Spatial Sorting of Workers and Firms*

Ryungha Oh†
Northwestern University

May 2025

Abstract

This paper develops a theory of two-sided spatial sorting in which 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 colocate in cities due to complementarities in production, and cities with higher-quality search pools become densely populated. Although positive assortative matching improves allocative efficiency, concentration in cities is inefficiently high. Since workers and firms embody productivity, relocating them can mitigate congestion without reducing output. Comparing two-sided sorting to models that abstract from either worker or firm sorting reveals 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:* `ryunghaoh@gmail.com`

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, substantial spatial wage differences in the data suggest heterogeneity in the quality of local matching pools, which arises endogenously from the location decisions of workers and firms. How does the presence of geography influence assortative matching? Can such sorting patterns 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. Importantly, I contrast this result 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 show its capacity to match cross-sectional patterns and to evaluate real-world policies, which illustrates how two-sided sorting shapes policy outcomes in practice. 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, which jointly shapes the value of each location. In each local labor market, workers and firms engage in random matching under search frictions, and upon matching, bargain over the match surplus. Thus, both high productivity and a large number of partners—which ensures a high matching probability—increase the attractiveness of local labor markets. However, concentration in a location leads to higher local costs. The costs of living or operating a business are higher in denser locations.

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

I show that positive assortative matching (PAM) arises across space, and that locations with more productive workers and firms become densely populated. Locations with either more productive firms or higher job arrival rates attract a greater number of workers until local congestion, in the form of high housing rents, outweighs the benefits of local labor markets. In particular, due to worker-firm complementarity, highly productive workers benefit more from these labor markets and thus are willing to pay higher housing rents. Importantly, I show that productivity—not matching probability—is a driving force behind assortative matching. The intuition is that matching probability, which depends on the relative number of workers and firms, may favor either side, while 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 such as 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 leads to excessive concentration in dense areas. The key insight is that productivity is embodied in workers and firms, which highlights that what matters is *who* produces, not *where* production takes place. Thus, concentration in denser areas is not desirable because it would only raise aggregate 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 desired from a social point of view, and these externalities lead to excess congestion costs. I show that the government can restore efficiency by taxing workers and firms in dense areas.

The importance of geography in shaping matching outcomes stems from its role in determining the types of potential partners. To illustrate this, I consider an alternative environment with neoclassical local labor markets, following Becker (1973). In contrast to the baseline model, geography in this setting has no economic meaning. Specifically, the matching between workers and firms replicates the outcome of 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 within a local market—yields qualitatively identical results, and again, leaves no meaningful role for geography.

In the final part of the theory, I examine the extent to which the positive and normative implications critically depend on two-sided sorting. I extend the baseline model to incorporate additional sources of spatial disparities—exogenous local TFP and agglomeration forces—which nests alternative mechanisms commonly studied in the literature, such as one-sided sorting. I show that all these mechanisms are equally successful in matching cross-sectional data. In contrast, I find that efficiency implications are sensitive to the nature of two-sided sorting. For example, when productivity is partly embodied in locations, spatial concentration in areas with high local TFP 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 exploit a key trade-off in workers' location decisions between matching opportunities and housing rents. Nominal wages in denser areas are higher due to either the compensation for higher housing rents or the skill premium. Controlling for housing rents, I estimate worker productivity differentials. Then, the remaining spatial wage gap is attributed to firm sorting. 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 two-sided sorting relative to alternative mechanisms. In the two-sided sorting model, relaxing housing regulations in denser cities has a negative impact on welfare. This deregulation amplifies the concentration of economic activity in cities, thereby exacerbating congestion costs. However, because both workers and firms relocate, the matching between them remains unchanged, leaving total output 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 the gains outweigh the associated congestion costs.

Finally, I present evidence of the two-sided sorting mechanism using German employer-employee matched data. I test the unique prediction of two-sided sorting: 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 the standard two-way fixed-effects estimates. I then instrument changes in average worker productivity using predicted changes in the productivity of incoming domestic migrants. These predictions are constructed by interacting pre-period migration networks with contemporaneous changes in outflows from other locations. This instrument addresses endogeneity concerns arising from local economic 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 realized 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 by assuming an effectively elastic land supply. Consequently, the model characterizes not only the types of agents but also the measure of agents in each location.[3]

Within the literature on two-sided matching, my 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 an equilibrium outcome. With random search, equilibrium no longer exhibits perfect assortative matching. In other words, agents match with a distribution of types, rather than a single partner (e.g., Shimer and Smith, 2000; Atakan, 2016).[4] With directed search, although perfect assortative matching is attainable, it requires stronger complementarities. Eeckhout and Kircher (2010) shows that in directed or competitive

[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. Notably, when worker and firm densities of each location are exogenous, the results of Demange and Gale (1985) or Roth and Sotomayor (1990) apply.

[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.

search settings, PAM arises if the match payoff is root-supermodular in productivity. 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.

My 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., 2023). Another body of literature shows that sorting can happen across ex ante homogeneous or symmetric 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 spillovers, 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. 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 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 heterogeneity—such as geography and amenities—or agglomeration forces play a crucial role. In contrast, I provide two distinct insights that arise from the sorting mechanism. First, I show that dispersion in marginal labor productivity does not necessarily imply misallocation. Second, I highlight the possibility of improving welfare by reducing excessive spatial congestion, which stems 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 originates 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. Workers consume housing h_t and tradable goods g_t, which are used as the numeraire. Housing is a strict necessity, and flow utility is given by g_t when they rent $\bar{h}$ units of housing at rate $r_t(\ell)$. 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,[6] 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

[5] This assumption is without loss of generality to characterize the steady-state equilibrium. In Section A.4, I formally prove that the steady-state equilibrium location choices and wages are the same with or without mobility. Note that off the equilibrium path wages and values depend on mobility.

[6] This linear function is widely assumed in the macro-labor literature for tractability (e.g., Postel-Vinay and Robin, 2002).

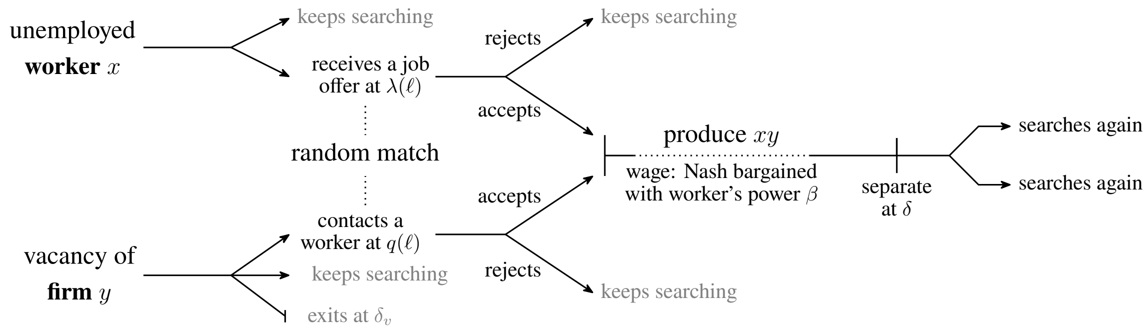


Figure 1. Local Labor Market: Timeline

$\underline{y} > b$. 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.[7] At each point in time, each firm posts δ_v units of new vacancy. Also, at each time t, 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.[8]

Technology. Workers and vacancies match one-to-one in a local labor market and 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 market. A vacancy in the search pool gets destroyed at rate δ_v.[9]

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.

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

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

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

The matching function $M(\cdot)$ is assumed to be strictly increasing, concave in both of its arguments, continuously differentiable, and $M(0,V) = M(U,0) = 0$. Defining labor market tightness as $\theta_t(\ell) = \frac{V_t(\ell)}{U_t(\ell)}$, contact 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.[10] 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, respectively, which represents their location decisions.[11] 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., cases where workers and firms randomly choose locations—they are not the focus of this paper. First, in Section 3.2, I show that the optimal assignment is pure. Therefore, restricting attention to pure assignment does not compromise the generality of the inefficiency result in Proposition 3.[12] Second, an equilibrium consistent with the empirical pattern, that is more productive workers and firms choose densely populated locations, should feature a pure assignment. See Proposition A.1 for details.

An assignment $(x(\ell), y(\ell))$ pins down the sorting patterns of workers and firms. If $y(\ell)$ increases in ℓ just like $x(\ell)$ as illustrated by $y_1(\cdot)$ in the right 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.

[10] To be precise, I assume that there exist finite $(\bar{L}, \bar{N})$ such that $\bar{h}C_r'(\bar{h}\bar{L}) \geq \underline{x}\underline{y} + \bar{h}C_r'(\bar{h}M_w)$ and $C_v'(\bar{N}) \geq \frac{1-\beta}{\tilde{\rho}} q\left(\frac{M_f}{\bar{L}}\right)\underline{x}(\underline{y} - b) + C_v'(M_f)$. These boundary conditions are required to establish the existence of equilibrium. Once an equilibrium exists, the results are independent of the degree of convexity of these functions.

[11] 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.

[12] 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.

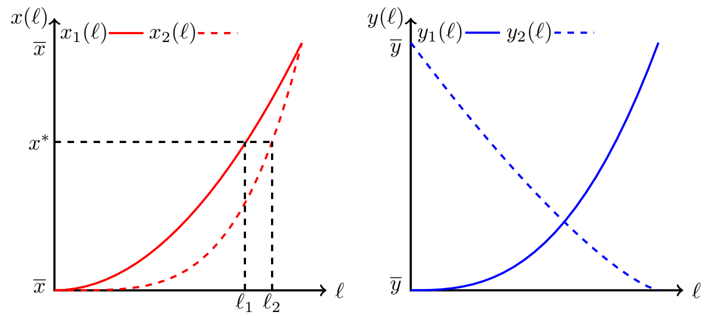


Figure 2. Assignment of Workers and Firms: Examples

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 measures of workers and firms choosing locations between 0 and ℓ equal the measures 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 $\frac{1}{M_w}\int_0^\ell L(\ell')\,\mathrm{d}\ell' = Q_w(x(\ell))$, and population density becomes,

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

Importantly, population density $L(\ell)$ is endogenously determined as an equilibrium outcome, which distinguishes this model from the standard assignment problem. For example, consider the two assignments, $x_1(\cdot)$ and $x_2(\cdot)$, in the left panel of Figure 2. Under $x_2(\cdot)$, higher-ℓ locations become denser than under $x_1(\cdot)$. The same measure of workers with productivity above x^* is concentrated in $[\ell_2, 1]$ under $x_2(\cdot)$ but dispersed over $[\ell_1, 1]$ under $x_1(\cdot)$. Specifically, population density at ℓ where x^* selects under $x_2(\cdot)$ is higher than under $x_1(\cdot)$, i.e., $M_w Q_w'(x_1(\ell_1))x_1'(\ell_1) < M_w Q_w'(x_2(\ell_2))x_2'(\ell_2)$, which follows from the steeper slope of $x_2(\ell_2)$.[13] 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. 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 $y(\ell)$ denotes the equilibrium productivity of firms in the local labor market, and 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 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 that is posted by a firm of productivity y operating in location 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:

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

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

[13] Because population density depends on both $Q'_x(x(\ell))$ and $x'(\ell)$, focusing on a pure assignment does not pin down its distribution across ℓ, unlike in an economy with finitely many worker types or locations. In Section A.6, I present an 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).

worker and a firm that posted the job 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. A worker enjoys her reservation unemployment value $V^u(x,\ell)$ plus a share β of the surplus, and a 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 $1-\tilde{\beta} \equiv \frac{\rho}{\rho+\delta_v}(1-\beta)$. First, wages 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:

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

where $\tilde{\rho} \equiv \rho+\delta$ when the surplus is positive, i.e., $S(x,y(\ell),\ell) \geq 0$ or equivalently $x(\ell) \geq B_w(x(\ell),\theta(\ell))$. Otherwise, the worker remains unemployed, and $\rho V^u(x,\ell) = bx + \Pi - \bar{h}r(\ell)$. I summarize the marginal return to worker productivity in the local labor market as *job opportunities*, $A_w(\cdot)$, which increase in local firm productivity $y(\ell)$ and the job arrival rate $\lambda(\ell)$. However, better job opportunities do not necessarily translate into higher values for workers. Firms appropriate a portion of the surplus, as represented by the term $B_w(\cdot)$, which increases in local worker productivity $x(\ell)$ and the vacancy contact rate $q(\ell)$. The final term represents the local congestion costs associated with housing expenditure. Workers choose locations that maximize their value functions, $V^u(x,\ell)$, 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 $\bar{V}^v(y,\ell)$. The value of a firm equals the discounted sum of the values of vacancies posted at each point in time, net of overhead costs, i.e., $\delta_v V^v(y,\ell) - c(\ell)$, which is given by

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

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 define the marginal return to firm's productivity in local labor markets as *hiring opportunities*, $A_f(\cdot)$, which increase in local worker productivity $x(\ell)$ and the vacancy contact rate $q(\ell)$. The term $B_f(\cdot)$ represents the fact that firms share the surplus with workers. Firms take a smaller share of the surplus when the value of job search for workers is higher due to higher $y(\ell)$ and $\lambda(\ell)$. Finally, firms need to pay their overhead costs $c(\ell)$, which increase in a measure of local firms due to increased demand for business services. Firms choose locations that yield the highest value, $\bar{V}^v(y,\ell)$, which determine 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:

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

Laws of motions. I consider the laws of motion for the measure of vacancies $V(\ell)$ and unemployment rate $u(\ell)$. A measure $N(\ell)$ of firms in location ℓ posts a flow $\delta_v N(\ell)$ of vacancies. A flow of separated matches $\delta(1-u(\ell))L(\ell)$ reenters the search pool. Vacancies can either be matched with unemployed workers at rate $q(\ell)$ or be destroyed at rate δ_v. Unemployed workers find jobs at rate $\lambda(\ell)$, and employed workers lose their jobs at rate δ. The flow-balance conditions for steady-state unemployment rates and the measure of vacancies are given by

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

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) every worker and firm 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 the efficiency properties. I close this section with a brief discussion of alternative modeling choices.

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) satisfies a single-crossing condition in worker productivity x and job opportunities $A_w(\cdot)$, and firm's value in (5) satisfies a single-crossing condition 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 results on monotone comparative statics (Milgrom and Shannon, 1994). Firm productivity $y(\ell)$ also increases if hiring opportunities increase in ℓ. Importantly, these patterns can be self-fulfilling because workers and firms simultaneously sort across ℓ. An increasing $x(\ell)$ ($y(\ell)$, respectively) can account for increasing $A_f(x(\ell), q(\ell))$ ($A_w(y(\ell), \lambda(\ell))$, respectively). In other words, the complementarity of the output function can create worker-firm complementarity in the spatial sorting problem, and in turn, can lead to PAM between workers and firms.

To confirm that this is indeed the case, I need to rule out the possibility that job opportunities are higher in a high-ℓ location with lower firm productivity $y(\ell)$ but sufficiently higher job arrival rate $\lambda(\ell)$. Search frictions generate differences in local contact rates, which makes PAM non-trivial.[14] Although such a scenario is consistent with the decision of workers, it cannot also be consistent with that of firms. A higher job arrival rate requires a larger measure of firms, which increases overhead costs. At the same time, if the

[14] 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 enables PAM.

high-ℓ location is chosen by firms with lower productivity, this implies that hiring opportunities are smaller in this location. Therefore, firms would prefer to deviate to a low-ℓ location, which prevents a non-PAM equilibrium.[15]

Having established that an assignment exhibits PAM if an equilibrium exists, the sorting of workers and firms can be characterized by two strictly increasing functions, $x(\ell)$ and $y(\ell)$, and the optimal location decisions are characterized by 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, equilibrium assignment can be obtained by combining the above equations, the boundary conditions—$x(0) = \underline{x}, x(1) = \overline{x}, y(0) = \underline{y}$, and $y(1) = \overline{y}$—the housing and business services market clearing conditions (6), and the steady-state flow-balance conditions (7). Mathematically, an equilibrium is the solution to the differential equations of $(x(\ell), y(\ell), x'(\ell), y'(\ell))$, and existence follows from the standard ODE theorem.[16] 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.*

Next, I characterize the spatial disparity implied by the sorting of workers and firms. When workers choose a location, they trade off between gains from local labor markets and costs from housing rents $r(\ell)$. Since I focus on a pure assignment, there is only a single type of workers and a single type of firms in each ℓ, and thus in equilibrium whenever a worker contacts a vacancy, they form a match. This also holds when workers deviate slightly from their optimal locations, due to the continuity of $x(\cdot)$ and $y(\cdot)$. Hence, the first-order condition of workers in (8) leads to

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

[15] The model does not make a specific prediction on spatial variation in market tightness. Cross-sectional market tightness 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 job arrival rates (higher worker productivity, 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.

[16] 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.

When choosing a marginally higher-ℓ location, workers can benefit from a better local labor market with higher job opportunities $A_w(\cdot)$. This marginal gain is partially offset by the changes in $A_w(\cdot)B_w(\cdot)$, which capture the firms' threat point in wage determination. Additionally, workers need to consider changes in housing rents. The conditions for firms' optimal location decisions can be similarly analyzed. Importantly, labor market gains increase in ℓ for both workers and firms, and this pattern is primarily driven by differences in productivity. In contrast, the differences in market tightness can benefit only one side, but not both at the same time.

I first show that population density increases in ℓ. In higher-ℓ locations, labor market gains are higher due to better job opportunities $A_w(\cdot)$. Thus, a larger number of workers are attracted to those locations, which drives up housing demand and, in turn, housing rents. Equilibrium population density is determined at the point where the marginal benefit equals the marginal cost from higher housing rents, as shown in (9). Note that the vacancy destruction rate δ_v needs to be sufficiently large for this result. If it is too low, gains in higher-ℓ locations may be lower despite higher job opportunities, due to an excessively high threat point of firms, $B_w(\cdot)$. 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.

Random matching plays a key role for the spatial variation in population density by introducing imperfect screening in local labor markets. It ensures that workers' labor market gains increase in local firm productivity, all else equal. Since search is random, all workers have the opportunity to be matched with more productive firms, as long as they search in locations where these firms are located. In such a case, workers earn higher wages as they enjoy a share the surplus increase. In contrast, if firms could select 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 lead to an increase in wages. I present the results under these alternative labor market structures in Section 3.3 and discuss the differences.

Second, I show that wages also increase in ℓ. Plugging the equilibrium assignment $(x(\ell), y(\ell))$ into (3), the equilibrium wages become

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

Wages differ across locations due to differences in worker and firm productivity, $x(\ell)$ and $y(\ell)$, and market tightness, $\theta(\ell)$. Higher worker and firm productivity leads to greater output, which results in higher local wages. Also, higher market tightness raises wages, as it increases workers' threat points in bargaining. The condition on the vacancy destruction rate δ_v plays the same role as before. When it is large and thus $(1 - \tilde{\beta})$ is small, firms' threat point in wage bargaining is sufficiently low, so that even if market tightness decreases, wages still increase in ℓ.

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

Proposition 2. *Suppose that δ_v is sufficiently large. In any pure-assignment equilibrium, population density $L(\cdot)$ and wages $w(x(\cdot), y(\cdot), \cdot)$ 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 several key dimensions of the spatial disparities we observe in the data. It is important to note that 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. The planner maximizes net output, a standard approach in the literature on frictional labor markets with transferable utilities (e.g., Shi, 2001; Acemoglu, 2001).[17] In this section, I focus on the limit case where a discount rate ρ goes to zero to simplify the exposition.[18] I present the proofs of all results in Section A.5.

I first characterize the optimal assignment and compare it with the equilibrium assignment. As in the literature, complementarity in output is a key force ensuring that the optimal matching follows PAM. In this economy, while spatial differences in market tightness introduce an additional factor affecting output beyond worker and firm productivity, PAM remains the optimal outcome. 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.*

[17] It is equivalent to solving the problem of a planner who can freely use transfers. This equivalence arises from the stylized nature of the economy, particularly under the assumptions of risk neutrality and exogenous labor supply. I further show that the equilibrium is Pareto inefficient in Section A.5.

[18] This assumption eliminates any asymmetry in impatience between workers (firms) and the social planner. Workers (firms) choose locations that maximize their unemployment value (vacancy posting value). I assume that workers and firms are infinitely patient to ensure that the social planner's objective function is comparable to the value function of workers and firms.

Using Lemma 1, the planner's problem becomes choosing an assignment $(x^*(\ell), y^*(\ell))$, in which both components are increasing. Even among increasing assignments, the planner still needs to determine worker and firm densities across locations. Local output is determined by the number of employed workers $(1-u(\ell))L(\ell)$ and the match output $x(\ell)y(\ell)$, and congestion costs depend on worker and firm densities. The planner solves the following problem,

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

in addition to the boundary conditions on $x(\ell)$ and $y(\ell)$. Constraints determine how the assignment pins down population density, firm density, and unemployment rates as in Section 2.2.

I show that the optimal assignment is characterized by

$$\bar{h}^2 C_r'(\bar{h}L^*(\ell))(L^*)'(\ell)L^*(\ell) + C_v'(N^*(\ell))(N^*)'(\ell)N^*(\ell) = 0 \quad \forall \ell, \tag{11}$$

in addition to (A.14) in Section A.5.

First, the planner finds it unnecessary to generate congestion because productivity is embodied in workers and firms. Specifically, condition (11) implies that changes in population density and firm density at each location ℓ have the opposite signs, i.e., $(L^*)'(\ell)(N^*)'(\ell) \leq 0$ for all ℓ. When the planner makes marginal adjustments to the assignment between two locations, she does not make one location strictly denser than another. If a location is more concentrated, the planner can maintain the same level of output while reducing overall congestion costs by relocating both workers and firms to a less concentrated one.

The second condition (A.14) governs how the planner adjusts market tightness, i.e., whether to increase worker or firm density in response to productivity heterogeneity. For example, the planner chooses to allocate relatively more workers in higher-ℓ location—i.e., $(L^*)'(\ell) > 0, (N^*)'(\ell) < 0$, and $(\theta^*)'(\ell) < 0$—when heterogeneity in firm productivity is more pronounced. By concentrating highly productive firms and a large measure of relatively homogeneous workers in higher ℓ, which raises these firms' vacancy contact rates, the planner can increase output.

Next, I evaluate the efficiency of the equilibrium. Workers and firms are drawn to cities by opportunities to search for more productive counterparts, which leads to congestion in these regions. In particular, in contrast to the optimal allocation, the marginal change in congestion in equilibrium is positive, as given by

$$\bar{h}r'(\ell)L(\ell)+c'(\ell)N(\ell)=\bar{h}^2C_r'(\bar{h}L(\ell))L'(\ell)L(\ell)+C_v'(N(\ell))N'(\ell)N(\ell)>0 \quad \forall \ell.$$

When workers choose between two locations, they do not consider that their choice may reduce local firms' chances of hiring more productive workers. Similarly, firms do not internalize the effects of their decisions on workers in the same manner. Due to these negative externalities, they are willing to pay higher housing rents $r(\ell)$ or overhead costs $c(\ell)$ to participate in the labor markets in these higher-ℓ locations.

To formalize this idea, I assume frictional local labor markets with random matching, together with heterogeneous agents sorting across space. When workers and firms are heterogeneous, the extent of the externalities they impose on others varies. More productive workers generate larger positive externalities on firms but the same negative externalities on workers. Thus, a single condition—i.e., the Hosios condition—no longer ensures these externalities cancel out (e.g., Acemoglu, 2001; Shimer and Smith, 2001; Albrecht et al., 2010).[19] The below proposition summarizes the result.

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.*

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, a generalized condition (e.g., Mangin and Julien, 2021) cannot address the inefficiency. For example, higher workers' bargaining power in cities may lower firms' inflows but increase workers' inflow even more. In line with the literature, when I assume the Hosios condition—the elasticity of $M(U,V)$ with respect to U equals β—there is no need to correct for externalities arising from market tightness, and the planner needs to correct only externalities from type heterogeneity. In particular, under the Hosios condition and zero unemployment benefit $b=0$, an equilibrium is efficient if and only if workers and firms are homogeneous.

The planner can implement the optimal assignment by using the spatial transfers to workers and firms, $t_w(\ell)$ and $t_f(\ell)$, as given by (A.19) and (A.20). To focus on externalities that arise from sorting, assume the Hosios condition and zero unemployment benefit. Then, spatial transfers simplify to the following

[19] The most related application is Bilal (2023) who applies this idea to the spatial sorting model of heterogeneous firms.

expressions, which decrease in ℓ,

$$
\begin{aligned}
t_w(\ell) &= t_w^0 - \int_0^{\ell} \frac{\varepsilon_\lambda(1-u^*(t))u^*(t)}{1-\varepsilon_\lambda(1-u^*(t))} y^*(t) x^{*'}(t)\,\mathrm{d}t, \\
t_f(\ell) &= t_f^0 - \int_0^{\ell} (1-\varepsilon_\lambda)(1-u^*(t)) \frac{L^*(t)}{V^*(t)} x^*(t) y^{*'}(t)\,\mathrm{d}t,
\end{aligned}
$$

where the constants t_w^0 and t_f^0 balance the government budget; ε_λ denotes the constant elasticity of job finding rates with respect to market tightness; $(x^*(\ell), y^*(\ell))$ denotes the optimal assignment; and $L^*(\ell), V^*(\ell)$, and $u^*(\ell)$ are determined under these allocations. In other words, the planner subsidizes workers and firms in lower-ℓ locations where congestion is lower. Notably, the marginal changes in transfers, $t_w'(\ell)$ and $t_f'(\ell)$, are more pronounced when $x^{*'}(\ell)$ and $y^{*'}(\ell)$ are larger, which indicates greater heterogeneity. Moreover, the above expression confirms that spatial transfers become unnecessary if and only if workers and firms are homogeneous.

In sum, local labor markets and spatial sorting ensure that matching is not fully random, enabling assortative matching. However, 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, which goes against (11). Nevertheless, PAM outperforms random matching when the output gains from complementarity exceed the additional congestion costs.[20]

3.3 Discussion of Alternative Modeling Choices

To understand the role of random matching, I consider two alternative labor market structures: neoclassical labor markets and labor markets with directed search. I show that these two models fall short of explaining the spatial disparities observed in the data, particularly differential population densities across regions. Moreover, the equilibrium turns out to be efficient in both cases, in contrast to the baseline case shown in Proposition 3.

[20] In contrast to Hoppe et al. (2009), my model does not establish conditions on the productivity distribution that guarantee PAM outperforms random matching. Instead, the outcome 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.

Neoclassical local labor markets. I assume that there are no frictions in local labor markets while maintaining the other assumptions from the baseline model. In a local labor market in ℓ, for a given local wage schedule $w(x,\ell)$, each firm hires a worker of productivity x to maximize its profit, $xy - w(x,\ell)$. The matching process remains bilateral, so each firm can hire only one worker. However, firms are free to choose the type of worker they hire, as long as they pay the corresponding wage. Each local labor market ℓ has the same environment as the seminal work by Becker (1973). The only difference is that labor markets are segmented by location. The following proposition characterizes the pure assignment equilibrium. See Section A.7 for the proof.

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

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

(2) Population density $L(\cdot)$ is the same in all locations.

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

In equilibrium, workers' wages do not depend on their location decisions, i.e., $\frac{\partial}{\partial \ell} w(x,\ell) = 0$. If instead $\frac{\partial}{\partial \ell} w(x,\ell) > 0$, then higher wages attract a greater population density and induce higher housing rents. On the firm side, because firms optimally choose the worker type x, the derivative of profit with respect to x is zero, and thus higher wages must be compensated by lower overhead costs and firm density in equilibrium. However, the resulting imbalance in worker and firm densities cannot be sustained in the absence of search frictions. Since wages do not depend on locations, workers have no incentives to choose dense and expensive locations. In turn, population density is uniform across regions.[21]

More importantly, the third property indicates that locations are economically irrelevant in this model. Intuitively, as long as workers and firms could be selectively matched with each other without any frictions, the existence of locations does not change matching outcomes between the two.

[21] Alternatively, consider the scenario where search frictions disappear such that $\lambda(\ell) \to \infty$ and $q(\ell) \to 0$. In this case, workers capture the entire match surplus, and the population density continues to increase in ℓ in the limit.

Directed search. In Section A.8, I consider an economy in which each local labor market is subject to search frictions, and workers and firms engage in competitive or directed search (e.g., Moen, 1997). In each labor market, firms post a worker-type-specific wage, and workers optimally queue for firms. These decisions determine both wages and the probability of matching. 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 first select locations. Then, in each location, workers and firms competitively search among those who chose the same location.

The key distinction from the baseline model with random matching is that under directed search, firms would not match with less productive workers unless doing so yields the same value as matching with more productive ones in local markets. 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. However, the equilibrium does *not* feature spatial concentration: Workers and firms are uniformly distributed across space. The equilibrium is efficient, and most importantly, the introduction of location has no effect on matching outcomes. See Proposition A.3 for the formal result.

3.4 Discussion of Alternative Mechanisms Accounting for Spatial Disparity

This paper focuses on the two-sided sorting mechanism that endogenously generates spatial disparity. However, similar spatial patterns may also arise from alternative mechanisms considered in the literature. What distinguishes two-sided sorting from these alternatives? To address this question, I compare the two-sided sorting mechanism to others in both positive and normative aspects.

I introduce the two additional sources of productivity: heterogeneous location productivity and agglomeration forces. Locations differ in terms of local TFP, denoted by $A(\ell) = \bar{A}(\ell)A^x(x(\ell))$, where $\bar{A}(\ell)$ represents exogenous location productivity and $A^x(x(\ell))$ reflects agglomeration forces. I focus on agglomeration forces arising from knowledge spillovers and assume that $A^x(x(\ell))$ increases in the quality of local workers $x(\ell)$.[22] To preserve the structure of the baseline model, I incorporate these productivity components by modifying the output function. Specifically, the output of a worker of productivity x and a firm of productivity y in location ℓ is given by $A(\ell)xy$.

[22] There are various forms of agglomeration forces. A large strand of literature assumes that local TFP increases in population density or city size (e.g., Kline and Moretti, 2014). It is more difficult to compare this mechanism with the two-sided sorting mechanism because its normative implications depend heavily on the functional form.

With these modifications, the expressions for the values of workers and firms choosing location ℓ remain almost the same. The equilibrium wage also remains nearly unchanged, except for an additional source of heterogeneity from 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.9.

This extended model relates closely to several important mechanisms in the literature, depending on the sources of heterogeneity: worker productivity, firm productivity, exogenous location productivity, and agglomeration forces. For example, it returns to the *two-sided sorting* model when focusing on worker and firm heterogeneity without considering local TFP, i.e., $\bar{A}(\ell) = A^x(x(\ell)) = 1$. I compare this model to three alternative 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 heterogeneity $x(\ell)$ alongside agglomeration forces $A^x(x(\ell))$ yields the *spillovers* model.

The following proposition compares these different 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 imply different policy interventions.*

Each source of spatial disparities contributes to raising output and wages and thus plays a similar role in matching cross-sectional data. For example, as shown in (12), an increase in wages in higher-ℓ locations can result from higher firm productivity $y(\ell)$, worker productivity $x(\ell)$, or local TFP $A(\ell)$. Thus, distinguishing between models based only on cross-sectional data is difficult. This explains why previous studies could account for the same empirical patterns of spatial disparity based on different mechanisms. Proposition 1 shows that two-sided sorting performs comparably to other models.

However, each model yields distinct policy implications. I first compare the *no sorting* and *one-sided sorting* models, which feature exogenous location productivity, to the *two-sided sorting* model. 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 local TFP, and congestion in cities may emerge, i.e., $L'(\ell) > 0$ and $N'(\ell) > 0$.[23]

Compared to the previous two models, the *spillovers* model more closely resembles the two-sided sorting model but differs in two respects. On the one hand, congestion in cities is unnecessary as in the two-sided sorting model because it assumes ex ante homogeneous locations[24]. On the other hand, externalities originate from only workers, through the quality of matching pools and agglomeration forces. Thus, the planner imposes more aggressive spatial transfers to workers, while no spatial transfers to firms. Moreover, the welfare implications of specific policies may differ. For example, place-based policies that attract high-performing firms to low-income locations increase net output in the spillovers model. In contrast, in the two-sided sorting model, such policies disrupt PAM, and are thus undesirable.

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

4. Quantitative Analysis

In this section, I first calibrate the two-sided sorting model using cross-sectional data on U.S. MSAs, 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 alternative mechanisms. This analysis illustrates how the theoretical results in Proposition 5 translate into real-world policy evaluations.[25]

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

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

[24] Although some studies on spillovers examine efficiency, the results often depend on parameterization, or focus on interactions across worker types rather than between workers and firms (e.g., Fajgelbaum and Gaubert, 2020; Rossi-Hansberg et al., 2019; Behrens et al., 2014).

[25] Importantly, 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.

consumption. This change enables the model to accurately capture the impact of congestion costs due to 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 the local government taxes 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. This tax is progressive, that is, the tax rate increases in workers' income. To approximate this feature, I assume that the tax rate is specific to each region.[26] Federal income tax revenues are redistributed to all workers as a lump sum.

With these changes, the value of unemployed workers becomes

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

where Π is the sum of lump-sum redistributions from firms' profits, income taxes, landowners' profits Π_r, and intermediaries' profits Π_c, 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 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 rate, housing rents, and housing spending share 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

[26] 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.

IPUMS (Flood et al., 2022), the U.S. Census, and the Bureau of Economic Analysis (BEA). See Section B.2 for further details.

To prepare for estimation, I order MSAs by population density, which increases in ℓ.[27] For computation purposes, I discretize locations into 20 bins. I group MSAs such that each bin contains an equal share of the total population. Each bin corresponds to an interval of ℓ in the model, with its length proportional to the inverse of the population density in that bin.[28] I compute the averages of population density $L(\ell)$, wage $w(\ell)$, unemployment rate $u(\ell)$, housing rent $r(\ell)$, and housing spending share 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 $\alpha = 0.5$, following the standard values 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 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 the firms' threat point is zero.[29]

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 regions $\{x(\ell), y(\ell)\}_\ell$. 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 level 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 expected profit across firms is zero after estimation. 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 that apply economy-wide are calibrated by targeting standard moments. The main challenge is to estimate worker and firm

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

[28] For clarity in the figures, I re-index ℓ to represent the percentile rank of workers across locations. As locations are ex ante homogeneous, this relabeling has no effect on the results.

[29] 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.

Table 1: Parameter Values

	Parameter	Target	Value
ρ	discount rate	interest rate	0.003
α	matching elasticity	literature	0.5
δ	separation rate	EU transition rate	0.028
$\mathcal{A}$	matching efficiency	market tightness	0.62
$\bar{h}, \omega$	housing demand	spending shares on housing	10.10, 0.11
η_w	housing supply	housing rents	11.31
b	unemployment benefit	replacement rate	0.28
β	worker's bargaining power	labor share	0.04
H_f, η_f	business services supply	wages	3.62, 14.68
T	housing tax	housing rents	1028.35

productivity. To address this, I recover productivity schedules during the estimation process. Specifically, I find $\{x(\ell), y(\ell)\}_\ell$ that satisfy the sorting conditions of workers and firms in (8), and clear housing markets in (6), conditional on population densities and unemployment rates $\{L(\ell), u(\ell)\}_\ell$, both of which are included among the targeted moments.[30]

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

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

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 that discipline the structural parameters follows standard practice and is discussed in detail in Section B.3. Here, I focus on the productivity schedules.

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

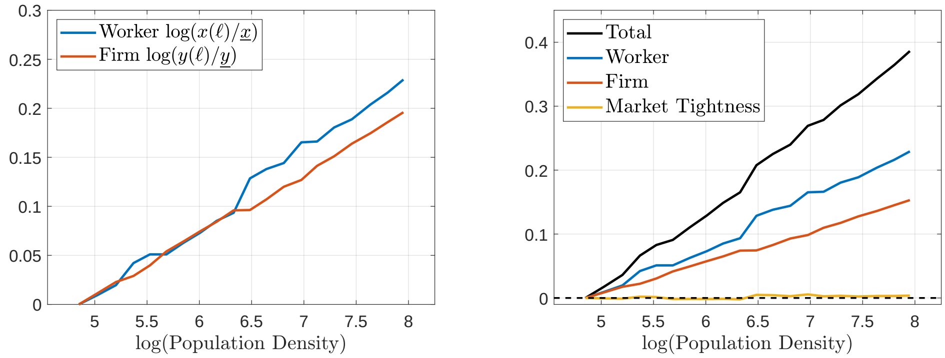


Figure 3. Worker and Firm Productivity (left) and Urban Wage Premium (right)

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.

Wage differences arise from dispersion in both worker and firm productivity, and separating the contribution of each is the key step. I infer worker productivity based on a revealed preference argument. Consider two locations, $\ell' < \ell''$, such that $\ell'' - \ell'$ is sufficiently small for workers in ℓ' to be nearly indifferent between the two. In this case, their hypothetical wage in ℓ'', $w(x(\ell'), y(\ell''), \ell'')$, must be high to compensate for the increase in housing rents. The gap between this hypothetical wage and the observed wage in ℓ'', $w(x(\ell''), y(\ell''), \ell'')$, reflects the difference in worker productivity. When the difference in housing rents is small, most of the wage gap is attributed to worker heterogeneity, thereby reducing the inferred role of firm productivity.

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. The estimated housing elasticity η_w is 11.31, which is higher than typical values in the literature (e.g., Saiz, 2010; Green et al., 2005). However, 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. The heterogeneity of both workers and firms is substantial, with worker productivity exhibiting greater dispersion.

Table 2: Model Fit

	Wage			Housing Share		Replac. rate	Labor share	Rent		
Quartile	2	3	4	mean	$4^{\text{th}} - 1^{\text{th}}$			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. Housing share difference is the average housing spending share in the last quartile minus that in the first quartile.

For example, workers and firms in the top 10% of cities are 24.4% and 20.4% more productive, respectively, than those in the bottom 10%.

Model fit. Despite its simplicity, the model successfully replicates the spatial disparities observed in the data. In Table 2, I report the fit of 10 targeted moments, excluding $\{L(\ell), u(\ell)\}_\ell$, for which the fit is documented in Figure A.2a. In Figure 4, I compare moments from the data and model for log wages, housing rents, and rent spending share across ℓ. As expected from Table 2, the model replicates the overall patterns of these moments well. Note that this stylized model performs well in explaining the overall pattern. However, it is not suitable for capturing irregular variation. In the right panel of Figure A.2a, the model aligns with the observed flat pattern in unemployment rates across ℓ, but it does not accurately fit the unemployment rates for each specific ℓ.

In sum, two-sided sorting alone suffices, not only theoretically but also quantitatively, even without region-specific factors such as local TFP or agglomeration forces.

Sources of spatial disparity and identification threat. The estimated model implies that workers and firms account for about 60.4% and 38.8% of the urban wage premium, respectively, as shown in the right panel of Figure 3. This is consistent with the literature, which emphasizes the dominant role of workers while recognizing the contribution of additional factors. See Section B.4 for the details.

Importantly, the model assumes that sorting fully accounts for spatial disparity. If other mechanisms exist, e.g., exogenous local TFP $\bar{A}(\ell)$, the estimation procedure attributes their effect to firm heterogeneity. While quantifying each mechanism is beyond the scope of this paper, I provide empirical evidence on the quantitative importance of two-sided sorting in Section 5.

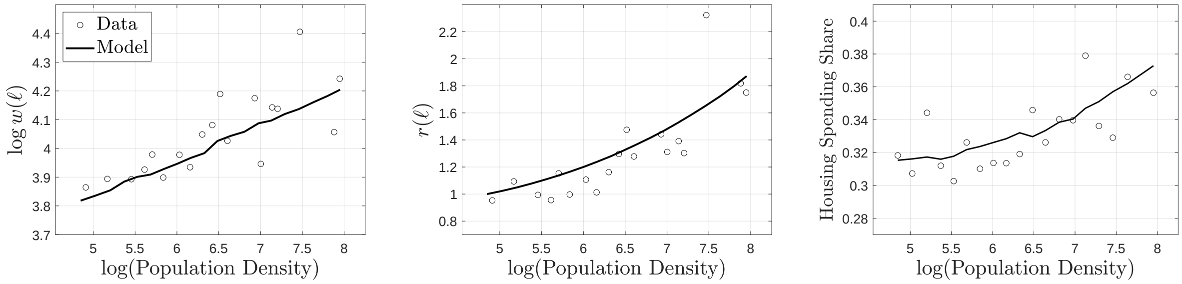


Figure 4. Model Fit: Wage, Housing Rent, and Housing Spending Share

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.

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 federal income tax of workers, discussed in Section B.6. In the U.S., housing regulations are much more stringent in dense cities, which leads to smaller housing elasticities (Saiz, 2010). 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).[31]

To implement a similar policy experiment, I relax the housing regulations in dense areas $\ell \in [0.9, 1]$ by reducing the housing tax $\tau(H;\ell)$ to the median level observed at $\ell = 0.5$. The left panel of Figure 5 shows that housing supply becomes more elastic in these high-ℓ regions. I compare two steady-states before and after the policy change.[32]

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 and are not constrained by migration frictions. Importantly, this inflow of workers increases vacancy contact rates and thus attracts

[31] Recent studies have examined how incorporating worker sorting and agglomeration forces can have additional impacts on similar policies. The conclusions vary. Some findings 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).

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

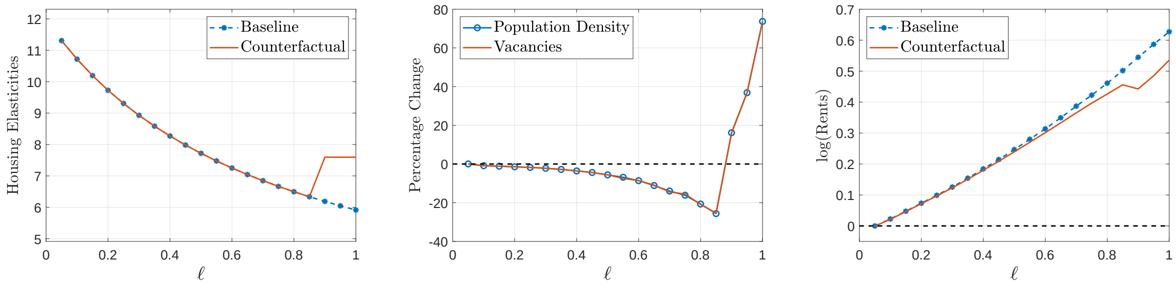


Figure 5. Impact of Relaxing Housing Regulations on Spatial Allocation

Notes: In the left panel, I plot housing elasticities before and after relaxing regulations. A red line shows the counterfactual housing elasticities when housing regulation of large cities, $\ell \in [0.9, 1]$, is relaxed to the level of the median city, $\ell = 0.5$.

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 -0.010%. Even after relocation, workers continue to be matched with similar firms, the output of each match and thus aggregate output remains roughly unchanged. In contrast, as workers move toward large cities, congestion costs from housing and business services markets increase by 1.36%.

Aggregate welfare of the utilitarian planner, i.e., an unweighted average of the values of workers, *decreases* by 0.21%. This result is consistent with Proposition 3. In the baseline, stringent housing regulations in dense areas served as spatial transfers that improved welfare. In the left panel of Figure 6, I plot changes in the value of unemployed workers by productivity percentile.[33] Welfare decrease primarily comes from housing markets. As concentration rises, the associated decline in landowner profits and transfers more than offsets the benefits of lower rents. In contrast, welfare changes from firm productivity and market tightness are modest and exhibit no clear pattern.

Comparison to one-sided sorting. I now examine the importance of incorporating both worker and firm sorting by comparing the two-sided sorting model with a 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. Estimated local TFP $\bar{A}(\ell)$ is identical to the firm productivity $y(\ell)$ in the two-sided sorting model, which ensures that both models produce the same

[33] Results for employed workers are qualitatively similar.

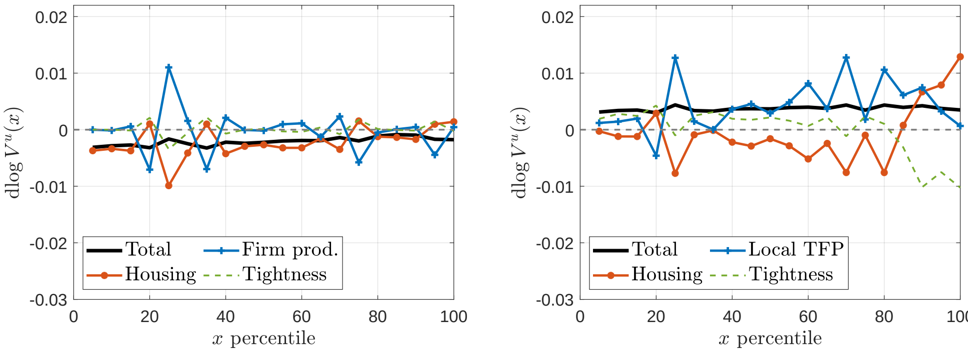


Figure 6. Impact of Relaxing Housing Regulation on Welfare

cross-sectional wage distribution. This confirms the result in Proposition 5, which shows that both models are equally successful in explaining spatial disparity.

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.[34] 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 from local TFP, whereas the left panel exhibits limited effects from firm productivity.[35]

5. Empirical Evidence of Two-Sided Sorting

In this section, I provide empirical evidence of two-sided sorting. As shown in Proposition 5, various mechanisms can yield observationally equivalent cross-sectional outcomes, complicating identification. Although a quantitative decomposition of these forces is beyond the scope of this paper, I show that two-sided sorting accounts for a significant portion of the observed variation. Specifically, I test the unique prediction of two-sided sorting: Worker sorting leads to firm sorting. As shown in (5), local labor markets with higher worker productivity $x(\ell)$ attract firms with higher productivity $y(\ell)$, conditional on the vacancy contact rate $q(\ell)$.

[34] Overhead costs in the one-sided sorting model are calibrated so that homogeneous firms' values are equalized across locations. The implied cost function $C_v(\cdot)$ is more convex than that of the two-sided sorting model, which results in a muted relocation response. Population density in the top 10% of locations increases by 10.9%. See Section B.6 for details.

[35] This is consistent with Hsieh and Moretti (2019), where productivity is fully embodied in local TFP.

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.

The dataset provides two-way fixed effects of workers and establishments from AKM wage regression.[36] These estimates are based on the IAB Employment History File (BEH), 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 for the main analysis.

Measuring productivity. Local worker productivity $x(\ell)$ in (5) represents the average worker productivity that firms expect when posting a vacancy. To construct an empirical counterpart, I compute the average productivity of workers in the local search pool.[37] I use the average productivity, which is consistent with random matching, and include all individuals eligible for new hires, whether they were unemployed or recently switched jobs. To measure the types of local jobs created, $y(\ell)$, I measure the average productivity of newly created jobs during the sample period. I focus on new jobs, as they represent the unit of production and correspond to vacancy postings. This includes jobs created by both incumbents and entrants, which allows variation driven by compositional shifts within incumbent establishments.

To estimate $x(\ell)$ and $y(\ell)$, I map standard 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). In Section C.4, I justify the strategy used in this section using an extended model that features a non-pure assignment and movers across firms and regions. Importantly, the value of firms retains the key property that I test. The wage equation also remains unchanged, which supports the validity of my productivity measures.

If local wages follow (10), then estimated worker fixed effects correctly represent the log their productivity. The average estimated fixed effect of newly matched jobs corresponds to the term in parentheses in (10).

[36] The dataset is provided at *establishment*-level. Since the model does not distinguish jobs, establishments, or firms, I use these terms interchangeably in this section.

[37] I only observe the CZ of establishments, and thus I define the CZ of migrants 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.

Since it depends on the job finding rate $\lambda(\ell)$ and other parameters, I invert the expression to obtain $y(\ell)$, which I explain in detail in Section C.1.[38]

Causal evidence of the two-sided sorting mechanism. I examine whether *changes* in worker sorting induce changes in firm sorting. Because it is difficult to identify two-sided sorting using cross-sectional data, I use a time-differenced regression, guided by (5):

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

where Δ denotes the change between period $t = 1$ (2003-2009) and $t = 2$ (2010-2016). To account for variation in the vacancy contact rate $q(\ell)$, I control for changes in the job finding rate $\lambda(\ell)$, which is observable and inversely related to $q(\ell)$.

A key concern is 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, the estimated coefficient may be biased upward. Another issue concerns the measurement of firm productivity $y(\ell)$. If local TFP $A(\ell)$ varies across ℓ, as implied by (12), then the estimated firm productivity reflects the product of $y(\ell)$ and $A(\ell)$, since the data do not follow establishments that relocate. In this case, the coefficient may partly reflect the interaction between worker sorting and local TFP.

To address these concerns, I construct an instrument based on *predicted* changes in the productivity of domestic *migrants*. These predictions rely on changes in their origin regions, not destinations. First, I compute the predicted share of migrants from each origin ℓ', $\hat{s}_t(\ell'|\ell)$, based on predicted flows $m_0(\ell' \to 0|\ell')O_{\ell',-\ell,t}$, where $m_0(\ell' \to 0|\ell')$ is the historical migration probability from ℓ' to ℓ in $t = 0$ (1991–2002), and $O_{\ell',-\ell,t}$ is the number of out-migrants from ℓ' to destinations other than ℓ in $t = 1, 2$.[39] Second, I compute the average productivity of migrants leaving ℓ' who choose destinations other than ℓ, $\hat{x}_t(\ell', -\ell)$. Combining these components, I predict the productivity of migrants arriving in ℓ as follows,

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

[38] 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., 2023). However, each strategy depends on a set of assumptions that is specific to its own context.

[39] 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 changes in the *number* of migrants, while I focus on changes in productivity.

Table 3: Changes in Local Firm Productivity

	(1) OLS	(2) IV	(3) IV	(4) IV	(5) IV
$\Delta \log x(\ell)$	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)
	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. Each observation is weighted by the number of workers.

Finally, I construct the instrument for $\Delta \log x(\ell)$ by differencing the above between the two periods.

Domestic *migrants* account for about 30% of local search pools in my data—a sizable share—and therefore directly affect local average worker productivity $x(\ell)$. Although the migration rate is limited, migrants are disproportionately represented among unemployed workers. In addition, an increase in migrant productivity indirectly affects that of non-migrants, for example, by increasing housing rents, which may in turn push out less productive workers.

The identification assumption is that shocks to origin locations ℓ' are uncorrelated with shocks to a given destination ℓ. By focusing on migrants moving to destinations other than ℓ, I exclude variation that may be driven by the pull factors of ℓ. This approach could be problematic if the historical network is influenced by factors that create correlated regional shocks, such as similar industry composition or geographic proximity. To address this concern, I either control for industry composition or exclude origin locations within a 100 km radius in robustness checks.

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 a potential time trend 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 regional unemployment rates.

Column (1) documents the coefficient of 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 the instrumental variable estimation. In the bottom panel, I report first-stage regression results. I regress changes in worker productivity $\Delta \log x(\ell)$ on the instrument $\Delta \log x^{\mathrm{m,IV}}(\ell)$ 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 10% increase in the log of worker productivity in the local search pool leads to an 11.3% increase in the log of productivity of new jobs in the same location.[40] To test robustness, in Column (3), I first residualize firm productivity with respect to 1-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 arising from agglomeration forces. 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 is driving the result. 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 changes in fixed effects of existing jobs $\Delta \log \hat{y}^{\mathrm{old}}$, i.e., the average change in fixed effects of jobs matched in both periods.[41] Note that this helps address concerns about other regional factors that may be influenced by the local workforce, such as shifts in the demand for local products, which may also affect firm productivity. Column (5) shows the result of my preferred specification with full controls. The coefficient of 1.045 implies that a 1 standard deviation increase in worker log productivity induces a 0.51 standard deviation increase in firm log productivity in the same location.

I provide further support for these findings with two additional results. First, I use a reduced-form approach and use changes in firm fixed effects $\Delta \log \hat{y}(\ell)$ as a dependent variable. The overall results remain largely unchanged: A 1 standard deviation increase in log worker productivity induces a 0.6 standard deviation increase in log firm productivity. Second, as a placebo test, I use the firm fixed effects of existing jobs, $\Delta \log \hat{y}^{\mathrm{old}}(\ell)$, as the dependent variable. In contrast to Table 3, the coefficients are statistically insignificant

[40] This coefficient is larger than that of OLS, which indicates that the attenuation bias due to measurement errors may be substantial. Additionally, it is well known that covariance of worker and firm fixed effects tends to biased downward, which may affect the estimation result.

[41] I use fixed effects rather than model-implied productivity because results are sensitive to the calibration of β for existing jobs.

and close to zero, which suggests that most changes in the productivity of new jobs arise from firm sorting rather than regional factors. All results are summarized in Section C.2.

Although the estimated coefficient is substantial, policy evaluation requires quantifying the relative contributions of each major mechanism driving spatial disparities. While a full decomposition is beyond the scope of this paper, I provide suggestive evidence on the quantitative importance of two-sided sorting in Section C.3. The analysis shows that worker-firm interactions, as measured in Table 3, account for over 60% of the magnitude predicted by the two-sided sorting model.

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 additional exogenous heterogeneity—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 performs 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.

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 (2023) "Firm Sorting and Spatial Inequality."

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.

Milgrom, Paul and Chris Shannon (1994) "Monotone Comparative Statics," *Econometrica: Journal of the Econometric Society*, 157–180.

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*.

Postel-Vinay, Fabien and Jean-Marc Robin (2002) "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 70 (6), 2295–2350.

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 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.

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, $S = V^e - V^u + V^p - V^v$,

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

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 $\delta_v V^v(y,\ell) = \frac{\delta_v}{\rho+\delta_v}q(\ell)(V^p(x(\ell),y,\ell) - V^v(y,\ell))$ 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 determining the outside option value. 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 surpluses, which can be expressed with the four terms,

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

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, as $x = x(\ell)$ 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 $x < B_w(\cdot)$, when x is much smaller than $x(\ell)$. Similarly, the value of a unit 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 job finding rates and vacancy contact rates as

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

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, it exhibits PAM, and then prove the existence.

PAM between workers and firms. To show that any equilibrium pure assignment features PAM, I proceed by contradiction. That is, 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 $A_w(\ell'') > A_w(\ell')$.[42] Given that firms are worse in ℓ'', for $A_w(\cdot)$ to be larger in ℓ'', 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 that $A_f(\ell') > A_f(\ell'')$. 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 $\mathcal{U}$ is supermodular and log-submodular in (x,ℓ), which yields

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

Since $\mathcal{U}(x(\ell'),\ell') < \mathcal{U}(x(\ell''),\ell')$, the above inequality implies that $\mathcal{U}(x(\ell'),\ell') < \mathcal{U}(x(\ell'),\ell'')$. Thus, 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 previous step, an 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 $z(\ell) = (x(\ell), y(\ell))$,

[42] To simplify the notation, I index the terms—$A_w(\cdot)$, $B_w(\cdot)$, $A_f(\cdot)$, $B_f(\cdot)$—with only ℓ in the proof unless it creates any confusions.

$z'(\ell) = (x'(\ell), y'(\ell))$, 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). Note that I have already plugged in housing and business services market clearing conditions (6).

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 conditions on $D_{z''}f(z, z', z'')$, I first obtain several useful expressions,

$$
\begin{aligned}
A_w'(\ell) &= A_w(\ell)\left(\varepsilon_\lambda(\theta)\frac{\theta'}{\theta}\frac{\tilde{\rho}}{\tilde{\rho} + \beta\lambda} + \frac{y'}{y - b}\right), && \text{(A.5)} \\
B_w'(\ell) &= B_w(\ell)\left(-\frac{-\tilde{\rho}\frac{\theta'}{\theta}\varepsilon_q(\theta) + \beta\lambda/\theta}{\tilde{\rho} + \beta\lambda(\theta) + (1 - \tilde{\beta})q} + \frac{x'}{x}\right), && \text{(A.6)} \\
A_f'(\ell) &= A_f(\ell)\left(\varepsilon_q(\theta)\frac{\theta'}{\theta}\frac{\tilde{\rho}}{\tilde{\rho} + (1 - \tilde{\beta})q} + \frac{x'}{x}\right), \\
B_f'(\ell) &= B_f(\ell)\left(\frac{\tilde{\rho}\varepsilon_\lambda(\theta)\frac{\theta'}{\theta} + (1 - \tilde{\beta})q/\theta}{\tilde{\rho} + \beta\lambda + (1 - \tilde{\beta})q} + \frac{y'}{y - b}\right), \\
\frac{\theta'(\ell)}{\theta(\ell)} &= \zeta(z, z')\left(\frac{N'}{N} - \frac{L'}{L}\right) \quad \text{where} \quad \zeta(z, z') = \frac{\lambda + \delta}{-\varepsilon_q(\theta)\lambda + \delta} > 0. && \text{(A.7)}
\end{aligned}
$$

Plugging the above expressions into f, 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 $a_w = \tilde{\rho}\zeta\frac{M_w q_w}{L}$, $a_f = \tilde{\rho}\zeta\frac{M_w q_f}{N}$, $b_w = \frac{A_w x}{\tilde{\rho} + \beta\lambda + (1 - \tilde{\beta})q}\left(\varepsilon_\lambda - \frac{(1 - \tilde{\beta})q}{\tilde{\rho} + \beta\lambda + (1 - \tilde{\beta})q}\varepsilon_q\right)$, 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 given by $\bar{h}^2 C_r'' \frac{\rho}{\delta_v} C_v'' + a_f b_f \bar{h}^2 C_r'' + a_w b_w \frac{\rho}{\delta_v} C_v'' > 0$, which is positive when cost functions are convex—i.e., $C_r'', C_v'' > 0$.

Step 1-(ii) 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 when $\min\{x', y'\} = 0$. In particular, I assign the limiting values of $g(\cdot)$ at these boundaries, i.e., $\lim_{\min\{x',y'\} \to 0} g(\cdot)$. To characterize these limit values, fix x, y, x', 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

$$[f_f] \quad x(y-b)\left(\frac{x'}{x} - \frac{\tilde{\rho}}{(1-\tilde{\beta})q_n}(-\varepsilon_q)\left(\frac{N_n'}{N_n} - \frac{L_n'}{L}\right) - \frac{\beta}{1-\tilde{\beta}}\right) - \frac{\rho}{\delta_v}C_v''(N_n)N_n' \to 0,$$

which implies $-\frac{1}{q_n}\left(\frac{N_n'}{N_n} - \frac{L_n'}{L}\right) - C_f C_v''(N_n)N_n' + D_f \to 0$ 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, and N_n', and thus L_n', should converge to zero. In sum, I conclude that $\lim_{y' \to 0} g(\cdot) = \left(-\frac{q_w'(x)(x')^2}{q_w(x)}, 0\right)$. Following a similar step, I can show that $\lim_{x' \to 0} g(\cdot) = \left(0, -\frac{q_f'(y)(y')^2}{q_f(y)}\right)$.

To construct $\tilde{g}(\cdot)$, I need to define two sets. First, I define a compact set, $Z \equiv [\underline{x}, \bar{x}] \times [\underline{y}, \bar{y}] \times [0, \bar{x}'] \times [0, \bar{y}']$. To find the upper bounds, $(\bar{x}', \bar{y}')$, find (i) $\bar{L}$ such that $\bar{h}C_r'(\bar{L}) = \underline{x}\underline{y} + \bar{h}C_r'(\bar{h}M_w)$, and (ii) $\bar{N}$ such that $C_v'(\bar{N}) = \frac{1-\beta}{\tilde{\rho}}q\left(\frac{M_f}{L}\right)\underline{x}(\underline{y} - b) + C_v'(M_f)$. These values will be used in Step 2 (ii). Using these values, I define $\bar{x}' = \frac{\bar{L}}{M_w q_w(\underline{x})}$ and $\bar{y}' = \frac{\bar{N}}{M_f q_f(\underline{y})}$. Then, define $(\bar{x}'', \bar{y}'') \equiv \sup_{(z,z') \in Z} g(\cdot)$, which are finite as $g(\cdot)$ is continuous and Z is compact. Second, I define $\bar{Z} \equiv [\underline{x}, \max\{\bar{x}, \underline{x} + (\bar{x}' + \bar{x}'')\}] \times [\underline{y}, \max\{\bar{y}, \underline{y} + (\bar{y}' + \bar{y}'')\}] \times [0, \bar{x}' + \bar{x}''] \times [0, \bar{y}' + \bar{y}'']$.

Finally, 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 such that each component corresponds to the closest value in $\bar{Z}$.

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

Importantly, for any initial values $(z(\underline{\ell}), z'(\underline{\ell})) \in Z$, the solution to the ODE problem $(x(\ell), y(\ell), x'(\ell), y'(\ell)) \in \bar{Z}$ 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 boundary does not bind, by the definition of $\bar{Z}$.

Step 2 The main purpose of this step is to find a solution of the ODE problem above that satisfies the two terminal boundary conditions, $x(1) = \bar{x}, y(1) = \bar{y}$.

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 from sorting conditions. I write $z'(0)$ to emphasize that the solution is obtained by using this initial point. The initial point $z(0)$ is always chosen to satisfy the boundary conditions, i.e., $z(0) = (\underline{x}, \underline{y})$.

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.

In Step 2 (i), I show that $\mathcal{P}_w(0, N) < 0$ for all N and $\mathcal{P}_f(L, 0) < 0$ for all L. In Step 2 (ii), I show that $\mathcal{P}_w(\bar{L}, N) > 0$ for all N and $\mathcal{P}_f(L, \bar{N}) > 0$ for all L. 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 $(\underline{x}, \underline{y}, \bar{x}', y'(0))$. Since the solution $(z(\ell), z'(\ell)) \in \bar{Z}$ for all ℓ, the sorting conditions for workers and firms are satisfied, which ensures that each worker and firm chooses ℓ that maximizes the 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 $\rho V^u(\underline{x}, 0) \geq \rho V^u(\underline{x}, \ell)$. Observe that $\rho V^u(\underline{x}, \ell) \geq b\underline{x} - \bar{h} C_r'(\bar{h} L(\ell))$ and $\rho V^u(\underline{x}, 0) \leq b\underline{x} + \underline{x}\underline{y} - \bar{h} C_r'(\bar{h}\bar{L})$, which implies

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

Thus, $L(\ell) \geq M_w$ for all ℓ, and $\mathcal{P}_w(\bar{L}, N) > 0$ 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 implies

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

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

Remark: Although for a given $(x'(0), y'(0))$, the equilibrium assignment is uniquely pinned down, it is still possible that there exist multiple equilibria. There can be multiple initial boundary conditions that satisfy $\mathcal{P}_w = \mathcal{P}_f = 0$.

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 gains from the labor market rise with ℓ, housing rents—and consequently, population density—also increase. 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 $\{1-\tilde{\beta}_n, \ell_n\}$ such that $1-\tilde{\beta}_n \to 0$, $r'(\ell_n;\tilde{\beta}_n) < 0$, $L'(\ell_n;\tilde{\beta}_n) < 0$, 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 case as $A'_w(\cdot) > 0$ under PAM. Given that $y'(\ell_n;\tilde{\beta}_n) > 0$, there exists $\bar{n}$ 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 $n \geq \bar{n}$. The firm-side sorting condition then gives

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

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. Therefore, 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 also strictly increase in ℓ if δ_v is sufficiently large. From (10), a sufficient condition for this is, for sufficiently small $1-\tilde{\beta}$,

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

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

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

The first term increases in ℓ as $A'_w(\ell) > 0$, and the second term increases in ℓ as $\lambda'(\ell) < 0$ and $y'(\ell) > 0$ when $1-\tilde{\beta}$ is close to zero.

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 a 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 $\mathbb{E}[x|\ell]$. In turn, the value functions have the same expression as in (4) and (5), except $x(\ell)$ and $y(\ell)$ are replaced with $\mathbb{E}[x|\ell]$ and $\mathbb{E}[y|\ell]$.

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 continuity of $Q_w(\cdot)$. Workers of x^* choose both locations, and thus they are indifferent between the two, i.e.,

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

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, $\lambda(\ell') < \lambda(\ell'')$ and $\mathbb{E}[y|\ell'] < \mathbb{E}[y|\ell'']$, leading to $A_w(\ell') < A(\ell'')$. Contradiction.

Second, suppose that there are firms of y^* that choose both $\ell' < \ell''$. From the same logic, firms' hiring opportunities are equal in two locations, i.e., $A_f(\ell') = A_f(\ell'')$. In turn, the condition under which firms of y^* are indifferent between ℓ' and ℓ'', which is analogous to the indifference condition of workers above, is given by

$$A_f(\ell')B_f(\ell') + c(\ell') = A_f(\ell'')B_f(\ell'') + c(\ell'').$$

This equality requires that $\theta(\ell'') < \theta(\ell')$ and $V(\ell'')/L(\ell'') < V(\ell')/L(\ell')$. To see this, suppose not. Since $A_w(\ell'') \geq A_w(\ell')$ to ensure that $\mathbb{E}[x|\ell]$ is increasing, $A_f B_f = \frac{\tilde{\rho}+\beta\lambda}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q} A_w$ is weakly greater in ℓ''. Moreover, the assumption on population density $L(\ell'') > L(\ell')$ requires that $V(\ell'') > V(\ell')$ and thus $c(\ell'') > c(\ell')$, and the above equality cannot hold. As a result, $q(\ell') < q(\ell'')$ in addition to $\mathbb{E}[x|\ell'] < \mathbb{E}[x|\ell'']$, which leads to $A_f(\ell') < A_f(\ell'')$. Contradiction.

A.4 Mobility of Workers and Firms

In this section, I formally discuss how the assumption of mobility affects the equilibrium. I show that equilibrium outcomes—e.g., assignment, wages—do not depend on whether workers and firms can move freely.

Proposition A.2. *The location choices of workers and firms $(x(\ell), y(\ell))$ are optimal in an economy in which they remain in the same location (no mobility) if and only if $(x(\ell), y(\ell))$ are optimal in an economy where they can move freely (free mobility). Moreover, wages $w(x(\ell), y(\ell), \ell)$ are identical along the equilibrium path.*

This proposition implies that the equilibrium characterized in the draft can be interpreted as outcomes from an economy with free mobility. The key idea is that as long as I focus on steady state, no workers or firms have incentives to move to another location, and thus assumption on mobility does not change equilibrium results (on the equilibrium path).

I first formally characterize the location choices of workers under two cases. I focus on the limit case where $\delta_v \to 0$ to save the notation. Consider **Problem 0** (P0), an economy with free mobility. I include subscript 0 to represent the outcomes in (P0). All assumptions remain the same, except the assumption that workers and firms can change locations at each point in time. For given allocations $(x(\ell), y(\ell))$, the values of workers (unemployed, employed) are characterized by

$$V_0^u(x,\ell) = \mathrm{d}t(bx - \bar{h}r(\ell)) + \lambda(\ell)\,\mathrm{d}t \cdot e^{-\rho\,\mathrm{d}t} V_0^e(x, y(\ell), \ell) + (1 - \lambda(\ell)\,\mathrm{d}t) e^{-\rho\,\mathrm{d}t} V_0^u(x, \ell_x),$$
$$V_0^e(x,y,\ell) = \mathrm{d}t(w_0(x,y,\ell) - \bar{h}r(\ell)) + (1 - \delta\,\mathrm{d}t) e^{-\rho\,\mathrm{d}t} V_0^e(x,y,\ell) + \delta\,\mathrm{d}t \cdot e^{-\rho\,\mathrm{d}t} V_0^u(x, \ell_x),$$

where $\mathrm{d}t$ represents a small interval of time, and $\ell_x = \mathrm{argmax}_\ell\, V_0^u(x,\ell)$ represents the optimal location choice of workers. I focus on the case $V_0^e(x, y(\ell), \ell) \geq V_0^u(x, \ell_x)$. The key difference from the baseline without mobility is that the continuation value depends on $V_0^u(x, \ell_x)$, rather than $v_0^u(x,\ell)$, reflecting the possibility of migration. The values of vacant and filled positions are

$$V_0^v(y,\ell) = q(\ell)\,\mathrm{d}t \cdot e^{-\rho\,\mathrm{d}t} V_0^p(x(\ell), y, \ell) + (1 - q(\ell)\,\mathrm{d}t) e^{-\rho\,\mathrm{d}t} V_0^v(y, \ell_y),$$
$$V_0^p(x,y,\ell) = \mathrm{d}t(xy - w_0(x,y,\ell)) + (1 - \delta\,\mathrm{d}t) e^{-\rho\,\mathrm{d}t} V_0^p(x,y,\ell) + \delta\,\mathrm{d}t \cdot e^{-\rho\,\mathrm{d}t} V_0^v(y, \ell_y).$$

where $\ell_y = \mathrm{argmax}_\ell\, V_0^v(y,\ell)$ denotes the optimal location choice of firms. I focus on the case $V_0^p(x(\ell), y, \ell) \geq V_0^v(y, \ell_y)$. The match surplus is defined as before: $S_0(x,y,\ell) \equiv V_0^e(x,\ell) - V_0^u(x,\ell_x) + V_0^p(x,y,\ell) - V_0^v(y,\ell_y)$.

Next, consider **Problem 1** (P1), the baseline economy without mobility described in the draft. I rewrite the key equations for comparison:

$$
\begin{aligned}
V^u(x,\ell) &= \mathrm{d}t(bx - \bar{h}r(\ell)) + \lambda(\ell)\,\mathrm{d}t e^{-\rho\,\mathrm{d}t}V^e(x,y(\ell),\ell) + (1-\lambda(\ell)\,\mathrm{d}t)e^{-\rho\,\mathrm{d}t}V^u(x,\ell)\\
V^e(x,y,\ell) &= \mathrm{d}t(w(x,y,\ell) - \bar{h}r(\ell)) + (1-\delta\,\mathrm{d}t)e^{-\rho\,\mathrm{d}t}V^e(x,y,\ell) + \delta\,\mathrm{d}t e^{-\rho\,\mathrm{d}t}V^u(x,\ell),\\
V^v(y,\ell) &= q(\ell)\,\mathrm{d}t\cdot e^{-\rho\,\mathrm{d}t}V^p(x(\ell),y,\ell) + (1-q(\ell)\,\mathrm{d}t)e^{-\rho\,\mathrm{d}t}V^v(y,\ell),\\
V^p(x,y,\ell) &= \mathrm{d}t(xy - w(x,y,\ell)) + (1-\delta\,\mathrm{d}t)e^{-\rho\,\mathrm{d}t}V^p(x,y,\ell) + \delta\,\mathrm{d}t e^{-\rho\,\mathrm{d}t}V^v(y,\ell).
\end{aligned}
$$

I first prove the property summarized in the below lemma, which is the key to the proposition.

Lemma A.1. *The values of unemployed workers in P1 (without mobility) are smaller than the maximum value of unemployed workers in P0 (free mobility), i.e., $V^u(x,\ell) \le V_0^u(x,\ell_x)$ for all (x,ℓ).*

Proof. I proceed by contradiction. Suppose, to the contrary, there exists (x,ℓ) such that $\Delta \equiv V^u(x,\ell) - V_0^u(x,\ell_x) > 0$.

Step 1 The solutions to Problem P0 and P1 are characterized by

$$e = \mathrm{d}t(w - \bar{h}r(\ell)) + \beta_4 e + \beta_5\hat{u}, \tag{e}$$

$$p = \mathrm{d}t(f - w) + \beta_4 p + \beta_5 v, \tag{p}$$

$$e = \hat{u} + \beta S, \tag{e2}$$

where $(e,\hat{u},w,p,S)$ differs between the two problems. Then, I solve for w as follows:

$$
\begin{aligned}
\mathrm{d}t(w - \bar{h}r(\ell)) &= (1-\beta_4-\beta_5)\hat{u} + (1-\beta_4)\beta S && \text{(e, e2)}\\
&= (1-\beta_4-\beta_5)\hat{u} + \beta\,\mathrm{d}t(f - \bar{h}r(\ell)) - \beta(1-\beta_4-\beta_5)(v+\hat{u})
\end{aligned}
$$

where the second line follows from $(1-\beta_4)(e+p) = \mathrm{d}t(f-\bar{h}r(\ell)) + \beta_5(v+\hat{u})$, and $S = e - \hat{u} + p - v = \mathrm{d}t(f-\bar{h}r(\ell)) + \beta_4(e+p) + \beta_5(v+\hat{u})$. Note that v can be treated as a constant in this case as it does not vary between problems. Therefore, $\mathrm{d}(\mathrm{d}tw) = (1-\beta)(1-\beta_4-\beta_5)\,\mathrm{d}\hat{u}$. Under this assumption, $w(x,\ell)\,\mathrm{d}t - w_0(x,\ell)\,\mathrm{d}t = (1-\beta)(1-\beta_4-\beta_5)\Delta = (1-\beta)(1-e^{-\rho\,\mathrm{d}t})\Delta$.

Step 2 Define the auxiliary values of workers who can move but (irrationally) stay in the same location, with a subscript $0a$. Because there is no commitment technology, workers and firms, when bargaining over wages, believe that workers would deviate to the optimal ℓ_x.

$$
\begin{aligned}
V_{0a}^u(x,\ell) &= \mathrm{d}t(bx - \bar{h}r(\ell)) + \lambda(\ell)\,\mathrm{d}t e^{-\rho\,\mathrm{d}t}V_{0a}^e(x,y(\ell),\ell) + (1-\lambda(\ell)\,\mathrm{d}t)e^{-\rho\,\mathrm{d}t}V_{0a}^u(x,\ell),\\
V_{0a}^e(x,\ell) &= \mathrm{d}t(w_0(x,y,\ell) - \bar{h}r(\ell)) + (1-\delta\,\mathrm{d}t)e^{-\rho\,\mathrm{d}t}V_{0a}^e(x,y,\ell) + \delta\,\mathrm{d}t e^{-\rho\,\mathrm{d}t}V_{0a}^u(x,\ell).
\end{aligned}
$$

Note that the surplus of this match remains the same, and in turn, the wage is equal to $w_0(x, y, \ell)$. Observe that the above values are characterized by the same equations as in P1, and the only difference lies in wages. For both problems, the relation between wages and the unemployed values is given by

$$\mathrm{d}V^u_{\cdot} = \square\,\mathrm{d}(\mathrm{d}tw_{\cdot}) \quad \text{where } \square \equiv \frac{\lambda(\ell)\,\mathrm{d}te^{-\rho\,\mathrm{d}t}}{(1-(1-\lambda(\ell)\,\mathrm{d}t)e^{-\rho\,\mathrm{d}t})(1-(1-\delta\,\mathrm{d}t)e^{-\rho\,\mathrm{d}t}) - \lambda(\ell)\,\mathrm{d}te^{-\rho\,\mathrm{d}t}\delta\,\mathrm{d}te^{-\rho\,\mathrm{d}t}},$$

for both (V^u_{0a}, w_0) and (V^u, w). Combining this with the result from Step 1,

$$V^u - V^u_{0a} = \square(w(x,\ell)\,\mathrm{d}t - w_0(x,\ell)\,\mathrm{d}t) = \square(1-\beta)(1-e^{-\rho\,\mathrm{d}t})\Delta.$$

Step 3 Note that $V^u_{0a}(x,\ell) \le V^u_0(x,\ell_x)$ as ℓ_x is the optimal location choice, and wages are the same in P0 and the auxiliary problem. Therefore, $V^u_{0a}(x,\ell) \le V^u_0(x,\ell_x) < V^u(x,\ell)$ leads to $1 < \square(1-\beta)(1-e^{-\rho\,\mathrm{d}t})$, which implies

$$0 > \delta\,\mathrm{d}te^{-\rho\,\mathrm{d}t} + (1-\delta\,\mathrm{d}t)e^{-2\rho\,\mathrm{d}t} > 0,$$

which leads to a contradiction. Thus, I conclude that $V^u_0(x,\ell_x) \ge V^u(x,\ell)$ for all (x,ℓ).[43] $\square$

Proof. I now prove Proposition A.2. The value of unemployed workers when choosing ℓ_x under P1 is equal to its counterpart under P0, i.e., $V^u(x,\ell_x) = V^u_0(x,\ell_x)$ on the equilibrium path. One can check this by plugging in $x = x(\ell)$ and observe that these values are characterized by the same set of equations. Plugging $x = x(\ell)$ and $y = y(\ell)$ into $S_0(x,y,\ell)$, the surplus on the equilibrium path is equal to that of P1.

$$S_0(x(\ell), y(\ell), \ell) = \frac{1}{\rho+\delta+\beta\lambda(\ell)+(1-\beta)q(\ell)}(x(\ell)y(\ell) - bx(\ell)) + o(1), \tag{A.8}$$

where $o(1)$ is the term that converges to 0 when $\mathrm{d}t \to 0$. As before, denoting $S_0(\ell) = S(x(\ell), y(\ell), \ell)$, plugging the above into $S_0(x,y,\ell)$ yields the following:

$$(\rho+\delta)S_0(x, y(\ell), \ell) = xy(\ell) - bx - \bar{h}(r(\ell) - r(\ell_x)) - \lambda(\ell_x)\beta S_0(\ell_x) - (1-\beta)q(\ell)S_0(\ell) + o(1),$$
$$(\rho+\delta)S_0(x(\ell), y, \ell) = x(\ell)y - bx(\ell) - \beta\lambda(\ell)S_0(\ell) - q(\ell_y)(1-\beta)S_0(\ell_y) + o(1).$$

Using the above results, the wage of workers choosing ℓ is given by

$$w_0(x,y,\ell) = (1-\beta)bx + (1-\beta)\bar{h}(r(\ell)-r(\ell_x)) + \beta xy + \beta\lambda(\ell_x)(1-\beta)S(\ell_x) - \beta(1-\beta)q(\ell_y)S(\ell_y) + o(1).$$

[43] In fact, I can further show that $V^u_0(x,\ell) \ge V^u(x,\ell)$ for all (x,ℓ). From $V^u_0(x,\ell_x)$ and $V^u(x,\ell)$ and the result from Step 1, $w_0(x,y(\ell),\ell) \ge w(x,y(\ell),\ell)$. Because the continuation values of unemployment and wages are both larger in P0, the value of unemployed workers is higher in P0.

The difference from the baseline is the option values of workers and firms are evaluated at their optimal locations, ℓ_x and ℓ_y, instead of their current location ℓ. In equilibrium, the above simplifies to

$$w_0(x(\ell), y(\ell), \ell) = (1-\beta)bx(\ell) + \beta x(\ell)y(\ell) + \beta(1-\beta)(\lambda(\ell) - q(\ell))S_0(x(\ell), y(\ell), \ell),$$

which coincides with (10) when (A.8) is substituted.

Combining this with the lemma, the solutions to P0 and that of P1 coincide. Similarly, one can show the symmetric result for the firm side. □

Remark. Workers choose the optimal location that maximizes the following:

$$\begin{aligned} V_0^u(x,\ell) &= \mathrm{d}t(bx - \bar{h}r(\ell)) + \mathrm{d}t e^{-\rho\,\mathrm{d}t}\beta\lambda(\ell)S(x,\ell) + e^{-\rho\,\mathrm{d}t}V_0^u(x,\ell_x) \\ &= \mathrm{d}t(bx - \bar{h}r(\ell)) + \frac{\mathrm{d}t}{\tilde{\rho}}e^{-\rho\,\mathrm{d}t}\beta\lambda(\ell)(y(\ell)-b)x - \frac{\mathrm{d}t}{\tilde{\rho}}e^{-\rho\,\mathrm{d}t}[\beta\lambda(\ell)(1-\beta)q(\ell)S(\ell) + \beta\lambda(\ell)\beta\lambda(\ell_x)S(\ell_x)] \\ &\quad - \frac{\mathrm{d}t}{\tilde{\rho}}e^{-\rho\,\mathrm{d}t}\beta\lambda(\ell)\bar{h}(r(\ell) - r(\ell_x)) + e^{-\rho\,\mathrm{d}t}V_0^u(x,\ell_x). \end{aligned} \tag{A.9}$$

Equation (A.9) is the counterpart to (4), which depends not only on (x,ℓ), but also on the worker's optimal location ℓ_x (specifically, $S(\ell_x)$ and $r(\ell_x)$). Due to this difference, compared to (4) which exhibits single-crossing property in $(x, A_w(\ell))$, the above equation is more difficult to directly understand the complementarity in worker's location choices, which is why I present P1 in the main draft.

A.5 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:

$$\begin{aligned} \max_{\bar{m}_w(\ell|x),\bar{m}_f(\ell|y)} \quad & \int_0^1 \left[\mathbb{E}[x|\ell]\,\mathbb{E}[y|\ell](1-u(\ell))L(\ell) - C_r(\bar{h}L(\ell)) - C_v(N(\ell))\right]\mathrm{d}\ell \\ \text{s.t.} \quad & \int_0^\ell \frac{L(\ell')}{M_w}\,\mathrm{d}\ell' = \int_{\underline{x}}^{\overline{x}} \bar{m}_w(\ell|x)q_w(x)\,\mathrm{d}x, \int_0^\ell \frac{N(\ell')}{M_f}\,\mathrm{d}\ell' = \int_{\underline{y}}^{\overline{y}} \bar{m}_f(\ell|y)q_f(x)\,\mathrm{d}y, u(\ell) = \frac{\delta}{\delta+\lambda(\ell)}, \quad \forall \ell \end{aligned} \tag{A.10}$$

where $\mathbb{E}[x|\ell]$ and $\mathbb{E}[y|\ell]$ are the average productivity of workers and firms assigned to location ℓ.

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 the population density across ℓ, with the objective of maximizing (A.10). In other words, a new worker assignment $\bar{m}_{w1}(\ell|x)$ must satisfy $\int_{\underline{x}}^{\overline{x}} \bar{m}_{w1}(\ell|x) q_w(x)\,\mathrm{d}x = \int_0^{\ell} \frac{L(\ell')}{M_w}\,\mathrm{d}\ell'$. 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 $(1-u(\ell))\,\mathbb{E}[y|\ell]L(\ell)$ 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 from the previous paragraph, 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[44], $x(\ell)$ 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, $\theta(\ell) = \frac{V(\ell)}{u(\ell)L(\ell)}$, where $V(\ell) = N(\ell)$ in steady state.

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

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

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

where $\varepsilon_\lambda(\theta) = \frac{\lambda'(\theta)}{\lambda(\theta)}\theta$. Similarly, first-order conditions of firm allocation are given by,

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

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

[44] I omit the superscript $(*)$, which denotes the optimal assignment, in the proof while retaining it in the main draft.

Finally, the constraint on the unemployment rate 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

$$\begin{aligned} C_r''\bar{h}^2 L' &= \frac{(1-u)(1-\varepsilon_\lambda)}{1-\varepsilon_\lambda(1-u)}y'x - \frac{\varepsilon_\lambda u(1-u)}{1-\varepsilon_\lambda(1-u)}x'y - \frac{(1-\varepsilon_\lambda)xy}{(1-\varepsilon_\lambda(1-u))^2}u' - \frac{u(1-u)xy}{(1-\varepsilon_\lambda(1-u))^2}\frac{\partial \varepsilon_\lambda}{\partial \ell} \\ &= \frac{(1-u)(1-\varepsilon_\lambda)}{1-\varepsilon_\lambda(1-u)}y'x - \frac{\varepsilon_\lambda u(1-u)}{1-\varepsilon_\lambda(1-u)}x'y + \frac{u(1-u)xy}{(1-\varepsilon_\lambda(1-u))^2}\left((1-\varepsilon_\lambda)\varepsilon_\lambda - \frac{\lambda''}{\lambda'}\theta + \varepsilon_\lambda - 1\right)\frac{\theta'}{\theta}, \end{aligned} \tag{A.11}$$

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 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

$$C_v''N' = -\frac{(1-\varepsilon_\lambda)(1-u)}{1-\varepsilon_\lambda(1-u)}xy'\frac{L}{N} + \frac{\varepsilon_\lambda(1-u)u}{1-\varepsilon_\lambda(1-u)}\frac{L}{N}x'y - \frac{u(1-u)xy}{(1-\varepsilon_\lambda(1-u))^2}\left((1-\varepsilon_\lambda)\varepsilon_\lambda - \frac{\lambda''}{\lambda'}\theta + \varepsilon_\lambda - 1\right)\frac{L}{N}\frac{\theta'}{\theta} \tag{A.12}$$

Multiply (A.11) by $L(\ell)$ and (A.12) by $N(\ell)$ respectively, and then sum them up to obtain the first equation that characterizes the optimal assignment,

$$C_r''(\bar{h}L(\ell))\bar{h}^2 L'(\ell)L(\ell) + C_v''(N(\ell))N'(\ell)N(\ell) = 0, \tag{A.13}$$

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.11) and then plugging in (A.13), I obtain the second equation,

$$\begin{aligned} &\frac{(1-u)}{1-\varepsilon_\lambda(1-u)}\left((1-\varepsilon_\lambda)\frac{y'}{y} - \varepsilon_\lambda u \frac{x'}{x}\right)xy \\ &= \left[C_r''\bar{h}^2 L + \frac{\varepsilon_\lambda u(1-u)}{(1-\varepsilon_\lambda(1-u))^3}\left(\varepsilon_\lambda(1-\varepsilon_\lambda) - \frac{\lambda''\theta}{\lambda'} + \varepsilon_\lambda - 1\right)xy\left(1 + \frac{C_r''}{C_v''}\bar{h}^2\left(\frac{L}{N}\right)^2\right)\right]\frac{L'}{L} \end{aligned} \tag{A.14}$$

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.13) and (A.14), together with boundary conditions.

Because housing and business services cost functions are convex, the condition (A.13) implies $L'(\ell)N'(\ell) \leq 0$. In addition, the second condition (A.14) provides the condition under which the optimal population density increases. The planner optimally chooses to increase population density while decreasing firm density when the below inequality holds at the optimal allocation,

$$(1-\varepsilon_\lambda(\theta(\ell)))\frac{y'(\ell)}{y(\ell)} > \varepsilon_\lambda(\theta(\ell))u(\ell)\frac{x'(\ell)}{x(\ell)}$$

if $\varepsilon_\lambda(1-\varepsilon_\lambda) - \frac{\lambda''\theta}{\lambda'} + \varepsilon_\lambda - 1 \geq 0$. For example, this condition holds if ε_λ is constant because $1-\varepsilon_\lambda + \frac{\lambda''\theta}{\lambda'} = 0$.

Proof of Proposition 3. Worker and firm sorting conditions, $f_w = f_f = 0$ in Section A.2, are given by

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

$$[f_f] \quad \frac{\partial}{\partial \ell}\left(\frac{\delta_v}{\rho+\delta_v}x(\ell)\frac{(1-\beta)q(\ell)}{\tilde{\rho}+(1-\tilde{\beta})q(\ell)}\left(y - b - \frac{\beta\lambda}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}(y(\ell)-b)\right)\right) = c'(\ell).$$

To evaluate the efficiency, I compute the worker sorting condition multiplied by $L(\ell)$ and the firm sorting condition multiplied by $N(\ell)$. 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 $x(\ell)(y(\ell)-b)$),

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

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

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

Combining the above expressions with the remaining terms, the sorting conditions, $f_w = f_f = 0$, are given by:

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

When ρ converges to 0, which leads to $1-\tilde{\beta} = \frac{\rho}{\rho+\delta_v}(1-\beta) \to 0$, the above expressions simplify to

$$L\bar{h}r' = \frac{\beta\lambda}{\delta+\beta\lambda}xy'L + \frac{\delta}{\delta+\beta\lambda}\frac{\beta\lambda}{\delta+\beta\lambda}\varepsilon_\lambda\frac{\theta'}{\theta}x(y-b)L, \tag{A.15}$$
$$Nc' = \frac{N(1-\beta)q}{\delta+\beta\lambda}x'(y-b) - \frac{\beta\lambda}{\delta+\beta\lambda}(1-\beta)(1-u)xy'L + \frac{\delta(1-u)(1-\beta)}{\delta+\beta\lambda}\left(\frac{\delta}{\delta+\beta\lambda}\varepsilon_\lambda - 1\right)\frac{\theta'}{\theta}x(y-b)L, \tag{A.16}$$

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

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

$$\bar{h}^2 C_r''(\bar{h}L(\ell))L(\ell)L'(\ell) + C_v''(N(\ell))N'(\ell)N(\ell) > 0, \tag{A.17}$$

which violates (A.13). I consider four cases, depending on the sign of $L'(\ell)$ and $N'(\ell)$. First, if both $L'(\ell)0$ and $N'(\ell)$ are strictly positive, (A.17) immediately follows. Next, consider the opposite case, $L'(\ell) \leq 0$ and $N'(\ell) \leq 0$. From (A.15), $\theta' < 0$. Moreover, adding (A.15) multiplied with $(1-\beta)(1-u)$ and (A.16),

$$(1-\beta)(1-u)\bar{h}^2 C_r'' LL' + C_v'' N'N = N\frac{(1-\beta)q}{\delta+\beta\lambda}x'(y-b) + \frac{\delta(1-\beta)(1-u)}{\delta+\beta\lambda}(\varepsilon_\lambda - 1)\frac{\theta'}{\theta}x(y-b)L. \tag{A.18}$$

Then, the LHS is negative from the assumption while the RHS 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.18),

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

The final case, $L'(\ell) \leq 0$ and $N'(\ell) > 0$, implies $\theta'(\ell) > 0$, which contradicts (A.15) and cannot be an equilibrium.

This conclusion does not depend on the assumption of the matching elasticity, and the inefficiency cannot be resolved by the Hosios condition. Under the Hosios condition, i.e., $\varepsilon_\lambda(\theta(\ell)) = 1-\beta$, the combined impact of the market tightness in the sorting conditions cancels out. In particular, the sum of the last terms in (A.15) and (A.16) becomes zero. However, the equilibrium remains inefficient due to worker and firm sorting. The remaining terms are given by

$$\bar{h}^2 C_r'' LL' + VC_v'' = V(1-\beta)q\frac{\delta}{\delta+\beta\lambda}x'(y-b) + \frac{\beta\lambda}{\delta+\beta\lambda}x(y')L > 0,$$

which confirms the conclusion of (A.17). Notably, when workers and firms are homogeneous, the above expression becomes zero, aligning with (A.13).

The planner can correct externalities by using spatial transfers. To compute spatial transfers, I will compare the worker and firm sorting conditions to the counterparts of the planner. Rearrange (A.15), I obtain

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

Comparing the above to the planner's solution, I conclude that the following spatial transfer can equalize the worker sorting condition to the planner's solution (A.11),

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

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 the market tightness externalities is standard, which is well studied under homogeneous agents. In addition, under heterogeneous agents sorting across markets, this condition additionally ensures that there are no externalities from heterogeneous firm types from worker sorting. Workers internalize the impact of firm heterogeneity because the potential match surplus of these workers depends on the productivity of local firms. The Hoisos condition ensures that this marginal benefit is equal to the marginal benefit evaluated from the planner's perspective. In turn, the spatial transfer simplifies to $t_w(\ell)$ in Section 3.2. Even under the Hosios condition, externalities arise from the negative impact of workers on local firms, whose value depends on the average productivity of workers in the local market. This intuition can be confirmed by observing that the second term is equal to $-\frac{N}{L}\frac{\partial\rho\bar{V}^v}{\partial x}x'$.

Similarly, I will assess the firm sorting condition. Rearrange (A.16),

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

Analogous to the above discussion, by comparing the above to the planner's solution (A.12), I obtain the spatial transfer for firms.

$$
\begin{aligned}
t_f(\ell) = t_f^0 + \int_0^\ell \Bigg[& (1-u)\left(\frac{(1-\beta)\beta(1-u)}{u+\beta(1-u)} - \frac{1-\varepsilon_\lambda}{1-\varepsilon_\lambda(1-u)}\right)xy' + (1-u)u\left(\frac{\varepsilon_\lambda}{1-\varepsilon_\lambda(1-u)}y - \frac{(1-\beta)}{u+\beta(1-u)}(y-b)\right)x' \\
& + \frac{u(1-u)}{(1-\varepsilon_\lambda(1-u))^2}\left(\frac{\lambda''}{\lambda'}\theta - \varepsilon_\lambda + 1 - \varepsilon_\lambda(1-\varepsilon_\lambda)\right)xy\frac{\theta'}{\theta} \\
& - \frac{u(1-u)(1-\beta)}{(u+\beta(1-u))^2}(u\varepsilon_\lambda - u - \beta(1-u))x(y-b)\frac{\theta'}{\theta}\Bigg]\frac{L}{N}\,\mathrm{d}t,
\end{aligned} \tag{A.20}
$$

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 Section 3.2. 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 the higher firm productivity in local labor markets increases the threat point of workers in wage bargaining, leading to a decrease in their values. This bargaining channel is merely the transfers between workers and firms, and thus should not be considered from the planner's perspective. The firm spatial transfer confirms that the former is always larger than the latter, so firms choose higher-ℓ than the optimal level.

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 the values of unemployed workers, where the weight $a(x)$ depends on productivity x,[45]

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

subject to constraints in the original problem, with $\int_{\underline{x}}^{\overline{x}} a(x)\,\mathrm{d}Q_w(x) = 1$. 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—and 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 output. For notational simplicity, I omit ℓ from this point onward. The Hamiltonian of this problem is given by

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

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

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

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

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

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

[45] I focus on the values of unemployed workers to align the planner's problem with equilibrium, as workers select locations ℓ to maximize their unemployment values.

By differentiating the conditions (L) and (N) w.r.t. ℓ, and plugging in the conditions (x) and (y) respectively, I obtain the following results,

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

Multiplying the above equations by L and N, respectively, and summing them yields a modified version of (A.13),

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

In contrast to (A.13), this expression is not necessarily zero. The planner may prefer concentration in high ℓ while reducing congestion in low ℓ, for example, if the weight is decreasing, i.e., $a'(x(\ell)) < 0$, favoring lower-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 RHS 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 RHS 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.13), and the right-hand side should be zero. Contradiction.

A.6 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 $x(\ell)$, $\bar{m}(\ell|x(\ell_0)) = \mathbb{1}\{\ell \geq \ell_0\}$. 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 $\bar{m}(\ell|x') = \bar{m}(\ell|x'') = 1$. Unlike pure assignment, which would impose restrictions on population density in this discrete economy, the assignment represented by $\bar{m}(\ell|x)$ allows multiple types to locate in a single location ℓ, or a single type to locate in multiple locations.

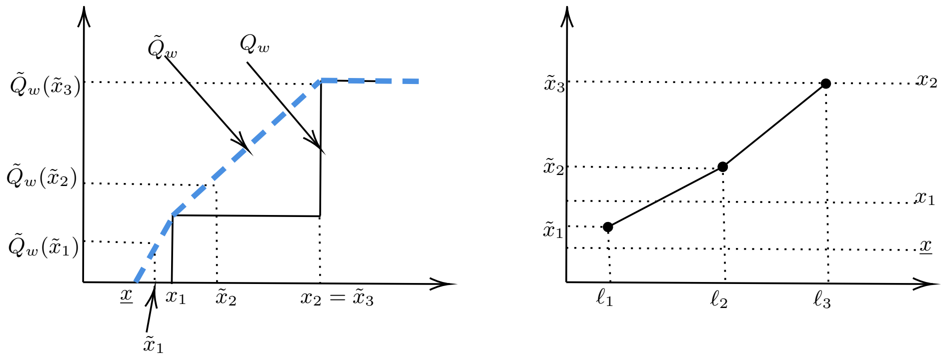


Figure A.1. Measure-Preserving Constraint

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 CDF Q_w, with a positive measure for each type. Choose x_0 which is smaller than x_1, implying $Q_w(x_0) = 0$. Also, let $\bar{m}(\ell|x_0) = 1$ for all ℓ to simplify the notation.

Consider *a finite number of locations* $\{\ell_1, ..., \ell_N\}$, where $\ell_n = n/N$. Without loss of generality, assume that the land distribution is uniform with total measure of 1, i.e., the measure of land in each location ℓ_n is given by a number $\ell_n - \ell_{n-1}$. As locations are ex ante homogeneous, the distribution is not identifiable, and the location index can be reformulated so that the distribution of land is uniform.

I will introduce two auxiliary functions, $\tilde{Q}_w$ and $\tilde{x}_n$. First, define a *(strictly) increasing* function $\tilde{Q}_w$ on $[\underline{x}, \bar{x}]$ such that (1) $\tilde{Q}_w = Q_w$ for all x_i, and (2) it is linearly increasing in (x_{i-1}, x_i) for all i. Next, to define an *increasing* function $\tilde{x}_n$ on $n \in \{1, 2, ...N\}$, I first define $j(n) \equiv \max\{j|\bar{m}(\ell_n|x_j) = 1\}$, i.e., the best type among workers who locate only in $\{\ell_1, ..., \ell_n\}$. Using this notation, 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 the original functions, Q_w and x_n, and the auxiliary functions, $\tilde{Q}_w$ and $\tilde{x}_n$, are related.

The total measure of workers from ℓ_1 to ℓ_n is $\tilde{Q}_w(\tilde{x}_n)$: All workers of types $x_1, ..., x_{j(n)}$ and a fraction $\bar{m}(\ell_n|x_{j(n)+1})$ of $x_{j(n)+1}$ workers are included. Thus, $\sum_{k=1}^{n} L(\ell_k) = Q_w(x_{j(n)}) + \bar{m}(\ell_n|x_{j(n)+1})(Q_w(x_{j(n)+1}) - Q_w(x_{j(n)})) = \tilde{Q}_w(\tilde{x}_n)$, 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 $x \in [\underline{x}, \bar{x}]$. Moreover, I can find the cutoff types $x(\ell_n) \in [\underline{x}, \bar{x}]$ such that all workers of $x \leq x(\ell_n)$ choose $\ell \leq \ell_n$, and all others choose $\ell > \ell_n$. By the continuity of $Q_w(x)$, $\tilde{x}_n = x_{j(n)} = x(\ell_n)$ is strictly increasing in n. In other words, workers of $x \in (x(\ell_{i-1}), x(\ell_i)]$ sort into location ℓ_i, and workers of a given type choose the same location with probability 1, which yields $x(\ell_n) = Q_w^{-1}(\ell_n)$. Thus, population density becomes

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

<u>*Step 3*</u> 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.7 Neoclassical Local Labor Markets: Proof of Proposition 4

In this section, I assume that there is no search friction and that each local labor market is competitive. I focus on a differentiable pure assignment equilibrium, in which the assignment and wages are differentiable.

I first show that workers and firms positively sort across space. 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, $\bar{V}^v(y,\ell) = x(\ell)y - w(x(\ell),\ell) - c(\ell)$. 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, I obtain

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

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

I will 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 a location ℓ such that $L(\ell') > N(\ell')$. Then, there must exist another location 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 assignment $(x(\ell), y(\ell))$, for local labor market clearing, it is optimal for firms of $y(\ell)$ to hire $x(\ell)$.

$$\frac{\partial}{\partial x}(xy(\ell) - w(x,\ell)) = 0 \Rightarrow y(\ell) = \frac{\partial w(x(\ell),\ell)}{\partial x}. \tag{A.21}$$

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.21) 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.8 Local Labor Market with Directed Search

In this section, I consider the directed or 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 directed search in each local labor market. 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.[46]

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

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

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

[46] 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.

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 the worker's 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

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

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,

$$\frac{f_{xy}f}{f_x f_y} \geq \sup_{\theta} \frac{\lambda'(\theta)(\lambda'(\theta)\theta - \lambda(\theta))}{\theta\lambda(\theta)\lambda''(\theta)}. \tag{A.22}$$

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.3. *Suppose that* (A.22) *holds. If a differentiable pure assignment equilibrium exists, then the equilibrium exhibits the following properties:*

1 Positive assortative matching (PAM) between workers and firms obtains in 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 profit of each firm type are equal to those of an economy with a single, nationwide labor market.

A.9 Alternative Mechanisms Accounting for Spatial Disparities: 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 the schedules of $r(\ell) = C_r'(\bar{h}L(\ell))$ and $c(\ell) = C_v'(N(\ell))$ that ensure the sorting conditions of workers and firms, which correspond to value equalization across ℓ. The *spillovers* model can be treated analogously.

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 PAM allocation, i.e., positive sorting among workers, firms, and locations. In other words, I focus on $x(\ell)$, $y(\ell)$, and $\bar{A}(\ell)$ that 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.5. 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.5, 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,

$$(1-u(\ell))x(\ell)y(\ell)\bar{A}'(\ell)A^x(x(\ell))L(\ell) = C_r''(\bar{h}L(\ell))\bar{h}^2 L'(\ell)L(\ell) + C_v''(N(\ell))N'(\ell)N(\ell). \tag{A.23}$$

On the one hand, the planner allocates more workers and firms in higher ℓ with higher exogenous location productivity $\bar{A}(\ell)$. On the other hand, other factors (workers, firms, knowledge spillovers) 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 planner increases population density L relatively more than that of firm N when firm heterogeneity y' is significantly larger than that of workers x', as in the baseline.

Efficiencies of the equilibrium. Derivations of the sorting conditions are explained in detail in Section A.5. To avoid repetition, I omit the intermediate steps and present the final expressions. When ρ goes to zero, the sorting conditions are given by

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

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:

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

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. The final term in $t_w(\ell)$ corrects externalities of workers on local TFP through spillovers. For example, in the *no sorting* model, $x'(\ell) = y'(\ell) = 0$, and the spatial transfers vanish.

B. Quantitative Analysis

B.1 Quantitative Model

Values. With a Stone-Geary utility function, income taxes, and housing regulations, the value of unemployed and employed workers change 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 $T_r(\ell)$ is locally the 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 solution. Consider a parallel time for bargaining between a worker and a firm, where the firm makes the first offer. The 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 the worker and firm 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 the preferences of each player are represented by $\delta_i^t v_i$ for some $0 < \delta_i < 1$ for $i = w, f$. In this setting, I have $v_f = A - w$, and $v_w = B + r^{-\omega}(1-\tau_w)w$ where $A = xy - q(V^p - V^v) - \delta_v V^v$ and $B = -r^{-\omega}bx - \lambda(V^e - V^u)$. It follows that $g(v_f) = B + r^{-\omega}(1-\tau_w)(A - v_f)$, which is decreasing and weakly concave.

Thus, a subgame perfect equilibrium is characterized by v_f^* and v_f' such that $\delta_f v_f^* = v_f'$ and $g(v_f^*) = \delta_w g(v_f')$. Applying this result and defining $\beta = \frac{1-\delta_f}{1-\delta_f\delta_w}$, the bargaining leads to the surpluses:

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

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 elasticity by running the following MSA-level regression:

$$
\gamma(\ell) = \gamma + 0.28\log(3 + \mathrm{WRI}(\ell)) + \Gamma_\gamma X_\gamma(\ell) + \varepsilon(\ell), \tag{A.24}
$$

where $\gamma(\ell)$ is the inverse of $\eta_w(\ell)$, i.e., the price elasticity, and $\mathrm{WRI}(\ell)$ is the local regulation index. I microfound the above relation by introducing taxes on housing production costs, $C_r(\cdot)$. Specifically, I assume that the tax rate depends on the total housing quantity, $\tau(H;\ell) = (H/T)^{t_h(\ell)} - 1$. Here, the term $t_h(\ell)$ quantifies the stringency of housing regulations. The constant T is a scaling parameter chosen so that τ is unit-free.[47]

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

$$
r(H,\ell) = \frac{1+\gamma+t_h(\ell)}{1+\gamma}\left(\frac{H}{H_w}\right)^{\gamma}\left(\frac{H}{T}\right)^{t_h(\ell)}. \tag{A.25}
$$

This functional form is consistent with (A.24), 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, $t_h(\ell) = 0.28[\log(3 + \mathrm{WRI}(\ell)) - \log(3 + \mathrm{WRI}(\underline{\ell}))]$.

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

The housing supply can be written as

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

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

$$r(\ell)H(r(\ell);\ell) = (1-\omega)\bar{h}L(\ell) + \omega((1-\tau_w(\ell))(u(\ell)bx(\ell) + (1-u(\ell))w(x(\ell),y(\ell),\ell)) + \Pi + T_r(\ell))L(\ell). \quad \text{(A.26)}$$

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 the local average housing share using raw labor income and rents. Next, I target the average spending shares on housing 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 infer location-specific tax rates and assess their impact on spatial disparities. 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 avoid any distortions caused by changes in the wage 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 regulation. I borrow estimates on differential housing supply elasticities across MSAs from Saiz (2010). In particular, he estimates how these 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).

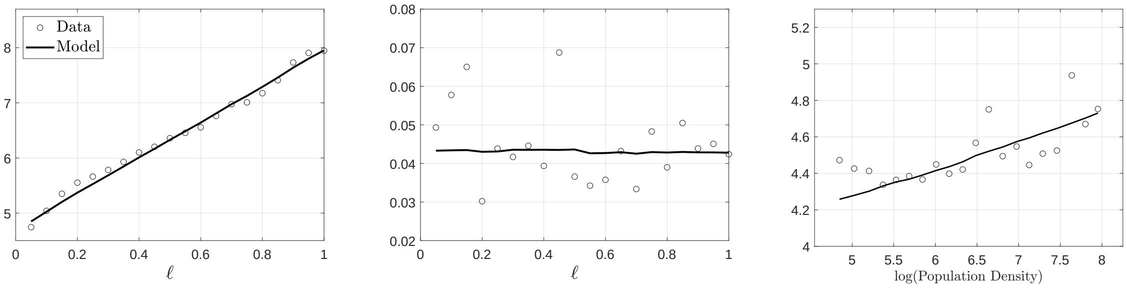
B.3 Estimation: Intuition for Identification

Housing market. Preferences for housing $(\bar{h}, w)$ can be obtained from the aggregate average housing spending share and its variation across regions. 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 the spending shares of all regions. I target average spending shares on housing in the U.S. and the ratio of spending shares between the first and last quartile to pin down $(\bar{h}, w)$.

Next, the housing supply elasticity, η_w, and the housing tax parameter, T, are informed by rents, housing demand, and housing regulations across locations. Equation (A.25) reveals that a faster increase in housing rents due to higher housing demand leads to a smaller η_w, and a larger effect of housing regulations on housing rents indicates a smaller T. I target the rent ratio 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 can 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 in (5) reveals how firms balance gains from the labor market, such as higher worker productivity, with increased overhead costs. Given worker and firm productivity, I can back out overhead costs that satisfy the sorting conditions of firms, which then pin down (H_f, η_f). A large increase in worker productivity leads to a significant increase in implied overhead costs, which indicates a higher H_f. If this increase is particularly pronounced in regions in which a measure of vacancy is significant, an implied housing supply elasticity η_f will be smaller.



(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

B.4 Estimation Result

Model fit. In addition to moments presented in Table 2, I also target population density and unemployment rates across locations. Figure A.2a shows the fit of these two sets of moments. In Figure A.2b, I present the fit of log GDP per capita, which is non-targeted. It is not surprising that the model replicates the variation in GDP per capita, since I target the cross-section of wages.

Sources of spatial disparity. Using the estimated model, 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. I compare these results to those of Card et al. (2023), who run a two-way fixed-effects regression of log wages on workers and CZ, using matched employer-employee data from the U.S. As explained in Section 5, their worker and CZ fixed effects can be mapped to $x(\ell)$ and $w(\ell)/x(\ell)$ in my model, respectively. This allows for a meaningful comparison between the two decomposition exercises. When they regress the estimated fixed effects on log population density, they obtain the coefficients of 0.04 and 0.034, which suggests that workers account for about 54% of spatial wage variation.

Despite the methodological differences—structural vs. reduced form—the overall conclusion is consistent. Worker heterogeneity is the primary driver of spatial wage disparity, although other factors remain quantitatively substantial. Differences in magnitude may arise from variations in geographical definitions, sample periods, or control variables. They may also reflect other factors that affect the sorting of workers such as amenities.

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.

Table A.1: Parameter Values

	Parameter	Target	Value
ρ	discount rate	interest rate	0.003
α	matching elasticity	literature	0.5
δ	separation rate	EU transition rate (LIAB)	0.0178
$\mathcal{A}$	matching efficiency	market tightness	0.49
$\bar{h}, \omega$	housing demand	spending shares on housing	0.80, 0.01
η_w	housing supply	housing rents	9.13
b	unemployment benefit	replacement rate	0.35
β	worker's bargaining power	labor share	0.05
H_f, η_f	business services supply	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.

Data. I primarily rely on regional statistics from the German Federal Statistical Office (GFSO). I obtain the district-level data for the year 2017: the average wage, unemployment rates, GDP per capita, and population density for each district. The average wages are computed based on the compensation of employees, which includes gross wages and salaries as well as 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). Housing rents are obtained from Ahlfeldt et al. (2023), who predicted housing prices for the centroids of postal codes using residential rental rates. 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 two variables—population density 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.

Table A.2: Model Fit

	Wage			Housing Share		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. Housing share difference is the average housing spending share between 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.

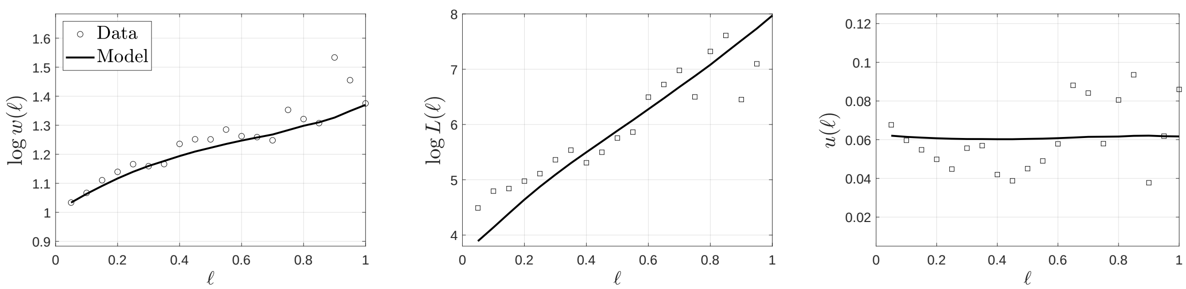


Figure A.3. Model Fit: Wage, Population Density, Unemployment rate

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.

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.

B.6 Policy Evaluation

Decomposition of welfare changes. To investigate the sources of welfare changes, I decompose the total changes in welfare into three components: firm productivity, housing market, and labor market tightness. Specifically, I compute the changes in *each* worker's value due to the changes in each component using (13), which are plotted in Figure 6. I then average across workers to obtain the aggregate contribution of each component.

Let $\hat{F}(x)$ denote the value of a variable F encountered by a worker of x in a new steady state, and define $\check{F}(y)$ analogously for firms. First, to isolate the impact of firm productivity, I compute $\hat{y}(x)$. Since firm productivity also affects workers through redistribution, I construct a new transfer, $\hat{\Pi}^y$, which includes tax revenues evaluated at $\hat{y}(x)$ and firm profits evaluated at $\check{x}(y)$. I then calculate (13) using $\hat{y}(x(\ell))$ and $\hat{\Pi}^y$, holding all other inputs constant. Second, to

Figure A.4. Welfare Impact of Relaxing Housing Regulation (% Change)

assess the role of housing markets, I incorporate changes in housing rents $\hat{r}(\ell)$ and transfers $\hat{T}_r(x)$, along with a new transfer $\hat{\Pi}^r$ that reflects changes in firm profits computed using $\check{c}(y)$, landowner profits, and intermediary profits. Third, for tightness, I compute $\hat{\theta}(x)$, and construct $\hat{\Pi}^{\theta}$, which incorporates changes in tax revenues computed using $\hat{\theta}(x)$ and firm profits computed using $\check{\theta}(y)$.

Results are summarized in Figure A.4. The primary source of the welfare decline lies in housing markets. By contrast, the contributions from firm productivity and tightness are modest. Intuitively, reallocating better firms to some workers necessarily worsens matches for others, and its aggregate gains are limited. Similarly, changes in tightness have a limited impact on aggregate welfare.

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. However, overhead costs $c^{\text{one}}(\ell)$ are calibrated differently to ensure that *homogeneous* firms are indifferent across ℓ,

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

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.

I decompose welfare changes into three components: *local TFP*, housing market, and labor market tightness. I consider changes in local TFP $\hat{A}(x)$—a local TPF encountered by a worker of x in a new steady state—instead of $\hat{y}(x)$. For computing associated $\hat{\Pi}^A$, I incorporate changes in firms' profits based on $\check{A}(q)$, which represents the local TFP of q-th quantile firm ordered by ℓ. All other steps follow the same procedure as before.

Federal income tax. Even when spatial policies are uniform nationwide, they 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

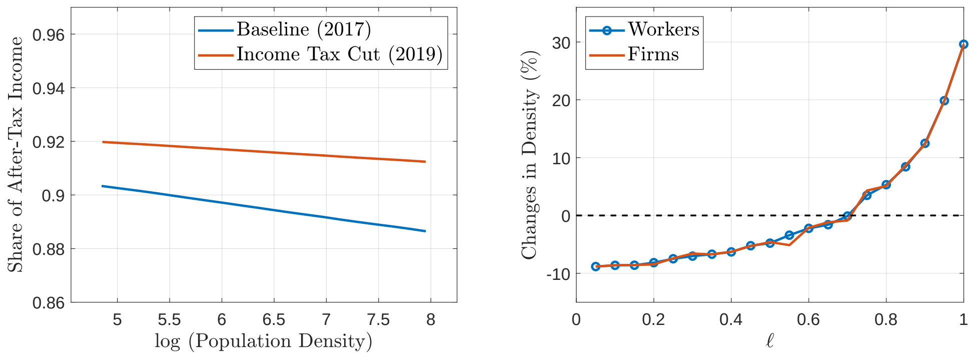


Figure A.5. Federal Income Tax Across Regions

Figure A.6. Welfare Impact of Tax Cut (% Change)

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.

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.5. Before the reform, after-tax income shares are lower for workers with higher incomes living in dense cities. 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.[48]

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.

[48] Changes in tax rates also affect welfare. However, the aggregate impact is negligible, because they merely redistribute income across workers.

C. Empirical Evidence of Two-Sided Sorting

C.1 Estimation of Worker and Firm Productivity

I continue to assume $\frac{\rho}{\rho+\delta_v} = 0$ as in Section 4.2, which implies

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

where $\hat{x}$ and $\hat{y}$ denote worker and firm fixed effects, respectively.[49] Local job finding rate $\lambda(\ell)$ is directly observable, 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 $\hat{y}(\ell)$. This in turn allows me to compute the labor share as:

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

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

The dispersion of local productivity among both workers and firms is substantial. The standard deviations of local averages of 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 finding rates.

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 1 standard deviation increase in the quality of the local workforce results in a 0.6 standard deviation increase in log firm productivity in the

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

Table A.3: Changes in Local Firm Fixed Effects

	(1) OLS	(2) IV	(3) IV	(4) IV	(5) IV
$\Delta \log x(\ell)$	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)
		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. Each observation is weighted by the number of workers.

same location. This finding aligns with the conclusion in Table 3.[50] 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 (s.e. 0.238).

Second, I conduct a robustness check by regressing changes in the fixed effects of *existing* jobs, that is jobs created in $t = 1$, $\Delta \log \hat{y}^{\text{old}}(\ell)$. 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 region-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 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 region-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 two-way fixed effects estimated using two different approaches. In Figure A.7, I plot a binned scatter and a fitted line of estimated worker and firm fixed effects across space, alongside their counterparts implied by the calibrated model in Section B.5. The two estimation

[50] To reach this conclusion, I first compute the variation in firm fixed effects arising from differences in firm productivity, by controlling for local job finding rate. I then compute the dispersion of firm productivity based on the residualized firm fixed effects.

Table A.4: Changes in Local Firm Fixed Effects of Existing Jobs

	(1) OLS	(2) IV	(3) IV	(4) IV
$\Delta \log x(\ell)$	0.012	0.155	0.094	-0.027
	(0.040)	(0.132)	(0.128)	(0.195)
		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. Each observation is weighted by the number of workers. The dependent variable is the change in firm fixed effects of existing jobs.

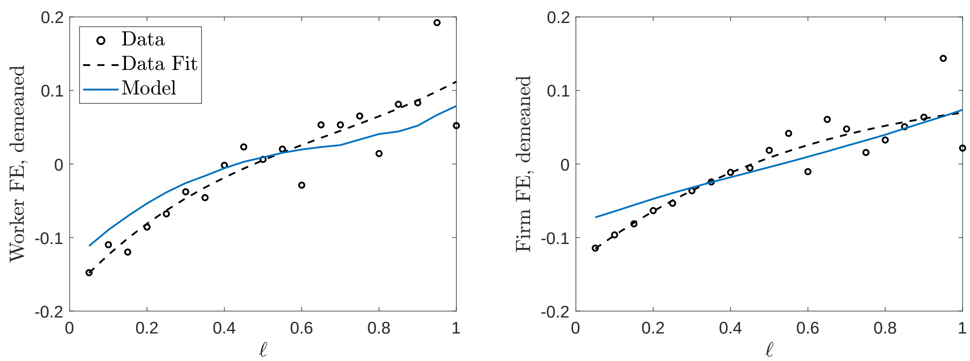


Figure A.7. Workers and Firms AKM Fixed Effects: Data vs. Model

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 German population. The fitted line is estimated using a third-order polynomial regression.

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 cross-section differences in wages and housing rents.

Overall, the estimated fixed effects (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 compute the endogenous changes in firm sorting that satisfy the sorting condition of firms and the measure-preserving constraint, which yields a new steady-state distribution. By comparing the two distributions, I obtain the changes in worker productivity, firm productivity, and job finding 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.[51]

On average, I find that a 1 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. The remaining 40% may stem from alternative mechanisms such as 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 two-way fixed effects estimation and a migration instrument in Section 5. Importantly, this model yields the same wage equation, and the prediction on the impact of worker sorting on firm sorting remains the same.

Model. I extend the model in several dimensions. I keep focusing 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, where the continuous F_ε has large support, ensuring that the probability of drawing a sufficiently large $\varepsilon(\ell)$ is positive. I assume that the 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, but they lose their current jobs. With the introduction of preference shocks, a major departure from the baseline is that an equilibrium is no longer characterized by a pure assignment. Workers with low productivity can choose a high ℓ if a preference shock is sufficiently large.

Migration is costly, and when a worker moves from ℓ' to ℓ, her value is discounted by a factor of $m_{\ell',\ell}^{-1}$. For example, these costs may arise from the distance between two regions (Schwartz, 1973). It is less costly for workers to migrate to nearby areas, and migration networks are stronger between geographically proximate regions. Given these two components, the value of a worker moving from ℓ' to ℓ is $\varepsilon(\ell)m_{\ell',\ell}^{-1}$ times the value of searching for a job in location ℓ, $V^u(x,\ell)$, which I describe below.

[51] Another potential approach is to develop a fully dynamic model that incorporates additional mechanisms and to estimate this model by matching the results in Table 3. The 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. However, I take a simpler approach that relies on the parsimonious baseline model.

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 $V^m(x,\ell) = \mathbb{E}[\max_{\ell'}\{m^{-1}_{\ell,\ell'}\varepsilon(\ell')V^u(x,\ell')\}]$. 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 $\tilde{\rho}$,

$$\log w(x,y,\ell) = \log x + \log\left(b + (1-\beta)\frac{\beta\lambda(\ell)}{\tilde{\rho} + \beta\lambda(\ell)}(y(\ell) - b) + \beta(y - b)\right). \tag{A.27}$$

It is important to note that the wage is 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 for two-way fixed-effects wage regression.

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

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

Observe that firm's value exhibits single crossing in $A_f(\cdot)$ and y. Thus, I have the following prediction.

Lemma A.2 (Prediction on firm sorting). *If a location ℓ experiences an exogenous increase in worker productivity $\mathbb{E}[x|\ell]$, holding the vacancy contact rate constant, then it will attract more productive firms.*

The location decision of a worker with type x in ℓ' is characterized by the probability $m(\ell|x,\ell')$ that she chooses a location in the interval $[0,\ell]$. 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 $\bar{m}(\ell|x,\ell') = 1\{x(\ell) \geq x\}$. I do not introduce any frictions to firm location decisions, and thus a firm assignment remains pure and can be characterized by a function $y(\ell)$.

Migration network. Workers' location decisions and their productivity distribution together determine the distribution of workers across locations. Let $\mu_\ell(x)$ denote the measure of workers of productivity x who choose 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 $m(\ell|x)$. Then, in steady state, I obtain $\mu_\ell(x) = m(\ell|x)M_w q_w(x)$.

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.

Due to preference shocks $\varepsilon(\ell)$, there are positive flows between locations. The migration network can be characterized by location choice probabilities and the measure of workers. The probability that workers are from ℓ' conditional on arriving at ℓ, $\Pr(\ell' \to \ell|\ell)$, are given by

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

The intensity of migration flows between a pair of regions can differ due to the migration costs $m_{\ell',\ell}$. If $m_{\ell',\ell}$ is smaller, then more workers from ℓ' will move to ℓ. It also depends on the values of other regions k for workers from ℓ'. If other regions are less attractive than ℓ, a larger share of workers move to ℓ. In particular, if more workers from ℓ' decide to leave, the number of out-migrants increases, leading to a higher share of migrants from ℓ', $\Pr(\ell' \to \ell|\ell)$.

The average productivity of migrants arriving at ℓ is given by

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

Workers from ℓ' tend to be more productive if the average productivity of workers moving out of ℓ', $\mathbb{E}[x|\ell' \to k, k \neq \ell']$, is higher. Hence, if a location with higher worker productivity sends out more workers (leading to a higher $\Pr(\ell' \to \ell|\ell)$), or if a location that is closely connected via the migration network—i.e., has a higher $\Pr(\ell' \to \ell|\ell)$—experiences an increase in worker productivity (leading to a higher $\mathbb{E}[x|\ell' \to \ell]$), then the average productivity of migrants is likely to increase, thereby raising the quality of the local search pool.