Two-Sided Sorting of Workers and Firms: Implications for Spatial Inequality and Welfare*

Guangbin Hong†

Michigan State University

March 15, 2026

Abstract

High-skilled workers and high-productivity firms co-locate in large cities. In this paper, I study how the two-sided sorting of workers and firms affects spatial earnings inequality, efficiency of the allocation of workers and firms across cities, and the welfare consequences of place-based policies. I build a general equilibrium model in which heterogeneous workers and firms sort across cities and match within cities. Using Canadian matched employer–employee data, I estimate the model and find that the urban earnings premium is almost entirely explained by worker and firm sorting. The laissez-faire equilibrium is inefficient as low-productivity firms overvalue locating in high-skilled cities. The optimal spatial policy would incentivize high-skilled workers and high-productivity firms to co-locate to an even greater extent while redistributing income towards low-earning cities. Model counterfactuals underscore the importance of two-sided sorting when evaluating distributional and aggregate outcomes of place-based policies.

Keywords: Two-sided sorting, spatial inequality, optimal spatial policy, place-based policies.

JEL codes: R12, R13, E25.

*I am deeply indebted to Nate Baum-Snow, Joseph Steinberg, Kevin Lim, Serdar Ozkan, and Kory Kroft for their continuous guidance and support. I have benefited significantly from discussions with Rodrigo Adão, Treb Allen, Costas Arkolakis, Patrick Bayer, David Berger, Adrien Bilal, Mons Chan, Arnaud Costinot, Sebastian Dyrda, Maximiliano Dvorkin, Jonanthan Dingel, Min Fang, Cecile Gaubert, Thibaut Lamadon, Rasmus Lentz, Enrico Moretti, Peter Morrow, Dávid Nagy, Ezra Oberfield, Fernando Parro, Charly Porcher, Diego Restuccia, Esteban Rossi-Hansberg, Hannah Rubinton, Ana Maria Santacreu, Martin Schneider, Will Strange, Felix Tintelnot, Matthew Turner, Xincheng Qiu, and Yue Yu, and from many seminar and conference presentations. I thank Kalan Couch for his excellent research assistance. The empirical analysis of this paper is conducted in the Statistics Canada Research Data Center. I thank Canadian Research Data Centre Network for its generous financial support. The results and views expressed in this paper are those of the author and do not represent Statistics Canada. All remaining errors are on my own.

†Michigan State University. Address: Old Botany 204B, 478 W Circle Dr, East Lansing, MI 48824. Email: ghongecon@gmail.com.

1 Introduction

Spatial earnings inequality is accompanied by the sorting of high-skilled workers and high-productivity firms into larger cities. Local governments around the world spend billions of dollars annually on place-based policies aimed at attracting the best workers and firms.[1] What are the roles of worker and firm sorting in shaping spatial inequality? Is the spatial allocation of workers and firms efficient? Can place-based policies improve welfare by reallocating workers and firms? Answering these questions requires a model with two-sided spatial sorting of heterogeneous workers and firms. However, the literature has studied worker sorting and firm sorting in isolation.

In this paper, I develop a two-sided sorting general equilibrium model and estimate it using Canadian matched employer–employee data. I find that worker sorting and firm sorting together explain nearly the entire urban earnings premium, accounting for 66.8% and 26.8% of it, respectively. I uncover novel firm-sorting externalities that emerge from the interaction of two-sided sorting and local matching: low-productivity firms overvalue locating in large cities, where they inefficiently compete with high-productivity firms for high-skilled workers. The optimal spatial policy promotes even greater co-location of high-skilled workers and high-productivity firms, while redistributing income toward lower-earning cities. I show that accounting for two-sided sorting amplifies both the aggregate and distributional impacts of place-based policies. Overall, the analysis highlights the crucial role of two-sided sorting in explaining spatial inequality and in guiding the design of effective place-based policies.

The model is motivated by well-documented empirical patterns on the spatial sorting of heterogeneous workers and firms across cities (Diamond, 2016; Combes, Duranton, Gobillon, Puga, and Roux, 2012), variation in the degrees of assortative matching within cities (Dauth, Findeisen, Moretti, and Suedekum, 2022), and wage inequality both between and within cities (Baum-Snow and Pavan, 2013). In the model, heterogeneous workers and firms choose among cities that differ in exogenous productivity and skill-specific amenities. City productivity increases with population as a result of agglomeration spillovers. Within each city, firms post skill-specific wages, and workers choose firms based on wages, non-wage amenities, and idiosyncratic preferences, the latter of which are unobserved by firms as in Lamadon, Mogstad, and Setzler (2022) (hereafter, LMS). Workers and firms demand floor space—for residential and production purposes, respectively—which is supplied by a local housing developer. In equilibrium, local labor and housing market interactions, together

[1] For instance, many U.S. states compete to attract high-skilled workers by offering lower tax rates (Moretti and Wilson, 2017). Similarly, many local governments offer subsidies to attract productive firms (Greenstone, Hornbeck, and Moretti, 2010). For reviews on place-based policies, see Glaeser and Gottlieb (2008) and Duranton and Venables (2018) among others.

with city fundamentals, determine heterogeneous workers' utilities and firms' profits. Consequently, the spatial sorting of workers and firms is interdependent, with their location decisions jointly shaping the spatial wage structure.

The model features production complementarities between worker skill and firm productivity. Following Bonhomme, Lamadon, and Manresa (2019) (hereafter BLM), I incorporate firm-specific skill-augmenting productivity. These complementarities encourage the co-location of high-skilled workers and high-productivity firms, as positive assortative matching yields higher returns. In addition, workers' idiosyncratic preferences generate horizontal differentiation across firms, resulting in two key implications. First, it gives rise to monopsony power, allowing firms to set wages along upward-sloping labor supply curves, as in Card, Cardoso, Heining, and Kline (2018) and LMS.[2] Second, because workers view different employers as imperfect substitutes, they have a love-of-variety preference for more firms in a city, aligning with the insights of Helsley and Strange (1990).

I identify two novel firm sorting externalities arising from the interaction between imperfect labor market competition and two-sided spatial sorting: labor market stealing and love-of-variety externalities. Specifically, the wedge between the social and private value of a firm's location choice comprises both a local and a national component. The local component includes the familiar stealing and love-of-variety externalities found in entry models under imperfect competition (Mankiw and Whinston, 1986). The national component—novel to the model—captures productivity changes associated with the spatial reallocation of workers caused by each firm's location decision. Welfare can thus be improved through the spatial reallocation of firms, even when local externalities net to zero, as in Dixit and Stiglitz (1977). With production complementarities, it is welfare-improving to increase the degree of positive assortative matching between workers and firms in the economy. Consequently, the planner would reallocate high-productivity firms to larger, higher-skilled cities and low-productivity firms to smaller, lower-skilled ones.

I estimate the structural model using Canadian matched employer-employee data, which provide annual earnings information for all Canadian workers across firms and cities. The main challenge of model estimation is to separately identify city, firm, and worker productivity parameters. To address this, I combine a worker movers design with a revealed preference approach based on firms' location choices. First, I follow BLM to identify worker skills and firm skill-augmenting productivities using earnings changes when workers switch firms. Then, I identify city productivity from firm skill-neutral productivity from firms' revealed location

[2]Imperfect labor market competition is supported by substantial within-city wage dispersion among workers with the same observed and unobserved characteristics (Baum-Snow and Pavan, 2013; Dauth et al., 2022).

decisions. Intuitively, a city is inferred to have a productivity advantage if it attracts more firms despite high floor space rents and a tight labor market.

Using the estimated model, I find that between-city worker and firm differences account for 66.8% and 26.8% of the urban earnings premium, respectively.[3] I estimate strong production complementarities between worker skill and firm productivity, which play a key role in shaping sorting patterns and the spatial earnings structure. Counterfactual simulations show that removing these complementarities reduces the covariance between city-average worker skill and firm productivity by 44.2%, and lowers the urban earnings premium by 40.6%. I also find, consistent with Diamond (2016), that larger cities offer greater and skill-biased amenities, which substantially contribute to skill sorting.

I show that the optimal spatial policy would further strengthen the co-location of high-skilled workers and high-productivity firms. By harnessing production complementarities to a greater extent, this reallocation would lead to a 9.6% increase in total output. To mitigate the negative labor market stealing effect, the optimal policy reduces the number of low-productivity firms in larger, higher-skilled cities while reallocating high-productivity firms towards such places. The optimal policy also includes spatial transfers to low-earning cities to equitably redistribute the efficiency gains. Overall, the optimal policy yields a consumption-equivalent Pareto welfare gain of 5.0% for all agents in the economy.

Finally, I show that two-sided sorting significantly amplifies the aggregate and distributional effects of place-based subsidies. I simulate a 5% wage subsidy for firms in the top 5% of the productivity distribution that choose to locate in Toronto, emulating the city's bid to attract Amazon's HQ2. In the full model, the subsidy increases total output by 0.8% but significantly exacerbates spatial earnings inequality. While high-skilled workers in Toronto benefit, low-skilled local workers face welfare losses due to rising housing costs and the outflow of low-productivity firms. In contrast, models with only one-sided sorting generate much smaller aggregate gains and limited distributional impacts.

This paper is closely related to three strands of literature. The first is the literature on spatial sorting, which includes both worker sorting (Baum-Snow and Pavan, 2013; Behrens et al., 2014; Eeckhout et al., 2014; Diamond, 2016; Davis and Dingel, 2019) and firm sorting (Gaubert, 2018; Ziv, 2019; Lindenlaub et al., 2022; Bilal, 2023), and examines their respective roles in explaining the spatial wage disparities. In a slightly different context, a notable exception is Hoelzlein (2023), who studies two-sided sorting of workers and firms across neighborhoods and the implications for neighborhood-level policies. I contribute to the literature by developing a two-sided sorting model and analyzing how the location choice

[3]This decomposition holds the spatial distribution of workers and firms fixed, assigning mean city earnings to different productivity components.

interaction between workers and firms shapes the spatial earnings distribution and sorting efficiency.

The second is the literature on optimal spatial policies (Kline and Moretti, 2014; Fajgelbaum and Gaubert, 2020; Rossi-Hansberg, Sarte, and Schwartzman, 2019), which has focused on correcting externalities caused by local productivity and amenity spillovers. I find that the interaction between spatial sorting and imperfect competition in local labor markets generates additional inefficiencies. The labor market stealing effect in my framework with firm monopsony power resembles the pooling externality in Bilal (2023), who studies firm sorting in a setting of search frictions and homogeneous workers. Building on this insight, I further show that when incorporating worker heterogeneity, the extent of firm sorting inefficiency depends critically on the match returns between heterogeneous workers and firms.

Lastly, I contribute to the literature on earnings inequality by applying the two-way fixed effects framework developed by Abowd, Kramarz, and Margolis (1999) (hereafter AKM). Dauth et al. (2022) and Card et al. (2025) use this approach to compare average worker and firm fixed effects across cities while Carry, Kleinman, and Nimier-David (2025) further exploit establishment mobility to isolate the location effect.[4] These studies consistently highlight the importance of the spatial sorting of workers and firms in explaining the spatial earnings disparities. I contribute to the literature by separately identifying the spatial heterogeneity of workers, firms, and cities in a manner consistent with a general equilibrium model.

The rest of the paper is organized as follows. In Section 2, I formulate the two-sided sorting spatial equilibrium model. Section 3 examines spatial sorting inefficiencies and proposes the optimal policy design. Section 4 describes the data and the model estimation strategy. Section 5 presents the quantitative analysis. The paper concludes in Section 6.

2 Model

In this section, I present a two-sided sorting spatial equilibrium model that extends the Rosen-Roback framework (Rosen, 1974; Roback, 1982) by incorporating rich heterogeneity among workers and firms.[5] The model characterizes the Nash equilibrium of location choices, where each agent's decision is the best response to the choices of all others in equilibrium.

[4] In Canada, firm movements often involve continuance procedures or changes in legal structure, under which the business registry assigns new firm identifiers. Hence, firm movers are not used for identification.

[5] Hong (2024) documents four empirical facts using Canadian matched employer-employee data that support the model: (1) larger cities pay higher average earnings and exhibit greater within-city earnings inequality, (2) larger cities have both higher-earning workers and higher-paying firms, (3) high-earnings workers and high-paying firms spatially follow each other over time, and (4) the degree of positive assortative matching is greater in larger cities.

2.1 Environment

The economy is populated by heterogeneous workers and firms. Workers, indexed by i, differ in skill $a_i \in (\underline{a}, \bar{a})$. Firms, indexed by j, differ in production technology, characterized by (z_j, θ_j), and in skill-specific amenities $\{G_j(a)\}_a$. The measure of workers at each skill level, $\{L(a)\}_a$, and the set of firms, $\mathcal{J}$, are exogenously given.

There are C cities in the economy. Cities, each indexed by c, differ in exogenous productivity A_c, skill-specific amenities $\{R_c(a)\}_a$, the housing supply elasticity γ_c, and land available for housing production $\bar{H}_c$. City productivity increases with population L_c due to agglomeration spillovers, with the agglomeration elasticity denoted by μ.[6] Let $\mathcal{L}_c$ and $\mathcal{J}_c$ denote the sets of workers and firms located in city c, respectively.

All firms produce a homogeneous, tradable final good that serves as the numeraire of the economy, and the model abstracts from industry heterogeneity. Table I.1 in the Online Appendix shows that the estimated urban earnings premium remains largely unchanged after controlling for city-level industry composition. Similarly, Card et al. (2025) find that location-based wage premia across U.S. cities vary little across industries. Together, these findings indicate that the cross-sectional urban earnings premium is driven primarily by spatial variation within industries.[7]

2.2 Worker's problem

2.2.1 Preferences

The utility of a worker i with skill a, in city c, working for firm $j \in \mathcal{J}_c$ and earning wage $W_{jc}(a)$ is given by:

$$U_i(j, c; a) = \log\left(\frac{R_c(a)}{r_c^{\eta}}\right) + \log \tau W_{jc}(a) + \log G_{jc}(a) + \beta_w^{-1}\epsilon_{ij}, \tag{2.1}$$

where r_c is the housing rent in city c, η is the expenditure share on housing, and $\tau > 1$ represents the rebate of land rents. I assume that each worker owns a share of land in the economy, with the share proportional to their labor earnings. As will become clear, this assumption ensures that the rebate does not affect workers' choices over firms and cities. Finally, ϵ_{ij} is worker i's idiosyncratic preference for firm j, which is drawn from a Type-I

[6] In Section F.1 of the Online Appendix, I extend the model to allow city amenities to respond to the local skill composition (Diamond, 2016). This endogenous amenities channel can amplify worker sorting.

[7] City-level industry composition has been shown to affect urban wage growth (Eckert, Ganapati, and Walsh, 2022) and local response to industrial-specific shocks (Autor, Dorn, and Hanson, 2013).

Extreme Value distribution:

$$F(\vec{\epsilon_i}) = \exp\left[-\sum_c \left(\sum_{j\in\mathcal{J}_c} exp\left(-\frac{\epsilon_{ij}}{\rho_w}\right)\right)^{\rho_w}\right], \tag{2.2}$$

where $\rho_w \in [0,1]$ governs the correlation of idiosyncratic preferences across firms within each city: $\rho_w = \sqrt{1-\text{corr}(\epsilon_{ij},\epsilon_{ij'})}$ for $j, j' \in \mathcal{J}_c$. A smaller ρ_w corresponds to more correlated within-city preferences and reflects stronger attachment to a particular city. The parameter β_w controls the dispersion of idiosyncratic preferences. A higher β_w results in less dispersion and thus weaker influence of idiosyncratic tastes in affecting workers' firm and city choices.

This utility specification extends LMS by incorporating preferences for local characteristics, and extends Diamond and Gaubert (2022) by incorporating preferences over heterogeneous firms. Within each city, equation (2.1) allows for both vertical and horizontal differentiation across firms. Vertical differentiation arises from firm-specific wages and amenities, with the latter introduced to ensure the model can capture within-city labor market matching patterns—central to both location decisions of workers and firms and the spatial wage structure. Horizontal differentiation arises from workers' idiosyncratic preferences, which give rise to a love-of-variety preference for local firms, as in Helsley and Strange (1990).

2.2.2 Labor supply

Workers observe each city's amenity $\{R_c(a)\}_a$, housing rents r_c, the set of firms $\mathcal{J}_c$, and their skill-specific wages $\mathbf{W}_c(a) = \{W_{jc}(a)\}_{a,j\in\mathcal{J}_c}$. With such information, each worker decides which city to locate in and which firm in that city to work for. Given the property of the Type-I Extreme Value distribution, the measure of skill a workers choosing firm j can be solved as:

$$S_{jc}(a; W_{jc}(a)) = L_c(a) \cdot \frac{(W_{jc}(a)G_{jc}(a))^{\frac{\beta_w}{\rho_w}}}{\mathbb{W}_c(a)}, \ \forall a, j \in \mathcal{J}_c, c, \tag{2.3}$$

where $L_c(a)$ is the measure of skill-a workers in city c and $\mathbb{W}_c(a) \equiv \sum_{j\in\mathcal{J}_c}(W_{jc}(a)G_{jc}(a))^{\frac{\beta_w}{\rho_w}}$ is the skill-specific wage index in city c. Equation (2.3) characterizes the firm-level labor supply curves by skill, with the supply elasticity given by β_w/ρ_w. Analogously, the measure of skill-a workers choosing city c can be solved as:

$$L_c(a) = L(a) \cdot \frac{U_c(a)^{\beta_w}}{\sum_{c'\in C} U_{c'}(a)^{\beta_w}}, \ \forall a, c, \tag{2.4}$$

where $U_c(a) \equiv R_c(a) \cdot r_c^{-\eta} \cdot \mathbb{W}_c(a)^{\rho_w/\beta_w}$. Workers trade off the wage index and city amenities against housing rents when choosing a location, with city-level labor supply elasticity given

by β_w. Within $U_c(a)$, the exponent on the wage index, ρ_w/β_w, is the love-of-variety elasticity and is equal to the inverse of the firm-level labor supply elasticity. Love-of-variety is stronger when the idiosyncratic preferences are more dispersed (low β_w) or less correlated (high ρ_w).

2.3 Firm's problem

2.3.1 Production technology

A firm j in city c produces the final good by combining a set of workers, $\mathbf{D}_{jc} = \{D_{jc}(a)\}_a$, and floor space, h_{jc}, using the Cobb-Douglas production technology:

$$Y_{jc}\left(\mathbf{D}_{jc}, h_{jc}\right) = \left[\int_{\underline{a}}^{\bar{a}} A_c L_c^{\mu} \cdot z_j \cdot a^{\theta_j} \cdot D_{jc}(a)\ da\right]^{1-\alpha} \cdot {h_{jc}}^{\alpha}, \tag{2.5}$$

where α is the output elasticity of floor space in the production function.

I assume that workers with different skills are perfect substitutes within a firm.[8] Labor productivity combines city productivity $A_c L_c^{\mu}$, firm common productivity z_j, and worker-firm interaction a_j^{θ}. Heuristically, z represents a firm's absolute advantage for all workers, and θ represents a firm's comparative advantage in utilizing higher-skilled workers. If θ is monotonically increasing in z, then worker skill a and firm productivity z are complements to each other. This specification is developed by BLM as an extension of the log-additive assumption in AKM.

2.3.2 Wage setting and input choices

Firms observe each city's productivity A_c, housing rent r_c, and the sets of firms $\mathcal{J}_c$ and workers $\mathcal{L}_c$. Each firm makes a two-stage decision. First, a firm j calculates the maximized profits in each city c by choosing skill-specific wage offers $\mathbf{W}_{jc} = \{W_{jc}(a)\}_a$, skill-specific labor inputs $\mathbf{D}_{jc} = \{D_{jc}(a)\}_a$, and the housing input h_{jc}. Second, it chooses a city based on these maximized profits. The first-stage profit maximization problem is:

$$\pi_c(j) = \max_{\mathbf{W}_{jc}, \mathbf{D}_{jc}, h_{jc}} \left\{ \left[\int_{\underline{a}}^{\bar{a}} A_c L_c^{\mu} z_j a^{\theta_j} \cdot D_{jc}(a)\ da\right]^{1-\alpha} {h_{jc}}^{\alpha} - r_c h_{jc} - \int_{\underline{a}}^{\bar{a}} W_{jc}(a) D_{jc}(a)\ da \right\}, \tag{2.6}$$

[8] Eeckhout et al. (2014) find extreme-skill complementarities at the city level based on the thicker tails of the skill distribution in large cities. My model incorporates firm heterogeneity, within-city matching, and skill-specific city amenities to provide a more micro-founded explanation for the spatial distribution of skills and wages.

which is subject to the labor supply curves:

$$D_{jc}(a) = S_{jc}\left(a; W_{jc}(a)\right) = L_c(a) \cdot \frac{\left(W_{jc}(a) G_{jc}(a)\right)^{\frac{\beta_w}{\rho_w}}}{\mathbb{W}_c(a)}, \ \forall a, j \in \mathcal{J}_c, c.$$

Following Card et al. (2018) and LMS, I assume that firms are infinitesimal in the local labor market.

Assumption 1. *All firms are infinitesimal in a city. Given a set of a firm in the city, $\mathcal{J}_c$, any firm's wage decision $W_{jc}(a)$ does not affect the city-level wage index $\mathbb{W}_c(a)$, that is*

$$\frac{\partial \mathbb{W}_c(a)}{\partial W_{jc}(a)} = 0, \ \forall j \in \mathcal{J}_c, c, a.$$

With this assumption, the optimal wage offers can be obtained as a constant wage markdown multiplied by the marginal product of labor:

$$W_{jc}(a) = \underbrace{\frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}}_{\text{Wage markdown}} \cdot \underbrace{(1-\alpha)\alpha^{\frac{\alpha}{1-\alpha}} \cdot A_c L_c^{\mu} r_c^{\frac{\alpha}{\alpha-1}} \cdot z_j \cdot a^{\theta_j}}_{\text{Marginal Product of Labor (MPL)}}. \tag{2.7}$$

The wage markdown $\frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}$—which is the worker's take-home share of their marginal revenue product—increases with the firm-level labor supply elasticity β_w/ρ_w. When workers perceive firms as closer substitutes, labor supply becomes more elastic, reducing firms' monopsony power. Local housing rent r_c also affects the marginal product of labor through the firm's housing input choice. Given optimal wage-setting and input choices, firm j's profits in city c can be expressed as:

$$\pi_c(j) = \Psi \cdot A_c L_c^{\mu} r_c^{\frac{a}{\alpha-1}} \cdot (z_j)^{1+\beta_w/\rho_w} \cdot \phi_{jc}, \tag{2.8}$$

where $\Psi \equiv \frac{1}{1+\beta_w/\rho_w} \cdot (1-\alpha)\alpha^{\frac{\alpha}{1-\alpha}}$ and

$$\phi_{jc} = \int_a \left(a^{\theta_j}\right)^{1+\beta_w/\rho_w} L_c(a) \cdot \frac{G_{jc}(a)^{\frac{\beta_w}{\rho_w}}}{\sum_{j' \in \mathcal{J}_c} \left(z_{j'} a^{\theta_{j'}} \cdot G_{j'c}(a)\right)^{\frac{\beta_w}{\rho_w}}} da. \tag{2.9}$$

Firm j's maximized profits depend on city c's adjusted productivity $A_c L_c^{\mu} r_c{}^{\frac{\alpha}{\alpha-1}}$, its common productivity z_j, and a labor composite term ϕ_{jc}. The composite term relates to the total efficiency units of labor that firm j can employ in city c, combining local labor market conditions and each firm's characteristics. Notably, variations in firms' skill-augmenting productivity θ_j and amenities $\{G_j(a)\}_a$ shape the profitability of different firms in the same

city through ϕ_{jc}, ultimately driving firm sorting.

2.3.3 Firm spatial sorting

Each firm j is owned by an entrepreneur, who spends profits only on the final good. The entrepreneur's preference is:

$$u_c(j) = \ln \pi_c(j) + {\beta_f}^{-1} \nu_{jc}. \tag{2.10}$$

Each entrepreneur draws a city-specific idiosyncratic preference ν_{jc} from a Type-I Extreme Value distribution, with its dispersion scaled by β_f. Given the expected profits in each city specified in equation (2.8), the probability that entrepreneur j chooses city c is:

$$p_c(j) = \frac{\left(\left(A_c L_c^{\mu} \bar{r}_c^{\frac{\alpha}{\alpha-1}}\right) \cdot \phi_{jc}\right)^{\beta_f}}{\sum_{c'} \left(\left(A_{c'} L_{c'}^{\mu} \bar{r}_{c'}^{\frac{\alpha}{\alpha-1}}\right) \cdot \phi_{jc'}\right)^{\beta_f}}. \tag{2.11}$$

Equation (2.11) reveals that firms are more likely to locate in cities with higher productivity (adjusted by local housing rent) and better access to workers based on each firm's own characteristics. Importantly, a firm's own common productivity z does not affect its location choice because it proportionally scales the profits it can earn in all cities. Moreover, according to equation (2.9), ϕ_{jc} is independent of city productivity A_c because it does not affect within-city matching of workers and firms. As I will show in Section 4, these features are important for separately identifying city and firm common productivity parameters.

2.4 Housing supply

Housing in city c is produced by an absentee housing developer, who uses the final good Y and land $\bar{H}_c$ to produce floor space. Solving the developer's problem yields the housing supply curve for each city:

$$H_c^S(r_c) = \bar{H}_c^0 \cdot {r_c}^{\frac{1}{\gamma_c}}, \tag{2.12}$$

where $\bar{H}_c^0$ is an exogenous housing supply shifter and $1/\gamma_c$ is the housing supply elasticity. I assume that the land rents are aggregated into a national portfolio and rebated to workers through τ in equation (2.1).

2.5 Spatial equilibrium

I now formally define the spatial equilibrium.

Definition 1. *Given city characteristics* $\{A_c, \{R_c(a)\}_a, \bar{H}_c, \gamma_c\}_{\forall c}$*, worker measures* $\{L(a)\}_{\forall a}$*, the set of firm* $\mathcal{J}$ *with their production technologies* $\{z_j, \theta_j\}_{\forall j}$ *and skill-specific amenities* $\{G_j(a)\}_{\forall a,j}$*, the spatial equilibrium is defined as a set of allocations including the set of workers and firms in each city,* $\{\mathcal{L}_c, \mathcal{J}_c\}_{\forall c}$*, firm employment* $\{D_{jc}(a)\}_{\forall a,j,c}$*, housing input* $\{h_{jc}\}_{\forall j,c}$*, and a set of prices including wage* $\{W_{jc}(a)\}_{\forall a,j,c}$*, and housing rents* $\{r_c\}_{\forall c}$ *that satisfy the following conditions:*

1. *Workers choose cities and firms by equations (2.3) and (2.4);*
2. *Firms choose cities optimally by equation (2.11), and choose wage offers and housing inputs by equations (2.7) and (B.2);*
3. *The housing market clears in each city such that housing demand equals supply;*
4. *The final good market clears;*
5. *Land rents are redistributed to workers through* τ*.*

One potential concern is the possibility of multiple equilibria, as workers and firms mutually respond to each other's location decisions—particularly in the presence of production complementarities. I prove in Appendix D that there exists a unique equilibrium when idiosyncratic preferences, which serve as congestion forces as in Redding and Rossi-Hansberg (2017), are sufficiently dispersed. I further test for multiplicity by solving the estimated model from different initial guesses, all of which converge to the same equilibrium.

3 Efficiency of Two-Sided Sorting

This section analyzes the efficiency of two-sided sorting. I begin by formulating the social planner's problem in Section 3.1. I compare the social and private values associated with location choices to highlight spatial sorting externalities in Section 3.2, which then inform the design of the optimal spatial policy in Section 3.3.

3.1 Social planner's problem

Without loss of generality, suppose there are N^w worker types, indexed by $a = 1, ..., N^w$, and N^f firm types (owned by entrepreneurs), indexed by $k = 1, ..., N^f$. Let $\varphi^w(a)$ and $\varphi^f(k)$ be the Pareto weights assigned to worker type a and firm type k, respectively. The social

welfare function is specified as:

$$\mathcal{W} = \sum_{a=1}^{N^w} \varphi^w(a) L(a) U(a) + \sum_{k=1}^{N^f} \varphi^f(k) J(k) \Pi(k), \tag{3.1}$$

where $U(a)$ and $\Pi(k)$ denote the expected utility of type-a workers and type-k entrepreneurs, respectively. With a slight abuse of notation, I use a to denote both the worker type and skill level.

The planner chooses the spatial allocation of workers $\{L_c(a)\}$ and firms $\{J_c(k)\}$, the matching of workers to firms within cities $\{D_{kc}(a)\}$,[9] the consumption allocations for workers and entrepreneurs $\{c_{kc}(a)\}$ and $\{c_c(k)\}$, the housing allocations for workers and firms $\{h_{kc}(a)\}$ and $\{h_c(k)\}$, and the final good allocations for floor space production $\{Y_c\}$. These decisions are subject to constraints on spatial mobility, local matching, and resource feasibility. I assume that the social planner does not observe workers' and entrepreneurs' idiosyncratic preferences. This extends the planning problems in Fajgelbaum and Gaubert (2020) and Rossi-Hansberg et al. (2019) by incorporating heterogeneous firms and worker–firm matching. Further details are available in Section C in the Online Appendix.

3.2 Sorting externalities

From the planner's problem, I derive the social values of location choices and compare them to the private ones in the following proposition.

Proposition 1. *The optimality conditions of the social planner's problem imply:*

1. *The social marginal product of a type-a worker employed by a type-k firm in city c is:*

$$\tilde{W}_{kc}(a) = \frac{1 + \beta_w/\rho_w}{\beta_w/\rho_w} \cdot \Big[\underbrace{W_{kc}(a)}_{\text{wage}} + \underbrace{\mu \bar{W}_c}_{\substack{\text{agglomeration} \\ \text{spillovers}}} \Big], \tag{3.2}$$

where $W_{kc}(a)$ is the competitive equilibrium wage defined in (2.7), and $\bar{W}_c$ is the average wage in city c.

[9] Each type-k firm j in city c is allocated $D_{jc}(a) = D_{kc}(a)/J_c(k)$ workers of type a.

2. The social marginal value of a type-k firm in city c is:

$$\tilde{\pi}_c(k) = \underbrace{\sum_a W^*_{kc}(a)\frac{D_{kc}(a)}{J_c(k)}}_{\textit{firm surplus}} - \underbrace{(1-\rho_w)\sum_a \bar{W}^*_c(a)\frac{D_{kc}(a)}{J_c(k)}}_{\textit{local labor stealing}} - \underbrace{\rho_w \sum_a \bar{W}^*(a)\frac{D_{kc}(a)}{J_c(k)}}_{\textit{national labor stealing}} + \underbrace{\frac{\rho_w}{\beta_w}\sum_a \varphi^w(a)\frac{D_{kc}(a)}{J_c(k)}}_{\textit{love-of-variety preference}}, \tag{3.3}$$

where $W^*_{kc}(a) = \tilde{W}_{kc}(a) - W_{kc}(a)$ is the shadow value of a type-a worker at a type-k firm in city c, and $\bar{W}^*_c(a)$ and $\bar{W}^*(a)$ are the skill-specific means at the city and national levels, respectively.

Proof. See Section C in the Online Appendix. □

The differences between social and private values reveal the sources of inefficiency in the laissez-faire equilibrium. Workers are paid according to their marginal product, though their social value also includes the agglomeration spillovers they generate locally. As a result, workers fail to internalize the positive externalities associated with their location choices. Such spatial sorting externalities are well studied in the literature, including Fajgelbaum and Gaubert (2020) and Rossi-Hansberg et al. (2019), among others.

Firms base their location choices solely on the profits they earn in a city, as captured by the firm surplus component in equation (3.3). However, they do not internalize two types of externalities. The first is the labor market stealing externality, which includes both local labor stealing—as firms attract workers from other firms within the same city—and national labor stealing—as firms draw workers from other cities. The second is the love-of-variety externality, which reflects workers' preference for a greater number of horizontally differentiated firms in the local labor market.

These firm sorting externalities arise from imperfect competition in the labor market. In the model, workers' idiosyncratic preferences are unobserved by firms. The presence of such information asymmetry gives rise to a market failure. As firms cannot condition wages based on workers' idiosyncratic preferences, they compete for workers through the common wage indices and set constant wage markdowns. Hence, an additional firm entering a city has two effects through its impact on local wage indices $\mathbb{W}_c(a)$: it congests the local labor market, making it harder for other firms to hire workers (equation (2.3)), while it also expands local employment opportunities, thereby improving local workers' welfare and attracting new workers from other cities (equation (2.4)). This mechanism is analogous to the labor market pooling externality in Bilal (2023), wherein posted wages cannot price the

impact of firm sorting on local vacancy-filling rates.

Equation (3.3) is the local labor market analog of the inefficient firm entry result in Mankiw and Whinston (1986), who show that firm entrants do not consider business stealing from other firms and consumers' love-of-variety preference in the presence of imperfect competition in the output market. In the spatial context, labor market segmentation and worker mobility imply that the labor market stealing externality manifests both locally and nationally. To make this clear, I rewrite equation (3.3) for a firm j as (assuming $\mu = 0$):

$$\begin{aligned}\tilde{\pi}_c(j) - \pi_c(j) = &\underbrace{\int_a \left(\frac{\rho_w}{\beta_w}\varphi^w(a) - \frac{1}{1+\beta_w/\rho_w}\bar{\tilde{W}}_c(a)\right) D_{jc}(a)da}_{\text{Local wedge: Dixit-Stiglitz}} \\ &+ \underbrace{\frac{\rho_w}{1+\beta_w/\rho_w}\int_a \left[\bar{\tilde{W}}_c(a) - \bar{\tilde{W}}(a)\right] D_{jc}(a)da}_{\text{National wedge: worker mobility}},\end{aligned} \tag{3.4}$$

where $\bar{\tilde{W}}_c(a)$ and $\bar{\tilde{W}}(a)$ denote the city-level and national-level mean social values of type-a workers, respectively. The local wedge mirrors the classic trade-off in Dixit and Stiglitz (1977), while the national wedge captures an additional inefficiency driven by worker mobility. Because firms influence workers' location choices, firm j's decision to locate in city c affects the productivity of workers who follow the firm. The measure of affected workers is given by ρ_w times firm j's employment $\mathbf{D}_{jc}$, where ρ_w is the ratio of city- to firm-level labor supply elasticities.[10] From the planner's perspective, the national stealing externality is the gap between city- and national-average social values of these workers. Guided by the national wedge, the planner should reallocate firms to cities where the workers who follow them earn above the national average wages. This first implies that the planner would allocate more firms to cities with higher city productivity. Moreover, due to production complementarities, high-skilled workers are more likely to follow high-productivity firms. As a result, assigning high-productivity firms to cities with higher skill premia—typically larger cities—enhances spatial sorting efficiency.

The lemma shows how firm sorting inefficiency arises from the interaction between monopsony power, two-sided sorting, and spatial inequality.

Lemma 1. *The following hold when $\mu = 0$:*

1. *When $\beta_w \to \infty$, firm sorting is efficient.*
2. *With β_w/ρ_w fixed at a finite value, when $\rho_w \to 0$, firm sorting is efficient under city-skill specific Pareto weights $\varphi_c^w(a) = \bar{W}_c(a),\ \forall a, c$.*

[10] Fixed rent-sharing ensures that workers indirectly internalize their decisions on firm profits.

3. *With β_w/ρ_w fixed at a finite value, when $\rho_w \to 1$, firm sorting is efficient under Pareto weights $\varphi^w(a) = \bar{W}(a),\ \forall a$.*

4. *When there is no spatial inequality, i.e. $\bar{W}_c(a) = \bar{W}(a),\ \forall a, c$, firm sorting is efficient under Pareto weights $\varphi^w(a) = \bar{W}(a),\ \forall a$.*

The first case shows that firm sorting is efficient under perfectly competitive labor markets. The next two cases highlight the role of two-sided sorting interdependence. When $\rho_w \to 0$, labor stealing is purely local; when $\rho_w \to 1$, labor markets are fully integrated across cities. In both scenarios, workers do not spatially follow firms, and firm sorting can be efficient. The fourth case demonstrates that spatial inequality is also a necessary condition for inefficiency: without within-skill spatial earnings differences, the national labor stealing externality disappears, as relocating workers does not alter their productivity.[11] In all three latter cases, the two wedges in equation (3.4) collapse into a single term, which vanishes under appropriate Pareto weights. Any policies to spatially reallocate firms would then be solely justified by the redistributive motive.

3.3 Efficiency conditions and the optimal policy

I now characterize the efficiency conditions of the spatial economy. Following Fajgelbaum and Gaubert (2020), I derive necessary conditions for the expenditure distributions $\{x_{kc}(a)\}$ and $\{x_c(k)\}$ in an efficient allocation, where $x_{kc}(a) \equiv c_{kc}(a) + r_c h_{kc}(a)$ denotes the total expenditure of a type-a worker employed by a type-k firm in city c and $x_c(k) \equiv c_c(k)$ denotes the total expenditure of a type-k firm in city c. Comparing the competitive equilibrium distributions with those implied by the planner's solution yields the following proposition.

Proposition 2. *The following hold in an efficient allocation:*

$$\underbrace{x_{kc}(a)}_{\substack{\text{private}\\ \text{consumption cost}}} = \frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}\Big[\underbrace{\tilde{W}_{kc}(a)}_{\substack{\text{social marginal}\\ \text{product of labor}}} - \underbrace{\mathcal{O}_c^w(a)}_{\text{opportunity cost}}\Big] + \frac{1}{1+\beta_w/\rho_w}\underbrace{\varphi^w(a)}_{\text{Pareto weight}} \tag{3.5}$$

for $D_{kc}(a) > 0\ \forall a, k, c$:

$$\underbrace{x_c(k)}_{\substack{\text{private}\\ \text{consumption cost}}} = \frac{\beta_f}{1+\beta_f}\Big[\underbrace{\tilde{\pi}_c(k)}_{\substack{\text{social marginal}\\ \text{value of firm}}} - \underbrace{\mathcal{O}^f(k)}_{\text{opportunity cost}}\Big] + \frac{1}{1+\beta_f}\underbrace{\varphi^f(k)}_{\text{Pareto weight}} \tag{3.6}$$

[11] This theoretical prediction also highlights the importance of identifying the source of spatial inequality: if it is driven solely by between-city skill sorting and $\bar{W}_c(a) = \bar{W}(a),\ \forall a, c$, then it does not warrant policy intervention to spatially reallocate firms on efficiency grounds.

for $J_c(k) > 0\ \forall k, c$. The opportunity costs, $\mathcal{O}_c^w(a)$ and $\mathcal{O}^f(k)$, are related to the shadow values of each worker and firm type in the planner's problem. With Pareto weights $\{\varphi^w(a)\}_a$ and $\{\varphi^f(k)\}_k$, if the planner's problem is globally concave and (3.5) and (3.6) hold, then the associated competitive equilibrium is efficient.

Proof. See Section C in the Online Appendix. □

Proposition 2 describes the relationship between private consumption and the spatial allocation of workers and entrepreneurs that must hold in any efficient allocation. Private consumption levels in the efficient allocation differ from their counterparts in the laissez-faire equilibrium in several ways. First, the efficient consumption allocations in (3.5) and (3.6) are based on the social values. Second, the efficient consumption allocations account for the social opportunity costs associated with allocating workers and firms.[12] Third, efficient consumption allocations increase less than one-for-one with the social marginal value for both workers and firms. This reflects the planner's redistributive motive. Within the same type, the planner seeks to reallocate resources toward lower-income agents, who have a higher marginal utility of consumption. Finally, efficient consumption allocations also incorporate between-type distribution, leading to higher consumption for agents with greater welfare weights.

Equipped with the efficiency conditions, I now design the optimal policy to implement the optimal allocation.

Proposition 3. *The optimal policy comprises a set of proportional taxes, including labor income taxes, $\{t_{kc}^w(a)\}_{\forall a,k,c}$, and corporate income taxes, $\{t_c^f(k)\}_{\forall k,c}$, which are specified as:*

$$t_{kc}^w(a) = -\frac{\frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}\tilde{W}_{kc}(a) + T_c^w(a) - W_{kc}(a)}{W_{kc}(a)}, \quad \forall a, k, c \tag{3.7}$$

$$t_c^f(k) = -\frac{\frac{\beta_f}{1+\beta_f}\tilde{\pi}_c(k) + T^f(k) - \pi_c(k)}{\pi_c(k)}, \quad \forall k, c \tag{3.8}$$

where $\tilde{W}_{kc}(a), \tilde{\pi}_c(k)$ are defined in Proposition 2, $T_c^w(a) \equiv -\frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}\mathcal{O}_c^w(a) + \frac{1}{1+\beta_w/\rho_w}\varphi^w(a)$, and $T^f(k) \equiv -\frac{\beta_f}{1+\beta_f}\mathcal{O}^f(k) + \frac{1}{1+\beta_f}\varphi^f(k)$.

Proof. See Section C in the Online Appendix. □

[12] See Section C.2 in the Online Appendix for the expressions of the opportunity cost terms, which can be interpreted as the "price of congestion". From the planner's perspective, allocating more workers or firms to a city implies worse idiosyncratic preferences for the marginal worker or firm. Within each type, workers' opportunity costs, $\mathcal{O}_c^w$, vary across cities, while firms' opportunity costs, $\mathcal{O}^f$, do not. Spatial variation in workers' opportunity costs arises from correlated idiosyncratic preferences within cities.

These optimal taxes combine Pigouvian corrections to address inefficiencies and redistributive transfers. I assume that firms treat worker income taxes as exogenous when setting wages, and that land rents are rebated to workers in proportion to their after-tax income. These assumptions ensure that labor supply at the firm level remains iso-elastic, thereby leaving firms' wage-setting decisions unaffected.

4 Empirical Implementation

I now present the estimation strategy. I describe the data in Section 4.1 and introduce additional assumptions for identification in Section 4.2. I then discuss the estimation strategy in Section 4.3 and present the estimation results in Section 4.4.

4.1 Data

The main dataset for the empirical analysis is the Canadian Employer–Employee Dynamics Database (CEEDD), which is a set of linkable administrative tax files maintained by Statistics Canada. The CEEDD covers the universe of tax-paying workers and firms. I obtain employer–employee linkages and annual earnings from the Statement of Remuneration Paid (T4); worker location, age, and gender from the T1 Personal Master File (T1–PMF); and firm location from the National Accounts Longitudinal Microdata File (NALMF). I use data from 2010 to 2017. All monetary values are expressed in 2002 Canadian dollars.

The baseline sample consists of workers aged 25 to 60 residing in cities, defined as Census Metropolitan Areas (CMAs) or Census Agglomerations (CAs) in the 2016 Census—analogous to U.S. commuting zones. I keep cities with at least 15,000 full-time workers and exclude one small outlier city with average earnings exceeding 150% of the national average, yielding a sample of 66 cities. For workers with multiple employers in a year, I retain only the highest-paying job. I drop observations with annual earnings below the equivalent of 20 hours per week for 13 weeks at the minimum wage. I exclude firms in industries including agriculture, mining, utilities, education, healthcare, non-profits, and public administration, as these firms' location decisions may follow different considerations.[13]

In the CEEDD, a firm is defined as a tax and accounting entity with an enterprise identifier in the Business Registry. For multi-location firms, only the headquarters' location is observed. I leverage workers' residential locations to form enterprise–city units within multi-location firms, assuming each unit operates with its own production technology and independently chooses its location and wage policies. I refer to these units as firms hereafter.

[13]These excluded industries account for 28% of total employment in Canada.

I construct two additional samples for estimation purposes: the stayers sample and the movers sample. The selection procedure follows LMS. The stayers sample includes workers who are associated with the same firm for at least 7 years. In addition, I restrict the stayers sample to firms with at least 10 worker stayers. The movers sample includes workers who switch between firms in any year. Following Kline et al. (2020), I restrict the movers sample to firms with at least two movers. See more details on the data in Section A of the Online Appendix.

4.2 Additional assumptions

4.2.1 Discretization

Following BLM, I assume that firm technology parameters (z_j, θ_j) and amenity parameters $\mathbf{G}_j = \{G_j(a)\}_a$ are drawn from a discrete distribution. Firms sharing the same $\{z, \theta, \mathbf{G}\}$ belong to the same cluster, indexed by $k \in \{1, 2, \ldots, K\}$. The clustering procedure helps mitigate limited mobility bias, a common concern with the mover design (Andrews et al., 2008).

4.2.2 Stochastic processes

I incorporate time-varying shocks into the model to exploit the panel structure of the matched employer–employee data for identification. I first introduce workers' idiosyncratic preference shocks and skill shocks.

Assumption 2. *Workers' idiosyncratic preference shocks are drawn from a Type-I Extreme Value Distribution with the cumulative distribution function:*

$$F(\vec{\epsilon}_{it}) = exp\left[-\sum_{c}\left(\sum_{j\in\mathcal{J}_{ct}} exp\left(-\frac{\epsilon_{ijt}}{\rho_w}\right)\right)^{\rho_w}\right]. \tag{4.1}$$

This assumption follows recent structural labor literature on worker sorting (e.g., Card et al. (2018), LMS). Changes in worker preferences can induce mobility across firms and cities. Note that equation (4.1) imposes no restrictions on the time-series properties of ϵ.

Assumption 3. *A worker's skill at time t, denoted a_{it}, consists of a permanent component a_i and a transient shock $\hat{a}_{it}$, where $\hat{a}_{it}$ follows a stationary, mean-zero stochastic process independent of a_i. The transient shock does not interact with a firm's skill-augmenting productivity, that is*

$$\theta_j \log a_{it} = \theta_j \log a_i + \log \hat{a}_{it}, \tag{4.2}$$

and it does not affect workers' preferences over firm and city amenities, that is, $G_j(a_{it}) = G_j(a_i)$, $R_c(a_{it}) = R_c(a_i)$.

Following LMS, I assume that transient skill shocks neither interact with skill-augmenting productivity nor affect preferences for amenities. These restrictions imply that the transient skill shocks generate earnings changes without affecting mobility decisions. Next, I introduce the firm- and city-level productivity shocks.

Assumption 4. *The productivity of firm* j *and of city* c *at time* t, $\{z_{jt}, A_{ct}\}$, *both contain a permanent part* $\{z_j, A_c\}$ *and a transient shock* $\{\hat{z}_{jt}, \hat{A}_{ct}\}$*:*

$$\log z_{jt} = \log z_j + \log \hat{z}_{jt} \tag{4.3}$$

$$\log A_{ct} = \log A_c + \log \hat{A}_{ct} \tag{4.4}$$

where $\log \hat{z}_{jt}$ *and* $\log \hat{A}_{ct}$ *follow first-order Markov processes with innovations that are i.i.d. across firms, cities, and time. Firm and city amenities are time-invariant.*

Assumption 5. *The stochastic process for the transient worker skill shock* $\hat{a}_{it}$*, firm productivity shock* $\hat{z}_{jt}$*, and city productivity shock* $\hat{A}_{ct}$ *are independent of each other.*

The dynamic productivity processes follow the well-known works on production function estimation in the industrial organization literature (e.g. Olley and Pakes (1996)). Productivity shocks create labor demand variation needed to identify the labor supply elasticities. Finally, I introduce the measurement errors in the observed firm wage bills.

Assumption 6. *The observed wage bill* $\dot{E}_{jt}$ *is measured with error relative to its model counterpart* E_{jt}*:*

$$\log E_{jt} = \log \dot{E}_{jt} + e_{jt}, \tag{4.5}$$

where the measurement error e_{jt} *follows a moving average (MA) process of order* q*:* $e_{jt} = \sum_{s=0}^{q} \delta_s u^e_{j(t-s)}$*, with* u^e_{jt} *i.i.d. across firms and time.*

4.3 Identification Strategy

In this section, I describe the identification strategy for the model parameters. These include the sorting elasticity parameters β_w, ρ_w, β_f; the productivity parameters for every worker a, firms $\{z_k, \theta_k\}_k$ by cluster, and cities $\{A_c\}_c$, along with the agglomeration elasticity μ; the amenity parameters for firms $\{G_k(a)\}_{k,a}$ and cities $\{R_c(a)\}_{c,a}$; and the housing supply parameters $(\{\gamma_c, \bar{H}^0_c\}_c)$. More details on estimation strategy can be found in Section E of the Online Appendix.

4.3.1 Sorting elasticity parameters

Worker sorting elasticity parameters I identify the elasticity of worker sorting parameters using the passthrough design of Kline et al. (2019) and LMS. This approach exploits variation in wage bill shocks and recovers the elasticity parameters from earnings changes among worker stayers. Let $\log \bar{E}_{ct}$ and $\log \bar{W}_{ct}$ denote the city c's log average firm wage bill and average log worker earnings in year t, respectively, and $\log \hat{\dot{E}}_{jt} \equiv \log \dot{E}_{jt} - \log \bar{\dot{E}}_{ct}$ and $\log \hat{W}_{ijt} \equiv \log W_{ijt} - \log \bar{W}_{ct}$ denote the log demeaned wage bill and earnings, respectively.[14]

The firm-level passthrough parameter $\delta_w \equiv \frac{1}{1+\beta_w/\rho_w}$ is identified from the following regression using the stayers sample:

$$\Delta \log \hat{W}_{ijt} = \delta_w \Delta \log \hat{\dot{E}}_{jt} + \Delta \hat{a}_{it} + \delta_w \left(\Delta e_{jt} - \Delta \log \frac{\hat{\phi}_{jct}}{\hat{\phi}_{ct}} \right). \tag{4.6}$$

The extent of passthrough from a firm-level shock $\Delta \log \hat{\dot{E}}_{jt}$ to the earnings of worker stayers $\Delta \log \hat{W}_{ijt}$ is governed by the firm-level labor supply elasticity β_w/ρ_w. When a firm experiences a positive productivity shock that raises labor demand, a more elastic labor supply implies that the firm can expand employment with smaller wage increases—resulting in less rent sharing with incumbent workers.

There are three residual terms in the net passthrough equation (4.6): the change in the *i.i.d.* skill shock $\Delta \hat{a}_{it}$, the change in the firm wage bill measurement error Δe_{jt}, and the change in the relative supply shifter $\Delta \log(\hat{\phi}_{jct}/\hat{\phi}_{ct})$. The last term captures changes in the composition of firms of city c in year t, which affect the firm-specific labor supply curve (2.3). The potential correlation between the wage bill shock and either the measurement error or the relative supply shifter raises endogeneity concerns.

To address the first endogeneity concern, I follow LMS and instrument the net wage bill shock $\Delta \log \hat{\dot{E}}_{jt}$ with its lags prior to year $t - q - 1$. These lagged shocks are correlated with the current shock due to the persistence of firm-level productivity but uncorrelated with the contemporaneous measurement error, which is assumed to follow an $MA(q)$ process. To address the second concern, I adopt a control function approach. Assuming that $\Delta \log(\hat{\phi}_{jct}/\hat{\phi}_{ct})$ follows a first-order Markov process with i.i.d. innovations—same as the dynamic process for city productivity A_c—I nonparametrically control for $\Delta \log \hat{W}_{ijt-1}$ and $\Delta \log \hat{\dot{E}}_{jt-1}$ to account for $\Delta \log(\hat{\phi}_{jct-1}/\hat{\phi}_{ct-1})$.

[14] To control for life-cycle earnings patterns and macroeconomic shocks, I estimate a Mincer-type regression of log earnings on a cubic age polynomial and year dummies. Following Card et al. (2013), I assume a flat earnings profile at age 40. The residuals from this regression are retrieved as $\log W_{ijt}$.

The city-level passthrough parameter $\delta_c \equiv \frac{1}{1+\beta_w}$ is identified from the city-level regression:

$$\Delta \log \left(\frac{\bar{W}_{ct}}{r_{ct}^{\eta}} \right) = \delta_c \Delta \log \left(\frac{\bar{\dot{E}}_{ct}}{r_{ct}^{\eta}} \right) + \delta_c \left(\Delta e_{ct} - \Delta \log \hat{\phi}_{ct} \right), \tag{4.7}$$

where e_{ct} denotes the measurement error in the city-level mean firm wage bill assumed to follow a $MA(q)$ process as e_{jt}. Importantly, δ_c needs to be estimated using changes in real wage bills and earnings that adjust for housing rents. Because city-level productivity shocks are capitalized into local housing costs, estimating the equation (4.7) using nominal terms would result in an upward bias in the city-level labor supply elasticity.

I apply the same strategy as in the firm-level passthrough estimation to address endogeneity concerns arising from Δe_{ct} and $\Delta \log \hat{\phi}_{ct}$. An additional concern is that city-level wage bill shocks may be correlated with changes in amenities. To address this, I include a specification that nonparametrically controls for changes in the share of high-skilled workers, which serves as a proxy for amenity shifts driven by the endogenous amenity channel. High-skilled workers are defined as those in the top three deciles of the worker fixed effect distribution, estimated using a reduced-form two-way fixed effect AKM regression.

Firm sorting elasticity parameter The firm sorting elasticity parameter β_f is identified from the passthrough of city-level total wage bill shocks to the mean wage bill of incumbent firms within the city:

$$\Delta \log \bar{E}_{ct} = \delta_f \Delta \log \dot{E}_{ct} + \delta_f \left(\Delta e_{ct} - \Delta \log \frac{\hat{\phi}_{ct}}{\Phi_{ct}} \right), \tag{4.8}$$

where $\delta_f \equiv \frac{1}{1+\beta_f}$. This regression leverages city-level productivity shocks to estimate the responsiveness of firm location decisions. It is estimated using a sample of firms that have operated in a city for at least seven years and employ at least ten workers in each year. The identification assumption is that the share of firms in each city adjusts annually to city productivity shocks in accordance with equation (2.11). I follow the same strategy as before to address these endogeneity concerns.

4.3.2 Productivity and amenity parameters

I now turn to the identification strategy for the productivity and amenity parameters. I follow BLM and leverage worker movers to identify worker skill and firm skill-augmenting productivity. I then separately identify city and firm common productivity using firms' revealed location decisions. Note that relying solely on worker movers does not permit the

separate identification of firm and city productivity, even with between-city moves. I do not use firm movers for identification because firm movements are often associated with changes in the identifier. I estimate the amenity parameters by matching the observed sorting patterns of workers across cities and firms conditional on (real) earnings.

To mitigate limited mobility bias, I follow BLM and classify firms based on moments of their empirical earnings distributions. However, heterogeneous city productivity complicates this classification: because both firm and city productivity influence earnings, firms in different cities can have similar earnings distributions despite having different technologies. For example, a low-productivity firm in Toronto may offer the same wages as a high-productivity firm in Hamilton if the city-level productivity advantage of Toronto offsets the firm-level productivity gap.

To isolate true firm heterogeneity, the clustering procedure must account for local productivity differences. I address this by developing an iterative estimation procedure that guesses and sequentially updates the vector of city composite productivity, $\mathbb{A}_c \equiv A_c \bar{L}_c^{\mu}$, which incorporates both exogenous and (average) endogenous components of city productivity. The agglomeration elasticity μ is estimated after the procedure converges. The iterative procedure includes three steps, and I now describe each step in detail.

Worker skill and firm productivity In the first iteration, start with a guess of a vector of $\{\mathbb{A}_c\}_c$. With the guess, I estimate worker skill and firm productivity in the first step. In Section E.3 of the Online Appendix, I construct a measure of adjusted log earnings, denoted by $\log \check{W}_{ijt}$, by netting out the city-level earnings determinants (including the city composite productivity guess and the housing rent component in the earnings equation) and time-varying firm-level and city-level shocks from log earnings $\log W_{ijt}$. The adjusted log earnings $\log \check{W}_{ijt}$ only contains worker permanent skills a_i, firm technology (z_j, θ_j), and the transient skill shocks $\hat{a}_{it}$:

$$\log \check{W}_{ijt} = \log \chi + \underbrace{\log z_j}_{\text{firm productivity}} + \underbrace{\theta_j \log a_i}_{\text{worker-firm interaction}} + \hat{a}_{it}, \tag{4.9}$$

where $\chi \equiv \frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}$. This equation follows the same empirical specification as in BLM. I classify firms into $K = 10$ clusters using the k-means clustering algorithm based on firms' empirical distributions of adjusted log earnings.[15] Having clustered firms, the moment con-

[15] I use a vector of 20 percentiles from each firm's adjusted log earnings distribution as the input for k-means clustering. In addition, the estimation is robust to alternative numbers of clusters, $K = 20, 30, 40, 50$. Despite the parsimony, the 10 (50) cluster fixed effects account for 87 (90) percent of the between-firm variance in log earnings. All firms remain in the same cluster throughout the sample period.

dition for identifying $\{z_k, \theta_k\}_{\forall k}$ is:

$$\mathbb{E}\left[\left(\frac{\log \bar{W}_{ij(t+1)}}{\theta_{k'}} - \frac{\log z_{k'}}{\theta_{k'}}\right) - \left(\frac{\log \bar{W}_{ijt}}{\theta_k} - \frac{\log z_k}{\theta_k}\right) \mid k \neq k'\right] = 0. \tag{4.10}$$

Equation (4.10) indicates that $\{z_k, \theta_k\}_k$ are identified from wage changes of between-cluster movers, and it gives $K \times K$ moments to identify $2K$ parameters. As discussed in BLM, identification of the skill-augmenting productivity θ exploits differences in earnings changes of worker movers in opposite directions, i.e. mover from k to k' and from k' to k, provided that they differ in skills, i.e. $\mathbb{E}_{kk'}(a) \neq \mathbb{E}_{k'k}(a)$.[16] With the firm productivity parameters identified, each worker's permanent skill a_i can be retrieved using a plug-in estimator: $a = \mathbb{E}_t\left[\frac{1}{\theta_{j(i,t)}}\left(\log \check{W}_{ij(i,t)t} - \log z_{j(i,t)}\right)\right]$.

The key identification assumption in equation (4.10) is that worker mobility across firms is uncorrelated with transitory skill shocks—commonly referred to as the "exogenous mobility" assumption. This assumption is satisfied under Assumption 3. Empirically, violations may arise due to negative earnings shocks prior to job changes (e.g., Ashenfelter dip) or selection on unobserved match-specific productivity. To assess this assumption, I follow Card et al. (2013) and construct an event-study analysis, grouping movers by their origin and destination firm clusters. Figure I.1 in the Online Appendix shows parallel pre-move earnings trends and nearly symmetric changes for workers moving in opposite directions, supporting the exogenous mobility assumption.[17]

Firm and city amenities In the next step, I identify the firm and city amenity parameters from the following equation:

$$R_c(a) \cdot G_k(a) = \frac{\bar{r}_c^{\eta}}{\mathbb{A}_c r_c^{\frac{\alpha}{\alpha-1}} \cdot z_k a^{\theta_k}} \cdot \bar{J}_c(k)^{\frac{\beta_w}{\rho_w}} \cdot \Lambda_{kc}(a)^{\frac{\rho_w}{\beta_w}} \Lambda_c(a)^{\frac{1}{\beta_w}}, \ \forall a, k, c, \tag{4.11}$$

where $\bar{J}_c(k)$ is the number of cluster-k firms in city c, $\Lambda_{kc}(a)$ is the share of skill-a workers residing in city c that are employed in cluster-k firms, $\Lambda_c(a)$ is the share of skill-a workers that reside in city c.

Equation (4.11) provides $K \times C$ moments to identify $K + C$ amenity parameters for each skill, subject to one normalization. Intuitively, a firm cluster (city) is inferred to have higher amenities if it attracts more workers despite offering lower nominal (real) earnings. For each

[16] This asymmetry is supported by data available upon request.

[17] As noted by BLM, symmetric earnings changes alone are not sufficient to rule out skill-augmenting productivity. Borovičková and Shimer (2024) find that, in the presence of search frictions, selective hiring can generate both assortative matching and endogenous mobility, the latter of which cannot be detected by an event-study design.

worker skill, cluster-level amenities are identified by comparing worker sorting shares within a city, while city-level amenities are identified from worker shares of those working in the firm cluster. Since the amenities are assumed to be time-invariant, all the empirical measures are averaged over the sample period.

City composite productivity The last step of the iteration is to update the guess of city composite productivity $\mathbb{A}_c$, where I leverage the revealed firm location choices using the empirical counterpart of equation (2.11):

$$\bar{p}_c(k) = \frac{\left(\mathbb{A}_c \bar{r}_c^{\frac{\alpha}{\alpha-1}} \cdot \phi_{kc}\right)^{\beta_f}}{\sum_{c'} \left(\mathbb{A}_{c'} \bar{r}_{c'}^{\frac{\alpha}{\alpha-1}} \cdot \phi_{kc'}\right)^{\beta_f}}, \ \forall k, c. \tag{4.12}$$

As I have clustered firms in the previous step, I can construct $\bar{p}_c(k)$, the average share of cluster-k firms located in city c, from the data. The labor composite term ϕ_{kc} can be computed using the estimated labor supply parameters β_w and ρ_w, worker skill a, firm productivity parameters $\{z_k, \theta_k\}_k$, and amenity parameters $\{G_k(a)\}_{k,a}$ The term ϕ_{kc} captures the labor supply and competition conditions faced by a cluster-k firm in city c and does not depend on $\mathbb{A}_c$.

Equation (4.12) provides $C \times K$ moments to identify the C city-level composite productivity parameters, subject to one normalization. A city is inferred to have high city productivity if it can attract a larger share of firms relative to other cities with lower housing cost or better access to labor. The identification relies on the assumption that all firm types perceive city productivity identically. As a result, the model cannot perfectly match the cluster-specific sorting shares $\bar{p}_c(k)$ and the variation in $\{\bar{p}_c(k)\}_k$ across clusters within a city c is solely driven by differences in ϕ_{kc}. Nevertheless, I show below that the estimated model can closely match average firm productivity across cities, which lends support to the identifying assumption and highlights the central role of labor market access in shaping the spatial sorting of firms.

The iterative procedure concludes once the $\{\mathbb{A}_c\}_c$ vector has converged. Then, I estimate the agglomeration elasticity μ by regressing $\log \mathbb{A}_c$ on $\log \bar{L}_c$, which is instrumented by immigration-based instrumental variable following Card (2001). The exogenous city productivity is then retrieved as $A_c = \mathbb{A}_c \cdot \bar{L}_c^{-\mu}$.

Discussions on the iterative procedure Two points about the iterative procedure merit discussion. First, convergence of the iterative procedure is guided by the observed firm sorting shares. Suppose that in a given iteration, the composite productivity of city c is

guessed to be greater than its true value. This leads to underestimated firm productivities z_j for firms in c, implying a counterfactual less competitive labor market and inflated $\{\phi_{kc}\}_k$ for firms in the city. Consequently, city c appears excessively profitable, attracting more firms in the model than observed in the data. The algorithm would then adjusts city c's productivity downward.

Second, prevalent measurement error—particularly in smaller cities—hampers convergence of the iterative procedure. To mitigate this, I smooth the updated guess of $\{\mathbb{A}_c\}_c$ against city population using a cubic polynomial after each iteration and assess convergence based on the resulting smoothed vector. While this approach preserves the model's ability to capture the role of city productivities in the urban earnings premium, it artificially compresses the variance in estimated city productivities, rendering a variance decomposition exercise unreliable.

4.3.3 Housing supply and demand parameters

Lastly, I estimate the housing supply elasticity using the following regression:

$$\Delta \log r_{ct} = \Gamma \Delta \log EH_{ct} + \Delta e^r_{ct} \tag{4.13}$$

where $\Gamma \equiv \frac{\gamma}{1+\gamma}$, r_{ct} is the average housing rent of city c in year t, and EH_{ct} is the total housing expenditure in city c in year t. Equation (4.13) is estimated using five-year city-level changes between 2002 and 2007, which is chosen to avoid the 2008–2009 housing bust and the subsequent surge in foreign investment in Canada's real estate market. I instrument for $\Delta \log EH_{ct}$ using a shift-share Bartik instrument. Then, I calibrate $\bar{H}^0_c$ to match the average housing rent of each city. Following Saiz (2010), I also allow γ to vary across cities by the share of undevelopable land.

For the parameters governing housing demand, I calibrate the worker share of expenditure on housing $\eta = 0.24$ following Davis and Ortalo-Magné (2011),[18] and the output elasticity of housing in the production function $\alpha = 0.06$ using information on firms' housing and wage bill expenditures.[19]

[18] Following Eeckhout et al. (2014), I also calibrate heterogeneous housing expenditure shares as a robustness check for the counterfactual analysis.

[19] I calculate total housing services as value of buildings $\times$ (depreciation rate + property tax rate + mortgage rate − capital gain rate) + real estate rental. Then, I calculate α as the share of total housing services over (total housing services+total wage bill).

Table 1: Sorting elasticity and housing supply and demand parameters

Parameter	Description	Value (S.E.)
β_w/ρ_w	Firm-level labor supply elasticity	6.1 (1.0)
β_w	City-level labor supply elasticity	2.1 (0.03)
β_f	Firm sorting elasticity	5.6 (1.0)
$1/\gamma$	Housing supply elasticity	1.2 (2.0)
α	Production elasticity of housing	0.06
η	Expenditure share on housing	0.24

Notes: The preferred specifications for the parameter estimates are discussed in the text. The standard errors (S.E.) are calculated using the delta method.

4.4 Estimation results

4.4.1 Sorting elasticity and housing supply elasticity parameters

I now discuss the estimation results of the sorting elasticity and housing supply elasticity parameters. Table 1 summarizes the estimated parameter values under the preferred specifications along with the calibrated parameters related to housing demand.

Table E.1 in the Online Appendix reports passthrough estimates for workers. For the firm-level passthrough, the IV estimate is 0.13 (column 2), about half the OLS estimate (0.26, column 1). Controlling for changes in labor demand (column 3) raises it slightly to 0.14, implying a labor supply elasticity of 6.1—consistent with prior work (e.g., Card et al. (2018), LMS).[20] For the city-level passthrough, OLS and IV estimates are similar at 0.28 and 0.30 (columns 1–2). Controlling for labor demand changes lowers the estimate to 0.25 (column 6), while further controlling for changes in high-skill shares raises it to 0.32 (column 7). This implies a city-level labor supply elasticity of 2.1, within the range found in the literature (e.g., Suárez Serrato and Zidar (2016); Hornbeck and Moretti (2024); Bilal (2023)).

Table E.2 in the Online Appendix reports passthrough estimates for firms. The IV estimate is 0.21 (column 2), well below the OLS estimate of 1.06 (column 1), suggesting substantial measurement error correlated with city-level wage bills. Controlling for labor competition (column 3) lowers the estimate to 0.15, implying a firm sorting elasticity of 5.6—within the range reported by Gaubert (2018) estimated for different tradable industries, though slightly above estimates reported in Suárez Serrato and Zidar (2016) and Fajgelbaum

[20]The estimate rises to 0.15 when assuming $\rho_w = 1$, in line with studies that abstract from the spatial variation in labor supply (e.g., Card et al. (2018)).

Table 2: Firm productivity parameter estimates by cluster

Cluster	1	2	3	4	5	6	7	8	9	10
$\log z$	0.00	0.67	1.26	1.45	1.46	1.74	1.78	1.82	1.88	2.08
θ	1.00	1.27	1.60	1.65	1.54	1.74	1.90	1.75	1.74	1.84
Count	83700	109200	103800	90400	71900	62700	43400	49500	35600	36800

Notes: Firm productivity parameter estimates by cluster, ranked by z. The number of firms are rounded to the nearest hundred. The estimates are normalized to $z = \theta = 1$ for the lowest-paying cluster.

et al. (2019).

Table E.3 in the Online Appendix reports housing supply elasticity estimates. The OLS estimate is 0.53 (column 1), while the IV estimate decreases slightly to 0.45 (column 2) and is not statistically significant, likely due to limited instrument strength (first-stage F = 24.7). This implies a housing supply elasticity of 1.25, close to the mean of 1.5 estimated Saiz (2010). The elasticity does not vary significantly by the share of undevelopable land (column 3), so I use a uniform elasticity across cities in the counterfactual analyses.

4.4.2 Productivity parameters

Table 2 reports the estimated firm productivity parameters, with a normalization of $z = \theta = 1$ of the lowest-paying firm cluster. There is substantial variation in both common and skill-augmenting productivity. In particular, these two productivity terms are highly correlated (coefficient = 0.97), indicating strong worker-firm complementarity.

The estimation produces a continuous skill distribution. In practice, I discretize the skill distribution into 100 groups, further splitting the top percentile into 21 bins to capture the spatial sorting of top-skilled workers, yielding 120 groups in total. Figure I.2 in the Online Appendix validates the worker skill estimates, showing a strong correlation of city-level high-skilled worker shares based on the model estimates with the city-level shares of college-educated workers using the 2016 Census data.

Figure I.4 in the Online Appendix shows that larger cities have limited exogenous productivity advantages. Table E.3 reports a small and statistically insignificant agglomeration elasticity μ. Together, these results indicate that differences in city productivity play a minor role in explaining the urban earnings premium—a point I revisit later in the paper. Table I.3 reveals that cities with warmer summers tend to be more productive, and those in eastern and northern Canada tend to have greater productivity advantages.

4.4.3 Amenity parameters

I present the firm and city amenity estimates in Figure E.1 in the Online Appendix. Panel (a) shows that high-productivity firm clusters tend to offer lower amenities, especially for high-skilled workers, possibly due to a more demanding work environment. Panel (b) shows that larger cities provide higher amenities for all workers and are especially preferred by higher-skilled workers, aligning with the findings in Diamond (2016). Table I.3 shows that better city amenities are associated with milder winters, breezy summers, safe streets, and mountainous terrain.

4.5 Model fit

The model cannot perfectly match the worker sorting shares for each city-cluster bin or the cluster-specific firm sorting shares for each city in the data. Since the spatial sorting of workers and firms is highly interdependent, any discrepancies in matching the sorting shares could weaken the model's ability to explain the relationship between two-sided sorting and spatial inequality. It is therefore important to assess how well the model aligns with key sorting and inequality patterns in the data.

Figure I.3 in the Online Appendix shows that the model fits the data well across all the key dimensions. Panels (a) and (b) demonstrate that the model closely reproduces the city-level mean worker skill and firm common productivity, with a less precise fit for the latter as expected. Panel (c) shows that the model captures the within-city matching pattern well, measured by the share of top-decile workers employed by the three highest-paying firm clusters. This, in turn, helps the model accurately replicate the distributions of city mean earnings and population, as shown in panels (d) and (e). Finally, panel (f) shows that the model generates an urban earnings premium of 0.018, closely matching the empirical estimate of 0.021. It also shows that the model provides a better fit for larger cities than smaller ones.

4.6 Discussions

Skill sorting and learning Baum-Snow and Pavan (2012) and De La Roca and Puga (2017) show that greater learning environments in big cities play a crucial role in driving the urban wage premium. However, this paper does not aim to distinguish between static sorting and dynamic learning.[21] Instead, I focus on characterizing cross-sectional heterogeneity

[21]Table I.1 in the Online Appendix shows that controlling for big city experience slightly decreases the urban earnings premium. One can rightfully think that the local learning benefits are absorbed into the city amenities in a static model.

in workers and firms across cities and evaluating the effects of place-based policies given their spatial distributions. As higher-skilled workers have been shown to benefit more from learning spillovers (Davis and Dingel, 2019), an optimal spatial policy that accounts for these dynamic gains would promote even stronger spatial sorting.

Endogenous labor market power Hirsch et al. (2022) show that larger cities have more competitive labor markets, resulting in smaller wage markdowns. While the model deliberately abstracts from endogenous labor market power, the average wage markdown is absorbed into the city composite productivity term, $\mathbb{A}_c$, which does not vary significantly across the city-size distribution.[22] I also assess this issue by estimating firm-level passthrough separately for the five largest cities and smaller ones. The estimated passthrough is 0.15 (0.02) in large cities and 0.12 (0.02) in smaller ones, a difference that is not statistically significant, suggesting limited spatial variation in monopsony power in Canada.

Multi-location firms Recent work has studied the location decisions of multi-location firms (Oberfield et al., 2024; Kleinman, 2022; Jiang, 2023), which is beyond the scope of this paper. Since establishments within a firm share non-rival inputs like managerial expertise, their productivity should be highly correlated. As a validation, I decompose productivity variance for multi-city firms and find that 89.7% is explained by within-firm, between-city differences. This confirms the strong within-firm correlation recovered by the estimation, while highlighting the model's ability to capture heterogeneity across cities within firms.

5 Quantitative Analysis

This section presents the quantitative analyses based on the estimated model, including: (1) a structural decomposition of the urban earnings premium that separates the roles of worker and firm sorting in Section 5.1; (2) the implementation of an optimal spatial policy to maximize social welfare in Section 5.2; and (3) the evaluation of a place-based subsidy in Section 5.3.[23]

[22] See Berger et al. (2022) for a formal analysis of endogenous labor market power in a setting where firm locations are fixed. See also Berry and Reiss (2007) for a discussion of the challenges in studying entry games with endogenous market power and strategic interactions.

[23] I present additional counterfactual exercises in Section H of the Online Appendix, including loosening housing supply constraints in high-productivity cities and the rise of remote work. Both exercises demonstrate that two-sided sorting amplifies the distributional and aggregate consequences of place-based policies and space-related technological shocks.

Table 3: Structural decomposition of the urban earnings premium

	Mean log earnings	*Earnings components:*			
		City	Worker	Firm	Interaction
	(1)	(2)	(3)	(4)	(5)
Log Population	0.021**	0.001	0.014*	0.006	−0.000
	(0.008)	(0.006)	(0.005)	(0.005)	(0.001)
% Explained	100.0%	6.6%	66.8%	26.8%	−0.2%
Num. obs.			66		
R^2	0.100	0.071	0.023	0.010	0.000

Notes: This table shows the decomposition results of city-size regressions of mean city log earnings, based on equation (5.1). The productivity parameters are estimated using the iterative procedure described in Section 4.3. All regressions are weighted by city population. *p<0.1; **p<0.05; ***p<0.01.

5.1 Decomposition of the urban earnings premium

With the estimation results of the productivity parameters, I now decompose city mean log earnings into the city, firm, worker, and interaction components as follows:

$$\mathbb{E}_c \log(W_{ijt}) = \log \chi + \underbrace{\mathbb{E}_c \log\left(A_c \cdot L_{ct}^{\mu} \cdot r_{ct}^{\alpha/(\alpha-1)}\right)}_{\text{city characteristics}} + \underbrace{\bar{\theta} \cdot \mathbb{E}_c \left(\log a_i - \log \bar{a}\right)}_{\text{worker sorting}} + \underbrace{\mathbb{E}_c \left(\log z_j + \theta_j \log \bar{a}\right)}_{\text{firm sorting}} + \underbrace{\mathbb{E}_c \left[\left(\theta_{j(i,t)} - \bar{\theta}\right) \cdot \left(\log a_i - \log \bar{a}\right)\right]}_{\text{interaction component}}. \tag{5.1}$$

The decomposition takes the spatial allocations of workers and firms as given; thus, it does not require solving the model. The interaction component captures the contribution of positive assortative matching via skill-augmenting productivity to city mean log earnings. Accordingly, the worker component is evaluated at average skill-augmenting productivity and the firm component includes the average worker skill multiplied by its skill augmenting productivity.

To quantify the contributions of each component to the urban earnings premium, I regress city mean log earnings and each right-hand-side component on log city population in Table 3.[24] The results show that worker and firm sorting together explain over 90% of the urban earnings premium—66.8% and 26.8%, respectively—highlighting the central role of two-sided

[24] I also decompose within-city inequality and examine how the variance components correlate with city size. Table I.2 in the Online Appendix shows that higher earnings dispersion in larger cities is primarily driven by greater variance in worker skill and stronger covariances between worker skill and firm productivity.

sorting in spatial earnings inequality. In contrast, the city component accounts for only 6.6%, indicating that the productivity advantages of larger cities are modest and partially offset by higher housing costs. The prominent role of worker sorting is consistent with Dauth et al. (2022) and Card et al. (2025), and the limited role of city productivity aligns with the findings of Carry et al. (2025). Perhaps surprisingly, the interaction component shows no systematic relationship with city population. This finding is consistent with BLM and LMS: while variation in θ drives positive assortative matching, the interaction term explains a small share of the overall earnings variance. Nevertheless, as I show next, worker–firm complementarities are crucial for explaining the observed co-location pattern and the spatial earnings structure.

Three of the four coefficients on the city earnings components in Table 3 are not statistically significant. This is likely due to the measurement errors and the limited number of cities in the sample. In addition, the estimated urban earnings premium of 0.02 in Canada is substantially lower than in other developed economies—for example, 0.07 in the U.S. (Card et al., 2025), 0.04 in Germany (Dauth et al., 2022), and 0.05 in Spain (De La Roca and Puga, 2017). The limited variation makes it more challenging to precisely identify the contributions of the different components.

5.1.1 Understanding spatial sorting using a Shapley value approach

The decomposition in the previous section is based on the equilibrium spatial distribution of workers and firms. I now turn to use model counterfactuals, employing the Shapley value approach, to uncover the forces driving the spatial sorting patterns and spatial inequality.

In the model, spatial earnings disparities arise from variation in three parameter groups: (i) city productivity parameters A_c, μ, and α; (ii) skill-specific amenities $\{R_c(a)\}$ and $\{G_k(a)\}$; and (iii) firm skill-augmenting parameters θ_k. To quantify the contributions of each parameter, I apply the Shapley value method, which involves solving counterfactual economies where variation in each parameter is sequentially removed.[25] The contribution of each parameter is then computed as a weighted average of the resulting changes in equilibrium outcomes across all possible combinations. See Section G in the Online Appendix for further details. I focus on four outcomes: the urban earnings premium, worker and firm sorting gradients, and the covariance between city-level mean worker skill and firm productivity. Table 4 reports the results.

Column (1) shows that city characteristics explain 17.0% of the urban earnings premium.

[25]There is no between-city variation in μ and α. I set them to zero to compute their respective Shapley values. I preserve heterogeneity in worker skill and firm common productivity in the Shapely value calculation.

Table 4: Shapley decomposition of the urban earnings premium and spatial sorting measures

	Urban Premium	Spatial Sorting		Sorting Cov.
		Worker	Firm	
	β^w	β^a	β^z	$\text{Cov}(\bar{a}_c, \bar{z}_c)$
	(1)	(2)	(3)	(4)
City prod. characteristics:	17.0%	−17.8%	−26.8%	−4.6%
City prod. A_c	13.2%	−19.7%	−30.8%	−4.7%
Agglomeration μ	25.9%	−2.5%	−4.1%	−0.4%
Housing share in prod. α	−22.1%	4.4%	8.2%	0.5%
Amenities:	42.4%	113.5%	−62.9%	60.4%
City amenity $R_c(a)$	71.4%	124.5%	121.0%	58.9%
Firm amenity $G_k(a)$	−29.1%	−11.0%	−183.9%	1.5%
Skill-augmenting prod. θ_k:	40.6%	4.2%	189.6%	44.2%

Notes: This table reports the Shapley decomposition of four equilibrium outcomes. The *city prod. characteristics* row aggregates the three parameters listed below it, and the *amenities* row aggregates the two below.

Exogenous productivity A_c and agglomeration elasticity μ contribute 13.2% and 25.9%, respectively, while the housing share α contributes −22.1%, meaning the premium would rise if production did not require housing. Variation in amenities accounts for 42.3%, with city and firm amenities each contributing 71.4% and 29.1%, respectively. Larger cities offer better amenities for high-skilled workers, whereas more productive firms offer worse, shaping worker sorting patterns. Variation in skill-augmenting productivity θ_k explains 40.6% by concentrating high-skilled workers and high-productivity firms in larger cities.

Columns (2)–(3) reveal that the worker skill gradient is primarily driven by better amenities in larger cities. The variation in skill-augmenting productivity θ_k has little effect on worker sorting but significantly impacts firm sorting. Because of the worker-firm production complementarity, high-productivity firms sort into larger cities in which the high-skilled workers tend to locate. This provides a micro-foundation of the complementarity between city size and firm productivity previously found in Combes et al. (2012), Gaubert (2018) and Bilal (2023). City productivity characteristics only affect the sorting gradients through their impacts on the city size distribution. Finally, column (4) shows that amenities account for 60.4% of the sorting covariance, while production complementarity contributes another 44.6%.

These Shapley value decomposition results yield two insights on the drivers of two-sided

sorting. First, production complementarity plays a critical role in shaping how workers and firms sort across cities, although it does not directly affect the urban earnings premium through the interaction term in Table 3. Second, the spatial sorting of firms is more influenced by workers than the reverse, highlighting the central role of worker-driven dynamics in shaping spatial disparities.

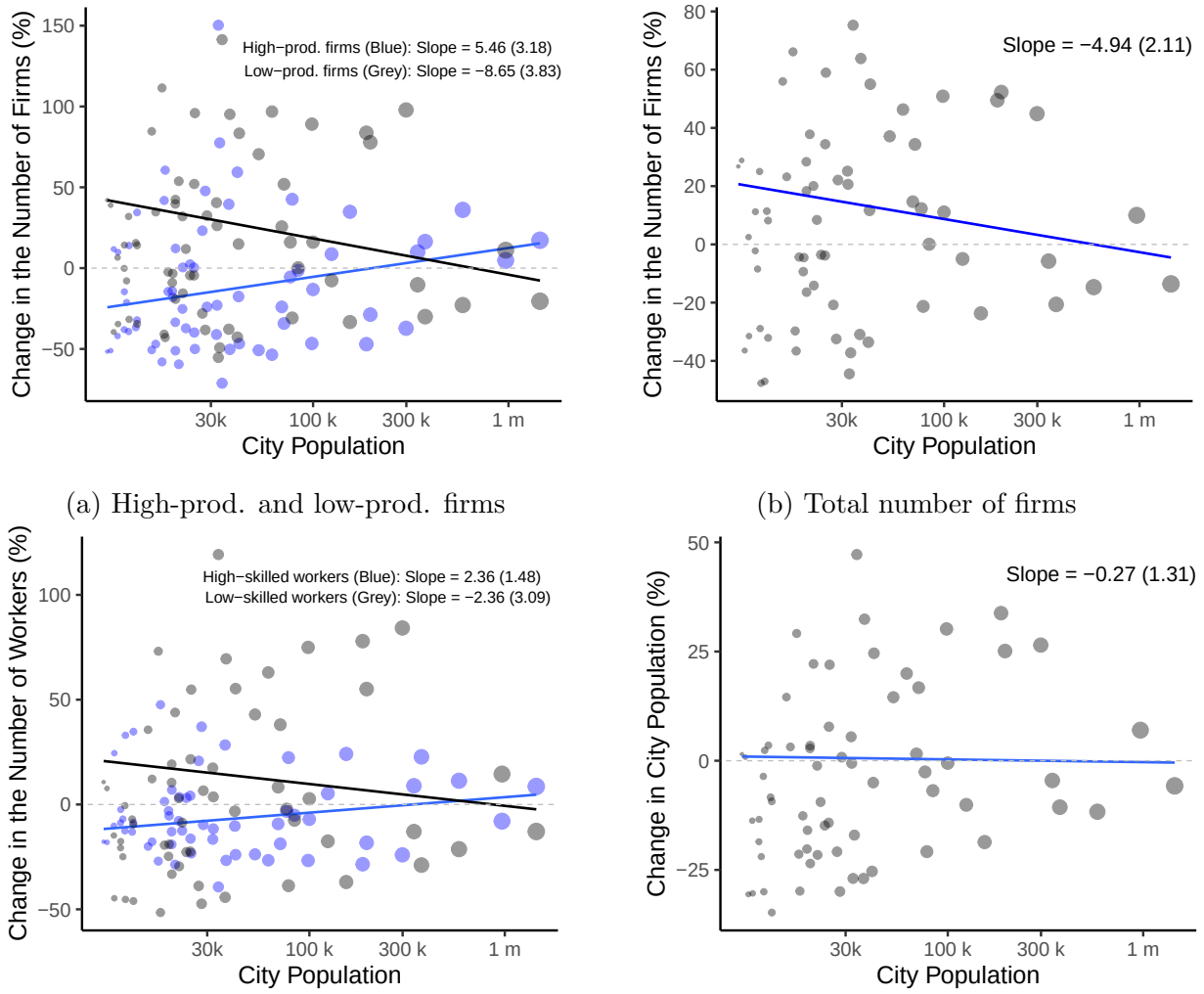
5.2 Optimal spatial policy

I now discuss the results of implementing the optimal spatial policy specified in Proposition 3. Recall that the optimal spatial policy consists of a set of tax instruments, $\{t^w_{kc}(a)\}_{\forall k,c,a}$ for workers and $\{t^f_c(k)\}_{\forall k,c}$ for firms. These tax instruments maximize social welfare by 1) taxing agents to ensure they internalize sorting externalities, 2) redistributing within types to equalize the marginal utility of consumption, and 3) redistributing between types depending on welfare weights.

I begin by assuming a utilitarian government, setting $\varphi^w(a) = 1, \forall a$ and $\varphi^f(k) = 1, \forall k$, and then conduct robustness checks using alternative weighting schemes. I assume that the social welfare function is sufficiently concave to guarantee a unique equilibrium under the implemented policy. For a formal treatment of equilibrium uniqueness in the context of optimal spatial policies, see Fajgelbaum and Gaubert (2020).

5.2.1 Sorting efficiency and spatial reallocation

Figure 1 plots the change in the number of firms and workers induced by the optimal policy against initial city population, classifying the top three firm clusters by z as high-productivity and the top three skill deciles of workers as high-skilled. Panel (a) shows that the optimal policy reallocates high-productivity firms toward larger, more skilled cities while shifting low-productivity firms away. With optimal profit taxes, firms internalize both labor market stealing and love-of-variety externalities. For low-productivity firms, the negative labor market externality dominates in large cities, as they compete with high-productivity firms for workers, while the love-of-variety benefit remains limited due to their small employment size. In contrast, high-productivity firms generate larger social gains in larger cities by complementing high-skilled workers and employing more labor. They also draw more high-skilled workers to these cities, better aligning with the planner's objective (Equation (3.4)). Overall, the optimal policy reduces the number of firms in large cities while raising average firm productivity. These findings are consistent with Gaubert (2018) and Bilal (2023), both of which advocate for a greater concentration of productive firms in larger cities. I contribute to the literature by introducing two-sided heterogeneity and emphasizing the importance of



(a) High-prod. and low-prod. firms

(b) Total number of firms

(c) High-skilled and low-skilled population

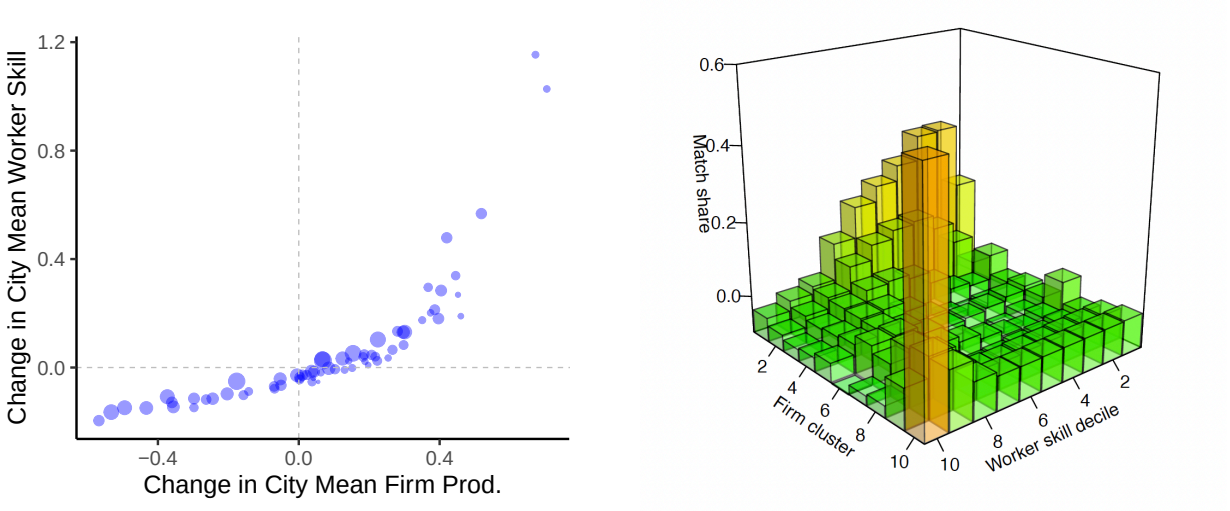
(d) Total population

Figure 1: Changes in the spatial allocation of workers and firms: optimal policy versus laissez-faire equilibrium

Notes: These figures display how the optimal policy shifts the spatial distribution of firms and workers for each city. City population are calculated in the laissez-faire equilibrium. Population-weighted regression coefficients and standard errors are reported.

worker–firm complementarity in the design of the optimal spatial policy.

Panel (c) of Figure 1 shows modest increases in skill sorting, while panel (d) displays slight population declines in larger cities. Due to weak agglomeration spillovers, the optimal policy provides limited incentives for relocating to large cities. As shown later, redistributive spatial transfers from large to small cities further reduce the return of locating in large cities. Furthermore, recall that the national wedge in Equation (3.4) incentivizes the planner to reallocate more workers toward productive cities. Although changes in population are not



(a) Changes in $\bar{a}_c$ and $\bar{z}_c$

(b) Changes in worker-firm match shares

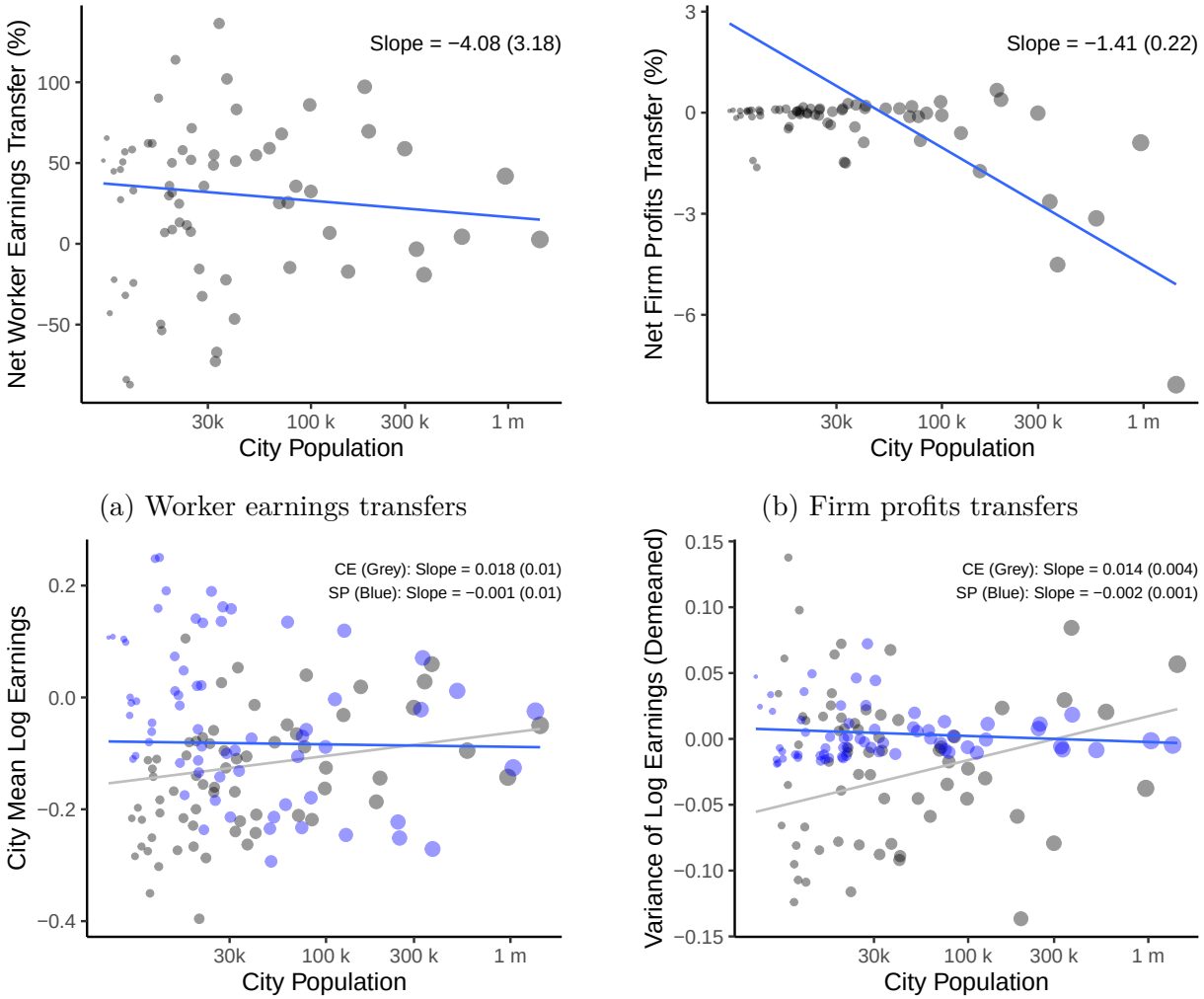
Figure 2: Changes in spatial sorting and worker-firm matching: optimal policy versus laissez-faire equilibrium

Notes: These figures show how the optimal policy changes between-city spatial sorting and within-city assortative matching. Panel (a) displays the relationship between city-level changes in mean worker skill and mean firm productivity. Panel (b) displays changes in the share of workers of each skill decile matched with each firm cluster of the entire economy.

strongly correlated with city size, the policy induces substantial reallocation across cities, driven by large productivity differences among similarly sized cities. Specifically, seven cities experience population growth exceeding 25%, while nine see declines of more than 25%.

Figure 2 shows how the optimal policy affects spatial sorting and worker-firm matching. Panel (a) shows significantly stronger co-location of high-skilled workers and high-productivity firms. The reallocation results in a 68.3% increase in the covariance between $\bar{a}_c$ and $\bar{z}_c$, with a rank correlation of 0.98 in their changes. Together with Figure I.5 in the Online Appendix, these results indicate that the optimal policy significantly strengthens the sorting of workers and firms into high-productivity cities. Panel (b) displays a substantial rise in assortative matching, with the share of top-decile workers employed by the most productive firm cluster increasing by about 50 percentage points.

By leveraging production complementarities, the optimal policy's spatial reallocation raises total output by 9.6%. However, without further intervention, the resulting increase in spatial sorting would worsen spatial inequality, reducing social welfare by amplifying spatial disparities in marginal utility of consumption. I show next that the optimal policy mitigates this through spatial transfers.



(a) Worker earnings transfers

(b) Firm profits transfers

(c) Average city log earnings

(d) Within-city variance in log earnings

Figure 3: Spatial transfers and changes in spatial inequality

Notes: Panels (a) and (b) show the relationships of net transfers for workers and firms as the shares of total income to city population. Panels (c) and (d) relate average city log earnings and demeaned within-city variance in log earnings to city population. Population-weighted regression coefficients and standard errors are reported.

5.2.2 Spatial transfers and changes in spatial inequality

Figure 3 illustrates the spatial transfers of the optimal policy and the resulting changes in the spatial earnings structure. Panels (a)–(b) show that the optimal policy significantly redistributes from larger, high-wage cities to smaller, low-wage ones, for both workers and firms. The spatial transfers combine (1) within-type redistribution to align marginal utility of consumption across cities and (2) between-type redistribution based on Pareto weights. Under utilitarian weights, the optimal policy redistributes towards low-skilled workers and

low-productivity firms, which are disproportionately concentrated in smaller cities.

Panels (c)–(d) show that the optimal policy eliminates the positive relationships between city population and both city average earnings and within-city earnings variance.[26] The former is a direct result of spatial transfers from larger towards smaller cities; the latter is primarily driven by a reduction in firm heterogeneity within each city.

5.2.3 Social welfare

The optimal policy results in a 47.4% increase in consumption-equivalent welfare. This substantial improvement is primarily driven by the redistribution motive. In the laissez-faire equilibrium, the marginal utility of consumption is considerably higher for low-income agents than for high-income agents. Consequently, even holding total output fixed, the transfers to low-income agents alone can generate significant welfare gains.

To isolate the efficiency gains, I design a system of type-specific lump-sum transfers that equalizes welfare gains across all workers and entrepreneurs while maintaining the spatial allocation of workers and firms. See Section C.4 in the Online Appendix for more details. This exercise results in a welfare gain of 5.0%, a magnitude comparable to the 4.0% welfare gain from the optimal spatial policy in Fajgelbaum and Gaubert (2020).[27]

Given the minimal agglomeration spillover elasticity μ, the welfare gains stem primarily from correcting firm sorting externalities. The variation in average firm size under the laissez-faire equilibrium indicates potential gains from the love-of-variety channel. I find that the optimal policy reduces the variance of average firm sizes across cities by 5.1%, resulting in a love-of-variety gain of 0.6%.[28] The remaining welfare increase is attributable to correcting labor market stealing externalities.

5.2.4 Discussions

Alternative welfare weights The degree of redistribution across worker and firm types under the optimal policy depends on the welfare weights. As a robustness check, I implement the optimal spatial policy using two alternative weighting schemes that restrict between-type

[26] The utilitarian government also redistributes income from entrepreneurs to workers. I adjust worker earnings by proportionally removing these transfers to isolate the effects of spatial sorting and worker-to-worker transfers. Additionally, the policy substantially reduces overall inequality through between-type transfers. To examine its varying impact across cities, I demean within-city log earnings variances in both equilibria.

[27] Davis and Gregory (2021) caution that the within-type redistribution in an optimal allocation depends on the randomness of preference heterogeneity, which cannot be identified from data. This exercise also addresses this concern.

[28] This is calculated as the change in the number of firms in each city, multiplied by the love-of-variety elasticity, and then aggregated across cities.

redistribution. First, I calibrate weights to equalize welfare gains across all agents.[29] Second, I apply Negishi weights, set as each type's average income in the laissez-faire equilibrium. The resulting welfare gains—5.7% and 5.2%, respectively—suggest that the efficiency gains are robust to alternative welfare weight specifications.

The role of production complementarity To highlight the role of production complementarity, I implement the corresponding optimal spatial policy in a counterfactual economy where I set $\theta_j = \bar{\theta}, \forall j$. In this counterfactual economy, the optimal policy generates less co-location of high-productivity firms and high-skilled workers (the rank correlation between $\Delta\bar{z}_c$ and $\Delta\bar{a}_c$ decreases from 0.98 to 0.85). Consequently, the output gain drops from 9.7% to 5.1% and the welfare gain decreases from 47.4% to 45.7%.

The role of two-sided heterogeneity Lastly, I demonstrate the importance of accounting for both worker and firm heterogeneity in designing the optimal policy. To this end, I re-estimate the model under two restricted specifications—one with only worker heterogeneity and one with only firm heterogeneity—and solve for the corresponding optimal policies. The results are shown in Figure I.6 in the Online Appendix. When firms are assumed to be homogeneous, the optimal policy reallocates more firms—rather than fewer—into large cities (panel (a)), as this helps attract more workers to productive locations (panel (b)). When workers are assumed to be homogeneous, the reallocation pattern is qualitatively similar but quantitatively much stronger. This is because, without accounting for worker heterogeneity, large cities are estimated to be significantly more productive, prompting greater social returns from spatial reallocation of firms. Taken together, the results highlight that overlooking two-sided heterogeneity can lead to qualitatively and quantitatively misguided policy prescriptions.

5.3 Place-based subsidies

Lastly, I evaluate a 5% wage subsidy for firms in the most productive cluster if they locate in Toronto, which emulates the city's bid to attract Amazon HQ2. I assume the subsidies are financed by a flat proportional labor income tax on all workers in the economy. I simulate the policy under various model scenarios to examine the importance of two-sided sorting, with the results in Table 5.

In the full model (column 1), the subsidy—equivalent to 0.3% of total GDP—has sizeable effects on both distributional and aggregate outcomes. In Toronto, the policy attracts more

[29] In practice, I group workers into 20 categories and calibrate weights for 20 worker groups and 10 firm clusters.

Table 5: Counterfactual analysis: subsidizing productive firms in Toronto

			Limited Re-sorting		
% Changes in	Full model (1)	No worker het. (2)	None (3)	Only firm (4)	Only worker (5)
Var. city log earnings	14.5%	−0.1%	1.8%	6.2%	2.5%
Total output	0.8%	0.2%	0.3%	0.7%	0.3%
Total welfare	−0.4%	−0.1%	−0.2%	−0.4%	−0.2%
High-skill welfare	−0.2%	–	−0.1%	−0.2%	−0.1%
Low-skill welfare	−0.5%	–	−0.2%	−0.5%	−0.2%
Pop. in Toronto	−2.3%	0.4%	–	–	0.0%
High-skill pop.	0.6%	–	–	–	0.5%
Low-skill pop.	−3.8%	–	–	–	−0.3%
# Firms in Toronto	−0.8%	2.0%	–	0.8%	–
# high-prod. firms	30.1%	7.9%	–	25.4%	–
# low-prod. firms	−7.7%	1.0%	–	−4.6%	–
Rent in Toronto	2.8%	0.8%	0.7%	2.7%	0.9%

high-productivity firms, which increase by 30.1%, and more high-skilled workers, who grow by 0.6%. However, increased labor market competition and a 2.8% rise in housing rents adversely affect low-productivity firms and low-skilled workers, leading to declines of 7.7% and 3.8%, respectively. These results are consistent with the empirical findings of Qian and Tan (2021), who show that the entry of skill-intensive firms benefits high-skilled workers but harms low-skilled ones. As a result, Toronto's total population falls by 0.8%, and the number of firms declines by 2.3%. At the national level, output increases by 0.8%, but at the cost of heightened spatial inequality and reduced welfare—particularly for low-skilled workers.[30]

When worker heterogeneity is ignored (column (2)), the policy has much weaker effects: fewer high-productivity firms relocate to Toronto, rent increases are smaller, and the output gain is limited, with minimal changes in spatial inequality or overall welfare. Without spatial re-sorting (column (3)), the subsidy boosts employment only among targeted firms in Toronto, with negligible effects elsewhere aside from its fiscal cost. In the one-sided re-sorting cases (columns (4)–(5)), mobility responses are muted relative to the full model, resulting in smaller changes in output, welfare, and spatial inequality. These comparisons underscore the importance of incorporating both two-sided heterogeneity and mobility when evaluating the aggregate and distributional effects of place-based subsidies.

[30]The counterfactual results are robust to incorporating endogenous amenities, free entry, and heterogeneous housing expenditure shares by skill, which are available upon request.

6 Conclusion

A city's economic fortunes are closely tied to the types of workers and firms it attracts. To study the causes and consequences of spatial sorting, I develop a model featuring a system of cities and mobile, heterogeneous workers and firms. I find that both worker and firm sorting are key drivers of spatial inequality, and that production complementarity is crucial for explaining the co-location of high-skilled workers and high-productivity firms.

The model informs novel sources of sorting externalities as a result of imperfect labor market competition and two-sided sorting. I show that the optimal policy can substantially increase social welfare and the total output of the economy. This is attained by increasing the co-location of high-skilled workers with high-productivity firms, along with a spatial redistribution towards low-earning cities. Model counterfactuals underscore the importance of two-sided sorting for policy evaluation.

The static equilibrium approach is a first step in studying two-sided spatial sorting. However, it abstracts from potentially important dynamic considerations. For example, larger cities provide learning opportunities that affect individuals' lifetime location decisions and earnings trajectories. Moving costs also influence mobility and short-term responses to local shocks. Future work can integrate these dynamic elements for a more comprehensive understanding.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67, 251–333.

Andrews, M. J., L. Gill, T. Schank, and R. Upward (2008): "High Wage Workers and Low Wage firms: Negative Assortative Matching or Limited Mobility Bias?" *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171, 673–697.

Autor, D. H., D. Dorn, and G. H. Hanson (2013): "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 103, 2121–68.

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

——— (2013): "Inequality and City Size," *Review of Economics and Statistics*, 95, 1535–1548.

Behrens, K., G. Duranton, and F. Robert-Nicoud (2014): "Productive Cities: Sorting, Selection, and Agglomeration," *Journal of Political Economy*, 122, 507–553.

Berger, D., K. Herkenhoff, and S. Mongey (2022): "Labor Market Power," *American Economic Review*, 112, 1147–93.

——— (2025): "Minimum Wages, Efficiency, and Welfare," *Econometrica*, 93, 265–301.

Berry, S. and P. Reiss (2007): "Empirical models of entry and market structure," *Handbook of industrial organization*, 3, 1845–1886.

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

Bonhomme, S., T. Lamadon, and E. Manresa (2019): "A Distributional Framework for Matched Employer Employee Data," *Econometrica*, 87, 699–739.

Borovĭcková, K. and R. Shimer (2024): "Assortative Matching and Wages: The Role of Selection," Working Paper 33184, National Bureau of Economic Research.

Card, D. (2001): "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration," *Journal of Labor Economics*, 19, 22–64.

Card, D., A. R. Cardoso, J. Heining, and P. Kline (2018): "Firms and Labor Market Inequality: Evidence and Some Theory," *Journal of Labor Economics*, 36, S13–S70.

Card, D., J. Heining, and P. Kline (2013): "Workplace Heterogeneity and the Rise of West German Wage Inequality," *Quarterly Journal of Economics*, 128, 967–1015.

Card, D., J. Rothstein, and M. Yi (2025): "Location, Location, Location," *American Economic Journal: Applied Economics*, 17, 297–336.

Carry, P., B. Kleinman, and E. Nimier-David (2025): "Location Effects or Sorting? Evidence from Firm Relocation," Working Paper 33779, National Bureau of Economic Research.

Combes, P.-P., G. Duranton, L. Gobillon, D. Puga, and S. Roux (2012): "The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection," *Econometrica*, 80, 2543–2594.

Dauth, W., S. Findeisen, E. Moretti, and J. Suedekum (2022): "Matching in Cities," *Journal of the European Economic Association*, 20, 1478–1521.

Davis, D. R. and J. I. Dingel (2019): "A Spatial Knowledge Economy," *American Economic Review*, 109, 153–70.

Davis, M. and J. M. Gregory (2021): "Place-Based Redistribution in Location Choice Models," Working Paper 29045, National Bureau of Economic Research.

Davis, M. A. and F. Ortalo-Magné (2011): "Household Expenditures, Wages, Rents," *Review of Economic Dynamics*, 14, 248–261.

De La Roca, J. and D. Puga (2017): "Learning by Working in Big Cities," *Review of Economic Studies*, 84, 106–142.

Deng, Z., R. Nirussette, and D. Messacar (2020): "Running the Economy Remotely: Potential for Working from Home during and after COVID-19," StatCan COVID 19: data to insights for a better Canada 45-28-0001.

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

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

Dingel, J. I. and B. Neiman (2020): "How Many Jobs Can be Done at Home?" *Journal of Public Economics*, 189, 104235.

Dixit, A. K. and J. E. Stiglitz (1977): "Monopolistic Competition and Optimum Product Diversity," *The American Economic Review*, 67, 297–308.

Duranton, G. and A. J. Venables (2018): "Place-Based Policies for Development," Working Paper 24562, National Bureau of Economic Research.

Eckert, F., S. Ganapati, and C. Walsh (2022): "Urban-Biased Growth: A Macroeconomic Analysis," Working Paper 30515, National Bureau of Economic Research.

Eeckhout, J., R. Pinheiro, and K. Schmidheiny (2014): "Spatial Sorting," *Journal of Political Economy*, 122, 554–620.

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

Fajgelbaum, P. D., E. Morales, J. C. Suárez Serrato, and O. Zidar (2019): "State Taxes and Spatial Misallocation," *Review of Economic Studies*, 86, 333–376.

Gaubert, C. (2018): "Firm Sorting and Agglomeration," *American Economic Review*, 108, 3117–53.

Glaeser, E. L. and J. D. Gottlieb (2008): "The Economics of Place-Making Policies," *Brookings Papers on Economic Activity*, 2008, 155–239.

Greenstone, M., R. Hornbeck, and E. Moretti (2010): "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings," *Journal of Political Economy*, 118, 536–598.

Helsley, R. W. and W. C. Strange (1990): "Matching and Agglomeration Economies in a System of Cities," *Regional Science and Urban Economics*, 20, 189–212.

Hirsch, B., E. J. Jahn, A. Manning, and M. Oberfichtner (2022): "The Urban Wage Premium in Imperfect Labor Markets," *Journal of Human Resources*, 57, S111–S136.

Hoelzlein, M. (2023): "Two-sided Sorting and Spatial Inequality in Cities," *Working paper*.

Hong, G. (2024): "Essays in Spatial and International Economics," Ph.D. thesis, University of Toronto (Canada).

Hornbeck, R. and E. Moretti (2024): "Estimating Who Benefits from Productivity Growth: Local and Distant Effects of City Productivity Growth on Wages, Rents, and Inequality," *The Review of Economics and Statistics*, 106, 587–607.

Hsieh, C.-T. and E. Moretti (2019): "Housing Constraints and Spatial Misallocation," *American Economic Journal: Macroeconomics*, 11, 1–39.

Jiang, X. (2023): "Information and Communication Technology and Firm Geographic Rxpansion," CESifo Working Paper 10452.

Kleinman, B. (2022): "Wage Inequality and the Spatial Expansion of Firms," Working paper.

Kline, P. and E. Moretti (2014): "People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs," *Annual Review of Economics*, 6, 629–662.

Kline, P., N. Petkova, H. Williams, and O. Zidar (2019): "Who Profits from Patents? Rent-Sharing at Innovative Firms," *Quarterly Journal of Economics*, 134, 1343–1404.

Kline, P., R. Saggio, and M. Sølvsten (2020): "Leave-Out Estimation of Variance Components," *Econometrica*, 88, 1859–1898.

Lamadon, T., M. Mogstad, and B. Setzler (2022): "Imperfect Competition, Compensating Differentials, and Rent Sharing in the US Labor Market," *American Economic Review*, 112, 169–212.

Lindenlaub, I., R. Oh, and M. Peters (2022): "Firm Sorting and Spatial Inequality," Working Paper 30637, National Bureau of Economic Research.

Mankiw, N. G. and M. D. Whinston (1986): "Free Entry and Social Inefficiency," *The RAND Journal of Economics*, 17, 48–58.

Moretti, E. and D. J. Wilson (2017): "The Effect of State Taxes on the Geographical Location of Top Earners: Evidence from Star Scientists," *American Economic Review*, 107, 1858–1903.

Oberfield, E., E. Rossi-Hansberg, P.-D. Sarte, and N. Trachter (2024): "Plants in Space," *Journal of Political Economy*, 132, 867–909.

Olley, G. S. and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1297.

Qian, F. and R. Tan (2021): "The Effects of High-skilled Firm Entry on Incumbent Residents," Working Paper 21-039, Stanford Institute for Economic Policy Research (SIEPR).

Redding, S. J. and E. Rossi-Hansberg (2017): "Quantitative Spatial Economics," *Annual Review of Economics*, 9, 21–58.

Roback, J. (1982): "Wages, Rents, and the Quality of Life," *Journal of Political Economy*, 90, 1257–1278.

Rosen, S. (1974): "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 82, 34–55.

Rossi-Hansberg, E., P.-D. Sarte, and F. Schwartzman (2019): "Cognitive Hubs and Spatial Redistribution," Working Paper 26267, National Bureau of Economic Research.

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

Suárez Serrato, J. C. and O. Zidar (2016): "Who Benefits from State Corporate Tax Cuts? A Local Labor Markets Approach with Heterogeneous Firms," *American Economic Review*, 106, 2582–2624.

Ziv, O. (2019): "The Micro-Geography of Productivity Density and Sorting," Working paper.

Online Appendix for "Two-Sided Sorting of Workers and Firms: Implications for Spatial Inequality and Welfare"

Appendix A Data

A.1 Canadian Employer-Employee Dynamics Database (CEEDD)

The main data source for the analysis is the Canadian Employer-Employee Dynamics Database (CEEDD), which is a set of linkable administrative tax files. These files include T1 Personal Master File (T1–PMF), T1 Family File (T1–FF), the National Accounts Longitudinal Microdata File (NALMF), the Statement of Remuneration Paid Files (T4), and the Longitudinal Immigration Database (IMDB). I use the data from 2010–2017 for the empirical analysis. I report the summary statistics in Table A.1.

T1 files Individual characteristics are obtained from the T1–PMF and T1–FF files, which are derived from the Income Tax and Benefit Return (T1) that all Canadians are required to file annually. The T1–PMF provides information on age, gender, marital status, and residential location, while the T1–FF reports the number of children. Each individual is identified by a unique longitudinal Social Insurance Number (SIN). In the T1–PMF, location in year t is based on the postal code reported in year $t+1$, since tax forms are filed the following year. Therefore, I assign individuals to their location in year t using the postal code recorded in the T1–PMF file for year $t-1$.

T4 files Annual earnings data and employee-employer linkages are from the Statements of Remuneration Paid (T4). The T4 files provide job-level earnings information with individual and firm identifiers, where a job is defined as a worker-firm pairing.

NALMF Firm information is obtained from NALMF. A firm is defined as a tax and accounting entity that files income tax returns and/or payroll remittances to the Canada Revenue Agency (CRA). The NALMF dataset covers firm location, industry (4-digit NAICS), firm size, payroll, and other balance sheet information.

IMDB The IMDB files provide information on immigrants since 1980, including each individual's gender, country of birth, year of landing in Canada and destination location.

Table A.1: Summary Statistics

Sample	Baseline	Stayers	Movers
Number of Observations			
Worker-Years (in 1,000)	51,475	7,504	23,572
Workers (in 1,000)	10,655	938	4,127
Firms (in 1,000)	1,588	16	1,250
Cities	66	66	66
Worker Characteristics			
Mean Log Earnings	10.44	10.85	10.29
Mean Age	41.68	43.84	39.96
Percent Male (%)	56.7	56.8	58.9
Percent in Largest 5 Cities (%)	56.7	62.6	56.8
Firm Characteristics			
Mean Log Wage Bill per Worker	10.21	10.71	10.22
Mean Log Wage Bill	11.07	14.85	11.35
Log Firm Size	0.86	4.14	1.13

Notes: This table reports summary statistics for the baseline, stayers, and movers samples. See Section 4.1 for the selection criteria for each sample. The numbers of worker-years, workers, and firms are rounded to the nearest thousand.

A.2 Other Data Sources

CPI I use the Consumer Price Index of all items from StatsCan to denominate all monetary values in 2002 Canadian dollars.

Housing rent I obtain housing rent data from the Canada Mortgage and Housing Corporation, which reports average monthly rents for Census Metropolitan Areas (CMAs) and Census Agglomerations (CAs) by unit type. I use the average monthly rent of two-bedroom units as the measure of housing cost, given its better coverage across cities.

Minimum wage I collect a panel of minimum wage data from the official websites of provincial and territorial governments. I then construct the annual national minimum wage as the lowest minimum wage across all provinces and territories in each year, which I use to construct the earnings cutoffs of each year for the baseline sample.

The share of undevelopable land Following Saiz (2010), I define undevelopable land as areas with a slope greater than 15 degrees or those covered by water. Slope is calculated

using elevation data from the CanVec Elevation Features dataset maintained by Natural Resources Canada (NRCan). Geographic boundaries of cities and water bodies are from the 2016 Census boundary files. In addition, I classify land designated under Ontario's Greenbelt Plan as undevelopable. The Greenbelt Act prohibits the rezoning of agricultural, heritage, and environmentally sensitive lands for urban development. The boundary file is obtained from Ontario GeoHub.

Share of college graduates by city I compute the share of individuals holding a college or higher degree in each city using data from the 2016 Population Census.

Other city characteristics For each city, I collect geographic coordinates (longitude and latitude) from Google Maps; climate data and the Air Quality Index (AQI) from Environment Canada; education quality indicators from the Conference Board of Canada; and data on road coverage and the Crime Severity Index from Statistics Canada.

Appendix B Model

B.1 Firm's problem

Under Assumption 1, I derive the FOCs of the firm's profit maximization problem:

$$W_{jc}(a): (1-\alpha)\frac{\beta_w}{\rho_w}{Q_{jc}}^{-\alpha}W_{jc}(a)^{\frac{\beta_w}{\rho_w}-1}\cdot A_c L_c^{\mu} z_j a^{\theta_j}\cdot {h_{jc}}^{\alpha}-\left(1+\frac{\beta_w}{\rho_w}\right)\cdot W_{jc}(a)^{\beta_w/\rho_w}=0 \tag{B.1}$$

$$h_{jc}: \alpha Q_{jc}^{1-\alpha}{h_{jc}}^{\alpha-1}-r_c=0, \tag{B.2}$$

where $Q_{jc}\equiv\int_{\underline{a}}^{\bar{a}} A_c L_c^{\mu}\cdot z_j\cdot a^{\theta_j}\cdot D_{jc}(a)\,da$ is defined as the total efficiency unit of labor. I then obtain the optimal wage offers and housing inputs from the FOCs as:

$$W_{jc}(a)=\chi\cdot A_c L_c^{\mu} r_c^{\frac{\alpha}{\alpha-1}}\cdot z_j\cdot a^{\theta_j} \tag{B.3}$$

$$h_{jc}=\alpha^{\frac{1}{1-\alpha}} r_c^{\frac{1}{\alpha-1}}\cdot Q_{jc}, \tag{B.4}$$

where $\chi\equiv\frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}\cdot(1-\alpha)\alpha^{\frac{\alpha}{1-\alpha}}$. Plugging equation (B.3) into Q_{jc} yields

$$Q_{jc}=(A_c L_c^{\mu})\cdot(z_j)^{1+\beta_w/\rho_w}\cdot\phi_{jc}, \tag{B.5}$$

where $\phi_{jc} \equiv \int_a \left(a^{\theta_j}\right)^{1+\beta_w/\rho_w} L_c(a) \cdot \frac{G_j(a)^{\frac{\beta_w}{\rho_w}}}{\sum_{j' \in \mathcal{J}_c}\left(z_{j'} a^{\theta_{j'}} \cdot G_{j'}(a)\right)^{\frac{\beta_w}{\rho_w}}} da$. Firm j's wage bill is the sum of wages to all workers it employs

$$E_{jc} = \int_{\underline{a}}^{\bar{a}} W_{jc}(a) D_{jc}(a)\, da = \chi \left(A_c {L_c}^{\mu} {r_c}^{\frac{a}{a-1}}\right) \cdot (z_j)^{1+\beta_w/\rho_w} \cdot \phi_{jc}. \tag{B.6}$$

Then, firm j's maximized profits in city c are

$$\pi_c(j) = {Q_{jc}}^{1-\alpha} {h_{jc}}^{\alpha} - r_c h_{jc} - E_{jc} = \Psi \cdot \left(A_c {L_c}^{\mu} {r_c}^{\frac{a}{\alpha-1}}\right) \cdot (z_j)^{1+\beta_w/\rho_w} \cdot \phi_{jc}, \tag{B.7}$$

where $\Psi \equiv \frac{1}{1+\beta_w/\rho_w} \cdot (1-\alpha)\alpha^{\frac{\alpha}{1-\alpha}}$.

B.2 The housing market

The housing developer of each city combines land $\bar{H}_c$ and the final good Y_c to produce housing. Given the housing rent r_c, the developer's profit maximization problem is

$$\max_{Y_c} r_c \cdot \bar{H}_c {Y_c}^{\frac{1}{1+\gamma_c}} - Y_c. \tag{B.8}$$

From the FOC, I can solve $Y_c = \left(\frac{1}{1+\gamma_c} r_c \bar{H}_c\right)^{\frac{1+\gamma_c}{\gamma_c}}$, which leads to the housing supply curve:

$$H_c^S(r_c) = \bar{H}_c^0 \cdot r_c^{\frac{1}{\gamma_c}}, \tag{B.9}$$

where $\bar{H}_c^0 \equiv (1+\gamma_c)^{-1/\gamma_c} \cdot {\bar{H}_c}^{(1+\gamma_c)/\gamma_c}$. The housing market clearing condition is

$$H_c^D(r_c) = \left(\tau\eta + \frac{\alpha}{1-\alpha}\frac{1+\beta_w/\rho_w}{\beta_w/\rho_w}\right)\frac{E_c}{r_c} = H_c^S(r_c), \tag{B.10}$$

where I define E_c as the total wage bill of city c. I assume that developers' profits are pooled nationally and redistributed to workers in proportion to their wages, that is

$$(\tau - 1)\sum_i W_i = \sum_c (r_c H_c - Y_c) = \sum_c \frac{\gamma_c}{1+\gamma_c} r_c H_c. \tag{B.11}$$

It is also useful to specify the final good market clearing condition, that is

$$\sum_c \sum_j \int_a (1-\eta) \cdot D_{jc}(a)\, \tau W_{jc}(a)\, da + \sum_c \sum_j \pi_c(j) + \sum_c Y_c = \sum_c \sum_j Q_{jc}^{1-\alpha} h_{jc}^{\alpha}. \tag{B.12}$$

Appendix C Efficiency

C.1 Social planner's problem

The planner aims to maximize the social welfare function, specified as

$$\mathcal{W} = \sum_{a=1}^{N^w} \varphi^w(a) \cdot L(a) U(a) + \sum_{k=1}^{N^f} \varphi^f(k) \cdot J(k) \Pi(k), \tag{C.1}$$

where $\varphi^w(a)$ and $\varphi^f(k)$ are the Pareto weights for skill a workers and type k entrepreneurs. Utility of each type of worker and firm is

$$U(a) = \frac{1}{\beta_w} \cdot \log\left(\sum_c U_c(a)^{\beta_w}\right) + \bar{C}^w \tag{C.2}$$

$$\Pi(k) = \frac{1}{\beta_f} \log\left(\sum_c c_c(k)^{\beta_f}\right) + \bar{C}^f, \tag{C.3}$$

where

$$U_c(a) = \left[R_c(a) \sum_k J_c(k) \left(G_k(a) \cdot c_{kc}(a)^{1-\eta} h_{kc}(a)^{\eta}\right)^{\beta_w/\rho_w}\right]^{\rho_w/\beta_w}, \tag{C.4}$$

$\bar{C}^w$ is a unrecoverable constant and $\bar{C}^f = \gamma/\beta_f$ with γ being the Euler-Mascheroni constant.

The planner chooses the following to maximize equation (C.1): 1) the amount of the final good and housing allocated to skill-a workers working for a type-k firm and city c, $c_{kc}(a)$ and $h_{kc}(a)$; 2) the amount of the final good allocated to a type-k entrepreneur in city c, $c_c(k)$; 3) the amount of the housing allocated to a type-k firm in city c for production, $h_c(k)$; 4) the total number of skill-a workers in type-k firms located in city c, $D_{kc}(a)$; 5) the number of type-k firms located in city c, $J_c(k)$ and 6) the amount of final good used to produce housing for each city c, Y_c.

The planner is subject to the following constraints. First, there are the spatial mobility and local labor matching constraints of workers and firms. For workers, the planner is subject to both within-city and between-city allocation constraints

$$\frac{D_{kc}(a)}{L_c(a)} = \frac{J_c(k) \cdot R_c(a) \left(G_k(a) \cdot c_{kc}(a)^{1-\eta} h_{kc}(a)^{\eta}\right)^{\beta_w/\rho_w}}{U_c(a)^{\beta_w/\rho_w}}, \ \forall c, k, a \tag{C.5}$$

$$\frac{L_c(a)}{L(a)} = \frac{U_c(a)^{\beta_w}}{\sum_c U_c(a)^{\beta_w}}, \ \forall c, a. \tag{C.6}$$

Combining (C.5) and (C.6), we can express $D_{kc}(a)$ as a fraction of $L(a)$

$$\frac{D_{kc}(a)}{L(a)} = J_c(k) \cdot \left(c_{kc}(a)^{1-\eta} h_{kc}(a)^{\eta}\right)^{\beta_w/\rho_w} \cdot U_c(a)^{\beta_w - \beta_w/\rho_w} \cdot \left(\sum_c U_c(a)^{\beta_w}\right)^{-1}, \ \forall c, k, a. \quad \text{(C.7)}$$

For firms, the planner is subject to the spatial mobility constraint:

$$\frac{J_c(k)}{J(k)} = \frac{c_c(k)^{\beta_f}}{\sum_c c_c(k)^{\beta_f}}, \ \forall c, k. \quad \text{(C.8)}$$

Second, there are worker and firm allocation constraints, which are satisfied given (C.7) and (C.8):

$$\sum_c \sum_k D_{kc}(a) = L(a), \ \sum_c J_c(k) = J(k). \quad \text{(C.9)}$$

Third, the planner is subject to a set of resource constraints. These include the resource constraint of the final good for the whole economy:

$$\sum_c \left(\sum_k \left[\sum_a D_{kc}(a) c_{kc}(a) + J_c(k) c_c(k) \right] + Y_c \right) \leq \sum_c \sum_k J_c(k) \left(\sum_a \frac{D_{kc}(a)}{J_c(k)} f_c(k,a) \right)^{1-\alpha} h_c(k)^{\alpha} \quad \text{(C.10)}$$

where I denote $f_c(k,a) = A_c L_c^{\mu} \cdot z_k a^{\theta_k}$, and the resource constraint of housing in each city

$$\sum_k \sum_a D_{kc}(a) h_{kc}(a) + \sum_k J_c(k) h_c(k) \leq \bar{H}_c Y_c^{1/(1+\gamma_c)}, \ \forall c. \quad \text{(C.11)}$$

Lastly, the planner is subject to non-negativity constraints for all consumption and housing allocations.

C.2 Solving the social planner's problem

Proof of Proposition 1 and 2. The Lagrange function of the social planner's problem is

$$\begin{aligned}
\mathcal{L} = & \sum_{a} \varphi^{w}(a) L(a) U(a) + \sum_{k} \varphi^{f}(k) J(k) \Pi(k) \\
& - \sum_{c} \sum_{k} \sum_{a} W_{kc}^{*}(a) D_{kc}(a) \left[\log \left(\frac{D_{kc}(a)}{L(a)} \right) \right. \\
& \qquad \left. - \log \left(J_{c}(k) \cdot \left(c_{kc}(a)^{1-\eta} h_{kc}(a)^{\eta} \right)^{\beta_{w}/\rho_{w}} \cdot U_{c}(a)^{\beta_{w}-\beta_{w}/\rho_{w}} \cdot \left(\sum_{c} U_{c}(a)^{\beta_{w}} \right)^{-1} \right) \right] \\
& - \sum_{c} \sum_{k} \pi_{c}^{*}(k) J_{c}(k) \left[\log \left(\frac{J_{c}(k)}{J(k)} \right) - \log \left(c_{c}(k)^{\beta_{f}} \cdot \left(\sum_{c} c_{c}(k)^{\beta_{f}} \right)^{-1} \right) \right] \\
& - \sum_{c} R_{c}^{*} \left[\sum_{k} \sum_{a} D_{kc}(a) h_{kc}(a) + \sum_{k} J_{c}(k) h_{c}(k) - \bar{H}_{c} I_{c}^{1/(1+\gamma_{c})} \right] \\
& - P^{*} \sum_{c} \left(\sum_{k} \left[\sum_{a} D_{kc}(a) c_{kc}(a) + J_{c}(k) c_{c}(k) \right] + Y_{c} - \sum_{k} J_{c}(k) \left(\sum_{a} \frac{D_{kc}(a)}{J_{c}(k)} f_{c}(k,a) \right)^{1-\alpha} h_{c}(k)^{\alpha} \right) \\
& - ... \qquad \text{(C.12)}
\end{aligned}$$

where $W_{kc}^{*}(a)$, $\pi_{c}^{*}(k)$, P^{*}, and R_{c}^{*} are the multipliers associated the constraints (C.7), (C.8), (C.10) and (C.11), respectively. I omit the terms for the non-negativity constraints and focus on the interior solution.

Towards characterizing the planner's solution, from the FOCs with respect to $c_{kc}(a)$ and $h_{kc}(a)$, we have

$$\varphi^{w}(a) - \left(\frac{\beta_{w}}{\rho_{w}} - \beta_{w} \right) \sum_{k'} W_{k'c}^{*}(a) \frac{D_{k'c}(a)}{L_{c}(a)} - \beta_{w} \sum_{c'} \sum_{k'} W_{k'c'}^{*}(a) \frac{D_{k'c'}(a)}{L(a)} = x_{kc}(a) - \frac{\beta_{w}}{\rho_{w}} W_{kc}^{*}(a), \qquad \text{(C.13)}$$

where $x_{kc}^{*}(a) \equiv P^{*} c_{kc}(a) + R_{c}^{*} h_{kc}(a)$ and

$$\eta P^{*} c_{kc}(a) = (1-\eta) R_{c}^{*} h_{kc}(a). \qquad \text{(C.14)}$$

Then, from the FOC with respect to $D_{kc}(a)$, we have

$$W_{kc}^{*}(a) = -x_{kc}^{*}(a) + P^{*}(1-\alpha) \cdot \alpha^{\frac{\alpha}{1-\alpha}} R_{c}^{*\frac{\alpha}{\alpha-1}} \left[f_{c}(k,a) + \frac{\mu}{L_{c}} \sum_{k'} \sum_{a'} D_{k'c}(a') f_{c}(k',a') \right]. \qquad \text{(C.15)}$$

I can now obtain the optimality condition for worker consumptions

$$x_{kc}^{*}(a)=\frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}\Big(\tilde{W}_{kc}(a)-\mathcal{O}_c^w(a)\Big)+\frac{1}{1+\beta_w/\rho_w}\varphi^w(a) \tag{C.16}$$

where $\tilde{W}_{kc}(a)\equiv P^*(1-\alpha)\cdot\alpha^{\frac{\alpha}{1-\alpha}}R_c^{*\frac{\alpha}{\alpha-1}}\left[f_c(k,a)+\frac{\mu}{L_c}\sum_k\sum_a D_{kc}(a)\,f_c(k,a)\right]$ is the social value of a skill-a worker in type-k firm and city c, which comprises her marginal product of labor and the agglomeration spillovers generated for all workers in the city; $\mathcal{O}_c^w(a)\equiv(1-\rho_w)\sum_{k'}W_{k'c}^*(a)\frac{D_{k'c}(a)}{L_c(a)}+\rho_w\sum_{c'}\sum_{k'}W_{k'c'}^*(a)\frac{D_{k'c'}(a)}{L(a)}$ is the opportunity cost of a skill-a worker in city c, which is a weighted average of the worker's average shadow value in city c, $\sum_{k'}W_{k'c}^*(a)\frac{D_{k'c}(a)}{L_c(a)}$, and the average shadow value in the entire economy, $\sum_{c'}\sum_{k'}W_{k'c'}^*(a)\frac{D_{k'c'}(a)}{L(a)}$. To simplify notation, I define $T_c^w(a)\equiv-\frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}\mathcal{O}_c^w(a)+\frac{1}{1+\beta_w/\rho_w}\varphi^w(a)$ and rewrite the optimal condition as

$$x_{kc}^{*}(a)=\frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}\tilde{W}_{kc}(a)+T_c^w(a). \tag{C.17}$$

Similarly, I can obtain the optimality condition for entrepreneur consumptions as

$$x_c^*(k)=\frac{\beta_f}{1+\beta_f}\tilde{\pi}_c(k)+T^f(k). \tag{C.18}$$

where $x_c^*(k)\equiv P^*c_c(k)$ and $\tilde{\pi}_c(k)$ is the social value of a firm of type-k in city c, $T^f(k)\equiv-\frac{\beta_f}{1+\beta_f}\mathcal{O}^f(k)+\frac{1}{1+\beta_f}\varphi^f(k)$, and $\mathcal{O}^f(k)\equiv\sum_c\pi_c^*(k)\frac{J_c(k)}{J(k)}$ is the opportunity cost of a type-k firm. The social value $\tilde{\pi}_c(k)$ can be expressed as

$$\begin{aligned}\tilde{\pi}_c(k)=\sum_a W_{kc}^*(a)\frac{D_{kc}(a)}{J_c(k)}&-(1-\rho_w)\sum_a\bar{W}_c^*(a)\frac{D_{kc}(a)}{J_c(k)}\\&-\rho_w\sum_a\bar{W}^*(a)\frac{D_{kc}(a)}{J_c(k)}+\frac{\rho_w}{\beta_w}\sum_a\varphi^w(a)\frac{D_{kc}(a)}{J_c(k)},\end{aligned} \tag{C.19}$$

where $\bar{W}_c^*(a)\equiv\sum_k W_{kc}^*(a)\frac{D_{kc}(a)}{L_c(a)}$ and $\bar{W}^*(a)\equiv\sum_c\sum_k W_{kc}^*(a)\frac{D_{kc}(a)}{L(a)}$ are the average shadow value of the skill-a workers in city c and in the economy.

□

The last condition is the one with respect to Y_c:

$$R_c^*\overline{H}_c\frac{1}{1+\gamma_c}Y_c^{\frac{\gamma_c}{1+\gamma_c}}=P^*,\forall c. \tag{C.20}$$

C.3 Optimal spatial policy

Proof of Proposition 2. The optimal spatial policy consist of (1) a set of labor income taxes $\{t^w_{kc}(a)\}_{\forall a,k,c}$ that are specific to worker skill a, firm type k and city c, and (2) a set of corporate profit taxes $\{t^f_c(k)\}_{\forall k,c}$ that are specific to firm type k and city c. Specifically, the labor income taxes and the corporate income taxes are set as

$$t^w_{kc}(a) = -\frac{\frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}\tilde{W}_{kc}(a) + T^w_c(a) - W_{kc}(a)}{W_{kc}(a)}, \quad \forall a,k,c$$
$$t^f_c(k) = -\frac{\frac{\beta_f}{1+\beta_f}\tilde{\pi}_c(k) + T^f(k) - \pi_c(k)}{\pi_c(k)}, \quad \forall k,c$$

where $\{W_{kc}(a)\}_{\forall a,k,c}$ and $\{\pi_c(k)\}_{\forall k,c}$ are wages and profits that workers and firms earn in a competitive equilibrium; the other terms in the taxes include the social values and transfers that have been derived above. Hence, the total after-tax income of workers are firms are $I^w_{kc}(a) = (1 - t^w_{kc}(a))W_{kc}(a)$ and $I^f_c(k) = \left(1 - t^f_c(k)\right)\pi_{kc}(a)$, respectively. I assume that the housing developers' profits are rebated to workers proportional to $I^w_{kc}(a)$, which is nondistortionary. I now show that the government's budget balances with the specified taxes, that is

$$-\sum_c\sum_k\sum_a D_{kc}(a)\left(t^w_{kc}(a)W_{kc}(a)\right) - \sum_c\sum_k J_c(k)\left(t^f_c(k)\pi_c(k)\right) = 0, \tag{C.21}$$

or equivalently

$$\sum_c\sum_k\left(\sum_a D_{kc}(a)T^w_c(a) + J_c(k)T^f(k)\right) = \sum_c\sum_k\left(\sum_a D_{kc}(a)\left(W_{kc}(a) - \tilde{W}_{kc}(a)\right) + J_c(k)\left(\pi_c(k) - \tilde{\pi}_c(k)\right)\right). \tag{C.22}$$

Note that

$$\sum_c\sum_k\left(\sum_a \frac{1}{\tau}I^w_{kc}(a)D_{kc}(a) + I^f_c(k)J_c(k)\right) = \sum_c\sum_k\left(\sum_a W_{kc}(a)D_{kc}(a) + \pi_c(k)J_c(k)\right).$$

Defining $\mathcal{T} \equiv \sum_c \sum_k \sum_a D_{kc}(a) T_c^w(a) + \sum_c \sum_k J_c(k) T^f(k)$, we then have

$$\begin{aligned}\mathcal{T} &= -\sum_c \sum_k \left(\sum_a \tilde{W}_{kc}(a) D_{kc}(a) + \tilde{\pi}_c(k) J_c(k) \right) + \sum_c \sum_k \left(\sum_a W_{kc}(a) D_{kc}(a) + \pi_c(k) J_c(k) \right) \\ &= \sum_c \sum_k \sum_a \left(W_{kc}(a) - \tilde{W}_{kc}(a) \right) D_{kc}(a) + \sum_{c'} \sum_{k'} \left(\pi_c(k) - \tilde{\pi}_c(k) \right) J_c(k) .\end{aligned}$$

Hence, I have proved that the government budget is balanced as shown in equation (C.22). If the planner's problem is globally concave, then the equilibrium under the specified taxes implements the efficient allocation. Specifically, the after-tax budget constraints for workers are identical to those in the planner's problem (C.17). Land rents are redistributed in proportion to after-tax income, which does not distort workers' location or firm choices. Workers' consumption decisions over the final good and housing, as implied by equation (2.1), match the planner's optimality conditions (C.14). Moreover, workers' between-city location choices and within-city firm choices—given after-tax income—coincide with the planner's optimal allocation conditions (C.6) and (C.5). Similarly, the after-tax budget constraints for entrepreneurs match those in the planner's problem (C.18), and the resulting firm location decisions align with the optimal firm allocation condition (C.8). The first-order condition for the final good used by developers implied by (B.8), is identical to that of the social planner implied by (C.12). Finally, the market-clearing conditions for the final good and housing (B.12) and (B.10) are equivalent to the resource constraints in the planner's problem (C.10) and (C.11).

□

C.4 Lump-sum redistribution

The optimal policy generates heterogeneous welfare changes across different workers and firms, largely driven by unobserved welfare weights. To isolate the efficiency gain of the optimal policy, I design a set of type-specific lump-sum transfers that equalize the welfare gains of all agents, a scheme similar to Berger et al. (2025). The design consists of two steps.

First, remove the optimal taxes $\{t_{kc}^w(a)\}_{k,c,a}$ and $\{t_c^f(k)\}_{c,k}$ while maintaining the efficient spatial allocation. Allow workers and firms to rematch within each city,[31] and solve for the equilibrium housing rent in each city. Second, with spatial allocation and within-city matching fixed, determine type-specific lump-sum transfers $\tau^w(a)$ and $\tau^f(k)$ so that all agent types experience the same welfare gain relative to the laissez-faire equilibrium.

[31] Although the optimal policy affects within-city matching through the transfer terms $T_c^w(a)$, within-city matching is efficient, and the transfers serve solely redistributive purposes.

Appendix D Uniqueness

In Proposition 4 below, I provide a sufficient condition for the existence of a unique equilibrium of the spatial economy. To isolate the impact of two-sided sorting on equilibrium multiplicity, I set the agglomeration elasticity, μ, and the housing related parameters, α and η, to zero.

Proposition 4. *Suppose $\alpha = \eta = \mu = 0$, there exists a unique equilibrium when $\rho_w < \frac{1}{2}$ and $\beta_f < \frac{1}{4}$.*

Proof. For simplicity, I assume a discrete set of worker and firm types in this proof. Denote worker type as $i \in \mathcal{I}$ and firm type as $k \in \mathcal{K}$. Let $\vec{\sigma}$ be the vector of the logarithm of the spatial allocations of workers and firms, i.e. $\sigma = \left[\{\log L_c(i)\}_{\forall c,i}, \{\log J_c(k)\}_{\forall c,k}\right]$. The equilibrium allocation is the solution to the set of functions

$$\vec{f}(\vec{\sigma}; \mathcal{P}) = \vec{\sigma} - \vec{g}(\vec{\sigma}; \mathcal{P}).$$

where $\mathcal{P}$ is the set of model parameters and $\vec{g}$ is a set of decision rules of location choices. Specifically, the decision rules are

$$\log L_c(i) = \log\left(L(i) \cdot \frac{\left[R_c(i) \cdot \left(\sum_k J_c(k) \cdot (G_{kc}(i) W_{kc}(i))^{\beta_w/\rho_w}\right)^{\rho_w/\rho_w}\right]^{\beta_w}}{\sum_{c'}\left[R_{c'}(i) \cdot \left(\sum_k J_{c'}(k) \cdot (G_{kc'}(i) W_{kc'}(i))^{\beta_w/\rho_w}\right)^{\rho_w/\rho_w}\right]^{\beta_w}}\right)$$

for workers and

$$\log J_c(k) = \log\left(J(k) \cdot \frac{(\sum_i D_{kc}(i) W_{kc}(i))^{\beta_f}}{\sum_{c'}(\sum_i D_{kc'}(i) W_{kc'}(i))^{\beta_f}}\right)$$

for firms, where

$$\log D_{kc}(i) = \log\left(L_c(a) \cdot \frac{(G_{kc}(i) W_{kc}(i))^{\beta_w/\rho_w}}{\sum_k J_c(k) \cdot (G_{kc}(i) W_{kc}(i))^{\beta_w/\rho_w}}\right).$$

When the Jacobian matrix is strictly diagonally dominant, the system of equations has a unique solution and can be solved using the Jacobi method. To establish such a property, start from the derivatives of the worker allocation $L_c(i), \forall c, i$:

$$\frac{\partial \log L_c(i)}{\partial \log J_c(k)} = \rho_w \cdot s_{kc}(i), \forall k; \quad \frac{\partial \log L_c(i)}{\partial \log L_{c'}(i')} = 0, \forall c', i'$$

and the derivatives of the firm allocation $J_c(k), \forall c, k$:

$$\frac{\partial \log J_c(k)}{\partial \log L_c(i)} = \beta_f \cdot s_{kc}(i); \quad \frac{\partial \log J_c(k)}{\partial \log J_c(k')} = -\beta_f \cdot s_c(k'), \forall k'.$$

where $s_{kc}(i)$ is the type-i worker wage bill share for a type-k firm in city c, i.e. $s_{kc}(i) = W_{kc}(i) D_{kc}(i) / \sum_i W_{kc}(i) D_{kc}(i)$, $s_c(k)$ is the wage bill share of a type-k firm in city c, i.e. $s_c(k) = \sum_i W_{kc}(i) D_{kc}(i) / (\sum_k J_c(k) \sum_i W_{kc}(i) D_{kc}(i))$.

For the Jacobian matrix to be strictly diagonally dominant, we want to establish two sets of conditions: one set for the workers and one set for the firms. For the conditions with respect to workers, it requires

$$1 > \sum_{c'} \sum_{k'} \left| \frac{\partial \log L_c(i)}{\partial \log J_{c'}(k')} \right| + \sum_{c'} \sum_{i'} 1\{c \neq c' | i \neq i'\} \left| \frac{\partial \log L_c(i)}{\partial \log L_{c'}(i')} \right|, \forall c, i. \tag{D.1}$$

It can be shown that (1) $\sum_{k'} \left| \frac{\partial \log L_c(i)}{\partial \log J_c(k')} \right| = \rho_w$; (2) $\frac{\partial \log L_c(i)}{\partial \log J_{c'}(k')} < 0, \forall c' \neq c, k'$; (3) $\sum_{c'} \sum_{k'} \frac{\partial \log L_c(i)}{\partial \log J_{c'}(k')} = 0$ because increasing the number of firms by the same proportion everywhere has no effect on workers' location choice probabilities; (4) $\frac{\partial \log L_c(i)}{\partial \log L_{c'}(i')} = 0, \forall c', i'$. Hence, the diagonal dominance condition for workers is satisfied when $\rho_w < \frac{1}{2}$.

For the conditions with respect to firms, it requires

$$1 + \beta_f \cdot s_c(k) > \sum_{c'} \sum_{k'} 1\{c \neq c' | k \neq k'\} \left| \frac{\partial \log J_c(k)}{\partial \log J_{c'}(k')} \right| + \sum_{c'} \sum_{i} \left| \frac{\partial \log J_c(k)}{\partial \log L_{c'}(i)} \right|, \forall c, k. \tag{D.2}$$

It can be shown that (1) and $\sum_k \left| \frac{\partial \log J_c(k)}{\partial \log J_c(k')} \right| = \beta_f$; (2) $\sum_i \frac{\partial \log J_c(k)}{\partial \log L_c(i)} = \beta_f$; (3) $\sum_{c'} \sum_{k'} \frac{\partial \log J_c(k)}{\partial \log J_{c'}(k')} = \sum_{c'} \sum_i \frac{\partial \log J_c(k)}{\partial \log L_c(i)} = 0$; (4) $\frac{\partial \log J_c(k)}{\partial \log J_{c'}(k')} > 0, \forall c' \neq c, k'$ and $\frac{\partial \log J_c(k)}{\partial \log L_{c'}(i)} < 0, \forall c' \neq c, i$. Hence, the condition can be re-written as

$$1 + \beta_f \cdot s_c(k) > \sum_{k' \neq k} \left| \frac{\partial \log J_c(k)}{\partial \log J_c(k')} \right| + \sum_{c' \neq c} \sum_{k'} \left| \frac{\partial \log J_c(k)}{\partial \log J_{c'}(k')} \right| + \sum_{c'} \sum_{i} \left| \frac{\partial \log J_c(k)}{\partial \log L_{c'}(i)} \right|, \forall c, k.$$

$$1 > 4\beta_f - 2\beta_f s_c(k), \forall c, k,$$

which holds when $\beta_f < \frac{1}{4}$, as $s_c(k) > 0$. □

Appendix E Identification

E.1 Worker sorting elasticities

Here, I show how to obtain the two wage terms $\hat{W}_{it}$, $\bar{W}_{ct}$ and two wage bill terms $\hat{E}_{jt}$ and $\bar{E}_{ct}$ for constructing the passthrough equations. Start with the wage equation with the time-varying productivity shocks:

$$W_{ijt}(a) = \chi \cdot r_{ct}^{\frac{\alpha}{\alpha-1}} \cdot A_c \hat{A}_{ct} L_{ct}^{\mu} \cdot z_j \hat{z}_{jt} \cdot a_i^{\theta_j} \hat{a}_{it}. \tag{E.1}$$

Wage changes for the workers who do not move can be due to 1) city-level shocks due to changes in housing rent, agglomeration spillovers, and productivity shocks, 2) firm-level productivity shocks, and 3) worker skill transient shocks. In what follows, I construct the average city wage $\bar{W}_{ct}$ of stayers to isolate the city-level shocks and the demeaned wage $\hat{W}_{it}$ to isolate firm-level shocks. The mean and demeaned wage bill terms are constructed accordingly. First, with equation (E.1) the average wage for stayers is given by

$$\begin{aligned} \bar{W}_{ct} &= \mathbb{E}_{S(i)=1}\left[\chi \cdot r_{ct}^{\frac{\alpha}{\alpha-1}} \cdot A_c \hat{A}_{ct} L_{ct}^{\mu} \cdot z_j \hat{z}_{jt} \cdot a_i^{\theta_j} \hat{a}_{it}\right] \\ &= \chi \cdot r_{ct}^{\frac{\alpha}{\alpha-1}} \cdot A_c \hat{A}_{ct} L_{ct}^{\mu} \cdot \bar{\Xi}_c \end{aligned} \tag{E.2}$$

where $\bar{\Xi}_c \equiv \mathbb{E}_{S(i)=1}\left[z_j \cdot a_i^{\theta_j}\right]$ denotes the city wage index for worker stayers. This measure is time-invariant, as the stayer sample consists of a fixed set of workers and firms. Firm productivity and transient skill shocks average out in the cross-section within each city. The demeaned wage is constructed as:

$$\hat{W}_{ijt} = W_{ijt}/\bar{W}_{ct} = (\bar{\Xi}_c)^{-1} \cdot z_j \hat{z}_{jt} \cdot a_i^{\theta_j} \hat{a}_{it}. \tag{E.3}$$

For the mean and demeaned wage bill terms, start with firm wage bill:

$$\begin{aligned} E_{jct} &= \chi \left(A_c \hat{A}_{ct} L_{ct}^{\mu} r_{ct}^{\frac{\alpha}{\alpha-1}}\right) \cdot (z_j \hat{z}_{jt})^{1+\beta_w/\rho_w} \cdot \phi_{jct} \\ &= \left(\chi r_{ct}^{\frac{\alpha}{\alpha-1}} A_c \hat{A}_{ct} L_{ct}^{\mu} \cdot z_j \hat{z}_{jt}\right)^{1+\beta_w/\rho_w} \int_a \left(a^{\theta_j}\right)^{1+\beta_w/\rho_w} \hat{a}_{it} \hat{\kappa}_{jct}(a) da \end{aligned} \tag{E.4}$$

where $\hat{\kappa}_{jct}(a)$ is the firm-skill-time-specific labor supply curve shifter

$$\hat{\kappa}_{jct}(a) = \left(A_c \hat{A}_{ct} L_{ct}^{\mu} r_{ct}^{\frac{\alpha}{\alpha-1}}\right)^{\beta_w - \beta_w/\rho_w} \cdot (r_{ct})^{-\eta\beta_w} \left(G_{jc}(a)\right)^{\frac{\beta_w}{\rho_w}} \cdot \hat{\kappa}_{ct}(a) \tag{E.5}$$

and $\hat{\kappa}_{ct}(a)$ is a city-skill-specific labor supply shifter

$$\hat{\kappa}_{ct}(a) = \frac{L(a) \cdot \left[R_c(a) \left(\sum_{j \in \mathcal{J}_{ct}} \left(\Xi_c \cdot z_j a^{\theta_j} \cdot G_{jc}(a) \right)^{\beta_w/\rho_w} \right)^{\rho_w/\beta_w} \right]^{\beta_w}}{\sum_{c'} \left[R_{c'}(a) r_{c't}^{-\eta} \left(\sum_{j' \in \mathcal{J}_{c't}} \left(W_{j'c't}(a) G_{j'c}(a) \right)^{\beta_w/\rho_w} \right)^{\rho_w/\beta_w} \right]^{\beta_w} \cdot \sum_{j' \in \mathcal{J}_{ct}} \left(\Xi_c \cdot z_{j'} a^{\theta_{j'}} \cdot G_{j'c}(a) \right)^{\frac{\beta_w}{\rho_w}}}. \tag{E.6}$$

Using the expression for $\hat{\kappa}_{ct}(a)$, the firm wage bill E_{jct} can be rewritten as:

$$E_{jct} = \left(r_{ct}^{\frac{\alpha}{\alpha-1}} A_c \hat{A}_{ct} L_{ct}^{\mu} \cdot z_j \hat{z}_{jt} \right)^{1+\beta_w} \cdot \left(\chi z_j \hat{z}_{jt} \right)^{1+\beta_w/\rho_w} \cdot \left(r_{ct} \right)^{-\eta\beta_w} \hat{\phi}_{jct} \tag{E.7}$$

where $\hat{\phi}_{jct} \equiv \int_a \left(a^{\theta_j} \right)^{1+\beta_w/\rho_w} \left(G_{jc}(a) \right)^{\frac{\beta_w}{\rho_w}} \cdot \hat{\kappa}_{ct}(a) da$. Lastly, $\bar{E}_{ct}$, which is average wage bill of firms in the stayers sample, can be obtained as

$$\begin{aligned} \bar{E}_{ct} &= \frac{1}{|\mathcal{J}_c^S|} \sum_{j \in \mathcal{J}_c^S} E_{jct} \\ &= \frac{\chi^{1+\beta_w/\rho_w} \cdot \left(A_c \hat{A}_{ct} L_{ct}^{\mu} r_{ct}^{\frac{\alpha}{\alpha-1}} \right)^{1+\beta_w} \left(r_{ct} \right)^{-\eta\beta_w}}{|\mathcal{J}_c^S|} \cdot \hat{\phi}_{ct} \end{aligned} \tag{E.8}$$

where $\mathcal{J}_c^S$ as the set of firms in the stayer sample of city c and $\hat{\phi}_{ct} \equiv \sum_{j \in \mathcal{J}_c^S} \left(z_j \right)^{1+\beta_w/\rho_w} \cdot \hat{\phi}_{jct}$. I then construct the demeaned firm wage bill as:

$$\hat{E}_{jt} = \frac{E_{jt}}{\bar{E}_{ct}} = |\mathcal{J}_c^S| \cdot \left(z_j \hat{z}_{jt} \right)^{1+\beta_w/\rho_w} \cdot \frac{\hat{\phi}_{jct}}{\hat{\phi}_{ct}}. \tag{E.9}$$

I can now derive the passthrough regressions using (E.2), (E.3), (E.8) and (E.9). Taking log and first differences of (E.3) and (E.9), the net passthrough equation with the measurement error is:

$$\Delta \log \hat{W}_{ijt} = \delta_w \Delta \log \hat{\dot{E}}_{jt} + \Delta \hat{a}_{it} + \delta_w \left(\Delta e_{jt} - \Delta \log \frac{\hat{\phi}_{jct}}{\hat{\phi}_{ct}} \right), \tag{E.10}$$

where $\delta_w \equiv \frac{1}{1+\beta_w/\rho_w}$. The potential correlations of the net wage bill shock with the measurement error Δe_{jt} and with the relative wage bill shifter $\Delta \log(\hat{\phi}_{jct}/\hat{\phi}_{ct})$ give rise to two endogeneity concerns.

To deal with the first concern, I follow Lamadon et al. (2022) to instrument the net wage bill shock $\log \Delta \hat{\dot{E}}_{jt}$ with its lags before year $t-q-1$. These lagged shocks are correlated with the current shock because the productivity shock $\hat{z}_{jt}$, assumed to follow a Markov process, is persistent. In addition, changes in the measurement error, Δe_{jt}, depends only on

Table E.1: Passthrough of wage bill shocks to workers

	Firm Passthrough			City Passthrough			
	OLS	IV	IV	OLS	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Firm-level shock	0.27***	0.13***	0.14***				
	(0.00)	(0.02)	(0.02)				
City-level shock				0.28***	0.30***	0.25***	0.32***
				(0.00)	(0.00)	(0.00)	(0.00)
Lag changes			Yes			Yes	Yes
Changes in skill shares							Yes
R^2	0.01	0.01	0.10	0.31	0.31	0.41	0.42
Num. obs.				2534400			
F statistic (First Stage)		4466.9	3924.5		159646.2	148297.0	102495.5

Notes: This table reports estimation results of passthrough regressions, which are estimated using the 2010-2017 stayers sample. Columns (1) and (4) report the OLS estimates. Columns (2) and (5) report the IV estimates. The instruments are quadratic polynomials of three- to five-period lagged changes in log residualized/mean wage bill and log residualized/mean worker earnings. Columns (3) and (6) further control for changes in labor market competition by including a cubic polynomial of one-period lagged changes in log residualized/mean wage bill and log residualized/mean worker earnings in the regressions. Column (7) controls for city-level changes in the share of high-skilled workers as a proxy for changes in endogenous amenities. The number of observations is rounded to the nearest hundred. *p<0.1; **p<0.05; ***p<0.01.

measurement error shocks u^e_{jt} in $\{t-q-1,..,t\}$. Thus, lagged residualized firm wage bill shocks before year $t-q-1$ are orthogonal to Δe_{jt}, making them valid instruments.

To deal with the second concern, I use the control function approach. Specifically, I assume $\log(\hat{\phi}_{jct}/\hat{\phi}_{ct}) = F^N\left(\log(\hat{\phi}_{jct-1}/\hat{\phi}_{ct-1})\right)+\epsilon^N_{jct} = H^N(\log \hat{W}_{ijt-1}, \log \hat{\dot{E}}_{jt-1})+\epsilon^N_{jct}$, where ϵ^N_{jct} is an $i.i.d.$ innovation. The first part of the equation relates the current relative shifter with its lag, as all the shocks that generate changes in $\log(\hat{\phi}_{jct}/\hat{\phi}_{ct})$ follow Markov processes; the second half says the unobserved $\log(\hat{\phi}_{jct-1}/\hat{\phi}_{ct-1})$ can be inferred from $\log \hat{W}_{ijt-1}$ and $\log \hat{\dot{E}}_{jt-1}$, using equation (E.10) at $t-1$. When $F^N(\cdot)$ is a linear function, the first difference of the relative shifter can be written as $\Delta \log(\hat{\phi}_{jct}/\hat{\phi}_{ct}) = F^N\left(\Delta \log(\hat{\phi}_{jct-1}/\hat{\phi}_{ct-1})\right)+\Delta\epsilon^N_{jct} = H^N(\Delta \log \hat{W}_{ijt-1}, \Delta \log \hat{\dot{E}}_{jt-1})+\Delta\epsilon^N_{jct}$. In practice, $H^N(\cdot,\cdot)$ is specified as a cubic polynomial.

Similarly, using equations (E.2) and (E.8) and defining the city passthrough parameter as $\delta_c \equiv \frac{1}{1+\beta_w}$, I have the mean passthrough equation as:

$$\Delta \log\left(\frac{\bar{W}_{ct}}{r^\eta_{ct}}\right) = \delta_c \Delta \log\left(\frac{\bar{\dot{E}}_{ct}}{r^\eta_{ct}}\right) + \delta_c\left(\Delta e_{ct} - \Delta \log \hat{\phi}_{ct}\right). \tag{E.11}$$

The city-level labor supply elasticity is identified as the passthrough of *real* mean wage bill shocks to *real* mean wages. The change in log city-level wage bill shifter enters the equation as a structural residual, along with the measurement error in $\bar{\dot{E}}_{ct}$. I assume this error term, e_{ct}, follows the same $MA(q)$ process as the firm-level error e_{jt}. Following the identification strategy of the net passthrough equation, I instrument $\Delta \log \left(\bar{\dot{E}}_{ct}/r_{ct}^{\eta}\right)$ using lags dated prior to year $t-q-1$, and control nonparametrically for two lagged differences: $\Delta \log \left(\bar{W}_{ct-1}/r_{ct-1}^{\eta}\right)$ and $\Delta \log \left(\bar{\dot{E}}_{ct-1}/r_{ct-1}^{\eta}\right)$. Table E.1 shows the estimation results of equations (E.10) and (E.11) under different specifications.

E.2 Firm sorting elasticity

The method of identifying the firm sorting elasticity β_f is similar to the worker elasticities. First, I have mean firm wage bill for stayers firms in city c, year t from equation (E.8) as:

$$\bar{E}_{ct} = \frac{\chi^{1+\beta_w/\rho_w}\left(A_c \hat{A}_{ct} L_{ct}^{\mu} r_{ct}^{\frac{\alpha}{\alpha-1}}\right)^{1+\beta_w} (r_{ct})^{-\eta\beta_w}}{|\mathcal{J}_c^S|} \cdot \hat{\phi}_{ct}. \tag{E.12}$$

Then, I derive the total firm wage bill in city c, year t as

$$\begin{aligned} E_{ct} &= \sum_k J_{kct} \cdot \bar{E}_{kct} \\ &= \left(\chi \cdot r_{ct}^{\frac{\alpha}{\alpha-1}} A_c \hat{A}_{ct} L_{ct}^{\mu}\right)^{(1+\beta_f)(1+\beta_w/\rho_w)} (r_{ct})^{-\eta\beta_w(1+\beta_f)} \cdot \hat{\Phi}_{ct} \end{aligned} \tag{E.13}$$

where $\hat{\Phi}_{ct} \equiv \sum_k \left(\hat{\phi}_{kct} z_k^{1+\beta_w/\rho_w}\right)^{1+\beta_f} J_k \cdot \frac{1}{\sum_c \bar{E}_{kct}^{\beta_f}}$. The equation above makes uses of two properties of the model. First, firm profit is proportional to wage bill. Second, firms in the same cluster have the same probability of choosing a city.

Combining equations (E.12) and (E.13) and introducing a measurement error of $\bar{E}_{ct}$, I can obtain the following passthrough equation for firms:

$$\Delta \log \bar{\dot{E}}_{ct} = \delta_f \Delta \log \dot{E}_{ct} + \delta_f \left(\Delta e_{ct} - \Delta \log \frac{\hat{\phi}_{ct}}{\hat{\Phi}_{ct}}\right). \tag{E.14}$$

As before, there are endogeneity concerns due to potential correlation between $\Delta \log \dot{E}_{ct}$ and Δe_{ct} and $\Delta \log(\hat{\phi}_{ct}/\hat{\Phi}_{ct})$. I follow the same strategy here as the worker passthrough regressions, in that I instrument $\Delta \log \dot{E}_{ct}$ with its lags before year $t-q-1$ and control for the two lagged terms $\Delta \log \bar{\dot{E}}_{ct-1}$ and $\Delta \log \dot{E}_{ct-1}$. Table E.2 shows the results.

Table E.2: Passthrough of wage bill shocks to firms

	OLS	IV	IV
	(1)	(2)	(3)
City total wage bill shock	1.06***	0.21***	0.15***
	(0.00)	(0.01)	(0.02)
Control			Yes
R^2	0.31	−0.01	0.24
Num. obs.		94100	
F statistic (First Stage)		2237.8	1450.8

Notes: This table reports estimation results of passthrough of city total wage bill shocks to firms. Column (1) is the OLS estimate and Columns (2)-(3) are the IV estimates. The instruments are quadratic polynomials of three- to five-period lagged changes in the log total city wage bill. Column (3) controls for a cubic polynomial of one-period lagged changes in log city and firm mean wage bill. The number of observations is rounded to the nearest hundred. *p<0.1; **p<0.05; ***p<0.01.

E.3 Worker skill and firm productivity

I show here how I construct the adjusted log earnings $\log \check{W}_{ijt}(a)$ for each individual, which relates only to the worker's permanent skill a, transient skill shock $\hat{a}$ and firm's productivity parameters (z, θ). I do so by partialling out the time-varying firm and city productivity shocks from log earnings $\log W_{ijt}$. First, to isolate the firm-productivity shock, I construct a cluster-level mean firm wage bill and cluster-residualized firm wage bill as:

$$\begin{aligned}\bar{E}_{kct} &= \frac{1}{|\mathcal{J}^S_{kc}|} \cdot \sum_{j \in \mathcal{J}^S_{kc}} E_{jct} \\ &= \frac{\chi^{1+\beta_w/\rho_w} \cdot \left(A_c \hat{A}_{ct} L^{\mu}_{ct} r_{ct}^{\frac{\alpha}{\alpha-1}}\right)^{1+\beta_w} (r_{ct})^{-\eta\beta_w}}{|\mathcal{J}^S_{kc}|} (z_k)^{1+\beta_w/\rho_w} \cdot \hat{\phi}_{kct}\end{aligned} \tag{E.15}$$

and

$$\hat{E}^{k(j)}_{jct} = \frac{E_{jct}}{\bar{E}_{kct}} = (\hat{z}_{jt})^{1+\beta_w/\rho_w}. \tag{E.16}$$

The city-level productivity shock can also be inferred from the mean log earnings of worker stayers, $\bar{W}^S_{ct}$. Define $\mathbb{A}_c \equiv A_c \bar{L}^{\mu}_c$. Combining equations (E.1), (E.16), and (E.2), and assuming that firm and city productivity shocks average to zero over time, I can construct:

$$\begin{aligned}\log \check{W}_{ijt}(a) &= \log W_{ijt} - \left(\delta_w \left(\log\left(\hat{E}^k_{jct}\right) - \log\left(\bar{\hat{E}}^k_{jc}\right)\right) + \left(\log \bar{W}^S_{ct} - \log \bar{W}^S_c\right)\right) - \frac{\alpha}{\alpha-1} \log \bar{r}_c - \log \mathbb{A}_c \\ &= \log z_{j(i,t)} + \theta_{j(i,t)} \log a_i + \hat{a}_{it}.\end{aligned} \tag{E.17}$$

E.4 Firm and city amenities

From workers' labor supply decisions, I have

$$\Lambda_{kc}(a) = \frac{\bar{J}_c(k)\left(z_k a^{\theta_k} G_k(a)\right)^{\frac{\beta_w}{\rho_w}}}{\sum_{j' \in \mathcal{J}_c}\left(z_{j'} a^{\theta'_j} G_{j'}(a)\right)^{\frac{\beta_w}{\rho_w}}} \tag{E.18}$$

$$\Lambda_c(a) = \frac{\left[R_c(a)\bar{r}_c^{-\eta} \cdot \mathbb{A}_c r_c^{\frac{\alpha}{\alpha-1}}\left(\sum_{j \in \mathcal{J}_c}\left(z_j a^{\theta_j} G_j(a)\right)^{\frac{\beta_w}{\rho_w}}\right)^{\rho_w/\beta_w}\right]^{\beta_w}}{\sum_{c'}\left[R_{c'}(a)\bar{r}_{c'}^{-\eta} \cdot \mathbb{A}_{c'} r_{c'}^{\frac{\alpha}{\alpha-1}}\left(\sum_{j' \in \mathcal{J}_{c'}}\left(z_{j'} a^{\theta_{j'}} G_{j'}(a)\right)^{\frac{\beta_w}{\rho_w}}\right)^{\rho_w/\beta_w}\right]^{\beta_w}}, \tag{E.19}$$

where $\Lambda_{kc}(a)$ is the share of skill-a workers in cluster-k firms, conditional on living in city c; $\Lambda_c(a)$ is the share of skill-a workers in city c. From the two equations, I can obtain

$$R_c(a) \cdot G_k(a) = \frac{\bar{r}_c^{\eta}}{\mathbb{A}_c r_c^{\frac{\alpha}{\alpha-1}} z_k a^{\theta_k}} \cdot \bar{J}_c(k)^{\frac{\beta_w}{\rho_w}} \cdot \Lambda_{kc}(a)^{\frac{\rho_w}{\beta_w}} \Lambda_c(a)^{\frac{1}{\beta_w}}, \ \forall a, k, c, \tag{E.20}$$

with one normalization imposed for every worker skill a as

$$\sum_c \left[R_c(a)\bar{r}_c^{-\eta} \cdot \mathbb{A}_c r_c^{\frac{\alpha}{\alpha-1}}\left(\sum_{j \in \mathcal{J}_c}\left(z_j a^{\theta_j} G_j(a)\right)^{\frac{\beta_w}{\rho_w}}\right)^{\rho_w/\beta_w}\right]^{\beta_w} = 1. \tag{E.21}$$

Figure E.1 displays the estimated firm and city amenities by worker skill decile.

E.5 City productivity

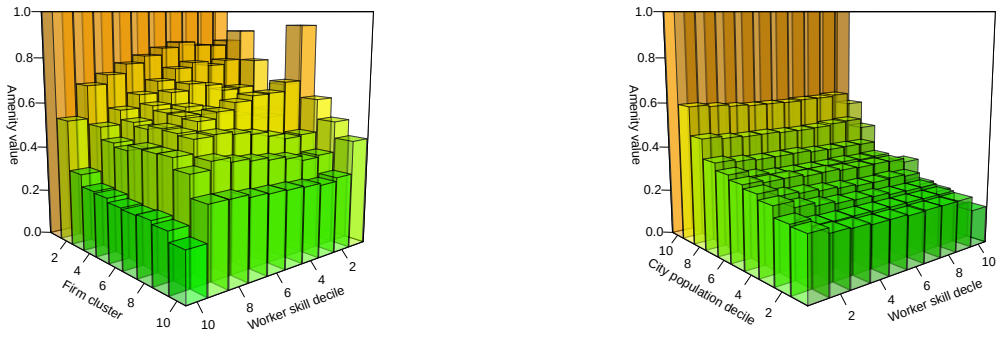
A cluster-k firm's average profits in city c is:

$$\bar{\pi}_c(k) = \Psi \cdot \left(\mathbb{A}_c \bar{r}_c^{\frac{\alpha}{\alpha-1}}\right) \cdot (z_k)^{1+\beta_w/\rho_w} \cdot \phi_{kc} \tag{E.22}$$

where

$$\phi_{kc} = \int_a \left(a^{\theta_k}\right)^{1+\beta_w/\rho_w} L_c(a) \cdot \frac{G_k(a)^{\beta_w/\rho_w}}{\sum_{j' \in \mathcal{J}_c}\left(z_{j'} a^{\theta_{j'}} \cdot G_{j'c}(a)\right)^{\frac{\beta_w}{\rho_w}}} da. \tag{E.23}$$

With the estimates of labor supply parameters, β_w, ρ_w, worker skill a, firm production parameters, $\{z_k, \theta_k\}_j$ and amenity parameters $\{G_k(a)\}_{k,a}$, I can obtain ϕ_{kc} for each city-cluster pair. To construct ϕ_{kc}, I use the average population and the set of firms that has operated in city c in the sample period. I can then identify the city composite productivities



(a) Firm amenities (b) City amenities

Figure E.1: Firm and city amenities

Notes: This set of figures presents skill-specific firm and city amenities. Workers are grouped into ten skill deciles, and cities are ranked by population and divided into ten equally sized groups. Amenities for firm cluster 1 and city group 10 are normalized to 1 across all skill deciles.

from the empirical firm sorting equation:

$$p_c(k) = \frac{\left(\left(\mathbb{A}_c \bar{r}_c^{\frac{\alpha}{\alpha-1}}\right) \cdot \phi_{kc}\right)^{\beta_f}}{\sum_{c'}\left(\left(\mathbb{A}_{c'} \bar{r}_{c'}^{\frac{\alpha}{\alpha-1}}\right) \cdot \phi_{kc'}\right)^{\beta_f}}, \tag{E.24}$$

where $p_c(k)$ is the average share of cluster-k firms in city c in the sample period. Given observed values of r_c and a calibrated α, equation (E.24) provides $K \times C$ firm sorting conditions to identify the C city-level composite productivity parameters, up to one normalization.

E.6 Agglomeration elasticity

Taking the log of the city composite productivity yields the estimating equation:

$$\log \mathbb{A}_c = A_0 + \mu \log \bar{L}_c + \epsilon_c^A \tag{E.25}$$

where A_0 is the intercept reflecting the normalization of exogenous city productivity parameters, and ϵ_c^A is the error term that captures the exogenous city productivity A_c. I construct the immigration-based instrument as:

$$IV_c^L = \sum_o \frac{IMM_{o,c,1980-2002}}{L_{c,2002}} \cdot \log I\hat{M}M_o \tag{E.26}$$

Table E.3: Housing Supply and Agglomeration Elasticity Estimates

Dependent variable:	$\Delta \log r_c$			$\log \mathbb{A}_c$	
	OLS	IV	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)
Changes in Log Housing Exp.	0.53***	0.45	0.47		
	(0.13)	(0.33)	(0.36)		
× Share of Undev. Land			−0.01		
			(0.01)		
Log Population				0.006	0.004
				(0.017)	(0.018)
R^2	0.30	0.29	0.29	0.00	0.00
Num. obs.			66		
F statistic (First Stage)		24.74	24.74		120.8

Notes: Columns (1)–(3) present estimates of housing supply elasticity. Column (1) reports OLS results, and Columns (2)–(3) show IV estimates using a Bartik-style instrument; Column (3) includes an interaction with the share of undevelopable land. Columns (4)–(5) report OLS and IV estimates of agglomeration elasticity. All regressions are weighted by city population. *p<0.1; **p<0.05; ***p<0.01.

where o denotes the origin country, $IMM_{o,c,1980-2002}$ is the number of immigrants from o arriving in city c from 1980 to 2002, and $\hat{IMM}_o$ is the total inflow from o to Canada since 2002. I construct the instrument using the 1980–2002 cumulative inflow relative to the city's 2002 population to proxy for baseline exposure, as IMDB data only track immigrants since 1980. The instrument captures exogenous variation in population inflows, with the assumption that $\hat{IMM}_o$ is orthogonal to city-level shocks. Table E.3 shows the results.

E.7 Housing supply parameters

From the housing market clearing condition, I obtain:

$$r_{ct}^{\frac{1+\gamma_c}{\gamma_c}} = \frac{1}{\bar{H}_c^0} \cdot \left(\tau\eta + \frac{\alpha}{1-\alpha} \frac{1+\beta_w/\rho_w}{\beta_w/\rho_w} \right) E_{ct} = \frac{1}{\bar{H}_c^0} \cdot EH_{ct} \tag{E.27}$$

where I define EH_{ct} as the total expenditure on housing in city c in year t. Taking the log and the difference over time yields the estimating equation:

$$\Delta \log r_{ct} = \Gamma_c \Delta \log EH_{ct} + \Delta e_{ct}^r \tag{E.28}$$

where $\Gamma \equiv \frac{\gamma_c}{1+\gamma_c}$ is the parameter of interest and e^r_{ct} represents the measurement error of observed housing rents. I allow the housing supply elasticity to be a function of the share of undevelopable land, denoted as $UNAVAL_c$, such that $\Gamma_c = \Gamma + \Gamma_L UNAVAL_c$. Changes in housing rent and expenditure may be due to unobserved labor and housing market shocks. To deal with the endogeneity concerns, I instrument $\Delta \log EH_{ct}$ using a shift-share bartik IV:

$$IV_c^{EH} = \sum_{ind} \frac{L_{c,ind,2002}}{L_{c,2002}} \times \Delta \log wage_{ind} \tag{E.29}$$

which is a base-year industry employment share weighted sum of national wage increases by industry. The identification assumption here is that industrial-level wage changes in the entire country are orthogonal to city-level unobserved labor and housing market shocks. Table E.3 shows the estimation results Γ_c. I then obtain $\gamma_c = \frac{\Gamma_c}{1-\Gamma_c}$ and choose $\bar{H}_c^0$ to match the average housing rent of each city.

Appendix F Model extensions

F.1 Endogenous amenity

The model can be extended to allow for endogenous amenity. Following Fajgelbaum and Gaubert (2020), I assume that skill-specific amenity $R_c(a)$ contains an exogenous part $\bar{R}_c(a)$ and an endogenous part determined by the local skill composition:

$$\begin{aligned} R_c(a) &= \bar{R}_c(a) \cdot L_c(H)^{\gamma_{HH}} L_c(L)^{\gamma_{LH}}, \ a \in H \\ R_c(a) &= \bar{R}_c(a) \cdot L_c(H)^{\gamma_{HL}} L_c(L)^{\gamma_{LL}}, \ a \in L \end{aligned} \tag{F.1}$$

where I define H and L as the high- and low-skilled groups depending on worker skill a, $L_c(H)$ and $L_c(L)$ are the numbers of high- and low-skilled workers in city c, and $\gamma_{gg'}$ is the amenity spillover elasticities from skill group g to g'.

F.2 Free entry

The model can be extended to have free entry of firms at the national level. Suppose each entrepreneur j has to pay a fixed cost c_e, denominated in the final good, to draw her production technology (z, θ) and amenities $\{G(a)\}_{\forall a}$ from a known distribution $\mathcal{H}$. Then,

the free entry condition can be written as

$$\mathbb{E}_{\mathcal{H}}\left(\Pi(j)\right) = \mathbb{E}_{\mathcal{H}}\left[\frac{1}{\beta_f}\log\left(\sum_c \pi_c(j)^{\beta_f}\right)\right] + \bar{C}^f = \log c_e, \tag{F.2}$$

where $\bar{C}^f = \gamma/\beta_f$ with γ being the Euler-Mascheroni constant.

F.3 Remote work

Assume that a share of workers can work remotely–that is, they can choose to work for a firm in city i while living in i'. The labor supply curve of remote workers can be written as

$$S_{jc}^R\left(W_{jc}(a), a\right) = L^R\left(a\right) \cdot \frac{\left(W_{jc}(a) \cdot G_{jc}\left(a\right)\right)^{\beta_w/\rho_w}}{\sum_c \sum_{j\in\mathcal{J}_c}\left(W_{jc}\left(a\right) \cdot G_{jc}\left(a\right)\right)^{\beta_w/\rho_w}},$$

where $L^R\left(a\right)$ is the measure of skill-$a$remote workers. The labor supply curve of on-site workers is

$$S_{jc}^O\left(W_{jc}(a), a\right) = L_c^O\left(a\right) \cdot \frac{\left(W_{jc}(a) \cdot G_{jc}\left(a\right)\right)^{\beta_w/\rho_w}}{\sum_{j\in\mathcal{J}_c}\left(W_{jc}\left(a\right) \cdot G_{jc}\left(a\right)\right)^{\beta_w/\rho_w}},$$

where $L_c^O\left(a\right)$ is the measure of skill-a on-site workers in city c. Firms post skill-specific wages to hire both remote and on-site workers and do not discriminate between them. A firm's wage posting decision is subject to

$$D_{jc}(W, a) = S_{jc}^R\left(W, a\right) + S_{jc}^O\left(W, a\right).$$

I assume that remote workers contribute to agglomeration spillovers in their city of residence. On-site workers choose locations in the same way as equation (2.4). Remote workers choose locations only based on local amenities and the cost of living:

$$L_c^R\left(a\right) = L^R\left(a\right)\frac{\left(R_c\left(a\right) r_c^{-\eta}\right)^{\beta_w}}{\sum_c\left(R_c\left(a\right) r_c^{-\eta}\right)^{\beta_w}}.$$

Appendix G Shapley value

This section formally describes the Shapley value approach. Let $\Omega = \cup_{n=1}^N \omega_n$ denote the full vector of estimated model parameters, where each ω_n is a parameter affecting the equilibrium outcome of interest. Let $X(\Omega)$ represent the value of an equilibrium outcome X under parameter vector Ω. To evaluate the contribution of each parameter ω_n, I construct counter-

Table H.1: Counterfactual analysis: remote work

	Short-run	Long-run
	(1)	(2)
Var city log earnings	−19.2%	−15.4%
Total output	8.1%	9.4%
Population in top 5 cities	−2.2%	−2.7%
High-skilled pop.	−3.8%	−4.3%
Low-skilled pop.	−1.4%	−1.9%
# firms in top 5 cities	0.0%	−8.0%
# high-prod. firms	0.0%	−15.9%
# low-prod. firms	0.0%	−6.2%

Notes: This table displays the counterfactual results of the rise of remote work. Column (1) shows short-run outcomes where only workers relocate; Column (2) shows long-run outcomes where both workers and firms relocate.

factual parameter values $\hat{\omega}_n$. Define $\mathcal{N} = \{1, 2, \ldots, N\}$, and let $\hat{\Omega}_S = \{\cup_{n\in S}\hat{\omega}_n\} \cup \{\cup_{n\notin S}\omega_n\}$ denote the parameter vector in which parameters in subset $S \subseteq \mathcal{N}$ are counterfactual and the rest are baseline. The Shapley value X_n for parameter ω_n with respect to outcome X is defined as

$$X_n = \sum_{S\subseteq\mathcal{N}\setminus\{n\}} \frac{|S|!(N! - |S|! - 1)}{N!} \left[X\left(\hat{\Omega}_{S\cup\{n\}}\right) - X\left(\hat{\Omega}_S\right)\right].$$

Then, the contribution of parameter ω_n to outcome X is given by $\frac{X_n}{X(\hat{\Omega})}$, which, by construction, sums to one across all $n \in \mathcal{N}$. When implementing the method, I define the counterfactual values as follows. For city characteristic parameters, I set $\hat{A}_c = 1$, $\hat{\mu} = 0$, and $\hat{\alpha} = 0$. For amenity parameters, I set $\hat{R}_c(a) = \hat{R}_c = \frac{\sum_a L_c(a)R_c(a)}{\sum_a L_c(a)}$ and $\hat{G}_k(a) = \hat{G}_k = \frac{\sum_c\sum_a D_{kc}(a)G_k(a)}{\sum_c\sum_a D_{kc}(a)}$, thereby removing skill-specific heterogeneity while preserving average amenities. For the skill-augmenting parameters, I set $\hat{\theta}_k = \frac{\sum_k J(k)\theta_k}{\sum_k J(k)}$.

Appendix H Additional Counterfactual Analysis

H.1 Remote work

I apply the model to assess the spatial implications of rising remote work. Following the methodology developed by Dingel and Neiman (2020), Deng et al. (2020) estimate that in Canada 60% of college-educated workers and 29% of non-college-educated workers are

Table H.2: Counterfactual analysis: loosening housing regulation in top 5 cities

% Changes in	Full model (1)	No worker het (2)	No firm het (3)	Only city het (4)
Total output	1.3%	2.0%	2.9%	3.9%
Total welfare	5.3%	5.1%	5.5%	7.0%
Pop. in top 5 cities	52.3%	51.9%	49.7%	46.4%
# of firms in top 5 cities	48.8%	42.0%	48.4%	41.6%

able to perform their jobs remotely. I simulate a scenario in which 25% of these eligible workers shift to remote work. Table H.1 reports the results. In the short run, with firms' locations fixed, remote work shifts workers—particularly the high-skilled—toward smaller, more affordable cities. The five largest cities lose 2.2% of their population, including 3.8% of high-skilled workers. This reallocation lowers the variance of city-level mean log earnings by 19.2%, and raises total output by 8.1%. In the long run, when firms also relocate, these effects magnify: output rises by 9.4%, large-city population losses reach 2.7%. Meanwhile, new high-earning mid-sized cities emerge, moderating the decline in spatial inequality.

H.2 Housing supply regulation

Lastly, I simulate a counterfactual in which the housing supply elasticity of the five highest-earning cities is increased by 25%. Such policies are widely recommended as a means to improve housing affordability and reduce spatial misallocation (Hsieh and Moretti, 2019). Table H.2 shows that the policy increases total output in all four model scenarios by encouraging the reallocation of workers and firms toward more productive cities and by expanding the supply of floor space, an input for production. However, the full model predicts the smallest gains. This reflects the diminished role of city fundamentals once worker and firm heterogeneity is accounted for—suggesting that simplified models may overstate the benefits of relaxing housing constraints.

Appendix I Additional Figures and Tables

Table I.1: City-Size Gradient of Mean Log Earnings

	(1)	(2)	(3)	(3)
Log Population	0.022***	0.023***	0.021***	0.017***
	(0.008)	(0.008)	(0.006)	(0.007)
Demographics		Yes	Yes	Yes
Industry			Yes	
Big city experience				Yes
Num. obs.		66		

Notes: Each column reports a population-weighted city-size regression of mean log earnings on log city population. Population is measured as the number of workers. Column (1) uses raw average log earnings. Columns (2)–(4) use residuals of different Mincer regressions to construct the average city log earnings. Column (2) controls for demographic characteristics, including a cubic age polynomial, gender, marital status, and the number of children. Column (3) controls for demographic characteristics and four-digit NAICS industry fixed effects. Column (4) controls for demographic characteristics and years of experience in the five largest cities since 2002 interacted with an indicator for whether the current city is among the five largest. Standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table I.2: Structural decomposition of city-size gradient of within-city inequality

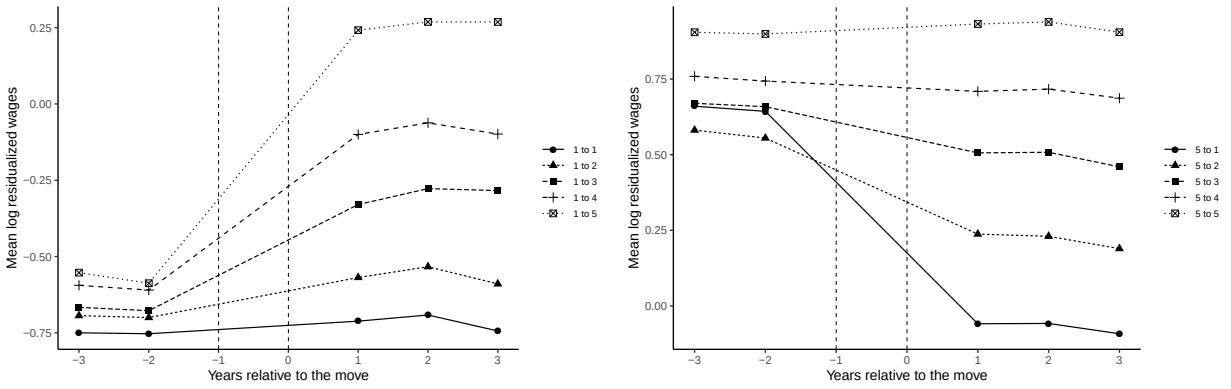
	Total	Var. Worker	Var. Firm	2× Covar.	Match Var.+Covars.
Log Population	0.025***	0.013***	0.003***	0.007***	0.000
	(0.006)	(0.003)	(0.001)	(0.002)	(0.001)
R^2	0.247	0.196	0.103	0.160	0.001
Num. obs.			66		

Notes: Variance of log earnings within a city can be written as: $Var_c(\log w_{it}) = \bar{\theta}_c^2 \cdot Var_c(a_i) + Var_c(z_{j(i,t)} + \theta_{j(i,t)} \cdot \bar{a}_c) + 2 \times Cov[\bar{\theta}_c \cdot a_i, z_{j(i,t)} + \theta_{j(i,t)} \cdot \bar{a}_c]$ + Match var.+covs., with the last term representing the variance and covariance terms related to variation in θ_j. I report city-size regressions of the LHS and each term in the RHS in the table. *p<0.1; **p<0.05; ***p<0.01.

Table I.3: Correlation of city productivity and amenity with local characteristics

Dependent variable:	Productivity (A)	Amenity (R)	
		High-skilled	Low-skilled
	(1)	(2)	(3)
Longitude	−0.006**	0.017	0.010
	(0.003)	(0.032)	(0.025)
Latitude	0.029	0.270	0.254
	(0.026)	(0.288)	(0.226)
January Low Temp.	−0.007	−0.368	−0.256
	(0.022)	(0.251)	(0.197)
January High Temp.	0.016	0.655*	0.476*
	(0.030)	(0.330)	(0.259)
July Low Temp.	−0.022	0.393*	0.332**
	(0.018)	(0.197)	(0.154)
July High Temp.	0.061***	0.073	0.073
	(0.021)	(0.239)	(0.188)
Avg Sunshine Hours	−0.00001	−0.001	−0.001
	(0.0001)	(0.001)	(0.001)
Avg Sunshine Days	−0.001	−0.003	−0.005
	(0.001)	(0.011)	(0.009)
January Wind Speed	0.053***	−0.314	−0.207
	(0.019)	(0.210)	(0.165)
July Wind Speed	−0.065**	0.797**	0.546**
	(0.028)	(0.313)	(0.245)
Air Quality Index	−0.003	0.078	0.055
	(0.005)	(0.051)	(0.040)
Crime Severity Index	−0.0005	−0.042***	−0.034***
	(0.001)	(0.012)	(0.009)
Share of Steep Land	−0.203	7.957*	6.011*
	(0.362)	(4.047)	(3.171)
Share of Water-bodies	−0.0001	0.072	0.032
	(0.008)	(0.088)	(0.069)
R^2	0.746	0.637	0.633
Num. obs.		55	

Notes: Regressions of estimated city productivity (A) and average amenities (R) for high- and low-skilled workers on local characteristics. See Section A for the data sources of the city characteristics. The location characteristics are available for only 50 out of the 66 cities in the baseline sample. *p<0.1; **p<0.05; ***p<0.01.



(a) Movers from the bottom group (b) Movers from the top group

Figure I.1: Event study figure of between-firm worker movers

Notes: These two figures display the earnings profile of between-firm worker movers before and after the move. The sample includes workers who have 1) moved between firms, 2) stayed in the previous firm for more than 4 years prior to the move and 2) stayed in the new firm for more than 3 years. The movement occurs between event years -1 and 0; I omit the average earnings for these two event years given that I only observe annual earnings and workers only work a part of each of these two years for the old and new firms. For ease of exposition, I collapse 10 firm clusters into 5 groups and show workers who move from the top and bottom ones.

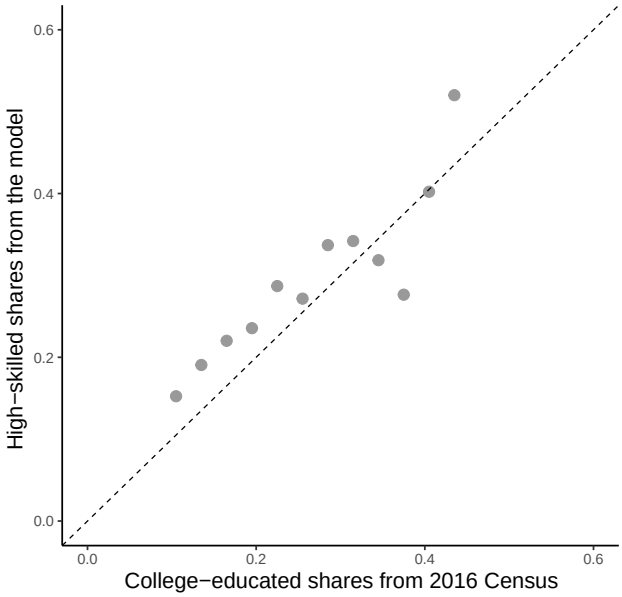
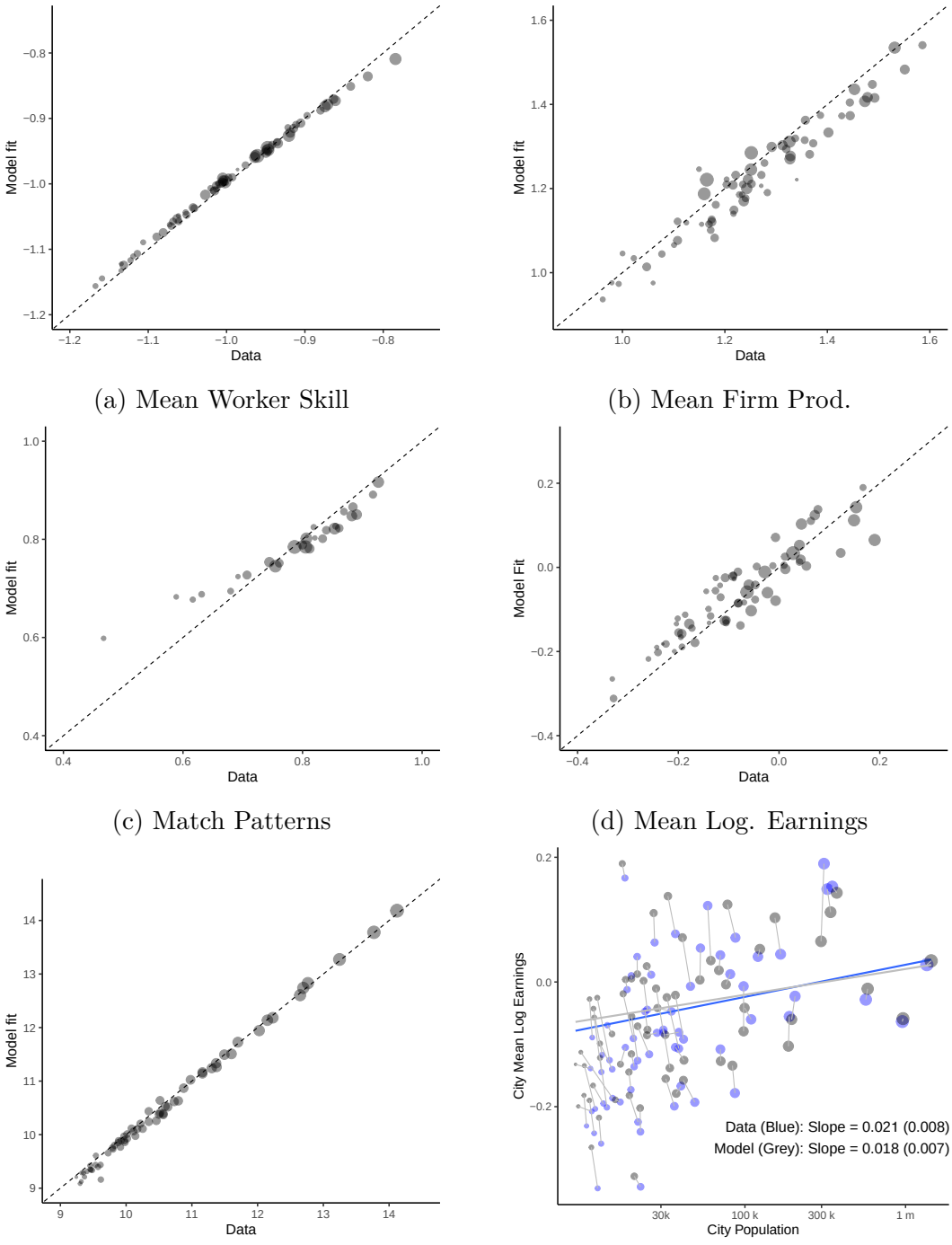


Figure I.2: Estimated high-skilled worker shares highly correlate with college-educated shares

Notes: This figure is a bin scatterplot of the model-estimated city high-skilled worker shares versus the city college-educated shares constructed using 2016 Population Census. High-skilled workers are those in the top three deciles of the worker skill distribution.



(a) Mean Worker Skill

(b) Mean Firm Prod.

(c) Match Patterns

(d) Mean Log. Earnings

(e) Log. Population

(f) Urban Earnings Premium

Figure I.3: Model fit

Notes: These figures present the model fit. Each dot is a city, with its size proportional to city population. Panels (a) and (b) show the average worker skill and firm common productivity for each city, respectively. Panel (c) plots the share of top-tercile workers matched with top-tercile firm clusters in each city, with the smallest 30 cities aggregated into a single point due to data release restrictions. Panels (d) and (e) present the average city log earnings and log city population, respectively. Finally, panel (f) compares the urban earnings premium estimated from the data with that predicted by the model.

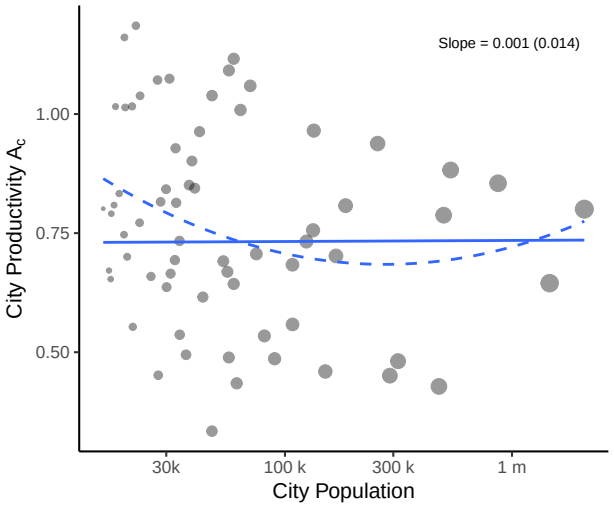
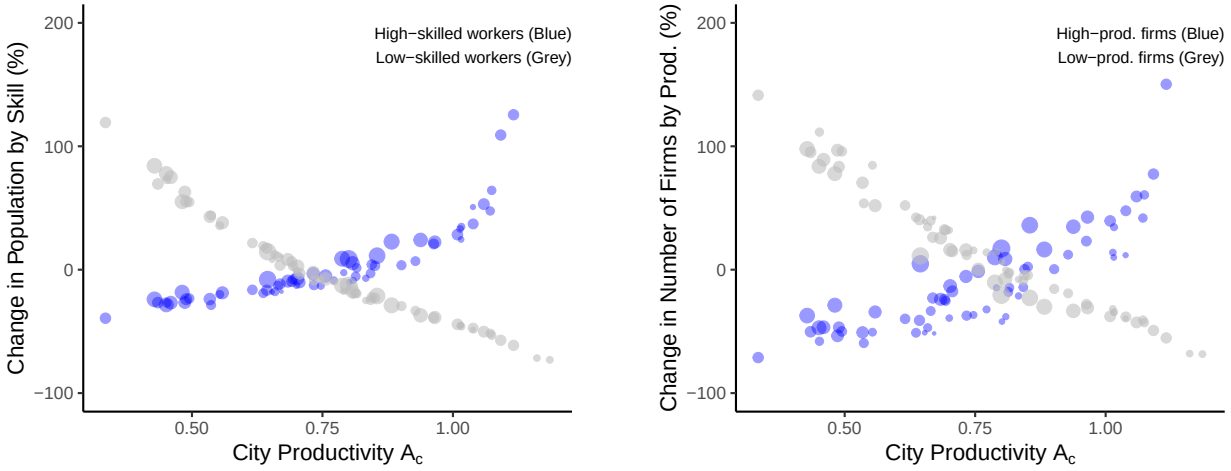


Figure I.4: City exogenous productivity versus city population

Notes: This figure is a scatterplot of estimated city productivity A_c on city population. The solid line is a weighted linear fit and the dashed line is a quadratic fit. The population-weighted OLS regression coefficient and the standard errors are reported.

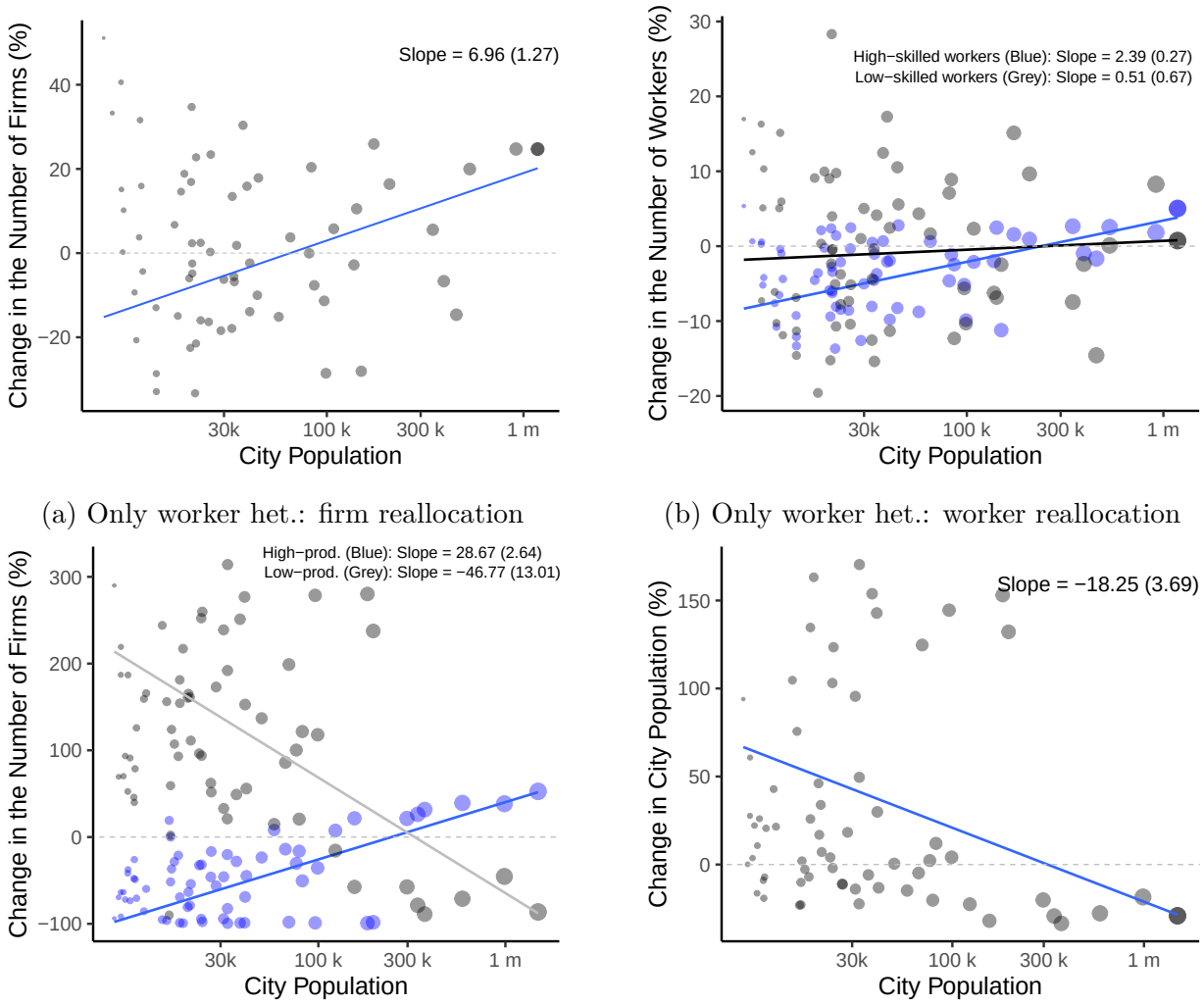


(a) Worker reallocation

(b) Firm reallocation

Figure I.5: Spatial reallocation of workers and firms of the optimal policy between cities with different exogenous productivities

Notes: The figures display the changes in the number of workers and firms induced by the optimal policy versus city exogenous productivity A_c. Each dot is a city, with its size proportional to city population. In panel (a), blue dots represent high-skilled workers and grey dots represent low-skilled workers; in panel (b), blue dots represent high-productivity firms and grey dots represent low-productivity firms.



(a) Only worker het.: firm reallocation

(b) Only worker het.: worker reallocation

(c) Only firm het.: firm reallocation

(d) Only firm het.: worker reallocation

Figure I.6: Spatial reallocation of workers and firms of the optimal policies with one-sided heterogeneity

Notes: The figures display the reallocation of workers and firms induced by optimal policies in models with only one-sided heterogeneity. Each dot is a city, with its size proportional to city population.