Spatial Sorting of Workers and Firms*

Ryungha Oh†
Northwestern University

September 2024

Abstract

Why do productive workers and firms locate together in dense cities? I develop a new theory of two-sided sorting in which both heterogeneous workers and firms sort across space. The location choices of workers and firms affect each other and endogenously generate spatial disparities in the presence of three essential forces: complementarity between worker and firm productivity, random matching within frictional local labor markets, and congestion costs. I demonstrate that the decentralized equilibrium exhibits excessive concentration of workers and firms, and dispersing them away from dense locations can mitigate congestion without reducing output. I then provide direct empirical evidence of the two-sided sorting mechanism using German administrative microdata. An exogenous increase in the quality of the workforce in a location results in more productive firms choosing that location. Finally, to quantify the implications of the model, I calibrate it to U.S. regional data and show that policies that relocate workers and firms toward less dense areas can increase welfare.

Keywords: Two-sided sorting, local labor markets, spatial inequality, spatial policies

*I am greatly indebted to my advisors, Giuseppe Moscarini, Ilse Lindenlaub, and Costas Arkolakis, for their guidance and support. I also thank 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, and participants at the Yale Macro Lunch, Yale International and Spatial Economics Lunch, SED 2023, UEA 2023 Toronto, The Federal Reserve Bank of Minneapolis Junior Scholar Conference, PSU New Faces in International Economics Conference, and the NBER Summer Institute 2024 for helpful comments. Data access was provided via on-site use at the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB) and subsequently remote data access.

†*Email:* ryungha.oh@northwestern.edu

1. Introduction

In the U.S., the share of college graduates is 40% in the top decile of Metropolitan Statistical Areas (MSAs) ranked by population density and only 26% in the bottom decile (ACS, 2017). In addition, almost half of Fortune 500 companies are headquartered in just 10 large cities (Fortune, 2022). This concentration of productive workers and firms results in spatial disparities that motivate various government policies, such as place-based policies and local regulations, which are designed to encourage or restrict the inflow of workers and firms. For policy evaluation, it is essential to consider the location choices of both workers and firms jointly, because the two are mutually dependent. For example, the effectiveness of government subsidies to firms in low-income regions hinges on the number and quality of workers attracted to these areas.[1]

In this paper, I propose a new theory of *two-sided sorting*: Workers and firms with different productivity sort across space, mutually incentivizing each other's sorting, and endogenously generate spatial disparities. Interaction between the location choices of workers and firms provides new insights for policies. Consider Austin, TX—a rising tech hub—as an example. The inflow of productive companies attracts a skilled workforce. Similarly, the presence of productive workers is one of the main reasons why high-tech firms are drawn to this area. If workers and firms are productive due to their own inherent quality, this relocation does not necessarily change aggregate output. However, if they come from more concentrated areas such as Silicon Valley, the aggregate costs due to congestion in dense areas could be reduced, leading to welfare gains. Empirically, I provide direct evidence of the two-sided sorting mechanism and quantify the implications of the mechanism by calibrating the model and evaluating real-world policies.

I begin the analysis by developing a parsimonious theory of spatial sorting of *heterogeneous* workers and *heterogeneous* firms across ex ante *homogeneous* locations. Their location decisions are interdependent, since they interact in local markets. In each local labor market, workers and firms randomly match subject to search frictions. Importantly, due to complementarity between their productivities, more productive workers produce more at the margin when they are matched with more productive firms, and vice versa. In addition, the number of workers and firms in each location is endogenously determined, and the costs of living or operating a business are higher in denser locations.

[1] Both the sorting of workers (e.g., Behrens, Duranton and Robert-Nicoud, 2014; Davis and Dingel, 2020; Martellini, 2022) and the sorting of firms (e.g., Bilal, 2023; Gaubert, 2018; Lindenlaub, Oh and Peters, 2023) have been recognized as significant sources of spatial inequality. However, the sorting of workers and firms has been separately investigated.

I show that more productive workers and more productive firms together choose denser areas. Due to random matching within each local market, labor markets with either better firms or a larger number of firms (relative to the number of workers) are more attractive to all workers. Thus, these locations attract a greater number of workers until the local congestion in the form of high housing rents outweighs the benefits from local labor markets. In particular, due to worker-firm complementarity, more productive workers benefit most from these labor markets, and thus are willing to pay higher housing rents. I show that the equilibrium exhibits positive assortative matching (PAM) between workers and firms across space. In principle, locations can be attractive to workers even with less productive firms if the number of firms is large enough. However, this possibility is ruled out in equilibrium. The decisions of less productive firms imply that local labor markets in this area are unattractive to firms. As a result, higher congestion costs due to a larger number of firms render these locations undesirable for any firm, which contradicts the firm sorting conditions. Importantly, the sorting of workers sustains the sorting of firms, and vice versa, which shows that two-sided sorting alone—without the presence of 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 features excessive concentration in dense areas. The key insight is that productivity is embodied in workers and firms, which emphasizes *who* produces rather than *where* the production occurs. Thus, there is no need to concentrate workers and firms in denser areas, since this only raises aggregate congestion costs. However, in the decentralized equilibrium, less productive workers overvalue the benefits of choosing denser locations more than a planner does. They do not internalize their negative impact on local firms that could have hired more productive workers when had they 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. The government can restore efficiency by taxing workers and firms in dense areas, which relocates them to less congested areas.

Before I calibrate the model, I present evidence of the two-sided sorting mechanism using German employer-employee matched data. Specifically, I test the main prediction that workers and firms interact with each other in their location choices. First, I use the model to recover the productivity of workers and firms in each location from two-way fixed-effects estimates obtained from a wage regression. In particular, I take into account that wages depend not only on the worker-firm productivity of a given match, but also on local labor market conditions—the average productivity of firms and job finding rates. I then instrument changes in local average worker productivity using predicted changes in incoming domestic migrant productivity, which is computed by interacting migration networks in the pre-period and contemporaneous shocks to other locations. This instrument addresses potential endogeneity concerns arising from changes in regional economic conditions such as local TFP growth that may attract more productive workers and firms at the same time. I further address concerns regarding alternative explanations, such as agglomeration forces. Unlike my mechanism, other mechanisms typically affect the productivity of not only new jobs but also preexisting jobs. Therefore, I control productivity changes in preexisting jobs. My analysis suggests that a 1 standard deviation increase in the productivity of the local workforce attracts firms that are about 0.5 standard deviation more productive to the same location. This finding supports the view that the location decisions of firms respond to those of workers, which highlihgts the importance of two-sided sorting as a driver of spatial disparities.

To quantitatively analyze how considering two-sided sorting can impact policy evaluation, 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), as well as housing rents. The key challenge to identification is to separately estimate the productivity of workers and firms across locations from the observed wages, which reflect both. First, to identify the magnitude of worker sorting, I rely on their location decisions—i.e., revealed preferences. Nominal wages in denser areas are higher due to either the compensation for higher housing rents or the skill premium. Controlling for housing rents in the data, I estimate worker productivity differentials. Second, the remaining spatial gap in wages is attributed to firm sorting. My estimation results reveal that the spatial heterogeneity of workers and firms are 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 those who are in the bottom decile, respectively.

Finally, I evaluate two real-world policies, which affect the location decisions of workers and firms. I begin by showing that relaxing housing regulations in denser cities is expected to have a negative impact on welfare by causing dense regions to be more congested. In the U.S., housing regulation is more stringent in dense cities, which reduces housing supply elasticity in these areas (Saiz, 2010). Therefore, relaxing regulation results in a significant inflow of workers into denser cities, and firms also move to these regions following workers. This substantial relocation further amplifies the concentration of economic activity in urban areas, which thus exacerbates congestion costs. However, because both workers and firms move at the same time, total output remains virtually unchanged, and welfare decreases as a consequence. Similarly, I find that the federal income tax cuts for workers result in a decrease in weflare by increasing the inflow of workers into urban areas.

To highlight the distinct implications of two-sided sorting, I compare these results with those from a model that lacks heterogeneity in either workers or firms. For example, what if firms are homogeneous, and productivity is embodied in locations rather than firms, so that it does not respond to policies? I show that the same policies have opposite effects on welfare. After the change in housing regulation, since a larger number of workers produce in more productive locations, output and weflare increase significantly. Similarly, the federal income tax cuts lead to a welfare gain without firm sorting. These comparisons highlight the fact that although different mechanisms may succeed in explaining spatial disparities in the data, incorporating two-sided sorting leads to qualitatively different policy implications.

Related literature. A large literature on spatial disparities focuses on the spatial sorting of either heterogeneous workers or heterogeneous firms across ex ante heterogeneous locations. Heterogeneous types of workers and firms value heterogeneous location fundamentals differently, and this affects 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, Ottaviano and Puga, 2023; Diamond, 2016). Other work finds that the spatial sorting of firms also plays an important role (e.g., Bilal, 2023; Lindenlaub, Oh and Peters, 2023). Another body of literature shows that sorting can happen across ex ante homogeneous or symmetric locations, via agglomeration forces that lead to complementarity between agent types and endogenous city characteristics (e.g., Davis and Dingel, 2019; Behrens, Duranton and Robert-Nicoud, 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, I am the first to show that two-sided sorting *alone* can endogenously generate dense areas populated by productive workers and firms.

There are a few studies that analyze 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, Manovskii and Qiu, 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, they are a driving force in generating differential population densities across space.

I also build on the literature on two-sided matching. Complementarity in payoffs has been the key source of assortative matching in competitive markets (e.g., Becker, 1973); frictional markets (e.g., Shimer and Smith, 2000; Eeckhout and Kircher, 2010); and game theoretic environments (e.g., Roth and Sotomayor, 1990). While complementarity between worker and firm productivity in output remains crucial for achieving PAM, this paper has two important differences. First, assortative matching is realized through location decisions. A worker's value depends not only on her employer, but also on other workers and firms within the same location who interact through local markets. In this context, locations serve as platforms on which these matches occur. Second, in my model, I allow the density of workers and firms to be endogenously determined 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.[2]

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. Examples include federal income taxes, stricter housing regulations in larger cities, and differences in state taxes (e.g, Albouy, 2009; Fajgelbaum et al., 2019; Hsieh and Moretti, 2019). Other studies claim that a decentralized equilibrium can often be inefficient and require policy interventions. For example, when agglomeration forces exist across heterogeneous workers, spatial transfers can increase welfare (e.g., Fajgelbaum and Gaubert, 2020; Rossi-Hansberg, Sarte and Schwartzman, 2019). They 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, it demonstrates that dispersion in marginal labor productivity does not necessarily imply misallocation in the presence of sorting. Second, it reveals that the sorting of heterogeneous workers and firms into frictional labor markets generates externalities, which suggests the possibility of improving welfare through spatial policies. This result is related to the discussion whereby externalities arising from search frictions cannot be canceled out by the Hosios condition when agents are heterogeneous (e.g., Acemoglu, 2001; Bilal, 2023; Shimer and Smith, 2001).

[2] More general matching models account for externalities, or they endogenize the number of agents. However, these models often make specific assumptions about values, which renders them not directly applicable to my context. For example, externalities arise from aggregate variables—such as economy-wide pollution—or values depend only on the types of agents, and are assumed to be independent of the number of agents. Notably, when the worker and firm density of each location is exogenous, I can apply the results of Demange and Gale (1985) or Roth and Sotomayor (1990).

The rest of the paper is organized as follows. Section 2 presents the model and solves for an equilibrium. Section 3 characterizes its properties and discusses the efficiency. Section 4 provides empirical evidence of two-sided sorting, and Section 5 discusses the identification strategy and presents estimation results. Finally, I evaluate policies in Section 6 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 I assume 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 $r_t(\ell)$. There is no utility from leisure, and workers inelastically supply one unit of labor. Workers discount the future at rate ρ. At time 0, workers choose locations in which to reside and work, and cannot migrate.[3] If workers are employed, they earn a flow wage. If workers are unemployed, they receive unemployment benefit bx,[4] 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 $y \in [\underline{y}, \overline{y}]$, drawn from the cdf $Q_f(\cdot)$, which is also assumed to be twice continuously differentiable. Firms choose a location to operate a business. Once firms post a vacancy, this job stays in the same location. At each point in time, they demand a unit of local business services at overhead cost $c_t(\ell)$. These costs are associated with renting commercial spaces, handling administrative procedures, advertising, or posting vacancies.

[3] This assumption is without loss of generality to characterize the steady-state equilibrium allocation of workers and firms and wages across locations. In Appendix A.3, 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 the mobility.

[4] This assumption is widely used in the macro-labor literature for tractability (e.g., Postel-Vinay and Robin, 2002).

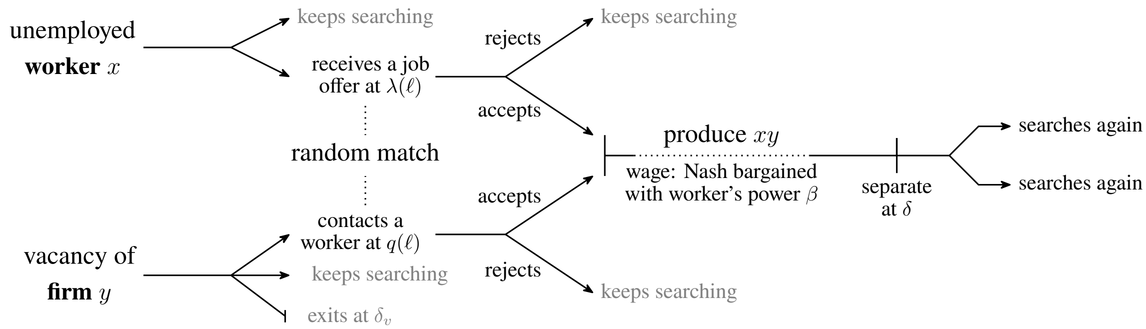


Figure 1. Local Labor Market: Timeline

Technology. Workers and firms 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. Each firm posts one vacancy at each point in time.[5] Posting an additional vacancy is prohibitively costly. At rate δ_v, a vacancy gets destroyed. At rate $q_t(\ell)$, a vacancy contacts workers, and if the match is acceptable, production begins. This match gets separated at rate δ and becomes a vacated position that reenters the search market.

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 measured by the efficiency unit. I assume that when a measure $N_t(\ell)$ of firms choose location ℓ, a measure $\delta_v N(\ell)$ of efficiency unit of vacancies are created at each point in time.[6] Thus, the effective measure of vacancies $V(\ell)$ equals δ_v times the number of vacancies posted. The matching function $M(\cdot)$ is assumed to be increasing and concave in both of its arguments. 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 $q_t(\ell) = q(\theta_t(\ell))$. I assume these functions are differentiable, and both $\frac{\lambda'(\theta)}{\lambda(\theta)}\theta$ and $\frac{q'(\theta)}{q(\theta)}\theta$ are bounded.

[5] The production unit is a match between a vacancy and a worker. Thus, the notion of a firm is simply a collection of vacancies and filled positions. Throughout the paper, I use the terms vacancies or jobs to denote individual positions, depending on whether it is unmatched or matched, respectively.

[6] Later, I consider a limit case in which vacancy destruction rate δ_v is sufficiently large. However, increasing destruction rate δ_v mechanically decreases the total number of vacancies. To avoid this problem, as a normalization, I assume that a single vacancy posting becomes more efficient in generating matches by introducing a notion of efficiency units of vacancies.

Local suppliers and ownership. Housing $H_t(\ell)$ is competitively supplied by landowners, and the costs of supplying H unit of housing are given by $C_r(H)$. In a business services market, competitive intermediaries provide services $S_t(\ell)$ for business operations, and it costs them $C_v(S)$ to provide units S of service. I assume that $C_r(\cdot)$ and $C_v(\cdot)$ are twice continuously differentiable, increasing, and convex. I further assume that they satisfy the Inada condition, $\lim_{H\to 0} C_r''(H) = \lim_{S\to 0} C_v''(S) = \infty$. 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.

In terms of the spatial sorting of workers and firms, I focus on the class of *pure assignments* between productivity x (y, respectively) and location ℓ. In other words, any two workers (firms, respectively) of the same productivity are assigned to the same location, and any two workers (firms, respectively) in the same location have the same productivity. Although many equilibria are not pure—e.g., assignments in which workers and firms randomly choose the location—they are not the focus of this paper. Mixed equilibria cannot explains the empirical pattern that we see in the data. In particular, in Proposition A.1, I show that any equilibrium allocation with more productive workers and firms choosing densely populated locations should be pure. Moreover, in Section 3.2, I show that the optimal assignment is pure.

When an assignment is pure, I can define a *location-matching function* $x(\ell)$ ($y(\ell)$, respectively),[7] which denotes the productivity of workers (firms, respectively) in local labor market ℓ. Because locations are ex ante homogeneous, without loss of generality, I label ℓ in such a way that $x(\ell)$ is strictly increasing. Importantly, since housing is elastically supplied in each location, a measure of workers who choose ℓ is endogenously determined as an equilibrium outcome, which is different from the standard assignment problem of Becker (1973). For example, consider the two assignment candidates, $x_1(\ell)$ and $x_2(\ell)$, in the left panel of Figure 2. The assignment $x_2(\ell)$ indicates that higher-ℓ locations supply more housing and become denser, compared with $x_1(\ell)$. The same measure of workers, who have productivity higher than x^*, are concentrated in $[\ell_2, 1]$ under $x_2(\ell)$ and spread out in $[\ell_1, 1]$ under $x_1(\ell)$. In contrast, if housing supply were inelastic, $x(\ell)$ would be uniquely pinned down by the housing market clearing condition. Also, note that I do not impose any

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

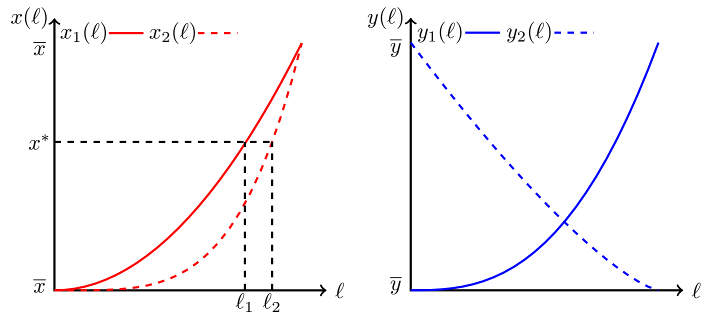


Figure 2. Location-Matching Functions: Examples

restrictions on $y(\ell)$ ex ante. Thus, it could either increase or decrease, as in the case of $y_1(\ell)$ or $y_2(\ell)$ in the right panel, or it could even be nonmonotonic.

Values of workers and firms. As a worker rents $\bar{h}$ units of housing and then uses the remaining income to purchase tradable goods, the flow indirect utility is given by $I - \bar{h}r(\ell)$, where I denotes the flow income, which depends on her employment state. 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:

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

where $y(\ell)$ denotes the productivity of firms in a local labor market in the candidate equilibrium with pure assignment, 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.[8] 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 a firm of productivity y enjoys when operating in location ℓ, and let $V^p(x,y,\ell)$ denote the value of a job of productivity y matched with a worker of productivity x in location ℓ. A vacancy either contacts a worker at rate $q(\ell)$ or gets destroyed at rate δ_v. From a matched job,

[8] Since I focus on the class of pure assignments, there are only firms of productivity $y(\ell)$ in labor market ℓ. Thus, workers always accept job offers in equilibrium.

a firm earns a flow profit $xy - w(x, y, \ell)$ until the match separates at rate δ. These values solve

$$\rho V^v(y, \ell) = q(\ell) \max\{V^p(x(\ell), y, \ell) - V^v(y, \ell), 0\} - \delta_v V^v(y, \ell), \tag{1}$$
$$\rho V^p(x, y, \ell) = xy - w(x, y, \ell) - \delta[V^p(x, y, \ell) - V^v(y, \ell)],$$

where $x(\ell)$ denotes the productivity of unemployed workers in location ℓ.

Define the surplus of the match between a worker and a job in location ℓ 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 workers and firms receive constant shares of the surplus:

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

Appendix A.1 presents all derivations. A worker of productivity x matched with a firm of productivity y enjoys her reservation unemployment utility $V^u(x, \ell)$ plus a share β of the surplus, and the firm takes the remaining share of the surplus. Nash bargaining between workers and firms gives the following flow wage:

$$w(x, y, \ell) = (1 - \beta)bx + \beta xy + (1 - \beta)\beta\lambda(\ell) S(x, y(\ell), \ell) - \beta(1 - \tilde{\beta}) q(\ell) S(x(\ell), y, \ell), \tag{2}$$

where $1 - \tilde{\beta} \equiv \frac{\rho}{\rho + \delta_v}(1 - \beta)$. First, wages increase in the output of a given match, xy, and the unemployment benefit, bx. In addition, wages depend on location ℓ, which determines the threat points in the bargaining game: If the job arrival rate $\lambda(\ell)$ or the productivity of firms in the local labor market $y(\ell)$ is higher, unemployed workers' values from search, captured by $\beta\lambda(\ell)S(x, y(\ell), \ell)$, is higher, which leads to an increase in wages. In contrast, if the vacancy contact rate $q(\ell)$ or the productivity of workers in the local labor market $x(\ell)$ is higher, the value of vacancies, captured by $(1 - \tilde{\beta})q(\ell)S(x(\ell), y, \ell)$, is higher, which leads to a decrease in wages.

Finally, I can solve for the value of workers of productivity x when choosing location ℓ,

$$\rho V^u(x, \ell) = bx + \underbrace{\frac{\beta\lambda(\ell)}{\tilde{\rho} + \beta\lambda(\ell)}(y(\ell) - b)}_{:=A_w(y(\ell), \lambda(\ell))} \left(x - \underbrace{\frac{(1 - \tilde{\beta})q(\ell)}{\tilde{\rho} + \beta\lambda(\ell) + (1 - \tilde{\beta})q(\ell)} x(\ell)}_{:=B_w(x(\ell), q(\ell))} \right) + \Pi - \bar{h}r(\ell), \tag{3}$$

where $\tilde{\rho} \equiv \rho + \delta$. I summarize the marginal return to a worker's productivity from local labor markets as *job opportunities*, A_w, 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, which increases in the vacancy contact rate $q(\ell)$ and local worker productivity $x(\ell)$. The final term represents the local congestion costs associated with housing expenditure, which is higher when there is a larger measure of worker, and results in increased housing demand. Workers of productivity x choose the location that maximizes value $V^u(x,\ell)$, and this decision defines the location-matching function of workers $x(\ell)$.

Next, I solve for the value of firms of productivity y when operating location ℓ, in which they post δ_v effective units of vacancies and pay local overhead costs $c(\ell)$ in each point in time.

$$\begin{aligned}\rho\bar{V}^v(y,\ell) &\equiv \delta_v V^v(y,\ell) - c(\ell) \\ &= \frac{\delta_v}{\rho}\underbrace{\frac{(1-\tilde{\beta})q(\ell)}{\tilde{\rho}+(1-\tilde{\beta})q(\ell)}x(\ell)}_{:=A_f(x(\ell),q(\ell))}\left(y - b - \underbrace{\frac{\beta\lambda(\ell)}{\tilde{\rho}+\beta\lambda(\ell)+(1-\tilde{\beta})q(\ell)}(y(\ell)-b)}_{:=B_f(y(\ell),\lambda(\ell))}\right) - c(\ell). \end{aligned} \tag{4}$$

The value of firms has a structure similar to that of workers. I define the marginal return to firm's productivity from local labor markets as *hiring opportunities*, A_f, which increase in local worker productivity $x(\ell)$ and the vacancy contact rate $q(\ell)$. The term B_f represents the fact that firms share the surplus with workers. Firms take a smaller share of the surplus when workers have a higher value for job search 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 of productivity y choose the optimal location ℓ that maximizes the value $\bar{V}^v(y,\ell)$, and this choice defines the location-matching function of firms $y(\ell)$.

Density of workers and firms. The location choices of workers determine population density $L(\ell)$, which represents the measure of workers per unit land in ℓ. Similarly, the location decisions of firms determine firm density $N(\ell)$, which is the measure of firms that choose location ℓ. The measures of workers and firms choosing locations between 0 and ℓ equal the measures of their types choosing these locations:

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

Formally, population (firm) density is defined as the Radon-Nikodym derivate of the measure of workers (firms) with respect to the area of land. In particular, under my labeling of locations, the first equation simplifies to $\frac{1}{M_w}\int_0^{\ell} L(\ell')\,\mathrm{d}\ell' = Q_w(x(\ell))$. Then, population density becomes[9]

$$L(\ell) = M_w Q_w'(x(\ell))x'(\ell). \tag{5}$$

Importantly, population density depends on an endogenous location-matching function $x(\ell)$, and is thus flexibly determined.[10]

Local housing rents $r(\ell)$ are determined by housing market clearing, $r(\ell) = C_r'(\bar{h}L(\ell))$. Similarly, local overhead costs $c(\ell)$ are pinned down by market clearing for business services, $c(\ell) = C_v'(N(\ell))$.

Laws of motions. I consider the laws of motion of 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 efficiency units of vacancies. A flow of separated matches $\delta(1-u(\ell))L(\ell)$ reenters the search pool. Vacancies can either be matched to 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. Steady-state unemployment rates and the measure of vacancies are given by

$$u(\ell) = \frac{\delta}{\delta + \lambda(\ell)}, \quad V(\ell) = N(\ell). \tag{6}$$

Definition of steady-state equilibrium. A pure-assignment steady-state equilibrium consists of location-matching functions $(x(\ell), y(\ell))$, 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 the wage is

[9] In Appendix A.5, I present an alternative derivation of this formula. I first obtain population density in a finite-worker-productivity and finite-location economy, and show that it converges to the same formula as the numbers of worker types and locations approach infinity.

[10] In other words, for a given worker distribution $Q_w(\cdot)$, a different location-matching function $x(\ell)$ leads to a different population density $L(\ell)$. The assumption on housing supply elasticity is crucial for this property. When housing supply is inelastic, the distribution of land equals the housing supply. Thus, the housing market clearing condition, $Q_w(x(\ell)) = \ell$, uniquely pins down $x(\ell)$, and population density becomes identical across ℓ.

determined by Nash bargaining; markets for housing and business services clear; the flow-balance conditions hold; and workers and firms optimally choose locations.

3. Equilibrium Analysis

In this section, I first characterize the positive properties of the decentralized 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

I first discuss the existence and several properties of the equilibrium.

Proposition 1. *A pure-assignment equilibrium exists. Any pure-assignment equilibrium exhibits the following properties:*

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

(2) When δ_v is sufficiently large, population density $L(\cdot)$ increases in ℓ.

(3) When δ_v is sufficiently large, wages $w(x(\cdot), y(\cdot), \cdot)$ increase in ℓ.

Proposition 1 establishes that an equilibrium with PAM between workers and firms exists.[11] The value of workers in (3) satisfies a single-crossing condition in worker productivity x and job opportunities A_w, and firms' value in (4) satisfies a single-crossing condition in firm productivity y and hiring opportunities A_f. 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. An increasing $x(\ell)$ ($y(\ell)$, respectively) can account for increasing $A_f(x(\ell), q(\ell))$ ($A_w(y(\ell), \lambda(\ell))$, respectively), when differences in $q(\ell)$ ($\lambda(\ell)$, respectively) are relatively small. This shows that supermodularity of the output function can lead to worker-firm complementarity in the spatial sorting problem, and in turn lead to PAM between workers and firms.

The challenge of showing PAM arises from the presence of search frictions, which generates differences in contact rates, $\lambda(\ell)$ and $q(\ell)$. In principle, job opportunities can be higher in higher-ℓ locations even when

[11] Note that the proposition does not guarantee uniqueness. Even among pure assignments, there can be multiple equilibria.

firm productivity $y(\ell)$ is lower if job arrival rates $\lambda(\ell)$ are sufficiently high, which breaks PAM. However, the introduction of congestion forces helps to resolve this tension. Higher job arrival rates require a larger measure of firms, which increases overhead costs. At the same time, hiring opportunities should be smaller in these locations, given that they are chosen by less productive firms by monotone comparative statics. Therefore, firms would prefer to locate elsewhere, which prevents a non-PAM equilibrium.

The second property of Proposition 1 shows that there is a positive relation between the productivity of workers and firms and the population density of locations. Workers benefit from better job opportunities in higher-ℓ locations chosen by more productive firms. Thus, a larger number of workers are attracted to those locations. Note that the vacancy destruction rate δ_v needs to be sufficiently large for this result. If it is too low, locations with more productive firms may be less attractive to workers despite better job opportunities due to an excessively high threat point of firms, B_w. In this scenario, these locations could be less densely populated.

Random matching in local labor markets is crucial for the second property of Proposition 1, which ensures that workers' value of choosing locations increases in local firm productivity, all other things being equal. Since search is random, all workers have the opportunity to be matched with more productive firms that pay higher wages, as long as they search for a job within these locations. In contrast, if firms could select the workers they hire based on workers' productivity, as in competitive 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.

Finally, the last part of the proposition claims that any pure-assignment equilibrium features an urban wage premium. Having characterized the sorting decisions, $x(\ell)$ and $y(\ell)$, I can examine the spatial disparity in local wages using (2):

$$w(x(\ell), y(\ell), \ell) = \left(b + \beta \frac{\tilde{\rho} + \lambda(\ell)}{\tilde{\rho} + \beta\lambda(\ell) + (1-\tilde{\beta})q(\ell)}(y(\ell) - b)\right)x(\ell). \tag{7}$$

Average wages differ across locations due to worker and firm productivity, $x(\ell)$ and $y(\ell)$, and market tightness $\theta(\ell)$. Workers and firms of higher productivity produce more output, which results in higher average wages. Also, higher market tightness leads to higher average wages, since it increases workers' threat points in bargaining. I show that the potential decrease in wages due to lower market tightness is smaller than the increase in wages due to higher productivity when firms' threat point is sufficiently low, which occurs when vacancy destruction rate δ_v is sufficiently large. It is important to note that the urban wage premium arises across homogeneous locations, and purely stems from the sorting mechanism.

Combining the results in Proposition 1, I conclude that an economy with search frictions in which heterogeneous workers and firms sort across local labor markets can explain several crucial dimensions of the spatial disparities we observe in the data.

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. I consider the problem of a social planner who can implement transfers across workers and can collect all profits from firms, landowners, and intermediaries, but is subject to search frictions within each labor market. Since the planner is allowed to use transfers, the problem essentially boils down to maximizing the flow of consumption goods that can be distributed across workers, which is total output minus the costs associated with housing and business services. In this section, I focus on the limit case where a discount rate ρ goes to zero to simplify exposition.[12] I present the proofs of all results in Appendix A.4.

I first characterize the optimal assignment and compare this with the spatial allocation obtained in the decentralized equilibrium. The following lemma demonstrates that the optimal spatial distribution features PAM between workers and firms, as in the decentralized equilibrium.

Lemma 1. *The optimal spatial allocation exhibits PAM between workers and firms and can be represented by two strictly increasing location-matching functions, $x^*(\ell)$ and $y^*(\ell)$.*

PAM between workers and firms across local labor markets results in higher output in the presence of complementarity. A potential complication arises from differences in market tightness across regions. For example, a planner would assign more productive firms to locations with less productive workers, when the probability of these firm' being matched is sufficiently higher. However, I show that this scenario does not

[12] 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) in location ℓ. I assume that workers and firms are infinitely patient and ensure that the social planner's objective function is comparable to the value function of workers and firms.

occur due to congestion forces.[13] These forces prevent the planner from assigning workers and firms such that regional differences in market tightness dominate those in productivity.

Using Lemma 1, the planning problem becomes choosing increasing worker and firm assignments $(x^*(\ell), y^*(\ell))$. Even among increasing functions, the planner still needs to decide worker and firm density across locations. Local output is determined by the number of employed workers $(1-u(\ell))L(\ell)$ and match output $x(\ell)y(\ell)$, and two congestion costs depend on worker and firm density. The planner solves the following problem,

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

in addition to the boundary conditions on $x(\ell)$ and $y(\ell)$. Constraints determine how assignments determine population density, firm density, and unemployment rates as in decentralized equilibrium. See (5) and (6) for more details.

The optimal assignment is characterized by

$$\bar{h}^2 C_r'(\bar{h}L(\ell))L'(\ell)L(\ell) + C_v'(V(\ell))V'(\ell)V(\ell) = 0, \tag{8}$$

in addition to (A.11) in Appendix A.4. First, the above condition (8) shows that the optimal assignment of workers and firms does not exhibit spatial concentration. Specifically, changes in population density $L'(\ell)$ and changes in the measure of vacancies $V'(\ell)$ at each location ℓ have the opposite sign, i.e., $L'(\ell)V'(\ell) \leq 0$ for all ℓ. When the planner makes marginal adjustments to the assignment of workers and firms between two locations, she does not make one location strictly denser than the other. 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 less concentrated ones. The second condition (A.11) illustrates how the planner decides whether to increase $L(\ell)$ or $V(\ell)$. For example, the planner chooses to allocate relatively more workers in the higher ℓ region—i.e., $L'(\ell) > 0$ and $V'(\ell) < 0$—when heterogeneity in firm productivity is more pronounced. By concentrating highly productive firms and a large measure of relatively homogeneous workers into a single location, the planner can increase output.

[13] Congestion forces in housing and business services are essential to show this result. Otherwise, for PAM to be optimal, one needs a stronger assumption on the matching function or stronger form of complementarity.

Next, I evaluate the efficiency of the decentralized equilibrium. The following proposition shows that the decentralized allocation is inefficient due to externalities arising from spatial sorting of heterogeneous workers and firms.

Proposition 2. *The decentralized equilibrium is inefficient. Each worker and each firm chooses a higher ℓ location than the social planner would designate, given the location decisions of all others. The social planner can implement the optimal assignment by using the spatial transfers to workers and firms, that are given by* (A.16) *and* (A.17).

Although workers and firms value denser cities, the planner finds that congestion costs are unnecessary because productivity is embodied in workers and firms. In this economy, output depends on the matching between workers and firms rather than their locations, and the planner does not prefer clustering workers and firms in cities. In contrast, workers and firms are drawn to cities by opportunities to search for more productive counterparts. 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. In particular, in contrast to the optimal allocation, the marginal change in congestion in decentralized equilibrium is positive, as given by

$$\bar{h}r'(\ell)L(\ell) + c'(\ell)V(\ell) = \bar{h}^2 C_r'(\bar{h}L(\ell))L'(\ell)L(\ell) + C_v'(V(\ell))V'(\ell)V(\ell) > 0 \quad \text{for all } \ell.$$

To formalize this idea, I assume frictional local labor markets with random matching, together with heterogeneous agents sorting acorss 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 that externalities cancel out (e.g., Shimer and Smith, 2001; Albrecht, Navarro and Vroman, 2010).[14]

Impoartantly, Proposition 2 holds independent of the elastcitiy of matching function and the workers' bargaining power.[15] In line with the literature, when I assume the Hosios condition, there is no need to correct

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

[15] Because *both* workers and firms inefficiently sort across space, a generalized condition (e.g., Mangin and Julien, 2021) cannot address the inefficiency. For example, higher worker's bargaining power in cities may lower firms' inflows but increase workers' inflow even more.

for externalities arising from changes in market tightness. In particular, let the elasticity of job finding rates with respect to market tightness be denoted by $\varepsilon_\lambda(\ell)$. Under the Hosios condition $\varepsilon_\lambda = 1 - \beta$ and zero unemployment benefit $b = 0$, an equilibrium is efficient if and only if workers and firms are homogeneous.

The social planner can implement the optimal assignment by using the spatial transfers to workers and firms, $t_w(\ell)$ and $t_f(\ell)$. To focus on externalities that arise from the sorting, assume the Hosios condition and zero unemployment benefit. Then, spatial transfers in (A.16) and (A.17) simlify as below.

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

where the constants t_w^0 and t_f^0 balance the government budget; $x^*(\ell)$ and $y^*(\ell)$ denote the optimal assignments; and $L^*(\ell)$, $V^*(\ell)$, and $u^*(\ell)$ are determined under these allocations. The above transfers decrease in ℓ, which emphasizes the fact that the decentralized equilibrium is overly crowded. By adjusting the spatial transfers, the planner can ensure that workers (firms) internalize the impact of their choices on firms' (workers') values. 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.

3.3 Discussion of Alternative Modeling Choices

To understand the role of random matching, I consider two alternative labor market structures: competitive markets and labor markets with directed search. I show that these two models fall short of explaining the spatial disparities observed in the data, and particularly differential population densities across regions. Moreover, the decentralized equilibrium turns out to be efficient in both cases—as opposed to the baseline as shown in Proposition 2.

Competitive local labor markets. In a local labor market in ℓ, each firm hires a worker of productivity x to maximize its profit, $xy - w(x,\ell)$, for a given wage schedule $w(x,\ell)$ without any frictions. 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 Appendix A.6 for the proof.

Proposition 3. *A 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) 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.

When a labor market is competitive without any frictions, firms can selectively hire workers and pay wages that are specific to each worker type, which are equal to the marginal contribution of workers on output. In equilibrium, wages of workers only depend on their type, i.e., $w(x, \ell) = w(x)$. Since wages are the same across locations, workers have no incentives to choose dense and expensive locations. In turn, population density is uniform across regions, which is not consistent with the data.

More importantly, the third property in Proposition 3 indicates that locations lack economic significance 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.

Directed search. In Appendix A.7, I also discuss an economy in which each local labor market has 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 a firm. 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. I incorporate the pre-stage in which workers and firms first make location choices. Next, in each location, workers and firms competitively search among those who chose the same location.

Focusing on the pure-assignment equilibrium, I show that when complementarity is strong enough, PAM between workers and firms arises.[16] However, the equilibrium does *not* exhibit spatial concentration: Population density and firm density move in the opposite direction. Thus, when population density increases in ℓ, the local unemployment rate increases in ℓ, which is inconsistent with the data, as illustrated in Figure A.2a. Therefore, this market structure also falls short in explaining empirical observations.

[16] I consider a more general production function in this case because PAM arises if complementarity in the output function is larger than the elasticity of the matching function, as in Eeckhout and Kircher (2010).

I further show that the decentralized equilibrium is efficient. With directed search, firms would not choose less productive workers unless their value is matched to that obtained when choosing more productive workers in local markets. Thus, less productive workers internalize their own negative impact on local firms. This differs from the baseline economy with random matching, in which workers gain a share of the increased surplus when matched with more productive firms. This finding aligns with the findings that directed search usually leads to the efficient allocation. See Proposition A.3 for the formal result.

3.4 Discussion of Alternative Mechanisms Accounting for Spatial Disparity

In this paper, I focus on the two-sided sorting mechanism to show that spatial disparity can endogenously arise. However, other mechanisms explored in previous studies can also produce similar patterns across space. What is the characteristic of two-sided sorting that distinguishes it from others? To answer this question, I compare the two-sided sorting mechanism to others in both positive and normative aspects.

I introduce the two additional mechanisms—heterogeneous location productivity and agglomeration forces—into the baseline model. To minimize changes, I modify the output function in a way that it depends directly on the location of production, in addition to the productivity of workers and firms. 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 $A^x(x(\ell))$ captures 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 $x(\ell)$.[17] I assume that output of a worker of productivity x and a firm of productivity y in location ℓ is given by $A(\ell)xy$.

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

$$\log w(x(\ell), y(\ell), \ell) = \log\left(b + \beta\frac{\tilde{\rho} + \lambda(\ell)}{\tilde{\rho} + \beta\lambda(\ell) + (1-\tilde{\beta})q(\ell)}(\bar{A}(\ell)A^x(x(\ell))y(\ell) - b)\right)x(\ell). \tag{9}$$

All derivations and proofs are in Appendix A.8.

This extended model relates closely to important models in the literature, depending on the importance of each source of heterogeneity: worker productivity, firm productivity, exogenous location productivity,

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

and agglomeration forces. For example, it returns to the *two-sided sorting* model when focusing on worker and firm heterogeneity without considering the heterogeneity in local TFP, i.e., $\bar{A}(\ell) = A^x(x(\ell)) = 1$. If the spatial disparity arises soley from exogenous location productivity $\bar{A}(\ell)$, the model corresponds to the *no sorting* model. When spatial disparity arises from worker productivity $x(\ell)$ and exogenous location productivity $\bar{A}(\ell)$, then it aligns with the *one-sided sorting* model of workers. Lastly, it corresponds to the *spillovers* model when incorporating the heterogeneity of workers $x(\ell)$ and agglomeration forces $A^x(x(\ell))$.[18]

The following proposition compares these different models, highlighting their similarities and differences.

Proposition 4. *Suppose that wages $w(\ell)$, population density $L(\ell)$, and unemployment rates $u(\ell)$ across locations are observable. Then, any model that assigns different contributions to each source of heterogeneity $\{x(\ell), y(\ell), \bar{A}(\ell), A^x(x(\ell))\}$ will be equally successful in explaining these observed cross-sectional moments. However, each model implies different policy interventions.*

Each sources of spatial disparity contributes to raising output and wages and thus plays a similar role in matching cross-sectional data. For example, as shown in (9), an increase in wages in higher-ℓ locations can result from either higher firm productivity $y(\ell)$ or higher exogenous location productivity $\bar{A}(\ell)$. Thus, distinguishing between models based only on cross-sectional data is difficult. This explains why previous studies have been able to 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 has different policy implications. I first compare the models with exogenous location productivity—*no sorting* and *one-sided sorting*—with 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 (A.10). In contrast, when locations are ex ante heterogeneous, the location of production directly affects aggregate output. In particular, I show that the planner allocates both more workers and firms to more productive locations in (A.22). In turn, congestion in cities—i.e., $L'(\ell) > 0$ and $V'(\ell) > 0$—may be desirable properties.[19]

While the *spillovers* model shares key similarities with the *two-sided sorting* model, the contribution of this paper lies in its focus on inefficiencies regarding spatial concentration. Both models assume ex

[18] Similarly, I can consdier a combination of firm sorting and another souce of heterogeneity.

[19] In the one-sided sorting model obtained from the model in this seciton, the decentralized equilibrium is inefficient due to exteranlities on the worker side. However, these extarnlities have differnet implicaitons. The key distinction lies in whether the location of production affects overall output or not. Additionally, if the objective is to study the one-sided sorting model, spatial disparity can be generated without introducing search frictions, in which case the decentralized economy can be efficient.

ante homogeneous locations, which limits the role of location in determining output. Specifically, these two models, as described above, conclude that congestion in dense cities is unnecessary. However, the literature has primarily emphasized the possibility of generating spatial disparity (e.g., Davis and Dingel, 2020; Behrens, Duranton and Robert-Nicoud, 2014)[20].

Furthermore, the spillovers model differs from the two-sided sorting model in two aspects related to policy implications. First, homogeneous firms no longer create negative externalities and are not subject to spatial transfers. Instead, workers generate additional externalities through local TFP due to agglomeration forces $A^x(x(\ell))$, which causes spatial transfers of workers to decrease more rapidly in ℓ. Second, the spillovers model does not consider the matching between workers and firms because firms are assumed to be homogeneous. For example, place-based policies that incentivize high-performing firms to locate in low-income locations are beneficial in the spillovers model. In contrast, in the two-sided sorting model, such policies disrupt PAM between workers and firms, and are thus less desirable.

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. Empirical Evidence of Two-Sided Sorting

Before turning to the structural estimation, I provide emprical evidence of two-sided sorting that suggests the importance of this mechanism. In Section 5 and Section 6, I return to the structural model for estimation and countefactual exercises.

I test the key prediction of the theory: Worker sorting leads to firm sorting. From the value of firms in (4), when a location ℓ' has higher worker productivity than a location ℓ, i.e., $x(\ell) < x(\ell')$, more productive firms post vacancies, i.e., $y(\ell) < y(\ell')$, provided that vacancy contact rates are the same. The responsiveness of firm sorting, particularly with respect to worker sorting, is a unique feature of two-sided sorting, which distinguishes it from others that produce similar cross-sectional patterns. Therefore, if the response of firm sorting is substantial in the data, it suggests that two-sided sorting significantly contributes to spatial disparity.

[20] Although Behrens, Duranton and Robert-Nicoud (2014) suggest that cities may be overly crowded under plausible parameter values, their conclusion depends on the values of parameters in contrast to the externalities arising from two-sided sorting. Another strand of literature studies efficiency but focuses on externalities arising from heterogeneous spillovers between different types of workers rather than between workers and firms (e.g., Fajgelbaum and Gaubert, 2020; Rossi-Hansberg, Sarte and Schwartzman, 2019).

Data. In this section, I use German administrative microdata. An empirical analysis in this section requires a matched employer-employee dataset, although it is not necessary for estimation in the next section. I use worker-level panel data from linked employer-employee data in Germany (LIAB) from the Institute for Employment Research (IAB) in Germany, which is generated by linking an annual establishment survey from IAB Establishment Panel and individual employment information from the Integrated Employment Biographies (IEB). The data provide identification numbers of establishments. Since the model does not distinguish jobs, establishments, or firms, I use these terms interchangeably in this section. I use 257 commuting zones (CZs) to capture Germany's local labor markets.

The variables I use most extensively are two-way fixed effects of workers and firms from AKM wage regression; two-way fixed effects are widely used in the literature since Abowd, Kramarz and Margolis (1999) to understand the dispersion of wages. I rely on estimates provided by the Research Data Centre (FDZ) of the German Federal Employment Agency.[21] These estimates are based on the IAB Employment History File (BEH), which represents the universe of workers subject to social security contributions. For example, the numbers of estimated worker and firm fixed effects over the years 2010–2017 are more than 30 million and 2 million, respectively. Fixed effects are provided for different periods and each period is about 7 years. I focus on estimates from 2003-2010 and 2010-2017 for the main analysis.

Measuring productivity. To measure worker productivity $x(\ell)$ that determines the values of posting vacancies, I compute the local average worker productivity in the search pool. To assess the response of firm sorting, I measure the average productivity of newly created jobs $y(\ell)$. Because firms are randomly matched with workers in the local search pool, I use the average of worker productivity. Moreover, I include all workers eligible for new hires, whether they were previously unemployed or have recently switched jobs. Similarly, I consider all new jobs, which represent the unit of my production, whether created by incumbents or entrants.

I estimate the local productivity of workers and firms by combining the model and a standard two-way fixed-effects wage regression. Measuring productivity is empirically challenging, because observable statistics such as wages depend on a variety of factors. My contribution is to map standard two-way fixed effects onto the wage components of my two-sided sorting model and show how worker and firm productivity can be recovered. In Appendix C, to justify using two-way fixed effects in my context, I present an extended model that features non-degenerated worker productivity within location ℓ and incorporates

[21] They follow the estimation strategy of Card, Heining and Kline (2013). See the paper for extensive discussion of the performance of fixed effects in approximating the wage in Germany.

migration frictions. Importantly, the value of firms retains the key property that I test in this section, i.e., higher local worker productivity attracts more productive firms. In addition, the wage equation remains unchanged from the baseline model, which supports my measure of productivity.[22]

If the data are generated by my model and local wages follow (7), then estimated worker fixed effects correctly represent their log productivity.[23] In contrast, firm fixed effects depend not only on firm productivity but also on the continuation value of local job search or the threat point of workers in bargaining. The average estimated fixed effect of newly matched jobs in location ℓ, which corresponds to the term in parenthesis in (7), depends not only on firm productivity $y(\ell)$ but also on local job finding rates $\lambda(\ell)$ and other parameters.[24] I can recover the productivity of firms $y(\ell)$ from estimated firm fixed effects and the local job finding rate $\lambda(\ell)$. In Appendix C.2, I explain how this can be achieved without estimating the full model.

Causal evidence of the two-sided sorting mechanism. To provide evidence of two-sided sorting, I examine whether *changes* in worker sorting lead to changes in firm sorting. I use the time-differenced regression due to the difficulty of identifying two-sided sorting from cross-sectional data, as discussed in Proposition 4. Guided by (4), I use the following regression:

$$\Delta \log y(\ell) = \gamma_0 + \gamma_1 \Delta \log x(\ell) + \gamma_2 \Delta \log \lambda(\ell) + u(\ell), \tag{10}$$

where Δ denotes the change in the value of a variable between the first period $t = 1$ (2003-2009) and the second period $t = 2$ (2010-2016). This choice is motivated by the fact that the FDZ estimated fixed effects separately for 2003-2010 and 2010-2017. Instead of changes in $q(\ell)$, I control for changes in the job finding rate, which are inversely related to changes in $q(\ell)$ and observable in the data.

One concern is that *changes* in location-specific factors—such as TFP, amenities, or infrastructure—may causally affect the location decisions of both workers and firms. For example, an increase in local TFP may attract both more productive workers and firms, if both worker and firm productivity are complementary to local TFP. Moreover, if the location directly affects production, estimated firm fixed effects contain differences

[22] This model features exhibits positive cross-regional worker flows, which are necessary to identify the differences in worker and firm productivity across locations. Also, the exogenous mobility assumption for two-way fixed-effects wage regression is guaranteed by a set of assumption that include random matching, exogenous separation, and the employment status of migrants.

[23] For the remainder of the analysis, I assume δ_v approaches to infinity. See Footnote 34 for further discussion of this assumption.

[24] Firms rarely move, and it is well acknowledged in the literature that identifying firm productivity separate from other factors is extremely difficult (e.g., Combes, Duranton and Gobillon, 2008). A few studies discuss strategies to identify firm productivity (e.g., Gaubert, 2018; Bilal, 2023; Lindenlaub, Oh and Peters, 2023). However, each strategy depends on a set of assumptions that is specific to its own context.

in local TFP because establishments do not change location in the data. From the wage equation of the extended model (9), estimated firm fixed effects correspond to the log of the term in the parentheses, which increases in both firm productivity and local TFP. Thus, if an increase in local TFP attracts more productive workers, it may bias the coefficient upward.

To address these concerns, I attempt to identify exogenous changes in worker productivity caused by shocks to other locations. Specifically, I instrument changes in the productivity of *all* workers in each location caused by changes in internal *migrants'* productivity that are induced by shocks to the origin locations of those migrants. Migrants account for about 30% of the local search pool in my sample period—a sizable share—and thus their productivity directly affects the local average of all workers.[25] Moreover, changes in the productivity of migrants indirectly also impact that of non-migrants. For example, an increase in migrant productivity may raise housing rents, which will push out less productive non-migrants who cannot afford those higher costs.

To identify plausibly exogenous changes in migrant productivity, I use predicted changes that arise from shocks to their origin locations instead of those to location ℓ. First, I compute predicted inflows from each origin ℓ' to ℓ. Specifically, I use migration flows in $t = 0$ (1991−2002) to estimate the probability that migrants from origin location ℓ' move to location ℓ, $m_0(\ell' \to \ell|\ell')$. I then multiply this probability by the total number of out-migrants from each origin location ℓ' to any locations except ℓ in period $t = 1, 2$, $O_{\ell',-\ell,t}$.[26] Dividing this predicted flow $m_0(\ell' \to \ell|\ell)O_{\ell',-\ell,t}$ by the total predicted flows into ℓ—i.e., the sum of these flows from all origin locations—I obtain the predicted share of migrants entering ℓ from each ℓ', $\hat{s}_t(\ell'|\ell)$. Second, I use the productivity of all migrants leaving ℓ' but choosing elsewhere, $\hat{x}_t(\ell', -\ell)$, instead of the productivity of migrants arriving at ℓ. Combining shares $\hat{s}_t(\ell'|\ell)$ and productivity $\hat{x}_t(\ell', -\ell)$, I predict the productivity of migrants arriving in ℓ by

$$\log x^{\mathrm{m,\,IV}}(\ell) = \sum_{\ell' \neq \ell} \hat{s}_t(\ell'|\ell) \log \hat{x}_t(\ell', -\ell) \quad \text{where} \quad \hat{s}_t(\ell'|\ell) = \frac{m_0(\ell' \to \ell|\ell')O_{\ell',-\ell,t}}{\sum_{k \neq \ell} m_0(k \to \ell|k)O_{k,-\ell,t}}. \tag{11}$$

[25] Although the overall migration rate is limited, I find that the importance of migrants is larger among unemployed workers. In my data, I only observe the CZ of establishments (i.e., firms), 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. This is consistent with the timing of the model, which assumes that workers first choose a location and then search for a job.

[26] This approach is similar to the strategy used to study the impact of international immigrants (e.g., Burchardi, Chaney and Hassan, 2019; Altonji and Card, 1991) or internal migrants (e.g., Howard, 2020; Boustan, Fishback and Kantor, 2010). However, compared with the literature that mostly focuses on changes in the *number* of migrants, I focus on changes in their productivity.

Finally, I obtain the instrument for $\Delta \log x(\ell)$ by taking the time differences of these values. When a significant fraction of the variation in migrant productivity arises from shocks to origin locations, the instrument can predict changes in the productivity of all workers.

The identification assumption is that shocks to origin locations are uncorrelated with shocks to a specific destination. My instrument exploits two types of variations. First, variation in the magnitude of push shocks among origin locations leads to changes in $\hat{s}(\ell'|\ell)$. For example, a positive shock to origin location ℓ' decreases its out-migration. If origin ℓ' has lower average productivity compared with other origin locations of ℓ, the composition of migrants arriving at ℓ improves, which increases worker productivity in ℓ. Next, changes in $\hat{x}(\ell', -\ell)$ arise from changes in the productivity of workers in origin location ℓ', and the impact is larger when the origin ℓ' historically sent a large number of migrants to ℓ. By focusing on migrants moving to other locations $-\ell$, I exclude influences attributable to pull factors of ℓ. This approach could be problematic if the historical network is influenced by factors that creates correlated regional shocks, such as similar industry composition or geographic proximity. To address this concern, I further control for industry or exclude origin location ℓ' within a 100 km radius in my robustness checks.

Table 1 shows the estimation results. In all specifications, I include the firm productivity and unemployment rates of location ℓ in $t = 1$ as controls. The former controls for a potential time trend in productivity due to regional convergence. The latter controls for 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, especially in regions with initially high unemployment rates.

Column (1) documents the result of 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, which helps fill this gap. In the bottom panel, I report first-stage regression results, in which 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. I find a statistically significant positive coefficient. The IV regression coefficient is statistically significant and economically meaningful: An exogenous 10% increase in the productivity of local workers in the search pool leads to an 11.3% increase in firm productivity of new jobs in the same location.[27] To test robustness, I construct a

[27] This coefficient is larger than that of OLS, which indicates that the attenuation bias due to measurement errors appears to be larger than the upward bias, for example, due to changes in local TFP.

Table 1: 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 employees. The dependent variable is the change in firm productivity of new jobs. In the first-stage regression, I regress the change in worker productivity on its instrument and the controls I include in the second stage.

dependent variable by first residualizing firm productivity against 1-digit industry fixed effects in Column (3) and drop all flows that are to or from locations within 100 km in Column (4). The coefficients of interest do not change much.

Even if instrumented changes in worker productivity are exogenous, it may be the case that agglomeration forces, in addition to firm sorting, are behind the positive coefficients. An increase in worker productivity may endogenously increase the productivity of all matched jobs in a location (e.g., Diamond, 2016; Rossi-Hansberg, Sarte and Schwartzman, 2019). In particular, going back to (9), an increase in worker productivity $x(\ell)$ leads to higher agglomeration forces $A^x(x(\ell))$, which would contaminate estimates of firm fixed effects and create positive measurement errors in firm productivity. To focus on the effects of firm sorting, I also control for changes in the firm fixed effects of *existing* jobs that are created in the first period, $\Delta \log \hat{y}^{\text{old}}$, which also benefit from agglomeration forces or any other form of region-level spillovers.[28] 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 productivity in a location induces a 0.51 standard deviation increase in firm productivity in the same location.

I further support these results with two additional results. First, I use a reduced-form approach and use changes in estimated firm fixed effects $\Delta \log \hat{y}(\ell)$ as a dependent variable. Overall results remain almost

[28] Specifically, I compute changes in the firm fixed effects of each job using fixed-effects estimates that are separately provided for two periods. I only include jobs in firms that survived in the second period and have firm fixed effects for both periods. I use the firm fixed effects of existing jobs due to the potential for a spurious negative correlation between their recovered productivity and that of new jobs. See Appendix C.3 for more discussion and the results.

the same: A 1 standard deviation increase in the worker productivity of a local labor market induces a 0.6 standard deviation increase in the productivity of firms choosing that location. Second, I also regress changes in the firm fixed effects of existing jobs, $\Delta \log \hat{y}^{\text{old}}(\ell)$, as a placebo test. In contrast to Table 1, the coefficients are insignificant and nearly zero, which implies that most changes in the productivity of new jobs arise from changes in firm sorting rather than changes in regional factors. All results are summarized in Appendix C.3.

In summary, this analysis shows that worker sorting has a significant impact on firm sorting. This finding supports the two-sided sorting mechanism as an important driver of spatial disparities. This evidence supports the use of my model for policy evaluation in the next section.

5. Quantitative Analysis

I now examine the quantitative implications of the model. In this section, I calibrate the model using cross-sectional data on U.S. MSAs for three reasons. First, the U.S. features more pronounced spatial disparities than any other country; extensive studies document spatial inequality in the U.S. (e.g., Diamond and Moretti, 2021; Card, Rothstein and Yi, 2023). Second, the U.S. has adopted various forms of spatial policies,[29] yet their welfare implications are unclear. Third, my estimation strategy requires only regional data, which are readily available in many countries.

5.1 Quantitative Model

Setting. To bring the model to the data, I generalize preferences to a Stone-Geary utility and assume that workers' flow utility is given by $g^{1-\omega}(h-\bar{h})^{\omega}$, where g denotes tradable goods and h denotes housing consumption. This change helps to accurately capture the quantitative impact of congestion costs from higher housing rents, which is a key component in workers' location decisions. Under this generalized utility, the tractability of the model is preserved. Furthermore, I normalize the average expected profit to zero by introducing firms' entry cost c_e that firms pay before drawing their productivity $y \sim Q_f$.[30] See Appendix B.2 for details.

[29] The U.S. not only adopts policies that are designed to address spatial concerns such as place-based policies (e.g., Busso, Gregory and Kline, 2013; Kline and Moretti, 2014), but also policies that indirectly affect the location choices of workers and firms such as local regulations (e.g., Hsieh and Moretti, 2019) and local taxes (e.g., Fajgelbaum et al., 2019).

[30] This normalization prevents income from firm portfolio from interacting with worker sorting under the Ston-Geary utility, maintaining consistency with the theoretical framework. In the counterfactual, lump-sum redistribution responds endogenously. I can additionally introduce free entry of firms before drawing y into the model, which ensures the average firms' profit remains zero

I introduce two policies that I evaluate in Section 6. First, I introduce local housing regulations. In particular, I assume that the local government taxes housing production at rate $\tau_h(H, \ell)$, where H is the total housing supply in location ℓ. For example, if the local government imposes a housing tax that increases in supply H, the housing supply elasticity will decline. I assume that local governments redistribute tax revenues to local workers as a lump sum. The second 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—i.e., the tax rate increases in workers' income. To approximate this feature, I assume that the tax rate is specific to each region.[31] Tax revenues are redistributed to all workers as a lump sum.

With these changes, the value of unemployed workers becomes

$$\rho V^u(x,\ell) = r(\ell)^{-\omega}((1-\tau_w(\ell))(b + A_w(y(\ell), \lambda(\ell)))x - \bar{h}r(\ell) + \Pi + T_r(\ell)), \tag{12}$$

where Π is the sum of lump-sum redistributions of income taxes, landowners' profits Π_r, and intermediaries' profits Π_c, net of taxes for unemployment benefits. Workers also receive housing tax revenues $T_r(\ell)$ that are redistributed locally. In addition, the housing market clearing condition in (A.25) takes into account that housing demand increases in both population density and after-tax average income—and given the demand, housing rents are higher when housing regulations are stricter.

Functional forms. I assume that matches are produced by a Cobb-Douglas matching function with the matching efficiency $\mathcal{A}$ and a constant matching elasticity α. I use a housing production cost function with constant elasticity, which leads to the housing supply function $H(\ell) = H_w r(\ell)^{\eta_w}$. This functional form is commonly used in the literature, and can be micro-founded by a housing production function that combines land and capital. Similarly, I use the same functional form for the business services supply function with parameters (H_f, η_f). Note that all functions are common across regions, so I do not assume any regional ex ante heterogeneity except for two policies: income taxes and housing regulations.

I assume that the tax rate on housing production is $1 + \tau_h(H, \ell) = (H/T)^{t_h(\ell)}$, where $t_h(\ell)$ captures the stringency of local housing regulations. In Appendix B.2, I show that the inverse of the housing supply elasticity becomes $1/\eta_w + t_h(\ell)$, and thus more stringent regulations result in lower housing supply elasticity.

in the counterfacutals. In this case, the total measure of firms M_f responds endogenously, but I find that the results remain almost the same.

[31] A more direct way is to assume that the tax rate depends on workers' income. However, in this case, the wage bargaining problem cannot be solved by Nash bargaining. By assuming a location-specific tax rate, I can maintain tractability.

5.2 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 residualize wages with demographics and 1-digit industry, unemployment status with demographics, and housing rents with characteristics of buildings.

I complement this dataset with a several additional sources. I use the Current Population Survey (CPS) from IPUMS (Flood et al., 2022) to compute the job separation rate and federal income taxes. Next, 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, Saiz and Summers (2008). Finally, I obtain population density from the U.S. Census and local GDP per capita from the Bureau of Economic Analysis (BEA) and average spending shares on housing from the Consumer Expenditure Survey (2017). See Appendix B.1 for more details.

5.3 Estimation

First, I order MSAs by population density, which is observable and increasing in ℓ.[32] For computation purposes, I group MSAs into 20 bins $\in \{\ell_1, ..., \ell_{20}\}$, with each bin containing the same proportion of the total population. Each bin ℓ_i corresponds to an interval of locations in the model, with its measure proportional to the inverse of the population in that bin.[33] For each bin, I compute the averages of population density $L(\ell)$, wage $w(\ell)$, unemployment rate $u(\ell)$, housing rent $r(\ell)$, and housing spending share by weighting each MSA with its population. Finally, I compute the equilibrium at monthly frequency.

I will now discuss estimation of the model. Although formal identification is not feasible, I will explain how each variable can be pinned down by certain moments, using heuristic arguments. I first discuss parameters that are externally set or calibrated directly from the data, which are summarized in the top panel of Table 2. Next, I proceed with the parameters, listed in the bottom panel, that are internally calibrated.

[32] As locations are ex ante homogeneous, choosing endogenous outcomes to order regions is unavoidable. In principle, I can use any variable that has a monotonic relation with $x(\ell)$—my criterion for ordering ℓ—such as the average wage or GDP per capita.

[33] Alternatively, I could group MSAs in such a way that each bin contains the same number of MSAs. I prefer this approach because it avoids over-assignment of bins to less densely populated MSAs, so that the number of workers assigned to each bin varies significantly.

Table 2: Parameter Values

	Parameter	Target	Value
δ	separation rate	EU transition rate	0.028
ρ	discount rate	interest rate	0.003
α	matching elasticity	literature	0.5
$\mathcal{A}$	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

Notes: The top panel shows parameters that are externally calibrated and the bottom panel shows parameters that are internally calibrated.

Externally set or calibrated parameters. I set the discount rate $\rho = 0.004$, which implies an annual real interest rate of 5%. I set the matching elasticity to $\alpha = 0.5$ based on standard values in the literature. The separation rate $\delta = 0.028$ is set to the average monthly transition probability from employment into unemployment. I compute the number of workers who report being unemployed for 5 weeks or less divided by employment 1 month before to avoid the time aggregation issue (Shimer, 2012). Given α and δ, using the flow-balance condition in (6), I can compute the measure of vacancies $V(\ell)$. As the normalization, I set $\mathcal{A}$ to match the ratio of the total number of vacancies over the total number of unemployed workers in the U.S. I assume the destruction rate of a vacancy, δ_v, is sufficiently large. In practice, I set $\frac{\rho}{\rho+\delta_v} = 0$, effectively assuming that firm's threat point is zero.[34] Finally, I find the bargaining power of workers β that matches the labor share.

Estimation strategy. I am now left with a set of parameters $\{\bar{h}, \omega, \eta_w, b, \beta, H_f, \eta_f, H_w, c_e\}$ and the productivity of workers and firms across regions $\{x(\ell), y(\ell)\}_\ell$ to be estimated internally. I normalize the minimum values of worker productivity $x(\underline{\ell})$, firm productivity $y(\underline{\ell})$, and housing rents $r(\underline{\ell})$, which together determine the scale of an economy. Specifically, I set $y(\underline{\ell})$ and $r(\underline{\ell})$ to 1, and choose $x(\underline{\ell})$ to match the level of wages

[34] This choice has three advantages. First, as stated in Proposition 1, δ_v needs to be sufficiently large to explain the urban wage premium. Second, it decreases the continuation value of vacancies to zero, comparable to the free entry assumption in Pissarides (2000) or no capacity constraint in Postel-Vinay and Lindenlaub (2023), which is widely used in the search literature mainly for tractability. This tractability is useful in Section 3.2 and Section 4. Lastly, separately identifying δ_v and β is challenging.

of the lowest bin, $w(\underline{\ell})$. The housing market clearing condition of region $\underline{\ell}$ pins down H_w. Entry cost c_e will be set to an average expected profit after estimation. This leaves me with 7 structural parameters, $\{\bar{h}, \omega, \eta_w, b, \beta, H_f, \eta_f\}$, and productivities $\{x(\ell), y(\ell)\}$.

I approach these two groups of parameters using different strategies. Structural parameters that apply to the entire economy can be calibrated by targeting standard moments. The main challenge is separately estimating the productivity of workers and firms across regions. To tackle this, I rely on equilibrium conditions as well as observed regional variables, which allows me to recover productivities $\{x(\ell), y(\ell)\}$ during the estimation process.

Specifically, I compute productivity using equilibrium conditions by targeting population density and unemployment rates $\{L(\ell), u(\ell)\}$ across all bins for given structural parameters.[35] First, using (6), I compute the measure of vacancies $\{V(\ell)\}$, job finding rates $\{\lambda(\ell)\}$, and vacancy contact rates $\{q(\ell)\}$. Given these variables and other parameters, I find $\{x(\ell), y(\ell)\}$ that satisfy first-order conditions of the location decisions of workers and firms, $V^u_\ell(x(\ell), \ell) = 0$ and $\bar{V}^v_\ell(y(\ell), \ell) = 0$, and housing market clearing.

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

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

The matrix $\mathscr{W}$ is a diagonal matrix, containing the reciprocals of the squared data moments.[36] I now discuss the relationship between parameters and specific moments. Table 2 summarizes parameters along with the most relevant targeted moments.

Housing market. Preferences for housing, $\bar{h}$ and ω, can be obtained from the aggregate average housing spending share and its variation across regions. The housing spending of a worker 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

[35] An alternative, more conventional way is to parameterize the distribution of workers and firms, Q_w and Q_f, and then search over its parameters. My approach has three advantages. First, it solves potential problems of multiple equilibria because I start directly from empirical moments. Second, I do not need to make parametric assumptions on the productivity distribution. I can compute implied Q_w and Q_f ex post, combining $\{x(\ell), y(\ell)\}$ and $\{L(\ell), V(\ell)\}$. Lastly, it is computationally much faster.

[36] Because all bins of $\{L(\ell), u(\ell)\}$ are included in $\hat{m}$, I divide the corresponding diagonal elements of $\mathscr{W}$ by the total number of bins to avoid overemphasizing these moments.

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.24) reveals that a faster increase 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, Michaillat and Saez, 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.

Productivity. Differences in wages arise from dispersion in the productivity of both workers and firms, and separating the contribution of the two is the key step in the identification. I can identify the contribution of workers based on a revealed preference argument—i.e., the sorting condition of workers. Specifically, the value of workers in (3) describes the trade-off between higher income and increased housing rents.

For example, consider two locations, $\ell' < \ell''$, where $\ell'' - \ell'$ is sufficiently small. Workers in ℓ' are almost indifferent regarding moving to ℓ''. This implies that if these workers choose ℓ'', they will earn higher nominal wages to offset higher housing rents. The gap between these hypothetical wages, $w(x(\ell'), y(\ell''), \ell'')$, and the observed wages of workers in ℓ'', $w(x(\ell''), y(\ell''), \ell'')$, reflects the difference in worker productivity between the two locations. When the difference in housing rents between locations is smaller, $w(x(\ell'), y(\ell''), \ell'')$ will be lower, which implies that workers in ℓ'' are much more productive. Thus, given the observed wage differences between two locations, the contribution of workers will be larger, while that of firms will be relatively smaller. This identification strategy aligns with Glaeser and Maré's claim (2001) that higher wages in large cities reflect either higher ability or higher prices. When housing rents are comparable across regions, this suggests that the higher ability of workers primarily accounts for higher wages, with firms playing a relatively limited role.

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 (4) reveals how firms balance gains from the labor market, such as higher worker productivity, with increased overhead costs. Given worker and firm productivity, as discussed above, 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 computed from the model is significant, implied housing supply elasticity η_f will be smaller.

5.4 Estimation Results

The right column of Table 2 summarizes the estimation results. The estimated housing subsistence level $\bar{h}$ is large, which indicates that the non-homotheticity of housing preferences is substantial. The estimated housing elasticity η_w is 11.31, which is relatively larger than estimates in the literature (e.g., Saiz, 2010; Green, Malpezzi and Mayo, 2005). However, the average housing elasticity after accounting for regulation is 7.95, which is much smaller. The bargaining power of workers β is estimated to be 0.04. This estimate is comparable to that of Bilal (2023), who assumes a similar setup of wage determination.

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 more prominent productivity dispersion observed among workers. To illustrate, workers and firms in the top 10% of cities are 24.4% and 20.4% more productive than those in the bottom 10%, respectively.

How does this productivity heterogeneity translate into spatial wage inequality? In the right panel of Figure 3, I break down the urban wage premium—i.e., the wage increase associated with population density—into three components using (7). Each line 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. Workers and firms account for about 60.4% and 38.8% of the total urban wage premium, respectively. In contrast, the role of market tightness is limited.

In Table 3, I report the fit of 10 targeted moments, excluding $\{L(\ell), u(\ell)\}$, whose fit is documented in Figure A.2a. Overall, the model performs well in matching the targeted moments. Note that the model is stylized and well suited for matching systematic patterns. However, matching each specific moment across ℓ is difficult, especially when the moments are noisy. In the right panel of Figure A.2a, the model aligns with the observed pattern of flat unemployment rates across ℓ, but it does not accurately fit the unemployment rates for each specific ℓ. In Figure 4, I further compare moments from the data and model for log wages,

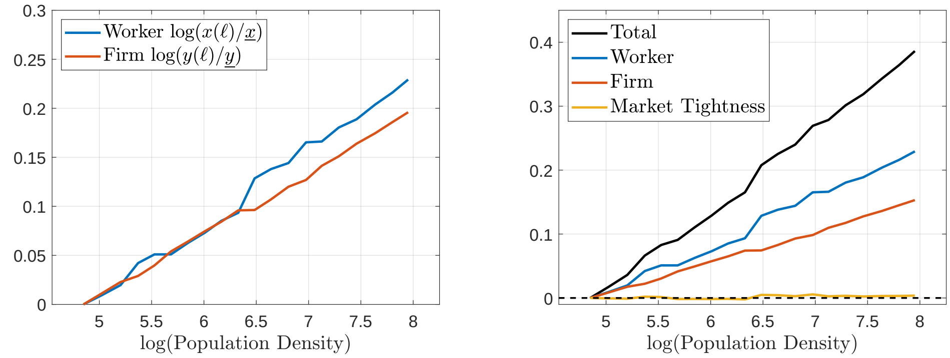


Figure 3. Worker and Firm Productivity (left) and Urban Wage Premium (right)

housing rents, and rent spending share across ℓ. As expected from Table 3, the model replicates the overall patterns of these moments well.[37]

Thus, I conclude that despite its simplicity, the model can successfully account for the spatial disparities observed in the data. Two-sided sorting alone suffices, not only theoretically but also quantitatively, without region-specific factors—such as local TFP and amenities—or agglomeration forces.

Overidentification. Using the estimated model, I regress log wages and its two components, log $x(\ell)$ and log $w(\ell)/x(\ell)$, on log population density. I estimate coefficients 0.13, 0.075, and 0.05, respectively. To validate the model, I compare these results to those of Card, Rothstein and Yi (2023), who run a two-way fixed-effects log wage regression of workers and CZ, using matched employer-employee data in U.S. As explained in Section 4, estimated worker fixed effects identify the productivity of workers. Thus, I can map their person effects and CZ place effects to $x(\ell)$ and $w(\ell)/x(\ell)$ in my model, respectively. Card, Rothstein and Yi run a similar regression and estimate coefficients of 0.075, 0.04, and 0.034, as reported in Table 3 in their paper. As in my results, the two components of wages are both quantitatively important and the coefficient of worker productivity is larger.[38]

In Figure A.2b, I also present a binned scatter of log GDP per capita across regions and the model counterpart. It is not surprising that the model replicates the variation in GDP per capita, since I target the cross-section of wages.

[37] The model is highly stylized, and thus is suitable for explaining the overall pattern. It is difficult, however, to replicate irregular variation if they are too volatile in ℓ. In addition, among targeted moments, the relative rent of the second quartile from the model of 0.147 exceeds the target of 0.079. In the data, both density and average wages increase substantially from the first to the second quartile, while rents remain relatively similar. Because the housing supply function is parameterized by a single elasticity parameter, matching all moments is challenging.

[38] The importance of worker heterogeneity is larger in my estimation, which may arise from differences in the geographical units, sample years, or controls. For example, they use Commuting Zones to define locations and population to measure the concentration of workers, while I use MSA and population density.

Table 3: Model Fit

Quartile	Wage 2	Wage 3	Wage 4	Housing Share mean	Housing Share $4^{\text{th}} - 1^{\text{th}}$	Replac. rate	Labor share	Rent 2	Rent 3	Rent 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 rents, I first compute the average of four quartile groups. Then, I target the average of the$i$-th quartile/the average of the first quartile-1, where $i = 2, 3, 4$. Housing share difference is the difference between the average housing spending share between the first and last quartile groups.

Threats to identification. My model is highly stylized and assumes that the spatial sorting of heterogeneous workers and firms alone accounts for spatial inequality. If other factors affect the value of workers or wages, they can pose a threat to my identification strategy. For example, if local TFP is higher in denser locations or agglomeration forces exist, they also contribute to spatial disparities but I will attribute them to firm heterogeneity. Similarly, if workers enjoy higher amenities in densely populated areas, all of these benefits will be attributed to firm heterogeneity.

Identifying alternative mechanisms from the two-sided sorting mechanism, however, is beyond the scope of this paper. Nonetheless, incorporating additional mechanisms in the model is straightforward as described in Section 3.4.

6. Policy Evaluation

To highlight the importance of the two-sided sorting mechanism, I now evaluate several real-world policies that affect the incentives of workers and firms to relocate across regions. The central idea behind the two-sided sorting mechanism is the fact that productivity is embodied in workers and firms. Therefore, it is crucial that we understand how heterogeneous workers and firms jointly respond to policies and, in turn, how particular workers are matched with particular firms in producing output in the new equilibrium. Many real-world institutions and policies influence the location decisions of workers and firms. Examples include place-based policies that provide subsidies to firms in economically disadvantaged areas, variation in local government tax systems, and discrepancies in the stringency of housing regulations. I focus on the two policies I incorporated in the model in Section 5.1. I first evaluate differential housing regulations across locations, then briefly discuss the impact of federal income tax of workers.

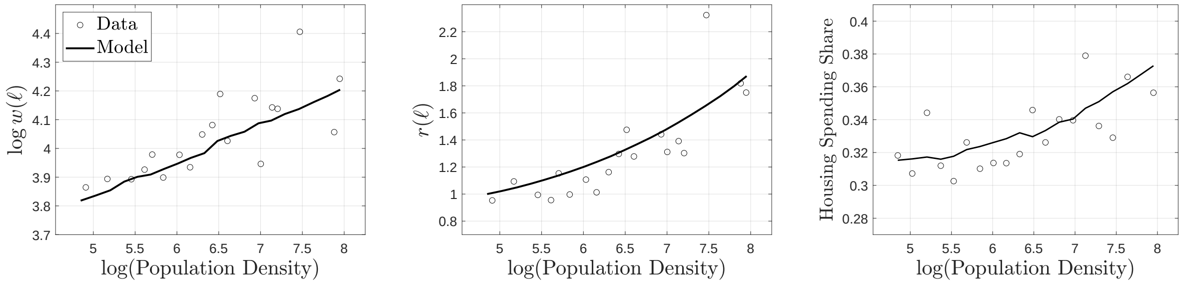


Figure 4. Model Fit: Wage, Housing Rent, and Housing Spending Share

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.

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). They assess the aggregate effect of land use restrictions by examining the counterfactual where they assume the level of regulations are set to the level found in the median city. They conclude that relaxing restrictions in large cities (New York, San Jose, San Fransisco) to the median US city would dramatically increase GDP (also increase welfare growth rates between 1964–2009).[39]

To implement the similar counterfactual exercise, I relax the housing regulations in dense areas $\ell \in [0.9, 1]$ where approximately 10% of workers reside. Specifically, I assume that these cities choose the same tax schedule $\tau(H; \ell)$ to that of the median location $\ell = 0.5$. The left panel of Figure 5 shows the baseline and counterfactual housing elasticities. The housing markets in high-ℓ locations become more elastic. I compare two steady-states before and after the policy change.[40]

Workers and firms move toward dense cities where the regulations are relaxed. The middle panel of Figure 5 illustrates percentage changes in population density and the measure of vacancies. Population density in the top 10% of regions increases by 55.3% as housing regulation is relaxed. The inflow of workers in higher-ℓ locations increases vacancy contact rates and thus attracts more firms, which leads to a 55.2% increase in the number of vacancies in the top 10% of regions.[41]

[39] Recent studies have examined how incorporating worker sorting and agglomeration forces can have additional impacts on similar policies. The conclusions, however, are not consistent. 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).

[40] Among the potential multiple equilibria, I choose the one that is closest to the calibrated equilibrium. Specifically, I choose the equilibrium with the population density in $\bar{\ell}$ that is closest to the population density before the policy change.

[41] Their spatial distribution changes susbtantially because I compute long-run changes and abstract away from migration frictions.

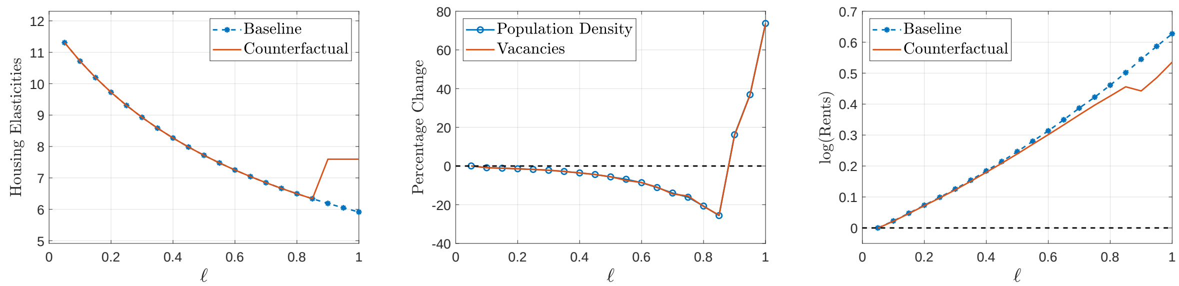


Figure 5. Impact of Relaxing Housing Regulations on Spatial Allocation

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, $\ell = 0.5$.

Yet, housing rents in dense regions are lower due to the relaxed regulations, as shown in the right panel of Figure 5. Moreover, housing rents in other regions decrease even without changes in regulations. As workers relocate toward the densest cities with relaxed regulations, other locations become less dense. In particular, the relocation of workers who are productive but not highly productive enough to make it to the densest regions with stringent housing regulation causes a sharp decline in population density in moderately crowded regions. This, in turn, causes a decline in housing rents across high-ℓ locations.

Despite the large-scale spatial reallocation, the change in aggregate output is marginal, at just -0.010%. Relocation patterns of workers and firms are highly similar, as shown in the middle panel of Figure 5. As workers are still matched with similar types of firms, in a two-sided sorting economy, the output of an individual worker remains almost the same, and so does an aggregate output. In contrast, as workers move toward large cities, congestion costs from housing and business services markets increase, leading to a decrease in GDP by 0.47%. This result is the opposite of the prediction of Hsieh and Moretti (2019). The main difference behind this discrepancy comes from the assumption of where the productivity is embodied.

While change in output is marginal, aggregate weflare, which is defined as an equally weighted average of the values of all workers,[42] *decreases* by 0.21%. This result is consistent with Proposition 2, which shows that reducing the density of dense areas enhances welfare. Stringent housing regulation in dense areas play a similar role to taxing workers and firms in those areas, and relaxing this regulation is opposite to what the optimal policy dictates.

[42] In other words, I assume the planner is utilitarian. In Section 3.2, it was unnecessary to choose specific weights of the planner. However, under Stone-Geary preferences, housing and final good consumption exhibit complementarity, and welfare depends on the distribution of tradable goods.

Figure 6. Welfare Impact of Relaxing Housing Regulation (% Change)

To investigate the sources of this welfare change, I decompose the total changes in welfare into three components: firm productivity; housing market; and labor market tightness. To quantify the contribution of each component, I first compute changes in the value of *each* worker due to changes in the relevant variable in the value of workers (12), holding all other variables constant. For example, in assessing the impact of firm productivity, I plug in firm productivity $y(\ell')$ in a new location chosen by a worker after the policy change, taking into account concurrent firm relocations. Then, I take the averages of these values to compute the aggregate welfare change attributable to each component. See Appendix D for details. Figure 6 summarizes the results.

A decline in welfare primarily comes from the housing market. As spatial concentration in the economy intensifies, aggregate housing costs increase, and this ultimately results in welfare loss. Specifically, the rise in aggregate housing costs leads to a reduction in the lump-sum income transferred from landowners' profits to workers, which more than offsets the benefits from lower housing rents. In contrast, the contributions of other mechanisms are limited. First, the aggregate welfare impact from changes in firm types is only -0.026%. Simply adjusting the matching schedule between workers and firms is insufficient to yield aggregate benefits because improving firm types for some workers inevitably results in worse firm types for others. Similarly, changes in market tightness cannot change aggregate welfare much.

In the left panel of Figure 7, I plot changes in the value of unemployed workers in each productivity percentile. The changes in the value of employed workers looks qualitatively the same. Since there are no migration frictions, although regulation is relaxed only in concentrated areas, the changes of welfare are uniform across different types of workers. However, the contribution of each mechanism differs across workers. The welfare changes from firm productivity and tightness do not have a systematic pattern across workers, and the corresponding lines move around zero. This is not surprising due to the smilarities in the relocation patterns of workers and firms. Even though workers move to different local labor markets, the relocation of firms in the new equilibrium leads to limited changes in firm productivity for each individual. There is no welfare gains for most of workers from housing market. Most of workers are worse off. Even workers of 90th percentile who experience upto 9% decrease in housing rents only marginally gain.

I now examine the importance of incorporating both worker and firm sorting in a unified framework. To illustrate this point, I consider a *one-sided* sorting model, in which heterogeneous workers sort across space while firms are assumed to be homogeneous. Instead, I introduce heterogeneity in exogenous location productivity $\bar{A}(\ell)$ as described in Section 3.4. When local TFP is estimated in the same way as firm productivity in my two-sided sorting model, estimated local TFP is identical to the estimated firm productivity $y(\ell)$ across locations. This confirms Proposition 4 which shows thath this one-sided sorting model is equally successful in explaining the documented spatial disparities across regions.[43]

Upon implementing the same housing deregulation within this one-sided sorting model, workers and firms relocate toward high-ℓ locations as before. Although the magnitude of changes is smaller because the calibrated overhead costs increase in the measure of vacancies more rapidly than those of the two-sided sorting model, the relocation patterns of workers and firms remain comparable. Population density and the measure of vacancies in the top 10% of locations increase by 10.9% and 10.0%, respectively.

Failing to account for firm sorting, however, leads to an overly optimistic assessment of changes in output and welfare. When workers relocate to high-ℓ locations, they are able to produce more output due to higher local TFP, and aggregate output increases by 0.40%. Thus, in a one-sided sorting economy, if congestion costs are not taken into account, it is always advantageous to have more workers and firms in high-ℓ locations, which is not the case in the two-sided sorting model. This gain in output ultimately results in an increase in aggregate welfare of 0.37%. The decomposition exercise in Figure 6 confirms this result. Housing market has a negative impact on welfare although its effect is much smaller than under two-sided sorting due to the lower concentration in high-ℓ locations. The right panel of Figure 7 also shows that most of workers are worse off because of housing market. The most notable difference is the substantial benefit from the location component, driven by a higher concentration of employment in productive areas.

Federal income tax. Besides the housing regulation, my model can be applied to examine various spatial policies. In Appendix D, I investigate the effects of federal income tax on workers. In the U.S., workers in the top 10% of dense regions pay, on average, 60% higher federal taxes than workers in the bottom 10% (CPS

[43] Overhead costs in the one-sided sorting model $c^{\text{one}}(\ell)$ are calibrated so that homogeneous firms' values are equalized across locations, and thus differ from $c(\ell)$. See Appendix D for more details.

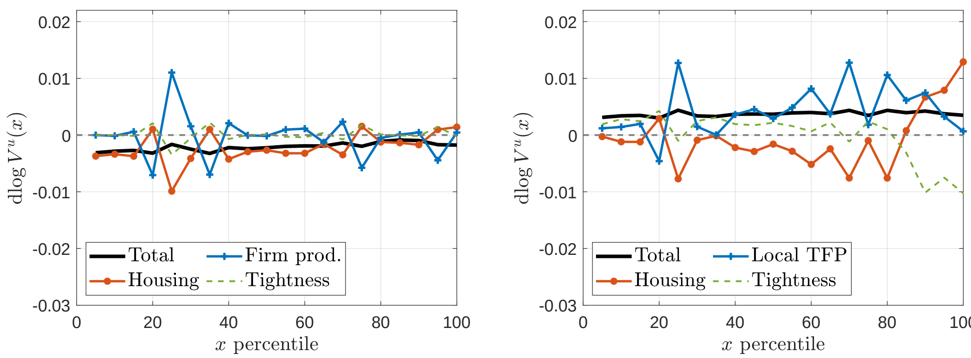


Figure 7. Impact of Relaxing Housing Regulation on Welfare

March, 2017). This tax discrepancy can have adverse effects on employment in dense areas for two reasons. First, higher rents in dense cities partially offset the benefits of higher nominal wages, which renders taxes based on nominal income disproportionately burdensome for workers in those regions. Second, if income tax rates are progressive, they further diminish the incentives for workers to choose densely populated areas with better job prospects and higher expected wages. Albouy (2009) argues that the federal income tax reduces employment in high-wage areas and leads to inefficiencies. I revisit this issue in the context of the two-sided sorting of workers and firms.

When income tax becomes less progressive and overall tax rates are reduced (See Figure A.4), both workers and firms simulatenously reloate to high-ℓ locations. Thus, output remains unchanged while aggregate congestion costs in housing markets increase, which leads to a decrease in welfare. In contrast, in the one-sided sorting model, the gains from higher local TFP dominate the increase in congestion costs. As a result, aggregate welfare improves in this model. This stark contrast in results underscores how two-sided sorting, which allows both heterogeneous workers and firms to respond to policies, can provide qualitatively different normative implications.

7. Conclusion

In this paper, I show that two-sided sorting of heterogeneous workers and firms alone can account for spatial disparities in several dimensions, including productivity, income, and population density. When both workers and firms simultaneously sort across space, the sorting of one side sustains the sorting of the other side, without requiring other factors as a source of sorting. In addition, allowing both groups to relocate across space provides unique normative implications for government policy.

I characterize the equilibrium that exhibits positive assortative matching between workers and firms across space. When workers and firms randomly match within local labor markets, their location choices determine their expected income and profits. These benefits are larger in locations inhabited by more productive workers and firms, which in turn draw a larger number of workers and firms. Due to production complementarity, highly productive agents gain relatively more and are willing to pay higher costs to reside or operate in these areas.

Two-sided sorting presents unique policy implications because both workers and firms, which embody productivity, can relocate in response to government policy. The government can reduce unnecessary congestion costs without reducing output by dispersing productive workers and firms to less densely populated locations. However, ignoring the sorting of one side leads to a different conclusion. On the one hand, spatial disparities can be equally well explained, for example, by introducing local TFP instead. On the other hand, because the productivity of one side is inherent to locations, dispersing economic activities to less densely populated areas leads to substantial output losses, which may outweigh any gains from reduced congestion costs.

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.

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.

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.

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.

Busso, Matias, Jesse Gregory, and Patrick Kline (2013) "Assessing the Incidence and Efficiency of a Prominent Place Based Policy," *American Economic Review*, 103 (2), 897–947.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline (2018) "Firms and Labor Market Inequality: Evidence and Some Theory," *Journal of Labor Economics*, 36 (S1), S13–S70.

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,"Technical report, National Bureau of Economic Research.

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*.

Dauth, Wolfgang, Sebastian Findeisen, Enrico Moretti, and Jens Suedekum (2022) "Matching in Cities," *Journal of the European Economic Association*, 20 (4), 1478–1521.

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

——— (2020) "The Comparative Advantage of Cities," *Journal of International Economics*, 123.

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 Enrico Moretti (2021) "Where is Standard of Living the Highest? Local Prices and the Geography of Consumption," *Working Paper*.

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]."

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

Glaeser, Edward L., and David C. Maré (2001) "Cities and Skills," *Journal of Labor Economics*, 19 (2), 316–342.

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.

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 Eevelopment, 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.

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*.

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.

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]."

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.

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).

Appendix

A. Omitted Proofs

A.1 Derivations

The match surplus of workers and vacancies solve

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

where $\tilde{\rho} \equiv \rho + \delta$. Combining the two gives the HJB equation for the joint surplus, $S = V^e - V^u + V^p - V^v$,

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

where I use the bargaining solution and (1) for the second line. Distinguishing the productivity of workers and firms of a given match (x,y) from productivity of local workers and firms $(x(\ell),y(\ell))$ is important for determining the outside option value. Under a pure assignment, $x = x(\ell)$ and $y = y(\ell)$, and the surplus of a match in ℓ simplifies to

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

where I define $1 - \tilde{\beta} \equiv \frac{\rho}{\rho+\delta_v}(1-\beta)$ to simplify the notation. Combining this equation with (A.2), I obtain the below:

$$\begin{aligned}
S(x(\ell),y,\ell) &= \frac{1}{\tilde{\rho} + (1-\tilde{\beta})q(\ell)}\left((y-b)x(\ell) - \frac{\beta\lambda(\ell)}{\tilde{\rho} + \beta\lambda(\ell) + (1-\tilde{\beta})q(\ell)}(y(\ell)-b)x(\ell)\right), \\
S(x,y(\ell),\ell) &= \frac{1}{\tilde{\rho} + \beta\lambda(\ell)}\left((y(\ell)-b)x - \frac{(1-\tilde{\beta})q(\ell)}{\tilde{\rho} + \beta\lambda(\ell) + (1-\tilde{\beta})q(\ell)}(y(\ell)-b)x(\ell)\right), \\
S(x,y,\ell) &= (y-b)x - \frac{\beta\lambda(\ell)}{\tilde{\rho} + \beta\lambda(\ell) + (1-\tilde{\beta})q(\ell)}x(y(\ell)-b) - \frac{(1-\tilde{\beta})q(\ell)}{\tilde{\rho} + \beta\lambda(\ell) + (1-\tilde{\beta})q(\ell)}(y-b)x(\ell) \\
&\quad + \frac{\beta\lambda(\ell)(1-\tilde{\beta})q(\ell)}{\tilde{\rho} + \beta\lambda(\ell) + (1-\tilde{\beta})q(\ell)}\left(\frac{1}{\tilde{\rho} + \beta\lambda(\ell)} + \frac{1}{\tilde{\rho} + (1-\tilde{\beta})q(\ell)}\right)(y(\ell)-b)x(\ell).
\end{aligned}$$

Using these expressions and (A.1), I can solve for wages,

$$
\begin{aligned}
w(x,y,\ell) &= bx + \beta(y-b)x + (1-\beta)\beta\lambda(\ell)S(x,y(\ell),\ell) - \beta(1-\tilde{\beta})q(\ell)S(x(\ell),y,\ell) \\
&= bx + \beta(y-b)x + \beta\left(\frac{\lambda(\ell)(1-\beta)}{\tilde{\rho}+\beta\lambda(\ell)}(y(\ell)-b)x - \frac{(1-\tilde{\beta})q(\ell)}{\tilde{\rho}+(1-\tilde{\beta})q(\ell)}(y-b)x(\ell)\right) \\
&\quad + \frac{\beta\lambda(\ell)(1-\tilde{\beta})q(\ell)}{\tilde{\rho}+\beta\lambda(\ell)+(1-\tilde{\beta})q(\ell)}\left(\frac{-(1-\beta)}{\tilde{\rho}+\beta\lambda(\ell)} + \frac{\beta}{\tilde{\rho}+(1-\tilde{\beta})q(\ell)}\right)(y(\ell)-b)x(\ell).
\end{aligned}
$$

Imposing $x = x(\ell)$ and $y = y(\ell)$, the equilibrium wages simplify to

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

Finally, I can solve for the value of workers and firms when choosing ℓ. First, plug in $S(x,y(\ell),\ell)$ ($S(x(\ell),y,\ell)$, respectively) to the equation $V^u(x,\ell)$ ($V^v(y,\ell)$, respectively). The value of workers choosing location ℓ simply equals the value of unemployed workers, $V^u(x,\ell)$. The flow value of firms choosing ℓ, $\rho\bar{V}^v(y,\ell)$, equals the value of posting δ_v effective units of vacancies, $\delta_v V^v(y,\ell)$, net of overhead costs, $c(\ell)$.

$$
\begin{aligned}
\rho V^u(x,\ell) &= bx + \underbrace{\frac{\beta\lambda(\ell)}{\tilde{\rho}+\beta\lambda(\ell)}(y(\ell)-b)}_{\equiv A_w(y(\ell),\lambda(\ell))}\left(x - \underbrace{\frac{(1-\tilde{\beta})q}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}x(\ell)}_{\equiv B_w(x(\ell),\theta(\ell))}\right) - \bar{h}r(\ell) + \Pi, \\
\rho\bar{V}^v(y,\ell) &= \underbrace{\frac{\delta_v}{\rho}x(\ell)\frac{(1-\tilde{\beta})q(\ell)}{\tilde{\rho}+(1-\tilde{\beta})q(\ell)}}_{\equiv A_f(x(\ell),q(\ell))}\left(y - b - \underbrace{\frac{\beta\lambda}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}(y(\ell)-b)}_{\equiv B_f(y(\ell),\theta(\ell))}\right) - c(\ell).
\end{aligned}
$$

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

A.2 Proof of Proposition 1

I prove Proposition 1 in steps. First, I show that a pure assignment equilibrium, if it exists, should be PAM. Then, the equilibrium can be characterized by two strictly increasing location-matching functions, $x(\ell)$ and $y(\ell)$. Second, I prove the existence. Finally, I show that population density and wages are increasing in ℓ.

PAM between workers and firms. I first show that if a pure assignment equilibrium exists, it has to be PAM. To show this, I proceed by contradiction. That is, suppose there exist two locations $\ell_1 > \ell_2$ and two types of workers and firms such that $x(\ell_1) = x^g > x(\ell_2) = x^b$ while $y(\ell_1) = y^b < y(\ell_2) = y^g$.

The location choices of workers and firms imply that $A_w(\ell_1) > A_w(\ell_2)$ and $A_f(\ell_2) > A_f(\ell_1)$. Given that firm is worse in ℓ_1, for $A_w(\ell)$ to be larger in ℓ_1, job arrival rate should be sufficiently higher, $\lambda(\ell_1) > \lambda(\ell_2)$, and thus $q(\ell_2) > q(\ell_1)$. These observations imply that $B_w(\ell) = \frac{\tilde{\rho}}{\delta_v}\frac{1-\tilde{\beta}}{1-\beta}\frac{\tilde{\rho}+(1-\tilde{\beta})q(\ell)}{\tilde{\rho}+\beta\lambda(\ell)+(1-\tilde{\beta})q(\ell)}A_f(\ell)$ is larger in ℓ_2. Thus, $A_w(\ell)(x - B_w(\ell))$ is supermodular in (ℓ, x), log-submodular in (ℓ, x), and increasing in x, implying that it is increasing in ℓ. I conclude that $r(\ell_1) > r(\ell_2)$ for ℓ_2 to be chosen by workers x^b.

Similarly, $B_f(\ell) = \frac{\tilde{\rho}+\beta\lambda(\ell)}{\tilde{\rho}+\beta\lambda(\ell)+(1-\tilde{\beta})q(\ell)}A_w(\ell)$ is larger in ℓ_1 from $A_w(\ell_1) > A_w(\ell_2)$ and $\lambda(\ell_1) > \lambda(\ell_2)$, which implies that $A_f(\ell)(y - b - B_f(\ell))$ is submodular in (ℓ, y), log-supermodular in (ℓ, y), and increases in y. Thus, I conclude that it is decreasing in ℓ, which then implies $c(\ell_2) > c(\ell_1)$ for firms g^b to choose ℓ_1.

As production costs of housing and business services are convex, two inequalities on $r(\ell)$ and $c(\ell)$ imply that $L(\ell_1) > L(\ell_2)$ and $N(\ell_1) < N(\ell_2)$. Differentiating $\theta(\ell) = \frac{V(\ell)}{u(\ell)L(\ell)}$ with respect to ℓ, and using steady-state flow-balance conditions (6),

$$\frac{\theta'(\ell)}{\theta(\ell)} = \frac{\lambda(\ell)+\delta}{-\varepsilon_q(\theta(\ell))\lambda(\ell)+\delta}\left(\frac{N'(\ell)}{N(\ell)} - \frac{L'(\ell)}{L(\ell)}\right), \tag{A.3}$$

where I define $\varepsilon_q(\theta) = \frac{q'(\theta)}{q(\theta)}\theta$ as the elasticity of vacancy contact rates. Similarly, I define the elasticity of job arrival rates, $\varepsilon_\lambda(\theta) = \frac{\lambda'(\theta)}{\lambda(\theta)}\theta$, for later. Observing $\varepsilon_\lambda > 0$ and $\varepsilon_q < 0$, θ is increasing in $\frac{N}{L}$ in steady state. Since I established $\frac{N(\ell_2)}{L(\ell_2)} > \frac{N(\ell_1)}{L(\ell_1)}$, the above equation implies $\theta(\ell_2) > \theta(\ell_1)$. This contradicts to $\lambda(\ell_1) > \lambda(\ell_2)$.

Thus, I conclude that workers and firms positively sort across space.

Existence. I show the existence of a pure assignment equilibrium in two steps. In the first step, I show that the equilibrium can be represented by a solution of differential equations. In the second step, I show that there exists an initial condition that satisfies the boundary conditions on worker and firm productivity.

Step 1 Given that a pure assignment equilibrium exhibits PAM, I can focus on $(x(\ell), y(\ell))$, which are increasing in ℓ. Because of the continuity of the productivity distributions, Q_w and Q_f, these two functions are strictly increasing.

Observe that $x(\ell)$ and $y(\ell)$ are differentiable almost everywhere as increasing. Guess that these two functions are twice differentiable, which will be verified later. Then, location choices can be characterized by the below first order conditions of workers and firms,

$$[f_w] \quad A_w'(\ell)x(\ell) - (A_w(\ell)B_w(\ell))' - \bar{h}C_r'(\bar{h}L(\ell))L'(\ell) = 0,$$
$$[f_f] \quad A_f'(\ell)(y(\ell) - b) - (A_f(\ell)B_f(\ell))' - C_v(N(\ell))N'(\ell) = 0,$$

together with the boundary conditions, $x(\bar{\ell}) = \bar{\ell}, x(\underline{\ell}) = \underline{x}$, $y(\bar{\ell}) = \bar{\ell}$, and $y(\underline{\ell}) = \underline{y}$, as well as steady state flow-balance conditions (6). Note that I already plugged in housing and business services market clearing conditions.

I can summarize two conditions by defining a function f, $f(z, z', z'') = 0 \in \mathbb{R}^2$ where $z(\ell) = (x(\ell), y(\ell))$. 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 such that $z'' = g(z, z')$ by Implicit Function Theorem. From now on, I omit location index ℓ unless it creates any confusions.

To show that $\det D_{z''}f(z,z',z'') = \det \begin{pmatrix} \frac{\partial f_w}{\partial x''}, \frac{\partial f_w}{\partial y''} \\ \frac{\partial f_f}{\partial x''}, \frac{\partial f_f}{\partial y''} \end{pmatrix} \neq 0$, obtain several useful expressions,

$$A'_w(\ell) = A_w(\ell)\left(\varepsilon_\lambda(\theta)\frac{\theta'}{\theta}\frac{\tilde{\rho}}{\tilde{\rho}+\beta\lambda} + \frac{y'}{y-b}\right),$$
$$B'_w(\ell) = B_w(\ell)\left(-\frac{-\tilde{\rho}\frac{\theta'}{\theta}\varepsilon_q(\theta)+\beta\lambda/\theta}{\tilde{\rho}+\beta\lambda(\theta)+(1-\tilde{\beta})q(\theta)} + \frac{x'}{x}\right),$$
$$A'_f(\ell) = A_f(\ell)\left(\varepsilon_q(\theta)\frac{\theta'}{\theta}\frac{\tilde{\rho}}{\tilde{\rho}+(1-\tilde{\beta})q} + \frac{x'}{x}\right),$$
$$B'_f(\ell) = B_f(\ell)\left(\frac{\tilde{\rho}\varepsilon_\lambda(\theta)\frac{\theta'}{\theta}+(1-\tilde{\beta})q/\theta}{\tilde{\rho}+\beta\lambda(\theta)+(1-\tilde{\beta})q(\theta)} + \frac{y'}{y-b}\right),$$

where I use $\frac{\partial L'}{\partial x''} = M_w q_w(x)$ and $\frac{\partial N'}{\partial y''} = M_f q_f(y)$.

Evaluating worker and firm sorting conditions at $x = x(\ell)$ and $y = y(\ell)$,

$$\frac{A_w x}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}\left((\tilde{\rho}+\beta\lambda)\frac{y'}{y-b} - (1-\tilde{\beta})q\frac{x'}{x} + \tilde{\rho}\frac{\theta'}{\theta}\left(\varepsilon_\lambda(\theta) - \frac{(1-\tilde{\beta})q}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}\varepsilon_q\right) - \frac{(1-\tilde{\beta})q}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}\beta\frac{\lambda}{\theta}\right) - \bar{h}r' = 0,$$
$$\frac{A_f(y-b)}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}\left((\tilde{\rho}+(1-\tilde{\beta})q)\frac{x'}{x} - \beta\lambda\frac{y'}{y-b} - \tilde{\rho}\frac{\theta'}{\theta}\left(-\varepsilon_q(\theta) + \frac{\beta\lambda}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}\varepsilon_\lambda\right) - \frac{\beta\lambda}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}(1-\tilde{\beta})\frac{q}{\theta}\right) - c' = 0.$$

From (A.3), $\frac{\theta'}{\theta} = \zeta(z,z')\left(\frac{N'}{N} - \frac{L'}{L}\right)$, where $\zeta(z,z') \equiv \frac{\lambda+\delta}{-\varepsilon_q(\theta)\lambda+\delta}$ which is always strictly positive. Differentiating this term with respect to x'' and y'', I obtain $\frac{\partial\theta'/\theta}{\partial x''} = -\zeta\frac{M_w q_w}{L}$ and $\frac{\partial\theta'/\theta}{\partial y''} = \zeta\frac{M_f q_f}{N}$, which then gives the below.

$$\det \begin{pmatrix} \frac{\partial f_w}{\partial x''}, \frac{\partial f_w}{\partial y''} \\ \frac{\partial f_f}{\partial x''}, \frac{\partial f_f}{\partial y''} \end{pmatrix} = \det \begin{pmatrix} -a_w b_w - \bar{h}M_w q_w C'_r(\bar{h}L) & a_f b_w \\ a_w b_f & -a_f b_f - M_f q_f C'_v(N) \end{pmatrix},$$

where I define four positive terms: $a_w = \zeta\frac{M_w q_w}{L}$, $a_f = \zeta\frac{M_f q_f}{N}$, $b_w = \frac{A_w x\tilde{\rho}}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}\left(\varepsilon_\lambda - \frac{(1-\tilde{\beta})q}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}\varepsilon_q\right)$, and $b_f = \frac{A_f(y-b)\tilde{\rho}}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}\left(-\varepsilon_q + \frac{\beta\lambda}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q}\varepsilon_\lambda\right)$. It is easy to check that the determinant of the matrix is positive when cost functions are convex—i.e., $C'_r, C'_v > 0$.

Finally, observe that $D_{z''}f$ is continuous when C_r, C_v, Q_w, and Q_f are twice continuously differentiable, and z' is continuously differentiable, which is assumed at this point. Then, I can verify that for a given $(z(\ell), z'(\ell))$, $z''(\cdot)$ exists and is continuous by the theorem.

Step 2 From step 1, I can represent the first order condition of location choices as differential equations of (z, z'), i.e., $(z', z'') = (z', g(z, z'))$. Find $\bar{z}'$ such that when $z' > \bar{z}'$, $C'_r(\bar{h}M_w q_w(x(\ell))x'(\ell)) > \overline{xy}$ and $C'_v(M_f q_f(y(\ell))y'(\ell)) > \overline{xy}$. Then, an equilibrium z' is bounded above by $\bar{z}'$. If not, workers or firms would always deviate to cheaper locations even at the expense of lower wages or profits. As $g(z, z')$ is continous and defined on a compact set, it is Lipschitz continous.

Given an initial condition $(\underline{z}, \underline{z}')$, the equilibrium exists from the standard ODE theorem. However, instead of two initial conditions, I have the initial and terminal conditions on z, which are $z(\underline{\ell}) = \underline{z}$ and $z(\bar{\ell}) = \bar{z}$. I will show that there exists at least one initial condition $z'(\underline{\ell})$ that gives $z(\bar{\ell}) = \bar{z}$. For this purpose, I show that guessing a too small (large) initial condition $z'(\underline{\ell})$ gives a too small (large) terminal point $z(\overline{\ell})$. Then, I use the Poincaré-Miranda theorem, which is a generalization of the intermediate value theorem.

Step 2 (i) The first part is to show that when an initial guess is too small, $z'(\underline{\ell}) \to 0$, a terminal point $z(\overline{\ell})$ becomes smaller than $\overline{z}$.

Let g^x represents the first component of g, i.e., x''. For now, suppose that $y(\ell)$ is given. By Lipschitz continuity, there exists K such that $|g^x(z,a) - g^x(z,b)| \le K|a-b|$ for all $a, b \in [0, (\overline{z}')^x]$. Make a guess on $x'(\underline{\ell}) = \varepsilon$. Then, $|g^x(z,z')| \le |g^x(z,\varepsilon)| + K|z' - \varepsilon|$. Define $\sup_{x\in[\underline{x},\overline{x}]} |g^x(z,\varepsilon)| = g_\varepsilon$. Then,

$$\begin{aligned} z'(\ell) &= z'(\underline{\ell}) + \int_{\underline{\ell}}^{\ell} z''(s)\,\mathrm{d}s \le z'(\underline{\ell}) + \int_{\underline{\ell}}^{\ell} (g_\varepsilon + K\varepsilon + K|z'(s)|)\,\mathrm{d}s \\ &= \varepsilon + (g_\varepsilon + K\varepsilon)\ell + K\int_{\underline{\ell}}^{\ell} z'(s)\,\mathrm{d}s = \varepsilon + (g_\varepsilon + K\varepsilon)\ell + K(z(\ell) - z(\underline{\ell})). \end{aligned}$$

Define an auxiliary function $\hat{z}(\ell)$ s.t. (1) $\hat{z}(\underline{\ell}) = z(\underline{\ell})$ and (2) $\hat{z}'(\ell) = K\hat{z}(\ell) + \varepsilon + (g_\varepsilon + K\varepsilon)\ell - Kz(\underline{\ell})$. A function defined as the difference between z and $\hat{z}$, $t(\ell) = \hat{z}(\ell) - z(\ell)$, satisfies $t(\underline{\ell}) = t'(\underline{\ell}) = 0$ and $t'(\ell) \ge Kt(\ell)$. Finally, define $T(\ell) = t(\ell)e^{-K(\ell-\underline{\ell})}$. As $T'(\ell) = (t'(\ell) - Kt(\ell))e^{-K\ell} \ge 0$, I conclude that $T(\ell) = t(\ell)e^{-K\ell} \ge T(0) = 0$, which leads to $t(\ell) \ge 0$. Thus, $\hat{z}(\ell) \ge z(\ell)$.

The solution of $\hat{z}'(\ell) = K\hat{z}(\ell) + a\ell + b$, where $a = g_\varepsilon + K\varepsilon, b = \varepsilon - K\underline{z}$, is given by

$$\hat{z}(\ell) = c_1 e^{K\ell} - \frac{a}{K}\ell + c_2.$$

I have two initial conditions $\hat{z}(\underline{\ell}) = \underline{z}, \hat{z}'(\underline{\ell}) = \varepsilon$, which give $c_1 = \frac{1}{K}(2\varepsilon + \frac{g_2}{K})$ and $c_2 = \underline{z} - \frac{1}{K}(2\varepsilon + \frac{g_2}{K})$. As a result, the terminal value of $\hat{z}$ becomes

$$\hat{z}(\overline{\ell}) = \underline{z} + \frac{2\varepsilon + g_\varepsilon}{K}(e^K - 1) - \Big(\frac{g_\varepsilon}{K} + \varepsilon\Big).$$

The final step of the first part is to show that $g_\varepsilon = \sup_z |g^x(z,\varepsilon)|$ converges to zero when ε goes to zero. It is sufficient to show that $\lim_{\varepsilon\to 0} g^x(z,\varepsilon) = 0$ for all x, as g^x is a continuous function defined on a compact set. The problem of firms is symmetric, and $\lim_{\varepsilon\to 0} g^y(z,\varepsilon) = 0$ for all y, where g^y represents the second component of g, i.e., y'', for a given $x(\ell)$, is sufficient.

I first show the limit holds for workers g^x. When x' goes to zero, population density $L = M_w q_w(x(\ell))x'(\ell)$ goes to zero as well, as $q_w(x(\ell)) \le \sup_{x\in[\overline{x},\overline{x}]} Q_w'(x)$ is bounded due to the assumption $Q_w \in \mathcal{C}^2$. Then, equation (A.3) gives $\frac{\theta'}{\theta} \to \frac{1}{-\varepsilon_q}\Big(\frac{N'}{N} - \frac{L'}{L}\Big)$, and the worker sorting condition becomes,

$$(y-b)x\frac{y'}{y-b} = (y-b)x\Big(\frac{\tilde{\rho}}{\beta}\frac{\varepsilon_\lambda}{-\varepsilon_q}\frac{1}{\lambda}\frac{1}{L} + \bar{h}C_r''(\bar{h}L)\Big)L'.$$

The left-hand side is bounded for a given y while C_r'' on right-hand side diverges to infinity when population density goes to zero, implying that $L' = M_w(q_w'(x)(x')^2 + q_w(x)x'')$ should converge to zero. Thus, I conclude that $g^x = x''$ should converge to zero.

I follow the same steps to obtain the result for firms g^y. The measure of firms N converges to zero when y' goes to zero, and thus $\frac{\theta'}{\theta} \to \frac{N'}{N} - \frac{L'}{L}$. The firm sorting condition becomes

$$\frac{\delta_v x}{\rho+\delta_v}\frac{1-\beta}{1-\tilde{\beta}}(y-b)\left(\frac{x'}{x} - \frac{\beta}{1-\tilde{\beta}}\right) = \left(\frac{\delta_v x}{\rho+\delta_v}\frac{1-\beta}{1-\tilde{\beta}}(y-b)\frac{-\varepsilon_q\tilde{\rho}}{(1-\tilde{\beta})q}\frac{1}{N} + C_v''(N(\ell))\right)N'(\ell).$$

As C_v'' diverges to infinity, the above equality indicates that $g^y = y''$ should converge to zero.

<u>*Step 2 (ii)*</u> The second part is to show that sending an initial condition $z'(\underline{\ell})$ to infinity leads to a terminal point that is larger than $\bar{z}$. First, I show the worker side. Consider a given $y(\ell)$ and a location ℓ. When $x'(\ell)$ goes to infinity, the firm sorting condition becomes,

$$x(y-b)\left(\frac{x'}{x} - \frac{\tilde{\rho}(-\varepsilon_q)}{(1-\tilde{\beta})q}\frac{\theta'}{\theta} - \frac{\beta}{1-\tilde{\beta}}\right) = C_v''(N)N'(\ell).$$

For the above equality to hold, I have $\frac{\tilde{\rho}(-\varepsilon_q)}{(1-\tilde{\beta})q}\frac{\theta'}{\theta} = \frac{x'}{x} - \frac{\beta}{1-\tilde{\beta}} + o(x')$ where $o(x')$ is the term that converges to 0 when $x' \to \infty$. Plugging this into the sorting condition of workers and taking the the limit,

$$\frac{\beta\lambda}{\tilde{\rho}+\beta\lambda}(y-b)\left(\frac{\varepsilon_\lambda}{-\varepsilon_q}\frac{x'}{x} - \left(2 + \frac{\varepsilon_\lambda}{-\varepsilon_q}\right)\frac{\beta}{1-\tilde{\beta}} + o(x')\right) = \bar{h}C_r''(\bar{h}L)L'.$$

In the limit, the left-hand side is positive, and I conclude that $L'(\ell)$ is weakly positive. In short, when worker productivity increases infinitely as ℓ increases, the market tightness also increases for the firm sorting condition to hold. Thus, workers prefer higher-ℓ locations, which implies increasing housing rents and population density.

Similarly, I can show N is increasing in ℓ when y' goes to infinity for a given $x(\ell)$, plugging in $-\tilde{\rho}\frac{\theta'}{\theta} = \frac{\beta\lambda}{\varepsilon_\lambda}\frac{y'}{y-b} + o(y')$, obtained from the worker sorting condition, into the firm sorting condition,

$$\frac{\delta_v}{\tilde{\rho}}\frac{(1-\beta)q}{\tilde{\rho}+(1-\tilde{\beta})q}x\left(\frac{-\varepsilon_q}{\varepsilon_\lambda}\frac{y'}{y-b} + o(y')\right) = C_v''(N)N',$$

where $o(y')$ is the term that converges to 0 when $y' \to \infty$.

From the above, $L(\ell) = M_w q_w(x(\ell))x'(\ell)$ is increasing at ℓ when $x'(\ell)$ diverges to infinity. As a result, if $x'(\underline{\ell})$ is sufficiently large, L' will be increasing in $\ell \in [\underline{\ell}, \bar{\ell}]$. Then, below inequality holds for all ℓ,

$$x'(\ell) \geq \frac{q(x(\underline{\ell}))}{q(x(\ell))}x(\underline{\ell}) \geq \frac{q_w(x(\underline{\ell}))}{\sup_{x\in[\underline{x},\bar{x}]} q_w(x)}x'(\underline{\ell}).$$

The above implies that $x(\overline{\ell}) > \overline{x}$. Following the same step, the same is true for firms. Thus, I conclude that if an initial condition is sufficiently large for a given schedule of the other, the terminal point would be larger than the maximum of the productivity level.

Step 3 Consider a continuous mapping $p_a(a'(\underline{\ell})) = a(\overline{\ell}) - \overline{a}$ for $a = x, y$, where $a(\overline{\ell})$ is given by the solution of ODE defined above by sorting conditions. The continuity of this function is guaranteed by the continuity of $g(\cdot)$. From the step 2-(i), the mapping p_a gives negative values when $a'(\underline{\ell})$ goes to zero. From the step 2-(ii), I can find a sufficiently large N such that the mapping p_a gives a positive value when a' is larger than N. By Pointcaré-Miranda theorem, there exists $z'(\underline{\ell}) \in [0, N]^2$ that gives $z(\overline{\ell}) = \overline{z}$. Using these values as an initial point, the standard ODE theorem gives the existence of an equilibrium. Note that this result does not guarantee the uniqueness.

Population density and wages. Housing rents and thus population density increase in ℓ, $r'(\ell) > 0$, when $A'_w(\ell)x(\ell) - (A_w(\ell)B_w(\ell))' > 0$. Otherwise, all workers would prefer higher-ℓ locations. The sufficient condition for this inequality is

$$A'_w(\ell)\underline{x} > A'_w(\ell)B_w(\ell) + A_w(\ell)B'_w(\ell).$$

Under PAM, $A'_w(\ell) > 0$ for all ℓ. Note that the right-hand side converges to zero when δ_v diverges to infinity so that $1 - \tilde{\beta}$ converges to zero. Thus, there exists $\overline{\delta}_v$ such that population density increases when $\delta_v > \overline{\delta}_v$.

I now show wages also increase in ℓ. Under PAM, wage increases in ℓ if

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

The above converges to $A'_w(\ell)x(\ell)$ when $1 - \tilde{\beta}$ converges to zero, which is positive as A_w is increasing in ℓ. Thus, I can find $\overline{\delta}'_v$ such that the above inequality holds if $\delta_v > \overline{\delta}'_v$.

Mixed Equilibrium. Although I focus on an equilibrium with pure assignment, there exist multiple equilibria. For example, an economy where all workers and firms randomly choose location ℓ is a mixed equilibrium. However, I show below that a mixed equilibrium, even if it exists, do not have equilibrium properties that are consistent with the data, as opposed to a pure assignment equilibrium as shown in Proposition 1.

As workers and firms are randomly matched within local market, when there can be multiple types of workers and firms in a given ℓ, they consider an average productivity. Workers' job opportunity $A_w(\mathbb{E}[y|\ell], \lambda(\ell))$ is a function of average firm productivity, $\mathbb{E}[y|\ell]$, instead of $y(\ell)$. Similarly, firms' hiring opportunity $A_f(\mathbb{E}[x|\ell], q(\ell))$, is determined by average worker types $\mathbb{E}[x|\ell]$ and vacancy contact rate $q(\ell)$. In turn, the value function of workers and firms choosing a location ℓ have the same expression as in (3) and (4), except $x(\ell)$ and $y(\ell)$ are replaced with $\mathbb{E}[x|\ell]$ and $\mathbb{E}[y|\ell]$.

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

Suppose that there exists a mixed equilibrium where $\mathbb{E}[x|\ell], \mathbb{E}[y|\ell]$, and $L(\ell)$ strictly increase in ℓ. Then, we can choose a pair of locations $\ell' < \ell''$ where either worker or firm distribution is not singleton, and there exists either x^* or y^* who chooses both locations, ℓ' and ℓ'', with positive probablities.

$A_w(\ell)$ and $A_f(\ell)$ are weakly larger in ℓ''. If not, i.e., $A_w(\ell') > A_w(\ell'')$ or $A_f(\ell') > A_f(\ell'')$, then due to single-crossing property of the value functions, $\mathbb{E}[x|\ell'] > \mathbb{E}[x|\ell'']$ or $E[y|\ell'] > E[y|\ell'']$, and this allocation violates the required properties.

Suppose that x^* chooses both ℓ' and ℓ''. Then, job opportunities must be equal in two locations, $A_w(\ell') = A_w(\ell'')$ because there exists at least one different type of workers. For example, if there are workers with higher productivity thatn x^* and $A_w(\ell'') > A_w(\ell')$, x^* workers prefers ℓ'. Since $r(\ell'') > r(\ell')$, vacancy contact rate should be smaller in ℓ'' so taht the cost of workers for choosing ℓ, i.e., $A_w B_w + r = A_w \frac{\tilde{\rho}}{\delta_v} \frac{1-\tilde{\beta}}{1-\beta} \frac{\tilde{\rho}+(1-\tilde{\beta})q}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q} A_f + r$, is not strictly larger in ℓ''. Then, $V(\ell)/L(\ell)$ is larger in ℓ'', and so does $c(\ell'')$. Moreover, $A_f B_f = \frac{\tilde{\rho}+\beta\lambda}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q} A_w$ is larger in ℓ''. Combining two observations, the firms' cost of choosing ℓ given by $A_f B_f + c$ is larger in ℓ''. For some firms to choose ℓ'', $A_f(\ell'')$ is larger than $A_f(\ell')$, and there exists only a unique type of firm in each of two locations. Because both $\lambda(\ell)$ and $\mathbb{E}[y|\ell]$ are higher in ℓ'', so does workers' job opportunity A_w. Contradiction.

Similarly, suppose that y^* chooses both ℓ' and ℓ''. In this case, firms' hiring opportunities are equal in two locations, i.e., $A_f(\ell') = A_f(\ell'')$. In turn, to ensure firm sorting condition holds, the cost of choosing two locations $A_f B_f + c = \frac{\tilde{\rho}+\beta\lambda}{\tilde{\rho}+\beta\lambda+(1-\tilde{\beta})q} A_w + c$ should be equal as well. This requires that vacancy contat rate $q(\ell)$ (job arrival rate $\lambda(\ell)$) should larger (smaller) in ℓ''. If not, $A_f B_f$ is larger due to $\lambda(\ell'') > \ell(\ell')$ and $A_w(\ell'') \geq A_w(\ell')$. Moreover, a higher V/L in ℓ'' and $L(\ell'') > L(\ell')$ together imply $V(\ell'') > V(\ell')$, which leads to $c(\ell'') > c(\ell')$. Then, for $A_f(\mathbb{E}[x|\ell], q(\ell))$ to be equal in two locations, $\mathbb{E}[x|\ell''] < \mathbb{E}[x|\ell']$, which contradicts the first required property.

A.3 Mobility of Workers and Firms

In this section, I formally discuss how the assumption on mobility changes the equilibrium. I show that equilibrium outcomes (worker/firm allocation and wages evaluated at the equilibrium allocation) do not depend on whether workers and firms can freely move or not.

Focus on the case that $\delta_v = 0$ to save the notation. Let $V_0^u(x)$ denote the value of an unemployed worker of x (when choosing the optimal location), and similarly, let $V_0^v(y)$ denote the value of a vacant position (when choosing the optimal location). Denote the optimal location choice of workers and firms as ℓ_x and ℓ_y with some abuse of notation.

Proposition A.2. *The location choices of workers and firms $(x(\ell), y(\ell))$ are optimal in an economy where they should stay in the same location (without 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 the same on the equilibrium path.*

The above 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).

To prove the above proposition, it is useful to first formally characterize the location choices of workers under two cases. Consider **Problem 0** (P0), an economy with free mobility. All assumptions remain the same, except the assumption that workers and firms can change their location at each point in time. For given allocations $(x(\ell), y(\ell))$, the values of workers (unemployed, employed) are characterized by

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

where $\mathrm{d}t$ represents the small interval of time, and I focus on the case that $V_0^e(x,\ell) \geq V_0^u(x,\ell_x)$. Note that the optimal choice ℓ_x given by $\ell_x = \mathrm{argmax}_\ell\, V_0^u(x,\ell)$ by definition. I omit $y(\ell)$ to emphasize that the above values are computed for a given $y(\ell)$. The values of firms (vacant, filled) on the equilibrium path under the assignment $(x(\ell), y(\ell))$ are

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

Note that the values of vacant position takes into account that the type of worker in location ℓ is given by $x(\ell)$ (not x). Then, the match surplus is defined as

$$S_0(x, y(\ell)\ell) \equiv V_0^e(x,\ell) - V_0^u(x,\ell_x) + V_0^p(x,y(\ell),\ell) - V_0^v(y(\ell),\ell).$$

Note that the threat point of workers is given by their values when they deviate to ℓ_x.

Next, consider **Problem 1** (P1) without mobility. Workers and firms should stay in the same location once they choose their locations. Then, the values of workers (unemployed, employed) and firms (vacant, filled) are given by (again, I focus on firms on the equilibrium path, i.e., $y = y(\ell)$ to focus on workers' location chocies)

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

With all the notations introduced so far, I am ready to prove the property summarized in the below lemma, which is the key for the proof of 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.,* $V^v(x,\ell) \le V_0^v(x,\ell_x)$ *for all* (x,ℓ).

Proof. I proceed by contradiction. Suppose not, i.e., there exists (x,ℓ) such that $\Delta \equiv V^u(x,\ell) - V_0^u(x,\ell_x) > 0$.

Step 1 The solution of 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 can solve for w as below

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

where the second line comes 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 considered as a constant in this case as it does not vary between problems. Therefore, $\mathrm{d}(\mathrm{d}tw) = (1-\beta)(1-\beta_4-\beta_5)\,\mathrm{d}\hat{u}$. So, I conclude that, under the 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 can move but (irrationally) stay in the same location, $V_{0a}^{\cdot}$. Because there is no commitment technology, when workers and firms bargain over wages, they believe that workers would deviate to the optimal ℓ_x.

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

Note that the surplus of this match remains the same, i.e., S_0, because the threat point of workers (when this match does not form) goes back to the environment of P0 where all workers deviate to ℓ_x. Therefore, the wage is equal to $w_0(x,\ell)$, the one under P0. Observe that the above values are characterized by the same equations as in P1, and the only difference comes from wages. For both problems, the relation between wages and the unemployed values is given by

$$\mathrm{d}V_{\cdot}^u = \underbrace{\frac{\lambda(\ell)\,\mathrm{d}t e^{-\rho\,\mathrm{d}t}}{(1-(1-\lambda(\ell)\,\mathrm{d}t)e^{-\rho\,\mathrm{d}t})(1-(1-\delta\,\mathrm{d}t)e^{-\rho\,\mathrm{d}t}) - \lambda(\ell)\,\mathrm{d}t e^{-\rho\,\mathrm{d}t}\delta\,\mathrm{d}t e^{-\rho\,\mathrm{d}t}}}_{\equiv\square}\,\mathrm{d}(\mathrm{d}tw_{\cdot}),$$

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

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

<u>Step 3</u> Note that $V^u_{0a}(x,\ell) \le V^u_0(x,\ell_x)$ as ℓ_x is the optimal location choices 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 then gives

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

which leads to the contradiction. Thus, I conclude that $V^0(x,\ell_x) \ge V(x,\ell)$ for all (x,ℓ).[44] □

Proof. I now prove Proposition A.2. The value of unemployed workers when choosing ℓ_x under P1 is equal to the 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 the values are characterized by the same set of equations. Also, the wages on the equilibrium path under P0 is given by

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

which is equal to (7) when I plug in $S_0(x(\ell),y(\ell),\ell)$ that is presented in the below, (A.4). Combining with the lemma, the solutions of P0 and that of P1 coincide. In other words, the solution obtained in the draft (under P1)—$(x(\ell),y(\ell))$ and $w(x(\ell),y(\ell),\ell)$—can also be interpreted as an equilibrium with full mobility. Similarly, one can show the symmetric result for the firm side. □

Remark. Another way to check Proposition A.2 is to directly compare the FOC of workers in P0 and P1, and show that they are identical. Plugging in $x = x(\ell)$ and $y = y(\ell)$ to $S(x,y,\ell)$, the surplus on the equilibrium path is equal to that of P1.

$$S_0(x(\ell),y(\ell),\ell) = \frac{1}{\rho+\delta+\beta\lambda(\ell)+(1-\beta)q(\ell)}(f(x(\ell),y(\ell)) - bx(\ell)) + o(1), \tag{A.4}$$

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

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

[44] 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 of Step 1, $w_0(x,\ell) \ge w(x,\ell)$. Because the continuation values of unemployment and wages are both larger in P0, it has to be the case V^u should be higher in P0 as well.

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

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

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.

Workers choose the optimal location that maximizes the below.

$$\begin{aligned} V_0^u(x,\ell) &= \mathrm{d}t(bx - \bar{h}r(\ell)) + \mathrm{d}t e^{-\rho\,\mathrm{d}t}\beta\lambda(\ell)S(x,\ell) + e^{-\rho\,\mathrm{d}t}V_0^u(x,\ell_x) \\ &= \mathrm{d}t(bx - \bar{h}r(\ell)) + \frac{\mathrm{d}t}{\tilde{\rho}}e^{-\rho\,\mathrm{d}t}\beta\lambda(\ell)(y(\ell)-b)x - \frac{\mathrm{d}t}{\tilde{\rho}}e^{-\rho\,\mathrm{d}t}[\beta\lambda(\ell)(1-\beta)q(\ell)S(\ell) + \beta\lambda(\ell)\beta\lambda(\ell_x)S(\ell_x)] \\ &\quad - \frac{\mathrm{d}t}{\tilde{\rho}}e^{-\rho\,\mathrm{d}t}\beta\lambda(\ell)\bar{h}(r(\ell) - r(\ell_x)) + e^{-\rho\,\mathrm{d}t}V_0^u(x,\ell_x). \end{aligned} \tag{A.5}$$

The above equation (A.5) is the counterpart of (3), 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 (3) which exhibits single-crossing property in $(x, A_w(y(\ell),\lambda(\ell))$, the above equation is more difficult directly understand the complementarity in worker's location choices, which is why I present P1 in the draft.

Differentiating w.r.t. ℓ (imposing $e^{-\rho\,\mathrm{d}t} = 1$), the optimal location choice of workers ℓ_x solves the below.

$$(\beta\lambda(\ell_x)(y(\ell_x)-b))'x - (\beta\lambda(\ell_x)(1-\beta)q(\ell_x)S(\ell_x))' - (\beta\lambda(\ell_x))'\beta\lambda(\ell_x)S(\ell_x) - (\tilde{\rho} + \beta\lambda(\ell_x))\bar{h}r'(\ell) = 0. \tag{A.6}$$

After some rearrangement, the FOC of P1 becomes identical to that of P0 (A.6),

$$\begin{aligned} 0 &= (\beta\lambda(\ell)(y(\ell)-b))'x - (\beta\lambda(\ell)(1-\beta)q(\ell)S(\ell))' - (\tilde{\rho}+\beta\lambda(\ell))\bar{h}r'(\ell) - \frac{\beta\lambda'(\ell)\beta\lambda(\ell)}{\tilde{\rho}+\beta\lambda(\ell)}((y(\ell)-b)x - (1-\beta)q(\ell)S(\ell)) \\ &= (\beta\lambda(\ell)(y(\ell)-b))'x - (\beta\lambda(\ell)(1-\beta)q(\ell)S(\ell))' - (\tilde{\rho}+\beta\lambda(\ell))\bar{h}r'(\ell) - \beta\lambda'(\ell)\beta\lambda(\ell)S(\ell). \end{aligned}$$

This confirms Proposition A.2 which claims that the location choices of workers are the same between P0 (free mobility) and P1 (without mobility).

A.4 Proofs of Section 3.2

Proof of Lemma 1. The planner chooses the (potentially non-pure) assignment of workers and firms—shares of workers of x assigned to locations smaller than ℓ, $\bar{m}_w(\ell|x)$, and shares of firms of y assigned to locations smaller than

ℓ, $\bar{m}_f(\ell|y)$ that solves the following problem:

$$\max_{\bar{m}_w(\ell|x),\bar{m}_f(\ell|y)} \quad \int_{\underline{\ell}}^{\bar{\ell}} \left[\mathbb{E}[x|\ell]\,\mathbb{E}[y|\ell](1-u(\ell))L(\ell) - C_r(\bar{h}L(\ell)) - C_v(V(\ell))\right] \mathrm{d}\ell \tag{A.7}$$

$$\text{s.t.} \quad \int^{\ell} \frac{L(\ell')}{M_w}\,\mathrm{d}\ell' = \int_x \bar{m}_w(\ell|x)q_w(x)\,\mathrm{d}x, \int^{\ell} \frac{V(\ell')}{M_f}\,\mathrm{d}\ell' = \int_y \bar{m}_f(\ell|y)q_f(x)\,\mathrm{d}y, u(\ell) = \frac{\delta}{\delta+\lambda(\ell)}, \quad \forall \ell$$

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. I proceed by contradiction. Suppose there exists two locations ℓ, ℓ' such that $\mathbb{E}[x|\ell] \leq \mathbb{E}[x|\ell']$ and $\mathbb{E}[y|\ell] \geq \mathbb{E}[y|\ell']$. Consider relocating workers and firms between two locations. I will simplify the notations $a(\ell), a(\ell')$ to a, a' for $a = x, y, L, V$. The planner problem, focusing on these two locations, is choosing (x, y, x', y') and (L, V, L', V') to maximize $[(1-u')x'y'L' - C_r(\bar{h}L') - C_v(V')] + [(1-u)yxL - C_r(\bar{h}L) - C_v(V)]$ such that $L + L' = \bar{L}, V + V' = \bar{V}$, $\frac{L}{L+L'}x + \frac{L'}{L+L'}x' = X$, and $\frac{V}{V+V'}y + \frac{V'}{V+V'}y' = Y$ where $\bar{L}$ ($\bar{V}$, respectively) is total measure of workers (firms, respectively), and X (Y, respectively) is an average productivity of workers (firms, respectively) in two locations.

There are four possible cases. First, consider the case such that $\mathbb{E}[x|\ell] < \mathbb{E}[x|\ell']$ and $\mathbb{E}[y|\ell] > \mathbb{E}[y|\ell']$. For $x' > x$ and $y' < y$ to be the optimal choice, the following conditions should hold:

$$(1-u')x'\frac{L'}{V'} - (1-u)x\frac{L}{V} < 0, \tag{y'}$$

$$(1-u')y' - (1-u)y > 0, \tag{x'}$$

$$-\frac{\partial u'}{\partial V'}y'x'L' - C_v'(V') = -\frac{\partial u}{\partial V}yxL - C_v'(V), \tag{V'}$$

$$-\frac{\partial u'}{\partial L'}y'x'L' + (1-u')y'x' - \bar{h}C_r'(\bar{h}L') = -\frac{\partial u}{\partial L}yxL + (1-u)yx - \bar{h}C_r'(\bar{h}L). \tag{L'}$$

Condition (y') reads to $q'x' < qx$, implying $\frac{V'}{L'} > \frac{V}{L}$. From (V') and (L'),

$$\Delta C_v'(V) = \Delta \alpha u y(1-u)x\frac{L}{V},$$

$$\Delta \bar{h}C_r'(\bar{h}L) = \Delta(-\alpha u(1-u)xy + (1-u)xy) = \Delta(1-u)xy(1-\alpha u),$$

where Δ denotes the difference between ℓ' and ℓ. Since αu is decreasing in θ, the first equation is negative while the second equation is positive. Thus, by the convexity of cost functions, $V' < V$ and $L' > L$, implying $\frac{V'}{L'} < \frac{V}{L}$. *Contradiction.*

Next, consider $\mathbb{E}[x|\ell] = \mathbb{E}[x|\ell']$ and $\mathbb{E}[y|\ell] > \mathbb{E}[y|\ell']$. From (y'), I obtain $\frac{V'}{L'} > \frac{V}{L}$ as the first case. As worker productivity is identical between two locations, the condition (x') becomes $(1-u')y' = (1-u)y$, which implies $u' < u$ and thus $\frac{V'}{L'} < \frac{V}{L}$. *Contradiction.* When $\mathbb{E}[x|\ell] < \mathbb{E}[x|\ell']$ and $\mathbb{E}[y|\ell] = \mathbb{E}[y|\ell']$, the similar argument gives the contradiction.

Lastly, when $\mathbb{E}[x|\ell] = \mathbb{E}[x|\ell']$ and $\mathbb{E}[y|\ell] = \mathbb{E}[y|\ell']$, conditions (x') and (y') give $V' = V$ and $L' = L$. Then, increasing $\mathbb{E}[x|\ell']$ and $\mathbb{E}[y|\ell']$ while holding V', L', V, L constant increases the net output due to supermodularity in the output function. *Contradiction*.

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

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

where μ_w and μ_f are co-state variables of x and y, respectively, μ_u is the multiplier of unemployment rates, and the market tightness is defined as before, $\theta = \frac{V}{uL}$. I already imposed steady state condition, $N(\ell) = V(\ell)$.

Let $\varepsilon_\lambda(\ell) = \frac{\lambda'(\theta)}{\lambda(\theta)}\theta$. First order conditions of worker allocation are given by,

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

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

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

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

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

Finally, the constraint on the unemployment rate leads to

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

From now on, I omit ℓ for notational simplicity unless it causes confusions. Differentiate (L) with respect to ℓ, and plug in (x), I obtain

$$\begin{aligned} C_r''\bar{h}^2L' &= \frac{(1-u)(1-\varepsilon_\lambda)}{1-\varepsilon_\lambda(1-u)}y'x - \frac{\varepsilon_\lambda u(1-u)}{1-\varepsilon_\lambda(1-u)}x'y - \frac{(1-\varepsilon_\lambda)xy}{(1-\varepsilon_\lambda(1-u))^2}u' - \frac{u(1-u)xy}{(1-\varepsilon_\lambda(1-u))^2}\varepsilon_\lambda', \\ &= \frac{(1-u)(1-\varepsilon_\lambda)}{1-\varepsilon_\lambda(1-u)}y'x - \frac{\varepsilon_\lambda u(1-u)}{1-\varepsilon_\lambda(1-u)}x'y + \frac{u(1-u)xy}{(1-\varepsilon_\lambda(1-u))^2}\left((1-\varepsilon_\lambda)\varepsilon_\lambda - \frac{\lambda''}{\lambda'}\theta + \varepsilon_\lambda - 1\right)\frac{\theta'}{\theta} \end{aligned} \tag{A.8}$$

where I use $1+\frac{\partial}{\partial u}\left(\frac{\alpha(1-u)u}{1-\alpha(1-u)}\right)=\frac{1-\alpha}{(1-\alpha(1-u))^2}$ for first line. For the second line, I use $\frac{u'(\ell)}{u(\ell)}=-(1-u)\varepsilon_\lambda\frac{\theta'(\ell)}{\theta(\ell)}$ and $\frac{\varepsilon_\lambda'(\ell)}{\varepsilon_\lambda(\ell)}=\left(\frac{\lambda''(\theta)}{\lambda'(\theta)}\theta-\varepsilon_\lambda+1\right)\frac{\theta'(\ell)}{\theta(\ell)}$.

Similarly, differentiate (V) with respect to ℓ, and plug in (y) gives

$$C_v''(V)V'=-\frac{(1-\varepsilon_\lambda)(1-u)}{1-\varepsilon_\lambda(1-u)}xy'\frac{L}{V}+\frac{\varepsilon_\lambda(1-u)u}{1-\varepsilon_\lambda(1-u)}\frac{L}{V}x'y-\frac{u(1-u)xy}{(1-\varepsilon_\lambda(1-u))^2}\left((1-\varepsilon_\lambda)\varepsilon_\lambda-\frac{\lambda''}{\lambda'}\theta+\varepsilon_\lambda-1\right)\frac{L}{V}\frac{\theta'}{\theta} \quad (A.9)$$

Multiply (A.8) by $L(\ell)$ and (A.9) by $V(\ell)$ respectively, and then sum them up,

$$C_r''(\bar{h}L(\ell))\bar{h}^2L'(\ell)L(\ell)+C_v''(V(\ell))V'(\ell)V(\ell)=0, \quad (A.10)$$

where I use $\frac{\theta'}{\theta}=\frac{V'}{V}-\frac{L'}{L}-\frac{u'}{u}$ by the definition of market tightness, which gives $\frac{\theta'}{\theta}=\frac{1}{1-\varepsilon_\lambda(1-u)}\left(\frac{V'}{V}-\frac{L'}{L}\right)$. This is the first equation that characterizes the optimal assignment.

Furthermore, rearrange (A.8) and then plug in (A.10), I obtain the second equation,

$$\begin{aligned}&\frac{(1-u)}{1-\varepsilon_\lambda(1-u)}\left((1-\varepsilon_\lambda)\frac{y'}{y}-\varepsilon_\lambda u\frac{x'}{x}\right)xy\\&=\left[C_r''\bar{h}^2L+\frac{\varepsilon_\lambda u(1-u)}{(1-\varepsilon_\lambda(1-u))^3}\left(\varepsilon_\lambda(1-\varepsilon_\lambda)-\frac{\lambda''\theta}{\lambda'}+\varepsilon_\lambda-1\right)xy\left(1+\frac{C_r''}{C_v''}\bar{h}^2\left(\frac{L}{V}\right)^2\right)\right]\frac{L'}{L}\end{aligned} \quad (A.11)$$

where I use $\frac{L'}{L}-\frac{V'}{V}=\frac{L'}{L}\left(1+\frac{C_r''}{C_v''}\bar{h}^2\left(\frac{L}{V}\right)^2\right)$.

In sum, the optimal allocation $\{x(\ell),y(\ell)\}$ is characterized by (A.10) and (A.11), together with boundary conditions.

Because housing and business services cost functions are convex, the condition (A.10) implies $L'(\ell)V'(\ell)\le 0$. In addition, the second condition (A.11) provides the condition under which the optimal population density increases. The planner optimally chooses to increase population density while decrease a measure of vacancy when the below inequality holds at the optimal allocation,

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

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

Proof of Proposition 2. Worker and firm sorting conditions in decentralized equilibrium are given by

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

To evalute if DE is efficient, I will compute the worker sorting condition multiplied by $L(\ell)$ and the firm sorting condition multiplied by $V(\ell)$. The expression is relatively long, and thus I will consider each component of this equation one by one. First, I gather the terms related to market tightness of the firm sorting condition (omitting $x(\ell)(y(\ell)-b)\theta'(\ell)$),

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

Next, gather the terms related to market tightness of the worker sorting condition (omitting $x(\ell)(y(\ell)-b)$),

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

Combining the above expressions with the remianing terms, the sorting conditions are given by:

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

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

$$L\bar{h}r'=\frac{\beta\lambda}{\delta+\beta\lambda}xy'L+\frac{\delta}{\delta+\beta\lambda}\frac{\beta\lambda}{\delta+\beta\lambda}\varepsilon_\lambda\frac{\theta'}{\theta}x(y-b)L, \tag{A.12}$$
$$Vc'=\frac{V(1-\beta)q}{\delta+\beta\lambda}x'(y-b)-\frac{\beta\lambda}{\delta+\beta\lambda}(1-\beta)(1-u)xy'L+\frac{\delta(1-u)(1-\beta)}{\delta+\beta\lambda}\left(\frac{\delta}{\delta+\beta\lambda}\varepsilon_\lambda-1\right)\frac{\theta'}{\theta}x(y-b)L, \tag{A.13}$$

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

If L and V move in the same direciton, i.e., $L'V'>0$, $L\bar{h}r'(\ell)+c'(\ell)$ is strictly positive. An allocation where both L and V decrease in ℓ, which makes this term negative, cannot be an equilibrium. To see this, multiply (A.12) by $(1-\beta)(1-u)$, and add (A.13),

$$(1-\beta)(1-u)\bar{h}^2C_r''LL'+VC_v''=V\frac{(1-\beta)q}{\delta+\beta\lambda}x'(y-b)+\frac{\delta}{\delta+\beta\lambda}(1-\beta)(1-u)(\varepsilon_\lambda-1)\frac{\theta'}{\theta}x(y-b)L. \tag{A.14}$$

If $L'<0$ and $V'<0$, the LHS is negative. In contrast, the RHS is positive from the worker sorting condition (A.12), which implies $\theta'<0$. Thus, I obtain contradiction.

Next, I consider the two cases where L and V move in the opposite direction. First, $\theta' > 0$ cannot arise under DE if ρ is sufficiently small. If $L' < 0$, $V' > 0$, and $\theta' > 0$, this directly contradicts the worker sorting condition. The RHS of worker sorting condition is positive under PAM if $\theta' > 0$. Second, consider the opposite case, $L' \geq 0$, $V' \leq 0$, $\theta' \leq 0$. Again, using (A.14),

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

where I use PAM $x' > 0$ and increasing market tightness $\theta' > 0$.

Therefore, I conclude that in DE,

$$\bar{h}^2 C_r'' LL' + VC_v'' > 0, \tag{A.15}$$

which violates (A.10), the planner's condition of efficient allocation. Thus, DE cannot be efficient.

Note that I do not impose any restriction on the matching elasticity, ε_λ. The Hosios condition, either standard or modified, cannot resolve the inefficiency. Under the Hosios condition, i.e., $\varepsilon_\lambda = 1-\beta$, the combined impact on the market tightness to workers and firms becomes zero. This can be confirmed from the sum of the last terms in (A.12) and (A.13)

$$\frac{\delta}{\delta+\beta\lambda}(1-u)(\varepsilon_\lambda - (1-\beta))\frac{\theta'}{\theta}x(y-b)L = 0,$$

which becomes zero under the Hosios condition. However, DE is still inefficient due to worker and firm sorting. The remaining terms becomes

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

which confirms (A.15).

The planner can correct externalities by using spatial transfers. To compute spatial transfers that can correct externalities, I will compare the worker and firm sorting condition to the counterparts of the planner. Rearrange the worker sorting condition $\rho V_\ell^u(x(\ell), \ell) = 0$,

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

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.8),

$$t_w(\ell) = t_w^0 + \int_{\underline{\ell}}^{\ell} \Bigg[(1-u)\left(\frac{1-\varepsilon_\lambda}{1-\varepsilon_\lambda(1-u)} - \frac{\beta}{u+\beta(1-u)} \right) y'x - \frac{\varepsilon_\lambda(1-u)u}{1-\varepsilon_\lambda(1-u)} x'y - \frac{u(1-u)\beta\varepsilon_\lambda}{(u+\beta(1-u))^2} x(y-b)\frac{\theta'}{\theta} + \frac{u(1-u)}{(1-\varepsilon_\lambda(1-u))^2}\left((1-\varepsilon_\lambda)\varepsilon_\lambda - \frac{\lambda''}{\lambda'}\theta + \varepsilon_\lambda - 1 \right) xy\frac{\theta'}{\theta} \Bigg] \mathrm{d}t \tag{A.16}$$

where t_w^0 is a constant that ensures the budget balance of the government. All assignment 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 tightnes—disapper. 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 differential impacts through the heterogeneity in firm productivity 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. However, 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 intution can be confirmed by observaing that the integrand of $t_w(\ell)$ in Section 3.2 equals to $-\frac{V}{L}\frac{\partial \rho \bar{V}^v}{\partial x} x'(\ell)$.

Similarly, I will assess the firm sorting condition. Rearrange the firm sorting condition $\rho \bar{V}_\ell^v(y(\ell), \ell) = 0$,

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

Similar to the above, by comparing the above to the planner's solution (A.9), I obtain the spatial transfer for firms.

$$t_f(\ell) = t_f^0 + \int_{\underline{\ell}}^{\ell} \Bigg[(1-u)\left(\frac{(1-\beta)\beta(1-u)}{u+\beta(1-u)} - \frac{1-\varepsilon_\lambda}{1-\varepsilon_\lambda(1-u)} \right) xy' + (1-u)u\left(\frac{\varepsilon_\lambda}{1-\varepsilon_\lambda(1-u)} y - \frac{(1-\beta)}{u+\beta(1-u)}(y-b) \right) x' + \frac{u(1-u)}{(1-\varepsilon_\lambda(1-u))^2}\left(\frac{\lambda''}{\lambda'}\theta - \varepsilon_\lambda + 1 - \varepsilon_\lambda(1-\varepsilon_\lambda) \right) xy\frac{\theta'}{\theta} - \frac{u(1-u)(1-\beta)}{(u+\beta(1-u))^2}(u\varepsilon_\lambda - u - \beta(1-u))x(y-b)\frac{\theta'}{\theta} \Bigg] \frac{L}{V} \mathrm{d}t \tag{A.17}$$

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

When the Hosios condition holds, $\varepsilon_\lambda = 1-\beta$, and unemployment benefit is zero, $b = 0$, all terms except the first term of the integrad cancel out, i.e., no externalities to be corrected, and the transfer simplifes to $t_f(\ell)$ in Section 3.2.

On the one hand, firms do not internalize the fact that local firms in higher ℓ locations are more productive, as are workers. On the other hand, they take into account that the higher average 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 on net firms choose higher ℓ than the optimal level with respect to externalities arising from the sorting channel.

A.5 Population Density

Let $\bar{m}(\ell|x)$ be the probability that a type x worker chooses a location smaller than ℓ. For example, under a pure assignment $x(\ell)$, $\bar{m}(\ell|x(\ell_0)) = \mathbb{1}\{\ell \geq \ell_0\}$. With this general (non-pure) assignment, I define population density $L(\ell)$ as below,

$$\int_{\underline{\ell}}^{\ell} L(t)\,\mathrm{d}t = \int_{\underline{x}}^{\overline{x}} \underbrace{\bar{m}(\ell|x)}_{x\to\ell}\underbrace{q_w(x)}_{\text{type dis.}}\,\mathrm{d}x.$$

Note that the above condition returns to $\int_{\underline{\ell}}^{\ell} L(t)\,\mathrm{d}t = \int_{\underline{x}}^{x(\ell)} q(x)\,\mathrm{d}x = Q(x(\ell))$ under PAM allocation. If $x(\ell)$ is strictly increasing and differentiable as shown in Proposition 1, I obtain $L(\ell) = M_w q_w(x(\ell))x'(\ell)$.

In a discrete economy with a finite number of worker types and locations, I can define population density in a conventional way, i.e., a measure of workers per unit of land in each location. I will start with this notion and then show that the above formula can be obtained as a limit.

Step 1 Consider an economy with a finite number of worker types and location. Consider *a finite number of worker types* $\{x_1, ...x_M\}$ with CDF Q_w. Assume that each type has a positive measure, and choose x_0 which is smaller than x_1. Then, $Q_w(x_0) = 0$.

Consider *a finite number of locations* $\{\ell_1, ..., \ell_N\}$. Without loss of generality, assume that the land distribution is uniform, i.e., the measure of land in each ℓ_n is given by a number $\mathrm{d}\ell$.[45] Consider a **weak PAM** equilibrium allocation $\bar{m}(\ell|x)$. This weak PAM allows multiple types to be in a single location ℓ, or a single type of worker x_i to be in multiple locations. Also, let $\bar{m}(\ell_n|x_0) = 1$ for all n to simplify the notation.

I will introduce two auxiliary functions, $\tilde{Q}$ and $\tilde{x}_n$. First, define a *(strictly) increasing* function $\tilde{Q}$ on $[\underline{x}, \overline{x}]$ such that (1) $\tilde{Q} = Q_w$ 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 $j(n) \equiv \max_j\{x_j|\bar{m}(\ell_n|x_j) = 1\}$, 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.

[45] As locations are ex ante homogeneous, the distribution is not identifiable, and the location index can be reformulated so that R is uniform.

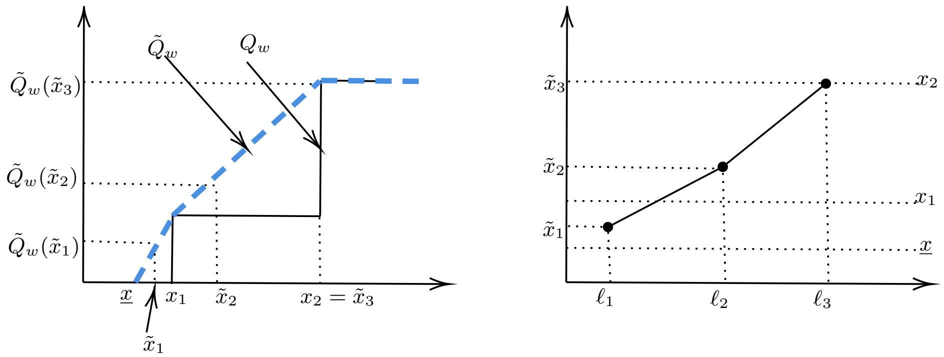


Figure A.1. Measure-Preserving Constraint

The total number of workers from ℓ_1 to ℓ_n is $\tilde{Q}_w(\tilde{x}_n)$: All worker of types $x_1, ..., x_{j(n)}$ and a fraction $\bar{m}(\ell_n|z_{j(n)+1})$ of $x_{j(n)+1}$ workers are included. Thus, $\sum_{k=1}^{n} L(\ell_k) = Q_w(x_{j(n)}) + \bar{m}(\ell_n|x_{j(n)+1})(Q_w(x_{j(n)+1}) - Q_w(x_{j(n)})) = \tilde{Q}_w(\tilde{x}_n)$. As a result, population density in location ℓ_n equals

$$L(\ell) = \frac{\tilde{Q}_w(\tilde{x}_n) - \tilde{Q}_w(\tilde{x}_{n-1})}{\ell_n - \ell_{n-1}} M_w \quad \text{for } \ell \in (\ell_{n-1}, \ell_n],$$

where M_w is the total measure of workers. The denominator comes from an assumption of uniform land distribution.

Step 2 Let the number of worker types goes to infinity, and the distribution of workers Q_w is strictly increasing. Then, $\tilde{Q}_w = Q_w$ for all $x \in [\underline{x}, \overline{x}]$, and moreover, I can characterize the cutoff types $x(\ell_n) \in [\underline{x}, \overline{x}]$ such that all workers of $x < x(\ell_n)$ choose $\ell \leq \ell_n$, and all others choose $\ell > \ell_n$. This cutoff should be strictly increasing as each type has 0 measure, and population density $L(\ell)$ is positive in equilibrium, which implies that $\tilde{x}_n = x_{j(n)} = x(\ell_n)$. 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. Note that I can also define Q_w^{-1} as Q_w is strictly increasing by assumption, and thus $x(\ell_n) = Q_w^{-1}(\ell_n)$.

Population density becomes

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

Step 3 Let the number of locations goes to infinity, and ℓ is uniformly distributed on $[\underline{\ell}, \overline{\ell}]$. As $\mathrm{d}\ell = \ell_n - \ell_{n-1}$ converges to zero, $x(\ell_n) - x(\ell_{n-1}) = Q_w^{-1}(\ell_n) - Q_w^{-1}(\ell_{n-1}) \to 0$ by the continuity of Q_w. From the above equation, two fractions become derivatives, and I obtain the following population density in the limit,

$$L(\ell) = M_w q_w(x(\ell)) x'(\ell),$$

which coincides with the definition that I use in the continuous economy.

A.6 Competitive Local Labor Markets: Proof of Proposition 3

In this section, I assume that there is no search friction and that each local labor market is competitive. I first show that workers and firms positively sort across space. I focus on the pure assignment represented by $x(\ell)$ and $y(\ell)$.

Workers of productivity x choose location ℓ that maximizes $w(x,\ell) - \bar{h}r(\ell)$. Firms of productivity y choose location ℓ that maximizes profits, $\pi(y,\ell) = x(\ell)y - w(x(\ell),\ell) - c(\ell)$. Consider two locations, ℓ and ℓ'. The location choice of firms implies that $\pi(y(\ell'),\ell') \geq \pi(y(\ell'),\ell)$ and $\pi(y(\ell),\ell) \geq \pi(y(\ell),\ell')$. Combining two inequalities, I obtain

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

Thus, $y(\ell') > y(\ell)$ implies $x(\ell') > x(\ell)$, implying PAM between workers and firms.

I will focus on the case the total measures of workers and firms are the same, i.e., $M_w = M_f$. First, the measure of workers should be equal to the measure of firms in each ℓ in equilibrium. As before, denote the measure of workers and firms in location ℓ by $L(\ell)$ and $V(\ell)$, respectively. To show by contradiction, assume that there exists location ℓ such that $L(\ell) > V(\ell)$, then, there must be another region ℓ' such that $L(\ell') < V(\ell')$. If $L(\ell) < L(\ell')$, $V(\ell')$ is larger than $V(\ell)$, and unmatched firms in ℓ' deviate to ℓ with lower costs $c(\ell)$, where they can also potentially hire workers and earn positive profits. Contrarily, when $L(\ell) > L(\ell')$, unmatched workers in ℓ deviate to ℓ' with lower rents $r(\ell')$ where they can also earn wages.

For each local labor market to be cleared, given location choices of workers and firms, it is optimal for firms of $y(\ell)$ to hire $x(\ell)$, i.e.,

$$\frac{\partial}{\partial x}(xy(\ell) - w(x,\ell)) = 0 \Rightarrow y(\ell) = \frac{\partial w(x(\ell),\ell)}{\partial x}. \tag{A.18}$$

Also, the firm sorting conditions give,

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

where I use (A.18) and the worker sorting condition, $\frac{\partial w(x(\ell),\ell)}{\partial \ell} = \bar{h}r'(\ell)$, in the second line. As $L(\ell) = V(\ell)$, in equilibrium, the signs of $r'(\ell)$ and $c'(\ell)$ should be equal, and thus $r'(\ell)$ should be zero for all ℓ. This implies that $L(\ell)$ is the same across locations, i.e., population density is uniform across locations.

An equilibrium wage of a worker of productivity x is given by

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

where I use (A.18) and the worker sorting condition for the second line, along with the change of variable $x'(\ell)\,\mathrm{d}\ell = \mathrm{d}x$. This wage formula is exactly the same as the wage function of the economy where there is only a single integrated labor market. The presence of segregated local labor markets does not affect equilibrium matching, wages, and profits. Given this observation, it is clear 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 $y(\ell)$,

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

Thus, $w(x(\ell),\ell) = w(x(\ell),\ell')$ for all ℓ,ℓ'.

A.7 Local Labor Market with Directed Search

In this section, I consider the directed search model: There exist search frictions, but search is competitive (e.g., Moen, 1997). In each local labor market, workers and firms engage in directed (competitive) search following Eeckhout and Kircher (2010). The only difference from their model is that, workers and firms first make location choices, and then engage in directed search in each local labor market. Their location choices first endogenously determine the distribution of workers and firms for each location ℓ. There are congestion costs in the housing and business services markets, which they pay when choosing locations but before participating in local labor markets.

I focus on pure assignment, $(x(\ell), y(\ell))$. The values of workers of productivity x and firms of productivity y from local labor market are given by[46]

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

where $\lambda(\theta)$ is job arrival rate, w is a wage, $q(\theta)$ is vacancy contact rate, and $f(x,y)$ is output. I use a general output function to ensure that I obtain PAM equilibrium without imposing a restriction on the matching elasticity. I will show

[46] Details of the model can be found in Eeckhout and Kircher (2010).

below the condition under which PAM arises. Alternatively, I can assume $f(x,y) = xy$ and impose a more restrictive assumption on the matching function. In location ℓ, workers of productivity x optimally choose $\theta(x,\ell)$ subject to $V^v(\ell) = q(\ell)(f(x,y(\ell)) - w(x,y(\ell),\ell))$, where $V^v(\ell) \equiv V^v(y(\ell),\ell)$.

When search is competitive, workers must ensure that firms in market ℓ enjoy their value $V^v(\ell)$ for making firms hire them. Plugging in $w(x,y(\ell),\ell) = f(x,y(\ell)) - V^v(\ell)/q(\theta(x,\ell))$,

$$V^u(x,y,\theta,\ell) = \lambda(\theta(x,\ell))f(x,y(\ell)) - \theta(x,\ell)V^v(\ell),$$

and the first order condition with respect to the market tightness is $\lambda'(\theta(x,\ell))f(x,y(\ell)) = V^v(\ell)$. Similarly, in location ℓ, firms of productivity y optimally chooses the market tightness subject to $V^u(\ell) = \lambda(\theta(y,\ell))w(x(\ell),y,\ell)$ where $U(\ell) \equiv V^u(x(\ell),\ell)$.

$$\max_{\theta} q(\theta)\left(f(x(\ell),y) - \frac{V^u(\ell)}{\lambda(\theta(y,\ell))}\right),$$

and the first order condition with respect to the market tightness equals $q'(\theta(y,\ell))f(x(\ell),y) + \frac{1}{\theta(y,\ell)^2}V^u(\ell) = 0$.

The worker (firm, respectively) value of choosing location ℓ, $\bar{V}^u(x,\ell)$ ($\bar{V}^v(y,\ell)$, respectively), is the sum of the value from the local labor market minus housing rents (overhead costs, respectively).

$$\begin{aligned}
&\bar{V}^u(x,\ell) = \lambda(\theta(x,\ell))f(x,y(\ell)) - \theta(x,\ell)V^v(\ell) - r(\ell) && \text{s.t. } \lambda'(\theta(x,\ell))f(x,y(\ell)) = V^v(\ell),\\
&\bar{V}^v(y,\ell) = q(\theta(y,\ell))\left(f(x(\ell),y) - \frac{V^u(\ell)}{\lambda(\theta(y,\ell))}\right) - c(\ell) && \text{s.t. } \theta(y,\ell)^2 q'(\theta(y,\ell))f(x(\ell),y) + V^u(\ell) = 0.
\end{aligned}$$

From the first order condition of the market tightness, I obtain $\theta(\ell)V^v(\ell) + V^u(\ell) = \lambda(\ell)f(x(\ell),y(\ell))$. Differentiating both sides with respect to ℓ,

$$\theta'(\ell)V^v(\ell) - \lambda'(\theta)\theta'(\ell)f(\ell) = \lambda(\theta(\ell))(f_x x'(\ell) + f_y y'(\ell)) - \theta(\ell)(V^v)'(\ell) - V^u(\ell). \tag{A.19}$$

The worker and firm sorting conditions are given by

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

Summing over the two,

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

where the second equality uses (A.19), and the third equality uses the worker's FOC of $\theta(x,\ell)$. The above equality implies that $L'(\ell)V'(\ell) \leq 0$, and I conclude that workers and firms do not concentrate in certain regions together.

Next, I characterize the condition under which the equilibrium exhibits PAM between workers and firms. Assuming that $x(\ell)$ is increasing without loss, PAM arises if and only if $\bar{V}^u_{x\ell} > 0$ and $\bar{V}^v_{y\ell} > 0$. Guess $y'(\ell)$. Then, $(V^v)'(\ell) = q(\ell)f_y y'(\ell)$ by the envelope theorem, and $\theta_x(x,\ell)\lambda''(\theta)f + \lambda'(\theta)f_x = 0$ by differentiating the worker's FOC of $\theta(x,\ell)$ w.r.t. x. Plugging these two,

$$\bar{V}^u_{x\ell} = \theta_x(x,\ell)(\lambda'(\theta)f_y y'(\ell) - (V^v)'(\ell)) + \lambda f_{xy} y'(\ell) = y'(\ell)\lambda\left(f_{xy} - \frac{\theta q'(\theta)\lambda'(\theta)}{\lambda''(\theta)\lambda f} f_x f_y\right),$$

which is positive if

$$\frac{f_{xy}f}{f_x f_y} \geq \sup_\theta \frac{\lambda'(\theta)\theta q'(\theta)}{\lambda''(\theta)\lambda(\theta)}. \tag{A.20}$$

Following the similar process,

$$\bar{V}^v_{y\ell} = \frac{\partial\theta(y,\ell)}{\partial y}\left(q'(\theta)f_x x'(\ell) + \frac{1}{\theta^2}(V^u)'(\ell)\right) + q(\theta)f_{xy}x'(\ell) = \left(\theta\frac{-\theta q'(\theta)}{\lambda\lambda'' f}\left(q'(\theta) + \frac{q}{\theta}\right)f_x f_y + f_{xy}\right)\lambda x'(\ell),$$

which is positive under the same condition. Under this condition, in equilibrium, $y(\ell)$ will be (weakly) increasing, and the pure assignment assumption yields a strictly increasing $y(\ell)$, which verifies the assumption.[47]

I now show that the decentralized equilibrium is efficient. I focus on the output function that leads to PAM equilibrium and the pure assignment.[48] The planner solves

$$\max \int_0^1 \lambda(\theta(\ell))L(\ell)f(x(\ell),y(\ell))\,\mathrm{d}\ell - \int_0^1 [C_r(L(\ell)) + C_v(V(\ell))]\,\mathrm{d}\ell,$$
$$\text{where } \ L(\ell) = M_w Q_w'(x(\ell))x'(\ell), V(\ell) = M_f Q_f'(y(\ell))y'(\ell), \theta(\ell) = V(\ell)/L(\ell).$$

Note that the definition is different from the baseline model as the problem is static under the assumption of Eeckhout and Kircher (2010). Set up the Hamiltonian,

$$\mathcal{H} = \lambda(\ell)L(\ell)f(x(\ell),y(\ell)) - C_r(L(\ell)) - C_v(V(\ell)) + \mu_w\frac{L(\ell)}{M_w q_w(x(\ell))} + \mu_f\frac{V(\ell)}{M_f q_f(y(\ell))},$$

[47] This condition is identical to Eeckhout and Kircher (2010). Location choice in the first stage does not interact with complementarity in the labor market because congestion costs arising from location choice affect their values additively.

[48] In contrast to the baseline model with random matching, pure assignment is no longer optimal. When the planner has a choice of non-pure assignment, the planner assigns workers and firms uniformly and matches workers and firms within each local labor market to minimize congestion costs.

which gives the following first order conditions,

$$0 = -\theta(\ell)^2 q'(\theta(\ell)) f(x(\ell), y(\ell)) - C_r'(L) + \frac{\mu_w}{M_w q_w(x(\ell))}, \tag{L}$$

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

$$0 = \lambda'(\theta) f(x(\ell), y(\ell)) - C_v'(V(\ell)) + \frac{\mu_f}{M_f q_f(y(\ell))}, \tag{V}$$

$$\mu_f'(\ell) = -\lambda(\theta(\ell)) L(\ell) f_y(x(\ell), y(\ell)) + \frac{\mu_f(\ell) V(\ell)}{M_f q_f^2(y(\ell))} q_f'(y(\ell)). \tag{y}$$

Differentiate (L) and (V) with respect to ℓ, and plug in (x) and (y) respectively,

$$C_r''(L(\ell))L'(\ell) = -\theta'(\ell)(2\theta q'(\theta) + \theta^2 q''(\theta)) f(\ell) - (\lambda(\theta) + \theta^2 q'(\theta)) f_x x'(\ell) - \theta^2 q'(\theta) f_y y'(\ell),$$
$$C_v''(V(\ell))V'(\ell) = \theta'(\ell)\lambda''(\theta) f + \lambda'(\theta) f_x x'(\ell) + \theta(\ell) q'(\theta) f_y y'(\ell).$$

To compare these conditions to those in decentralized equilibrium, differentiate the first order conditions of the market tightness of workers and firms with respect to ℓ,

$$(V^v)'(\ell) = \theta'(\ell)\lambda''(\theta) f + \lambda'(f_x x'(\ell) + f_y y'(\ell)),$$
$$-(V^u)'(\ell) = \theta'(\ell)(2\theta q' + \theta^2 q'') f + \theta^2 q'(f_x x'(\ell) + f_y y'(\ell)).$$

Plug the above to the worker and firm sorting conditions,

$$\begin{aligned} r'(\ell) &= \lambda f_y y' - \theta'(\ell)\lambda''(\theta) f(\ell) - \lambda'(\ell)(f_x x'(\ell) + f_y y'(\ell)) \\ &= -\theta'(\ell)(2\theta q'(\theta) + \theta^2 q'') f - \theta(q + \theta q') f_x x'(\ell) - \theta^2 q'(\theta) f_y y'(\ell), \\ c'(\ell) &= q(\theta(\ell)) f_x x'(\ell) - \frac{1}{\theta}(V^u)'(\ell) \\ &= (q + \theta q'(\theta)) f_x x'(\ell) + \theta q'(\theta) f_y y'(\ell) + \theta'(\ell)(2q'(\theta) + \theta q''(\theta)) f. \end{aligned}$$

These conditions on location choice are identical to those of the planner characterized above. Thus, I conclude that the decentralized equilibrium is efficient. In directed search models, workers (or firms) are constrained to provide the expected value of firms (or workers), and this condition ensures that all externalities are internalized. I confirm that the intuition remains valid even when I introduce the pre-stage where workers and firms choose locations. The following proposition summarizes the result.

Proposition A.3. *Suppose that complementarity in output is sufficiently large such that* (A.20) *holds. If a 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 not concentrated in space: $L'(\ell)V'(\ell) \leq 0$.

3 Decentralized equilibrium is efficient.

Contrary to Proposition 3, matching between workers and firms differs when labor markets are segmented across locations. Differential housing rents and overhead costs additionally shape location choices of workers and firms, and thus affect their matching. For example, if the cost functions of housing and business services are extremely convex, the equilibrium will have a smaller variation in the market tightness across locations. In contrast, if these functions are linear, the equilibrium will be identical to the one with an integrated labor market.

A.8 Alternative Mechanisms Accounting for Spatial Disparities: Proof of Proposition 4

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

$$\begin{aligned} S(x(\ell),y(\ell),\ell) &= \frac{1}{\tilde{\rho}+\beta\lambda(\ell)+(1-\tilde{\beta})q(\ell)}(A(\ell)y(\ell)-b)x(\ell), \\ w(x(\ell),y(\ell),\ell) &= \left(b+\beta(A(\ell)y(\ell)-b)+(1-\beta)\frac{\beta\lambda(\ell)-(1-\tilde{\beta})q(\ell)}{\tilde{\rho}+\beta\lambda(\ell)+(1-\tilde{\beta})q(\ell)}(A(\ell)y(\ell)-b)\right)x(\ell). \end{aligned} \tag{A.21}$$

Other expressions can be modified in a similar manner by adding local TFP $A(\ell)$ accordingly. Importantly, the values of workers and firms choosing a location ℓ are given by

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

Observe that all of the above equations return to those in the baseline wihtout heterogeneity in local TFP, i.e., $A(\ell) = 1$.

Matching Cross-Sectional Moments. Given unemployment rates across regions $\{u(\ell)\}$, I can compute local market tightness and thus $\{\lambda(\ell), q(\ell)\}$. Moreover, suppose that the two out of three $\{x(\ell), y(\ell), \bar{A}(\ell)\}$ are given.

First, I can always choose the remaining one so that (A.21) matches cross-sectional wages in the data. Assuming the population density in ℓ is equal to $\{L(\ell)\}$ in the data, the model matches spatial disparities across locations, i.e., $\{L(\ell), w(\ell)\}$. Second, I can choose production costs $C_r(\cdot)$ and $C_v(\cdot)$ to ensure that the equilibrium conditions are satisfied. If housing rents $r(\ell)$ and overhead costs $c(\ell)$ are estimated such that $\frac{\partial}{\partial\ell}V^u(x(\ell),\ell) = \frac{\partial}{\partial\ell}V^p(y(\ell),\ell) = 0$, then the sorting of workers and firms are optimal. Next, if I calibrate marginal production costs such that $C_r'(\bar{h}L(\ell)) = r(\ell)$ and $C_v'(N(\ell)) = c(\ell)$, housing and business services markets clear.

Planner problem. Focus on PAM allocation, i.e., positive sorting among workers, firms, and locations. In other words, I focus $x(\ell)$, $y(\ell)$, and $\bar{A}(\ell)$ that are increasing in ℓ. The Hamiltonian can be fomulated as below,

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

where I use the same notation as in Appendix A.4. First order conditions of the planner are given by

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

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

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

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

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

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

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

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

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

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 costs. Finally, the below equation in addition to (A.22) fully characterizes the planner's solution.

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

where $\square = \frac{u(1-u)}{(1-\varepsilon_\lambda(1-u))^3}\Big(\varepsilon_\lambda(1-\varepsilon_\lambda) - \frac{\lambda''}{\lambda'}\theta + \varepsilon_\lambda - 1\Big)xyA$. In particular, the above equation demonstrates that the planner increases population density L relatively more than that of firm V when firm heterogeneity y' is significantly larger than that of workers x', as in the baseline.

Efficiencies of the decentralized equilibrium. Derivations of the sorting conditions is explained in detail in Appendix A.4. So, I omit the steps and will give the final expressions. When ρ goes to zero, the sorting conditions are given by

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

To focus on the externalities arising from the sorting, I assume the Hosios condition $\varepsilon_\lambda = 1-\beta$ and zero unemployment rate $b=0$. Comparing the sorting condition of workers and firms and those to the planner, I obtain the spatial transfers that can correct externalities in the decentralized economy.

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

where t_w^0 and t_f^0 are constants that ensure the government budget balance. All functions are evaluated at the optimal assignment. Transfers that correct externalities arising from the sorting of workers and firms remain the same as before, which internalize their impact on the quality of local labor market. In addition, the last term in workers' transfer corrects the ignorance of workers on their impact on local TFP through spillovers.

B. Quantitative Analysis

B.1 Data

Locations. I base my analysis on metropolitan statistical areas (MSAs). Population density of the year 2010 across MSAs is obtained from the Census data. 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 Ruggles et al. (2023) for the year 2017. 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.

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 Job 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 year 2017 and 2019 using adjusted gross income (*adjginc*) and federal income taxation (*fedtaxac*) in the CPS from IPUMS, 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 project the tax rates on locations to understand 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, Storesletten and Violante, 2017). The estimates for the year 2017 are $\log(\text{disp. income rates}) = -0.072 - 0.0061 \log(\text{pop. density})$. When I estimate the same regression for the year 2019, to avoid any distortions caused by changes in the wage distribution, 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. Using this predicted after-tax income of the year 2017 under the 2019 tax rates, the regression yields $\log(\text{disp. income rates}) = -0.071 - 0.0026 \log(\text{pop. density})$.

B.2 Quantitative Model

I explain how the key equations change after several modifications that I make in Section 5.1.

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

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

The value of firms remains almost the same. The only change is that firms additionally pay entry cost c_e, so that the value becomes $\bar{V}^v(y,\ell) = \frac{1}{\rho}(\delta_v V^v(y,\ell) - c(\ell)) - c_e$.

Rubinstein bargaining and wages. I discuss the bargaining solution when workers have a Stone-Geary utility function and workers pay income taxes as assumed in Section 5. I show that the solution still has a simple solution with a slight modification.

Consider a parallel time for bargaining where a flow *surplus* of firms (workers, respectively) is $v_f = \tilde{\rho}(V^p(x,y,\ell) - V^v(y,\ell))$ ($v_w = \tilde{\rho}(V^e(x,y,\ell) - V^u(x,\ell))$, respectively). A worker with a discount factor δ_w and a firm with a discount factor δ_f are involved in a bargaining game where a firm is the first mover. I will check four assumptions to apply proposition 122.1 in Osborne and Rubinstein (1994). It is straightforward to check that the bargaining problem satisfies the first three assumptions (no redundancy, the indifference when the opponent enjoys the best agreement, monotone Pareto frontier) when workers discount their value by $r^{-\omega}(1-\tau_w)$. The final assumption is satisfied if the Pareto frontier of the set of agreements is the set of $\{v \in \mathbb{R}^2 : v_w = g(v_f)\}$ for some decreasing concave function g, and the preference of each player is represented by $\delta_i^t v_i$ for some $0 < \delta_i < 1$ for $i = w, f$. In my context, $v_f = A - w$, and $v_w = B + r(\ell)^{-\omega}(1-\tau(w))w$ where $A = xy - q(V^p - V^v) - \delta_v V^v$ and $B = -r^{-\omega}bx - \lambda(V^e - V^u)$. It is clear that $g(v_f) = B + r(\ell)^{-\omega}(1-\tau_w)(A - v^f)$ is decreasing and weakly concave. Thus, from the proposition, 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, setting $\beta = \frac{1-\delta_f}{1-\delta_f\delta_w}$ leads to

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

The outcome is equivalent to the solution of the Nash bargaining in which workers and firms split the adjusted surplus $S = r^{\omega}(1-\tau_w)^{-1}(V^e - V^u) + V^p - V^v$ with worker's bargaining power β.

Housing markets. I assume that total cost of providing housing $H(\ell)$ is $C_r(H(\ell)) = \frac{1}{(1+\gamma)H_w^{\gamma}} H(\ell)^{1+\gamma}$, where $\gamma = 1/\eta_w$ is an inverse of the housing supply elasticity η_w, i.e., a price elasticity. Because the housing market is competitive, housing rents $r(\ell)$ equal the marginal cost $(H(\ell)/H_w)^{\gamma}$, and the housing supply $H_w(r(\ell))$ equals $H_w r(\ell)^{\eta_w}$.

Saiz (2010) finds that more stringent regulations reduce housing supply elasticity η_w. He first estimates the price elasticity $\gamma(\ell)$ across MSAs and regresses estimated elasticitieson the local regulation index, WRI(ℓ),

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

I microfound the above relation by introducing taxes on housing production. Suppose the government imposes taxes on housing production. Specifically, I assume that the tax rate depends on the total housing supply, $\tau(H;\ell) = (H/T)^{t_h(\ell)} - 1$. Here, the term $t_h(\ell)$ quantifies the stringency of housing regulations.[49] The sum of production costs

[49] I include a parameter T so that tax rates $1+\tau$ become unit free by appropriately choosing T based on the unit of housing. Moreover, the term H/T being larger than 1 for all ℓ guarantees that housing rents $r(\ell)$ are increasing in tax rates $\tau(H;\ell)$.

and taxes of housing supply becomes $(1+\tau(H;\ell))C_r(H(\ell)) = \frac{1}{1+\gamma}\frac{1}{H_w^\gamma}\frac{1}{T^{t_h(\ell)}}H^{1+\gamma+t_h(\ell)}$, and housing rents become

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

First, observe that this functional form is consistent with (A.23). The elasticity of housing rents $r(\ell)$ with respect to housing supply H is $\gamma + t_h(\ell)$, and more stringent regulations (i.e., higher $t_h(\ell)$) lead to a higher price elasticity or a smaller housing elasticity, $\eta_w(\ell) = 1/(\gamma + t_h(\ell))$. I normalize tax rates so that $\min_\ell t_h(\ell)$ and $\min_\ell \tau(H;\ell)$ are zero. In other words, $t_h(\ell) = 0.28[\log(3+\text{WRI}(\ell)) - \log(3+\text{WRI}(\underline{\ell}))]$. The housing supply becomes

$$H_w(r(\ell);t_h(\ell)) = \underbrace{\left(\frac{1+\gamma}{1+\gamma+t_h(\ell)}\right)^{\frac{1}{\gamma+t_h(\ell)}}\left(H_w^\gamma T^{t_h(\ell)}\right)^{\frac{1}{\gamma+t_h(\ell)}}}_{\equiv H_w(t_h(\ell);\ell)} r(\ell)^{\frac{1}{\gamma+t_h(\ell)}}.$$

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

$$r(\ell)H_w(r(\ell),\ell) = (1-\omega)\bar{h}L(\ell) + \omega((1-\tau_w(\ell))(u(\ell)bx(\ell) + (1-u(\ell))w(x(\ell),y(\ell),\ell)) + \Pi + T_r(\ell))L(\ell), \tag{A.25}$$

where $H_w(r(\ell),\ell)$ is implicitly determined by (A.24) and $T_r(\ell)$ is the local tax revenue per capita.

Business services market. Assuming the same functional form for the cost function of business services as in housing, $C_v(S(\ell)) = S(\ell)^{1+1/\eta_f}/\left((1+1/\eta_f)H_f^{1/\eta_f}\right)$, profits of intermediaries are given by

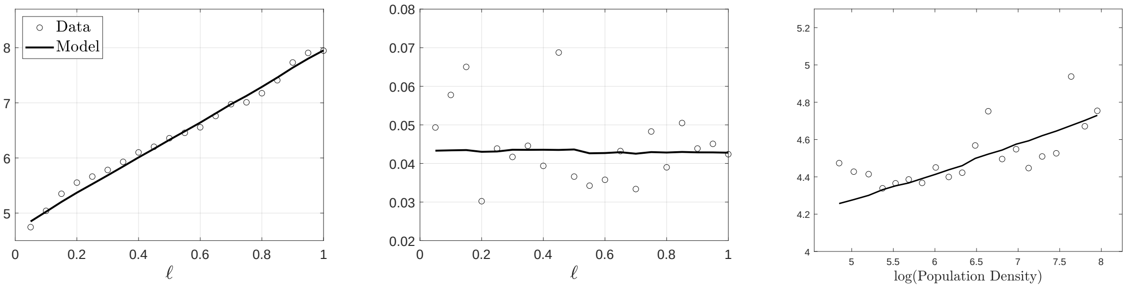
$$\pi^v(N(\ell)) = c(\ell)N(\ell) - C_v(N(\ell)) = c(\ell)N(\ell)/(1+\eta_f),$$

where I impose the market clearing assumption, $S(\ell) = N(\ell)$, for the second equality.

B.3 Model Fit

In addition to moments presented in Table 3, 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.

C. Empirical Evidence of Two-Sided Sorting



(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

C.1 Extended Model

In this section, I extend the model to justify two-way fixed effects estimation and a migration instrument that I use in Section 4. Importantly, this model predicts 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 limit case where δ_v goes to infinity as in Section 3.2. I introduce migration frictions. Workers, either employed or unemployed, can migrant only when they receive the migration shock at rate δ_m. Receiving this shock, workers first draw preference shocks $\varepsilon \overset{iid}{\sim} F_\varepsilon$ for each location ℓ, and this location-specific shock $\varepsilon(\ell)$ increases the value of workers multiplicatively of moving to ℓ. I assume that F_ε is continuous, and it has a large support so that workers can draw a sufficiently large $\varepsilon(\ell)$ with a positive probability. Workers then decide where to go based on this realization. After migrating, they start searching for a job. They are allowed to choose the current location and stay, but they lose their current jobs if they were employed. 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 location populated by productive firms and higher housing rents if the preference shocks are sufficiently large.

I also assume that there are migration costs, $m_{\ell',\ell}$. When a worker moves from ℓ' to another location ℓ, her value is discounted by a discount factor $m_{\ell',\ell}^{-1}$. 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. With these two components, the value of a worker moving from ℓ' to ℓ equals $\varepsilon(\ell) m_{\ell',\ell}^{-1}$ multiplied by the value of searching for a job in location ℓ, $V^u(x, \ell)$, which I explain below.

Location decisions. The flow value of unemployed and employed in location ℓ, excluding preference shocks $\varepsilon(\ell)$ and migration costs $m_{\ell',\ell}^{-1}$, remains almost the same, but I additionally take into account the migration option value $V^m(x, \ell)$,

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

where $V^m(x,\ell) = \mathbb{E}[\max_k\{m_{\ell,k}^{-1}\varepsilon(k)V^u(x,k)\}]$. 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, but I replace $x(\ell)$ with $\mathbb{E}[x|\ell]$ to account for the fact that the distribution of worker productivity is non-degenerate. The wage of a match is same as before with modified $\tilde{\rho}$,

$$\log w(x,y,\ell) = \log x + \log\left(b + (1-\beta)\frac{\beta\lambda(\ell)}{\tilde{\rho} + \beta\lambda(\ell)}(y(\ell) - b) + \beta(y - b)\right). \tag{A.26}$$

It is important to observe that the wage is log additive in worker productivity and the remaining part. This feature will turn out to be useful later for two-way fixed effects estimation.

With some algebra following Appendix A.1, the value of workers and firms become

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

Observe that the value of a firm is single-crossing in hiring opportunities $A_f(\mathbb{E}[x|\ell], q(\ell)) = q(\ell)\,\mathbb{E}[x|\ell]$ and the productivity of firm y. Thus, I have the following prediction.

Lemma A.2 (Prediction on firm sorting). *If a location ℓ suddenly experiences an exogenous increase in worker productivity $\mathbb{E}[x|\ell]$, while the vacancy contact rate does not change, then this location will attract more productive firms than before.*

Location decisions of workers are characterized by the probability that a worker of x chooses a location between 0 and ℓ, $M(\ell|x,\ell')$. Define the corresponding density $m(\ell|x,\ell')$, which exists under the regularity assumption on F_ε. Note that when the dispersion of preference shocks goes to zero, the location choice can be characterized by a location-matching function, $x(\ell)$, and the decision simplifies to $M(\ell|x,\ell') = 1\{x(\ell) \geq x\}$. In contrast, I do not introduce any frictions to location decisions of firms, thus the firm sorting remains pure, and thus can be characterized by a function $y(\ell)$.

Migration network. Location decisions of workers $M(\ell|x,\ell')$ and their productivity distribution $Q_w(x)$ together determine the distribution of workers across locations. Let $\mu_\ell(x)$ be the measure of workers who have productivity x and choose ℓ. Then, this measure satisfies the following the law of motions,

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

For example, if $m_{\ell',\ell} = 1$, it is clear that $m(\ell|x,\ell')$ becomes independent of ℓ' and simplifies to $m(\ell|x)$. Then, I obtain $\mu_\ell(x) = m(\ell|x)M_w q_w(x)$ due to the measure-preserving constraintm $M_w q_w(x) = \int_\ell \mu_\ell(x)\,\mathrm{d}\ell$. Population density of ℓ is given by

$$\int_0^\ell L(k)\,\mathrm{d}k = \int_0^\ell \int_x \mu_\ell(x)\,\mathrm{d}x\,\mathrm{d}k \Rightarrow L(\ell) = \int_x \mu_\ell(x)\,\mathrm{d}x.$$

Note that the above definition of population density returns to $\int^\ell L(k)\,\mathrm{d}k = \int_x 1\{x(\ell) \geq x\} q_w(x)\,\mathrm{d}x = \int^{x(\ell)} q_w(x)\,\mathrm{d}x = Q_w(x(\ell))$ if the assignment is pure.

I now examine the migration network, which is crucial information for constructing an instrument. 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 (11) will be highly correlated with the observed changes in migrant productivity.

Due to the preference shocks $\varepsilon(\ell)$, there will be positive flows between locations. This migration network can be characterized by the location choice probability 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

$$\Pr(\ell' \to \ell|\ell) = \frac{\int_x m(\ell|x,\ell')\mu_{\ell'}(x)\,\mathrm{d}x}{\int_k \int_x m(\ell,\ell'|x)\mu_k(x)\,\mathrm{d}x\,\mathrm{d}k}.$$

The intensity of migration flows between a pair of two 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.[50] Also, observe that if a larger measure of workers from origin ℓ' choose to leave their location, $m(\ell|x,\ell')$ increases, leading to an increase in $\Pr(\ell' \to \ell|\ell)$. This shows how an increase in out-migrants $O_{\ell',-\ell,t}$ leads to an increase in share of migrants from ℓ'.

The average productivity of migrants in ℓ is given by

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

and the average productivity of non-migrants is the same in steady state. Workers from ℓ' tend to be more productive, if the average productivity of workers in ℓ', $\mathbb{E}[x|\ell']$, is higher.

[50] In addition, it can also depend on the composition $\mu_\ell(x)$. For example, the fraction of workers moving to ℓ conditional on leaving ℓ', $\Pr(\ell' \to \ell|\ell')$, is higher if the composition of workers is biased toward types x with higher values in region ℓ. In the model, this is the case when $|\ell - \ell'|$ is smaller, controlling for migration costs.

C.2 Estimation of Worker and Firm Productivity

Estimating parameters. In this section, I assume that δ_v approaches infinity, in accordance with the condition needed to ensure that the population density and wages increase in ℓ as shown in Proposition 1. This assumption is imposed also in Section 5.3 and Section 5.

I recover the productivity of firms $y(\ell)$ in two steps.[51] First, I measure the local job finding rate $\lambda(\ell)$ directly from the panel and calibrate $\tilde{\rho}$, which is the sum of discount rate and separation rate. Next, I estimate two unknown parameters—worker's bargaining power β and an unemployment benefit b—by targeting two moments—the replacement rate and aggregate labor share. Given these parameters, I can compute $y(\ell)$ from

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

where $\log \hat{y}(\ell)$ is the average estimated fixed effect of newly matched jobs in location ℓ. This allows me to compute two implied corresponding moments using worker and firm productivity and estimated fixed effects. Importantly, this estimation strategy recovers productivity independently of the other parameters of the model.

The average unemployment benefit of unemployed workers equals b multiplied by average worker productivity, which is simply the average of worker fixed effects. Wage equals worker productivity multiplied by firm fixed effects, and its average can be computed from the panel data. The replacement rate is the ratio between the two, and I target 0.6 following the out-of-work dataset provided in the Social Policy Indicator (SPIN) database. Expressing replacement rate as a function of observable variables,

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

which pins down $b = 0.65$. Focusing on the newly created jobs that determine the option value, under the pure assignment of firms, the productivity can be expressed as below,

$$y(\ell) = \frac{\tilde{\rho}+\beta\lambda(\ell)}{\beta(\tilde{\rho}+\lambda(\ell))}\hat{y}(\ell) - b\frac{\tilde{\rho}(1-\beta)}{\beta(\tilde{\rho}+\lambda(\ell))}.$$

For a given value of the bargaining power β, I can compute the firm productivity and thus the labor share as below,

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

The labor share increases in β because output decreases in β for given fixed effects $\hat{y}(\ell)$. I find that $\beta = 0.068$ gives the labor share of 0.63 (FRED, 2017).

[51] Two-way fixed effects can identify only the relative productivity gap among workers and among firms. I normalize the average level of $\log \hat{x}, \log \hat{y}$ to 0. This is without loss of generality because the model is independent of the scale.

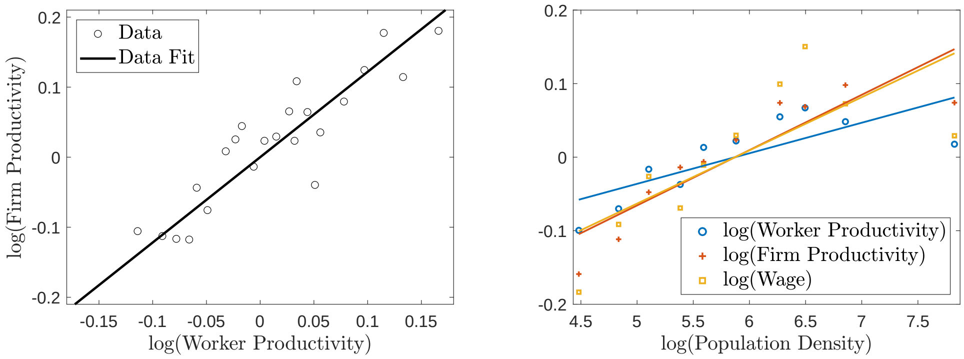


Figure A.3. Average Productivity of Workers and Firms in Space

Notes: Data source: LIAB (2003-2016). I compute the local average of the log productivity of employed workers and their firms, as well as the log wage.

Estimation results. The dispersion of local productivity of 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 discrepancy between the two arises from differences in job arrival rates, which lead to differential bargaining threat points of workers.

Descriptive evidence on spatial disparities. Using these estimated productivity, I show that the data is qualitatively consistent with the predictions made in Proposition 1. The left panel of Figure A.3 shows a binscatter plot of worker and firm productivity across regions. Worker and firm productivity are positively correlated across regions, consistent with the first property, i.e., PAM between workers and firms across locations.[52] A 10% increase in worker productivity is associated with an 12% increase firm productivity across space.

The right panel of Figure A.3 displays a binscatter of productivity and average wages against log population density. The results support the second and third properties of Proposition 1, showing that denser locations have higher average productivity and wages. The fitted lines for log productivity of workers, firms, and wages have slopes of 0.042, 0.075, and 0.072, respectively.

C.3 Additional Results

I provide two additional results that complement the findings in Section 4. First, I regress changes in estimated firm fixed effects of new jobs, $\Delta \log \hat{y}(\ell)$, as opposed to changes in recovered firm productivity. The results are presented in Table A.1, with each column following the same specification to that in Table 1. Focusing on the preferred specification in Column (5), a 1 standard deviation increase in the quality of the local workforce results in a 0.6 standard deviation

[52] The positive correlation between their productivities in Germany is also documented in Card et al. (2018) and Dauth et al. (2022), but they focus on estimated fixed effects.

Table A.1: 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 employees. The dependent variable is the change in firm fixed effects of new jobs. In the first-stage regression, I regress the change in worker productivity on its instrument and the controls I include in the second-stage.

increase in firm productivity in the same location. This finding aligns with the conclusion in Table 1, which suggests that the estimation results are robust.[53,54]

For a second robustness check, I also regress changes in estimated firm fixed effects of *existing* jobs, $\Delta \log \hat{y}^{\text{old}}(\ell)$, created in the first period.[55] If a rise in the quality of the local workforce equally increases both new and existing jobs, it indicates that the productivity growth is primarily driven by region-specific factors. I present the results in Table A.2, following the same specifications but omitting Column (5) as it is not applicable in this context. None of the coefficients are statistically significant, and all are near zero. The stark contrast between the coefficients in Table A.1 and those in Table A.2 suggests that the sorting mechanism is mainly responsible for the positive effect resulting from a rise in worker productivity. Taking a step further, the coefficients being close to zero indicate that instrumented changes in worker productivity are not correlated with region-specific shocks, thereby confirming the validity of my instrument.

[53] As I control for changes in the job finding rate, changes in firm fixed effects can be attributed to changes in firm productivity. To measure the dispersion of estimated firm fixed effects arising from firm productivity, I compute the standard deviation of residualized firm fixed effects.

[54] I obtain a similar result using relative changes in estimated firm fixed effects of new jobs compared with those of existing jobs, i.e., $\Delta \log \hat{y}(\ell) - \Delta \log \hat{y}^{\text{old}}$, as the dependent variable, instead of adding $\Delta \log \hat{y}^{\text{old}}(\ell)$ as a control. It is reassuring that the coefficient decreases only to 0.632 (s.e. 0.238).

[55] I use estimated firm fixed effects of existing jobs, instead of their recovered productivity. As shown in (A.26), the productivity of existing jobs is negatively correlated with the productivity of new jobs $y(\ell)$ for given firm fixed effects. Thus, I focus on estimated firm fixed effects which do not rely on my wage equation and the estimated parameter values. This does not pose a concern when I focus on new jobs as in Table 1.

Table A.2: 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			
$\Delta \log x^{\text{IV}}(\ell)$		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 employees. The dependent variable is the change in firm fixed effects of existing jobs. In the first-stage regression, I regress the change in worker productivity on its instrument and the controls I include in the second-stage.

D. Policy Evaluation

Decomposition of Welfare Changes. I decompose the welfare changes into three componetns: firm productivity, housing, and market tightness. The rest of welfare changes may arise, for example, from changes in income tax rates, which are not the focus of this paper. Define the term $\hat{F}(x)$ as the F that a worker of productivity x encounters in a new steady state. Also, I define $\check{F}(y)$ analogously. For example, $\hat{y}(x)$ ($\check{x}(y)$) is the productivity of firm (worker) matched with a worker (firm) of productivity x (y) in a new steady state.

First, I compute the welfare changes attributable to firm type. I find a mapping $\hat{y}(x)$ which directly enters in the values of workers. In addition, firm type indirectly affects workers' values through a lump-sum redistribution Π, which comes from tax revenues, firm portfolio, and profits of landowners and intermediaries. I compute a new lump-sum redistribution $\hat{\Pi}^y$ by making two changes. I use the new tax revenue using $\hat{y}(x)$, and new firms' profit using $\check{x}(y)$. By plugging in $\hat{y}(x(\ell))$ and $\hat{\Pi}^y$ to (12) instead of $y(\ell)$ and Π, I compute the values for workers of productivity $x(\ell)$, while fixing all others. The difference between these values and those from the baseline is the welfare change due to firm type.

Second, I consider the impact of hosuing market by utilizing $\hat{r}(x)$ and $\hat{T}_r(x)$. In addition, I also compute a new lump-sum redistribution $\hat{\Pi}^r$. I consider changes in firm's profit due to changes in overhead costs, $\check{c}(y)$. I also use the new profits of landowners and intermediaries.

Lastly, for tightness, I use $\hat{\lambda}(x)$ and $\hat{q}(x)$ which directly affects (12). Additionally, I compute a new $\hat{\Pi}^\lambda$. To be specific, I compute the new tax revenues, which are obtained under $\hat{\lambda}(x)$, $\hat{q}(x)$, and $\hat{u}(x)$, and new firms' profit using $\check{\ell}(y)$ and $\check{q}(y)$.

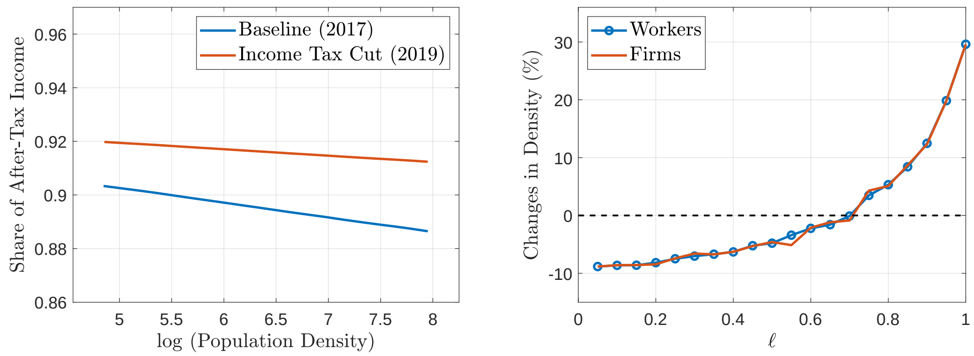


Figure A.4. Federal Income Tax Across Regions

Notes: Data source: CPS March (2017, 2019). The ratio between after-tax income and adjusted gross income is approximated as a function of log population density.

One-Sided Sorting Model. To calibrate the one-sided sorting model, I choose the same schedules of $\{x(\ell), \lambda(\ell), q(\ell), L(\ell)\}$ as in the two-sided sorting model. In addition, I set $\bar{A}(\ell)$ equal to $y(\ell)$ from the two-sided sorting model, which ensures that the both models yield the same cross-sectional wages. I also maintain the housing rents $r(\ell)$, which clear the housing market in each location and satisfy the sorting conditions of workers. However, I choose different overhead costs $c^{\text{one}}(\ell)$ as below to ensure that the sorting of *homogeneous* firm is optimal,

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

where $\bar{V}^p$ is a constant independent of ℓ. From $\{c^{\text{one}}(\ell)\}$ and $\{N(\ell)\}$, I can calibrate the margianl cost of providing business services using $c^{\text{one}}(\ell) = C_v'(N(\ell))$. I set $C_v(N) = C_v^0 + \int_0^N C_v'(N)\,\mathrm{d}N$ where C_v^0 is chosen so that the total profit of intermediaries of the one-sided sorting model is equal to that of the two-sided sorting model. I additionally adjust c_e^{one} so that the average profit of firms becomes zero as in the two-sided sorting model. These adjustments ensure that a lump-sum income of workers, Π, remains the same.

For the welfare decomposition, I adjust the definition of location component. Instead of $\hat{y}(x)$, I consider changes in local TFP $\hat{A}(x)$, a local TPF that a worker of productivity x encounters in a new steady state. Moreover, I consider changes in a lump-sum redistribution Π from firm portfolio. Specifically, I first order homogeneous firms with respect to local TFP $\bar{A}(\ell)$ because I can no longer index them with their productivity. Next, for the profit of firms of q-th quantile, I plug in local TFP of q-th quantile in a new steady state, i.e., $\check{A}(q)$.

Federal Income Tax. I examine the impact the federal income tax cuts which lower the overall tax rate and makes tax schedule less progressive. This exercise is motivated by changes in workers' income tax after the Tax Cuts and Job Acts enacted in December 2017, which includes changes in tax brackets, mostly lower tax rates, and an increase in standard deductions. Specifically, I consider the tax refrom in the left panel of Figure A.4, which plots the shares of after-tax

Figure A.5. Welfare Impact of Tax Cut (% Change)

income of adjusted gross income before and after the reform. Workers on average pay about 11% of their income as taxes, and those in dense cities pay higher rates due to tax progressivity. After the tax cuts, the approximated after-tax income share is higher and exhibits less regional variation.

The main message remains the same as the housing deregulation exercise in Section 6. In my model, the tax cuts induce significant relocation of workers and firms toward large cities. Both population density and the measure of vacancies in the top 10% regions increase by 25%. Despite the large-scale spatial reallocation, change in aggregate output is marginal, increasing only by 0.004%.

Although there is no change in output, aggregate welfare of the utilitarian social planner *decreases* by 0.228% because reducing the progressivity of the federal income tax contradicts the optimal policy. To quantify the contribution of each component, I perform decomposition exercise as in Section 6.[56] A decline in welfare primarily comes from the housing market. As spatial concentration in the economy intensifies, aggregate housing costs increase. In contrast, the contributions of other mechanisms including firm type and market tightness are limited.

I find that the welfare effect is opposite when I apply the same tax reform to the *one-sided sorting* model. After the tax reform, workers and firms move toward high$-\ell$ locations, similar to the two-sided sorting. Assuming exogenous local TFP instead of firm sorting, however, results in the opposite welfare implication. As workers and firms migrate toward productive locations, aggregate output increase by 0.388%, which leads to a 0.239% rise in aggregate welfare. Figure A.5 confirms that the primary source of this welfare gain comes is the benefits derived from location productivity, which outweights the negative impact from the housing market.

[56] There is an additional impact resulting from changes in tax rates. However, the aggregate impact of this component is negligible, as it merely involves the redistribution among workers.