape. As home concentration $N_h$ falls, $G^k_h$ falls (through $\partial U^k_{d,h}/\partial N_h > 0$). The FOC for the home market binds at lower $e^k_h$, releasing some effort to the away market. The convex cost $\psi(e^k_h + e^k_a)^2$ means the away FOC also tightens as $E$ shifts, capping the reallocation.

**Concentration and the allocation of search.** The main result characterizes how home-market concentration reallocates search effort across markets.

**Proposition 1** (Concentration in the home market)**.** *For a type-$k$ worker, holding $U^k$, $\{\theta_m\}$, and $\{N_m\}_{m\neq k}$ fixed (partial equilibrium), a fall in $N_k$ produces:*

1. *a decline in home effort $e^k_k$, and*
2. *a rise in effort directed at every other market $j \neq k$.*

*Proof sketch.* A fall in $N_k$ lowers $G^k_k$ via $\partial U^k_{d,k}/\partial N_k > 0$, while $G^k_j$ for $j \neq k$ is unaffected. The implicit function theorem applied to $\sum_j e_j(E) = E$ delivers $dE/dN_k > 0$, hence $de^k_j/dN_k < 0$ for $j \neq k$. For $j = k$, the direct channel (lower $G^k_k$) dominates the indirect channel (lower $E$), yielding $de^k_k/dN_k > 0$. See Appendix A. □

The economic chain runs through three steps. Greater home concentration raises $f_k$ and tightens the embargo on home-market firms; the threat point $\widetilde{U}^k_k$ compresses, shrinking the worker's Nash share of the home surplus and dragging the embargo correction in $G^k_k$ further into the red; the marginal return to a home application falls, the FOC binds at a lower $e^k_k$, and effort spills over to away markets. The outward substitution is the source of the cross-market spillovers tested in Section 5.

**Corollary 2** (Concentration in the away market)**.** *Under the hypotheses of Proposition 1, for any $j \neq k$, a fall in $N_j$ raises home effort $e^k_k$ and lowers effort directed at $j$.*

**Corollary 3** (Total effort declines with concentration)**.** *Under the hypotheses of Proposition 1, for any market $m$, a fall in $N_m$ lowers a type-$k$ worker's total search effort $E^k$.*

The convex effort cost prevents perfect substitution. When $G^k_m$ shrinks, the worker withdraws effort from $m$, but reallocating elsewhere raises $\psi E$ in every market, not just $m$. The FOCs for non-$m$ markets tighten through the higher $\psi E$, capping how much the worker reallocates. Equilibrium restores the FOCs at strictly lower total effort: reallocation is partial, and aggregate search contracts.

### 3.8 Equilibrium: firm entry, tightness, and contact-weighted exposure

To close the model we pin down tightness $\{\theta_m\}$ and unemployment $\{u_k\}$ through free entry and flow balance. This subsection also makes precise the role of the distance cost $\kappa$ at interior values and introduces the contact-weighted exposure that aggregates within- and cross-market effects of monopsony.

**Role of the distance cost $\kappa$.** The interior gradient of $\kappa$ pins down how much of the embargo's bite is absorbed by reallocation. When $\kappa$ is high, workers are effectively trapped at home: cross-market access $\Lambda^{m,k}_d$ shrinks, $\tau^k_m$ rises, and monopsony is amplified. When $\kappa$ is low, away applications are cheap, the expected punishment duration is short, and wage gaps across markets compress. The limiting cases are collected in Section 3.9.

**Free entry.** Vacancies in market $m$ are posted at flow cost $c > 0$. Under the at-most-one-match regime, each matched worker corresponds to one filled vacancy, so

$$C_m \;=\; \sum_k \lambda^k_m\, u_k L_k$$

counts total matches and total filled vacancies alike. The vacancy fill rate is $\xi_m = C_m/v_m$, and a filled vacancy matches with a type-$k$ worker with probability $s^k_m = \lambda^k_m\, u_k L_k / C_m$. Free entry equates the discounted expected return on a vacancy to its posting cost:

$$c \;=\; \beta\,\xi_m \sum_k s^k_m\, J^k_m \;=\; \beta\,\xi_m\,(1-\eta) \sum_k s^k_m\, S^k_m. \tag{7}$$

The free-entry condition plays the role of a labor-demand schedule: $\theta_m$ adjusts so that the expected return on a vacancy covers its posting cost.

**Flow-balance unemployment.** Match surpluses are positive, so workers accept every offer and type $k$'s job-finding rate equals her aggregate contact rate $\Lambda_k = \sum_m \lambda^k_m$. Equating the flow into unemployment from separations with the flow out from job finding, $\delta(1-u_k) = \Lambda_k u_k$, yields

$$u_k \;=\; \frac{\delta}{\delta + \Lambda_k}.$$

**Definition 1** (Steady-state equilibrium)**.** *A steady-state equilibrium consists of values for market tightness in each market $\{\theta_m\}_{m=1}^M$, value functions $\{U^k, U^k_{d,m}\}_{m,k}$, search efforts $\{e^k_m\}_{m,k}$, wages $\{w^k_m\}_{m,k}$, and unemployment rates $\{u_k\}_k$ such that*

1. *search efforts $\{e^k_m\}$ solve the worker's Bellman in Section 3.4, given $\{\theta_m, U^k, U^k_{d,m}\}$;*
2. *value functions $\{U^k, U^k_{d,m}\}$ satisfy the Bellmans of Section 3.4;*
3. *wages $\{w^k_m\}$ satisfy the Nash-bargaining condition (2);*
4. *free entry (7) holds in each market;*
5. *unemployment satisfies flow balance: $u_k = \delta/(\delta + \Lambda_k)$.*

**Contact-weighted exposure.** A useful aggregation of the within- and cross-market forces in equilibrium is the contact-weighted exposure

$$\mathcal{C}^k_{\text{eff}} \;=\; \sum_m \frac{\lambda^k_m}{\Lambda^k}\,\varphi\,\tau^k_m.$$

This object is endogenous: workers in concentrated home markets lower $\mathcal{C}^k_{\text{eff}}$ by shifting effort toward competitive markets. It will reappear in Section 5.6 when we discuss the welfare cost of local labor-market policies.

### 3.9 Nesting familiar models

The model nests several familiar special cases. At $\varphi = 0$ the embargo never bites: the threat point collapses to $\widetilde{U}^k_m = U^k$, the Nash wage reduces to $w^k_m = \eta y_m + (1-\eta)(1-\beta)U^k$, and wages are independent of concentration—this is standard multi-market search. At $\varphi = 1$ the embargo is deterministic, and—holding other parameters fixed—the per-unit markdown and the per-unit effort reallocation in response to home concentration are at their largest. At the other limit of the distance cost, $\kappa \to \infty$, away applications drop out ($e^k_a \to 0$) and markets become isolated islands, reducing the economy to $M$ independent single-market Jarosch, Nimczik, and Sorkin [2024] monopsony problems. The opposite limit, $\kappa = 0$, makes cross-market reallocation frictionless and lets workers fully arbitrage concentration through effort. Sending $N_m \to \infty$ for all $m$ drives $f_m \to 0$ and $\lambda^{d,k}_m \to \lambda^k_m$, so $\tau^k_m \to 0$ and the competitive Nash outcome obtains. Finally, with $M = 1$ and $\varphi = 1$ the framework nests Jarosch, Nimczik, and Sorkin [2024] directly, with $\tau = \beta\Delta/[1 - \beta(1 - \lambda^d)]$: the numerator uses the contact-rate penalty $\Delta$ while the denominator uses $\lambda^d$, reflecting release on any contact.

### 3.10 Testable predictions

The model delivers four testable predictions, which we bring to the data with U.S. commuting zones (CZs) as the unit of analysis.

P1. **Spatial reallocation.** Workers in concentrated home markets send a larger share of applications to away markets (Proposition 1 and Corollary 2).

P2. **Within-market embargo.** Within a market, workers re-apply to large firms less often than a random-search benchmark predicts, and the dampening is sharpest where firm vacancy shares are highest (firm-by-firm logic of Section 3.3).

P3. **Within-firm wages drop with distance to market boundaries.** Wages decline with concentration; conditional on the match surplus, the markdown scales linearly with $\varphi\,\tau_m^k$ (Lemma 2 and Corollary 1).

P4. **Pass-through dampening.** Productivity pass-through into wages is attenuated in concentrated markets, and the attenuation is stronger where cross-market escape is harder (high $\kappa$, low $\Lambda_d^{m,k}$), through the markdown equation (3).

The observable counterparts of the model's parameters are: $N_m$ (firm counts within CZ-quarter, summarized by HHI), $M$ (number of CZs in a worker's application portfolio), effort shares $e_m^k/E^k$ (application shares), $\kappa$ (home-market share), and $\eta$ (wage-productivity slope at low HHI).

# 4 Quantitative Illustration

We illustrate the mechanisms at work in our model numerically, in the two-market case ($M = 2$). This is a simple exercise meant to lay bare the interaction between spatial frictions and labor market monopsony.

Productivity is equalized across markets ($x_A = x_B$), so the only primitive difference between $A$ and $B$ is the number of firms $N_m$. Table 1 lists the parameter values. Conventional parameters ($\beta$, $\sigma$, $\eta$) take standard search-literature values; the model-specific parameters (c, $\varphi$, $\kappa$, $\psi$, and the firm counts $N_m$) are set illustratively to highlight the mechanisms rather than calibrated to data targets.

Table 1: Baseline parameters

| **Parameter** | **Value** | **Interpretation** |
|---|---|---|
| $\beta$ | 0.96 | Discount factor |
| $\delta$ | 0.08 | Separation rate |
| $b$ | 0.20 | Flow value of non-employment |
| $\eta$ | 0.50 | Worker bargaining power |
| $\varphi$ | 0.35 | Embargo probability |
| $\kappa$ | 0.50 | Distance cost (away search) |
| $\psi$ | 1.00 | Effort-cost scale |
| $c$ | 1.20 | Vacancy posting cost |
| $N_A, N_B$ | 2, 10 | Firms in $A$ (concentrated) and $B$ (competitive) |
| $x_A, x_B$ | 1.5, 1.5 | Market-specific productivity |
| $L_A, L_B$ | 1, 1 | Mass of workers by home market |

### 4.1 Wages and concentration

Figure 3 illustrates the two wage predictions, P3 and P4. Wages in $A$ lie strictly below wages in $B$ whenever $N_A < N_B$ and meet wages in $B$ at $N_A = N_B$ by symmetry. The embargo penalty $|w_A - w_A^0|$ rises as $N_A$ falls—equivalently, as the embargoing firm's vacancy share $f_A = 1/N_A$ rises. The penalty does not vanish at $N_A = N_B$: both markets remain subject to the embargo at $f = 1/10$, so the $A$ penalty converges to the (still-positive) $B$ penalty rather than to zero.

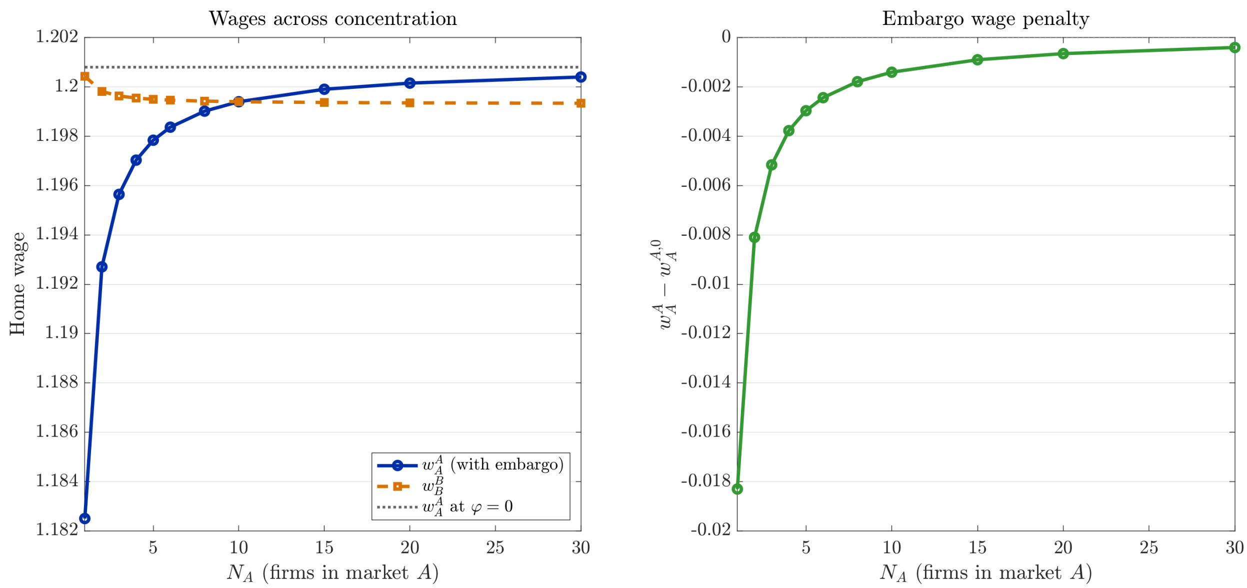


Figure 3: Wages and the embargo penalty as $N_A$ varies, holding $N_B = 10$ fixed. The wage in $A$ rises with $N_A$ and meets the wage in $B$ at $N_A = N_B$; the embargo penalty in $A$ shrinks but converges to the (still-positive) penalty in $B$, not to zero.

### 4.2 Effort reallocation

Figure 4 traces the effort response of type-$A$ workers as $N_A$ varies. The home share $\text{hs}_A = e_A^A/E^A$ falls as $A$ becomes more concentrated, with effort reallocating toward the less concentrated market $B$ (P1). Total effort $E^A$ declines in parallel: the convex cost prevents full substitution, so the contraction in home applications is not offset one-for-one in $B$. Monopsony thus operates jointly on price (the wage markdown) and quantity (effort and, in equilibrium, unemployment).

### 4.3 Discussion: From the Model to Estimation

The quantitative illustration of Section 4 is purposely stylized: it fixes parameters at values borrowed from the granular-search literature and traces out the model's comparative statics in a two-market environment with symmetric productivity. The reduced-form empirical work in Section 5 provides corroborating evidence for the model's qualitative predictions, but does not estimate the structural parameters $\{\eta, \varphi, \kappa, \psi, c, \beta\}$. Before turning to the empirics, we briefly discuss how the model's parameters map to observable moments in the data, sketching a roadmap for the structural

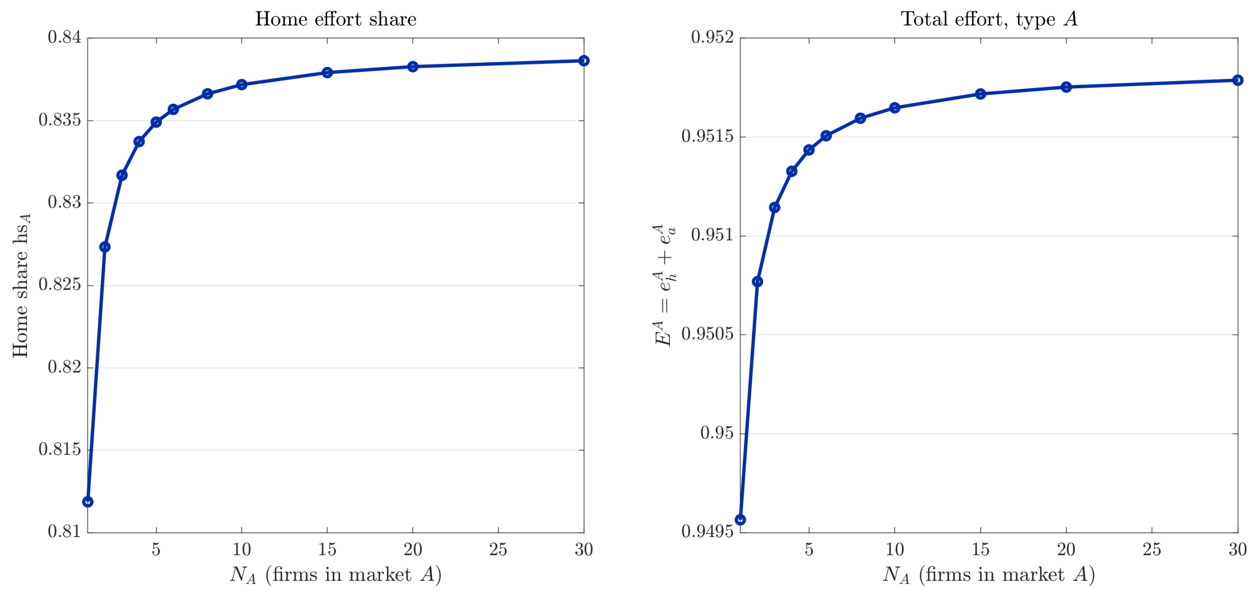


Figure 4: Effort allocation of type-$A$ workers as $N_A$ varies, holding $N_B = 10$ fixed. (Left) home share $\mathrm{hs}_A = e_A^A/E^A$ falls as $A$ becomes more concentrated (Proposition 1). (Right) total effort $E^A$ declines (Corollary 3).

exercise we plan in follow-up work.

**Identification of $\eta$ and $\varphi$.** In the limit of perfect competition ($N_m \to \infty$), the wage equation (2) reduces to $w_m^k = \eta y_m + (1-\eta)(1-\beta)U^k$. The worker bargaining share $\eta$ is then identified from the elasticity of wages to productivity in highly competitive markets. The embargo probability $\varphi$ is identified, given $\eta$, from the wage-HHI elasticity through the markdown equation (3). Joint identification of $(\eta, \varphi)$ thus requires both the level of pass-through and its interaction with HHI, which we provide in Section 5 below.

**Identification of $\kappa$ and $\psi$.** The distance cost $\kappa$ is identified from the home-share of applications among workers in home markets of varying concentration. Specifically, $\kappa$ governs the FOC wedge between home and away markets in (4): a higher $\kappa$ leads to higher home shares for any given $G^k$. Holding $G^k$ fixed (for instance by conditioning on home-market HHI), variation in away-market HHI traces out $\kappa$. The empirical out-of-CZ share of applications is the relevant moment. The effort-cost scale $\psi$ is identified from total application volume in response to home-market concentration: Corollary 3 predicts that a higher home-market HHI lowers $E^k$, and the slope of total applications on HHI (controlling for productivity) identifies $\psi$.

**Identification of $\beta$ and $c$.** The discount factor $\beta$ is conventionally calibrated to the inverse of the monthly real interest rate. The vacancy posting cost $c$ is identified from the free-entry condition (7): given the other parameters and observed $\theta_m$, the equation pins down $c$.

**Validation moments.** Following standard practice in the search-and-matching literature, we plan to validate the calibrated model against three out-of-sample moments: (i) the cyclical variation in leakage documented in Section 2; (ii) the cross-CZ relation between unemployment and HHI; and (iii) the cumulative number of distinct SOC codes per applicant by application volume. The latter is an indirect check on the model's prediction that workers expand on multiple margins simultaneously as home opportunities deplete.

# 5 Empirical Tests

We bring the model's four predictions from section 3.10 to the data, with U.S. commuting zones (CZs) as the geographic unit of analysis.

- P1. Spatial reallocation of search effort away from concentrated markets.
- P2. Application pattern are consistent with within-market embargo.
- P3. Within-firm wages drop with distance to market boundaries.
- P4. Productivity pass-through into wages is attenuated in concentrated markets.

Throughout, specifications cluster standard errors at the CZ level and include year fixed effects (with richer fixed effects where noted).

## 5.1 Measurement and mapping to the model

Before turning to the tests, we describe how the empirical objects map to model quantities. We measure concentration with the HHI of vacancy shares within a CZ-quarter, $\text{HHI}c, t = \sum_f sf, c, t^2$. The model's comparative statics are partial-equilibrium statements about how a single worker's choices respond to a change in $N_m$ holding other markets fixed; the empirical tests below compare CZs at a point in time. We interpret cross-CZ variation in HHI as proxying for variation in $1/N_m$, abstracting from sorting of workers and firms into markets, and we read the cross-CZ correlations as suggestive of the model's predictions rather than as direct estimates of the partial-equilibrium objects. Distance from the home CZ is constructed from the residential ZIP code and the geographic boundary of the worker's CZ. Specifically, for each applicant we compute (i) the geodesic distance from her residential ZIP centroid to each applied-job's ZIP centroid; (ii) the geodesic distance from her residential ZIP centroid to the nearest CZ-boundary point. The *border-excess distance* is the difference between mean applied-job distance and the distance to the CZ edge. A negative value means the typical applied job is closer than the CZ edge (interior search); positive values indicate cross-CZ search.

Re-application is identified as a binary indicator equal to one if the application in question is to a firm to which the same applicant has previously applied, within the observation window. The base rate of re-application is high—workers often apply to multiple postings at a single firm—which is why we compare to a random-search benchmark rather than to zero.

Finally, we construct productivity shocks *à la* Bartik to test the pass-through to wages. These are constructed at the CZ-year level. We use $g^{\text{Bartik}}_{cz,t} = \sum_n s_{n,cz,2019}\, g^{\text{nat}}_{n,t}$, where $s_{n,cz,2019}$ is the QCEW NAICS-2 industry $n$ share of total employment in CZ $cz$ in 2019, and $g^{\text{nat}}_{n,t}$ is national productivity growth in industry $n$ from KLEMS in year $t$.[5]

### 5.2 P1: Workers reallocate search effort away from concentrated markets

The reduced-form specification underlying P1 is

$$y_{i,t} \;=\; \alpha + \beta\, \text{HHI}_{c(i),t} + \lambda_t + \varepsilon_{i,t}, \tag{8}$$

where $y$ takes four spatial-breadth outcomes: search radius $\bar{d}_p$, application-set width $D_p$, border-excess distance (applicant-to-job minus distance from residence to CZ edge), and the share of applications sent out of the home CZ. CZ-clustered standard errors and year fixed effects throughout.

Table 2 reports results for the first three outcomes. A 1-standard-deviation increase in home-CZ HHI is associated with a +2.52 mile increase in the average applicant-to-job distance, a +0.96 mile increase in average pairwise distance across the application set, and a +1.93 mile increase in distance net of the home-CZ edge. The first two effects are about 37% and 15% of their respective sample medians. The border-excess effect is more striking: the sample median is −1.60 miles, indicating that the typical applicant searches *within* the home CZ by about 1.6 miles. The 1-SD HHI rise of +1.93 miles flips the median from inside- to outside-CZ search—a direct quantitative counterpart to Proposition 1.

Table 2: P1: Search portfolios widen geographically with home-market concentration

| | **Avg. applicant-to-job distance** | **Avg. pairwise distance, applied jobs** | **Applicant-to-job dist. minus home-to-CZ edge** |
|---|---|---|---|
| HHI ($\beta$) | +314.65*** | +120.42*** | +241.50*** |
| | (95.68) | (38.44) | (62.68) |
| 1-SD HHI effect | **+2.52 mi** | **+0.96 mi** | **+1.93 mi** |
| Sample median | 6.89 mi | 6.48 mi | −1.60 mi |
| % of \|median\| | 37% | 15% | 121% |
| $N$ | 12,289,221 | 8,366,269 | 12,201,258 |

Notes: OLS with year fixed effects; standard errors clustered at the CZ level. $^{*}p < .10,^{**}p < .05,^{***}p < .01$.

Table 3 examines the two remaining margins. A conditional-logit specification for the share of applications sent out of the home CZ finds that a 1-SD HHI rise tilts the outward share by +0.55 percentage points among workers who search out of CZ at all. A complementary CZ-year-level OLS regression in logs of total applications to CZ $c$ on $\text{HHI}_{c,t}$ finds that a 1-SD HHI increase is

[5] The 2019 base year predates the COVID-19 shock, mitigating endogenous responses in the share structure. Robustness specifications use QCEW wage growth and OEWS-platform employment growth as the national shifter.

associated with a $-15.5\%$ reduction in the log-flow of applications received. Concentrated CZs thus both lose home-CZ workers and attract fewer applicants from elsewhere—local labor supply is lower on both intensive and extensive margins. The pattern is consistent with the asymmetric effects of Proposition 1 (home workers leave) and Corollary 2 (workers in other CZs do not increase their applications to a concentrated market).

Table 3: P1: Thinner labor supply in concentrated CZs

| | **Out-of-CZ appl. share** | **Appl. to CZ, log-flow** |
|---|---|---|
| HHI (coef.) | $+2.112^{**}$ | $-1.515^{***}$ |
| | (0.988) | (0.115) |
| 1-SD HHI effect | **+0.55 pp** | **-15.5%** |
| $N$ | 1,300,705 | 3,511 |

Notes: Column 1 is a conditional logit; column 2 is OLS in logs at the CZ-year level. Year fixed effects; CZ-clustered standard errors.

### 5.3 P2: Within-market application patterns are consistent with the embargo

The embargo model predicts that workers re-apply to firms they have applied to before less often than a random-search benchmark would predict, and that the dampening is stronger at larger firms. Let $r_{i,n}$ denote the predicted re-application rate under random search to a firm of vacancy share $f_i$, with $n$ prior applications in the spell: $r_{i,n} = 1 - (1 - f_i)^n$.

Figure 5 plots the ratio of actual to predicted re-application rates by firm-size quartile. The predicted re-application rate under random search is $1 - (1 - f_i)^n$.

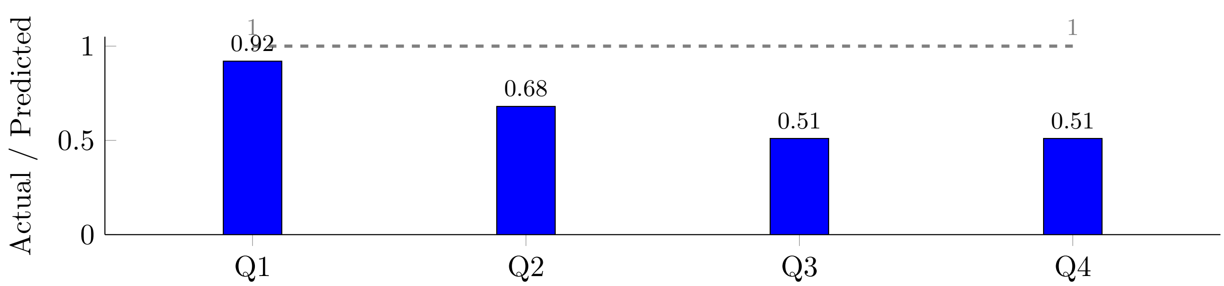


Figure 5: Re-application rates are below the random benchmark in every firm-size quartile. The gap is largest for the largest employers — consistent with a firm-level embargo that bites harder at larger firms. Firm size quartiles on the $x$-axis. Actual re-application rates over predicted on the $y$-axis. Re-application is measured across search spells, as the number of applications to an employer already applied to in the past over all applications in that spell. Q1: 0.92; Q2: 0.68; Q3: 0.51; Q4: 0.51. The dashed line is the random-search benchmark of 1.

Re-application rates are below the random benchmark in every quartile, with the gap largest at the top: 0.92 in Q1, 0.68 in Q2, 0.51 in Q3, and 0.51 in Q4. The qualitative pattern lines up with a firm-level embargo whose grip tightens with vacancy share.

**Within-market test.** We estimate a conditional-logit choice model in which each application contributes the chosen firm plus 14 randomly sampled non-chosen firms in the same CZ-quarter as alternatives [McFadden, 1978]. The model is

$$\Pr(\text{chosen}_{ijmt}) \;=\; \Lambda\big(\beta_1\,\text{reapply}_{ijmt} + \beta_2\,f_{jmt} + \beta_3\,f_{jmt}\times\text{reapply} + \alpha_i + \delta_{mt}\big).$$

Table 4 reports estimates. The main effect of having previously contacted firm $j$ is large and positive (+2.54), as expected if workers tend to return to firms they have already explored. Vacancy share enters negatively in the conditional logit, reflecting the relative attractiveness of small firms within the choice set. The interaction $f_{imt}\times$ reapply is large and negative ($-26.20$): the gradient of re-application with respect to firm size is strongly negative once selection effects are absorbed. Concentrated markets generate a mechanical tendency to revisit large firms, but the embargo correction more than offsets this tendency.

Table 4: P2: Re-application is dampened at large firms within markets

| | Pr(**chosen**) |
|---|---|
| re-apply (= 1 if prior contact) | $+2.54^{***}$ |
| | (0.008) |
| vacancy share $f_{imt}$ | $-0.87^{***}$ |
| | (0.04) |
| $f_{imt}\times$ re-apply | $\mathbf{-26.20}^{***}$ |
| | (0.38) |
| FEs | Worker + CZ×quarter |
| $N$ | 28.7M |

Notes: FE conditional logit; each application contributes the chosen firm plus 14 randomly sampled non-chosen firms in the same CZ-quarter as alternatives [McFadden, 1978]. Year, worker, and CZ-quarter fixed effects. SEs clustered at the CZ level.

**HHI-firm-share interaction.** A simpler OLS specification on the binary indicator $\text{reapply}_{ijmt}$ at the application level confirms the pattern:

$$\text{reapply}_{ijmt} \;=\; \alpha + \beta_1\,\text{HHI}_{mt} + \beta_2\,f_{jmt} + \beta_3\,\text{HHI}\times f + \lambda_t + \varepsilon_{ijmt}.$$

Table 5 reports $\beta_1 = 0.201$, $\beta_2 = 0.027$, and $\beta_3 = -0.396$, all highly significant. Re-application is mechanically more common in concentrated markets (positive $\beta_1$), but the negative interaction $\beta_3$ shows that the embargo cost is sharpest precisely where firm-level vacancy shares are high. Concentration thus interacts with firm size in the way the model predicts.

Table 5: P2: Re-application is lower for large firms in concentrated markets

| | **Re-apply** ($= 1$ **if prior contact)** |
|---|---|
| HHI | $+0.201^{***}$ |
| | (0.034) |
| vacancy share $f_{imt}$ | $+0.027^{***}$ |
| | (0.002) |
| $f_{imt}\times$ HHI | $-\mathbf{0.396}^{***}$ |
| | (0.063) |
| FEs | Year |
| $N$ | 41.9M |

Notes: OLS with year fixed effects; SEs clustered at the CZ level.

### 5.4 P3: Wages decline with concentration; within-firm wages drop with distance to the market boundary

We test P3 with two complementary specifications. The first asks whether the within-firm wage gradient with respect to distance from the CZ boundary steepens in concentrated markets:

$$\log w_{fmt} \;=\; \alpha + \beta_1\,\mathrm{dist}_{fm} + \beta_2\,\mathrm{dist}\times\mathrm{HHI} + \gamma_f + \delta_{cz,t} + \varepsilon.$$

The second asks whether tightness moderates the wage-HHI gradient—the empirical counterpart of the escape channel ($\Lambda_d^{m,k}\uparrow\Rightarrow \tau_m^k\downarrow$):

$$\log w_{fmt} \;=\; \alpha + \beta_1\,\mathrm{HHI}_{mt} + \beta_2\,\tilde{\theta}_{m,t-1} + \beta_3\,\mathrm{HHI}\times\tilde{\theta} + \gamma_f + \delta_{soc3} + \mu_{cz,t} + \varepsilon.$$

Table 6: P3: Within-firm wage gradient steepens with HHI; wage-HHI gradient is flatter where escape is easier

| | **Within-firm gradient** | **Tightness-HHI "escape"** |
|---|---|---|
| HHI | — | $-0.0245^{***}$ |
| | | (0.0093) |
| dist. to CZ edge (mi) | $-3.8\times10^{-5}$ | — |
| | $(1.9\times10^{-4})$ | |
| dist. $\times$ HHI | $-\mathbf{0.062}^{***}$ | — |
| | (0.013) | |
| $\tilde{\theta}_{t-1}$ | — | $-0.0238^{***}$ |
| | | (0.0031) |
| HHI $\times\,\tilde{\theta}_{t-1}$ | — | $+\mathbf{0.0197}^{**}$ |
| | | (0.0090) |
| FEs | Firm + CZ$\times$year | Firm + SOC3 + CZ$\times$year |
| $N$ | 3.63M | 34.9M |

Notes: Dependent variable is log-wage at the establishment-CZ-year level. CZ-clustered SEs. CZ size (number of jobs) controlled for in column 1.

Table 6 reports the two specifications. Two findings emerge.

*Within-firm gradient.* For a given firm, establishments located farther from the CZ boundary pay lower wages, and the gradient *steepens* with HHI ($\beta_2 = -0.062$, $p < 0.01$). Workers in interior locations of concentrated CZs are paid less than otherwise comparable workers near the CZ edge—consistent with the embargo bite being attenuated by easier physical access to outside-CZ alternatives.

*Escape channel.* The tightness-HHI interaction is positive ($\beta_3 = +0.0197$, $p < 0.05$): the wage-HHI gradient is flatter in tighter markets. Higher tightness raises contact rates in all markets, increasing the cross-market escape rate $\Lambda_d^{m,k}$, lowering $\tau_m^k$, and shrinking the markdown—exactly the model's escape mechanism.

**First-differences.** To isolate within-firm-within-CZ variation, we estimate

$$\Delta \log w_{fmt} \;=\; \alpha + \beta_1 \,\Delta\mathrm{HHI}_{mt} + \beta_2\,\mathrm{HHI}_{m,t-1} + \varsigma_t + \varepsilon.$$

Table 7 reports $\beta_1 = -0.161$ in the main specification and $-0.131$ with firm fixed effects, both significant at the 5–10% level. A 1-SD rise in ΔHHI is associated with about $-0.1\%$ in wage growth. First-differencing removes time-invariant firm-level and CZ-level wage heterogeneity, supporting a causal-style interpretation of P3.

Table 7: P3: Changes in HHI lower firm wages

| | **Main** | **Firm-FE robustness** |
|---|---|---|
| Δ HHI | $-\mathbf{0.161}^{**}$ | $-\mathbf{0.131}^{*}$ |
| | (0.070) | (0.067) |
| FEs | Year | Year + firm |
| $N$ | 354,591 | 354,591 |

Notes: First-differences. Lagged HHI controlled for throughout. CZ-clustered SEs.

## 5.5 P4: Productivity pass-through is dampened by concentration

The fourth prediction concerns the interaction of productivity shocks and concentration: $\partial w/\partial y$ should be smaller in concentrated markets, where the embargo absorbs a larger share of the surplus. We construct a CZ-level Bartik shock from QCEW NAICS shares times KLEMS national productivity growth, with 2019 as the base year, and estimate

$$\log w_{fmt} \;=\; \alpha + \beta_1\, g^{\mathrm{Bartik}}_{cz,t} + \beta_2\,\mathrm{HHI}_{m,t-1} + \beta_3\, g\times\mathrm{HHI} + \gamma_f + \delta_{cz} + \eta_t + \varepsilon.$$

Table 8 reports the main result. Productivity shocks raise wages on average ($\beta_1 = +0.0036$, $p < 0.05$). The interaction with lagged HHI is negative and significant ($\beta_3 = -0.052$, $p < 0.05$): pass-through is significantly smaller in concentrated CZs. Two alternative Bartik measures (QCEW wage growth and OEWS-platform employment growth) reported in the appendix yield the same sign on the interaction, though noisier with the platform-SOC weighting.

Table 8: P4: Productivity pass-through is attenuated by concentration

| | **log wage** |
|---|---|
| Bartik shock $g_{cz,t}$ | $+0.0036^{**}$ |
| | (0.0016) |
| HHI lag | $-0.038$ |
| | (0.087) |
| $g_{cz,t}\times$ HHI lag | $-\mathbf{0.052}^{**}$ |
| | (0.021) |
| FEs | Firm + CZ + year |
| $N$ | 1,209,776 |

Notes: Base shares from QCEW; shock from KLEMS national productivity growth (2019 base). CZ-clustered SEs.

The dampening of pass-through has a direct mapping to the wage equation (2). Differentiating with respect to productivity, holding $\tau_m^k$ fixed at the normal-state effort optimum (the convention used throughout the model's comparative statics),

$$\frac{\partial w_m^k}{\partial y_m} = \frac{\partial w_m^{k,0}}{\partial y_m} - (1-\eta)\,(1-\beta(1-\delta))\,\varphi\,\tau_m^k \cdot \frac{\partial (W_m^k - U^k)}{\partial y_m},$$

where the second term is the pass-through attenuation. The prefactor $\varphi\,\tau_m^k$ is larger in concentrated markets, so a larger share of any improvement in $y_m$ is absorbed by the markdown and a smaller share is passed through to wages. (Allowing $\tau_m^k$ itself to respond to $y_m$ through endogenous effort would introduce an additional second-order term, which we abstract from for transparency.)

## 5.6 Discussion

The empirical patterns in Section 5 support the model's qualitative predictions, but the full welfare and policy implications of endogenous boundaries warrant further structural work. We offer four observations to frame that work and plan to develop them in future work.

First, endogenous boundaries provide a natural lens on the long-standing gap between micro and macro estimates of how labor reallocates in response to local shocks. Local studies see only within-market adjustment; aggregate studies pick up cross-market reallocation. Whenever $\kappa$ is finite, the cross-market substitution margin is non-trivial, and aggregate measures of labor reallocation exceed local ones. Our finding that leakage rises in recessions suggests, further, that this wedge is itself countercyclical.

Second, local labor-market policies operate in a regime that depends on *neighbors*. The classical monopsony case for a city-level minimum wage is that a binding minimum forces firms to internalize their market power. Through the lens of our framework, the incidence of a city-level minimum depends on the home-market HHI of the affected city *and* on the HHIs of the neighboring cities to which workers can apply. The relevant statistic is the contact-weighted exposure $\mathcal{C}^k_{\text{eff}} = \sum_m (\lambda^k_m/\Lambda^k)\,\varphi\,\tau^k_m$ defined in Section 3.8. By the same logic, place-based subsidies behave like worker-based subsidies when $\kappa$ is high but leak across markets when $\kappa$ is low; the substantial outward reallocation in Section 5 suggests the latter regime is empirically relevant [Gaubert, Kline, Vergara, and Yagan, 2025].

Third, the bite of monopsony has both price and quantity margins. The wage gap $w^{k,0}_m - w^k_m$ measures the per-period transfer from worker to firm *conditional on a match*. Workers in concentrated markets, however, also exert less total search effort in partial equilibrium (Corollary 3). Whether this partial-equilibrium effort decline translates into longer equilibrium unemployment durations is a general-equilibrium question: lower local search lowers $u_m$, which raises $\theta_m$ through free entry (7) and partially offsets the effort decline through higher contact rates. Quantifying the net effect requires a calibrated GE exercise, which we leave to follow-up work. Wage-based markdown estimators [Yeh, Macaluso, and Hershbein, 2022, Berger, Herkenhoff, and Mongey, 2022] capture the price margin only; the quantity margin requires data on applications and unemployment durations across CZs of different concentration. Our cross-CZ finding of $-15.5\%$ fewer applications flowing into the typical 1-SD-HHI CZ is consistent with such a quantity margin, though we do not separately identify within-worker effort contraction from across-worker reallocation.

Finally, several caveats apply. The exposure measure $\mathcal{C}^k_{\text{eff}}$ and the welfare statements above depend on parameters we do not separately identify in the reduced-form analysis; the roadmap of Section 4.3 indicates how a structural exercise would pin them down. Our empirical setting also focuses on hourly low- and middle-wage workers via a single platform, so external validity to other segments of the labor market or to other data environments is an open question that we leave for future work.

# 6 Conclusions and Next Steps

We propose a search-and-bargaining model in which labor market boundaries are endogenous, monopsony exposure is summarized by a single sufficient statistic $\tau^k_m$, and the wage markdown emerges from the interaction of within-market concentration and cross-market escape. Using a large dataset of online applications to U.S. hourly jobs, we document substantial heterogeneity across workers and cyclical variation in the geographic breadth of search. We then show that four of the model's predictions find empirical support: workers in concentrated home markets reallocate effort outward, re-application to large firms is dampened relative to a random-search benchmark (and the dampening sharpens with HHI), wages decline with concentration along an interior gradient that steepens with home-market HHI, and productivity pass-through is attenuated where escape is harder.

Three contributions warrant emphasis. First, the empirical findings refine our understanding of how local labor markets work. Search is local on average, but not uniformly so: concentration in the home market induces outward search by a non-trivial fraction of workers, and the cyclical pattern of leakage suggests that the boundaries identified in cross-sectional analyses understate cyclical variation in those boundaries. Place-based and concentration-based policy evaluation that takes labor market boundaries as fixed risks missing both heterogeneity and cyclicality. Second, the theoretical framework provides a tractable summary of monopsony exposure that depends jointly on within-market and cross-market structure. The sufficient-statistic interpretation of $\tau_m^k$ economizes on parameters: a researcher who can measure the firm-level vacancy share, market-level tightness, and cross-market access can characterize the role of market structure in the wage markdown—that is, $\tau_m^k$—without estimating a full structural model. Mapping $\tau_m^k$ to the level of the markdown additionally requires the bargaining share $\eta$, the embargo probability $\varphi$, and the match surplus $W_m^k - U^k$. Third, the framework is well-suited to policy analysis. Local minimum wages, place-based subsidies, and labor market mergers all have welfare implications that depend on the cross-market access component of $\tau_m^k$.

Two specific extensions stand out for future work. First, we will expand the model beyond $M = 2$ and calibrate it to the empirical distribution of CZ-level HHI and worker portfolios. Second, we plan to use the calibrated model to evaluate two policy questions: a replication of the spatial-discontinuity literature on the employment effects of state-level minimum wages, and the cyclical incidence of local versus federal minimum-wage changes through the search-reallocation channel.

## References


David H. Autor, David Dorn, and Gordon H. Hanson. The China syndrome: Local labor market effects of import competition in the United States. *American Economic Review*, 103(6):2121–2168, 2013. doi: 10.1257/aer.103.6.2121. URL https://www.aeaweb.org/articles?id=10.1257/aer.103.6.2121.

José Azar, Ioana Marinescu, and Marshall Steinbaum. Labor market concentration. *Journal of Human Resources*, 57(S):S167–S199, 2022. doi: 10.3368/jhr.monopsony.1218-9914R1. URL https://jhr.uwpress.org/content/57/S/S167.

David W. Berger, Kyle Herkenhoff, and Simon Mongey. Labor market power. *American Economic Review*, 112(4):1147–1193, 2022. doi: 10.1257/aer.20191521. URL https://www.aeaweb.org/articles?id=10.1257/aer.20191521.

David W. Berger, Thomas Hasenzagl, Kyle F. Herkenhoff, Simon Mongey, and Eric A. Posner. Merger guidelines for the labor market. *Journal of Monetary Economics*, 153:103762, 2025. doi: 10.1016/j.jmoneco.2025.103762. URL https://www.sciencedirect.com/science/article/abs/pii/S030439322500056X. Published online; NBER Working Paper No. 31147.

Adrien Bilal. The geography of unemployment. *Quarterly Journal of Economics*, 138(3):1507–1576,

2023. doi: 10.1093/qje/qjad014. URL https://academic.oup.com/qje/article-abstract/138/3/1507/7085015.

Olivier Jean Blanchard and Peter Diamond. Ranking, unemployment duration, and wages. *Review of Economic Studies*, 61(3):417–434, 1994. doi: 10.2307/2297897. URL https://www.jstor.org/stable/2297897.

R. Jason Faberman and Marianna Kudlyak. The intensity of job search and search duration. *American Economic Journal: Macroeconomics*, 11(3):327–357, 2019. doi: 10.1257/mac.20170315. URL https://www.aeaweb.org/articles?id=10.1257/mac.20170315.

Jesús Fernández-Villaverde, Federico Mandelman, Yang Yu, and Francesco Zanetti. The "Matthew effect" and market concentration: Search complementarities and monopsony power. *Journal of Monetary Economics*, 121:62–90, 2021. doi: 10.1016/j.jmoneco.2021.04.010. URL https://www.sciencedirect.com/science/article/pii/S030439322100050X.

Cecile Gaubert, Patrick Kline, Damian Vergara, and Danny Yagan. Place-based redistribution. *American Economic Review*, 115(10):3415–3450, 2025. doi: 10.1257/aer.20202067. URL https://www.aeaweb.org/articles?id=10.1257/aer.20202067.

Erik Hurst, Patrick J. Kehoe, Elena Pastorino, and Thomas Winberry. The macroeconomic dynamics of labor market policies. NBER Working Paper 33614, National Bureau of Economic Research, 2025. URL https://www.nber.org/papers/w33614.

Gregor Jarosch, Jan Sebastian Nimczik, and Isaac Sorkin. Granular search, market structure, and wages. *Review of Economic Studies*, 91(6):3569–3607, 2024. doi: 10.1093/restud/rdae004. URL https://academic.oup.com/restud/article-abstract/91/6/3569/7564585.

John Kennan and James R. Walker. The effect of expected income on individual migration decisions. *Econometrica*, 79(1):211–251, 2011. doi: 10.3982/ECTA4657. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA4657.

Alan Manning and Barbara Petrongolo. How local are labor markets? Evidence from a spatial job search model. *American Economic Review*, 107(10):2877–2907, 2017. doi: 10.1257/aer.20131026. URL https://www.aeaweb.org/articles?id=10.1257/aer.20131026.

Ioana Marinescu and Roland Rathelot. Mismatch unemployment and the geography of job search. *American Economic Journal: Macroeconomics*, 10(3):42–70, 2018. doi: 10.1257/mac.20160312. URL https://www.aeaweb.org/articles?id=10.1257/mac.20160312.

Daniel McFadden. Modelling the choice of residential location. In Anders Karlqvist, Lars Lundqvist, Folke Snickars, and Jörgen W. Weibull, editors, *Spatial Interaction Theory and Planning Models*, pages 75–96. North-Holland, Amsterdam, 1978. URL https://eml.berkeley.edu/reprints/mcfadden/location.pdf.

Ferdinando Monte, Stephen J. Redding, and Esteban Rossi-Hansberg. Commuting, migration, and local employment elasticities. *American Economic Review*, 108(12):3855–3890, 2018. doi: 10.1257/aer.20151507. URL https://www.aeaweb.org/articles?id=10.1257/aer.20151507.

Benoît Schmutz and Modibo Sidibé. Frictional labour mobility. *Review of Economic Studies*, 86 (4):1779–1826, 2019. doi: 10.1093/restud/rdy056. URL https://academic.oup.com/restud/article-abstract/86/4/1779/5102816.

Robert Shimer. The assignment of workers to jobs in an economy with coordination frictions. *Journal of Political Economy*, 113(5):996–1025, 2005. doi: 10.1086/444551. URL https://www.journals.uchicago.edu/doi/10.1086/444551.

Danny Yagan. Employment hysteresis from the great recession. *Journal of Political Economy*, 127 (5):2505–2558, 2019. doi: 10.1086/701809. URL https://www.journals.uchicago.edu/doi/10.1086/701809.

Chen Yeh, Claudia Macaluso, and Brad Hershbein. Monopsony in the U.S. labor market. *American Economic Review*, 112(7):2099–2138, 2022. doi: 10.1257/aer.20200025. URL https://www.aeaweb.org/articles?id=10.1257/aer.20200025.

## A Proofs

This appendix collects the algebra for Lemmas 1 and 2, Corollary 1, Proposition 1, and Corollaries 2–3. All comparative statics are partial-equilibrium: $N_m$ varies while $U^k$, $\{U^k_{d,m'}\}_{m'\neq m}$, and the market tightnesses $\{\theta_{m'}\}$ are held fixed.

### A.1 Proof of Lemma 1

Subtract the embargo-state Bellman from the unemployment Bellman. The flow terms and the continuations in markets $j \neq m$ cancel. Using the Nash relation $W^k_m - U^k_{d,m} = (1-\varphi)(U^k - U^k_{d,m}) + \eta S^k_m$, the difference $x \equiv U^k - U^k_{d,m}$ satisfies

$$x \cdot D_1 \;=\; \beta(\lambda^k_m - \lambda^{d,k}_m)\,\eta\, S^k_m,$$

where

$$D_1 \;\equiv\; 1 - \beta\Big(1 - \lambda^k_m - \sum_{j\neq m} \lambda^k_j\Big) - \beta(\lambda^k_m - \lambda^{d,k}_m)(1-\varphi).$$

Expanding $D_1$:

$$D_1 \;=\; (1-\beta) + \beta\lambda^k_m\varphi + \beta\lambda^{d,k}_m(1-\varphi) + \beta\sum_{j\neq m}\lambda^k_j \;\geq\; 1-\beta \;>\; 0,$$

using $\beta < 1$, $\lambda \geq 0$, and $\varphi \in [0,1]$. The right-hand side is strictly positive when $\lambda^{d,k}_m < \lambda^k_m$ and $S^k_m > 0$, so $x > 0$. □

### A.2 Derivation of $\tau^k_m$

At the optimum,

$$U^k = \frac{A}{D}, \quad A = b - \tfrac{\psi}{2}E^2 - \sum_j d(k,j)e_j + \beta\sum_m \lambda^k_m W^k_m, \quad D = 1 - \beta(1-\Lambda^k).$$

For the embargo state,

$$U^k_{d,m} = \frac{A - \beta\Delta_m W^k_m}{D - \beta\Delta_m}.$$

Therefore

$$U^k - U^k_{d,m} = \frac{\beta\Delta_m\,(DW^k_m - A)}{D\,(D-\beta\Delta_m)} = \frac{\beta\Delta_m}{D-\beta\Delta_m}(W^k_m - U^k),$$

using $A/D = U^k$ and $D - \beta\Delta_m = 1 - \beta(1-\Lambda^{m,k}_d)$. The bracketed coefficient is $\tau^k_m$. □

### A.3 $\partial U^k_{d,m}/\partial N_m > 0$

The closed form in Section 3.4 expresses $U^k_{d,m}$ in terms of $W^k_m$. Substituting $W^k_m = (1-\eta)\widetilde{U}^k_m + \eta(y_m + \beta\delta U^k)/(1-\beta(1-\delta))$ and $\widetilde{U}^k_m = (1-\varphi)U^k + \varphi U^k_{d,m}$, and solving,

$$U^k_{d,m} \;=\; \frac{C}{D_2 - \beta\lambda^{d,k}_m(1-\eta)\varphi},$$

where $D_2 \equiv 1-\beta(1-\lambda^{d,k}_m - \sum_{j\neq m}\lambda^k_j)$ and $C$ collects all terms independent of $U^k_{d,m}$. Differentiating with respect to $N_m$ and using $W^k_m - U^k_{d,m} = (1-\eta)[(1-\varphi)U^k + \varphi U^k_{d,m}] + \eta(y_m+\beta\delta U^k)/(1-\beta(1-\delta)) - U^k_{d,m}$,

$$\frac{\partial U^k_{d,m}}{\partial N_m} \;=\; \frac{\beta\,(\partial\lambda^{d,k}_m/\partial N_m)\,(W^k_m - U^k_{d,m})}{D_2 - \beta\lambda^{d,k}_m(1-\eta)\varphi} \;>\; 0,$$

since $W^k_m > U^k_{d,m}$ (Lemma 1) and $\partial\lambda^{d,k}_m/\partial N_m > 0$. To verify the latter, write $\pi^d_m = \pi_m + f_m(q_m - \pi_m)$ with $f_m = 1/N_m$ and $\pi_m > q_m$ for $\theta_m > 0$:

$$\frac{\partial \pi^d_m}{\partial N_m} \;=\; \frac{\partial f_m}{\partial N_m}(q_m - \pi_m) \;=\; -\frac{1}{N_m^2}(q_m-\pi_m) \;>\; 0,$$

so $\lambda^{d,k}_m = 1-\exp(-\pi^d_m e^k_m)$ is increasing in $N_m$. The denominator

$$D \;\equiv\; D_2 - \beta\lambda^{d,k}_m(1-\eta)\varphi \;=\; (1-\beta) + \beta\lambda^{d,k}_m[1-(1-\eta)\varphi] + \beta\sum_{j\neq m}\lambda^k_j \;\geq\; 1-\beta > 0,$$

using $(1-\eta)\varphi \leq 1$. □

### A.4 $\partial G^k_m/\partial N_m > 0$

From $G^k_m = \beta\pi_m\eta S^k_m + \beta\pi_m\varphi(U^k_{d,m} - U^k)$ and $\partial S^k_m/\partial N_m = -\varphi\,\partial U^k_{d,m}/\partial N_m$ (with $U^k$ fixed),

$$\frac{\partial G^k_m}{\partial N_m} \;=\; \beta\pi_m\varphi(1-\eta)\,\frac{\partial U^k_{d,m}}{\partial N_m} \;>\; 0. \quad \square$$

### A.5 Comparative statics in $N_m$

Define $h(E;N_m) \equiv \sum_j e_j(E;N_m) - E$, where $e_j(E;N_m)$ is the closed-form effort from (5). The equilibrium condition $h=0$ pins down $E$. Differentiating,

$$\frac{\partial h}{\partial E} \;=\; -1-\psi S, \qquad \frac{\partial h}{\partial N_m} \;=\; \frac{1}{\pi_m G^k_m}\,\frac{\partial G^k_m}{\partial N_m} \;>\; 0,$$

where $S \equiv \sum_j 1/[\pi_j(\psi E + d(k,j))] > 0$. By the implicit function theorem,

$$\frac{dE}{dN_m} \;=\; \frac{(\partial G^k_m/\partial N_m)\,/\,(\pi_m G^k_m)}{1+\psi S} \;>\; 0.$$

For $j \neq m$, $\partial e_j/\partial N_m\big|_E = 0$, so

$$\frac{de_j}{dN_m} \;=\; \frac{\partial e_j}{\partial E}\,\frac{dE}{dN_m} \;=\; -\frac{\psi}{\pi_j(\psi E + d(k,j))}\,\frac{dE}{dN_m} \;<\; 0.$$

For $j = m$, both channels operate. Combining $\partial e_m/\partial E$ with $\partial e_m/\partial N_m\big|_E = (\pi_m G_m^k)^{-1}\,\partial G_m^k/\partial N_m$,

$$\frac{de_m}{dN_m} \;=\; \frac{1}{\pi_m G_m^k}\,\frac{\partial G_m^k}{\partial N_m}\left[1 - \frac{\psi}{\pi_m(\psi E + d(k,m))(1+\psi S)}\right].$$

Since $S \geq 1/[\pi_m(\psi E + d(k,m))]$, the bracket is strictly positive and $de_m/dN_m > 0$.

### A.6 Proof of Proposition 1

Set $m = k$ in the comparative statics above. Then $de_k^k/dN_k > 0$ gives part (1), and $de_j^k/dN_k < 0$ for $j \neq k$ gives part (2). □

### A.7 Proof of Corollary 2

Set $m \neq k$. Applying the $de_j/dN_m$ formula at $j = k$ gives $de_k^k/dN_m < 0$, so home effort rises as $N_m$ falls; the $de_m/dN_m$ formula gives $de_m^k/dN_m > 0$, so effort directed at $m$ falls. □

### A.8 Proof of Corollary 3

Immediate from $dE/dN_m > 0$. □

### A.9 Solution algorithm

We solve the model by Gauss-Seidel iteration on tightness $(\theta_A, \theta_B)$:

1. Fix $\theta_B$. Solve a 1D root for $\theta_A$ satisfying free entry; the inner problem iterates on value-function scalars with closed-form wages, surplus, employment values, and 1D effort.

2. Fix $\theta_A$ (new). Solve for $\theta_B$ analogously.

3. Iterate until $\max|\log\theta_m^{\text{new}} - \log\theta_m^{\text{old}}| < \varepsilon$ and free-entry residuals $\approx 0$.

The inner problem reduces to six scalars $(U_A, U_B, U_{d,AA}, U_{d,AB}, U_{d,BA}, U_{d,BB})$ with dampened updates.

## B Robustness and external validity

This appendix discusses the external validity of our estimates and the robustness of the main findings to alternative specifications.

## B.1 Selection into the platform

Because the platform is the largest job-search website for hourly work, its coverage is broad, but it is not representative of the universe of U.S. low- and middle-wage hourly job seekers. Two selection concerns warrant attention.

First, the platform's user base may be tilted toward applicants who are comfortable with online job search. This is more concerning for some questions than for others. For our P2 test (re-application within market), the relevant comparison is between actual and random-search predicted re-application rates, both computed within the platform. Platform-specific selection should not bias the comparison so long as the embargo mechanism operates within the platform sample. For P1 and P3, the relevant comparison is between concentrated and competitive CZs within the platform; the same logic applies.

Second, the platform's employer base is concentrated in industries that hire heavily at the lower end of the wage distribution: retail, hospitality, healthcare support, transportation and warehousing.

## B.2 Robustness to alternative trimming

Our baseline drops the top 0.5% of applicants by application volume and the top 0.5% by application distance. Alternative thresholds (0.1%, 1%, 5%) yield qualitatively identical patterns for all four main tests. The 1-SD HHI effect on search radius (P1) is +2.52 miles at our baseline and ranges from +2.3 to +2.8 miles under alternative trimming, with statistical significance unchanged. Similar stability holds for P2, P3, and P4.

## B.3 Identification threats and clustering

Standard errors are clustered at the CZ level throughout, which is the level at which our key explanatory variable (HHI) varies. Two-way clustering by CZ and year leaves all main coefficients statistically significant at the same conventional levels.

The main threat to identification of P3 is that high-HHI CZs may differ systematically from low-HHI CZs along dimensions correlated with both concentration and wages. The first-differences specification (Table 7) addresses time-invariant confounders, and the firm-fixed-effects robustness leaves the sign and significance of $\beta_1$ unchanged. For P4, the Bartik shock relies on the assumption that national-level industry shocks are exogenous to local CZ-level wages, conditional on the chosen fixed-effects structure. As a robustness check, we re-estimate P4 with QCEW wage growth as the national shifter (Table 9); the interaction with HHI lag retains its sign in both alternative specifications.

## B.4 Robustness to alternative market definitions

The baseline uses commuting zones (CZs) as the geographic unit. We have also explored CBSA-MSA definitions and CZ×SOC3 cells as alternative market definitions. Results for P1 and P2 are qualitatively unchanged. For P3, results are also robust, though the wage-HHI gradient is somewhat steeper in CZ×SOC3 cells than in CZ-only cells, consistent with sharper identification of monopsony power within occupations.

The CZ×SOC3 specification is particularly informative for the application-share/vacancy-share tracking benchmark (Table 10). At the CZ level alone, application shares exceed vacancy shares by 12% (slope 1.12), consistent with workers tilting applications toward larger firms within CZ. At the CZ×SOC3 level, the slope is 0.95, close to the random-search null of 1. The model is consistent with both findings: workers direct effort across SOC categories but search undirectedly within them.

## C Additional figures and tables

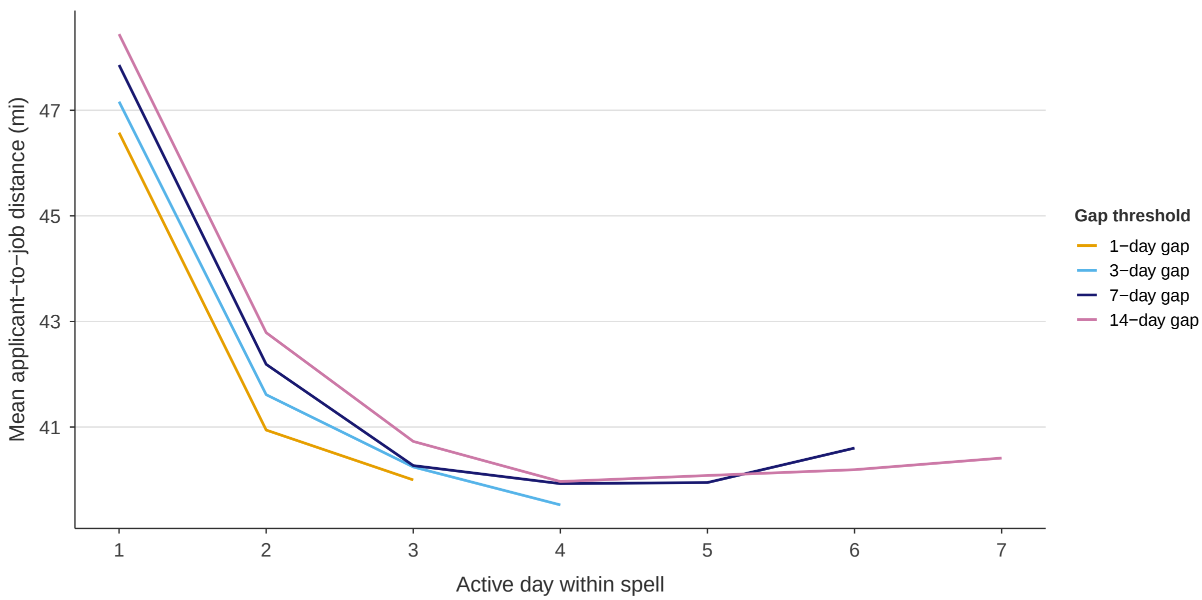


Figure 6: Search radius versus day of search within a spell. Selected sample averages by allowed inactivity gap between applications.

**Within-firm and within-spell dynamics.**

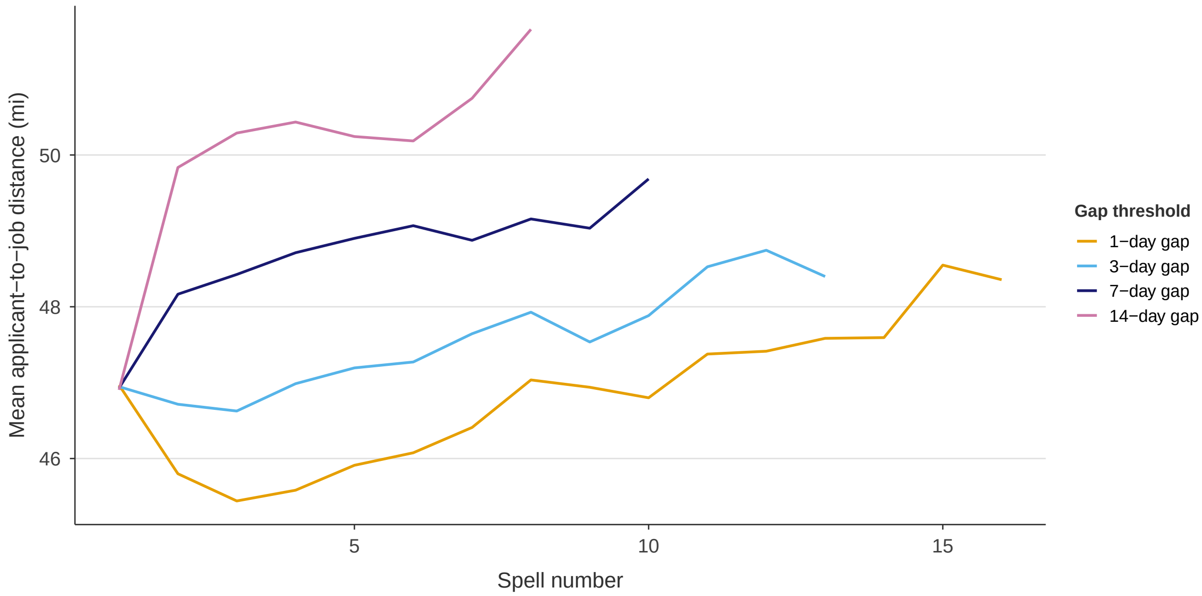


Figure 7: Search radius across search spells. Selected sample averages by allowed inactivity gap between applications.

Table 9: P4 robustness: alternative Bartik measures

| | **QCEW wage Bartik** | **OEWS-platform emp. Bartik** |
|---|---|---|
| $g^{\text{Bartik}}_{cz,t}$ | $+0.41^{*}$ | $+0.49^{***}$ |
| | (0.22) | (0.19) |
| HHI lag | +0.042 | −0.090 |
| | (0.141) | (0.105) |
| $g^{\text{Bartik}}_{cz,t} \times$ HHI lag | −3.77 | −0.053 |
| | (3.92) | (0.79) |
| FEs | Firm + CZ + year | Firm + CZ + year |
| $N$ | 1,209,776 | 1,209,752 |

Notes: Both columns confirm positive pass-through; sign of the interaction term is in the predicted direction in both, statistically distinguishable from zero in neither with the alternative weightings.

Table 10: Auxiliary: application shares track vacancy shares within markets

| Market definition | CZ | CZ×SOC3 |
|---|---|---|
| Vacancy share | $1.12^{***}$ | $\mathbf{0.95}^{***}$ |
| | (0.01) | (0.003) |
| $t(\hat{\beta} = 1)$ | $+12.94^{***}$ | $-18.57^{***}$ |
| $N$ | 3.6M | 2.9M |

Notes: Specification $s^{a}_{imt} = \alpha + \beta\, s^{v}_{imt} + \gamma_{mt} + \varepsilon_{imt}$. Within CZ×SOC3 markets, the slope is 0.95, close to the random-search null of 1. Rejection of $\beta = 1$ is driven by very large $N$; magnitude is the more meaningful summary.

Table 11: Auxiliary: workers with broader portfolios re-encounter the same firm less often

| | **OLS** | **Logit (AME)** |
|---|---|---|
| search breadth (variance, $\text{mi}^2$) | $-4.51\times10^{-8\,***}$ | $-6.9\times10^{-3}$ |
| | $(1.8\times10^{-9})$ | (per 1-SD) |
| HHI | $+0.673^{***}$ | $+5.29^{***}$ |
| | (0.088) | (0.74) |
| FEs | Year | Year |
| $N$ | 45.8M | 45.8M |

Notes: Specification $\text{reapply}_{ijmt} = \alpha + \beta_1\,\text{breadth}_i + \beta_2\,\text{HHI}_{mt} + \lambda_t + \varepsilon$. Predicted $\beta_1 < 0$ under the model: broader $\Lambda_d^{m,k} \Rightarrow$ shorter punishment duration $\Rightarrow$ less re-encounter.

Table 12: Search behavior across spell duration

| **Spell phase** | **Mean geo. dist. (mi)** | **Mean skill dist.** | **Daily apps (mean)** | **Sample (thousands)** |
|---|---|---|---|---|
| Day 1 | 44.82 | 0.20 | 3.75 | 11,754.58 |
| Day 2 | 38.34 | 0.16 | 3.03 | 3,514.55 |
| Day 3 | 36.73 | 0.16 | 2.89 | 1,395.86 |
| Days 4–7 | 36.94 | 0.14 | 2.55 | 662.64 |
| Days 8–14 | 39.21 | 0.15 | 2.62 | 77.23 |
| Days 15+ | 39.20 | 0.16 | 2.80 | 7.61 |

Table 13: Search patterns by application volume

| **Apps/Individual** | **Mean geo. dist. (mi)** | **Mean skill dist.** | **Distinct SOC** | **Obs. (thousands)** | **Pct.** |
|---|---|---|---|---|---|
| 2–3 | 44.88 | 0.23 | 1.80 | 3,784.70 | 45.93% |
| 4–5 | 40.04 | 0.33 | 2.91 | 1,558.95 | 18.92% |
| 6–7 | 37.60 | 0.36 | 3.86 | 830.26 | 10.08% |
| 8–10 | 36.67 | 0.38 | 4.86 | 692.97 | 8.41% |
| 11–15 | 36.74 | 0.39 | 6.20 | 569.14 | 6.91% |
| 16–20 | 37.87 | 0.40 | 7.71 | 281.05 | 3.41% |
| 21–30 | 39.71 | 0.40 | 9.46 | 259.94 | 3.15% |
| 31+ | 46.96 | 0.41 | 14.36 | 263.65 | 3.20% |