Spatial Sorting of Workers and Firms*

Ryungha Oh†
University of Chicago

August 2025

Abstract

This paper develops a theory of two-sided spatial sorting where heterogeneous workers and firms choose locations and match randomly within local labor markets. Spatial disparities arise endogenously across ex ante homogeneous locations. Productive workers and firms co-locate in cities due to production complementarities, making cities with high-quality search pools densely populated. Although spatial sorting improves allocative efficiency by facilitating positive assortative matching, concentration in cities is inefficiently high. Since workers and firms embody productivity, relocating them can mitigate congestion without reducing output. A comparison with models without either worker or firm sorting reveals that two-sided sorting yields distinct policy implications.

Keywords: assortative matching, spatial sorting, search frictions, spatial disparities, place-based policies

*I am greatly indebted to my advisors, Giuseppe Moscarini, Ilse Lindenlaub, and Costas Arkolakis, for their guidance and support. I also thank George-Marios Angeletos, Job Boerma, Hector Chade, Daniel Haanwinckel, Patrick Kehoe, Sam Kortum, Michael Peters, Fabien Postel-Vinay, Simon Mongey, Bernardo Ribeiro, Jaeeun Seo, Martin Souchier, Takatoshi Tabuchi, Aleh Tsyvinski for helpful comments. I am grateful to audiences at SED 2023, UEA 2023 Toronto, The Federal Reserve Bank of Minneapolis Junior Scholar Conference, PSU New Faces in International Economics Conference, USC Marshall Macro Day 2024, the NBER Summer Institute 2024, SEA Annual Meeting 2024, Hitotsubashi-Gakushuin Conference 2024, and St. Louis Fed-WashU-LAEF Macro-Labor Conference 2025 for helpful comments; and to seminar participants at Boston University, Bocconi, Chicago Booth, Cornell, CUHK, EIEF, EUI, Harvard, HKUST, LSE, Michigan, Northwestern, NYU, Princeton, Seoul National University, Stanford, Stanford GSB, UC Berkeley, the University of Tokyo, the University of Virginia, WashU, Wharton, Wisconsin, and Yale for valuable feedback. Data access was provided via on-site use at the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB) and subsequently remote data access.

†*Email:* `ryungha.oh@chicagobooth.edu`

1. Introduction

More productive workers tend to be employed by more productive firms (e.g., Card et al., 2013). This positive assortative matching (PAM) arises naturally from complementarities in production. However, search is costly, and thus the degree of sorting is shaped by the process in which workers and firms match with one another. For example, when all agents search randomly in a single market, the extent of PAM may be limited (e.g., Shimer and Smith, 2000). In reality, geography plays a crucial role in determining the potential pool of matches, as workers and firms primarily search locally. Moreover, the quality of local pools—determined by the types of workers and firms present—varies across regions as suggested by substantial spatial wage differentials. How does the presence of geography influence the pattern of assortative matching? Can such a sorting pattern account for spatial disparities in the data? What are the implications for efficiency?

This paper addresses these questions by proposing a new theory of *two-sided sorting*. First, I show that two-sided sorting leads to PAM and generates spatial disparities: productive workers and firms colocate in cities characterized by high population density and wages. Second, I show that congestion in cities stemming from better matching opportunities is inefficient. Since output depends solely on worker and firm productivity, jointly relocating workers and firms can reduce congestion without lowering output. This result is in stark contrast with those from benchmark models that abstract from either worker or firm sorting, which are widely used in the literature to explain spatial disparities.[1] Third, I calibrate the model to demonstrate its capacity to match cross-sectional patterns and to evaluate real-world policies. Finally, I provide evidence that worker sorting causally affects firm sorting, which highlights the empirical relevance of two-sided sorting.

I begin the analysis by developing a parsimonious theory of spatial sorting of *heterogeneous* workers and *heterogeneous* firms across ex ante *homogeneous* locations. A key feature of the model is that the matching between workers and firms occurs through their location decisions. In this model, the type and number of both workers and firms across regions are endogenously determined. In each local labor market, workers and firms engage in random matching under search frictions, and upon matching, bargain over the match surplus. The attractiveness of local labor markets increases in the productivity and number of agents on the other side of the market. In contrast, the concentration of agents on the same side increases local congestion costs: housing rents for workers and business operating costs for firms.

[1] See Diamond and Gaubert (2022) for a detailed review.

I show that PAM arises across space and that locations with more productive workers and firms become densely populated. For example, more workers are drawn to locations with either more productive firms or a higher job arrival rate, until housing rents rise enough to offset the gains from better local labor markets. A similar mechanism applies to firms. In particular, due to worker-firm complementarity, highly productive workers and firms benefit more from these labor markets and thus are willing to pay higher local congestion costs. Importantly, assortative matching is primarily driven by productivity, rather than by the probability of matching. The intuition is that the probability of matching, which depends on the relative number of workers and firms, may attract one side of the market but not both at the same time, whereas productivity may benefit both workers and firms. As a result, worker sorting sustains firm sorting, and vice versa, which implies that two-sided sorting alone—without location heterogeneity arising from local TFP differences or agglomeration forces—can explain the spatial disparities observed in the data.

I then evaluate the efficiency of the decentralized equilibrium and demonstrate that it features excessive concentration in dense areas. The key insight is that productivity is embodied in workers and firms; what matters is *who* produces, not *where* production takes place. Thus, concentration in denser areas is not desirable because it would only raise total congestion costs. However, in equilibrium, less productive workers overvalue the benefits of choosing denser locations. They do not internalize their negative impact on local firms that could have hired more productive workers if they had not chosen these locations. Similarly, less productive firms choose denser locations than is socially optimal, and these externalities lead to excessive congestion costs. I show that the government can restore efficiency by taxing workers and firms choosing dense areas.

The importance of geography in shaping matching outcomes stems from its role in determining the types of potential partners, as screening is imperfect within local labor markets due to random matching. To illustrate this, I consider an alternative environment with neoclassical local labor markets, following Becker (1973). In contrast to the baseline, geography in this setting has no economic meaning. Specifically, the matching between workers and firms replicates the outcome of an economy with a single, nationwide labor market. Moreover, population density is uniform across space, and the equilibrium is efficient. The key difference from the baseline is that without random matching, workers and firms can selectively choose their partners within each local market. As a result, location decisions no longer directly influence the matching process. I confirm that introducing directed search—where agents can target specific types of partners within a local market—yields qualitatively identical results, and again, leaves no meaningful role for geography.

In the final part of the theory, I show that two-sided sorting yields distinct normative implications compared to alternative explanations of spatial disparities. To this end, I extend the baseline model to incorporate additional sources of spatial disparities: exogenous local TFP and agglomeration forces. This extended model nests alternative mechanisms commonly studied in the literature, such as one-sided sorting. I show that these mechanisms and the two-sided sorting mechanism are equivalent in matching cross-sectional data. In contrast, I find that efficiency implications crucially depend on the nature of two-sided sorting, highlighting the importance of incorporating it into policy analysis. For example, unlike the baseline, when productivity is partly embodied in locations, spatial concentration in more productive areas is socially optimal.

To test the quantitative potential of two-sided sorting, I calibrate the model using cross-sectional data from the U.S. Despite its parsimony, the model successfully replicates spatial disparities in nominal wages and population density, namely the urban wage premium. For estimating worker and firm heterogeneity, I rely on workers' location decisions. Higher wages driven by firm heterogeneity benefit any workers in local markets. As a result, workers flow into such locations until higher housing rents offset these gains. By controlling for these compensating differentials, I separately identify worker and firm productivity. My estimation reveals that the spatial heterogeneity of workers and firms is both significant: workers and firms in the top decile of locations in terms of population density are 24.4% and 20.4% more productive than their counterparts in the bottom decile, respectively.

Using this calibrated model, I quantify the effects of spatial policy interventions, focusing on how and to what extent they differ under the two-sided sorting mechanism relative to alternative mechanisms. I study a policy that relaxes housing regulations only in denser cities. In the two-sided sorting model, this deregulation lowers welfare by amplifying the concentration of economic activity in cities and thus exacerbating congestion costs. However, because both workers and firms relocate, the matching between them remains nearly unchanged, leaving total output largely unaffected. In contrast, I consider a benchmark one-sided sorting model in which heterogeneous workers sort across space with exogenous local TFP, while firms are homogeneous. In this case, the same policy has the opposite effect on welfare. Because workers relocate to cities with higher local TFP, aggregate output rises, and these gains outweigh the associated congestion costs. The defining feature of two-sided sorting underlying these divergent results is that local productivity is policy-dependent, as it is endogenously shaped by the mutually dependent sorting of workers and firms.

Finally, I test this key feature using German employer-employee matched data to present empirical evidence of two-sided sorting. Specifically, I show that an increase in the productivity of workers in a local search pool attracts more productive jobs to that location. To implement this, I use the model to recover the productivity of workers and firms in each location from worker and firm fixed-effects estimates from a log wage regression. I then instrument changes in worker productivity using predicted changes in the productivity of incoming domestic migrants, constructed from pre-period migration networks and contemporaneous outflows from other locations. This instrument addresses endogeneity concerns arising from changes in location-specific conditions, such as local TFP shocks, which could affect both worker and firm sorting. To ensure that my estimates are not confounded by alternative forces such as agglomeration, I further control for changes in the productivity of preexisting jobs. My estimates show that a one standard deviation increase in the log productivity of the local workforce attracts jobs with approximately 0.5 standard deviations higher log productivity. This finding implies that worker sorting substantially affects firm sorting, which underscores the role of two-sided sorting as a key driver of spatial disparities.

Related literature. A large literature on two-sided matching shows that complementarity in payoffs serves as the key source of assortative matching in different environments, such as competitive neoclassical markets (e.g., Becker, 1973); game theoretic environments (e.g., Roth and Sotomayor, 1990); and markets with search frictions (e.g., Shimer and Smith, 2000; Eeckhout and Kircher, 2010; Smith, 2006; Shi, 2001). This paper also characterizes assortative matching arising from complementarity, but the matching process differs in two important respects. First, assortative matching is facilitated through location decisions. Locations serve as platforms: workers and firms determine their potential pool of matches by selecting a location, but within that location, their search is no longer directed. Hence, search is neither fully random nor fully directed.[2] Second, I endogenize the density of workers and firms across space. Consequently, the model characterizes not only the types but also the measures of agents in each location.[3]

Within the literature on two-sided matching, this paper builds on studies that adopt a search framework. In a seminal paper, Becker (1973) shows that supermodular production leads to a competitive equilibrium that exhibits PAM. Subsequent studies show that under search frictions, it is more difficult to obtain PAM as

[2] This aspect of search resembles the environment in Menzio (2007), where heterogeneous firms can communicate using nonbinding contractual messages prior to homogeneous workers' application.

[3] More general models endogenize the number of agents of a match, but often under specific assumptions that limit applicability in my context. For example, values are assumed to depend only on agent types and to be independent of the number of agents. If worker and firm densities at each location were assumed to be exogenous, I could apply the results of Demange and Gale (1985) or Roth and Sotomayor (1990).

an equilibrium outcome. With random search, equilibrium no longer exhibits perfect assortative matching. In other words, each agent type is matched with a non-degenerate distribution of partner types, rather than a single type (e.g., Shimer and Smith, 2000; Atakan, 2016).[4] With directed (competitive) search, although perfect assortative matching is attainable, it requires stronger complementarities. Eeckhout and Kircher (2010) characterize the condition for such complementarities, referred to as root-supermodularity of match payoffs, which depends on the elasticity of substitution of the search technology. In this paper, I show that when workers and firms can direct their search through location choices before engaging in random search, supermodular production is sufficient to generate perfect PAM.

This paper also closely contributes to the literature on spatial disparities. Studies have focused on the spatial sorting of heterogeneous workers or firms across ex ante heterogeneous locations. Heterogeneous workers or firms value location-specific fundamentals differently, which in turn shapes their location choices. The importance of worker sorting in spatial inequalities has been extensively studied, both empirically and theoretically (e.g., Baum-Snow and Pavan, 2012; De la Roca et al., 2023; Diamond, 2016; Martellini, 2022). Other work finds that the spatial sorting of firms also plays an important role (e.g., Bilal, 2023; Lindenlaub et al., 2025). Another body of literature shows that sorting can happen across ex ante homogeneous locations, through agglomeration forces that lead to complementarity between agent types and endogenous city characteristics (e.g., Davis and Dingel, 2019; Behrens et al., 2014; Gaubert, 2018). I also assume that locations are ex ante homogeneous. However, instead of agglomeration forces, the location choices of workers and firms mutually support each other. In contrast to those papers, to the best of my knowledge, this is the first paper to show that two-sided sorting *alone* can endogenously generate dense areas populated by productive workers and firms.

A small but growing literature analyzes frictional local labor markets across space. Kline and Moretti (2013) present a model that combines the Diamond-Mortensen-Pissarides framework (e.g., Pissarides, 2000) and the Roback (1982) framework. Subsequent studies extend this model to account for spatial differences in unemployment rates through firm sorting and the resulting differential separation rates (Bilal, 2023) or through endogenous separations and on-the-job search (Kuhn et al., 2022). I also combine the standard Diamond-Mortensen-Pissarides framework with a spatial equilibrium model but emphasize an additional

[4] Similarly, Smith (2006) shows that when utility is non-transferable, block segregation may arise under supermodular production. See Chade et al. (2017) for the related discussion.

role of search frictions. In addition to generating the spatial unemployment gap, search frictions are a driving force in generating differential population densities across space.

Finally, my findings are related to studies on spatial policies and spatial misallocation. Some studies show that spatial policies can introduce distortions in the spatial distribution of economic activities (e.g, Albouy, 2009; Fajgelbaum et al., 2019; Hsieh and Moretti, 2019). Other studies support policy interventions due to sources of externalities such as agglomeration forces (e.g., Fajgelbaum and Gaubert, 2020; Rossi-Hansberg et al., 2019). These studies typically address these questions by building a quantitative spatial model (e.g., Allen and Arkolakis, 2014; Redding, 2016) in which location fundamentals or agglomeration forces are key factors to be considered for policy evaluation. In contrast, this paper emphasizes the sorting of workers and firms, which in turn leads to distinct insights. Because both workers and firms can endogenously relocate, the model predicts responses to policies that differ from those implied by other mechanisms. I also highlight the potential for welfare improvements through the reduction of excessive spatial congestion arising from search frictions.

The rest of the paper is organized as follows. Section 2 presents the model and characterizes an equilibrium. Section 3 analyzes its properties and evaluates efficiency. Section 4 estimates the model and conducts counterfactual policy analysis. Finally, I provide empirical evidence on two-sided sorting in Section 5 and conclude.

2. The Economy

This section presents a model of spatial disparities that originate from the sorting of heterogeneous workers and firms. I first present the model and derive the equilibrium conditions.

2.1 Environment

Time indexed by t is continuous. There are ex ante homogeneous locations indexed by $\ell \in [0, 1]$. Each location is endowed with a unit measure of land.

Workers and firms. The economy is populated by a measure M_w of infinitely lived risk-neutral heterogeneous workers. Workers differ in productivity $x \in [\underline{x}, \overline{x}]$, drawn from the CDF $Q_w(\cdot)$. I assume that Q_w is twice continuously differentiable. Housing is a strict necessity, and workers rent $\bar{h}$ units of housing at rate $r_t(\ell)$.

They consume tradable goods g_t, which serve as the numeraire, and their flow utility is given by g_t. They inelastically supply one unit of labor and discount the future at rate ρ. Workers choose locations in which to reside and work, and cannot migrate.[5] If workers are employed, they earn a flow wage. If workers are unemployed, they receive unemployment benefit bx, which is financed by a lump-sum tax.

The economy is populated by a measure M_f of risk-neutral firms with a discount rate ρ. Firms differ in productivity $y \in [\underline{y}, \overline{y}]$, which is drawn from the twice continuously differentiable CDF $Q_f(\cdot)$. I assume that $\underline{y} > b$. Firms choose a location to operate a business, where they remain permanently. In this model, a firm is represented as a collection of vacant and filled positions.[6] At each point in time, each firm posts δ_v units of new vacancies. Also, firms demand a unit of local business services at overhead cost $c_t(\ell)$ for all t. These costs are associated with renting office spaces, handling administrative procedures, advertising, or posting vacancies.[7]

Technology. A worker and a vacancy match one-to-one in a local labor market to produce a flow xy of tradable goods.

Search and wage. At rate $\lambda_t(\ell)$, unemployed workers receive a job offer, become employed if they accept an offer, and earn a flow wage. Employed workers become unemployed at rate δ when the match is separated. At rate $q_t(\ell)$, a vacancy contacts a worker, and production begins if the match is acceptable. When this match is separated at rate δ, the position becomes vacant and reenters the search pool. A vacancy in the search pool gets destroyed at rate δ_v.[8]

Upon matching, a flow wage $w_t(x, y, \ell)$ is determined by Nash bargaining with worker's bargaining power β. The timing of events in each local labor market is summarized in Figure 1.

Matching. Search is random and matches are created by a constant-returns-to-scale matching function $M(U_t(\ell), V_t(\ell))$, where $U_t(\ell)$ is the measure of unemployed workers and $V_t(\ell)$ is the measure of vacancies.

[5] This assumption is without loss for characterizing a steady-state equilibrium in which neither workers nor firms have an incentive to relocate. In steady state, the values of workers and firms as well as wages along the equilibrium path are unaffected by the possibility of migration. This assumption is useful because it simplifies the exposition, making the interaction between worker and firm sorting more transparent in their values.

[6] Throughout the paper, I use the terms vacancies and jobs to denote individual positions, depending on whether a position is unmatched or matched, respectively.

[7] Business services serve as the counterpart to housing for workers, playing an equivalent role. While separating these markets is crucial for establishing existence, other results hold even with a single local real estate market serving both workers and firms.

[8] I assume, as a normalization, that the unit of vacancy posted by each firm equals the vacancy destruction rate. This ensures that the total measure of vacancies does not mechanically decrease when the vacancy destruction rate increases. I focus on the case in which δ_v is positive, although most key equations can be derived by setting $\delta_v = 0$.

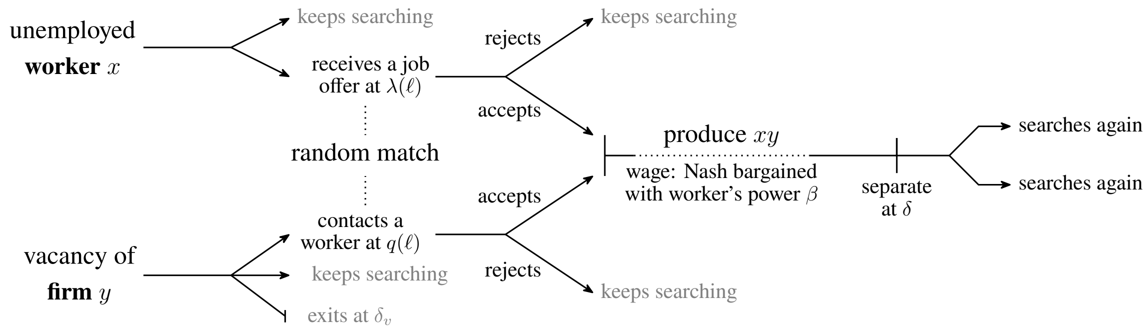


Figure 1. Local Labor Market: Timeline

The matching function $M(\cdot)$ is assumed to be strictly increasing, concave in both arguments, continuously differentiable, and $M(0,V) = M(U,0) = 0$. Defining labor market tightness as $\theta_t(\ell) = \frac{V_t(\ell)}{U_t(\ell)}$, the two matching rates can be represented as functions of market tightness, $\lambda_t(\ell) = \lambda(\theta_t(\ell))$ and $q_t(\ell) = q(\theta_t(\ell))$.

Local suppliers and ownership. Housing $H_t(\ell)$ is competitively supplied by landowners, and the costs of supplying H units of housing are given by $C_r(H)$. In a business services market, competitive intermediaries provide services $S_t(\ell)$ for firms, and it costs them $C_v(S)$ to provide S units of service. I assume that $C_r(\cdot)$ and $C_v(\cdot)$ are twice continuously differentiable, increasing, and convex. I further assume that $C_r'(\cdot)$ and $C_v'(\cdot)$ grow sufficiently large as their arguments go to infinity.[9] All workers own identical diversified portfolios of firms, landowners, and intermediaries.

2.2 Equilibrium

I assume the economy is in steady state. Thus, all equilibrium objects are time-invariant, and I will drop the time subscript t from this point onward. I focus on the class of *pure assignments*, where there is a one-to-one mapping between worker productivity x and location ℓ, as well as between firm productivity y and location ℓ. Let $(x(\ell), y(\ell))$ denote the assignment of workers and firms, which represents their location decisions.[10] Because locations are ex ante homogeneous, without loss of generality, I label ℓ in a way that $x(\ell)$ is strictly increasing. Although non-pure assignment equilibria exist—e.g., uniform random assignment across locations—they are not the focus of this paper. In Section 3.2, I show that the optimal assignment is pure. Therefore, restricting attention to pure assignments does not compromise the generality of the

[9] I provide the precise condition in (A.8). These boundary conditions are required to establish the existence of equilibrium. Once existence is established, the results are independent of the degree of convexity of these functions.

[10] With slight abuse of notation, I will use (x, y) to indicate the productivity of each worker and firm, and $(x(\ell), y(\ell))$ to denote an assignment.

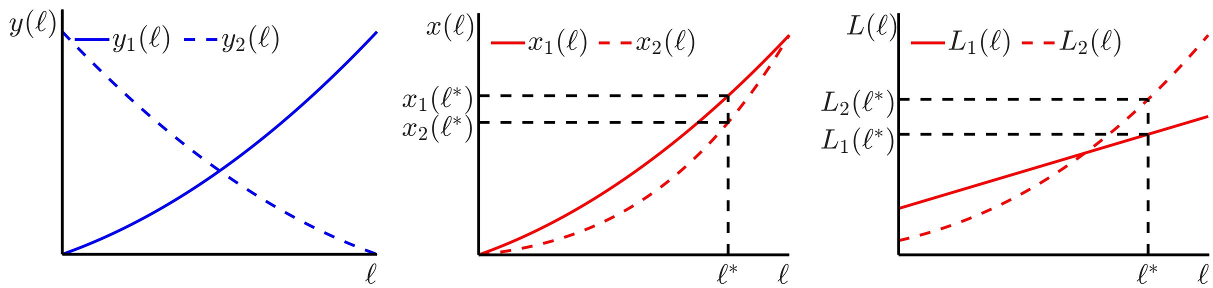


Figure 2. Assignment of Workers and Firms: Examples

inefficiency result in this paper.[11] I also show that any equilibrium consistent with the empirical pattern in which more productive workers and firms locate in densely populated locations must involve a pure assignment. See Proposition A.1 for details.

An assignment $(x(\cdot), y(\cdot))$ pins down the sorting of workers and firms. If $y(\ell)$ increases in ℓ just like $x(\ell)$ as illustrated by $y_1(\cdot)$ in the left panel of Figure 2, the equilibrium exhibits PAM. In contrast, if $y(\ell)$ decreases in ℓ as in the case of $y_2(\cdot)$, the equilibrium exhibits negative assortative matching.

In addition, an assignment characterizes population density $L(\ell)$ and firm density $N(\ell)$, which represent the measure of workers and firms per unit of land. The measure of workers and firms choosing locations between 0 and ℓ equals the measure of their types choosing these locations:

$$\int_0^{\ell} L(\ell')\,\mathrm{d}\ell' = M_w \int_{\{x(\ell'):\ell'\in[0,\ell]\}} \mathrm{d}Q_w(x),$$
$$\int_0^{\ell} N(\ell')\,\mathrm{d}\ell' = M_f \int_{\{y(\ell'):\ell'\in[0,\ell]\}} \mathrm{d}Q_f(y).$$

In other words, population (firm) density is defined as the Radon-Nikodym derivative of the measure of workers (firms) with respect to the area of land. In particular, since $x(\ell)$ is increasing, the first equation simplifies to $\int_0^{\ell} L(\ell')\,\mathrm{d}\ell' = M_w Q_w(x(\ell))$, and population density becomes,

$$L(\ell) = M_w Q_w'(x(\ell))x'(\ell). \tag{1}$$

Population density $L(\ell)$ is endogenously determined in equilibrium, which distinguishes this model from the standard assignment problem.[12] To illustrate, consider the two assignments, $x_1(\cdot)$ and $x_2(\cdot)$, in the middle

[11] In addition, focusing on pure assignments implicitly assumes that all locations are occupied, which is also a property of the optimal assignment. This paper does not address the determination of the total measure of occupied land.

[12] Because population density depends on both $Q_w'(x(\ell))$ and $x'(\ell)$, focusing on pure assignments does not mechanically pin down its distribution across ℓ, unlike in an economy with finitely many worker types and locations. In Section A.5, I present an

panel of Figure 2, with corresponding population densities, $L_1(\cdot)$ and $L_2(\cdot)$, in the right panel. Under $x_2(\cdot)$, more workers are assigned to the interval $[\ell^*, 1]$, as reflected in the fact that $L_2(\cdot)$ exceeds $L_1(\cdot)$ in higher-ℓ locations. This property relies on elastic housing supply; if housing supply were inelastic, $x(\cdot)$ would be uniquely pinned down by the housing market clearing condition.

Values of workers and firms. To characterize the conditions under which a given assignment $(x(\ell), y(\ell))$ constitutes a steady-state equilibrium, I consider the value of an individual worker and firm, taking as given the location choices of all others.

As workers rent $\bar{h}$ units of housing and then use the remaining income to purchase tradable goods, their flow indirect utility is given by flow income net of housing expenditure, $\bar{h}r(\ell)$. Let $V^u(x,\ell)$ denote the value of an unemployed worker of productivity x in location ℓ, and let $V^e(x,y,\ell)$ denote the value of an employed worker of productivity x in location ℓ who is matched with a firm of productivity y. These values are characterized by the following equations:

$$\rho V^u(x,\ell) = bx + \Pi - \bar{h}r(\ell) + \lambda(\ell)\max\{V^e(x,y(\ell),\ell) - V^u(x,\ell), 0\},$$
$$\rho V^e(x,y,\ell) = w(x,y,\ell) + \Pi - \bar{h}r(\ell) - \delta[V^e(x,y,\ell) - V^u(x,\ell)],$$

where Π is the profit from a portfolio minus the lump-sum tax used to finance unemployment benefits. At rate $\lambda(\ell)$, she receives a job offer from a firm of productivity $y(\ell)$ in the local market and becomes employed if the offer is acceptable. An employed worker earns wage $w(x,y,\ell)$ until she becomes unemployed at rate δ.

Let $V^v(y,\ell)$ denote the value of a vacancy (i.e., vacant position) that is posted by a firm of productivity y in ℓ, and let $V^p(x,y,\ell)$ denote the value of a job (i.e., filled position) of productivity y matched with a worker of productivity x in location ℓ. A vacancy contacts a worker at rate $q(\ell)$, and this position is filled with a worker of $x(\ell)$ in the local search pool. It gets destroyed at rate δ_v, which yields no payoff to the firm. A job yields a flow profit $xy - w(x,y,\ell)$ to the firm until it separates at rate δ, at which point the position becomes vacant. These values solve:

$$\rho V^v(y,\ell) = q(\ell)\max\{V^p(x(\ell),y,\ell) - V^v(y,\ell), 0\} - \delta_v V^v(y,\ell), \tag{2}$$
$$\rho V^p(x,y,\ell) = xy - w(x,y,\ell) - \delta[V^p(x,y,\ell) - V^v(y,\ell)].$$

alternative derivation of (1) by first analyzing a *non-pure* assignment in a finite-worker-productivity and finite-location economy, then showing that as the numbers of worker types and locations approach infinity, the assignment becomes pure, and population density converges to (1).

Define the surplus of a match by $S(x,y,\ell) \equiv V^e(x,y,\ell) - V^u(x,\ell) + V^p(x,y,\ell) - V^v(y,\ell)$—i.e., the sum of the worker and job surplus. Then the bargaining problem has a well-known solution, in which a worker and a firm receive constant shares of the surplus:

$$V^e(x,y,\ell) = V^u(x,\ell) + \beta S(x,y,\ell),$$
$$V^p(x,y,\ell) = V^v(y,\ell) + (1-\beta) S(x,y,\ell).$$

Section A.1 presents all derivations. The worker enjoys her reservation unemployment value $V^u(x,\ell)$ plus a share β of the surplus, and the firm takes the remaining share of the surplus. This bargaining rule gives the following flow wage:

$$w(x,y,\ell) = (1-\beta)bx + \beta xy + (1-\beta)\beta\lambda(\ell)S(x,y(\ell),\ell) - \beta(1-\tilde{\beta})q(\ell)S(x(\ell),y,\ell), \tag{3}$$

where $1-\tilde{\beta} \equiv \frac{\rho}{\rho+\delta_v}(1-\beta)$. Wages first increase in the output of a given match, xy, and in the unemployment benefit, bx. In addition, wages depend on location ℓ, which determines the threat points in the bargaining game: when either the job arrival rate $\lambda(\ell)$ or local firm productivity, $y(\ell)$, is higher, the value of search for unemployed workers, $\beta\lambda(\ell)S(x,y(\ell),\ell)$, is greater, which leads to an increase in wages. In contrast, when either the vacancy contact rate $q(\ell)$ or local worker productivity, $x(\ell)$, is higher, the value of vacancies, $(1-\tilde{\beta})q(\ell)S(x(\ell),y,\ell)$, increases, which leads to lower wages.

I combine these equations and obtain the value of a worker of productivity x when choosing location ℓ:

$$\rho V^u(x,\ell) = bx + \underbrace{\frac{\beta\lambda(\ell)}{\tilde{\rho}+\beta\lambda(\ell)}(y(\ell)-b)}_{\equiv A_w(y(\ell),\lambda(\ell))}\left(x - \underbrace{\frac{(1-\tilde{\beta})q(\ell)}{\tilde{\rho}+\beta\lambda(\ell)+(1-\tilde{\beta})q(\ell)}x(\ell)}_{\equiv B_w(x(\ell),\theta(\ell))}\right) + \Pi - \bar{h}r(\ell), \tag{4}$$

when the surplus is positive, i.e., $S(x,y(\ell),\ell) \geq 0$ or equivalently $x \geq B_w(x(\ell),\theta(\ell))$ where $\tilde{\rho} \equiv \rho+\delta$. Otherwise, the worker remains unemployed, and $\rho V^u(x,\ell) = bx + \Pi - \bar{h}r(\ell)$. I refer to the marginal return to worker productivity in the local labor market as *job opportunities*, $A_w(y(\ell),\lambda(\ell))$, which increase in local firm productivity $y(\ell)$ and the job arrival rate $\lambda(\ell)$. The benefits of higher job opportunities, however, may be partially offset by firms' stronger bargaining position, as captured by the term $B_w(\cdot)$. This effect is stronger with higher worker productivity $x(\ell)$ or lower market tightness $\theta(\ell)$. The final term represents the local congestion costs associated with housing expenditure. Workers choose a location that maximizes their value, $V^u(x,\ell)$, and these decisions define the equilibrium assignment of workers, $x(\ell)$.

Next, I solve for the value of a firm of productivity y when operating in location ℓ, denoted $\bar{V}^v(y,\ell)$. At each point in time, a firm posts δ_v units of vacancies, which yield $\delta_v V^v(y,\ell)$, and demands business services at a cost of $c(\ell)$. Hence, the value of a firm equals the discounted sum of these two terms, which is given by

$$\rho\bar{V}^v(y,\ell) = \frac{\delta_v}{\rho}\underbrace{\frac{(1-\tilde{\beta})q(\ell)}{\tilde{\rho}+(1-\tilde{\beta})q(\ell)}x(\ell)}_{\equiv A_f(x(\ell),q(\ell))}\left(y-b-\underbrace{\frac{\beta\lambda(\ell)}{\tilde{\rho}+\beta\lambda(\ell)+(1-\tilde{\beta})q(\ell)}(y(\ell)-b)}_{\equiv B_f(y(\ell),\theta(\ell))}\right)-c(\ell), \tag{5}$$

when the match surplus is positive, i.e., $S(x(\ell),y,\ell)\geq 0$ or equivalently $y-b\geq B_f(y(\ell),\theta(\ell))$. Otherwise, $\rho\bar{V}^p(y,\ell)=-c(\ell)$. The firm's value has a structure similar to that of a worker. I refer to the marginal return to firm's productivity in the local labor market as *hiring opportunities*, $A_f(x(\ell),q(\ell))$, which increase in local worker productivity $x(\ell)$ and the vacancy contact rate $q(\ell)$. The term $B_f(\cdot)$ represents the effect of workers' bargaining position in firm's value. Firms take a smaller share of the surplus when the value of job search for workers is higher due to higher $y(\ell)$ or $\theta(\ell)$. Finally, the value decreases in overhead costs $c(\ell)$, i.e., the local congestion costs of firms. Firms choose a location that yields the highest value, $\bar{V}^v(y,\ell)$, which in turn determines the equilibrium assignment of firms $y(\ell)$.

Housing and business services markets. Local housing rents $r(\ell)$ and overhead costs $c(\ell)$ are determined by market clearing conditions:

$$r(\ell)=C_r'(\bar{h}L(\ell)), \quad c(\ell)=C_v'(N(\ell)). \tag{6}$$

These costs are increasing in $L(\ell)$ and $N(\ell)$, respectively.

Flow-balance conditions. The outflow from and inflow into unemployment, i.e., $\lambda(\ell)u(\ell)L(\ell)$ and $\delta(1-u(\ell))L(\ell)$, must balance, which determines the steady-state unemployment rate $u(\ell)$. Similarly, flows into and out of the vacancy pool must be in balance. The outflows from matching and destruction, i.e., $q(\ell)V(\ell)$ and $\delta_v V(\ell)$, must equal the inflows from separation and new postings, i.e., $\delta(1-u(\ell))L(\ell)$ and $\delta_v N(\ell)$. Since the outflow from matching equal the inflow from separations, it follows that the measure of vacancies

$V(\ell)$ equals firm density $N(\ell)$. The resulting flow-balance conditions are summarized as follows:

$$u(\ell) = \frac{\delta}{\delta + \lambda(\ell)}, \quad V(\ell) = N(\ell). \tag{7}$$

Definition of equilibrium. A pure-assignment steady-state equilibrium consists of the location decisions of workers and firms $(x(\ell), y(\ell))$, population density $L(\ell)$, firm density $N(\ell)$, measure of vacancies $V(\ell)$, unemployment rates $u(\ell)$, housing rents $r(\ell)$, overhead costs $c(\ell)$, and wages $w(x, y, \ell)$, such that (i) workers and firms optimally decide a location, taking the decisions of others as given, (ii) wages are determined by Nash bargaining, (iii) the markets for housing and business services clear, and (iv) the flow-balance conditions hold.

3. Equilibrium Analysis

In this section, I first characterize the positive properties of an equilibrium and then analyze its efficiency. I then discuss the role of random matching, a key assumption, in shaping these results, and close this section by comparing the two-sided sorting mechanism with alternative explanations for spatial disparities.

3.1 Spatial Sorting and Spatial Disparities

The location choices of workers and firms are interdependent because the two interact through a local labor market. Worker's value in (4) has increasing differences in worker productivity x and job opportunities $A_w(\cdot)$, and firm's value in (5) has increasing differences in firm productivity y and hiring opportunities $A_f(\cdot)$. Because I assume that $x(\ell)$ is increasing, job opportunities increase in ℓ in equilibrium based on the standard result from monotone comparative statics (Topkis, 1978). By the same logic, firm productivity $y(\ell)$ and hiring opportunities $A_f(\ell)$ should move in the same direction. However, whether $y(\ell)$ is increasing remains ambiguous, since job opportunities increase in both firm productivity and the job arrival rate.

The analysis below demonstrates that $y(\ell)$ is indeed increasing in ℓ, and hence an equilibrium features PAM between workers and firms across space. To illustrate, suppose instead that firm productivity $y(\ell)$ is lower in a higher-ℓ location, and job opportunities are higher due to a sufficiently higher job arrival rate

$\lambda(\ell)$.[13] Such an allocation is consistent with workers' incentives but violates firms' incentives. A higher job arrival rate requires a larger measure of firms, which increases overhead costs. At the same time, if a higher-ℓ location is chosen by firms with lower productivity, hiring opportunities are smaller in this location. Therefore, firms would prefer to deviate to a lower-ℓ location, which contradicts the assumption and rules out such a non-PAM assignment.[14]

Importantly, the resulting PAM is self-fulfilling. The sorting of workers, represented by $x(\ell)$, and of firms, represented by $y(\ell)$, accounts for increasing $A_f(x(\ell), q(\ell))$ and $A_w(y(\ell), \lambda(\ell))$, respectively, which in turn induces the sorting of the other side. In this equilibrium, productivity heterogeneity, not variation in market tightness, is the primary driver of sorting, as it can raise both job and hiring opportunities, whereas differences in market tightness increase one and decrease the other.

Having established that an assignment exhibits PAM, an equilibrium can be characterized by two strictly increasing functions, $x(\ell)$ and $y(\ell)$, and the optimal location decisions satisfy the following first-order conditions,

$$\left.\frac{\partial \rho V^u(x,\ell)}{\partial x}\right|_{x=x(\ell)} = 0, \quad \left.\frac{\partial \rho \bar{V}^p(y,\ell)}{\partial \ell}\right|_{y=y(\ell)} = 0, \quad \forall x, y. \tag{8}$$

Then, an equilibrium assignment can be obtained by combining the above equations, the boundary conditions—$x(0) = \underline{x}, x(1) = \bar{x}, y(0) = \underline{y}$, and $y(1) = \bar{y}$—the housing and business services market clearing conditions (6), and the steady-state flow-balance conditions (7). The existence of the solution follows from the standard ODE theorem.[15] I summarize the results in the proposition below. See Section A.2 for the proof.

Proposition 1. *A pure-assignment equilibrium exists. Any pure-assignment equilibrium exhibits positive assortative matching between workers and firms across space.*

[13] Search frictions generate differences in local matching rates, which make PAM non-trivial. This is a recurring finding in the literature. Under search frictions, PAM requires a stronger degree of complementarities than supermodularity (e.g., Shimer and Smith, 2001; Eeckhout and Kircher, 2010). In this paper, I assume a supermodular output function, but introducing the spatial sorting of workers and firms allows the model to generate PAM.

[14] The model does not make a specific prediction on spatial variation in market tightness, which depends on productivity distribution, housing market elasticity, and so on. For example, if firm heterogeneity is much smaller than worker heterogeneity, market tightness is likely to increase in ℓ. In this case, more productive workers (firms, respectively) prefer higher-ℓ locations mainly due to higher $\lambda(\ell)$ (higher $x(\ell)$, respectively). This result contrasts with the mechanisms in Bilal (2023) and Kuhn et al. (2022), which predict specific spatial patterns of market tightness. The key distinction in my model is that two-sided heterogeneity in *productivity* serves as the primary driver of the sorting.

[15] A model may have multiple equilibria because the values depend on both productivity and market tightness, similar to Burdett and Coles (1997) and Borovicková and Shimer (2024). This does not pose any concerns for the quantitative analysis in Section 4 because the calibrated model has a unique equilibrium.

Next, I show that higher-ℓ locations chosen by more productive workers and firms are more densely populated. From the workers' first-order condition in (8),

$$\underbrace{\left(\frac{\partial}{\partial \ell} A_w(y(\ell), \lambda(\ell))\right) x(\ell) - \frac{\partial}{\partial \ell}\left(A_w(y(\ell), \lambda(\ell)) B_w(x(\ell), \theta(\ell))\right)}_{\text{changes in labor market benefits}} = \underbrace{\bar{h}\frac{\partial}{\partial \ell} r(\ell)}_{\text{changes in housing rents}}. \quad (9)$$

When choosing a marginally higher-ℓ location, workers benefit from a local labor market with higher job opportunities $A_w(\cdot)$. Due to greater benefits, a larger number of workers are attracted to this location, and they need to pay higher housing rents. This result requires the vacancy destruction rate δ_v to be sufficiently large so that the gain from higher job opportunities is not offset by an increase in firms' threat point $B_w(\cdot)$.[16]

Screening in local labor markets is imperfect under random matching, and this imperfection is central to generating the spatial variation in population density. Because search is random, even less productive workers have the opportunity to be matched with productive firms, as long as they search in locations where such firms are located. In such a case, workers earn higher wages as they enjoy a share of surplus. In contrast, if firms could perfectly screen the workers they hire based on worker productivity, as in neoclassical labor markets or labor markets with directed search, simply participating in a local labor market with more productive firms would not raise wages. I characterize the equilibrium under these alternative labor market structures in Section 3.3 and discuss the differences.

Higher-ℓ locations feature not only greater population density but also higher wages. Substituting the equilibrium assignment $(x(\ell), y(\ell))$ into (3) yields

$$w(x(\ell), y(\ell), \ell) = \left(b + \beta \frac{\tilde{\rho} + \lambda(\ell)}{\tilde{\rho} + \beta\lambda(\ell) + (1 - \tilde{\beta})q(\ell)}(y(\ell) - b)\right) x(\ell). \quad (10)$$

Higher productivity of workers or firms raises output, and greater market tightness improves workers' bargaining positions; both effects increase wages. Under PAM, because both workers and firms are more productive in higher-ℓ locations, wages are correspondingly higher. For this result, the vacancy destruction rate δ_v must again be large enough so that changes in workers' and firms' bargaining positions do not dominate an increase in output.

The following proposition summarizes the two results. See Section A.3 for the proof.

[16] Mathematically, large δ_v ensures that $(1 - \tilde{\beta})$, and thus $B_w(\cdot)$, is small, so that changes in $A_w(\cdot)B_w(\cdot)$ are small enough.

Proposition 2. *Suppose that δ_v is sufficiently large. In any pure-assignment equilibrium, population density $L(\ell)$ and wages $w(x(\ell), y(\ell), \ell)$ increase in ℓ.*

Combining the results in Proposition 1 and Proposition 2, I conclude that an economy with search frictions, where heterogeneous workers and firms sort across local labor markets, can explain key dimensions of the spatial disparities we observe in the data. Importantly, in this economy, the urban wage premium arises across homogeneous locations, and purely stems from the sorting mechanism.

3.2 Efficiency Properties of Equilibrium

In this section, I characterize the properties of the optimal spatial allocation and evaluate the efficiency of the decentralized equilibrium. For expositional simplicity, I focus on the limiting case in which the discount rate ρ goes to zero.[17] The planner maximizes the steady-state flow net output, i.e., the total output minus total congestion costs, which is the standard approach in the literature on frictional labor markets (e.g., Hosios, 1990; Shi, 2001; Acemoglu, 2001).[18] All proofs are presented in Section A.4.

I first characterize the optimal assignment and compare it with the equilibrium assignment. As in the literature, due to complementarity in output, the optimal assignment features PAM, despite the presence of market tightness, which is an additional factor affecting output. The key insight is that, conditional on population and firm densities, the planner's problem reduces to a classic assignment problem, which admits PAM as a solution. Lemma 1 formalizes this result.

Lemma 1. *The optimal spatial allocation exhibits PAM between workers and firms.*

Using Lemma 1, the planner's problem reduces to choosing two increasing functions, $x(\ell)$ and $y(\ell)$. Local output equals the product of employment $(1-u(\ell))L(\ell)$ and the output of a match $x(\ell)y(\ell)$, while congestion costs depend on worker and firm densities. The planner solves

$$\begin{aligned} \max_{x(\ell),y(\ell)} \quad & \int_0^1 \left[(1-u(\ell))L(\ell)x(\ell)y(\ell) - C_r(\bar{h}L(\ell)) - C_v(N(\ell))\right] \mathrm{d}\ell \\ \text{s.t.} \quad & x'(\ell) = \frac{L(\ell)}{M_w Q_w'(x(\ell))},\ y'(\ell) = \frac{N(\ell)}{M_f Q_f'(y(\ell))},\ u(\ell) = \frac{\delta}{\delta+\lambda(\ell)} \quad \forall \ell, \end{aligned}$$

[17] This assumption makes workers and firms infinitely patient, which aligns their location choices with the planner's assignment problem. If not, since location decisions precede search, workers and firms overvalue locations with higher matching rates compared to the planner.

[18] This formulation is equivalent to the problem that maximizes the total surplus when the planner can freely use transfers among agents due to the stylized nature of the economy, e.g., risk neutrality and exogenous labor supply. I further show that the equilibrium is Pareto inefficient in Section A.4.

in addition to the boundary conditions on $x(\ell)$ and $y(\ell)$. Constraints determine how an assignment pins down population density, firm density, and unemployment rates as in Section 2.2. The optimal assignment, denoted by a superscript *, is characterized by two equations described below.

First, the planner spreads out workers and firms across space to reduce congestion costs. Specifically, the planner sets the weighted sum of $\frac{\partial}{\partial \ell} C_r(\bar{h}L^*(\ell))$ and $\frac{\partial}{\partial \ell} C_v(N^*(\ell))$, representing changes in congestion costs, to zero across all ℓ:

$$\underbrace{\bar{h}^2 C_r'(\bar{h}L^*(\ell))(L^*)'(\ell)L^*(\ell) + C_v'(N^*(\ell))(N^*)'(\ell)N^*(\ell)}_{\text{changes in congestion costs of the optimal assignment}} = 0 \quad \forall \ell. \tag{11}$$

The above condition implies that the planner does not make one location strictly denser than another: changes in population and firm densities have opposite signs, $(L^*)'(\ell)(N^*)'(\ell) \leq 0$. Importantly, productivity is embodied in workers and firms. Thus, the planner can reduce congestion costs while maintaining the total output by relocating them from more to less congested locations.

The second condition governs how the planner adjusts market tightness, i.e., whether to increase worker or firm density, according to productivity heterogeneity:

$$\begin{aligned} &\frac{(1-u^*(\ell))}{1-\varepsilon_\lambda^*(\ell)(1-u^*(\ell))}\left((1-\varepsilon_\lambda^*(\ell))\frac{(y^*(\ell))'(\ell)}{y^*(\ell)} - \varepsilon_\lambda^*(\ell)u^*(\ell)\frac{(x^*(\ell))'}{x^*(\ell)}\right)x^*(\ell)y^*(\ell) \\ &= \left[C_r''(\bar{h}L^*(\ell))\bar{h}^2 L^*(\ell) + \frac{\varepsilon_\lambda^*(\ell)u^*(\ell)(1-u^*(\ell))}{(1-\varepsilon_\lambda^*(\ell)(1-u^*(\ell)))^3}\mathcal{E}(\ell)\left(1+\frac{C_r''(\bar{h}L^*(\ell))}{C_v''(N^*(\ell))}\bar{h}^2\left(\frac{L^*(\ell)}{N^*(\ell)}\right)^2\right)x^*(\ell)y^*(\ell)\right]\frac{(L^*(\ell))'}{L^*(\ell)} \end{aligned}$$

where $\varepsilon_\lambda(\ell) = \frac{\lambda'(\theta(\ell))}{\lambda(\theta(\ell))}\theta(\ell)$ is the elasticity of the job arrival rate with respect to market tightness, and $\mathcal{E}(\ell)$ is defined as $\varepsilon_\lambda^*(\ell)(1-\varepsilon_\lambda^*(\ell)) - \frac{(\lambda^*)''(\theta^*(\ell))\theta^*(\ell)}{(\lambda^*)'(\theta^*(\ell)))} + \varepsilon_\lambda^*(\ell) - 1$. If $\mathcal{E}(\ell) \geq 0$, e.g., under constant matching elasticity, the sign of $(L^*)'(\ell)$ follows that of the left-hand side, so more workers and fewer firms are allocated to high-ℓ locations when firm productivity is more heterogeneous. By doing so, the vacancy contact rate of highly productive firms rises, leading to greater output.

Next, I compare the planner's solution to the decentralized equilibrium to evaluate efficiency. In equilibrium, high-ℓ locations provide more attractive labor markets with greater $A_w(\cdot)$ and $A_f(\cdot)$. Thus, the changes in congestion costs are strictly positive,

$$\bar{h}r'(\ell)L(\ell) + c'(\ell)N(\ell) = \underbrace{\bar{h}^2 C_r'(\bar{h}L(\ell))L'(\ell)L(\ell) + C_v'(N(\ell))N'(\ell)N(\ell)}_{\text{changes in congestion costs of the equilibrium assignment}} > 0 \quad \forall \ell.$$

This congestion is already suggested by Proposition 1, which shows that population density is greater in high-ℓ locations. The above is in stark contrast to (11): in equilibrium, workers and firms are inefficiently concentrated in high-ℓ locations.

These inefficiencies arise from externalities in frictional local markets with random matching. As in standard search models with homogeneous agents, individuals do not internalize the impact of their matches on others' matching rates. More importantly, in this economy, there are additional externalities due to worker and firm heterogeneity. When workers choose between two locations, they ignore the effect of their choice on the quality of local search pools. To elaborate, if less productive workers choose a marginally higher-ℓ location, they reduce local firms' chances of hiring more productive workers. The same logic applies to firms (e.g., Acemoglu, 2001; Shimer and Smith, 2001; Albrecht et al., 2010; Bilal, 2023).[19] Because participation in local labor markets is determined through location choices, these externalities can be addressed through location-specific transfers. The expressions for these transfers are lengthy and are provided in the appendix; see (A.18) and (A.19). The below proposition summarizes the results.

Proposition 3. *The decentralized equilibrium is inefficient. Each worker and each firm chooses a higher-ℓ location than the social planner would designate, given the location decisions of all others. The planner can implement the optimal assignment by using the spatial transfers to workers and firms.*

Importantly, Proposition 3 holds independent of the elasticity of matching function and the workers' bargaining power. Because *both* workers and firms sort inefficiently across space, neither the standard nor the generalized Hosios condition (e.g., Mangin and Julien, 2021) can restore efficiency. For example, increasing workers' bargaining power in a higher-ℓ location may lower firm inflows but increase worker inflows even more. To see this more clearly, I turn off the externalities arising from changes in matching rates and focus only on those arising from changes in the quality of search pools. In this case, the equilibrium is efficient if and only if both workers and firms are homogeneous, and the planner taxes those choosing high-ℓ locations to reduce unnecessary congestion. In particular, the magnitude of these taxes increases with heterogeneity, measured by $(x^*)'(\ell)$ and $(y^*)'(\ell)$.

Corollary 1. *Assume the Hosios condition, i.e., $\frac{\lambda'(\theta)}{\lambda(\theta)}\theta = 1-\beta$, and zero unemployment benefit $b = 0$. Then, an equilibrium is efficient if and only if workers and firms are homogeneous. Moreover, the spatial transfers*

[19] In most papers, due to externalities, the composition of jobs is inefficiently biased toward low-productivity jobs. In this paper, because I focus on pure assignments, the distributions of workers and firms are degenerate in each ℓ. Instead, externalities arise from choosing a higher-ℓ location than is socially optimal. The most related application is Bilal (2023) who applies this idea to the spatial sorting model of heterogeneous firms.

to workers and firms, $t_w(\ell)$ and $t_f(\ell)$, decrease in ℓ,

$$t_w(\ell) = t_w^0 - \int_0^\ell \frac{(1-\beta)(1-u^*(t))u^*(t)}{1-(1-\beta)(1-u^*(t))} y^*(t) x^{*'}(t)\,\mathrm{d}t,$$
$$t_f(\ell) = t_f^0 - \int_0^\ell \beta(1-u^*(t)) \frac{L^*(t)}{V^*(t)} x^*(t) y^{*'}(t)\,\mathrm{d}t,$$

where the constants t_w^0 and t_f^0 balance the government budget.

The above results offer a new way to think about the role of geography in shaping labor market outcomes. Although PAM is facilitated by spatial sorting across local labor markets, this comes at a price in the form of congestion costs. This discussion is reminiscent of the signaling equilibrium in Spence (1973), where more productive workers send a costly signal to distinguish themselves from less productive ones. Workers and firms match randomly within each local labor market, with location serving as the only source of information about each other's type. More productive workers and firms signal their higher productivity by choosing more expensive locations. However, the presence of these congested areas suggests inefficiencies. Nevertheless, PAM outperforms random matching when the output gains from complementarity exceed the additional congestion costs.[20]

3.3 The Role of Random Matching in Frictional Labor Markets

To understand the role of imperfect screening under random matching, I compare the baseline to cases in which firms can screen workers. Specifically, I consider two models with alternative labor market structures: neoclassical labor markets and labor markets with directed search. Unlike the baseline model, these two models fall short of explaining the spatial disparities observed in the data, in particular differential population densities across regions. Moreover, in both cases, the resulting equilibrium is efficient.

Neoclassical local labor markets. The matching remains bilateral, that is, each firm hires only one worker, but labor markets are now frictionless: firms can choose the type of workers they hire in each ℓ, as long as they pay the corresponding wage. Specifically, for a given local wage schedule $w(x,\ell)$, each firm hires a worker of productivity x that maximizes its profit, $xy - w(x,\ell)$. I maintain all other assumptions. Each

[20] Unlike Hoppe et al. (2009), this paper cannot identify conditions on the productivity distribution under which PAM dominates random matching, since the result also depends on the convexity of $C_r(\cdot)$ and $C_v(\cdot)$. For example, when total housing and business services are nearly fixed, the differences in total congestion costs between PAM and random matching become negligible, which ensures that PAM is optimal.

local labor market has the same environment as the seminal work by Becker (1973), except that markets are segmented by location. The following proposition characterizes a pure assignment equilibrium. See Section A.6 for the proof.

Proposition 4. *A differentiable pure assignment equilibrium exists and is unique with the following properties:*

(1) Positive assortative matching between workers and firms arises across space: firm productivity $y(\ell)$ increases in ℓ just like worker productivity $x(\ell)$.

(2) Population density $L(\ell)$ is the same across ℓ.

(3) The equilibrium is efficient. Moreover, matching between workers and firms, wages by worker type, and profits by firm type are equal to those of an economy with a single, nationwide labor market.

In the absence of search frictions, PAM arises straightforwardly in equilibrium. Given workers' location choices, characterized by an increasing function $x(\ell)$, the firm's value has increasing differences in firm productivity y and location ℓ, from which PAM follows directly.

Population density is uniform across ℓ because in equilibrium wages are constant across ℓ, i.e., $\frac{\partial}{\partial \ell} w(x, \ell) = 0$, eliminating workers' incentives to choose dense and expensive locations. Suppose, to the contrary, that $\frac{\partial}{\partial \ell} w(x, \ell) > 0$ and hence $r'(\ell) > 0$. Because firms optimally choose the worker type x in ℓ, the derivative of profit with respect to x is zero, and thus these higher wages must be compensated by lower overhead costs $c'(\ell) < 0$ in equilibrium. With $L'(\ell) > 0$ and $N'(\ell) < 0$, the resulting imbalance in worker and firm densities cannot be sustained without search frictions, leading to a contradiction.

More importantly, the third property indicates geography is economically irrelevant in this model. Intuitively, when workers and firms can selectively match with each other without frictions, the presence of locations does not change matching outcomes.

Directed search. In Section A.7, I consider an economy in which each local labor market is subject to search frictions, but workers and firms engage in directed (competitive) search (e.g., Moen, 1997). In each labor market, firms post wages specific to worker types, and workers optimally queue for firms. These decisions determine both wages and matching rates. Specifically, I extend Eeckhout and Kircher (2010), who study directed search under two-sided heterogeneity, by incorporating a preliminary stage in which workers and firms select locations before competitively searching among those in the same location.

The key distinction from the baseline is that firms would not hire less productive workers unless doing so yields the same surplus as hiring more productive ones in the local market. Thus, less productive workers internalize their negative impact on the surplus of local firms.

The equilibrium exhibits the same properties as in Proposition 4. PAM between workers and firms arises, and the equilibrium does *not* feature spatial concentration: workers and firms are uniformly distributed across space. The equilibrium is efficient, as is typically the case under directed search. Finally, the introduction of location has no effect on all outcomes. See Proposition A.2 for the formal result.

3.4 Comparing Two-Sided Sorting with Alternative Mechanisms Explaining Spatial Disparity

Although this paper shows that two-sided sorting can account for spatial disparity, the literature has demonstrated that other mechanisms can do so as well. In this section, I distinguish the two-sided sorting mechanism from these alternatives along both positive and normative dimensions.

While maintaining the structure of the baseline model, I introduce the two additional sources of productivity heterogeneity in the output function: exogenous location productivity $\bar{A}(\ell)$ and agglomeration forces $A^x(x(\ell))$. Specifically, the output of a worker of productivity x and a firm of productivity y in location ℓ is given by $A(\ell)xy$, where local TFP $A(\ell)$ equals $\bar{A}(\ell)A^x(x(\ell))$. I focus on agglomeration forces arising from knowledge spillovers and assume that $A^x(x(\ell))$ increases in the quality of local workers $x(\ell)$.[21]

With these modifications, the expressions for the value of workers and firms choosing location ℓ remain almost the same. The wage equation is also similar, except for the additional terms reflecting local TFP,

$$\log w(x(\ell), y(\ell), \ell) = \log\left(b + \beta \frac{\tilde{\rho} + \lambda(\ell)}{\tilde{\rho} + \beta\lambda(\ell) + (1-\tilde{\beta})q(\ell)} (\bar{A}(\ell)A^x(x(\ell))y(\ell) - b)\right)x(\ell). \tag{12}$$

All derivations and proofs are in Section A.8.

This extended model relates closely to several important mechanisms in the literature, depending on the source(s) of heterogeneity. I compare the baseline *two-sided sorting* model to three benchmarks. First, when the spatial disparity arises solely from exogenous location productivity $\bar{A}(\ell)$, the model corresponds to the *no sorting* model. Second, when spatial disparity arises from worker productivity $x(\ell)$ and exogenous location productivity $\bar{A}(\ell)$, it reduces to the *one-sided sorting* model of workers. Lastly, incorporating worker

[21] Agglomeration forces can take various forms. The most commonly used assumption is that local TFP increases in population density or city size (e.g., Kline and Moretti, 2014). Comparing this mechanism with the two-sided sorting mechanism is more difficult because its normative implications depend heavily on the functional form.

heterogeneity $x(\ell)$ alongside agglomeration forces $A^x(x(\ell))$ yields the *spillovers* model. The following proposition compares these models, highlighting their similarities and differences.

Proposition 5. *All models fit the observed cross-sectional spatial disparities—wages, population density, and unemployment rates—equally well. However, they yield different optimal allocations and require different spatial transfers to restore efficiency.*

The two-sided sorting mechanism matches cross-sectional moments just as well as alternative mechanisms because output and wage levels do not depend on where productivity is embodied. For example, in the one-sided sorting model, higher wages arise from higher $\bar{A}(\ell)$ rather than from higher $y(\ell)$, as in the two-sided sorting model. In terms of matching data, the mapping between the two models simply involves relabeling $y(\ell)$ as $\bar{A}(\ell)$. Hence, cross-sectional data alone cannot identify the underlying mechanism, which explains why prior studies could explain the same empirical patterns using different mechanisms.

However, each model calls for different policy interventions. I first compare the *two-sided sorting* model with the *no sorting* and *one-sided sorting* models, which feature exogenous location productivity. The key property of the two-sided sorting model is that productivity is embodied in workers and firms, and thus the planner avoids creating unnecessary congestion as shown in (11). In contrast, when locations are ex ante heterogeneous, the location of production directly affects output. In particular, the planner assigns more workers and firms to locations with higher $\bar{A}(\ell)$, which creates congestion in cities, i.e., $L'(\ell) > 0$ and $N'(\ell) > 0$, even in the efficient allocation.[22]

Although both the two-sided sorting and *spillovers* models feature ex ante homogeneous locations and thus call for no congestion in the efficient allocation,[23] a key difference is that the spillovers model does not incorporate firm heterogeneity. As a result, the planner imposes no spatial transfers on firms, while applying them to workers more aggressively. Moreover, this distinction matters for real-world policies that explicitly target firms. For example, place-based policies that attract high-performing firms to low-income locations increase net output in the spillovers model, but disrupt PAM and reduce allocative efficiency in the two-sided sorting model. In general, even when the two models share certain features, it is more efficient to design policies that directly address firms whenever firm heterogeneity is present.

[22] The equilibrium in the one-sided sorting model is inefficient because of search frictions, which are retained to make minimal changes to the baseline model. Alternatively, one can develop a one-sided sorting model without search frictions that replicates spatial disparity and yields an efficient equilibrium.

[23] Although the two models appear similar in this section, this need not be the case. In general, efficient allocations in spillover models crucially depend on parameterization, and spillovers may operate among worker types rather than between workers and firms (e.g., Fajgelbaum and Gaubert, 2020; Rossi-Hansberg et al., 2019; Behrens et al., 2014).

In sum, while different mechanisms can be equally effective in matching the cross-sectional data, the normative implications and the corresponding policy requirements can vary significantly.

4. Quantitative Analysis

In this section, I first calibrate the two-sided sorting model using U.S. cross-sectional data, which illustrates its quantitative potential to account for spatial disparity. Next, I evaluate spatial policies using the calibrated model and contrast these results with those predicted by the models with alternative mechanisms in Section 3.3. This analysis illustrates how the theoretical results in Proposition 5 translate into real-world policy evaluations.[24]

4.1 Quantitative Model

Setting. To bring the model to the data, I generalize preferences to a Stone-Geary utility and assume that workers' flow utility is given by $g^{1-\omega}(h-\bar{h})^{\omega}$, where g denotes tradable goods and h denotes housing consumption. This change enables the model to accurately capture the impact of higher housing rents, which are central to workers' location decisions.

I introduce two policies that I evaluate in Section 4.4. First, I introduce local housing regulations. In particular, I assume that local governments tax housing production at rate $\tau_h(H,\ell) = (H/T)^{t_h(\ell)} - 1$, which depends on housing supply H and the stringency of local regulation $t_h(\ell)$. Local governments redistribute tax revenues to workers in the same region as a lump sum. The second policy is the federal income tax. Workers pay a fraction $\tau_w(\ell)$ of their labor income, either wages $w(x,y,\ell)$ or unemployment benefit bx. To approximate the progressivity of the tax schedule, I assume that the tax rate is location-specific and increasing in ℓ.[25] Federal income tax revenues are redistributed to all workers as a lump sum.

With these changes, the value of unemployed workers becomes

$$\rho V^u(x,\ell) = r(\ell)^{-\omega}((1-\tau_w(\ell))(bx + A_w(y(\ell),\lambda(\ell))(x - B_w(x(\ell),\theta(\ell)))) - \bar{h}r(\ell) + \Pi + T_r(\ell)), \quad (13)$$

[24] This section does not aim to quantify the contribution of two-sided sorting to spatial disparity, which is not feasible using cross-sectional data alone as established in Proposition 5. I return to this issue in Section 5.

[25] Assuming location-specific tax rates preserves tractability. If, instead, tax rates vary with individual income, the wage determined by the bargaining process cannot be characterized by the Nash bargaining solution.

where Π is the sum of lump-sum redistributions from firms' profits, income taxes, landowners' profits, and intermediaries' profits, net of taxes for unemployment benefits. Workers also receive locally redistributed housing tax revenues $T_r(\ell)$. See Section B.1 for details.

Functional forms. I assume that $M(U,V) = \mathcal{A}U^{\alpha}V^{1-\alpha}$. I use a housing production cost function with a constant elasticity, $C_r(H) = H_w^{-1/\eta_w} H^{1+1/\eta_w}/(1+1/\eta_w)$, which is commonly used in the literature. I use a similar function for business services $C_v(S) = H_f^{-1/\eta_f} S^{1+1/\eta_f}/(1+1/\eta_f) + c_e S$. Note that all functions are common across regions, so I do not assume any ex-ante heterogeneity across space except for two policies.

4.2 Estimation

Data. My primary data source is the American Community Survey (ACS) 2017 from IPUMS (Ruggles et al., 2023). I compute the average annual earnings, unemployment rates, housing rents, and housing spending shares for each MSA. Before computing local averages, I regress wages on demographics and 1-digit industry, unemployment status on demographics, and housing rents on building characteristics, and use the residuals. I complement this dataset with additional sources, including the Current Population Survey (CPS) from IPUMS (Flood et al., 2022), the U.S. Census, and the Bureau of Economic Analysis (BEA). See Section B.2 for further details.

To map each MSA to ℓ, I order MSAs by population density, which increases in ℓ.[26] These ranked MSAs are grouped into 20 equally populated bins, each corresponding to an interval of ℓ of equal length.[27] I compute the averages of population densities, wages, unemployment rates, housing rents, and housing spending shares for each bin, weighting each MSA with its population. Finally, I compute the equilibrium at a monthly frequency.

Externally set or calibrated parameters. I first discuss the parameters that are externally calibrated, which are summarized in the top panel of Table 1. I set the discount rate $\rho = 0.004$, which implies an annual real interest rate of 5%. The matching elasticity is set to $\alpha = 0.5$, the standard value in the literature. The separation rate $\delta = 0.028$ is chosen to match the average monthly transition probability from employment into unemployment. Given α and δ, I compute the measure of vacancies $V(\ell)$ using (7). I calibrate $\mathcal{A}$ to

[26] Since locations are ex ante homogeneous, ordering regions by an endogenous outcome is unavoidable.

[27] For improved figure visibility, I generalize the distribution of ℓ so that ℓ represents the percentile rank of workers across locations. This entails a simple re-indexing, as locations are ex ante homogeneous.

Table 1: Parameter Values

Parameter		Target	Value
Discount rate	ρ	Interest rate	0.004
Matching elasticity	α	Literature	0.5
Separation rate	δ	EU transition rate	0.028
Matching efficiency	$\mathcal{A}$	Market tightness	0.62
Housing demand	$\bar{h}, \omega$	Spending shares on housing	10.10, 0.11
Housing supply	η_w	Housing rents	11.31
Unemployment benefit	b	Replacement rate	0.28
Worker's bargaining power	β	Labor share	0.04
Business services supply	H_f, η_f	Wages	3.62, 14.68
Housing tax	T	Housing rents	1028.35

Notes: The top panel shows parameters that are externally calibrated, and the bottom panel shows parameters that are internally calibrated. In addition, worker and firm productivities $\{x(\ell), y(\ell)\}_\ell$ are calibrated.

match the ratio of the number of vacancies over the number of unemployed workers in the U.S. I assume the destruction rate of a vacancy, δ_v, is sufficiently large. Specifically, I set $\frac{\rho}{\rho+\delta_v} = 0$, which effectively implies that firms' threat point in bargaining is zero.[28]

Internally calibrated parameters. The remaining parameters include $\{\beta, b, \bar{h}, \omega, \eta_w, H_w, \eta_f, H_f, c_e\}$ and the productivity of workers and firms across space $\{x(\ell), y(\ell)\}_\ell$. I normalize the minimum values of worker productivity, firm productivity, and housing rents, which together determine the scale of an economy. Specifically, I set $y(0)$ and $r(0)$ to 1, and choose $x(0)$ to match the wage of the lowest bin. The housing market clearing condition at $\ell = 0$ pins down H_w. As a normalization, I choose c_e such that the average profit of firms is zero. This leaves seven structural parameters, $\{\bar{h}, \omega, \eta_w, b, \beta, H_f, \eta_f\}$, and productivity schedules $\{x(\ell), y(\ell)\}_\ell$.

I adopt different strategies for estimating these two groups. Structural parameters are calibrated by targeting standard moments, as described below. Given these parameters, I solve for $\{x(\ell), y(\ell)\}_\ell$ that satisfy

[28] This choice has three advantages. First, as stated in Proposition 1, the parameter δ_v needs to be sufficiently large to rationalize the urban wage premium. Second, it sets the continuation value of a vacancy to zero, comparable to the standard assumptions such as free entry in Pissarides (2000) or no capacity constraint in Postel-Vinay and Lindenlaub (2023), both of which are widely used for tractability. This tractability is particularly useful in Section 5. Lastly, separately identifying δ_v and β is challenging.

equilibrium conditions (8) and (6), conditional on population densities and unemployment rates $\{L(\ell), u(\ell)\}_\ell$, both of which are included among the targeted moments.[29]

I estimate parameters by minimizing the distance between the vector of targeted moments, $\hat{m}$, and the model counterpart, $m(\Theta)$,

$$(\hat{m} - m(\Theta))' \mathscr{W} (\hat{m} - m(\Theta)), \quad \text{where } \Theta = \{\bar{h}, \omega, \eta_w, b, \beta, H_f, \eta_f\} \cup \{L(\ell), u(\ell)\}_\ell.$$

The matrix $\mathscr{W}$ is diagonal, containing the reciprocals of the squared data moments. Because the values of $L(\ell)$ and $u(\ell)$ at each bin are included in $\hat{m}$, I divide the corresponding diagonal elements of $\mathscr{W}$ by the total number of bins to avoid overweighting these moments.

Intuition for identification. Table 1 summarizes parameters along with the most relevant moments. Although formal identification is not feasible, I explain how each variable can be pinned down by certain moments, using heuristic arguments. The selection of moments used to calibrate structural parameters is summarized here, with a more formal treatment deferred to Section B.3. The average housing spending share determines ω, and spatial differentials across ℓ determine $\bar{h}$. Conditioning on the housing demand and regulations, housing rents identify η_w and T. The average replacement rate and the labor share increase in b and β, respectively. Finally, the sorting condition of firms disciplines (H_f, η_f).

Next, the identification of productivity schedules $\{x(\ell), y(\ell)\}$ follows a revealed preference argument. Consider workers of $x(\ell')$ who are almost indifferent between ℓ' and a nearby location ℓ''. Using their wages in ℓ' and the housing rent differential, I compute their hypothetical wages in ℓ'', which must be higher to compensate for higher housing rents there. The gap between these hypothetical wages, $w(x(\ell'), y(\ell''), \ell'')$, and the observed wages in ℓ'', $w(x(\ell''), y(\ell''), \ell'')$, identifies the difference between $x(\ell')$ and $x(\ell'')$. After accounting for worker heterogeneity, the remaining wage gap between ℓ' and ℓ'' is attributed to differences in firm productivity. For example, when the difference in housing rents is large, the implied worker heterogeneity is small, suggesting greater heterogeneity among firms.

[29] An alternative approach is to parameterize the distributions of worker and firm productivity and estimate their parameters. Compared to this more conventional approach, my method avoids parametric assumptions and is computationally much faster.

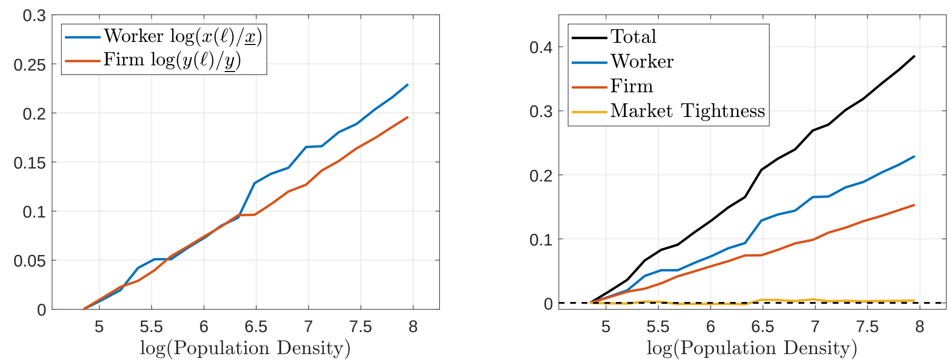


Figure 3. Worker and Firm Productivity (left) and Urban Wage Premium (right)

Notes: Each line in the right panel represents the cross-sectional wage resulting from differences in worker productivity $x(\ell)$, firm productivity $y(\ell)$, and market tightness $\theta(\ell)$, while keeping all other variables at their average values. Worker and firm heterogeneity account for approximately 60% and 40% of spatial wage differentials, respectively, while variation in market tightness plays a negligible role.

4.3 Estimation Results

Parameters. The right column of Table 1 summarizes the estimation results. The estimated housing subsistence level $\bar{h}$ suggests a high degree of non-homotheticity in housing preferences. Although the estimated housing elasticity η_w, at 11.31, exceeds typical estimates in the literature (e.g., Saiz, 2010; Green et al., 2005), once accounting for regulation, the average effective housing elasticity is 7.95, which is significantly smaller. The bargaining power of workers β is estimated to be 0.04, which is comparable to the estimate of Bilal (2023), who uses a similar wage-setting framework.

In the left panel of Figure 3, I plot the estimated productivity of workers and firms across regions. Productivity varies substantially across locations for both, with worker productivity displaying greater dispersion. For example, workers and firms in the top 10% cities are 24.4% and 20.4% more productive, respectively, than those in the bottom 10%. The right panel shows how this heterogeneity translates into the urban wage premium. Workers and firms account for about 60.4% and 38.8% of the urban wage premium, respectively.

Model fit. Despite its simplicity, the model successfully replicates the spatial disparities observed in the data. Table 2 reports the fit of 10 targeted structural moments, and Figure 4 plots moments from the data and model for log wages, housing rents, and housing spending shares across ℓ. The model matches these moments well. It also reproduces the overall patterns of population densities and unemployment rates $\{L(\ell), u(\ell)\}_\ell$; see

Table 2: Model Fit

	Wage			Housing		Replac. rate	Labor share	Rent		
Quartile	2	3	4	mean	$4^{th} - 1^{th}$			2	3	4
Target $\hat{m}$	0.091	0.229	0.351	0.330	0.044	0.500	0.600	0.079	0.343	0.641
Model $m(\Theta)$	0.092	0.226	0.353	0.320	0.046	0.500	0.604	0.147	0.353	0.644

Notes: For wages and housing rents, I first compute the average for each four quartile group. Then, for $i = 2, 3, 4$, I target $(\text{avg}_i/\text{avg}_1) - 1$, where avg_i denotes the average for the i-th quartile group. For variation in housing spending shares, I target the difference between those in the last quartile and the first quartile.

Figure A.2a. The slope of $\{L(\ell)\}_\ell$ and the flat profile of $\{u(\ell)\}_\ell$ are captured well, though unemployment rates at individual locations are not, since the model is not suitable for capturing irregular spatial variation.

In sum, two-sided sorting alone can account for spatial disparity, not only theoretically but also quantitatively, even without location-specific factors.

Identification threat. Because the model attributes spatial disparities solely to worker and firm heterogeneity, the omission of other contributing factors may bias estimation results. For example, if local TFP varies across ℓ as in the extended model in Section 3.3, the estimation strategy based on a revealed preference argument would attribute this variation to firm heterogeneity. As shown in Proposition 5, since the model is estimated by matching cross-sectional moments, worker and firm heterogeneity cannot be separately identified from local TFP. While quantifying the relative contribution of each mechanism is beyond the scope of this paper, in Section 5, I use a time-difference specification and an instrumental variable strategy to provide suggestive evidence on the quantitative importance of two-sided sorting.

4.4 Policy Evaluation

Spatial policies. Many real-world policies influence the location decisions of workers and firms. Examples include place-based policies, variations in local government tax systems, and discrepancies in the stringency of housing regulations. I focus on the two policies that are incorporated in Section 4.1: housing regulations, analyzed below, and the federal income tax of workers, discussed in Section B.6. In the U.S., housing regulations are much more stringent in dense cities, leading to smaller housing elasticities (Saiz, 2010).

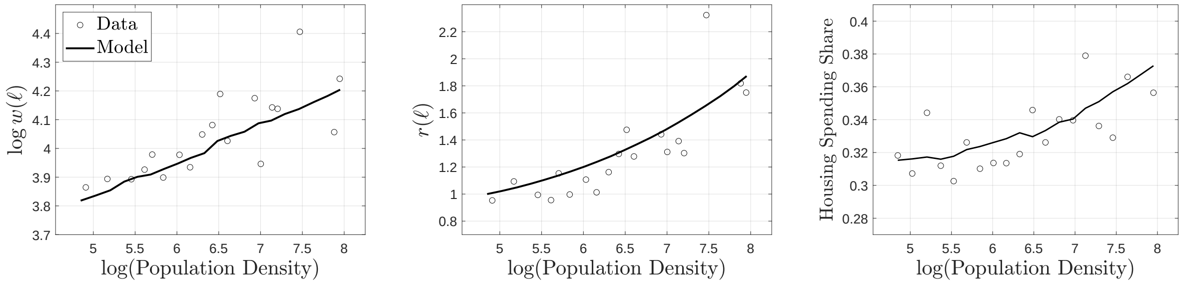


Figure 4. Model Fit: Wage, Housing Rent, and Housing Spending Share

Notes: Data source: ACS (2017). Each dot represents 5% of the population. I compute average values for each dot by weighting each MSA with its population.

Previous studies claim that relaxing housing regulations in major cities can attract numerous workers into these 'productive' regions, leading to a substantial increase in GDP (e.g., Hsieh and Moretti, 2019).[30]

To implement a similar policy experiment, I relax the housing regulations in dense areas $\ell \in [0.9, 1]$ by reducing the housing tax rate $\tau(H;\ell)$ to the median level observed at $\ell = 0.5$. The left panel of Figure 5 plots housing elasticities, taking regulations into account; housing supply becomes more elastic in dense regions after deregulation. I compare two steady-states before and after the policy change.[31]

Results. Following the policy change, workers and firms move toward dense cities where housing becomes more affordable. The middle panel of Figure 5 illustrates that population density in the top 10% of regions increases by 55.3%. The changes are substantial, as they reflect long-run adjustments. Importantly, this inflow of workers increases the vacancy contact rate and thus attracts more firms to high-ℓ locations, which leads to a similarly large relocation of firms. Housing rents are lower either because of relaxed regulations or lower population density across ℓ, as shown in the right panel of Figure 5.

Despite the reallocation, the change in aggregate output is marginal, at just -0.01%. Even after relocation, workers continue to be matched with similar firms, so the output of each match and thus aggregate output remain roughly unchanged. In contrast, as workers move toward dense cities, congestion costs from housing and business services markets increase by 1.36%.

[30] Recent studies have examined how incorporating worker sorting and agglomeration forces can have additional impacts on similar policies. The conclusions vary; some suggest that agglomeration forces can mitigate the effects of policies due to endogenous changes in local productivity (e.g., Martellini, 2022), while others indicate that they can further contribute to an increase in the aggregate growth rate (e.g., Crews, 2023).

[31] While the model admits the possibility of multiple equilibria in theory, the calibrated version yields a unique equilibrium.

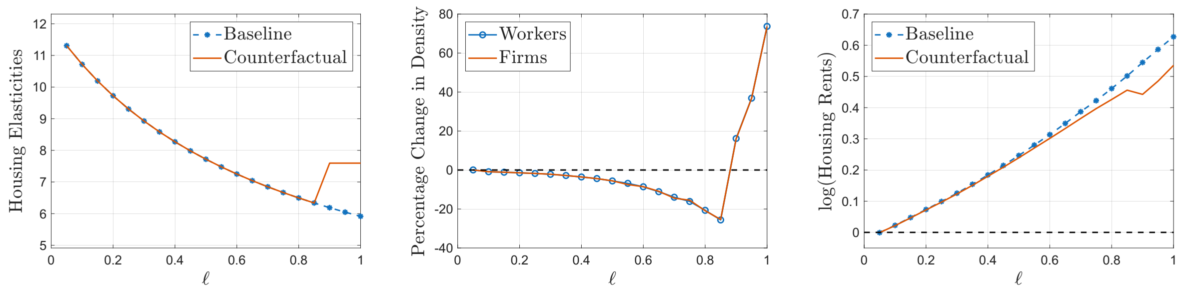


Figure 5. Impact of Relaxing Housing Regulations on Spatial Allocation

Aggregate welfare of the utilitarian planner, i.e., an unweighted average of the values of workers, *decreases* by 0.21%, consistent with Proposition 3. In the estimated model, stringent housing regulations in dense areas serve as lower spatial transfers, and thus lifting regulations lowers welfare. To investigate the sources of this decline, I decompose the total welfare changes into three components: firm productivity, housing market, and labor market tightness. Specifically, I compute changes in *each* worker's value due to changes in outcomes relevant to each component using (13). As shown in the left panel of Figure 6, for most workers, the welfare loss is driven primarily by housing markets: as concentration increases, the reduction in transfers due to lower landowner profits more than offsets the benefit of lower housing rents. In contrast, welfare changes from firm productivity and market tightness are modest and exhibit no clear pattern.

Comparison with one-sided sorting. I now examine the importance of incorporating both worker and firm sorting by comparing the two-sided sorting model with the benchmark *one-sided* sorting model described in Section 3.4. In the one-sided sorting model, heterogeneous workers sort across locations with exogenous location productivity $\bar{A}(\ell)$, while firms are homogeneous. The estimated $\bar{A}(\ell)$ is identical to firm productivity $y(\ell)$ in the two-sided sorting model, and thus both models produce identical cross-sectional wages, as implied by Proposition 5. See Section B.6 for details.

Failing to account for firm sorting, however, leads to an overly optimistic assessment. Following the same policy change, workers and firms relocate toward high-ℓ locations, as in the two-sided sorting case. However, the key difference is that workers produce more output due to higher local TFP, which leads to an aggregate output increase of 0.40%. This gain in output is large enough to raise aggregate welfare by 0.37%. The right panel of Figure 6 confirms that workers benefit mostly from local TFP, whereas the left panel exhibits limited effects from firm productivity. The result from the one-sided sorting model is consistent with the conclusion of previous studies, which assume that productivity is fully or partly embodied in locations.

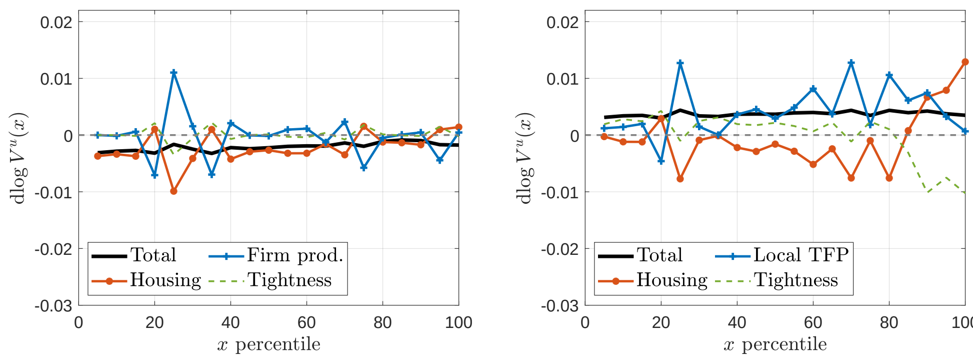


Figure 6. Welfare Effects of Relaxing Housing Regulations: Two-Sided (Left) vs. One-Sided (Right)

Notes: The left panel plots changes in unemployed workers' value in the two-sided sorting model, while the right panel plots those in the one-sided sorting model. Results for employed workers are qualitatively similar. To isolate the contribution of each component, I vary the relevant variables together with the associated changes in transfers, holding all other elements fixed. For the firm productivity component, I vary $y(\ell)$. The housing market component reflects changes in $r(\ell)$ and $c(\ell)$. For the labor market tightness component, I adjust $\theta(\ell)$. In the one-sided sorting model, the decomposition remains analogous, except for the local TFP component, in which I vary $\bar{A}(\ell)$, rather than $y(\ell)$.

5. Empirical Evidence of Two-Sided Sorting

In this section, I provide empirical evidence of two-sided sorting. Since various mechanisms can yield observationally equivalent cross-sectional outcomes, identification of two-sided sorting is challenging. To address this, I exploit spatial variation in changes in worker and firm sorting between two periods together with an instrumental variable strategy. Although a quantitative decomposition of these mechanisms is beyond the scope of this paper, I show that two-sided sorting accounts for a significant portion of the observed spatial disparities.

Data. An empirical analysis in this section requires a matched employer-employee dataset, and I therefore use German administrative microdata. I use linked employer-employee data (LIAB) from the Institute for Employment Research (IAB), which is generated by linking an annual establishment survey and individual employment information. I use 257 commuting zones (CZs) to capture Germany's local labor markets.[32]

The dataset provides two-way fixed effects, which are estimated from a log-wage regression with additive worker and firm fixed effects.[33] These estimates are based on the IAB Employment History File (BEH),

[32] I only observe the CZ of establishments, and thus I define the CZ of workers based on their employment locations. For the period of unemployment, I assign CZ based on their next job, which is consistent with the timing of the model.

[33] The dataset is provided at *establishment*-level. Since this paper does not distinguish between firms and establishments, both of which are treated as collections of jobs, I use the terms firms, establishments, and jobs interchangeably throughout this section.

which represents the universe of workers, subject to social security contributions, and are obtained following the strategy of Card et al. (2013). Estimates are provided for different periods, and I focus on estimates of 2003-2010 and 2010-2017.

Measuring productivity. Local worker productivity in firms' value (5), $x(\ell)$, represents the average worker productivity that firms expect when posting vacancies. To construct its empirical counterpart, I compute the average productivity of workers in the local search pool; I use the average productivity to be consistent with random matching, and include all workers eligible for new hires, whether they were unemployed or recently switched jobs. Similarly, firm productivity choosing a location ℓ, $y(\ell)$, reflects the quality of vacancies posted in that local labor market. Its empirical counterpart is the average productivity of newly created jobs; I focus on jobs since this paper treats jobs as the unit of production, and include all new jobs, whether filled by incumbents or entrants. Hence, variation in $y(\ell)$ partly reflects compositional shifts among incumbent establishments.

To estimate $x(\ell)$ and $y(\ell)$, I map two-way fixed effects to the wage components of the model. Two-way fixed effects are widely used in the literature as proxies for worker and firm productivity (Abowd et al., 1999). From equilibrium wages (10), worker fixed effects correctly represent the log of their productivity. By contrast, fixed effects of new jobs correspond to the term in parentheses in (10), which depends on the job arrival rate $\lambda(\ell)$ and other parameters. I therefore invert the expression to obtain $y(\ell)$, which I explain in detail in Section C.1.[34]

In this section, I employ empirical strategies that rely on features absent from the baseline model, e.g., using two-way fixed effects and an instrumental variable approach. In Section C.4, I justify these strategies using an extended model that features a non-pure assignment and movers across firms and regions. Although the results from Section 2.2 are challenging to establish, this model preserves the key properties essential for the empirical analysis: the wage equation remains unchanged, validating my productivity measures, and firms' value continues to reflect complementarity between $x(\ell)$ and $y(\ell)$, which I test below.

Causal evidence of the two-sided sorting mechanism. I test the defining prediction of two-sided sorting: worker sorting leads to firm sorting. This prediction, formalized in (5), reveals the complementarity between worker and firm types in their spatial sorting; local labor markets with higher worker productivity $x(\ell)$ attract

[34] It is well acknowledged that identifying firm productivity separately from other factors is difficult (e.g., Combes et al., 2008). A few studies discuss strategies to identify firm productivity (e.g., Gaubert, 2018; Bilal, 2023; Lindenlaub et al., 2025). However, each strategy depends on a set of assumptions that is specific to its own context.

firms with higher productivity $y(\ell)$, conditional on the vacancy contact rate $q(\ell)$. It is distinctive in that firm sorting endogenously responds to economic conditions, and in particular to worker sorting, rather than other location-specific factors.

Because cross-sectional variation cannot identify two-sided sorting, I exploit time differences and examine whether *changes* in worker sorting induce *changes* in firm sorting. Specifically, I estimate the following time-differenced regression:

$$\Delta \log y(\ell) = \gamma_0 + \gamma_1 \Delta \log x(\ell) + \mathbf{X}(\ell)'\gamma_2 + u(\ell), \tag{14}$$

where Δ denotes changes between period $t = 1$ (2003-2009) and $t = 2$ (2010-2016). The control vector $\mathbf{X}(\ell)$ includes changes in the job arrival rate $\Delta \log \lambda(\ell)$ to control for changes in the vacancy contact rate $\Delta \log q(\ell)$, since $\lambda(\ell)$ is observable and inversely related to $q(\ell)$. The coefficient γ_1 is the primary parameter of interest, and I test whether it is positive.

A key concern is time-varying location-specific factors, such as local TFP or amenities, that may affect both worker and firm sorting. For example, if worker and firm productivity are both complementary to local TFP, changes in local TFP could induce both worker and firm sorting, rather than one causing the other, thereby biasing the coefficient upward. Moreover, potential problems in identifying firm productivity $y(\ell)$ exacerbate this concern. If local TFP $A(\ell)$ varies across ℓ as in the extended model in Section 3.3, the wage equation (12) implies that the estimated firm productivity reflects the product of $y(\ell)$ and $A(\ell)$.

To address this concern, I instrument $\Delta \log x(\ell)$ by exploiting changes in the quality of domestic migrants arriving in ℓ, driven by changes in their origins ℓ' rather than those in ℓ. Specifically, I construct the *predicted* changes in migrant productivity using historical migration patterns between ℓ and ℓ' and contemporaneous outflows from ℓ'. First, I compute the predicted share of migrants originating from ℓ' conditional on arriving in ℓ, $\hat{s}_t(\ell'|\ell)$, by interacting the historical migration probability of choosing ℓ from ℓ' in $t = 0$ (1991−2002), $m_0(\ell' \to \ell|\ell')$, with the number of out-migrants from ℓ' to destinations other than ℓ in $t = 1, 2$, $O_{\ell',-\ell,t}$.[35] Second, I calculate the average productivity of migrants from ℓ' to all destinations other than ℓ,

[35] This approach is closely related to studies on migrants across and within countries (e.g., Burchardi et al., 2019; Altonji and Card, 1991; Howard, 2020; Boustan et al., 2010). They primarily examine the effect of changes in the *number* of migrants, while I focus on changes in productivity.

$\hat{x}_t(\ell', -\ell)$. Combining these yields the predicted productivity of migrants as follows,

$$\log x^{\mathrm{m,\,IV}}(\ell) = \sum_{\ell' \neq \ell} \hat{s}_t(\ell'|\ell) \log \hat{x}_t(\ell', -\ell) \quad \text{where } \hat{s}_t(\ell'|\ell) = \frac{m_0(\ell' \to \ell|\ell') O_{\ell', -\ell, t}}{\sum_{k \neq \ell} m_0(k \to \ell|k) O_{k, -\ell, t}}. \tag{15}$$

Finally, I construct the instrument $\Delta \log x^{\mathrm{IV}}(\ell)$ by differencing the above between the two periods.

The productivity of domestic migrants is relevant to the productivity of all workers in the local search pool, directly and indirectly. Although the migration rate is modest, migrants are disproportionately represented among unemployed workers, and they account for about 30% of local search pools—a sizable share. In addition, an increase in migrant productivity indirectly affects that of non-migrants in the search pool, for example, by increasing housing rents, which may in turn push out less productive workers.

The identification assumption is that shocks to origins ℓ' are uncorrelated with shocks to destinations ℓ. This assumption could be violated if a single underlying factor drives both historical migration patterns and the correlation of current shocks across locations. For example, similarity in industry composition between two locations may have generated substantial past migration, and if it moreover exposes them to similar industry-specific shocks today, the predicted migrant productivity will be correlated with local shocks. Geographic proximity may generate a similar problem if workers historically migrated to nearby regions and location-specific shocks are spatially correlated. To address this concern, I control for industry composition and exclude variation arising from changes in nearby regions as a robustness check.

Table 3 shows the estimation results. In all specifications, I include firm productivity and unemployment rates in $t = 1$ as controls. The former controls for the potential time trend of firm productivity due to regional convergence, while the latter captures the potential impact of the German Hartz reform, a major unemployment insurance reform in the mid-2000s. This reform led to a substantial decline in unemployment rates, particularly in regions with higher initial unemployment.

Column (1) documents the coefficient from an ordinary least squares estimation. An increase in worker productivity is associated with an increase in firm productivity in the same region. However, this result is subject to endogeneity concerns.

In Column (2), I report the results of an instrumental variable approach. In the bottom panel, I report first-stage regression results. I regress changes in worker productivity $\Delta \log x(\ell)$ on the instrument $\Delta \log x^{\mathrm{IV}}(\ell)$ and the same controls as in the second stage, which yields a statistically significant positive coefficient. The IV regression coefficient in the top panel is statistically significant and economically meaningful: an exogenous

Table 3: The Response of Firm Sorting to Changes in Worker Sorting

	OLS (1)	IV (2)	IV (3)	IV (4)	IV (5)
$\Delta \log x(\ell)$	0.686	1.127	1.075	1.080	1.045
	(0.133)	(0.384)	(0.307)	(0.724)	(0.333)
$\Delta \log \hat{y}^{\text{old}}(\ell)$					0.625
					(0.182)
Industry controls			✓		
Geography controls				✓	
	2SLS First-stage Estimates				
$\Delta \log x^{\text{IV}}(\ell)$		0.610	0.621	0.680	0.608
		(0.184)	(0.181)	(0.342)	(0.172)

Notes: $N = 257$. Robust standard errors are shown in parentheses. Columns (2)-(5) in the top panel report second-stage estimates, with the dependent variable equal to the change in firm productivity, between 2003-2009 and 2010-2016. All regressions include firm productivity and unemployment rates in the first period. First-stage regressions in the bottom panel include the same controls as the corresponding second-stage specification in the top panel. Each observation is weighted by the number of workers. In Column (3), firm productivity is residualized with respect to one-digit industry fixed effects. In Column (4), flows to or from locations within 100 km are excluded when constructing $\Delta \log x(\ell)$ and its instrument. In Column (5), the specification additionally controls for the average change in fixed effects of jobs matched in both periods.

10% increase in worker productivity in the local search pool leads to an 11.3% increase in the productivity of new jobs in the same location.[36] To test robustness, in Column (3), I first residualize firm productivity with respect to one-digit industry fixed effects before constructing the dependent variable. In Column (4), I drop all flows to or from locations within 100 km. The coefficients of interest do not change much.

Finally, I address concerns related to agglomeration forces that the instrument may not eliminate. The literature suggests that an increase in worker productivity may endogenously raise the productivity of local jobs (e.g., Diamond, 2016; Rossi-Hansberg et al., 2019). However, a distinction in my setting is that I focus on workers in the search pool rather than the entire employed workforce, which reduces the likelihood that agglomeration forces are driving the results. More importantly, unlike two-sided sorting, agglomeration forces typically benefit not only new jobs but also *existing* jobs. Thus, the gap between changes in new and preexisting jobs reflects the impact of firm sorting. Motivated by this observation, I control for $\Delta \log \hat{y}^{\text{old}}$,

[36] This coefficient estimate exceeds that obtained under OLS, which suggests that the attenuation bias due to measurement errors may be substantial. Measurement issue is particularly important in my context. Because researchers observe the sum of changes in worker and fixed effects—namely, changes in wages—any mismeasurement of changes in worker fixed effects mechanically translates into mismeasurement of firm fixed effects in the opposite direction, thereby inducing a downward bias.

defined as the average change in firm fixed effects of existing jobs matched in both periods.[37] This control also helps address additional concerns about other location-specific factors influenced by the local workforce, such as shifts in the demand for local products, which may also bias the coefficient. The coefficient in Column (5) is 1.045, essentially identical to earlier estimates.

Overall, these findings imply that changes in worker sorting generate substantial changes in firm sorting; the two-sided sorting mechanism is sufficiently large to be economically meaningful. Interpreting this coefficient in a cross-sectional context—by normalizing the coefficient by the ratio of the cross-sectional standard deviations of the log of worker and firm productivity—implies that a one standard deviation increase in log worker productivity induces a 0.51 standard deviation increase in log firm productivity in the same location.

In Section C.2, I provide further support for these findings with two additional results. First, I rely on changes in worker and firm fixed effects, rather than changes in model-implied productivity (Table A.3). The overall results remain largely unchanged: a one standard deviation increase in log worker productivity induces a 0.6 standard deviation increase in log firm productivity. Second, I implement a placebo exercise in which the dependent variable is changes in firm fixed effects of existing jobs, $\Delta \log \hat{y}^{\text{old}}(\ell)$ (Table A.4). In contrast to Table 3 and Table A.3, coefficients are statistically insignificant and close to zero, which indicates that most changes in the productivity of new jobs arise from changes in firm sorting rather than those in location-specific factors.

Finally, based on these results, I provide suggestive evidence on the quantitative importance of two-sided sorting in explaining spatial disparities. The coefficients in Table 3 are approximately 60% of the value implied by the analogous exercise using the two-sided sorting model, which is estimated using German cross-sectional data; see Section C.3 for details. While a formal decomposition of all mechanisms is beyond the scope of this paper, the analysis indicates that two-sided sorting is a major contributor to spatial disparities and should be given careful consideration in policy evaluation.

[37] I use firm fixed effects rather than model-implied firm productivity, because the productivity of existing jobs, unlike that of new jobs, is sensitive to the calibration of β.

6. Conclusion

In this paper, I show that when heterogeneous workers and firms match through their location choices, productive workers and firms self-select into cities, thereby giving rise to PAM. The interaction between worker and firm sorting—absent exogenous heterogeneity across locations—endogenously generates spatial disparities in productivity, income, and population density. Furthermore, I highlight that two-sided sorting presents distinctive policy implications, as both workers and firms—who embody productivity—can relocate in response to government policy.

The contribution of this paper is to propose a parsimonious framework that isolates the role of two-sided sorting in shaping spatial disparities and emphasizes its relevance for policy analysis. A promising direction for future research is to construct a quantitative framework for policy assessment that incorporates all relevant mechanisms behind spatial disparities and to perform a careful decomposition. This framework should also incorporate and quantify other important determinants of location decisions, such as migration frictions and amenities.

References

Abowd, John M, Francis Kramarz, and David N Margolis (1999) "High Wage Workers and High Wage Firms," *Econometrica*, 67 (2), 251–333.

Acemoglu, Daron (2001) "Good Jobs Versus Bad Jobs," *Journal of Labor Economics*, 19 (1), 1–21.

Ahlfeldt, Gabriel M, Stephan Heblich, and Tobias Seidel (2023) "Micro-geographic Property Price and Rent Indices," *Regional Science and Urban Economics*, 98.

Albouy, David (2009) "The Unequal Geographic Burden of Federal Taxation," *Journal of Political Economy*, 117 (4), 635–667.

Albrecht, James, Lucas Navarro, and Susan Vroman (2010) "Efficiency in a Search and Matching Model with Endogenous Participation," *Economics Letters*, 106 (1), 48–50.

Allen, Treb and Costas Arkolakis (2014) "Trade and the Topography of the Spatial Economy," *Quarterly Journal of Economics*, 129 (3), 1085–1140.

Altonji, Joseph G. and David Card (1991) *The Effects of Immigration on the Labor Market Outcomes of Less-skilled Natives*, 201–234: University of Chicago Press.

Atakan, Alp E. (2016) "Assortative Matching with Explicit Search Costs," *Econometrica*, 74 (3), 667–680.

Baum-Snow, Nathaniel and Ronni Pavan (2012) "Understanding the City Size Wage Gap," *Review of Economic Studies*, 79 (1), 88–127.

Becker, Gary S. (1973) "A Theory of Marriage: Part I," *Journal of Political Economy*, 81 (4), 813–846.

Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud (2014) "Productive Cities: Sorting, Selection, and Agglomeration," *Journal of Political Economy*, 122 (3), 507–553.

Bilal, Adrien (2023) "The Geography of Unemployment," *Quarterly Journal of Economics*, 138 (3), 1507–1576.

Borovicková, Katarına and Robert Shimer (2024) "Assortative Matching and Wages: The Role of Selection," *NBER Working Paper*.

Boustan, Leah Platt, Price V Fishback, and Shawn Kantor (2010) "The Effect of Internal Migration on Local Labor Markets: American Cities During the Great Depression," *Journal of Labor Economics*, 28 (4), 719–746.

Burchardi, Konrad B, Thomas Chaney, and Tarek A Hassan (2019) "Migrants, Ancestors, and Foreign Investments," *Review of Economic Studies*, 86 (4), 1448–1486.

Burdett, Ken and Melvyn G Coles (1997) "Marriage and Class," *Quarterly Journal of Economics*, 112 (1), 141–168.

Card, David, Jörg Heining, and Patrick Kline (2013) "Workplace Heterogeneity and the Rise of West German Wage Inequality," *Quarterly Journal of Economics*, 128 (3), 967–1015.

Card, David, Jesse Rothstein, and Moises Yi (2023) "Location, Location, Location," *NBER Working Paper*.

Chade, Hector, Jan Eeckhout, and Lones Smith (2017) "Sorting through Search and Matching Models in Economics," *Journal of Economic Literature*, 55, 493–544.

Combes, Pierre Philippe, Gilles Duranton, and Laurent Gobillon (2008) "Spatial Wage Disparities: Sorting Matters!," *Journal of Urban Economics*, 63 (2), 723–742.

Crews, Levi Garrett (2023) "A Dynamic Spatial Knowledge Economy," *Working Paper*.

Davis, Donald R. and Jonathan I. Dingel (2019) "A Spatial Knowledge Economy," *American Economic Review*, 109 (1), 153–170.

Demange, Gabrielle and David Gale (1985) "The Strategy Structure of Two-Sided Matching Markets," *Econometrica*, 53 (4), 873–888.

Diamond, Rebecca (2016) "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980–2000," *American Economic Review*, 106 (3), 479–524.

Diamond, Rebecca and Cecile Gaubert (2022) "Spatial Sorting and Inequality," *Annual Review of Economics*, 14, 795–819.

Eeckhout, Jan and Philipp Kircher (2010) "Sorting and Decentralized Price Competition," *Econometrica*, 78 (2), 539–574.

Fajgelbaum, Pablo D and Cecile Gaubert (2020) "Optimal Spatial Policies, Geography, and Sorting," *Quarterly Journal of Economics*, 135 (2), 959–1036.

Fajgelbaum, Pablo D, Eduardo Morales, Juan Carlos Suárez Serrato, and Owen Zidar (2019) "State Taxes and Spatial Misallocation," *Review of Economic Studies*, 86 (1), 333–376.

Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, and Michael Westberry (2022) "Integrated Public Use Microdata Series, Current Population Survey: Version 10.0 [dataset]," https://doi.org/10.18128/D030.V10.0.

Gaubert, Cecile (2018) "Firm Sorting and Agglomeration," *American Economic Review*, 108 (11), 3117–3153.

Green, Richard K, Stephen Malpezzi, and Stephen K Mayo (2005) "Metropolitan-Specific Estimates of the Price Elasticity of Supply of Housing, and Their Sources," *American Economic Review*, 95 (2), 334–339.

Gyourko, Joseph, Albert Saiz, and Anita Summers (2008) "A New Measure of the Local Regulatory Environment for Housing Markets: The Wharton Residential Land Use Regulatory Index," *Urban Studies*, 45 (3), 693–729.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L Violante (2017) "Optimal Tax Progressivity: An Analytical Framework," *Quarterly Journal of Economics*, 132 (4), 1693–1754.

Hoppe, Heidrun C, Benny Moldovanu, and Aner Sela (2009) "The Theory of Assortative Matching Based on Costly Signals," *Review of Economic Studies*, 76 (1), 253–281.

Hosios, Arthur J (1990) “On the Efficiency of Matching and Related Models of Search and Unemployment,” *Review of Economic Studies*, 57 (2), 279–298.

Howard, Greg (2020) “The Migration Accelerator: Labor Mobility, Housing, and Demand,” *American Economic Journal: Macroeconomics*, 12 (4), 147–179.

Hsieh, Chang Tai and Enrico Moretti (2019) “Housing Constraints and Spatial Misallocation,” *American Economic Journal: Macroeconomics*, 11 (2), 1–39.

Kline, Patrick and Enrico Moretti (2013) “Place Based Policies with Unemployment,” *American Economic Review: Papers & Proceedings*, 103 (3), 238–243.

——— (2014) “Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority,” *Quarterly Journal of Economics*, 129 (1), 275–331.

Kuhn, Moritz, Iourii Manovskii, and Xincheng Qiu (2022) “The Geography of Job Creation and Job Destruction,” *SSRN Electronic Journal*.

Landais, Camille, Pascal Michaillat, and Emmanuel Saez (2018) “A Macroeconomic Approach to Optimal Unemployment Insurance: Theory,” *American Economic Journal: Economic Policy*, 10 (2), 152–181.

Lebuhn, Henrik, Andrej Holm, Stephan Junker, Stephan Junker, and Kevin Neitzel (2017) “Wohnverhältnisse in Deutschland,” *Hans Böckler Stiftung*.

Lindenlaub, Ilse, Ryungha Oh, and Michael Peters (2025) “Spatial Firm Sorting and Local Monopsony Power.”

Mangin, Sephorah and Benoit Julien (2021) “Efficiency in Search and Matching Models: A Generalized Hosios Condition,” *Journal of Economic Theory*, 193, 105208.

Martellini, Paolo (2022) “Local Labor Markets and Aggregate Productivity,” *Working Paper*.

Menzio, Guido (2007) “A Theory of Partially Directed Search,” *Journal of Political Economy*, 115 (5), 748–769.

Moen, Espen R (1997) “Competitive Search Equilibrium,” *Journal of Political Economy*, 105 (2), 385–411.

Nelson, Kenneth, Daniel Fredriksson, Tomas Korpi, Walter Korpi, Joakim Palme, and Ola Sjöberg (2020) “The Social Policy Indicators (SPIN) Database,” *International Journal of Social Welfare*, 29 (3), 285–289.

Osborne, Martin J and Ariel Rubinstein (1994) *A Course in Game Theory*: MIT press.

Pissarides, Christopher A (2000) *Equilibrium Unemployment Theory*: MIT press.

Postel-Vinay, Fabien and Ilse Lindenlaub (2023) “Multi-Dimensional Sorting under Random Search,” *Journal of Political Economy*, 131 (12).

Redding, Stephen J (2016) “Goods Trade, Factor Mobility and Welfare,” *Journal of International Economics*, 101, 148–167.

Roback, Jennifer (1982) “Wages, Rents, and the Quality of Life,” *Journal of Political Economy*, 90 (6), 1257–1278.

De la Roca, Jorge, Gianmarco IP Ottaviano, and Diego Puga (2023) "City of Dreams," *Journal of the European Economic Association*, 21 (2), 690–726.

Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Felipe Schwartzman (2019) "Cognitive Hubs and Spatial Redistribution," 19 (16), 1–83.

Roth, Alvin E and Marilda Sotomayor (1990) *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*: Cambridge University Press.

Ruf, Kevin, Lisa Schmidtlein, Stefan Seth, Heiko Stüber, and Matthias Umkehrer (2021a) "Linked Employer-Employee Data from the IAB: LIAB Longitudinal Model (LIAB LM) 1975 – 2019.," FDZ-Datenreport, 06/2021 (en), Nuremberg. DOI: 10.5164/IAB.FDZD.2106.en.v1.

Ruf, Kevin, Lisa Schmidtlein, Stefan Seth, Heiko Stüber, Matthias Umkehrer, Stephan Gries*s* emer, and Steffen Kaimer (2021b) "Linked-Employer-Employee-Data of the IAB (LIAB): LIAB longitudinal model 1975-2019, version 1," Research Data Centre of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB). DOI: 10.5164/IAB.LIABLM7519.de.en.v1.

Ruggles, Steven, Sarah Flood, Matthew Sobek, Danika Brockman, Grace Cooper, Stephanie Richards, and Megan Schouweiler (2023) "IPUMS USA: version 13.0 [dataset]," https://doi.org/10.18128/D010.V13.0.

Saiz, Albert (2010) "The Geographic Determinants of Housing Supply," *Quarterly Journal of Economics*, 125 (3), 1253–1296.

Schwartz, Aba (1973) "Interpreting the Effect of Distance on Migration," *Journal of Political Economy*, 81 (5), 1153–1169.

Shi, Shouyong (2001) "Frictional Assignment. I. Efficiency," *Journal of Economic Theory*, 98 (2), 232–260.

Shimer, Robert (2012) "Reassessing the Ins and Outs of Unemployment," *Review of Economic Dynamics*, 15 (2), 127–148.

Shimer, Robert and Lones Smith (2000) "Assortative Matching and Search," *Econometrica*, 68 (2), 343–369.

——— (2001) "Matching, Search, and Heterogenity," *Advances in Macroeconomics*, 1 (1).

Smith, Lones (2006) "The marriage model with search frictions," *Journal of Political Economy*, 114 (6), 1124–1144.

Spence, Michael (1973) "Job Market Signaling," *Quarterly Journal of Economics*, 87 (3), 355–374.

Topkis, Donald M (1978) "Minimizing a submodular function on a lattice," *Operations research*, 26 (2), 305–321.

Appendix

A. Omitted Proofs

A.1 Derivations

The match surplus of workers and vacancies solves

$$\tilde{\rho}(V^e(x,y,\ell) - V^u(x,\ell)) = w(x,y,\ell) - bx - \lambda(\ell)\max\{V^e(x,y(\ell),\ell) - V^u(x,\ell),0\}, \tag{A.1}$$
$$\tilde{\rho}(V^p(x,y,\ell) - V^v(y,\ell)) = xy - w(x,y,\ell) - q(\ell)\max\{V^p(x(\ell),y,\ell) - V^v(y,\ell),0\} + \delta_v V^v(y,\ell),$$

where $\tilde{\rho} \equiv \rho + \delta$. Consider the case that $V^e(x,y(\ell),\ell) \geq V^u(x,\ell)$ and $V^p(x(\ell),y,\ell) \geq V^v(y,\ell)$. Combining the two gives the HJB equation for the joint surplus, $S(x,y,\ell) \equiv V^e(x,y,\ell) - V^u(x,\ell) + V^p(x,y,\ell) - V^v(y,\ell)$,

$$\begin{aligned}\tilde{\rho}S(x,y,\ell) &= xy - \lambda(\ell)[V^e(x,y(\ell),\ell) - V^u(x,\ell)] - q(\ell)[V^p(x(\ell),y,\ell) - V^v(y,\ell)] + \delta_v V^v(y,\ell)\\ &= xy - bx - \beta\lambda(\ell)S(x,y(\ell),\ell) - (1-\tilde{\beta})q(\ell)S(x(\ell),y,\ell),\end{aligned} \tag{A.2}$$

where I define $1-\tilde{\beta} \equiv \frac{\rho}{\rho+\delta_v}(1-\beta)$ to simplify the notation. To obtain the second line, I use the bargaining solution and $\delta_v V^v(y,\ell) = \frac{\delta_v}{\rho+\delta_v}q(\ell)(V^p(x(\ell),y,\ell) - V^v(y,\ell))$ derived from (2). Distinguishing the productivity of workers and firms of a given match (x,y) from local worker and firm productivity $(x(\ell),y(\ell))$ is important for characterizing equilibrium conditions. When $x = x(\ell)$ and $y = y(\ell)$, the surplus simplifies to

$$S(x(\ell),y(\ell),\ell) = \frac{1}{\tilde{\rho} + \beta\lambda(\ell) + (1-\tilde{\beta})q(\ell)}(x(\ell)y(\ell) - bx(\ell)). \tag{A.3}$$

I define four terms, A_w, A_f, B_w, and B_f, which will be referred to frequently in the proofs:

$$A_w(y(\ell),\lambda(\ell)) = \frac{\beta\lambda(\ell)}{\tilde{\rho}+\beta\lambda(\ell)}(y(\ell)-b), \quad B_w(x(\ell),\theta(\ell)) = \frac{(1-\tilde{\beta})q(\ell)}{\tilde{\rho}+\beta\lambda(\ell)+(1-\tilde{\beta})q(\ell)}x(\ell),$$
$$A_f(x(\ell),q(\ell)) = \frac{(1-\tilde{\beta})q(\ell)}{\tilde{\rho}+(1-\tilde{\beta})q(\ell)}x(\ell), \quad B_f(y(\ell),\theta(\ell)) = \frac{\beta\lambda(\ell)}{\tilde{\rho}+\beta\lambda(\ell)+(1-\tilde{\beta})q(\ell)}(y(\ell)-b).$$

Plugging (A.3) into (A.2), I obtain the following expressions,

$$\beta\lambda(\ell)S(x,y(\ell),\ell) = A_w(y(\ell),\lambda(\ell))(x - B_w(x(\ell),\theta(\ell))),$$
$$(1-\tilde{\beta})q(\ell)S(x(\ell),y,\ell) = A_f(x(\ell),q(\ell))(y - b - B_f(y(\ell),\theta(\ell))).$$

Plugging these expressions into (A.1), I can solve for wages,

$$\begin{aligned} w(x,y,\ell) &= bx + \beta(y-b)x + (1-\beta)\beta\lambda(\ell)S(x,y(\ell),\ell) - \beta(1-\tilde{\beta})q(\ell)S(x(\ell),y,\ell) \\ &= (1-\beta)bx + \beta xy + (1-\beta)A_w(\cdot)(x - B_w(\cdot)) - \beta A_f(\cdot)(y - b - B_f(\cdot)). \end{aligned}$$

In equilibrium, with $x = x(\ell)$ and $y = y(\ell)$, the above simplifies to

$$w(x(\ell),y(\ell),\ell) = \left(b + \beta(y(\ell)-b) + \beta\frac{(1-\beta)\lambda(\ell) - (1-\tilde{\beta})q(\ell)}{\tilde{\rho} + \beta\lambda(\ell) + (1-\tilde{\beta})q(\ell)}(y(\ell)-b)\right)x(\ell). \tag{A.4}$$

Finally, I can solve for the value of workers and firms when choosing ℓ. The value of workers choosing location ℓ simply equals the value of unemployed workers, $V^u(x,\ell)$.

$$\begin{aligned} \rho V^u(x,\ell) &= bx + \beta\lambda(\ell)\max\{S(x,y(\ell),\ell),0\} + \Pi - \bar{h}r(\ell) \\ &= bx + \max\{A_w(y(\ell),\lambda(\ell))(x - B_w(x(\ell),\theta(\ell))),0\} + \Pi - \bar{h}r(\ell), \end{aligned}$$

where I allow the possibility that the surplus is negative, or equivalently $x < B_w(\cdot)$, which happens when x is significantly smaller than $x(\ell)$. Similarly, the value of vacancy is given by

$$\rho V^v(y,\ell) = \frac{\rho}{\rho+\delta_v}(1-\beta)q(\ell)\max\{S(x(\ell),y,\ell),0\} = \max\{A_f(x(\ell),q(\ell))(y - b - B_f(y(\ell),\theta(\ell))),0\}.$$

When firms operate their business in ℓ, at each point in time, they post δ_v units of vacancies and pay overhead costs. The former yields a value of $\delta_v V^v(y,\ell)$ while the latter costs $c(\ell)$. The value of firms choosing ℓ equals the discounted sum of the two:

$$\begin{aligned} \rho\bar{V}^v(y,\ell) &= \rho\int_0^\infty e^{-\rho t}(\delta_v V^v(y,\ell) - c(\ell))\,\mathrm{d}t = \delta_v V^v(y,\ell) - c(\ell) \\ &= \frac{\delta_v}{\rho}\max\{A_f(x(\ell),q(\ell))(y - b - B_f(y(\ell),\theta(\ell))),0\} - c(\ell). \end{aligned}$$

In addition, for reference, I define the elasticity of the job arrival rate and the vacancy contact rate:

$$\varepsilon_\lambda(\theta) \equiv \frac{\lambda'(\theta)}{\lambda(\theta)}\theta, \quad \varepsilon_q(\theta) \equiv \frac{q'(\theta)}{q(\theta)}\theta,$$

which satisfy $\varepsilon_\lambda(\theta) - \varepsilon_q(\theta) = 1$, $0 < \varepsilon_\lambda(\theta) < 1$, and $-1 < \varepsilon_q(\theta) < 0$.

A.2 Proof of Proposition 1

I first show that a pure assignment equilibrium, if it exists, exhibits PAM, and then prove existence.

PAM between workers and firms. I prove that any pure assignment equilibrium must feature positive assortative matching by contradiction. Suppose there exist two locations $\ell' < \ell''$ and two types of workers and firms such that $x(\ell'') > x(\ell')$ but $y(\ell'') < y(\ell')$.

The location choices of workers imply that $A_w(\ell'') > A_w(\ell')$.[38] Given that firm types are worse in ℓ'', for $A_w(\cdot)$ to be larger in ℓ'', the job arrival rate must be sufficiently higher, $\lambda(\ell'') > \lambda(\ell')$, which implies that $q(\ell') > q(\ell'')$. In addition, the location choices of firms imply $A_f(\ell') > A_f(\ell'')$. Combining these two, I conclude that $B_w(\cdot) = \frac{\tilde{\rho}+(1-\tilde{\beta})q(\cdot)}{\tilde{\rho}+\beta\lambda(\cdot)+(1-\tilde{\beta})q(\cdot)}A_f(\cdot)$ is larger in ℓ'. Denoting $\mathcal{U}(x,\ell) \equiv A_w(\ell)(x - B_w(\ell))$, the discussion so far implies that $\mathcal{U}(x,\ell)$ is supermodular and log-submodular in (x,ℓ), which yields

$$\mathcal{U}(x(\ell''),\ell') + \mathcal{U}(x(\ell'),\ell'') < \mathcal{U}(x(\ell'),\ell') + \mathcal{U}(x(\ell''),\ell'') < \mathcal{U}(x(\ell'),\ell') + \frac{\mathcal{U}(x(\ell''),\ell')}{\mathcal{U}(x(\ell'),\ell')}\mathcal{U}(x(\ell'),\ell''),$$
$$\Rightarrow \mathcal{U}(x(\ell'),\ell')\left(\frac{\mathcal{U}(x(\ell''),\ell')}{\mathcal{U}(x(\ell'),\ell')} - 1\right) < \mathcal{U}(x(\ell'),\ell'')\left(\frac{\mathcal{U}(x(\ell''),\ell')}{\mathcal{U}(x(\ell'),\ell')} - 1\right).$$

Since $\mathcal{U}(x(\ell'),\ell') < \mathcal{U}(x(\ell''),\ell')$, the above inequality implies that $\mathcal{U}(x(\ell'),\ell') < \mathcal{U}(x(\ell'),\ell'')$. The sorting condition of workers, $V^u(x(\ell'),\ell') \geq V^u(x(\ell'),\ell'')$, implies $r(\ell'') > r(\ell')$.

Similarly, $B_f(\cdot) = \frac{\tilde{\rho}+\beta\lambda(\cdot)}{\tilde{\rho}+\beta\lambda(\cdot)+(1-\tilde{\beta})q(\cdot)}A_w(\cdot)$ is larger in ℓ'' due to $A_w(\ell'') > A_w(\ell')$ and $\lambda(\ell'') > \lambda(\ell')$, which implies that $A_f(\ell)(y-b-B_f(\ell))$ is submodular and log-supermodular in (ℓ, y), and increasing in y. Following the same logic, $A_f(\cdot)(y(\ell'') - b - B_f(\cdot))$ is smaller in ℓ''. Therefore, the firm sorting condition, $\bar{V}^p(y(\ell''),\ell'') \geq \bar{V}^p(y(\ell''),\ell')$, gives $c(\ell') > c(\ell'')$.

As production costs of housing and business services are convex, two inequalities involving $r(\cdot)$ and $c(\cdot)$ lead to $L(\ell'') > L(\ell')$ and $N(\ell'') < N(\ell')$, which together imply $\frac{N(\ell')}{L(\ell')} > \frac{N(\ell'')}{L(\ell'')}$. Note that in steady state, a higher firm density to worker density ratio, $\frac{N(\ell)}{L(\ell)}$, leads to a higher $\theta(\ell)$. To see this, plug (7) into the definition of the market tightness to obtain $\theta(\ell)\frac{\delta}{\delta+\lambda(\theta(\ell))} = \frac{N(\ell)}{L(\ell)}$, and then check that the left-hand side increases in $\theta(\ell)$ as $1 - \frac{\lambda(\ell)}{\delta+\lambda(\ell)}\varepsilon_\lambda(\theta(\ell)) \geq 0$. However, this contradicts $\lambda(\ell'') > \lambda(\ell')$. Thus, I conclude that workers and firms positively sort across space.

Existence. I show the existence of a pure assignment equilibrium in steps. In the first step, I write the sorting conditions of workers and firms as an ordinary differential equation problem (ODE). In the second step, I show that there exists a solution to ODE that satisfies all remaining equilibrium conditions.

Step 1-(i) From the discussion above, the equilibrium is characterized by two strictly increasing functions, $x(\ell)$ and $y(\ell)$. I will find an equilibrium assignment that is twice continuously differentiable. Define $z(\ell) = (x(\ell), y(\ell))$,

[38] To simplify the notation, I index the terms—$A_w(\cdot)$, $B_w(\cdot)$, $A_f(\cdot)$, $B_f(\cdot)$—by ℓ in the proofs, unless this creates ambiguity.

$z'(\ell) = (x'(\ell), y'(\ell))$, and $z''(\ell) = (x''(\ell), y''(\ell))$. Then, the following first-order conditions should hold:

$$\begin{aligned} f_w(z(\cdot), z'(\cdot), z''(\cdot)) \equiv & A'(\ell)x(\ell) - (A_w(\ell)B_w(\ell))' - \bar{h}^2 C_r''(\bar{h}L(\ell))L'(\ell) = 0, \\ f_f(z(\cdot), z'(\cdot), z''(\cdot)) \equiv & A_f'(\ell)(y(\ell) - b) - (A_f(\ell)B_f(\ell))' - \frac{\rho}{\delta_v} C_v''(N(\ell))N'(\ell) = 0, \end{aligned}$$

in addition to steady-state flow-balance conditions (7). Housing and business services market clearing conditions (6) are imposed above.

I stack the two conditions, $f(\cdot) \equiv (f_w(\cdot), f_f(\cdot))' = 0 \in \mathbb{R}^2$. If $D_{z''}f$ is continuous and $\det D_{z''}f(z, z', z'') \neq 0$, then there exists a unique continuously differentiable function $g(\cdot)$ such that $z''(\ell) = g(z(\ell), z'(\ell))$ by Implicit Function Theorem. From now on, I omit location index ℓ unless it creates any confusion.

To check the condition on $D_{z''}f(z, z', z'')$, I first obtain several useful expressions,

$$A_w'(\ell) = A_w(\ell)\left(\varepsilon_\lambda(\theta)\frac{\theta'}{\theta}\frac{\tilde{\rho}}{\tilde{\rho} + \beta\lambda} + \frac{y'}{y - b}\right), \tag{A.5}$$

$$B_w'(\ell) = B_w(\ell)\left(-\frac{-\tilde{\rho}\frac{\theta'}{\theta}\varepsilon_q(\theta) + \beta\lambda/\theta}{\tilde{\rho} + \beta\lambda(\theta) + (1 - \tilde{\beta})q} + \frac{x'}{x}\right), \tag{A.6}$$

$$A_f'(\ell) = A_f(\ell)\left(\varepsilon_q(\theta)\frac{\theta'}{\theta}\frac{\tilde{\rho}}{\tilde{\rho} + (1 - \tilde{\beta})q} + \frac{x'}{x}\right),$$

$$B_f'(\ell) = B_f(\ell)\left(\frac{\tilde{\rho}\varepsilon_\lambda(\theta)\frac{\theta'}{\theta} + (1 - \tilde{\beta})q/\theta}{\tilde{\rho} + \beta\lambda + (1 - \tilde{\beta})q} + \frac{y'}{y - b}\right),$$

$$\frac{\theta'(\ell)}{\theta(\ell)} = \zeta(z, z')\left(\frac{N'}{N} - \frac{L'}{L}\right) \quad \text{where} \;\; \zeta(z, z') = \frac{\lambda + \delta}{-\varepsilon_q(\theta)\lambda + \delta} > 0. \tag{A.7}$$

Plugging the above expressions into $f(\cdot)$, I obtain the following:

$$\begin{aligned} f_w(\cdot) = & \frac{A_w x}{\tilde{\rho} + \beta\lambda + (1 - \tilde{\beta})q}\Bigg((\tilde{\rho} + \beta\lambda)\frac{y'}{y - b} - (1 - \tilde{\beta})q\frac{x'}{x} \\ & + \tilde{\rho}\frac{\theta'}{\theta}\left(\varepsilon_\lambda(\theta) + \frac{(1 - \tilde{\beta})q}{\tilde{\rho} + \beta\lambda + (1 - \tilde{\beta})q}(-\varepsilon_q(\theta))\right) - \frac{(1 - \tilde{\beta})q}{\tilde{\rho} + \beta\lambda + (1 - \tilde{\beta})q}\beta\frac{\lambda}{\theta}\Bigg) - \bar{h}^2 C_r'' L', \\ f_f(\cdot) = & \frac{A_f(y - b)}{\tilde{\rho} + \beta\lambda + (1 - \tilde{\beta})q}\Bigg((\tilde{\rho} + (1 - \tilde{\beta})q)\frac{x'}{x} - \beta\lambda\frac{y'}{y - b} \\ & - \tilde{\rho}\frac{\theta'}{\theta}\left(-\varepsilon_q(\theta) + \frac{\beta\lambda}{\tilde{\rho} + \beta\lambda + (1 - \tilde{\beta})q}\varepsilon_\lambda(\theta)\right) - \frac{\beta\lambda}{\tilde{\rho} + \beta\lambda + (1 - \tilde{\beta})q}(1 - \tilde{\beta})\frac{q}{\theta}\Bigg) - \frac{\rho}{\delta_v}C_v'' N'. \end{aligned}$$

Differentiating the above with respect to z'', I obtain $D_{z''}f$:

$$D_{z''}f = \begin{pmatrix} \frac{\partial f_w}{\partial x''}, \frac{\partial f_w}{\partial y''} \\ \frac{\partial f_f}{\partial x''}, \frac{\partial f_f}{\partial y''} \end{pmatrix} = \begin{pmatrix} -a_w b_w - \bar{h}^2 M_w q_w C_r'' & a_f b_w \\ a_w b_f & -a_f b_f - \frac{\rho}{\delta_v} M_f q_f C_v'' \end{pmatrix},$$

where q_w and q_f represent the PDFs of worker and firm distributions, respectively; $a_w = \tilde{\rho}\zeta\frac{M_w q_w}{L}$; $a_f = \tilde{\rho}\zeta\frac{M_w q_f}{N}$; $b_w = \frac{A_w x}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}\left(\varepsilon_\lambda - \frac{(1-\tilde{\beta})q}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}\varepsilon_q\right)$; and $b_f = \frac{A_f(y-b)}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}\left(-\varepsilon_q + \frac{\beta\lambda}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}\varepsilon_\lambda\right)$.

Observe that $D_{z''}f$ is continuous under the regularity conditions imposed on $C_r, C_v, Q_w, Q_f, \lambda$, and q. Also, the determinant of the matrix is strictly positive, i.e., $\bar{h}^2 C_r'' \frac{\rho}{\delta_v} C_v'' + a_f b_f \bar{h}^2 C_r'' + a_w b_w \frac{\rho}{\delta_v} C_v'' > 0$, given that $C_r'', C_v'' > 0$.

<u>*Step 1-(ii)*</u> From the previous step, I have established that $g(\cdot)$ is a continuous function defined on $[\underline{x}, \bar{x}] \times [\underline{y}, \bar{y}] \times (0,\infty)^2$. To exploit the ODE theorem, I modify $g(\cdot)$ and obtain a Lipschitz continuous function $\tilde{g}(\cdot)$ defined on $\mathbb{R}^4$.

First, I extend the support of $g(\cdot)$ to include the boundaries in which $\min\{x'(\ell), y'(\ell)\} = 0$. In particular, I assign the limiting values of $g(\cdot)$ at these boundaries. To find these values, fix $(x(\ell), y(\ell), x'(\ell))$, and consider $\{y_n'(\ell)\}_n$ that converge to 0. Then, the condition $f_w(\cdot) = 0$ implies that

$$[f_w] \quad x(y-b)\frac{\beta}{1-\tilde{\beta}}\frac{\lambda_n}{q_n}\tilde{\rho}\left(\frac{N_n'}{N_n} - \frac{L_n'}{L}\right) - \bar{h}^2 C_r''(\bar{h}L)L' \to 0 \quad \Rightarrow \quad N_n' - C_w L_n' \to 0 \text{ for some } C_w > 0,$$

where I use $N_n \to 0, \lambda_n \to 0, q_n \to \infty, u_n \to 1$, and $\frac{\theta_n'}{\theta_n} \to \frac{N_n'}{N_n} - \frac{L_n'}{L}$. Next, the condition $f_f(\cdot) = 0$ simplifies to

$$[f_f] \quad x(y-b)\left(\frac{x'}{x} - \frac{\tilde{\rho}}{(1-\tilde{\beta})q_n}(-\varepsilon_q)\left(\frac{N_n'}{N_n} - \frac{L_n'}{L}\right) - \frac{\beta}{1-\tilde{\beta}}\right) - \frac{\rho}{\delta_v}C_v''(N_n)N_n' \to 0,$$
$$\Rightarrow \quad -\frac{1}{q_n}\left(\frac{N_n'}{N_n} - \frac{L_n'}{L}\right) - C_f C_v''(N_n)N_n' + D_f \to 0 \text{ for some } C_f > 0, D_f \in \mathbb{R}.$$

Plugging in the result from $[f_w]$ above, I conclude that $-N_n'\left(\frac{1}{q_n N_n} + \frac{1}{q_n}C_f C_v''(N_n)\right) + D_f \to 0$. Because $q_n N_n = M(U_n, N_n)$ goes to zero, the parenthesis diverges. Thus, N_n' should converge to zero, and so does L_n'. In sum, I conclude that $\lim_{y'\to 0} g(\cdot) = \left(-\frac{q_w'(x)(x')^2}{q_w(x)}, 0\right)$. An analogous derivation yields $\lim_{x'\to 0} g(\cdot) = \left(0, -\frac{q_f'(y)(y')^2}{q_f(y)}\right)$.

Next, I construct a compact set $\bar{Z}$ that is sufficiently large to contain the equilibrium $z(\ell)$ and $z'(\ell)$, as will be shown below. Since $C_r'(\cdot)$ and $C_v'(\cdot)$ grow sufficiently large, I can find a finite constant $\bar{L}$, and given this value, a finite constant $\bar{N}$, such that

$$\bar{h}C_r'(\bar{h}\bar{L}) = \underline{x}\underline{y} + \bar{h}C_r'(\bar{h}M_w), \text{ and } C_v'(\bar{N}) = \frac{1-\beta}{\tilde{\rho}}q\left(\frac{M_f}{\bar{L}}\right)\underline{x}(\underline{y}-b) + C_v'(M_f), \tag{A.8}$$

and, correspondingly, define $\bar{x}' = \frac{\bar{L}}{M_w q_w(\underline{x})}$ and $\bar{y}' = \frac{\bar{N}}{M_f q_f(\underline{y})}$. Define an auxiliary set $Z \equiv [\underline{x}, \bar{x}] \times [\underline{y}, \bar{y}] \times [0, \bar{x}'] \times [0, \bar{y}']$, and then find $(\bar{x}'', \bar{y}'') \equiv \sup_{(z,z')\in Z} g(\cdot)$, which are finite since $g(\cdot)$ is continuous and Z is compact. Finally, I define $\bar{Z} \equiv [\underline{x}, \max\{\bar{x}, \underline{x} + (\bar{x}' + \bar{x}'')\}] \times [\underline{y}, \max\{\bar{y}, \underline{y} + (\bar{y}' + \bar{y}'')\}] \times [0, \bar{x}' + \bar{x}''] \times [0, \bar{y}' + \bar{y}''].$

Lastly, I construct $\tilde{g}(\cdot)$. To extend $f(\cdot) = 0$ beyond the support of x and y, I extend $q_w(\cdot)$ and $q_f(\cdot)$ to $\mathbb{R}$ by assigning boundary values whenever x or y is outside its support. For $(z, z') \in \bar{Z}$, I set $\tilde{g}(z, z') = g(z, z')$. For $(z, z') \notin \bar{Z}$, I define $\tilde{g}(z, z') = g(\hat{z}, \hat{z}')$, where $(\hat{z}, \hat{z}')$ is chosen among the boundary of $\bar{Z}$ such that each component corresponds to the closest value in $\bar{Z}$.

Step 1-(iii) From Step 1-(ii), I can represent the sorting conditions $f(\cdot) = 0$ as the system of differential equations given by $z'' = \tilde{g}(z, z')$, which is Lipschitz continuous. Then, for a given initial condition $(\underline{x}, \underline{y}, x'(0), y'(0))$, the solution $(x(\ell), y(\ell))$ uniquely exists from the standard ODE theorem.

Importantly, for any initial values $(z(0), z'(0)) \in Z$, the solution to the ODE problem $(x(\ell), y(\ell), x'(\ell), y'(\ell))$ lies in $\bar{Z}$ for all $\ell \in [0, 1]$. To see this, observe that $x''(\ell) = 0$ when $x'(\ell) = 0$, which ensures $x'(\ell) \geq 0$ and $x(\ell) \geq \underline{x}$. The same is true for firms. The upper bounds of $\bar{Z}$ do not bind by construction.

Step 2 In this step, I show that there exists a solution of the ODE problem that satisfies the two terminal boundary conditions: $x(1) = \bar{x}$ and $y(1) = \bar{y}$.

Define two real-valued functions $\mathcal{P}_w, \mathcal{P}_f : [0, \bar{L}] \times [0, \bar{N}] \to \mathbb{R}$. Specifically, let $\mathcal{P}_w(z'(0)) = \frac{1}{M_w} \int_0^1 L(s; z'(0))\, \mathrm{d}s - 1$ and $\mathcal{P}_f(z'(0)) = \frac{1}{M_f} \int_0^1 N(s; z'(0))\, \mathrm{d}s - 1$ where $(L(\ell), N(\ell))$ is given by the solution to ODE defined above. I specify $z'(0)$ to emphasize that the solution depends on this initial point. The initial point $z(0) = (\underline{x}, \underline{y})$ is always chosen to satisfy the initial boundary conditions.

Observe that $\mathcal{P}_w$ and $\mathcal{P}_f$ are continuous. Lipschitz continuity of $\tilde{g}(\cdot)$ ensures that the solution (z, z') is continuous in the initial point, $z'(0)$, at each ℓ. Observe that $\mathcal{P}_w = \frac{1}{M_w} \int_0^1 M_w q_w(t)\, \mathrm{d}t - 1$ by change of variables, where the extended q_w is bounded by construction. Then, by the Dominated Convergence Theorem, $\mathcal{P}_w$ is continuous. Similarly, $\mathcal{P}_f$ is continuous.

Step 2-(i) shows that $\mathcal{P}_w(0, N) < 0\ \forall N \in [0, \bar{N}]$ and $\mathcal{P}_f(L, 0) < 0\ \forall L \in [0, \bar{L}]$. Step 2-(ii) shows $\mathcal{P}_w(\bar{L}, N) > 0\ \forall N \in [0, \bar{N}]$ and $\mathcal{P}_f(L, \bar{N}) > 0\ \forall L \in [0, \bar{L}]$. Then, by Pointcaré-Miranda theorem (multivariable version of intermediate value theorem), there exists $(x'(0), y'(0)) \in (0, \bar{x}') \times (0, \bar{y}')$ that gives $\mathcal{P}_w = \mathcal{P}_f = 0$.

Step 2-(i) Find the solution $(z(\ell), z'(\ell))$ given the initial point $(\underline{x}, \underline{y}, 0, y'(0)) \in Z$. From Step 1-(ii), $x''(0) = 0$, which implies that $x'(\ell)$ remains zero, so that $x(\ell) = \underline{x}$ for all ℓ. Thus, $\mathcal{P}_w(0, N) < 0$ for all $0 \leq N \leq \bar{N}$. Similarly, $\mathcal{P}_f(L, 0) < 0$ for all $0 \leq L \leq \bar{L}$.

Step 2-(ii) Consider the solution given the initial point $(\underline{x}, \underline{y}, \bar{x}', y'(0)) \in Z$. Since the solution $(z(\ell), z'(\ell)) \in \bar{Z}$ for all ℓ, the sorting conditions of workers and firms are satisfied, which ensures that each worker and firm chooses ℓ that maximizes their value. Note that at this point, although I allow $x(\ell) > \bar{x}$ or $y(\ell) > \bar{y}$, the problem outside of the original support of x and y is well-defined due to the construction of $\tilde{g}(\cdot)$.

The worker's value, evaluated at the solution, satisfies $\rho V^u(\underline{x}, 0) \geq \rho V^u(\underline{x}, \ell)$ for all ℓ. Observe that $\rho V^u(\underline{x}, \ell) \geq b\underline{x} - \bar{h}C_r'(\bar{h}L(\ell))$ and $\rho V^u(\underline{x}, 0) \leq b\underline{x} + \underline{x}\underline{y} - \bar{h}C_r'(\bar{h}\bar{L})$, together implying

$$\bar{h}C_r'(\bar{h}L(\ell)) \geq \bar{h}C_r'(\bar{h}\bar{L}) - \underline{x}\underline{y} = \bar{h}C_r'(\bar{h}M_w),$$

where the last equality comes from (A.8). Thus, $L(\ell) \geq M_w$ for all ℓ, and I conclude that $\mathcal{P}_w(\bar{L}, N) > 0$ for all $0 \leq N \leq \bar{N}$. Similarly, the solution also satisfies $\rho \bar{V}^v(\underline{y}, 0) \geq \rho \bar{V}^v(\underline{y}, \ell)$. Observe that $\rho \bar{V}^v(\underline{y}, \ell) \geq -C_v'(N(\ell))$ and

$\rho \bar{V}^v(\underline{y}, 0) \leq A_f(0)(\underline{y} - b) - C_v'(\bar{N})$, which together imply

$$C_v'(N(\ell)) \geq C_v'(\bar{N}) - \frac{\delta_v}{\delta_v + \rho} \frac{(1-\beta)q(0)}{\tilde{\rho} + (1-\tilde{\beta})q(0)} \underline{x}(\underline{y} - b) \geq C_v'(\bar{N}) - \frac{(1-\beta)}{\tilde{\rho}} q\left(\frac{M_f}{\bar{L}}\right) \underline{x}(\underline{y} - b) = C_v'(M_f),$$

where I use $q(0) = q\left(\frac{N(0)}{u(0)L(0)}\right) \leq q\left(\frac{M_f}{\bar{L}}\right)$ for the second inequality and (A.8) for the last equality. Thus, $N(\ell) \geq M_f$ for all ℓ, and I conclude that $\mathcal{P}_f(L, \bar{N}) > 0$ for all $0 \leq L \leq \bar{L}$.

Step 2-(iii) I confirm that the solution to ODE problem is indeed an equilibrium. Let $(x(\ell), y(\ell), x'(\ell), y'(\ell))$ be the solution to ODE such that $\mathcal{P}_w = \mathcal{P}_f = 0$. The condition $\mathcal{P}_w = \mathcal{P}_f = 0$ ensures that $x(1) = \bar{x}$ and $y(1) = \bar{y}$, so the boundary conditions are satisfied. Note that wage determination, housing and business services market clearing conditions, and the flow-balance conditions are already imposed.

Next, both $x'(\ell)$ and $y'(\ell)$ are strictly positive for all ℓ, so that $(z(\ell), z'(\ell))$ is entirely obtained from the sorting conditions of the model, rather than the extended ODE problem. To show this, suppose, to the contrary, that there exists $\ell^* < 1$ such that $x'(\ell^*) = L(\ell^*) = 0$. Then, from Step 1-(ii), $x''(\ell) = L(\ell) = L'(\ell) = 0$ for all $\ell \geq \ell^*$. Near ℓ^*, $\lambda(\ell)$ is bounded above. If not, workers of $\underline{x}$ would deviate to a higher ℓ close to ℓ^* where they can gain from both labor and housing markets. Thus, $y''(\ell) = N(\ell) = N'(\ell) = 0$ for all $\ell \geq \ell^*$. From the continuity of the solution, $x'(\ell), y'(\ell), x''(\ell)$, and $y''(\ell)$ converge to zero when $\ell \to \ell^*$. To rule out the deviation of firms, $q(\ell)$ is also bounded above near ℓ^*. Based on these results, the following limits should hold when $\ell \to \ell^*$,

$$[f_w] \quad \tilde{\rho}\zeta\left(\frac{N'}{N} - \frac{L'}{L}\right)\left(\varepsilon_\lambda(\theta) - \frac{(1-\tilde{\beta})q}{\tilde{\rho} + \beta\lambda + (1-\tilde{\beta})q}\varepsilon_q\right) - \frac{(1-\tilde{\beta})q}{\tilde{\rho} + \beta\lambda + (1-\tilde{\beta})q}\beta\frac{\lambda}{\theta} - C_w \lambda C_r''(\bar{h}L)L\frac{L'}{L} \to 0,$$

$$[f_f] \quad -\tilde{\rho}\zeta\left(\frac{N'}{N} - \frac{L'}{L}\right)\left(-\varepsilon_q(\theta) + \frac{\beta\lambda}{\tilde{\rho} + \beta\lambda + (1-\tilde{\beta})q}\varepsilon_\lambda\right) - \frac{\beta\lambda}{\tilde{\rho} + \beta\lambda + (1-\tilde{\beta})q}(1-\tilde{\beta})\frac{q}{\theta} - C_f C_v''(N)N\frac{N'}{N} \to 0,$$

for some $C_w > 0, C_f > 0$. As L goes to zero, so does $C_r''(\bar{h}L)L$. To see this, observe that $C_r'(\varepsilon) - C_r'(0) = \int_0^\varepsilon C_r''(H)H \cdot \frac{1}{H}\,\mathrm{d}H > 0.5(\lim_{H\to 0} C_r''(H)H)\int_0^\varepsilon \frac{1}{H}\,\mathrm{d}H$ when ε is sufficiently small. Since the left-hand side converges to zero from $C_r \in \mathcal{C}^2$, this inequality implies that the term in parentheses on the right-hand side should converge to zero. The same is true for $C_v''(N)N$. Then the above equations simplify to

$$\left(\frac{N'}{N} - \frac{L'}{L}\right) - D_w \to 0, \quad -\left(\frac{N'}{N} - \frac{L'}{L}\right) - D_f \to 0,$$

where D_w and D_f are strictly positive and bounded constants. It is easy to verify that the two equations cannot hold at the same time, which leads to a contradiction. Analogous reasoning implies that $y'(\ell)$ is strictly positive for all ℓ.

Lastly, $(x(\ell), y(\ell), x'(\ell), y'(\ell)) \in \bar{Z}$ for all ℓ by construction. Moreover, since $x(\ell) \leq \bar{x}$, $y(\ell) \leq \bar{y}$, $L(\ell) > 0$, and $N(\ell) > 0$, the solution to ODE problem satisfies the original problem in Step 1-(i) and thus the first-order conditions in (8). Note that $x'(\ell), y'(\ell) > 0$ ensure $A_w'(\ell), A_f'(\ell) > 0$, which implies that the value functions of workers and firms are supermodular in (x, ℓ) and (y, ℓ), respectively. Therefore, the solution to (8) attains the global maximum.

Note that while the assignment is uniquely determined by a given initial condition, multiple equilibria may exist if there are multiple initial conditions that satisfy the boundary conditions.

A.3 Proof of Proposition 2

Population density. I first show that population density strictly increases in ℓ if δ_v is sufficiently large. When benefits of labor market rise in ℓ, housing rents and consequently population density also increase in ℓ. Otherwise, all workers would strictly prefer higher-ℓ locations. It suffices to show that, for a sufficiently small $1-\tilde{\beta}$,

$$A_w'(\ell)x(\ell) > A_w'(\ell)B_w(\ell) + A_w(\ell)B_w'(\ell) \quad \forall \ell.$$

Suppose, to the contrary, that there exists a sequence $\{1-\tilde{\beta}_n, \ell_n\}_n$ such that $1-\tilde{\beta}_n \to 0$ and

$$A_w'(\ell_n;\tilde{\beta}_n)x(\ell_n;\tilde{\beta}_n) \leq A_w'(\ell_n;\tilde{\beta}_n)B_w(\ell_n;\tilde{\beta}_n) + A_w(\ell_n;\tilde{\beta}_n)B_w'(\ell_n;\tilde{\beta}_n),$$

where I index each function with $\tilde{\beta}_n$ to emphasize that each term is a function of an equilibrium assignment given $\tilde{\beta}_n$. In the limit, since $B_w(\ell_n;\tilde{\beta}_n)=0$ and also $B_w'(\ell_n;\tilde{\beta}_n)=0$ from (A.6), the above inequality implies $\lim A_w'(\ell_n;\tilde{\beta}_n) \leq 0$. From (A.5), this yields

$$\lim_n \left(\varepsilon_\lambda(\theta(\ell_n;\tilde{\beta}_n)) \frac{\theta'(\ell_n;\tilde{\beta}_n)}{\theta(\ell_n;\tilde{\beta}_n)} \frac{\tilde{\rho}}{\tilde{\rho}+\beta\lambda(\ell_n;\tilde{\beta}_n)} + \frac{y'(\ell_n;\tilde{\beta}_n)}{y(\ell_n;\tilde{\beta}_n)-b} \right) = 0,$$

where I rule out the negative sign as $A_w'(\cdot) > 0$ under PAM. Given that $y'(\ell_n;\tilde{\beta}_n) > 0$, there exists $\bar{n} \in \mathbf{N}$ such that $\theta'(\ell_n;\tilde{\beta}_n) < 0$ for all $n \geq \bar{n}$. Then, from (A.7), $N'(\ell_n;\tilde{\beta}_n) < 0$, implying $c'(\ell_n;\tilde{\beta}_n) < 0$ for all $n \geq \bar{n}$. To ensure that the firm sorting condition holds,

$$A_f'(\ell_n;\tilde{\beta}_n)(y(\ell_n;\tilde{\beta}_n)-b) < A_f'(\ell_n;\tilde{\beta}_n)B_f(\ell_n;\tilde{\beta}_n) + A_f(\ell_n;\tilde{\beta}_n)B_f'(\ell_n;\tilde{\beta}_n) \quad \forall n \geq \bar{n}.$$

As $1-\tilde{\beta}_n$ goes to zero, $B_f'(\cdot)$ converges to $A_w(\cdot)A_w'(\cdot)$, and in turn, the second term on the right-hand side vanishes. Since $y-b > B_f$, there exists $n^* \geq \bar{n}$, such that $A_f'(\ell_{n^*};\tilde{\beta}_{n^*}) < 0$. However, this is not possible as both $x'(\ell_{n^*};\tilde{\beta}_{n^*})$ and $q'(\ell_{n^*};\tilde{\beta}_{n^*})$ are strictly positive. Contradiction.

Wage. Next, I show that wages strictly increase in ℓ if δ_v is sufficiently large. From (10), it suffices to show

$$\left(\beta \frac{\tilde{\rho}+\lambda(\ell)}{\tilde{\rho}+\beta\lambda(\ell)+(1-\tilde{\beta})q(\ell)}(y(\ell)-b) \right)' \geq 0 \quad \forall \ell.$$

When $\lambda'(\ell) \geq 0$, it is easy to verify that the above holds when $1-\tilde{\beta}$ is close to zero. To consider the case $\lambda'(\ell) < 0$, I rewrite the above as

$$\left(A_w(\ell)\frac{\tilde{\rho}+\beta\lambda(\ell)}{\tilde{\rho}+\beta\lambda(\ell)+(1-\tilde{\beta})q(\ell)}+\frac{\beta\tilde{\rho}}{\tilde{\rho}+\beta\lambda(\ell)+(1-\tilde{\beta})q(\ell)}(y(\ell)-b)\right)' \geq 0 \quad \forall \ell.$$

When $1-\tilde{\beta}$ is close to zero, the first term converges to $A_w(\ell)$, which is increasing in ℓ in equilibrium. Moreover, the second term increases in ℓ as $\lambda'(\ell) < 0$ and $y'(\ell) > 0$.

Discussions on a non-pure assignment equilibrium. Beyond a pure assignment equilibrium which is the focus of this paper, an alternative equilibrium with non-pure assignment exists in which multiple worker or firm types coexist within a given location ℓ. For example, an economy in which all workers and firms randomly choose location ℓ is one mixed equilibrium. However, the proposition below shows that a non-pure assignment equilibrium does not exhibit equilibrium properties consistent with the data, as opposed to a pure assignment equilibrium as shown in Proposition 2.

Proposition A.1. *If the productivity of workers $\mathbb{E}[x|\ell]$, the productivity of firms $\mathbb{E}[y|\ell]$, and population density $L(\ell)$ strictly increase in ℓ, an equilibrium allocation is pure.*

Proof. When there are multiple types of workers and firms in a given ℓ, due to random matching, workers' job opportunities $A_w(\mathbb{E}[y|\ell], \lambda(\ell))$ depend on the average firm productivity, $\mathbb{E}[y|\ell]$, instead of $y(\ell)$. Similarly, firms' hiring opportunities $A_f(\mathbb{E}[x|\ell], q(\ell))$, are determined by average worker types $\mathbb{E}[x|\ell]$. Aside from these modifications, the value functions have the same expression as in (4) and (5).

I proceed by contradiction. Suppose that there exists a mixed equilibrium where $\mathbb{E}[x|\ell]$, $\mathbb{E}[y|\ell]$, and $L(\ell)$ strictly increase in ℓ. Then, I can choose two locations and either the productivity of workers or firms choosing two locations with positive probabilities.

First, consider an equilibrium with workers of x^* who choose both $\ell' < \ell''$. Job opportunities must be equal in two locations, $A_w(\ell') = A_w(\ell'')$. If not, all workers of $x > x^*$ choose $\ell > \ell''$ while others of $x < x^*$ choose $\ell < \ell'$ in equilibrium, and $[\ell', \ell'']$ are only chosen by workers of x^*, which is not possible due to the continuity of $Q_w(\cdot)$. Because workers of x^* choose both locations, they are indifferent between the two:

$$A_w(\ell')x^* - A_w(\ell')B_w(\ell') - \bar{h}r(\ell') = A_w(\ell'')x^* - A_w(\ell'')B_w(\ell'') - \bar{h}r(\ell'').$$

Because $r(\ell'') > r(\ell')$ from the assumption on population density, for the above equality to hold, $B_w(\ell'') < B_w(\ell')$. Given that $\mathbb{E}[x|\ell''] > \mathbb{E}[x|\ell']$, to ensure this inequality, it must be that $\theta(\ell') < \theta(\ell'')$. As a result, $\lambda(\ell') < \lambda(\ell'')$, leading to $A_w(\ell') < A(\ell'')$ since firm productivity is also higher in ℓ''. Contradiction.

Second, suppose that there are firms of y^* that choose both $\ell' < \ell''$. Following the same step, it must be $A_f(\ell') = A_f(\ell'')$ and the indifference condition for y^* implies that $A_f(\ell')B_f(\ell') > A_f(\ell'')B_f(\ell'')$ if and only if and

$c(\ell') < c(\ell'')$. Consider the case of $\theta(\ell'') > \theta(\ell')$. Then, $A_f B_f = \frac{\tilde{\rho}+\beta\lambda}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q} A_w$ is weakly greater in ℓ'' as $A_w(\cdot)$ is also higher in ℓ'' as implied by workers' sorting. Moreover, since $L(\cdot)$ is greater in ℓ'', so does $V(\cdot)$, implying $c(\ell'') > c(\ell')$; this cannot be true in equilibrium. As a result, market tightness is smaller in ℓ'', implying $q(\ell') \leq q(\ell'')$. Combined with the assumption $\mathbb{E}[x|\ell'] < \mathbb{E}[x|\ell'']$, $A_f(\ell') < A_f(\ell'')$. Contradiction.

A.4 Proofs of Section 3.2

Proof of Lemma 1. The planner chooses a (potentially non-pure) assignment, represented by the share of workers of x assigned to locations below ℓ, denoted by $\bar{m}_w(\ell|x)$, and the share of firms of y assigned to locations below ℓ, denoted by $\bar{m}_f(\ell|y)$, that solves the following problem:

$$\begin{aligned} \max_{\bar{m}_w(\cdot),\bar{m}_f(\cdot)} \quad & \int_0^1 \left[\mathbb{E}[x|\ell]\,\mathbb{E}[y|\ell](1-u(\ell))L(\ell) - C_r(\bar{h}L(\ell)) - C_v(N(\ell))\right] \mathrm{d}\ell \qquad \text{(A.9)} \\ \text{s.t.} \quad & \int_0^\ell \frac{L(\ell')}{M_w}\,\mathrm{d}\ell' = \int_{\underline{x}}^{\overline{x}} \bar{m}_w(\ell|x)\,\mathrm{d}Q_w(x), \int_0^\ell \frac{N(\ell')}{M_f}\,\mathrm{d}\ell' = \int_{\underline{y}}^{\overline{y}} \bar{m}_f(\ell|y)\,\mathrm{d}Q_f(y), u(\ell) = \frac{\delta}{\delta+\lambda(\ell)}, \ \forall \ell. \end{aligned}$$

I first show that the optimal allocation is PAM, i.e., the assignment is pure and strictly increasing. Suppose that the optimal assignment is characterized with $(\mathbb{E}[x|\ell], \mathbb{E}[y|\ell], L(\ell), N(\ell), u(\ell))$. Without loss of generality, suppose that $(1-u(\ell))L(\ell)\,\mathbb{E}[y|\ell]$ increases in ℓ. Consider a new problem where the planner assigns workers given the firm assignment while maintaining population densities across ℓ, with the objective of maximizing (A.9). In other words, a new worker assignment $\bar{m}_{w1}(\ell|x)$ must satisfy $\int_{\underline{x}}^{\overline{x}} \bar{m}_{w1}(\ell|x)\,\mathrm{d}Q_w(x) = \int_0^\ell \frac{L(\ell')}{M_w}\,\mathrm{d}\ell'$. The solution to this problem must coincide with the assignment in the original problem. Since this is a standard linear assignment problem, the optimal solution must be an extreme point of the feasible set, i.e., a pure assignment. Moreover, because of the complementarity between $(1-u(\ell))L(\ell)\,\mathbb{E}[y|\ell]$ and $\mathbb{E}[x|\ell]$, the optimal pure worker assignment $x(\ell)$ should be increasing in ℓ. Applying the same logic, the optimal pure firm assignment $y(\ell)$ should be increasing in $(1-u(\ell))x(\ell)L(\ell)$.

Next, I establish that not only $x(\ell)$ but also $y(\ell)$ increases in ℓ. Suppose, for contradiction, that there exist two locations ℓ, ℓ' such that $x(\ell) < x(\ell')$ but $y(\ell) > y(\ell')$. The latter condition implies

$$(1-u(\ell'))x(\ell')L(\ell') \leq (1-u(\ell))x(\ell)L(\ell) \Rightarrow \frac{(1-u(\ell'))L(\ell')}{(1-u(\ell))L(\ell)} \leq \frac{x(\ell)}{x(\ell')} < 1.$$

However, by the assumption that $(1-u(\ell))y(\ell)L(\ell)$ increases in ℓ, it follows that

$$(1-u(\ell'))y(\ell')L(\ell') \geq (1-u(\ell))y(\ell)L(\ell) \Rightarrow \frac{(1-u(\ell'))L(\ell')}{(1-u(\ell))L(\ell)} \geq \frac{y(\ell')}{y(\ell)} > 1,$$

which contradicts the first inequality.

Optimal allocation. Given that the optimal allocation is PAM, the planner's problem becomes finding two increasing functions, $x(\ell)$ and $y(\ell)$, and can be formulated by the following Hamiltonian:

$$\mathcal{H} = (1-u(\ell))y(\ell)x(\ell)L(\ell) - C_r(\bar{h}L(\ell)) - C_v(N(\ell)) + \frac{\mu_w(\ell)L(\ell)}{M_w q_w(x(\ell))} + \frac{\mu_f(\ell)N(\ell)}{M_f q_f(y(\ell))} + \mu_u(\ell)\left(u(\ell) - \frac{\delta}{\delta+\lambda(\theta(\ell))}\right),$$

where μ_w and μ_f are co-state variables of x and y, respectively, and μ_u is the multiplier of unemployment rates. The market tightness is defined as before, $\theta(\ell) = \frac{V(\ell)}{u(\ell)L(\ell)}$. I omit the superscript *, which denotes the optimal assignment, in the appendix while retaining it in the main draft.

First-order conditions regarding the allocation of workers are given by,

$$0 = (1-u(\ell))y(\ell)x(\ell) - \bar{h}C_r'(\bar{h}L(\ell)) + \frac{\mu_w(\ell)}{M_w q_w(x(\ell))} - \mu_u(\ell)\frac{\varepsilon_\lambda(\theta(\ell))}{L(\ell)}u(\ell)(1-u(\ell)), \tag{L}$$

$$\mu_w'(\ell) = -(1-u(\ell))y(\ell)L(\ell) + \frac{\mu_w(\ell)L(\ell)}{M_w q_w^2(x(\ell))}q_w'(x(\ell)). \tag{x}$$

Similarly, first-order conditions of firm allocation are given by,

$$0 = -C_v'(N(\ell)) + \frac{\mu_f(\ell)}{M_f q_f(y(\ell))} + \mu_u(\ell)\frac{\varepsilon_\lambda(\theta(\ell))}{N(\ell)}u(\ell)(1-u(\ell)), \tag{N}$$

$$\mu_f'(\ell) = -(1-u(\ell))x(\ell)L(\ell) + \frac{\mu_f(\ell)N(\ell)}{M_f q_f^2(y(\ell))}q_f'(y(\ell)). \tag{y}$$

Finally, the constraint on unemployment rates leads to

$$0 = -y(\ell)x(\ell)L(\ell) + \mu_u(\ell)(1-\varepsilon_\lambda(\theta(\ell))(1-u(\ell))). \tag{u}$$

From now on, I omit ℓ for notational simplicity unless it causes any confusion. Differentiating the condition (L) with respect to ℓ and substituting the conditions (x) and (u) yields

$$\begin{aligned} C_r''\bar{h}^2L' &= \frac{(1-u)(1-\varepsilon_\lambda)}{1-\varepsilon_\lambda(1-u)}y'x - \frac{\varepsilon_\lambda u(1-u)}{1-\varepsilon_\lambda(1-u)}x'y - \frac{(1-\varepsilon_\lambda)xy}{(1-\varepsilon_\lambda(1-u))^2}u' - \frac{u(1-u)xy}{(1-\varepsilon_\lambda(1-u))^2}\frac{\partial\varepsilon_\lambda}{\partial\ell} \\ &= \frac{(1-u)(1-\varepsilon_\lambda)}{1-\varepsilon_\lambda(1-u)}y'x - \frac{\varepsilon_\lambda u(1-u)}{1-\varepsilon_\lambda(1-u)}x'y + \frac{u(1-u)xy}{(1-\varepsilon_\lambda(1-u))^2}\left((1-\varepsilon_\lambda)\varepsilon_\lambda - \frac{\lambda''}{\lambda'}\theta + \varepsilon_\lambda - 1\right)\frac{\theta'}{\theta}, \end{aligned} \tag{A.10}$$

where I use $1 + \frac{\partial}{\partial u}\left(\frac{\varepsilon_\lambda(1-u)u}{1-\varepsilon_\lambda(1-u)}\right) = \frac{1-\varepsilon_\lambda}{(1-\varepsilon_\lambda(1-u))^2}$ for the first line. For the second line, I use $\frac{\partial u}{\partial \ell}\frac{1}{u} = -(1-u)\varepsilon_\lambda\frac{\partial\theta}{\partial\ell}\frac{1}{\theta}$ and $\frac{\partial\varepsilon_\lambda}{\partial\ell}\frac{1}{\varepsilon_\lambda} = \left(\frac{\lambda''(\theta)}{\lambda'(\theta)}\theta - \varepsilon_\lambda + 1\right)\frac{\partial\theta}{\partial\ell}\frac{1}{\theta}$. Similarly, by differentiating the condition (N) with respect to ℓ, and plugging in the conditions (y) and (u), I obtain

$$C_v''N' = -\frac{(1-\varepsilon_\lambda)(1-u)}{1-\varepsilon_\lambda(1-u)}xy'\frac{L}{N} + \frac{\varepsilon_\lambda(1-u)u}{1-\varepsilon_\lambda(1-u)}\frac{L}{N}x'y - \frac{u(1-u)xy}{(1-\varepsilon_\lambda(1-u))^2}\left((1-\varepsilon_\lambda)\varepsilon_\lambda - \frac{\lambda''}{\lambda'}\theta + \varepsilon_\lambda - 1\right)\frac{L}{N}\frac{\theta'}{\theta} \quad \text{(A.11)}$$

Multiply (A.10) by $L(\ell)$ and (A.11) by $N(\ell)$ respectively, and then sum them up to obtain the first equation that characterizes the optimal assignment:

$$C_r''(\bar{h}L(\ell))\bar{h}^2L'(\ell)L(\ell) + C_v''(N(\ell))N'(\ell)N(\ell) = 0, \quad \text{(A.12)}$$

where I use $\frac{\theta'}{\theta} = \frac{1}{1-\varepsilon_\lambda(1-u)}\left(\frac{N'}{N} - \frac{L'}{L}\right)$. Furthermore, by rearranging (A.10) and then plugging in (A.12), I obtain the second equation:

$$\begin{aligned}&\frac{(1-u)}{1-\varepsilon_\lambda(1-u)}\left((1-\varepsilon_\lambda)\frac{y'}{y} - \varepsilon_\lambda u\frac{x'}{x}\right)xy \\ &= \left[C_r''\bar{h}^2L + \frac{\varepsilon_\lambda u(1-u)}{(1-\varepsilon_\lambda(1-u))^3}\left(\varepsilon_\lambda(1-\varepsilon_\lambda) - \frac{\lambda''\theta}{\lambda'} + \varepsilon_\lambda - 1\right)xy\left(1 + \frac{C_r''}{C_v''}\bar{h}^2\left(\frac{L}{N}\right)^2\right)\right]\frac{L'}{L}\end{aligned} \quad \text{(A.13)}$$

where I use $\frac{L'}{L} - \frac{N'}{N} = \frac{L'}{L}\left(1 + \frac{C_r''}{C_v''}\bar{h}^2\left(\frac{L}{N}\right)^2\right)$. In sum, the optimal allocation $\{x(\ell), y(\ell)\}$ is characterized by (A.12) and (A.13), together with boundary conditions.

Proof of Proposition 3. Recall that worker and firm sorting conditions, $f_w = f_f = 0$, are given by (Section A.2),

$$\begin{aligned}[f_w] \quad & \frac{\partial}{\partial\ell}\left(\frac{\beta\lambda(\ell)}{\tilde{\rho}+\beta\lambda(\ell)}(y(\ell)-b)\left(x - \frac{(1-\tilde{\beta})q}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}x(\ell)\right)\right) = \bar{h}r'(\ell), \\ [f_f] \quad & \frac{\partial}{\partial\ell}\left(\frac{\delta_v}{\rho+\delta_v}x(\ell)\frac{(1-\beta)q(\ell)}{\tilde{\rho}+(1-\tilde{\beta})q(\ell)}\left(y - b - \frac{\beta\lambda}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}(y(\ell)-b)\right)\right) = c'(\ell).\end{aligned}$$

To compare these conditions to those of the optimal assignment, I multiply $[f_w]$ and $[f_f]$ by $L(\ell)$ and $N(\ell)$, respectively. The expression is relatively long, and thus I proceed in steps. I omit ℓ from now on. First, I gather the terms related to market tightness in $[f_w]$ (omitting $x(\ell)(y(\ell)-b)$),

$$\begin{aligned}&L\left[\left(\frac{\beta\lambda}{\tilde{\rho}+\beta\lambda}\right)'\frac{\tilde{\rho}+\beta\lambda}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q} - \frac{\beta\lambda}{\tilde{\rho}+\beta\lambda}\left(\frac{(1-\tilde{\beta})q}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}\right)'\right] \\ &= \frac{\beta\lambda(\ell)L(\ell)}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}\left[\frac{1}{\tilde{\rho}+\beta\lambda}\left(\tilde{\rho} + \frac{(1-\tilde{\beta})\beta\lambda q}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}\right)\varepsilon_\lambda - \frac{(1-\tilde{\beta})q}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}\varepsilon_q\right]\frac{\theta'}{\theta} \equiv \Theta_w.\end{aligned}$$

Next, I gather the terms related to market tightness in $[f_f]$ (omitting $x(\ell)(y(\ell)-b)$),

$$\begin{aligned}
&N\frac{\delta_v}{\rho+\delta_v}\left[\left(\frac{(1-\beta)q(\ell)}{\tilde{\rho}+(1-\tilde{\beta})q(\ell)}\right)'\frac{\tilde{\rho}+(1-\tilde{\beta})q}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}-\frac{(1-\beta)q(\ell)}{\tilde{\rho}+(1-\tilde{\beta})q}\left(\frac{\beta\lambda}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}\right)'\right]\\
&=\frac{\frac{\delta_v}{\rho+\delta_v}(1-\beta)qN(\ell)}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}\left[\frac{1}{\tilde{\rho}+(1-\tilde{\beta})q}\left(\tilde{\rho}+\frac{(1-\tilde{\beta})\beta q\lambda}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}\right)\varepsilon_q-\frac{\beta\lambda}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}\varepsilon_\lambda\right]\frac{\theta'}{\theta}\equiv\Theta_f.
\end{aligned}$$

Combining the above with the remaining terms, $[f_w]$ multiplied by $L(\ell)$ and $[f_f]$ multiplied by $N(\ell)$ are given by

$$\begin{aligned}
\bar{h}r'L&=-\frac{(1-\tilde{\beta})q}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}\frac{\beta\lambda}{\tilde{\rho}+\beta\lambda}x'(y-b)L+\frac{\beta\lambda}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}y'xL+\Theta_w,\\
c'N&=\frac{(1-\tilde{\beta})q}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}\frac{\delta_v}{\rho}x'(y-b)N-\frac{\delta_v}{\rho+\delta_v}\frac{(1-\beta)q}{\tilde{\rho}+(1-\tilde{\beta})q(\ell)}\frac{\beta\lambda(\ell)}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}y'xN+\Theta_f.
\end{aligned}$$

In particular, when ρ converges to 0, so does $1-\tilde{\beta}=\frac{\rho}{\rho+\delta_v}(1-\beta)$, so the above expressions simplify to

$$L\bar{h}r'=\frac{\beta\lambda}{\delta+\beta\lambda}xy'L+\frac{\delta}{\delta+\beta\lambda}\frac{\beta\lambda}{\delta+\beta\lambda}\varepsilon_\lambda\frac{\theta'}{\theta}x(y-b)L, \tag{A.14}$$

$$Nc'=\frac{N(1-\beta)q}{\delta+\beta\lambda}x'(y-b)-\frac{\beta\lambda}{\delta+\beta\lambda}(1-\beta)(1-u)xy'L+\frac{\delta(1-u)(1-\beta)}{\delta+\beta\lambda}\left(\frac{\delta}{\delta+\beta\lambda}\varepsilon_\lambda-1\right)\frac{\theta'}{\theta}x(y-b)L, \tag{A.15}$$

where I use $N=V=\theta uL$ and $\varepsilon_q=\varepsilon_\lambda-1$.

To prove the inefficiency of the equilibrium, I will show that

$$\bar{h}^2C_r''(\bar{h}L(\ell))L(\ell)L'(\ell)+C_v''(N(\ell))N'(\ell)N(\ell)>0, \tag{A.16}$$

which violates (A.12). I consider four cases, depending on the sign of $L'(\ell)$ and $N'(\ell)$. First, if both $L'(\ell)$ and $N'(\ell)$ are strictly positive, (A.16) immediately follows. Next, consider the opposite case, $L'(\ell)\leq 0$ and $N'(\ell)\leq 0$. From (A.14), $\theta'(\ell)<0$. Moreover, adding (A.14) multiplied with $(1-\beta)(1-u)$ and (A.15), I obtain

$$(1-\beta)(1-u)\bar{h}^2C_r''LL'+C_v''N'N=N\frac{(1-\beta)q}{\delta+\beta\lambda}x'(y-b)+\frac{\delta(1-\beta)(1-u)}{\delta+\beta\lambda}(\varepsilon_\lambda-1)\frac{\theta'}{\theta}x(y-b)L. \tag{A.17}$$

The left-hand side is negative from the assumption while the right-hand side is positive. Hence, this case cannot be an equilibrium.

Third, consider the case that $L'(\ell)>0$, $N'(\ell)\leq 0$, which leads to $\theta'(\ell)<0$. Again, using (A.17),

$$\begin{aligned}
&\bar{h}^2C_r''LL'+C_v''N'N>(1-\beta)(1-u)\bar{h}^2C_r''LL'+C_v''N'N\\
&\qquad=N\frac{(1-\beta)q}{\delta+\beta\lambda}x'(y-b)+\frac{\delta}{\delta+\beta\lambda}(1-\beta)(1-u)(\varepsilon_\lambda-1)\frac{\theta'}{\theta}x(y-b)L>0.
\end{aligned}$$

The final case, $L'(\ell) \leq 0$ and $N'(\ell) > 0$, implies $\theta'(\ell) > 0$, which contradicts (A.14), and cannot be an equilibrium.

The planner can correct externalities by using spatial transfers. To compute spatial transfers, I will compare the worker and firm sorting conditions to the planner's the counterparts. Rearranging (A.14) yields

$$\bar{h}r'(\ell) = \frac{\beta(1-u)}{u+\beta(1-u)}y'x + \frac{u(1-u)\varepsilon_\lambda\beta}{(u+\beta(1-u))^2}\frac{\theta'}{\theta}x(y-b).$$

When the terms in the planner's condition (A.10) but absent from the above are imposed as marginal changes in spatial transfers, workers internalize their externalities and hence their decisions coincide with those of the planner. Integrating these terms yields the following spatial transfer schedule to workers:

$$\begin{aligned} t_w(\ell) = t_w^0 + \int_0^\ell \Bigg[&(1-u)\left(\frac{1-\varepsilon_\lambda}{1-\varepsilon_\lambda(1-u)} - \frac{\beta}{u+\beta(1-u)}\right)y'x - \frac{\varepsilon_\lambda(1-u)u}{1-\varepsilon_\lambda(1-u)}x'y - \frac{u(1-u)\beta\varepsilon_\lambda}{(u+\beta(1-u))^2}x(y-b)\frac{\theta'}{\theta} \\ &+ \frac{u(1-u)}{(1-\varepsilon_\lambda(1-u))^2}\left((1-\varepsilon_\lambda)\varepsilon_\lambda - \frac{\lambda''}{\lambda'}\theta + \varepsilon_\lambda - 1\right)xy\frac{\theta'}{\theta}\Bigg]\,\mathrm{d}t, \end{aligned} \tag{A.18}$$

where t_w^0 is a constant that ensures the budget balance of the government. All functions $(x, y, u, \varepsilon_\lambda, \theta)$ are evaluated at the optimal assignment.

Under the Hosios condition $\varepsilon_\lambda = 1 - \beta$ and zero unemployment benefit $b = 0$, among the integrands, the first term—externalities from $y'x$—and the last two terms—related to inefficiencies from the market tightness—disappear. The role of the Hosios condition related to matching rates is standard, which is well studied under homogeneous agents. In addition, under worker and firm heterogeneity, this condition ensures that there are no externalities arising from heterogeneous firm types in worker sorting. Workers internalize the impact of firm heterogeneity because their match surplus depends on $y(\ell)$. The Hoisos condition ensures that this marginal benefit is equal to the counterpart evaluated by the planner. See Corollary 1 for simplified $t_w(\ell)$. Even under the Hosios condition, externalities arise from workers' negative impact on local firms, whose value depends on the average worker productivity in the local market. This intuition can be confirmed by observing that the second term is equal to $-\frac{N}{L}\frac{\partial\rho\bar{V}^v}{\partial x}x'$ after some rearrangement.

Similarly, I will assess the firm sorting condition. Rearranging (A.15) yields

$$c' = \frac{(1-\beta)(1-u)u}{u+\beta(1-u)}\frac{L}{N}x'(y-b) - \frac{(1-\beta)\beta(1-u^2)}{u+\beta(1-u)}\frac{L}{N}y'x + \frac{u(1-u)(1-\beta)}{u+\beta(1-u)}\left(\frac{u}{u+\beta(1-u)}\varepsilon_\lambda - 1\right)\frac{L}{N}x(y-b)\frac{\theta'}{\theta}.$$

Analogous to the above discussion, by comparing this to (A.11), I obtain the spatial transfer for firms:

$$
\begin{aligned}
t_f(\ell) = t_f^0 + \int_0^\ell \Bigg[& (1-u)\left(\frac{(1-\beta)\beta(1-u)}{u+\beta(1-u)} - \frac{1-\varepsilon_\lambda}{1-\varepsilon_\lambda(1-u)}\right)xy' + (1-u)u\left(\frac{\varepsilon_\lambda}{1-\varepsilon_\lambda(1-u)}y - \frac{(1-\beta)}{u+\beta(1-u)}(y-b)\right)x' \\
& + \frac{u(1-u)}{(1-\varepsilon_\lambda(1-u))^2}\left(\frac{\lambda''}{\lambda'}\theta - \varepsilon_\lambda + 1 - \varepsilon_\lambda(1-\varepsilon_\lambda)\right)xy\frac{\theta'}{\theta} \\
& - \frac{u(1-u)(1-\beta)}{(u+\beta(1-u))^2}(u\varepsilon_\lambda - u - \beta(1-u))x(y-b)\frac{\theta'}{\theta}\Bigg]\frac{L}{N}\,\mathrm{d}t, \qquad \text{(A.19)}
\end{aligned}
$$

where t_f^0 is a constant that ensures the budget balance of the government. All functions $(x, y, u, \varepsilon_\lambda, \theta, L, N)$ are evaluated at the optimal assignment.

Under the Hosios condition and zero unemployment benefit, the transfer simplifies to $t_f(\ell)$ in Corollary 1. On the one hand, firms do not internalize that local firms in higher-ℓ locations are more productive, as are workers. On the other hand, they take into account that higher $y(\ell)$ increases the threat point of workers in wage bargaining, leading to a decrease in their values. The firm spatial transfer confirms that the former is always larger than the latter, so firms choose higher-ℓ than the optimal level.

Note that $t_w'(\ell)$ is zero if and only if $x'(\ell) = 0$, and $t_f'(\ell)$ is zero if and only if $y'(\ell) = 0$. Thus, an equilibrium is efficient and requires no spatial transfers, if and only if workers and firms are homogeneous, $x'(\ell) = y'(\ell) = 0$.

Pareto Inefficiency. In Proposition 3, I show that any pure-assignment PAM equilibrium does not maximize the net output. To further establish its Pareto inefficiency, I consider a planner who maximizes a weighted sum of unemployed workers' value,[39]

$$\max_{x(\ell),y(\ell)} \int_0^1 a(x(\ell))\tilde{\rho}V^u(x(\ell), y(\ell), \ell)L(\ell)\,\mathrm{d}\ell$$

subject to constraints in the original problem, where $a(x)$ denotes a weight for workers of x such that $\int_{\underline{x}}^{\overline{x}} a(x)\,\mathrm{d}Q_w(x) = 1$. By substituting the expression for Π, which includes the profits of landowners, intermediaries, firms, and taxes collected for unemployment benefits, and then rearranging the terms, I decompose the welfare into the net output and additional components to leverage the previous results,

$$\int_0^1 (a(x(\ell)) - 1)\left(bx(\ell) + \frac{\beta\lambda(\ell)}{\delta + \beta\lambda(\ell)}x(\ell)(y(\ell) - b) - \bar{h}r(\ell)\right)L(\ell)\,\mathrm{d}\ell + Y - C,$$

where $Y - C$ is the net output—the objective function of the original problem. I continue to impose $\rho \to 0$. If $a(x) = 1$ for all x, the expression simplifies to $Y - C$, which implies that the utilitarian planner's welfare coincides with net

[39] I focus on the value of unemployed workers to align the planner's problem with the problem of workers who select ℓ that maximizes their unemployment values.

output. For notational simplicity, I omit ℓ from this point onward. The Hamiltonian of this problem is given by

$$\mathcal{H}^P = \mathcal{H} + (a-1)\left(bx + \frac{\beta\lambda}{\delta+\beta\lambda}x(y-b) - \bar{h}r\right)L,$$

where $\mathcal{H}$ is the Hamiltonian of the original problem. The planner's optimal choices satisfy the first-order conditions,

$$0 = \frac{\partial\mathcal{H}}{\partial L} + (a-1)\left(bx + \frac{\beta\lambda}{\delta+\beta\lambda}x(y-b) - \bar{h}r\right), \tag{L}$$

$$\mu_w'(\ell) = -\frac{\partial\mathcal{H}}{\partial x} - (a-1)\left(b + \frac{\beta\lambda}{\delta+\beta\lambda}(y-b)\right)L - a'\left(bx + \frac{\beta\lambda}{\delta+\beta\lambda}x(y-b) - \bar{h}r\right)L, \tag{x}$$

$$0 = \frac{\partial\mathcal{H}}{\partial N}, \tag{N}$$

$$\mu_f'(\ell) = -\frac{\partial\mathcal{H}}{\partial y} - (a-1)\frac{\beta\lambda}{\delta+\beta\lambda}xL. \tag{y}$$

Differentiating (L) and (N) with respect to ℓ, and plugging in (x) and (y), respectively, I obtain the following results,

$$0 = \frac{\partial}{\partial\ell}\frac{\partial\mathcal{H}}{\partial L} + (a-1)x\frac{\partial}{\partial\ell}\left(b + \frac{\beta\lambda}{\delta+\beta\lambda}(y-b)\right) - \frac{\partial}{\partial\ell}((a-1)\bar{h}r),$$
$$0 = \frac{\partial}{\partial\ell}\frac{\partial\mathcal{H}}{\partial N} - (a-1)\frac{\beta\lambda}{\delta+\beta\lambda}x\frac{L}{N}y'.$$

Multiplying the above equations by L and N, respectively, and summing them yields a modified version of (A.12),

$$\bar{h}^2C_r''L'L + C_v''N'N = -a'\bar{h}C_r'L + (a-1)\left(\frac{\beta\lambda\delta}{(\beta\lambda+\delta)^2}\varepsilon_\lambda\frac{\theta'}{\theta}x(y-b) - \bar{h}^2C_r''L'\right)L.$$

In contrast to (A.12), this expression is not necessarily zero. The planner may prefer concentration in high ℓ while reducing congestion in low ℓ, for example, if $a'(x(\ell)) < 0$, favoring low x workers.

To complete the proof, I proceed by contradiction. Suppose that the decentralized equilibrium is Pareto efficient. Observe that the first term in the parenthesis on the right-hand side is equal to the partial derivative of $\tilde{\rho}V^u(x, y(\ell), \ell)$ with respect to changes in market tightness. Substituting the workers' first-order condition, $f_w = 0$, yields

$$-(a(x(\ell))-1)\underbrace{\frac{\partial\tilde{\rho}V^u(x(\ell), y(\ell), \ell)}{\partial y}y'(\ell)}_{>0} - a'(x)\underbrace{x'(\ell)\bar{h}r(\ell)L(\ell)}_{>0} = L(\ell)\bar{h}r'(\ell) + N(\ell)c'(\ell) > 0.$$

From the previous results, the right-hand side is positive in equilibrium. If $a(x(\ell)) - 1 > 0$, the equation can hold only if $a'(x(\ell)) < 0$. Given that $\int_x a(x)\,\mathrm{d}Q_w(x) = 1$, it follows that $a(x) = 1$ for all x. Recall that under $a(x) = 1$, the objective function simplifies to net output, identical to that of the original problem. Therefore, the conditions satisfy (A.12), and the right-hand side should be zero. Contradiction.

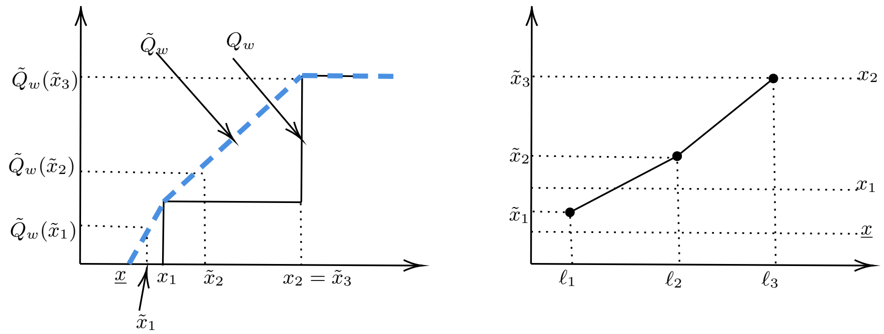


Figure A.1. Population Density in a Finite-Type, Finite-Location Economy

A.5 Population Density

In this section, I consider population density in a discrete economy with a finite number of worker types and locations, and then show that its formula converges to (1) as the numbers of types and locations grow to infinity. In a discrete economy, population density is conventionally defined as a measure of workers per unit of land. Let $\bar{m}(\ell|x)$ be the probability that a type x worker chooses a location less than or equal to ℓ. For example, under a pure assignment $x(\ell)$, $\bar{m}(\ell|x(\ell_0)) = \mathbb{1}\{\ell \geq \ell_0\}$. Note that, unlike the baseline, pure assignments in a discrete economy restrict population density to match the probability mass function of each worker type. I focus on a *weakly* increasing allocation $\bar{m}(\ell|x)$, which means that for any $x' < x''$, either $\bar{m}(\ell|x') > \bar{m}(\ell|x'')$ or $\bar{m}(\ell|x') = \bar{m}(\ell|x'') = 1$.

Step 1 Consider an economy with a finite number of worker types and locations. Consider *a finite number of worker types* $\{x_1, ..., x_M\}$ with the CDF Q_w, with a positive measure for each type. Choose $\underline{x}$ smaller than x_1 with $Q_w(\underline{x}) = 0$. Also, let $\bar{m}(\ell|\underline{x}) = 1$ for all ℓ to simplify the notation.

Consider *a finite number of locations* $\{\ell_1, ..., \ell_N\}$, where $\ell_n = n/N$. Assume that the land distribution is uniform with total measure of one, i.e., the measure of land in each location ℓ_n is given by $\ell_n - \ell_{n-1}$. This assumption is without loss of generality under ex ante homogeneous locations.

I introduce an auxiliary function and a sequence, $\tilde{Q}_w(\cdot)$ and $\{\tilde{x}_n\}_n$. First, define a *(strictly) increasing* $\tilde{Q}_w(\cdot)$ on $[\underline{x}, \bar{x}]$ such that (1) $\tilde{Q}_w(x_i) = Q_w(x_i)$ for all i, and (2) linearly increasing in $x \in (x_{i-1}, x_i)$ for all i. Next, for $\{\tilde{x}_n\}_n$, I first find $j(n) \equiv \max\{j|\bar{m}(\ell_n|x_j) = 1\}$, i.e., the best type among workers who locate only in $\{\ell_1, ..., \ell_n\}$. Then, I define $\tilde{x}_n = x_{j(n)} + \bar{m}(\ell_n|x_{j(n)+1})(x_{j(n)+1} - x_{j(n)})$. Figure A.1 provides an illustrative example of how $Q_w(\cdot)$ and $\{x_n\}_n$ are related to their counterparts, $\tilde{Q}_w(\cdot)$ and $\{\tilde{x}_n\}_n$.

The total measure of workers in $\{\ell_1, ..., \ell_n\}$ equals $\tilde{Q}_w(\tilde{x}_n)$, which comprises $Q_w(x_{j(n)})$, the measure of workers of $x_1, ..., x_{j(n)}$, and $\bar{m}(\ell_n|x_{j(n)+1})(Q_w(x_{j(n)+1}) - Q_w(x_{j(n)}))$, a fraction of the measure of $x_{j(n)+1}$. Thus, $\sum_{k=1}^{n} L(\ell_k) = \tilde{Q}_w(\tilde{x}_n)$, and in turn, population density equals

$$L(\ell_n) = M_w \frac{\tilde{Q}_w(\tilde{x}_n) - \tilde{Q}_w(\tilde{x}_{n-1})}{\ell_n - \ell_{n-1}}.$$

Step 2 Let the number of worker types go to infinity, and the distribution of workers $Q_w \in \mathcal{C}$ is strictly increasing. Then, $\tilde{Q}_w = Q_w$ for all $x \in [\underline{x}, \overline{x}]$. Moreover, I can find the cutoff types $x(\ell_n)$ such that all workers of $x \leq x(\ell_n)$ choose $\ell \leq \ell_n$, and others choose $\ell > \ell_n$. In other words, workers of $x \in (x(\ell_{i-1}), x(\ell_i)]$ sort into location ℓ_i. By the continuity of $Q_w(x)$, $\tilde{x}_n = x_{j(n)} = x(\ell_n)$, and population density becomes

$$L(\ell_n) = M_w \frac{Q_w(x(\ell_n)) - Q_w(x(\ell_{n-1}))}{\ell_n - \ell_{n-1}} = M_w \frac{Q_w(x(\ell_n)) - Q_w(x(\ell_{n-1}))}{x(\ell_n) - x(\ell_{n-1})} \frac{x(\ell_n) - x(\ell_{n-1})}{\ell_n - \ell_{n-1}}.$$

Step 3 Let the number of locations go to infinity, so that ℓ is uniformly distributed on $[0, 1]$. As $\ell_n - \ell_{n-1}$ converges to zero, two terms in $L(\ell_n)$ become derivatives, and the population density in the limit coincides with (1), i.e., $L(\ell) = M_w Q_w'(x(\ell))x'(\ell)$.

A.6 Neoclassical Local Labor Markets: Proof of Proposition 4

All assumptions from the baseline model are maintained, except that the labor market is now frictionless and competitive in each location. Workers of productivity x choose a location ℓ that maximizes $w(x, \ell) - \bar{h}r(\ell)$. Firms of productivity y choose a location ℓ that maximizes profits, $\bar{V}^v(y, \ell) = x(\ell)y - w(x(\ell), \ell) - c(\ell)$. Equilibrium wages $w(x, \ell)$ in ℓ equate the local demand and supply of type-x workers. I focus on a differentiable pure assignment equilibrium $(x(\ell), y(\ell))$, in which an assignment and wages are differentiable.

I first show that an equilibrium exhibits PAM. Consider two locations, ℓ' and ℓ''. The location choice of firms implies that $\bar{V}^v(y(\ell''), \ell'') \geq \bar{V}^v(y(\ell''),' \ell)$ and $\bar{V}^v(y(\ell'), \ell') \geq \bar{V}^v(y(\ell'), \ell'')$. Combining two inequalities yields

$$x(\ell'')(y(\ell'') - y(\ell')) \geq x(\ell')(y(\ell'') - y(\ell')).$$

Thus, $y(\ell'') > y(\ell')$ implies $x(\ell'') > x(\ell')$. Without loss of generality, I assume both $x(\ell)$ and $y(\ell)$ are increasing in ℓ.

I focus on the case in which the total measures of workers and firms are the same, i.e., $M_w = M_f$. First, worker and firm densities in each ℓ should be equal in equilibrium, i.e., $L(\ell) = N(\ell)$. To show by contradiction, assume that there exists ℓ' such that $L(\ell') > N(\ell')$. Then, there must exist another ℓ'' such that $L(\ell'') < N(\ell'')$. If $L(\ell') < L(\ell'')$, then $N(\ell'') > N(\ell')$, and firms in ℓ'', that fail to hire workers, deviate to ℓ', where overhead costs are lower and workers are available. Conversely, if $L(\ell') > L(\ell'')$, then unemployed workers in ℓ' deviate to ℓ'' with lower housing rents and better employment opportunities.

Under $(x(\ell), y(\ell))$, for local labor market clearing, it is optimal for firms of $y(\ell)$ to hire $x(\ell)$,

$$\frac{\partial}{\partial x}(xy(\ell) - w(x, \ell)) = 0 \Rightarrow y(\ell) = \frac{\partial w(x(\ell), \ell)}{\partial x}. \tag{A.20}$$

Moreover, the firm sorting condition reads to

$$\begin{aligned}0 &= \frac{\partial x(\ell)}{\partial \ell} y(\ell) - \left(\frac{\partial w(x(\ell),\ell)}{\partial x} \frac{\partial x(\ell)}{\partial \ell} + \frac{\partial w(x,\ell)}{\partial \ell} \right) - c'(\ell) \\ &= \frac{\partial x(\ell)}{\partial \ell} y(\ell) - \left(y(\ell) \frac{\partial x(\ell)}{\partial \ell} + \bar{h} r'(\ell) \right) - c'(\ell) = -(\bar{h} r'(\ell) + c'(\ell)),\end{aligned}$$

where I use (A.20) and the worker sorting condition, $\frac{\partial w(x(\ell),\ell)}{\partial \ell} = \bar{h} r'(\ell)$, in the second line. As $L(\ell) = N(\ell)$, in equilibrium, the signs of $r'(\ell)$ and $c'(\ell)$ must be the same, and thus $r'(\ell)$ should be zero for all ℓ. This implies that population density is uniform across locations, i.e., $L'(\ell) = 0$.

An equilibrium wage is given by

$$w(x(\ell),\ell) = w(\underline{x},0) + \int_0^{\ell} \left(\frac{\partial w(x,\ell)}{\partial x} \frac{\partial x(\ell)}{\partial \ell} + \frac{\partial w(x,\ell)}{\partial \ell} \right) \mathrm{d}\ell = w(\underline{x},0) + \int_{\underline{x}}^{x(\ell)} y(Q_w^{-1}(x)) \, \mathrm{d}x,$$

where I use the change of variable $x'(\ell)\,\mathrm{d}\ell = \mathrm{d}x$ and $Q_w(x(\ell)) = \ell$. This wage is exactly the same as the wage of the economy with a single integrated labor market. The presence of segregated local labor markets does not affect equilibrium matching, wages, and profits. Given this observation, it follows that the equilibrium is efficient.

Finally, an equilibrium wage $w(x,\ell)$ is unique, and it does not vary across ℓ. For a location ℓ to be the optimal choice for a worker of $x(\ell)$ and a firm of $y(\ell)$, the following conditions must hold.

$$\begin{aligned} w(x(\ell),\ell) - r(\ell) \geq w(x(\ell),\ell') - r(\ell') &\Rightarrow w(x(\ell),\ell) \geq w(x(\ell),\ell') \quad \forall \ell', \\ x(\ell)y(\ell) - w(x(\ell),\ell) - c(\ell) \geq x(\ell)y(\ell) - w(x(\ell),\ell') - c(\ell') &\Rightarrow w(x(\ell),\ell) \leq w(x(\ell),\ell') \quad \forall \ell', \end{aligned}$$

which together imply that $w(x(\ell),\ell) = w(x(\ell),\ell')$ for all ℓ, ℓ'.

A.7 Local Labor Market with Directed Search

In this section, I consider the directed (competitive) search model (e.g., Moen, 1997). In each local labor market, workers and firms engage in directed search following Eeckhout and Kircher (2010). The key distinction from their framework is that workers and firms first choose locations, then engage in search there. Upon choosing locations, both workers and firms demand a unit of housing at local rents $r(\ell)$, prior to entering local labor markets. Unlike the baseline model, I assume that workers and firms share a common local housing market.[40]

I focus on a differentiable pure assignment, $(x(\ell), y(\ell))$. Let $V^u(x, y(\ell), \ell)$ and $V^v(x(\ell), y, \ell)$ denote the values of workers of productivity x and firms of productivity y from participating in the local labor market at location ℓ. For further details on the labor market environment, see Eeckhout and Kircher (2010). The model is static, and if workers

[40] This assumption ensures the efficiency of pure assignment. If not, the planner would prefer a non-pure assignment to equalize worker and firm densities across space and thereby reduce congestion costs, while preserving the match between workers and firms.

or firms fail to match, their value is zero. The value of ℓ to a worker and a firm, $\bar{V}^u(x,\ell)$ and $\bar{V}^v(y,\ell)$, are given by the labor market value net of housing rents,

$$\bar{V}^u(x,\ell) = V^u(x,y(\ell),\ell) - r(\ell) \qquad \text{where } V^u(x,y(\ell),\ell) = \max \lambda(\theta)w,$$
$$\bar{V}^v(y,\ell) = V^v(x(\ell),y,\ell) - r(\ell) \qquad \text{where } V^v(x(\ell),y,\ell) = \max q(\theta)(f(x,y) - w),$$

where θ denotes market tightness, $\lambda(\theta)$ the job arrival rate, $q(\theta)$ the vacancy contact rate, w the wage, and $f(x,y)$ the output.

Workers choose market tightness $\theta(x,\ell)$ subject to $V^v(x(\ell),y(\ell),\ell) = q(\theta(x,\ell))(f(x,y(\ell)) - w(x(\ell),y(\ell),\ell))$. Substituting this condition into workers' labor market value $V^u(x,y(\ell),\ell)$ yields the first-order condition,

$$\max_\theta \lambda(\theta(x,\ell))f(x,y(\ell)) - \theta(x,\ell)V^v(x(\ell),y(\ell),\ell) \Rightarrow \lambda'(\theta(x,\ell))f(x,y(\ell)) = V^v(x(\ell),y(\ell),\ell).$$

Similarly, firms choose $\theta(y,\ell)$ subject to $V^u(x(\ell),y(\ell),\ell) = \lambda(\theta(y,\ell))w(x(\ell),y(\ell),\ell)$. Substituting into the firm's labor market value $V^v(x(\ell),y,\ell)$, gives the following first-order condition,

$$\max_\theta q(\theta)\left(f(x(\ell),y) - \frac{V^u(x(\ell),y(\ell),\ell)}{\lambda(\theta(y,\ell))}\right) \Rightarrow q'(\theta(y,\ell))f(x(\ell),y) + \frac{1}{\theta(y,\ell)^2}V^u(x(\ell),y(\ell),\ell) = 0.$$

The sorting conditions of workers and firms are given by

$$r'(\ell) = \lambda(\theta(x(\ell),\ell))f_y(x(\ell),y(\ell))y'(\ell) - \theta(x(\ell),\ell)\frac{\partial}{\partial \ell}V^v(x(\ell),y(\ell),\ell),$$
$$\theta(y(\ell),\ell)r'(\ell) = \lambda(\theta(y(\ell),\ell))f_x(x(\ell),y(\ell))x'(\ell) - \frac{\partial}{\partial \ell}V^u(x(\ell),y(\ell),\ell).$$

Summing these two expressions yields

$$r'(\ell) + \theta(\ell)r'(\ell) = \lambda(\theta(\ell))(f_x x'(\ell) + f_y y'(\ell)) - (\theta(\ell)(V^v)'(\ell) + (V^u)'(\ell)) = \theta'(\ell)(V^v(\ell) - \lambda'(\theta)f(\ell)) = 0,$$

where I simplify notation by indexing all equilibrium objects by ℓ. The first equality uses the two conditions for the market tightness, which together imply that $\theta(\ell)V^v(\ell) + V^u(\ell) = \lambda(\ell)f(x(\ell),y(\ell))$. Differentiating the both sides with respect to ℓ, I obtain $\theta'(\ell)V^v(\ell) - \lambda'(\theta)\theta'(\ell)f(\ell) = \lambda(\theta(\ell))(f_x x'(\ell) + f_y y'(\ell)) - \theta(\ell)(V^v)'(\ell) - (V^u)'(\ell)$. The second equality follows from the first-order condition for $\theta(x,\ell)$. From the above equation, it follows that $r'(\ell) = 0$, and thus workers and firms are uniformly distributed across space. That is, the equilibrium features no spatial congestion.

Since rents are uniform across space, the problem faced by workers and firms is identical to the one analyzed in Eeckhout and Kircher (2010), where agents match in a nationwide labor market. Accordingly, the same condition guarantees PAM,

$$\frac{f_{xy}f}{f_x f_y} \geq \sup_{\theta} \frac{\lambda'(\theta)(\lambda'(\theta)\theta - \lambda(\theta))}{\theta\lambda(\theta)\lambda''(\theta)}. \tag{A.21}$$

Moreover, the matching between workers and firms remains unchanged, and the efficiency result continues to hold. As in Proposition 4, locations have no economic content even in the presence of search frictions as long as workers (firms) can condition their search on firm (worker) types in each local labor market.

Proposition A.2. *Suppose that* (A.21) *holds. If a differentiable pure assignment equilibrium exists, then the equilibrium exhibits the following properties:*

1. *Positive assortative matching (PAM) between workers and firms arises across space: firm productivity $y(\ell)$ increases in ℓ just like worker productivity $x(\ell)$.*
2. *Workers and firms are uniformly distributed across space.*
3. *The equilibrium is efficient. Moreover, matching between workers and firms, the wage of each worker type, and the values of workers and firms from labor market are equal to those of an economy with a single, nationwide labor market.*

A.8 Model with Additional Sources of Productivity Heterogeneity: Proof of Proposition 5

Derivations. The derivations remain almost the same, except that the output function has an additional component $A(\ell)$. For example, the surplus of a match between a worker $x(\ell)$ and a firm $y(\ell)$ in location ℓ is given by

$$S(x(\ell), y(\ell), \ell) = \frac{1}{\tilde{\rho} + \beta\lambda(\ell) + (1-\tilde{\beta})q(\ell)}(A(\ell)y(\ell) - b)x(\ell),$$

and wages are given by (12). The values of workers and firms choosing a location ℓ are given by

$$\rho V^u(x,\ell) = bx + \frac{\beta\lambda(\ell)}{\tilde{\rho} + \beta\lambda(\ell)}(A(\ell)y(\ell) - b)\left(x - \frac{(1-\tilde{\beta})q(\ell)}{\tilde{\rho} + \beta\lambda + (1-\tilde{\beta})q(\ell)}x(\ell)\right) - \bar{h}r(\ell) + \Pi,$$

$$\rho \bar{V}^v(y,\ell) = \frac{1}{\rho}\frac{(1-\tilde{\beta})q(\ell)}{\tilde{\rho} + (1-\tilde{\beta})q(\ell)}x(\ell)\left(A(\ell)y - b - \frac{\beta\lambda(\ell)}{\tilde{\rho} + \beta\lambda(\ell) + (1-\tilde{\beta})q(\ell)}(A(\ell)y(\ell) - b)\right) - c(\ell).$$

Observe that all the above equations return to those in the baseline when $A(\ell) = 1$.

Matching Cross-Sectional Moments. Given unemployment rates across regions $\{u(\ell)\}$, I infer local market tightness and thus $\{\lambda(\ell), q(\ell)\}$. Furthermore, given population density $L(\ell)$, I can compute $V(\ell)$.

In the *no sorting* model, I calibrate $\bar{A}(\ell)$ such that (12) aligns with the cross-sectional wage data. Then, I recover $r(\ell) = C_r'(\bar{h}L(\ell))$ and $c(\ell) = C_v'(N(\ell))$ that ensure the sorting conditions of workers and firms, which correspond to the equalization of their values across ℓ. The *spillovers* model can be treated analogously by finding $x(\ell)$ rather than $\bar{A}(\ell)$.

Next, consider the *one-sided sorting* model. Suppose that housing rents $r(\ell)$ are observed. I then jointly identify $\{x(\ell), \bar{A}(\ell)\}$ as the solution to (12) and $\frac{\partial}{\partial \ell} V^u(x(\ell), \ell) = 0$. I then determine $c(\ell) = C_v'(N(\ell))$ to satisfy the sorting condition of firms. The *two-sided sorting* model can be calibrated analogously; see Section 4.2 for more discussion.

Planner problem. Focus on positive sorting among workers, firms, and locations, that is, $x(\ell), y(\ell)$ and $A(\ell)$ are increasing in ℓ. The Hamiltonian can be formulated as below,

$$\mathcal{H} = (1-u(\ell))x(\ell)y(\ell)A(\ell)L(\ell) - C_r(\bar{h}L(\ell)) - C_v(N(\ell)) + \frac{\mu_w(\ell)L(\ell)}{M_w q_w(x(\ell))} + \frac{\mu_f(\ell)N(\ell)}{M_f q_f(y(\ell))} + \mu_u(\ell)\left(u(\ell) - \frac{\delta}{\delta + \lambda(\theta(\ell))}\right),$$

where I use the same notation as in Section A.4. First-order conditions are given by

$$0 = (1-u(\ell))x(\ell)y(\ell)A(\ell) - \bar{h}C_r'(\bar{h}L(\ell)) + \frac{\mu_w(\ell)}{M_w q_w(x(\ell))} - \mu_u(\ell)\frac{\varepsilon_\lambda(\ell)}{L(\ell)}u(\ell)(1-u(\ell)), \tag{L}$$

$$\mu_w'(\ell) = -(1-u(\ell))y(\ell)A(\ell)L(\ell) - (1-u(\ell))x(\ell)y(\ell)\frac{\partial A(\ell)}{\partial x}L(\ell) + \frac{\mu_w(\ell)L(\ell)}{M_w q_w^2(x(\ell))}q_w'(x(\ell)), \tag{x}$$

$$0 = -C_v'(N(\ell)) + \frac{\mu_f(\ell)}{M_f q_f(y(\ell))} + \mu_u(\ell)\frac{\varepsilon_\lambda(\ell)}{N(\ell)}u(\ell)(1-u(\ell)), \tag{N}$$

$$\mu_f'(\ell) = -(1-u(\ell))x(\ell)A(\ell)L(\ell) + \frac{\mu_f(\ell)N(\ell)}{M_f q_f^2(y(\ell))}q_f'(y(\ell)), \tag{y}$$

$$0 = -x(\ell)y(\ell)A(\ell)L(\ell) + \mu_u(\ell)(1-\varepsilon_\lambda(\ell)(1-u(\ell))). \tag{u}$$

From now on, I omit ℓ for notational simplicity. Following the same step as in Section A.4, I obtain the below conditions:

$$\bar{h}^2 C_r'' L' = \frac{(1-\varepsilon_\lambda)(1-u)}{1-\varepsilon_\lambda(1-u)}x(y\bar{A})'A^x - \frac{\varepsilon_\lambda(1-u)u}{1-\varepsilon_\lambda(1-u)}\frac{\partial xyA}{\partial x}x' + \frac{u(1-u)xyA}{(1-\varepsilon_\lambda(1-u))^2}\left((1-\varepsilon_\lambda)\varepsilon_\lambda - \frac{\lambda''}{\lambda'}\theta + \varepsilon_\lambda - 1\right)\frac{\theta'}{\theta},$$

$$C_v''(N)N' = -\frac{(1-\varepsilon_\lambda)(1-u)}{1-\varepsilon_\lambda(1-u)}xy'A\frac{L}{N} + \frac{\varepsilon_\lambda(1-u)u}{1-\varepsilon_\lambda(1-u)}\frac{L}{N}(Ax)'y - \frac{u(1-u)xyA}{(1-\varepsilon_\lambda(1-u))^2}\left((1-\varepsilon_\lambda)\varepsilon_\lambda - \frac{\lambda''}{\lambda'}\theta + \varepsilon_\lambda - 1\right)\frac{L}{N}\frac{\theta'}{\theta}.$$

Combining the two conditions, I obtain the first equation that characterizes the optimal allocation,

$$(1-u(\ell))x(\ell)y(\ell)\bar{A}'(\ell)A^x(x(\ell))L(\ell) = C_r''(\bar{h}L(\ell))\bar{h}^2 L'(\ell)L(\ell) + C_v''(N(\ell))N'(\ell)N(\ell). \tag{A.22}$$

The planner allocates more workers and firms to higher-ℓ locations with higher exogenous location productivity $\bar{A}(\ell)$. In contrast, other factors, i.e., heterogeneity of workers and firms as well as agglomeration forces, do not depend on the location of production, and hence, the planner spreads out workers and firms to avoid unnecessary congestion. Next, the second equation is given by

$$\left(\Box + C_r''\bar{h}^2 L\right)\frac{L'}{L} - \left(\Box + C_v'' N\frac{N}{L}\right)\frac{N'}{N} = \frac{1-u}{1-\varepsilon_\lambda(1-u)}\left((1-\varepsilon_\lambda)xy'A - \varepsilon_\lambda u\left(yA + \frac{\partial A}{\partial x}\right)x'\right),$$

where $\Box = \frac{u(1-u)}{(1-\varepsilon_\lambda(1-u))^3}\left(\varepsilon_\lambda(1-\varepsilon_\lambda) - \frac{\lambda''}{\lambda'}\theta + \varepsilon_\lambda - 1\right)xyA$. The baseline intuition remains the same; the planner increases L relatively more than N when firm heterogeneity y' is significantly larger than that of workers x'.

Efficiency of the equilibrium. Derivations of the sorting conditions are explained in Section A.4. To avoid repetition, I omit the intermediate steps and present the final expressions for ρ close to zero,

$$\begin{aligned} L\bar{h}r' &= \frac{\beta\lambda}{\delta+\beta\lambda}x(Ay)'L + \frac{\delta\beta\lambda}{(\delta+\beta\lambda)^2}\varepsilon_\lambda\frac{\theta'}{\theta}xA(y-b)L, \\ Nc' &= \frac{N(1-\beta)q}{\delta+\beta\lambda}(xA)'y - \frac{\beta\lambda}{\delta+\beta\lambda}(1-\beta)(1-u)Axy'L + \frac{\delta(1-u)(1-\beta)}{\delta+\beta\lambda}\left(\frac{\delta}{\delta+\beta\lambda}\varepsilon_\lambda - 1\right)\frac{\theta'}{\theta}Ax(y-b)L. \end{aligned}$$

To focus on the externalities arising from the sorting mechanism, I assume the Hosios condition $\varepsilon_\lambda = 1-\beta$ and zero unemployment rate $b = 0$. Comparing the sorting conditions and the planner's solution yields the spatial transfers:

$$\begin{aligned} t_w(\ell) &= t_w^0 - \int_0^\ell \left(\frac{\varepsilon_\lambda(1-u)u}{1-\varepsilon_\lambda(1-u)}Ay + (1-u)\frac{\partial A^x}{\partial x}\bar{A}y\right)x'\,\mathrm{d}t, \\ t_f(\ell) &= t_f^0 - \int_0^\ell (1-\varepsilon_\lambda)(1-u)\frac{L}{N}Axy'\,\mathrm{d}t. \end{aligned}$$

where t_w^0 and t_f^0 are constants that ensure the government budget balance. All functions are evaluated at the optimal assignment. These transfers are consistent with those derived in Section 3.2. In addition to $t_w(\ell)$ in the baseline, the second term in parentheses corrects externalities of workers on local TFP through spillovers. In addition, the transfers not only depend on x and y, but also on A.

B. Quantitative Analysis

B.1 Quantitative Model

Values. With a Stone-Geary utility function, housing regulations, and income taxes, the value of unemployed and employed workers changes to

$$\rho V^u(x,\ell) = r(\ell)^{-\omega}((1-\tau_w(\ell))bx - \bar{h}r(\ell) + \Pi + T_r(\ell)) + \lambda(\ell)(V^e(x,y(\ell),\ell) - V^u(x,\ell)),$$

$$\rho V^e(x,y,\ell) = r(\ell)^{-\omega}((1-\tau_w(\ell))w(x,y,\ell) - \bar{h}r(\ell) + \Pi + T_r(\ell)) + \delta(V^u(x,\ell) - V^e(x,y,\ell)),$$

where $\tau_w(\ell)$ denotes the income tax on labor income and $T_r(\ell)$ denotes locally redistributed taxes on housing markets. The value of firms remains the same.

Rubinstein bargaining and wages. I show that the bargaining solution retains a simple form. Consider a parallel time for bargaining between a worker and a firm, where the firm makes the first offer. Firm's flow surplus is given by $v_f = \tilde{\rho}(V^p(x,y,\ell) - V^v(y,\ell))$, and worker's surplus is given by $v_w = \tilde{\rho}(V^e(x,y,\ell) - V^u(x,\ell))$. Let workers and firms have discount factors δ_w and δ_f, respectively.

To apply Proposition 122.1 in Osborne and Rubinstein (1994), I verify that the four required conditions are satisfied. The first three—no redundancy, the indifference when the opponent enjoys the best agreement, and monotone Pareto frontier—are straightforwardly met, assuming that workers discount their values by $r^{-\omega}(1-\tau_w)$. The final assumption is satisfied if the Pareto frontier of the set of agreements can be written as $\{v \in \mathbb{R}^2 : v_w = g(v_f)\}$ for some decreasing, concave function $g(\cdot)$, and each player's surplus is represented by $\delta_i^t v_i$ for some $0 < \delta_i < 1$ for $i = w, f$. In this setting, their surpluses take the form $v_f = A - w$ and $v_w = B + r^{-\omega}(1-\tau_w)w$ where $A = xy - q(V^p - V^v) - \delta_v V^v$ and $B = -r^{-\omega}(1-\tau_w)bx - \lambda(V^e - V^u)$. It follows that $v_w = g(v_f) = B + r^{-\omega}(1-\tau_w)(A - v_f)$, which is decreasing and weakly concave in v_f.

Thus, a subgame perfect equilibrium is characterized by v_f^* and v_f' such that $\delta_f v_f^* = v_f'$ and $g(v_f^*) = \delta_w g(v_f')$. Defining $\beta = \frac{1-\delta_f}{1-\delta_f\delta_w}$, the bargaining solution is given by

$$v_f = (1-\beta)(xy - bx - r^{\omega}\lambda(V^e - V^u) - q(V^p - V^v) + \delta_v V^v),$$
$$v_w = r^{-\omega}(1-\tau_w)\beta(xy - bx - r^{\omega}\lambda(V^e - V^u) - q(V^p - V^v) + \delta_v V^v).$$

This outcome is equivalent to the Nash bargaining solution where the surplus, $S = r^{\omega}(1-\tau_w)^{-1}(V^e - V^u) + V^p - V^v$, is split between a worker and a firm with worker's bargaining power β.

Housing markets. Saiz (2010) quantifies the impact of housing regulations on housing supply elasticities by running the following MSA-level regression:

$$\gamma(\ell) = \gamma + 0.28\log(3 + \text{WRI}(\ell)) + \Gamma_\gamma X_\gamma(\ell) + \varepsilon(\ell), \tag{A.23}$$

where $\gamma(\ell)$ denotes the inverse of the price elasticity $\eta_w(\ell)$ and $\text{WRI}(\ell)$ denotes the local regulation index. To borrow his estimates, I microfound the above relation by introducing taxes on housing production costs, $C_r(\cdot)$; tax rates $\tau(\cdot)$ depend on local housing supply H and the stringency of housing regulations $t_h(\ell)$, specifically, $\tau(H;\ell) = (H/T)^{t_h(\ell)} - 1$. The constant T is a scaling parameter chosen so that τ is unit-free.[41]

[41] Also, the term H/T being larger than one ensures that housing rents increase in tax rate.

The total cost of housing provision including taxes equals $(1+\tau(H;\ell))C_r(H) = (1+\tau(H;\ell))\frac{H^{1+\gamma}}{(1+\gamma)H_w^{\gamma}} = \frac{1}{1+\gamma}\frac{1}{H_w^{\gamma}}\frac{1}{T^{t_h(\ell)}}H^{1+\gamma+t_h(\ell)}$, so corresponding housing rents are given by

$$r(H,\ell) = \frac{1+\gamma+t_h(\ell)}{1+\gamma}\left(\frac{H}{H_w}\right)^{\gamma}\left(\frac{H}{T}\right)^{t_h(\ell)}. \tag{A.24}$$

This functional form is consistent with (A.23), as the elasticity of housing rents with respect to housing supply H is $\gamma + t_h(\ell)$. That is, more stringent regulations, i.e., higher $t_h(\ell)$, lead to a higher price elasticity. I normalize tax rates such that $\min_\ell t_h(\ell)$ and $\min_\ell \tau(H;\ell)$ are zero. In other words, $t_h(\ell) = 0.28[\log(3+\text{WRI}(\ell)) - \log(3+\text{WRI}(\underline{\ell}))]$. With some rearrangement, the housing supply is given by

$$H(r(\ell);\ell) = H_w(\ell)r(\ell)^{\frac{1}{\gamma+t_h(\ell)}} \quad \text{where } H_w(\ell) = \left(\frac{1+\gamma}{1+\gamma+t_h(\ell)}\right)^{\frac{1}{\gamma+t_h(\ell)}}\left(H_w^{\gamma}T^{t_h(\ell)}\right)^{\frac{1}{\gamma+t_h(\ell)}}.$$

Finally, housing rents are pinned down by the housing market clearing condition,

$$r(\ell)H(r(\ell);\ell) = (1-\omega)\bar{h}L(\ell) + \omega((1-\tau_w(\ell))(u(\ell)bx(\ell) + (1-u(\ell))w(x(\ell),y(\ell),\ell)) + \Pi + T_r(\ell))L(\ell).$$

B.2 Data

Locations. I base my analysis on metropolitan statistical areas (MSAs). I obtain the population density of the year 2010 from the U.S. Census and local GDP per capita from the Bureau of Economic Analysis (BEA). It is well known that population density across MSAs remains relatively stable over time.

Wages. I use *incwage* (nominal, wage and salary income) provided by the ACS from IPUMS for the year 2017 (Ruggles et al., 2023). I retain workers between the ages of 26 and 59 and exclude individuals employed in military occupations. I use sampling weights (*perwt*) to account for the survey sampling design. I first residualize log nominal wages controlling for age, sex, race (4 groups), and 1-digit industry (5 groups). I do not control for occupation as it is not an inherent characteristic of individuals. Then, I compute the local average wages using the residualized values.

Separation rate. I use IPUMS-CPS from 2016 to 2019 (Flood et al., 2022). I keep workers of ages between 26 and 59. I calculate the separation rate δ as the number of workers unemployed for 5 weeks or less divided by employment one month earlier to avoid the time aggregation issue (Shimer, 2012).

Housing markets. I run a hedonic regression of log housing rents on attributes of buildings (number of rooms, built year, the number of housing units in the structure) to compute the residualized housing rents from ACS (2017) for each individual and compute the average. I compute average housing spending shares across space using raw labor income and housing rents. Next, I target the average housing spending share in the total consumption expenditure of 33% (Consumer Expenditure Survey, 2017).

Federal income tax. From the March CPS, I compute the average tax rates for each MSA (2017, 2019). The Tax Cuts and Jobs Acts was passed at the end of December in 2017, with most of the changes taking effect in January 2018. Thus, I compare tax rates for the years 2017 and 2019 using adjusted gross income (*adjginc*) and federal income taxation (*fedtaxac*) in IPUMS-CPS, which are calculated by the Census Bureau's tax model and added to the data. I regress the log disposable income rates, i.e., after-tax income over income, on the log population density to compuite location-specific tax rates $\tau_w(\ell)$. A more commonly used approach is to regress the log disposable income rates on the log income (e.g., Heathcote et al., 2017). The estimates for the year 2017 are $\log(\text{disp. income rates}) = -0.072 - 0.0061 \log(\text{pop. density})$. To estimate the same regression for the year 2019, I first predict the after-tax income of the year 2017 under the 2019 tax rates to take into account changes in income distribution. In particular, I first approximate the log after-tax income in 2019 as a function of the log of income. I then predict the after-tax income for each individual using these estimates and the income in 2017. The regression yields $\log(\text{disp. income rates}) = -0.071 - 0.0026 \log(\text{pop. density})$.

Housing market regulations. I borrow estimates on differential housing supply elasticities across MSAs from Saiz (2010). He estimates how housing elasticities are affected by the Whorton Residential Urban Land Regulation Index (WRI), which is a measure of housing regulation developed by Gyourko et al. (2008).

B.3 Estimation: Intuition for Identification

Housing markets. Preferences for housing $(\bar{h}, \omega)$ can be obtained from housing spending shares. The housing spending of workers with income I equals $(1-\omega)r(\ell)\bar{h} + \omega I$. The local housing spending share increases faster in ℓ if the parameter $\bar{h}$ is larger, since housing rents rise in ℓ more rapidly than income in the data. In contrast, the parameter ω uniformly increases housing spending shares of all regions. I target the ratio of spending shares between the first and last quartile and the average housing spending share in the U.S.

Next, the housing supply elasticity, η_w, and the housing tax parameter, T, are informed by housing rents. Equation (A.24) reveals that a smaller η_w leads to a more pronounced increase in housing rents in response to higher housing demand, while a lower T magnifies the effect of regulatory changes on housing rents. I target the ratio of housing rents between the first quartile and three other quartiles to jointly determine (η_w, T).

Labor market. I pin down the unemployment benefit parameter, b, by targeting the average replacement rate of 50% (Landais et al., 2018), which is the ratio between the income of unemployed workers $b\,\mathbb{E}[x]$ and the average wage in my model. Next, I obtain the bargaining power of workers, β, from the labor share.

Business services market. I obtain business services supply function parameters, (H_f, η_f), using the sorting conditions of firms. The value of firms reveals how firms balance gains from labor market, such as higher worker productivity, with higher overhead costs. Given the distribution of productivity across workers and firms, I recover the marginal cost function $c'(\ell)$ implied by the firm sorting condition in (8). The elasticity η_f of the supply of business services is then inferred from the responsiveness of $c(\ell)$ to variation in $N(\ell)$. The scale parameter H_f, capturing the overall level of $c(\ell)$, is increasing in the dispersion of worker heterogeneity across ℓ.

B.4 Estimation Result

Figure A.2a shows the fit of population densities and unemployment rates across ℓ. In addition, in Figure A.2b, I present the fit of log GDP per capita, which is non-targeted. Overall, the model replicates how fast GDP per capita increases in population density.

To quantify the contribution of worker heterogeneity in explaining spatial wage differentials, I regress $\log x(\ell)$ and $\log(w(\ell)/x(\ell))$ on log population density and obtain coefficients 0.075 and 0.05, respectively, which implies that about 60% of spatial wage inequality is attributable to worker heterogeneity.

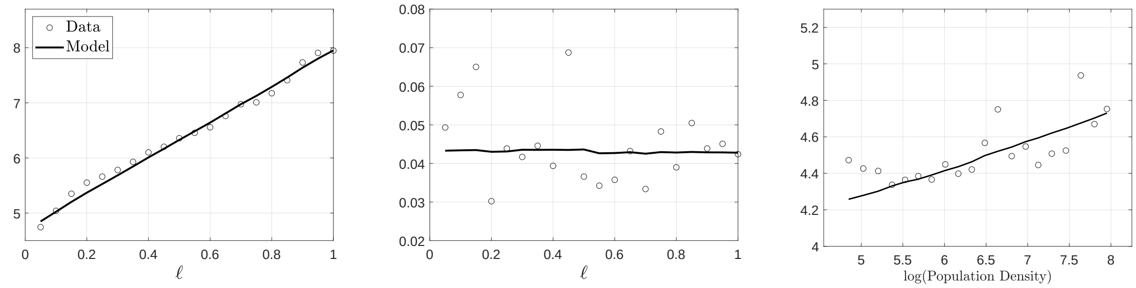
These results may be compared to those in Card et al. (2023), who estimate a reduced-form wage equation with additive worker and CZ fixed effects using matched employer-employee data. Under the mapping implied by (10), their estimates correspond to $\log x(\ell)$ and $\log w(\ell)/x(\ell)$ in this paper. Regressing worker and CZ fixed effects on log population density, they obtain coefficients of 0.04 and 0.034, which suggests that workers account for about 54% of spatial wage variation.

Despite the methodological differences—structural in this paper versus reduced form—the overall conclusion is consistent—60% versus 54%. Differences in magnitude may reflect variations in geographical units, time coverage, or control variables. They may also reflect other factors such as amenities, which influence worker sorting but are abstracted from in this paper.

B.5 Model Estimation using German Data

In this section, I calibrate the model using German cross-sectional data, following the steps described in Section 4.2. This exercise serves two purposes. First, it demonstrates the quantitative relevance of the two-sided sorting mechanism beyond the U.S. context. Second, it lays the groundwork for the analysis in Appendix C, where I provide suggestive evidence on the quantitative significance of two-sided sorting.

Data. I primarily rely on regional statistics from the German Federal Statistical Office (GFSO). I obtain the district-level data for the year 2017: average wages, unemployment rates, GDP per capita, and population densities. Average wages are computed based on the compensation of employees, which includes gross wages and salaries as well as



(a) Population Density (Left) and Unemployment Rates (Right) (b) log GDP pc

Figure A.2. Model Fit: (a) Targeted Moments and (b) Non-Targeted Moments

Notes: Data source: U.S. Census (2010), ACS (2017), and BEA (2017). Each dot represents 5% of the population. I compute average values for each dot by weighting each MSA with its population.

employers' social contributions, adjusted for total hours worked. My analysis is based on 257 commuting zones (CZs, *Arbeitsmarktregionen*). I aggregate all variables to the CZ level using a crosswalk provided by the Federal Office for Building and Regional Planning of Germany. The average spending share on housing is obtained from the rent-to-income of main tenant households (GFSO). Housing spending shares for a set of selected CZs are sourced from Lebuhn et al. (2017), and housing rents across CZs are obtained from Ahlfeldt et al. (2023). Finally, I obtain a replacement rate from the Out-of-Work Benefits Dataset provided as part of the Social Policy Indicator (SPIN) database (Nelson et al., 2020).

Estimation. I order CZs based on the log of population densities and average wages, both of which increase in ℓ in my model, as shown in Proposition 1. While these variables are generally correlated across Germany, the relationship weakens among the most densely populated CZs. Therefore, instead of relying on a single measure, I construct an index that combines the two using principal component analysis (PCA). The resulting index is highly correlated with each variable, with pairwise correlations of 0.87.

Estimation results. I summarize the estimation results in Table A.1. The fit of targeted moments is documented in Table A.2, and that of $\{L(\ell), u(\ell)\}_\ell$ is plotted in Figure A.3. Overall, the model successfully matches the targeted moments. In particular, it closely matches the observed magnitude of spatial disparities in wages and population density.

The heterogeneity of workers and firms across space is substantial. Workers in the top 10% of cities are 18.9% more productive than those in the bottom 10%, and the corresponding figure for firms is 22.9%. Worker and firm heterogeneity account for about 57.0% and 42.5% of spatial wage inequality, respectively.

Table A.1: Parameter Values

Parameter		Target	Value
Discount rate	ρ	Interest rate	0.004
Matching elasticity	α	Literature	0.5
Separation rate	δ	EU transition rate (LIAB)	0.0178
Matching efficiency	$\mathcal{A}$	Market tightness	0.49
Housing demand	$\bar{h}, \omega$	Spending shares on housing	0.80, 0.01
Housing supply	η_w	Housing rents	9.13
Unemployment benefit	b	Replacement rate	0.35
Worker's bargaining power	β	Labor share	0.05
Business services supply	H_f, η_f	Wages	6.16, 12.36

Notes: The top panel shows parameters that are externally calibrated, and the bottom panel shows parameters that are internally calibrated. The separation rate δ is calibrated to match the observed monthly transition probability from employment to unemployment, using data from LIAB. In addition, worker and firm productivity $\{x(\ell), y(\ell)\}_\ell$ are calibrated.

Table A.2: Model Fit

	Wage			Housing		Replac. rate	Labor share	Rent		
Quartile	2	3	4	Mean	Diff.			2	3	4
target $\hat{m}$	0.109	0.186	0.280	0.272	0.034	0.600	0.630	0.180	0.212	0.381
model $m(\Theta)$	0.109	0.184	0.273	0.313	0.023	0.542	0.585	0.129	0.257	0.409

Notes: For wages and housing rents, I first compute the average for each four quartile group. Then, for $i = 2, 3, 4$, I target $(\text{avg}_i/\text{avg}_1) - 1$, where avg_i denotes the average for the i-th quartile group. For variation in housing spending shares, I target the difference between those in the top and the bottom 50%. I split CZs into two groups, rather than four, because I only have observations of 18 CZs in total.

B.6 Policy Evaluation

One-sided sorting model. I estimate the model while keeping $\{x(\ell), \lambda(\ell), q(\ell), L(\ell), N(\ell), r(\ell)\}_\ell$ the same. I set $\bar{A}(\ell)$ equal to $y(\ell)$ from the two-sided sorting model, which ensures that the sorting conditions of workers hold. More importantly, the two models produce identical cross-sectional wages. In contrast, overhead costs $c^{\text{one}}(\ell)$ are calibrated differently to ensure that *homogeneous* firms are indifferent across ℓ,

$$\rho(\bar{V}^p)^{\text{one}} = \frac{(1-\beta)q(\ell)}{\tilde{\rho} + \beta\lambda(\ell)} x(\ell)(\bar{A}(\ell) - b) - c^{\text{one}}(\ell).$$

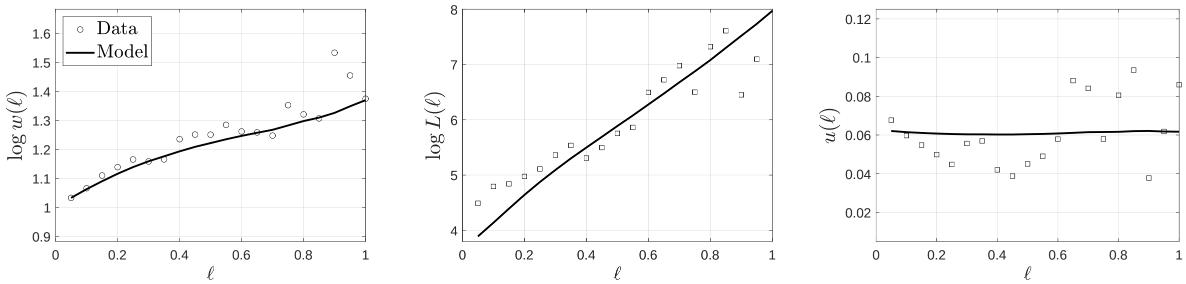


Figure A.3. Model Fit: Wage (Left), Population Density (Middle), Unemployment Rates (Right)

Notes: In the left panel, I plot the average wages from the data and the model. Three moments summarized in Table A.2 are targeted. In the middle and the right panels, I plot the scatter plots of population density and unemployment rates computed from the data, all of which are targeted, and the counterpart from the model. Each dot represents 5% of the population. I compute average values for each dot by weighting each CZs with its population.

Given $\{c^{\text{one}}(\ell)\}$, I recover the marginal cost function from $c^{\text{one}}(\ell) = C_v'(N(\ell))$, with a normalization of setting the average firm profit to zero. Moreover, I choose $C_v(0)$ to match the total profit of intermediaries in the two-sided model, ensuring that Π remains the same.

Federal income tax. Even policies that are common across space can interact with spatial disparities and thereby affect location decisions. One example is the federal income tax, which imposes higher tax rates on individuals with higher nominal incomes. Albouy (2009) argues that the federal income tax reduces employment in high-wage areas and generates inefficiencies. In the context of the two-sided sorting model, however, this tax can act similarly to the optimal spatial transfers, which effectively taxes workers in dense, high-wage cities.

In this section, I examine the impact of the federal income tax cuts motivated by the Tax Cuts and Jobs Act enacted in December 2017. Specifically, I consider the tax reform illustrated in the left panel of Figure A.4. The reform raises average after-tax income shares and reduces their spatial variation.

This tax reform relocates workers and firms in the same direction as the housing regulation reform, producing qualitatively similar effects. Both population and firm densities in the top 10% of regions increase by 25%. Despite this large-scale relocation, the change in aggregate output is marginal, increasing by only 0.004%. Yet, congestion costs increase, and the aggregate welfare of the utilitarian social planner *decreases* by 0.228%. The decomposition of welfare changes also yields qualitatively similar patterns. Housing markets remain the primary source of welfare losses, while firm productivity and market tightness have minimal roles.

I then contrast the results with those from the *one-sided sorting* model. The two models produce sharply different implications. As workers and firms migrate toward productive locations, aggregate output increase by 0.388%, which leads to a 0.239% rise in aggregate welfare. In sum, this exercise confirms that the two-sided and one-sided sorting provide qualitatively different policy implications.

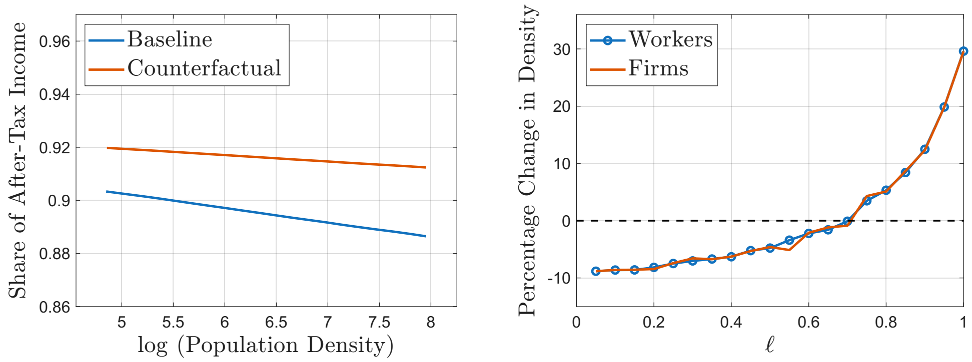


Figure A.4. Federal Income Tax Across Regions

C. Empirical Evidence of Two-Sided Sorting

C.1 Estimation of Worker and Firm Productivity

As in estimation in Section 4.2, I continue to assume $\frac{\rho}{\rho+\delta_v} = 0$. Equilibrium wages (10) implies

$$\log \hat{x}(\ell) = \log x(\ell), \ \log \hat{y}(\ell) = \log\left(b + \beta \frac{\tilde{\rho} + \lambda(\ell)}{\tilde{\rho} + \beta\lambda(\ell)} (y(\ell) - b) \right),$$

where $\log \hat{x}$ and $\log \hat{y}$ denote worker and firm fixed effects, respectively.[42] The job arrival rate $\lambda(\ell)$ can be directly computed from the data, and $\tilde{\rho}$ is externally calibrated. I calibrate the worker's bargaining power β and the unemployment benefit b by matching two moments: the replacement rate and the aggregate labor share. This approach recovers worker and firm productivity without estimating the full structural model.

I first calibrate b by targeting the replacement rate, which is the ratio of average unemployment benefits $\mathbb{E}[bx]$ to the average wage $\mathbb{E}[w(x, y, \ell)]$. This ratio can be expressed in terms of worker and firm fixed effects,

$$\text{Replacement rate} = \frac{\mathbb{E}[bx]}{\mathbb{E}[w(x, y, \ell)]} = \frac{b\,\mathbb{E}[\hat{x}]}{\mathbb{E}[\hat{x}\hat{y}|\text{ employed}]}.$$

Targeting a value of 0.6, as reported in the SPIN (Social Policy Indicator) database, I obtain $b = 0.65$. Given this value, and conditional on β, firm productivity $y(\ell)$ can be inferred from the above equation of $\log \hat{y}(\ell)$. This in turn allows me to compute the labor share,

$$\text{Labor share} = \frac{\mathbb{E}[w(x, y, \ell)]}{\mathbb{E}[xy]} = \frac{\mathbb{E}[\hat{x}\hat{y}]}{\mathbb{E}[\hat{x}y]}.$$

I find that $\beta = 0.068$ rationalizes a labor share of 0.63 (FRED, 2017).

[42] Two-way fixed effects identify only relative productivity differences. I normalize the averages fixed effects to zero. This normalization is without loss of generality, as the model is scale-invariant.

The variation in local worker and firm productivity across regions is substantial; the standard deviations of the average log productivity are 0.06 and 0.12 for workers and firms, respectively. Higher firm fixed effects indicate higher firm productivity, with a correlation of 0.96. The small discrepancy reflects heterogeneity in job arrival rate.

C.2 Additional Results

I provide two additional results that complement the findings in Section 5. First, I take a reduced-form approach by using changes in firm fixed effects of new jobs, $\Delta \log \hat{y}(\ell)$, instead of $\Delta \log y(\ell)$. The results are presented in Table A.3, with each column following the same specification as in Table 3. Focusing on Column (5), a one standard deviation increase in the quality of the local workforce results in a 0.6 standard deviation increase in log firm productivity in the same location.[43] This finding aligns with the conclusion in Table 3. Moreover, I obtain a similar result when using the relative changes in firm fixed effects of new jobs versus existing jobs, i.e., $\Delta \log \hat{y}(\ell) - \Delta \log \hat{y}^{\text{old}}$, as the dependent variable, instead of including $\Delta \log \hat{y}^{\text{old}}(\ell)$ as a control. It is reassuring that the coefficient decreases only modestly to 0.632 with a standard error of 0.238.

Second, I conduct a robustness check by regressing changes in the fixed effects of *existing* jobs, which are jobs operated in both periods, $\Delta \log \hat{y}^{\text{old}}(\ell)$. If an increase in worker productivity were to raise the productivity of both new and existing jobs equally, it would suggest that the observed effects are primarily driven by location-specific factors. I present the results in Table A.4, following the same specifications except that Column (5) is omitted. None of the coefficients is statistically significant, and all are close to zero. This stark contrast with the findings in Table A.3 indicates that firm sorting is the main channel behind the earlier results. Going a step further, the near-zero coefficients indicate that instrumented changes in worker productivity are uncorrelated with location-specific shocks, thereby confirming the validity of my instrument.

C.3 Quantitative Importance of Two-sided Sorting

In this section, I provide two pieces of evidence suggesting that the two-sided sorting mechanism is quantitatively important.

Worker and firm heterogeneity. I first compare the calibrated heterogeneity of workers and firms in the two-sided sorting model to that observed in the data by examining worker and firm fixed effects estimated using two different approaches. In Figure A.5, I plot a binned scatter and a fitted line of estimated worker and firm fixed effects from a log wage regression, alongside their counterparts implied by the calibrated model in Section B.5. The two estimation

[43] To reach this conclusion, I compute the variation in firm fixed effects arising from differences in firm productivity, by first controlling for the local job arrival rate. Based on the residualized firm fixed effects, I then compute the dispersion of firm productivity. For the counterpart of workers, I use worker fixed effects without any controls as guided by the theory.

Table A.3: The Response of Firm Sorting to Changes in Worker Sorting

	OLS (1)	IV (2)	IV (3)	IV (4)	IV (5)
$\Delta \log x(\ell)$	0.449	0.834	0.765	0.797	0.766
	(0.090)	(0.269)	(0.219)	(0.501)	(0.233)
$\Delta \log \hat{y}^{\text{old}}(\ell)$					0.438
					(0.136)
Industry controls			✓		
Geography controls				✓	
		2SLS FIRST-STAGE ESTIMATES			
$\Delta \log x^{\text{IV}}(\ell)$		0.606	0.619	0.677	0.604
		(0.185)	(0.181)	(0.342)	(0.183)

Notes: $N = 257$. Robust standard errors are shown in parentheses. Columns (2)-(5) in the top panel report second-stage estimates, with the dependent variable equal to the change in firm fixed effects, between 2003-2009 and 2010-2016. All regressions include firm fixed effects and unemployment rates in the first period. First-stage regressions in the bottom panel include the same controls as the corresponding second-stage specification in the top panel. Each observation is weighted by the number of workers. In Column (3), firm fixed effects are residualized with respect to one-digit industry fixed effects. In Column (4), flows to or from locations within 100 km are excluded when constructing $\Delta \log x(\ell)$ and its instrument. In Column (5), the specification additionally controls for the average change in fixed effects of jobs matched in both periods.

approaches rely on different identification strategies: the first exploits movers and changes in their wages, while the second is based on the sorting condition of workers and spatial differences in wages and housing rents.

Overall, the estimated fixed effects provided in the dataset (dashed black lines) and the model predictions (solid blue lines) are comparable. In particular, the model performs well in capturing the role of worker heterogeneity in explaining wage disparities, as evidenced by the similarity in the slopes of the two lines in the left panel.

Comparing reduced-form evidence to the model. To provide suggestive evidence that two-sided sorting is a major source of spatial disparity, I compare the model-implied worker-firm interaction with its empirical counterparts. I use a calibrated model in Section B.5 and compute the model counterpart of the estimation result in Section 5.

Specifically, I randomly perturb the spatial distribution of workers. Given this perturbed worker distribution, I solve for the location decisions of firms, which yields a new distribution of firms across space that reflects an endogenous response in firm sorting. By comparing the two distributions, I obtain the changes in worker productivity, firm productivity, and job arrival rates across space. Using this simulated data, I estimate the same regression as in (14) to obtain the model-implied coefficient. I repeat this simulation 100 times and compute the average estimate.[44]

[44] Another potential approach, which may be a promising direction for future research, is to develop a fully dynamic model that incorporates additional sources of productivity heterogeneity and to estimate this model by matching the results in Table 3. The

Table A.4: The Changes in the Productivity of Existing Jobs to Changes in Worker Sorting

	OLS (1)	IV (2)	IV (3)	IV (4)
$\Delta \log x(\ell)$	0.012	0.155	0.094	-0.027
	(0.040)	(0.132)	(0.128)	(0.195)
Industry controls			✓	
Geography controls				✓
	2SLS FIRST-STAGE ESTIMATES			
$\Delta \log x^{\mathrm{IV}}(\ell)$		0.606	0.619	0.677
		(0.185)	(0.181)	(0.342)

Notes: $N = 257$. Robust standard errors are shown in parentheses. Columns (2)-(5) in the top panel report second-stage estimates, with the dependent variable equal to the change in firm fixed effects of existing jobs, between 2003-2009 and 2010-2016. All regressions include firm fixed effects and unemployment rates in the first period. First-stage regressions in the bottom panel include the same controls as the corresponding second-stage specification in the top panel. Each observation is weighted by the number of workers. In Column (3), firm fixed effects are residualized with respect to one-digit industry fixed effects. In Column (4), flows to or from locations within 100 km are excluded when constructing $\Delta \log x(\ell)$ and its instrument.

On average, I find that a one standard deviation increase in log worker productivity leads to a 0.91 standard deviation increase in log firm productivity. This result suggests that firm sorting observed in the data accounts for approximately 60% of the magnitude predicted by the model, which supports the quantitative importance of two-sided sorting in explaining spatial disparities. The remaining 40% may stem from alternative mechanisms arising from local TFP heterogeneity or agglomeration forces. Another explanation is that firm sorting adjusts slowly over time, whereas the model's prediction—based on steady-state comparisons—reflects long-run equilibrium outcomes.

C.4 Extended Model

In this section, I extend the baseline model to justify empirical strategies in Section 5.

Model. I extend the baseline model in several dimensions. I continue to focus on the limiting case where δ_v approaches infinity. I introduce migration frictions. Workers, either employed or unemployed, can migrate only when they receive a migration shock at rate δ_m. Upon receiving this shock, workers draw preference shocks $\varepsilon(\ell) \overset{iid}{\sim} F_\varepsilon$ for each ℓ. This preference shock enters multiplicatively into the worker's value of moving to ℓ. Workers then move to the optimal location where they start searching for a job. They are allowed to stay in the current location; in such case, they lose

model would resemble the version described in Section C.4, where a key feature is the forward-looking migration decisions of workers, accounting for the changes in distribution of firms across space over time.

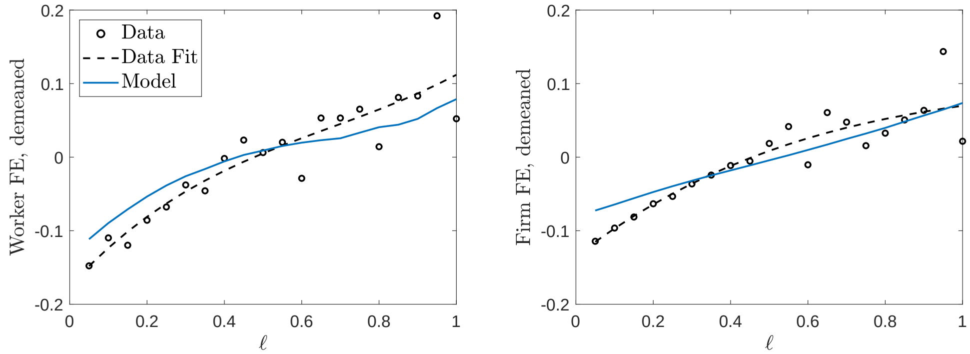


Figure A.5. Workers and Firms AKM Fixed Effects: Data vs. Model

Notes: Data source: LIAB (2010-2016). I first order CZs based on PCA of log wages and log population density. I then group them into 20 bins, each representing 5% of the population. The fitted line is estimated using a third-order polynomial regression.

their current jobs. With preference shocks, a major departure from the baseline is that an equilibrium is no longer characterized by a pure assignment. Workers with low productivity may choose a high-ℓ location if $\varepsilon(\ell)$ is sufficiently large. I assume that the CDF $F_\varepsilon \in \mathcal{C}$ has a large support, so the probability of choosing ℓ from ℓ' is positive for all ℓ, ℓ'.

Migration is costly, and when a worker moves from ℓ' to ℓ, her value is discounted by a factor of $m^{-1}_{\ell',\ell}$. For example, these costs may arise from the distance between two regions (Schwartz, 1973). Given these two components, the value of a worker moving from ℓ' to ℓ equals $\varepsilon(\ell) m^{-1}_{\ell',\ell}$ multiplied by the value of searching for a job in location ℓ, $V^u(x, \ell)$, which I describe below.

Location decisions. The flow value of workers, excluding preference shocks $\varepsilon(\ell)$ and migration costs $m^{-1}_{\ell',\ell}$, remains nearly the same, except for the added migration option value $V^m(x, \ell)$,

$$\rho V^u(x,\ell) = bx + \Pi - \bar{h}r(\ell) + \lambda(\ell)\max\{V^e(x,y(\ell),\ell) - V^u(x,\ell), 0\} + \delta_m(V^m(x,\ell) - V^u(x,\ell)),$$
$$\rho V^e(x,y,\ell) = w(x,y,\ell) + \Pi - \bar{h}r(\ell) + \delta(V^u(x,\ell) - V^e(x,y,\ell)) + \delta_m(V^m(x,\ell) - V^e(x,y,\ell)),$$

where $V^m(x,\ell) = \mathbb{E}[\max_{\ell'}\{m^{-1}_{\ell,\ell'}\varepsilon(\ell')V^u(x,\ell')\}]$.[45] Defining the match surplus, $S(x,y,\ell) = V^e(x,y,\ell) - V^u(x,\ell) + V^p(x,y,\ell) - V^v(y,\ell)$ as before, the only change from the baseline is that $\tilde{\rho} = \rho + \delta$ increases to $\tilde{\rho} = \rho + \delta + \delta_m$. The value of firms remains almost the same, except that $x(\ell)$ is replaced by $\mathbb{E}[x|\ell]$ to reflect the non-degenerate distribution of worker productivity. Wages have the same formula with modified $\tilde{\rho}$,

$$\log w(x,y,\ell) = \log x + \log\left(b + (1-\beta)\frac{\beta\lambda(\ell)}{\tilde{\rho} + \beta\lambda(\ell)}(y(\ell) - b) + \beta(y - b)\right).$$

[45] Values may differ due to a history of migration through $m^{-1}_{\ell',\ell}$ that remains after migration. Since this factor does not influence workers' decisions, tracking the cumulative migration cost offers no additional insight. Thus, I disregard it in the equilibrium characterization for expositional simplicity.

Importantly, wages are log-additive in worker productivity and the remaining components. Moreover, a set of assumptions—random matching, exogenous separation, and the employment status of migrants—guarantee the exogenous mobility assumption required for two-way fixed-effects estimation.

With some algebra, following Section A.1, the values of workers and firms are given by:

$$\rho V^u(x,\ell) = bx + \underbrace{\frac{\beta\lambda(\ell)}{\tilde{\rho}+\beta\lambda(\ell)}(y(\ell)-b)}_{\equiv A_w(y(\ell),\lambda(\ell))}x - \bar{h}r(\ell),$$

$$\rho \bar{V}^v(y,\ell) = \underbrace{(1-\beta)q(\ell)\,\mathbb{E}[x|\ell]}_{\equiv A_f(\mathbb{E}[x|\ell],q(\ell))}\left(y - b - \underbrace{\frac{\beta\lambda(\ell)}{\tilde{\rho}+\beta\lambda(\ell)}(y(\ell)-b)}_{\equiv B_f(y(\ell),\theta(\ell))}\right) - c(\ell).$$

Observe that the firm's value has increasing differences in $A_f(\cdot)$ and y, leading to the following prediction.

Lemma A.1. *If a location ℓ experiences an exogenous increase in worker productivity $\mathbb{E}[x|\ell]$, holding the vacancy contact rate constant, it will attract more productive firms than before.*

The location decision of workers of x in ℓ' is characterized by the probability that she chooses a location in the interval $[0,\ell]$, $\bar{m}(\ell|x,\ell')$. Define the corresponding density $m(\ell|x,\ell')$, which exists under the regularity assumption on F_ε. Note that as the dispersion of preference shocks and migration frictions converge to zero, location choices converge to a pure assignment characterized by a function $x(\ell)$, and the decision simplifies to $\bar{m}(\ell|x,\ell') = 1\{x \leq x(\ell)\}$. As firms' location decisions remain frictionless as in the baseline, a firm assignment remains pure and can be characterized by a function $y(\ell)$.

Migration network. Let $\mu_\ell(x)$ denote the measure of workers of productivity x who are located in ℓ. Then, this measure satisfies the following law of motion,

$$\dot{\mu}_\ell(x) = -\delta_m\mu_\ell(x) + \int_0^1 \delta_m\mu_{\ell'}(x)m(\ell|x,\ell')\,\mathrm{d}\ell'.$$

For example, if $m_{\ell',\ell} = 1$, then $m(\ell|x,\ell')$ becomes independent of ℓ' and simplifies to $m(\ell|x)$, and in steady state, $\mu_\ell(x) = m(\ell|x)M_w q_w(x)$.

I now examine the migration network, which provides crucial information for constructing the instrument in Section 5. The characterization below shows that different pairs of locations can have different magnitudes of migration flows. Moreover, it shows how changes in the number of out-migrants and their productivity from an origin ℓ' affect the average productivity of incoming migrants, which implies that my instrument in (15) is correlated with the observed changes in migrant productivity.

The migration network is characterized by the probability of departing ℓ' conditional on arriving in ℓ,

$$\Pr(\ell' \to \ell|\ell) = \frac{\int_{\underline{x}}^{\overline{x}} m(\ell|x,\ell')\mu_{\ell'}(x)\,\mathrm{d}x}{\int_0^1 \int_{\underline{x}}^{\overline{x}} m(\ell|x,k)\mu_k(x)\,\mathrm{d}x\,\mathrm{d}k}.$$

Migration flows between regions vary with the cost of moving, $m_{\ell',\ell}$. Lower migration costs induce greater migration flows. For example, if it is less costly for workers to migrate to nearby areas, migration networks are stronger between geographically proximate locations. These flows also depend on the relative attractiveness of alternative destinations, $k \neq \ell$. When other locations offer lower value to workers from ℓ', a greater share migrates to ℓ.

The average productivity of migrants arriving in ℓ can be computed by combining the migration network with the distribution of worker productivity in origins,

$$\mathbb{E}[x|\ell,\ \text{migrant}] = \int_{\underline{x}}^{\overline{x}} \Pr(\ell' \to \ell|\ell)\,\mathbb{E}[x|\ell' \to \ell]\,\mathrm{d}x \quad \text{where} \quad \mathbb{E}[x|\ell' \to \ell] = \int_{\underline{x}}^{\overline{x}} \frac{m(\ell|x,\ell')\mu_{\ell'}(x)}{\int_{\underline{x}}^{\overline{x}} m(\ell|x',\ell')\mu_{\ell'}(x')\,\mathrm{d}x'} x\,\mathrm{d}x.$$

Workers arriving in ℓ are likely to be more productive when a larger share of migrants originates from regions with higher average productivity.