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

Guangbin Hong†

University of Chicago

March 15, 2025

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 a 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: E25, R12, R13.

*I am deeply indebted to Nate Baum-Snow, Joseph Steinberg, Kevin Lim, Serdar Ozkan, and Kory Kroft for their 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, Hans Koster, 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, Chris Severen, Martin Schneider, Immo Schott, Gregor Schubert, 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 (RDC). I thank CRDCN 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.

†University of Chicago. Address: 1155 E 60th St, Chicago, IL 60637. Email: ghong7@uchicago.edu.

1 Introduction

Spatial earnings inequality is accompanied by the sorting of high-skilled workers and high-productivity firms into larger cities.[1] Local governments around the world spend billions of dollars annually on place-based policies aimed at attracting the best workers and firms.[2] 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; yet 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 with Canadian matched employer-employee data. I first use the estimated model to decompose the earnings premium of larger cities, finding that worker sorting and firm sorting each account for 66.8% and 26.8% of the premium, respectively. Second, I find novel firm sorting externalities that arise from the interplay of two-sided sorting and local matching: low-productivity firms tend to overvalue locating in large cities, where they inefficiently compete with high-productivity firms for high-skilled workers. The optimal spatial policy would encourage greater co-location of high-skilled workers and high-productivity firms while redistributing income towards low-earning cities. Third, by accounting for two-sided spatial sorting, I find that commonly considered place-based subsidies can lead to unintended adverse distributional or aggregate effects. Overall, the analysis underscores the crucial role of two-sided sorting in explaining spatial inequality and informing the design of effective place-based policies.

The model is motivated by well-established empirical facts on the spatial distributions of heterogeneous workers and firms across cities (Diamond, 2016; Combes et al., 2012), the variation in the degrees of assortative worker-firm matching within cities (Dauth et al., 2022), and wage inequality between and within cities (Baum-Snow and Pavan, 2013). In the model, heterogeneous workers and firms choose among a set of cities, each with different exogenous city productivity and skill-specific amenities. City productivity increases with population due to agglomeration spillovers. Within a city, each firm posts skill-specific wages, and each worker chooses a firm based on posted wages, non-wage firm amenities, and idiosyncratic

[1] See Combes et al. (2008), Eeckhout et al. (2014), and Diamond (2016) for evidence of worker sorting, and Combes et al. (2012), Gaubert (2018), Bilal (2023) for evidence of firm sorting.

[2] For instance, many U.S. states actively compete for high-skilled workers by offering lower tax rates (Moretti and Wilson, 2017). Similarly, many local governments offer subsidies to attract productive firms to locate in their jurisdictions (Greenstone et al., 2010), including the recent bidding war for Amazon HQ2. For reviews on place-based policies, see Glaeser and Gottlieb (2008) and Duranton and Venables (2018) among others.

firm-specific preferences that 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. Interactions in local labor and housing markets, along with fundamental city characteristics, affect workers' utilities and firms' profits in each city, ultimately shaping their location choices. As a result, the spatial sorting of workers and firms is interdependent, with their location choices jointly pinning down the wage structure within and across cities.

The model features production complementarities between worker skill and firm productivity. Specifically, following Bonhomme, Lamadon, and Manresa (2019) (hereafter BLM), I incorporate skill-augmenting productivity that differs across firms. These complementarities incentivize the co-location of high-skilled workers and high-productivity firms, as positive assortative matching generates greater returns. In addition, workers' preference heterogeneity induces horizontal differentiation between firms, leading to two key implications. First, it gives rise to firms' monopsony power, allowing them to set wages along the upward-sloping labor supply curves, as in Card et al. (2018).[3] Second, because workers perceive different employers as imperfect substitutes, they have a love-of-variety preference, valuing a greater number of firms in a city as in Helsley and Strange (1990).

I identify two novel firm sorting externalities arising from the interaction between imperfect labor market competition and 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 consists of a local and a national component. The local component includes the stealing and love-of-variety externalities that are common in entry models under imperfect competition (Mankiw and Whinston, 1986). The national component represents the productivity changes resulting from the spatial reallocation of workers induced by each firm's location choice. Consequently, the planner can still enhance welfare through spatial reallocation of firms, even when the local wedge is net zero, as in Dixit and Stiglitz (1977). With production complementarities, welfare-improving spatial reallocation should increase the degree of worker-firm assortative matching in the economy. Therefore, 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 records information on the annual earnings of all Canadian workers at their firms across cities. The main challenge is to separately identify city, firm, and worker productivity pa-

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

rameters. To achieve this, I combine the movers design of workers with a revealed preference approach based on firms' location choices. First, I follow BLM to identify worker skills and firm skill-augmenting productivities based on earnings changes when workers move between firms.[4] Then, I identify city productivities by leveraging firms' revealed location decisions. Intuitively, a city is inferred to be productive if it attracts a large share of firms despite having high floor space rents and a competitive labor market. Therefore, unobserved city productivities can be estimated as compensating differentials to explain the observed firm sorting shares, after controlling for other firm profits determinants.

I use the estimated model to decompose the spatial earnings differential, finding that differences in worker, firm, and city productivities across cities account for 66.8%, 26.8%, and 6.6% of the urban earnings premium, respectively.[5] Three additional insights are noteworthy. First, I estimate strong production complementarities between worker skill and firm productivity, which are crucial in shaping the sorting patterns and the spatial earnings differential. Counterfactual simulations show that eliminating these complementarities would reduce the covariance between city-average worker skill and firm productivity by 44.2% and the urban earnings premium by 40.6%. Second, while the estimated agglomeration elasticity is minimal after accounting for worker and firm heterogeneity, worker-firm complementarities help explain previous findings on the greater productivity benefits of larger cities for higher-skilled workers and higher-productivity firms (Davis and Dingel, 2020; Gaubert, 2018). Third, consistent with Diamond (2016), I estimate that larger cities offer greater and skilled-biased amenities, which further 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 leads to a 9.6% increase in total output. To minimize the negative labor market stealing effects, the optimal policy decreases the number of low-productivity firms in higher-skilled, larger cities, while increasing the number of high-productivity firms there. Furthermore, the optimal policy contains spatial transfers towards low-earnings cities, to spatially redistribute the efficiency gain in an equitable way. The policy generates a consumption-equivalent Pareto welfare gain of 4.9% for all agents in the

[4]These worker movers include those who move within a city and those who move across cities. To mitigate limited mobility bias in estimating firm effects (Andrews et al., 2008), I adopt the approach of BLM in grouping firms into clusters based on their empirical earnings distributions. The clustering exercise to uncover firm heterogeneity is non-trivial as earnings contain information on both firm and city characteristics. See Section 4 for an iterative algorithm I develop to overcome this challenge.

[5]This decomposition holds fixed the spatial distribution of workers and firms, allocating mean city earnings to different productivity components. The three shares do not sum to 100%: the remaining -0.2% reflects an interaction component that captures the effect of skill-augmenting productivity through worker-firm matching.

economy.

Lastly, I demonstrate that the interdependence of two-sided sorting significantly magnifies the aggregate and distributional effects of placed-based subsidies. To illustrate this, I compare the policy outcomes from the full model with those from models where either workers or firms are immobile. I simulate a 5% wage subsidy targeted at firms in the top five percentiles of productivity that choose to locate in Toronto, which emulates the bid proposed by the Toronto government to attract Amazon HQ2. In the full model, the subsidy increases total output of the Canadian economy by 0.8% but also amplifies the between-city variance in mean log earnings by 14.5%. Moreover, I find that while high-skilled workers in Toronto benefit from the policy, local low-skilled workers experience significant welfare losses due to rising housing costs and the exodus of low-productivity firms. In contrast, the aggregate and distributional impacts of the policy are much smaller when evaluated by models with only one-sided sorting.

This paper is closely related to three strands of literature. The first is the literature on spatial inequality (Combes et al., 2008; Moretti, 2013; Baum-Snow and Pavan, 2012; Behrens et al., 2014; Davis and Dingel, 2020), which has shown that larger cities exhibit higher average wages, high-skilled worker shares, skill premia, and housing costs. A growing body of literature examines the spatial sorting of heterogeneous agents across cities, including worker sorting (Baum-Snow and Pavan, 2013; Eeckhout et al., 2014; Diamond, 2016; Davis and Dingel, 2019) and firm sorting (Gaubert, 2018; Lindenlaub et al., 2022; Bilal, 2023), and quantifies their contributions to the urban wage premium. I contribute to this literature by developing a two-sided sorting model and by estimating the respective roles of worker and firm sorting in driving spatial inequality.[6]

The second is the literature on optimal spatial policies (Kline and Moretti, 2014; Fajgelbaum and Gaubert, 2020; Rossi-Hansberg et al., 2019), which has primarily focused on addressing sorting inefficiencies caused by local productivity and amenity spillovers (see Fajgelbaum and Gaubert (2025) for a recent review). I show that the interaction between spatial sorting and imperfect competition in local markets can cause inefficiencies. The firm sorting externalities I identify are analogous to the firm entry externalities studied by Dixit and Stiglitz (1977) and Mankiw and Whinston (1986). To my knowledge, my work is the first to formalize these externalities a spatial equilibrium model. In addition, the labor market stealing effect resembles the pooling externality in Bilal (2023), wherein low-productivity firms sub-optimally co-locate with high-productivity firms in a frictional labor market en-

[6] Hoelzlein (2023) examines how two-sided sorting of workers and firms within a city shapes the welfare effects of neighborhood-targeted policies through consumption and commuting channels. In comparison, I focus on spatial sorting across cities while modeling labor market matching within cities. I view these two studies as complementary.

vironment with homogeneous workers. I show that the extent of the inefficiency crucially depends on the match returns of heterogeneous workers and firms, as well as the spatial mobility of workers.

Lastly, I contribute to the literature on earnings inequality using the two-way fixed effects approach, which is introduced by Abowd, Kramarz, and Margolis (1999) (hereafter, AKM) and further developed in Card et al. (2013), Song et al. (2019), BLM, and LMS. Recent studies on spatial inequality have adopted this method. Both Dauth et al. (2022) and Card et al. (2025) find that larger cities have higher-earning workers and higher-paying firms, with city-level productivity determinants absorbed into the firm effect. I contribute to this literature by separately identifying the worker, firm, and city components from an earnings equation.

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 the two-sided sorting spatial equilibrium model, which extends the standard Rosen-Roback framework (Rosen, 1974; Roback, 1982) by incorporating rich heterogeneity.[7] 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.

2.1 Environment

Consider an economy populated by heterogeneous workers and firms. Workers, each indexed by i, differ in skill $a_i \in (\underline{a}, \bar{a})$. Firms, indexed by j, each differ in production technology, characterized by (z_j, θ_j), and skill-specific amenities $\{G_j(a)\}_a$. All firms produce a homogeneous, tradable final good, which serves as the economy's numeraire.[8] The measure of workers of each skill, $L(a)$, 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 the amount

[7] In Appendix B, I provide four empirical facts using Canadian matched employer-employee data to motivate the model. These facts are (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 assortative matching is greater in larger cities.

[8] I abstract from industry heterogeneity in the model. In Figure J.3, I show that controlling for the industry composition does not significantly affect the estimated urban earnings premium. Card et al. (2025) also find that the wage premia of each city vary little across industries.

of land used for housing production $\bar{H}_c$. A city's productivity increases with its population L_c through endogenous agglomeration spillovers, with the elasticity given by μ.[9] In each city, firms set skill-specific wages, and workers choose their employers accordingly. Both workers and firms demand floor space, which is supplied by a local housing developer. I denote $\mathcal{L}_c$ as the set of workers and $\mathcal{J}_c$ as the set of firms located in city c.

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 of city c, η is the expenditure share on housing, $\tau > 1$ represents the rebate of national land rents which is assumed to be proportional to a worker's labor earnings. As will become clear shortly, this assumption ensures that the rebate does not affect workers' firm or city choices. Moreover, ϵ_{ij} is worker i's idiosyncratic preference for firm j. I assume that ϵ_{ij} is drawn from a Type-I 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 degree of correlation of tastes for different firms within each city, i.e. $\rho_w = \sqrt{1 - corr(\epsilon_{ij}, \epsilon_{ij'})}$ if $j, j' \in \mathcal{J}_c$. A smaller ρ_w is associated with more correlated preference draws for firms within the same city. The correlation captures each worker's idiosyncratic preference for the city. For example, a worker strongly attached to a particular city may have high preferences for all firms located there. The dispersion of the idiosyncratic preference is controlled by parameter β_w. A higher β_w reduces the dispersion, leaving a smaller role of the preference idiosyncrasy in affecting a worker's preferred firm and city.

This utility function extends LMS by incorporating preferences for city amenities and housing rents, and builds on Diamond and Gaubert (2022) by adding preferences for firm amenities and wages within cities. The specification introduces both vertical and horizontal differentiation across firms and cities. The horizontal differentiation gives rise to a love-of-

[9] In Appendix G.1, I extend the model to allow city amenities to respond to the local skill composition (Diamond, 2016; Almagro and Dominguez-Iino, 2024). This endogenous amenity channel can amplify worker sorting.

variety preference of workers for more local firms as in Helsley and Strange (1990), since workers view different employers as imperfect substitutes.

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 to work for. Given the property of the Type-I Extreme Value distribution, the measure of skill a workers choosing firm j is:

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

where $L_c(a)$ is the measure of workers of skill a 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 curve by skill, with the firm-level labor supply elasticity given by β_w/ρ_w. Analogously, the measure of skill-a workers choosing city c is:

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

where $U_c(a) \equiv R_c(a){r_c}^{-\eta} \cdot \mathbb{W}_c(a)^{\frac{\rho_w}{\beta_w}}$. Workers trade off the wage index and city amenities against housing rents when making the location choice, with the city-level labor supply elasticity given by β_w. The exponent of the wage index $\mathbb{W}_c(a)$ in $U_c(a)$, ρ_w/β_w, reflects the love-of-variety elasticity for more local firms and corresponds to the inverse of the firm-level labor supply elasticity.

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 $D_{jc}(a)$ is the measure of skill-a workers employed by the firm j in city c, and α is the output elasticity of floor space in the production function.

There are several labor productivity terms in the production function, including a city part $A_c L_c^{\mu}$, a firm part z_j, and a worker-firm interaction part a^{θ_j}. Firm technology is characterized by the common productivity z and the skill-augmenting productivity θ. Heuristically speaking, common productivity z represents a firm's absolute advantage for all workers, and skill-augmenting productivity θ represents a firm's comparative advantages in utilizing higher-skilled workers. This specification is developed by BLM to extend the log-additive specification in AKM.

If skill-augmenting productivity θ is monotonically increasing in common productivity z, then worker skill a and firm productivity z are complements in production. Mathematically, in this case, the matched output of a worker-firm pair is log-supermodular in worker skill a and firm productivity z, leading to positive assortative matching. I will empirically estimate the two productivity parameters and show that they are nearly perfectly correlated across firms.

Two additional assumptions are made on the production technology. First, I assume that all firms in a given city share a common city-level productivity. Both Gaubert (2018) and Bilal (2023) find complementarities between firm productivity and city size that drive firm sorting. My model builds on this by introducing worker heterogeneity and the worker-firm complementarities. Since larger cities tend to attract a higher-skilled workforce, these features provide a micro-foundation for the firm-city complementarities identified in these previous studies.

Second, I assume that workers with different skills are perfect substitutes within a firm. Eeckhout et al. (2014) estimate extreme-skill complementarities at the city level based on thicker tails in the skill distributions of 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.

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 decides its optimal production and location choices in a backward manner. First, for each potential location c, it chooses the optimal 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} to maximize its expected profits. Second, it chooses a city based on these expected profits. The first step is a standard profit maximization problem:

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

subject to the labor supply curves specified by equation (2.3):

$$D_{jc}(a) = S_{jc}(a; W_{jc}(a)) = L_c(a) \cdot \frac{(W_{jc}(a)G_{jc}(a))^{\frac{\beta_w}{\rho_w}}}{\mathbb{W}_c(a)}, \ \forall a. \tag{2.7}$$

Following Card et al. (2018) and LMS, I assume that firms are infinitesimal in the local labor market.[10]

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

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

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

$$W_{jc}(a) = \underbrace{\frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}}_{\text{Wage markdown}} \cdot \underbrace{(1-\alpha)\alpha^{\frac{\alpha}{1-\alpha}} \cdot A_c L_c^{\mu} r_c^{\frac{\alpha}{\alpha-1}} \cdot z_j \cdot a^{\theta_j}}_{\text{Marginal Product of Labor (MPL)}}. \tag{2.8}$$

It can be seen that the wage markdown $\frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}$, which is workers' take-home share of their marginal revenue product, increases with the firm-level labor supply elasticity β_w/ρ_w. When workers view different firms as closer substitutes, labor supply becomes more elastic, which in turn diminishes firms' monopsony power. The local housing rent r_c affects the optimal wage offer as it changes the optimal housing input and thus the marginal product of labor. With the optimal wage setting and input choices, firm j's optimal profits in city c can be obtained as:

$$\pi_c(j) = \Psi \cdot \left(A_c L_c^{\mu} r_c^{\frac{a}{\alpha-1}}\right) \cdot (z_j)^{1+\beta_w/\rho_w} \cdot \phi_{jc} \tag{2.9}$$

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

$$\phi_{jc} = \int_a \left(a^{\theta_j}\right)^{1+\beta_w/\rho_w} L_c(a) \cdot \frac{G_{jc}(a)^{\frac{\beta_w}{\rho_w}}}{\sum_{j' \in \mathcal{J}_c} \left(z_{j'} a^{\theta_{j'}} \cdot G_{j'c}(a)\right)^{\frac{\beta_w}{\rho_w}}} da. \tag{2.10}$$

Firm j's maximized profits depend on city c's adjusted productivity $A_c L_c^{\mu} r_c^{\frac{\alpha}{\alpha-1}}$, the firm's common productivity z_j, and a labor composite term ϕ_{jc}. This labor composite term captures the total efficiency units of labor that firm j can employ in city c, synthesizing local

[10]This assumption abstracts from the strategic interaction of firms' wage-setting decisions, which is studied by Berger et al. (2022) with a fixed set of non-atomistic firms in each labor market.

labor supply and demand conditions along with the firm's ability to attract and complement workers with different skills. 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}. More details on the profit maximization problem can be found in Appendix C.1.

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:

$$u_c(j) = \ln \pi_c(j) + {\beta_f}^{-1} \nu_{jc}. \tag{2.11}$$

Analogous to workers, 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 (2.9), the probability that entrepreneur j chooses city c is:

$$p_c(j) = \frac{\left(\left(A_c L_c^{\mu} \bar{r}_c^{\frac{\alpha}{\alpha-1}}\right) \cdot \phi_{jc}\right)^{\beta_f}}{\sum_{c'} \left(\left(A_{c'} L_{c'}^{\mu} \bar{r}_{c'}^{\frac{\alpha}{\alpha-1}}\right) \cdot \phi_{jc'}\right)^{\beta_f}}. \tag{2.12}$$

Equation (2.12) 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 firms' 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 (2.10), 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 productivities and firm common productivities.

2.4 Housing supply

Housing in city c is produced by an absentee housing developer, who combines the final good Y and land $\bar{H}_c$ to produce floor space:

$$H_c = \bar{H}_c \cdot Y^{\frac{1}{1+\gamma_c}} \tag{2.13}$$

where $\gamma_c > 0$ governs the returns-to-scale of the housing production function. The developer chooses the final good input Y to maximize profits, taking the rental rate r_c as given. Solving the developer's problem yields the housing supply curve for each city:

$$H_c^S(r_c) = \bar{H}_c^0 \cdot {r_c}^{\frac{1}{\gamma_c}} \tag{2.14}$$

where $\bar{H}_c^0$ is an exogenous housing supply shifter. I assume that the developer's profits, which are equivalent to 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 worker location and firm choices* $\{c(i), j(i)\}_{\forall i}$*, firm location choices* $\{c(j)\}_{\forall j}$*, firm labor demand* $\{D_{jc}(a)\}_{\forall a,j,c}$*, housing input demand* $\{h_{jc}\}_{\forall j,c}$*, 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 (2.12), and choose wage offers and housing inputs by equations* (2.8) *and* (C.6)*;*

3. *The housing market clears in each city such that housing demand equals supply, which are given by equations (C.14) and (C.13);*

4. *The final good market clears by equation* (C.17)*;*

5. *Land rents are redistributed to workers by equation* (C.16)*.*

One concern with the two-sided sorting model is that it may result in multiple equilibria, as workers and firms mutually follow each other across cities. This interdependence is particularly pronounced in the presence of production complementarity. I prove in Appendix E that a unique equilibrium exists when idiosyncratic preferences of workers and firms, which serve as congestion forces in the model 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 are found to converge to the same equilibrium.

3 Efficiency of Two-Sided Sorting

In this section, I examine the impact of two-sided sorting on overall efficiency and design the optimal policy to maximize social welfare. I first solve the social planner's problem and compare the social and private values associated with workers' and firms' location choices. This comparison informs the sorting externalities in the laissez-faire economy. Building on these insights, I then develop the optimal spatial policy that maximizes social welfare.

The social planner maximizes the aggregate welfare of all agents in the economy, including workers and entrepreneurs. For simplicity, assume there are N^w types of workers, indexed by a, and N^f types of firms (entrepreneurs), indexed by k, though the result can be readily generalized to an infinite number of types. Suppose the planner assigns Pareto weights $\varphi^w(a)$ to each worker type and $\varphi^f(k)$ to each firm type. The social welfare function is then defined 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. The planner maximizes equation (3.1) by choosing the spatial allocations of heterogeneous workers and firms across cities, $\{L_c(a)\}$ and $\{J_c(k)\}$; the measure of worker-firm matches in each city, $\{D_{kc}(a)\}$; the consumption of the final good by workers and entrepreneurs, $\{c_{kc}(a)\}$ and $\{c_c(k)\}$; the final good used to produce floor space, $\{Y_c\}$; and the allocation of housing across workers and firms, $\{h_{kc}(a)\}$ and $\{h_c(k)\}$. These choices are subject to constraints on spatial mobility, local matching of workers and firms, and the resource constraints of final goods and housing.[11] This planning problem generalizes the frameworks in Fajgelbaum and Gaubert (2020) and Rossi-Hansberg et al. (2019) by incorporating decisions on the allocation of heterogeneous firms between cities and the worker-firm matching within cities. More details of the planning problem can be found in Appendix D.

3.1 Sorting externalities

From the planning problem, I first derive the social marginal value associated with workers' and firms' location choices, which are shown in the following proposition.

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

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

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

where $W_{kc}(a)$ is the competitive equilibrium wage defined in (2.8), and $\bar{W}_c$ is the average wage in city c.

[11] Here, $D_{kc}(a)$ represents the total measure of type-a workers employed by type-k firms in city c. The number of type-a workers employed by each type-k firm j in city c is $D_{jc}(a) = D_{kc}(a)/J_c(k)$.

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 worker a at firm k in city c, and* $\bar{W}^*_c(a)$ *and* $\bar{W}^*(a)$ *are the skill-specific mean values at the city and national levels, respectively.*

The difference in social and private values reveal the sources of sorting inefficiency in the laissez-faire equilibrium. For workers, they are compensated by firms at a fraction of their marginal labor product, while their social values reflect their full contribution to final good production, including both their direct production output and the agglomeration spillovers they generate. However, workers do not internalize the agglomeration spillovers associated with their location choices, which are represented by $\mu\bar{W}_c$ in equation (3.2). Such agglomeration externalities have been well studied in the literature, including Fajgelbaum and Gaubert (2020) and Rossi-Hansberg et al. (2019) among many others.

For firms, they consider only the profits earned in a city, represented by the firm surplus component in equation (3.3). However, firms do not internalize two key externalities. The first is the labor market stealing externality, which includes local labor stealing externality, where firms hire workers away from other firms within the same city, and national labor stealing externality, where firms attract workers from other cities. The second is the love-of-variety externality, arising from workers' preference for a greater number of horizontally differentiated firms.

These firm sorting externalities arise from imperfect labor market competition. 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 in a city has two effects through its impact on local wage indices: it congests the local labor market, making it harder for other firms to hire worker (equation (2.3)), while it also expands local employment opportunities, improving local workers' welfare and attracting new workers from other cities (equation (2.4)).

The source of inefficiency is analogous to the labor market pooling externality studied by Bilal (2023), who develops a firm sorting model with frictional labor markets characterized by random search. In Bilal (2023), where vacancies posted by high- and low-productivity firms are pooled in the matching function, firm sorting is inefficient because wages do not reflect the impact of firms' location choices on the local vacancy filling rates. Similarly, in the monopsony framework, the externality arises because wages posted by firms do not account for their impact on the local labor market through the wage indices, $\mathbb{W}_c(a)$.

Equation (3.3) is also 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 product market.[12] In the spatial context, two-sided sorting and the spatial segmentation of labor markets give rise to a novel source of efficiency through the national stealing effect. To make this clear, I rewrite equation (3.3) for any firm j as follows:

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

where $\bar{\bar{\tilde{W}}}_c(a)$ and $\bar{\bar{\tilde{W}}}(a)$ are the city-level and national-level mean social values of type-a workers, respectively, and I set $\mu = 0$ to isolate the wedges with respect to firm sorting inefficiency. The local wedge has the same trade-off as in Dixit and Stiglitz (1977). The national wedge reflects the additional efficiency effect associated with worker mobility. Workers' location decisions are affected by firms. As a result, the location choice of the firm j affects the productivity of workers induced to choose city c, while these workers do not necessarily work for that firm after locating in city c.[13] From the social planner's perspective, this productivity effect is reflected in the gap between the city-average and national-average social values of workers. The measure of relocated workers is given by ρ_w times firm j's employment $\mathbf{D}_{jc}$,

[12]There is no business stealing externality in this model because all firms are price takers in the product market, which is consistent with Corollary 1 in Mankiw and Whinston (1986). Moreover, to explicitly include the wage markdown in the expression, equation (3.3) can be re-written as

$$\tilde{\pi}_c(k) = \sum_a \frac{D_{kc}(a)}{J_c(k)} \left(\frac{1}{1+\beta_w/\rho_w} \tilde{W}_{kc}(a) - \frac{1}{1+\beta_w/\rho_w} \left[(1-\rho_w) \bar{\bar{\tilde{W}}}_c(a) + \rho_w \bar{\bar{\tilde{W}}}(a) \right] + \frac{\rho_w}{\beta_w} \varphi^w(a) + ... \right)$$

where the omitted terms account for the changes in agglomeration spillovers by re-locating workers.

[13]One might suspect an additional worker sorting externality: workers neglecting their impact on firms. However, the fixed rent-sharing implies that workers indirectly account for their impact on firms' profits.

where ρ_w is the ratio between city-level and firm-level labor supply elasticities.

The lemma below illustrates the role of two-sided sorting, the spatial segmentation of the labor markets, and spatial inequality in affecting firm sorting inefficiency.

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

1. *With β_w/ρ_w fixed, when $\rho_w \to 0$, firm sorting is efficient if the planner can set city-skill specific Pareto weights as $\varphi_c^w(a) = \bar{W}_c(a),\ \forall a, c$.*

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

3. *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 $\varphi^w(a) = \bar{W}(a),\ \forall a$.*

When ρ_w approaches 0, the labor stealing effects only happens locally, since workers no longer spatially follow firms. When ρ_w approaches 1, there is no distinction between local and national labor stealing, as local markets of different cities are spatially integrated. When there is no spatial inequality, national stealing is not associated with productivity changes of the relocated workers. Under all three cases, the two wedges in equation (3.4) collapse into one as in Dixit and Stiglitz (1977) with heterogeneity. Any remaining discrepancy between the optimal and laissez-faire allocations would be caused by redistributional considerations of the planner – the planner would allocate more firms to a city where the workers' Pareto weights exceed their average wages.

Equation (3.4) informs the determinants of the welfare losses from inefficient firm sorting and clarifies how the planner's solution should differ from the laissez-faire allocation. All else equal, the degree of inefficiency is greater when β_w/ρ_w is smaller, corresponding to greater monopsony power. In addition, with imperfect worker mobility across cities ($0 < \rho_w < 1$), the degree of inefficiency is greater when there exist greater within-type spatial earnings disparities.

Guided by the national wedge, the planner should reallocate firms to cities where workers who follow the firms can earn above national-average wages. First, this implies that the planner would want to reallocate more firms towards high-productivity cities. Furthermore, with worker-firm production complementarity, high-skilled workers are more likely to follow high-productivity firms. Therefore, reallocating high-productivity firms to cities with higher skill premia (greater $\bar{W}_c(a) - \bar{W}(a)$ for high a's)—typically larger cities—and reallocating low-productivity firms away from these locations can improve efficiency.

3.2 Efficiency conditions and the optimal policy

I now turn to characterizing the efficiency conditions of the spatial economy. Following Fajgelbaum and Gaubert (2020), I derive conditions on the expenditure distributions, $\{x_{kc}(a)\}$ and $\{x_c(k)\}$, that must hold in any efficient allocation. Here, I define $x_{kc}(a) \equiv c_{kc}(a) + r_c h_{kc}(a)$ as the total expenditure of a type-a worker employed by a type-k firm in city c, and $x_c(k) \equiv c_c(k)$ as the total expenditure of a type-k firm in city c. In Appendix D, I compare the competitive equilibrium characterized by the expenditure distributions $\{x_{kc}(a)\}$ and $\{x_c(k)\}$ with the outcomes of the social planner's problem, which leads to the following proposition.

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

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

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

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

for $J_c(k) > 0$ $\forall k, c$. The opportunity costs, $\mathcal{O}_c^w(a)$ and $\mathcal{O}_c^f(k)$, 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.

The two conditions in the proposition above describe the relationship between private consumption and the spatial allocation of workers and entrepreneurs that must hold in any efficient allocation. Private consumptions 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, efficient consumption allocations account for the social opportunity costs associated with allocating workers and firms. Within each type, workers' opportunity costs, $\mathcal{O}_c^w$, vary across cities, while firms' opportunity costs, $\mathcal{O}^f$, do not. The spatial variation in workers' opportunity costs arises from correlated idiosyncratic preferences within cities and different matching opportunities between cities.[14]

[14] See Appendix D for the expressions of the opportunity cost terms, which relate to the shadow costs in the planner's problem and can be interpreted as the "price of congestion". From the planner's perspective,

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 equity 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 weights.

Equipped with the efficiency conditions, I now design a set of tax instruments to implement the optimal allocation.

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

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

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

where $\tilde{W}_{kc}(a), \tilde{\pi}_c(k)$ are workers' and firms' social values defined in Proposition 2, and $T^w_c(a)$ and $T^f(k)$ are defined as $T^w_c(a) \equiv -\frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}\mathcal{O}^w_c(a) + \frac{1}{1+\beta_w/\rho_w}\varphi^w(a)$ and $T^f(k) \equiv -\frac{\beta_f}{1+\beta_f}\mathcal{O}^f(k) + \frac{1}{1+\beta_f}\varphi^f(k)$.

These optimal taxes incorporate Pigouvian adjustments for efficiency and redistribution adjustment for equity considerations. I show in Appendix D.4 that the policy described in the proposition is revenue neutral. Moreover, I assume that firms treat worker income taxes as fixed when setting wages and that land rents are rebated to workers proportionally to their after-tax income, $(1 - t^w_{kc}(a)) \cdot W_{kc}(a)$. These assumptions are made to preserve the iso-elastic labor supply curves at the firm level, which consequently does not affect the optimal wages posted by firms.

4 Empirical Implementation

I now turn towards identification and estimation of the model's parameters. I first describe the main sources of data used in the empirical analysis in Section 4.1. I then introduce additional assumptions useful for identification in Section 4.2, before discussing the estimation strategy in Section 4.3 and presenting the estimation results in Section 4.4.

assigning more workers or firms to a city implies worse idiosyncratic preferences for the marginal worker or firm.

4.1 Data

The main dataset for my 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, and I mainly use the data from 2010 to 2017 for the empirical analysis. On the worker side, I observe total annual earnings from each firm in each year, residential location, as well as demographic information such as age and gender. On the firm side, I observe location, industry (4-digit NAICS code), wage bill, revenue, and value-added. All monetary variables are converted to 2002 Canadian dollars.

The baseline sample of the analysis includes full-time working individuals between the ages of 25 and 60 who live in a city (see the definition of a city below). For workers receiving earnings from multiple firms within a year, I only keep the highest-paying job for that year. I drop worker-year observations with annual earnings less than the equivalent of working 20 hours per week for 13 weeks at the minimum hourly wage. Furthermore, I exclude firms in industries including agriculture, mining, utilities, education, healthcare, non-profit organizations, and public administration, as firms in these industries have other considerations when making location and wage-setting decisions.

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 location of their headquarters is observed. I exploit workers' residential location information to form enterprise–city units within multi-location firms. I assume that each unit has its own production technology and makes independent location choice and wage-setting decisions. I will refer to such units as firms hereafter.

A city is defined as a Census Metropolitan Area (CMA) or a Census Agglomerate (CA) delineated in the 2016 Census of Population. The concepts of CMA and CAs resemble U.S. commuting zones, as they group a population center with surrounding municipalities that are closely linked through commuting flows. I keep CMAs and CAs with no fewer than 15,000 full-time working individuals and further drop one small outlier city that has average earnings greater than 150% of the national average. This selection process leaves me with 66 cities.

I construct two additional samples for estimation purposes: the stayers sample and the movers sample. The selection procedure follows LMS. For the stayers sample, I only include 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. For the movers sample, I include workers who switch firms in any year. Following Kline et al. (2020), I restrict the movers sample to firms with at least two movers, which helps mitigate limited mobility bias. See more details on the data and sample selection in Appendix A and the summary statistics in Table J.1.

4.2 Additional assumptions

4.2.1 Discretization

I follow BLM and restrict firm technology parameters (z_j, θ_j) and amenity parameters $\mathbf{G}_j = \{G_j(a)\}_a$ to be drawn from a discrete distribution. I refer to each set of firms with the same $\{z, \theta, \mathbf{G}\}$ as a firm cluster, indexed by $k \in \{1, 2, 3, ..., K\}$. The classification of firms into clusters helps mitigate limited mobility bias, which is prevalent in the AKM estimation using the movers design (Andrews et al., 2008).

4.2.2 Stochastic processes

I introduce time-varying shocks to the model to utilize the panel feature of the matched employer-employee dataset for identification. I restate key model variables with these shocks in Appendix F.1. 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) and LMS). Changes in workers' preferences can lead to movements across firms and cities. Note that although equation (4.1) adds the time dimension to the preference shock in equation (2.2), it does not restrict the time-series properties of ϵ.

Assumption 3. *The skill of a worker i at time t, a_{it}, contains a permanent skill component a_i and a transient skill shock $\hat{a}_{it}$, where $\hat{a}_{it}$ follows a stationary mean-zero stochastic process that is 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 is does not affect workers' preference for firm and city amenities, that is

$$G_j(a_{it}) = G_j(a_i),\ \ R_c(a_{it}) = R_c(a_i). \tag{4.3}$$

Following LMS, I assume that the transient skill shock does not interact with the skill-augmenting productivity or affect workers' preferences for non-wage amenities. These restrictions imply the transient skill shocks generate earnings changes but do not cause movement across firms or cities. I next introduce productivity shocks at the firm and city level.

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

$$\log z_{jt} = \log z_j + \log \hat{z}_{jt} \tag{4.4}$$

$$\log A_{ct} = \log A_c + \log \hat{A}_{ct} \tag{4.5}$$

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, $\{G_k(a), R_c(a)\}$, do not change over time.

The dynamic productivity processes follow the well-known works on production function estimation in the industrial organization literature (e.g., Olley and Pakes (1996)). These productivity shocks generate labor demand shocks, which are necessary for identifying labor supply elasticities at both the city and firm levels. I assume that the time-varying worker skill shocks and firm and city productivity shocks are independent.

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.*

Finally, I introduce the measurement errors in the observed firm wage bills.

Assumption 6. *The observed wage bill of firm j in the data $\dot{E}_{jt}$ are related to their counterparts in the model E_{jt} with a measurement error e_{jt}:*

$$\log E_{jt} = \log \dot{E}_{jt} + e_{jt} \tag{4.6}$$

where the measurement error for wage bill follows a $MA(q)$ process given by $e_{jt} = \sum_{s=0}^{q} \delta_s u^e_{j(t-s)}$, and u^e_{jt} is i.i.d. across firms and time.

4.3 Identification Strategy

4.3.1 Sorting elasticity parameters

Worker sorting elasticity parameters

I identify the workers sorting elasticity across firms, β_w/ρ_w, and the elasticity across cities, β_w, using the passthrough design as in Kline et al. (2019) and LMS. This empirical design exploits wage bill shocks and identifies the elasticity parameters through changes in earnings of the worker stayers. I define $\bar{E}_{ct}$ and $\bar{W}_{ct}$ as city-level mean firm wage bills and mean worker earnings, and $\log \hat{\dot{E}}_{jt} \equiv \log \dot{E}_{jt} - \log \bar{\dot{E}}_{ct}$ and $\log \hat{W}_{ijt} \equiv \log W_{ijt} - \log \bar{W}_{ct}$ as the residualized wage bill and earnings.[15] As I show in Appendix F.2, the changes in the mean variables isolate city-level shocks, and the changes in residualized variables isolate firm-level shocks.

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:

$$\Delta \log \hat{W}_{ijt} = \delta_w \Delta \log \hat{\dot{E}}_{jt} + \Delta \hat{a}_{it} + \delta_w \left(\Delta e_{jt} - \Delta \log \frac{\hat{\phi}_{jct}}{\hat{\phi}_{ct}} \right). \tag{4.7}$$

The extent of passthrough from the firm-level shock $\Delta \log \hat{\dot{E}}_{jt}$ to workers stayers' earnings $\Delta \log \hat{W}_{ijt}$ is controlled by the firm-level labor supply elasticity β_w/ρ_w. Suppose that a firm experiences a positive productivity shock that increases its labor demand. If the labor supply is elastic, the firm can expand employment with smaller increases in earnings, resulting in less rent sharing for workers who stay at the firm. Moreover, equation (4.7) makes clear that estimating the passthrough elasticity requires using the earnings changes of the stayers, as the movers' earnings changes are associated with movement across firms and/or cities.

There are three residual terms in the net passthrough equation (4.7), which are the changes in i.i.d. transient skill shock $\Delta \hat{a}_{it}$, changes in the firm wage bill measurement error Δe_{jt}, and changes in the relative supply shifter $\Delta \log(\hat{\phi}_{jct}/\hat{\phi}_{ct})$. The last term relates to changes in the set of firms in the city c, which affects the firm-level labor supply curves (2.3) through the wage index $\mathbb{W}_c(a)$. The potential correlation of the wage bill shock with the measurement error and with the relative supply shifter leads to two endogeneity concerns.

To deal with the first concern, I follow LMS to instrument the net wage bill shock $\log \Delta \hat{\dot{E}}_{jt}$ with its lags before year $t-q-1$ The lagged shocks are correlated with the current shock as firm-level productivity shocks are persistent, and they are uncorrelated with contemporaneous measurement errors, which are assumed to follow $MA(q)$. To deal with the second concern, I apply 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, the same as the dynamic process of city producitivity A_c, I non-parametrically control for $\Delta \log \hat{W}_{ijt-1}$ and $\Delta \log \hat{\dot{E}}_{jt-1}$. In practice, this is implemented by including a cubic polynomial of these lagged variables. See Appendix

[15]To partial out the life-cycle component of the earnings profile and macroeconomic shocks from worker earnings, I run a Mincer-type regression of log worker earnings on an third-order age polynomial and year dummies, where I follow Card et al. (2013) in restricting the earnings-age profile to be flat at the age of 40. The residuals from this regression are then retrieved as $\log W_{ijt}$.

F.2 for more details.

The city-level passthrough parameter δ_c is identified from the city-level regression:

$$\Delta \log \left(\frac{\bar{W}_{ct}}{r_{ct}^{\eta}} \right) = \delta_c \Delta \log \left(\frac{\bar{\dot{E}}_{ct}}{r_{ct}^{\eta}} \right) + \delta_c \left(\Delta e_{ct} - \Delta \log \hat{\phi}_{ct} \right), \tag{4.8}$$

where e_{ct} is introduced as the measurement error of city-level mean firm wage bills, which I assumed to follow the same $MA(q)$ process as e_{jt}. Notably, workers' location choices are affected by changes in real, not nominal, wages. City-level productivity shocks are capitalized into housing rents, thereby dampening population responses. Hence, estimating the city-level equation using nominal earnings can upward bias the city-level labor supply elasticity.

I employ the same strategy as the firm-level passthrough estimation to address endogeneity concerns that come from Δe_{ct} and $\log \hat{\phi}_{ct}$. Another concern is that city-level wage bill shocks may be correlated with changes in amenities (e.g., Diamond (2016)). To address this, I include a specification where I non-parametrically control for changes in the share of high-skilled workers as a proxy for potential amenity shifts driven by the endogenous amenity channel.[16]

Firm sorting elasticity parameter

Analogously, the firm sorting elasticity parameter β_f is identified from the passthrough of city total wage bill shocks to mean wage bill of existing firms in the city:

$$\Delta \log \bar{E}_{ct} = \delta_f \Delta \log \dot{E}_{ct} + \delta_f \left(\Delta e_{ct} - \Delta \log \frac{\hat{\phi}_{ct}}{\Phi_{ct}} \right) \tag{4.9}$$

where I define the firm passthrough parameter $\delta_f \equiv \frac{1}{1+\beta_f}$. This passthrough regression is estimated using a sample of firms that operate in a city for at least 7 years and employ at least 10 workers in each year. Equation (4.9) utilizes city-level productivity shocks to estimate the responsiveness of firm location decisions. The identification assumption is that the share of firms in each city adjusts to city productivity every year according to equation (2.12), which can be achieved through systematic firm entry and exit. To address the endogeneity concerns associated with the measurement error and labor supply shifter, I employ the same strategy as before. See Appendix F.3 for more details.

[16] Here, I measure worker skill using the fixed effect from a reduced-form AKM regression of log earnings on worker and firm fixed effects, along with a third-order polynomial in worker age. High-skilled workers are defined as those in the top three deciles of this fixed effect distribution.

4.3.2 Productivity and amenity parameters

I now present the identification strategy for the productivity and amenity parameters. For the productivity parameters, I leverage the movers design to identify worker skill a and firm skill-augmenting productivity θ following BLM. I exploit firms' revealed location decisions to identify city productivity A, drawing on the concept of compensating differentials from the Rosen-Roback model. The remaining part of the worker earnings is attributed to firm productivity z.[17] For the amenity parameters, I estimate them by matching the observed worker sorting shares across cities and firms conditional on (real) worker earnings.

To address this issue of limited mobility bias, I follow BLM to classify firms based on moments of the empirical earnings distribution. However, heterogeneous city productivity complicates this process. Since both firm and city productivity affect workers' earnings, firms in different cities may exhibit similar earnings distributions despite differing in firm productivity. To capture true firm heterogeneity, the clustering approach must account for local productivity determinants when classifying firms.

To make progress, I develop an iterative estimation procedure that involves guessing and updating the vector of city composite productivity, $\mathbb{A}_c \equiv A_c \bar{L}_c^{\mu}$, which encapsulates both exogenous and endogenous components of city productivity. I estimate the agglomeration elasticity μ from $\mathbb{A}_c$ after convergence. In what follows, I describe each step of this procedure.

Worker skill and firm productivity

In Appendix F.4, I construct a measure of adjusted log earnings, denoted by $\log \check{W}_{ijt}$, by netting out the city-level earnings determinants and time-varying firm-level and city-level shocks. When the guess is correct, the adjusted log earnings $\log \check{W}_{ijt}$ is only affected by worker permanent skills a_i, firm technology (z_j, θ_j), and the transient skill shocks $\hat{a}_{it}$:

$$\log \check{W}_{ijt} = \log \chi + \underbrace{\log z_j}_{\text{firm productivity}} + \underbrace{\theta_j \log a_i}_{\text{worker-firm interaction}} + \hat{a}_{i,t}, \tag{4.10}$$

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, $\log \check{W}_{ijt}$, as inputs.[18] Having clustered firms,

[17] It is important to note that relying solely on worker movers does not enable separate identification of firm productivity and city productivity, even with between-city movers in the data.

[18] Specifically, 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, $K = 20, 30, 40, 50$ (for simplicity, I set $\mathbb{A} = 0, \forall c$ and $\theta_k = 0, \forall k$ in this robustness check). Despite the parsimony, the 10 (50) cluster fixed effects account for 87 (90) percent of the between-firm variance in log earnings.

the moment condition for identifying the cluster-level productivity parameters $\{z_k, \theta_k\}_{\forall k}$ is:

$$\mathbb{E}\left[\left(\frac{\log \bar{W}_{ij(t+1)}}{\theta_{k'}} - \frac{\log z_{k'}}{\theta_{k'}}\right) - \left(\frac{\log \bar{W}_{ijt}}{\theta_k} - \frac{\log z_k}{\theta_k}\right) \mid k \neq k'\right] = 0. \tag{4.11}$$

Equation (4.11) indicates that $\{z_k, \theta_k\}_k$ are identified with wage changes of between-cluster movers, regardless of whether the move happens within or between cities. The identification assumption is that workers' across-firm movements are uncorrelated with unobserved skill shocks, which is satisfied with the assumptions of $\hat{a}_{it}$ in Assumption 3. Equation (4.11) 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. $\mathbb{E}_{kk'}(a) \neq \mathbb{E}_{k'k}(a)$.[19] I show in Figure J.7 that such an asymmetry is supported by data. 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 (4.11) is that worker movements between firms are not correlated with transitory skill shocks, known as the "exogenous mobility" assumption. Violations may arise from negative earnings shocks before moves (i.e., Ashenfelter dip) or selection in unobserved match-specific productivity. To assess this, I follow Card et al. (2013) and construct an event-study figure that group movers by origin-destination firm clusters. As shown in Figure J.5, the earnings trends before moves are parallel and the changes in log earnings between the firm groups are nearly symmetric, supporting the exogenous mobility assumption.[20]

Firm and city amenities

I estimate the firm and city amenity parameters based on the observed worker distribution across firms and cities, conditional on their real earnings. Specifically, I show in Appendix F.5 that:

$$R_c(a) \cdot G_k(a) = \frac{\bar{r}_c^{\eta}}{\mathbb{A}_c r_c^{\frac{\alpha}{\alpha-1}} \cdot z_k a^{\theta_k}} \cdot \bar{J}_c(k)^{\frac{\beta_w}{\rho_w}} \cdot \Lambda_{kc}(a)^{\frac{\rho_w}{\beta_w}} \Lambda_c(a)^{\frac{1}{\beta_w}}, \ \forall a, k, c \tag{4.12}$$

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 employed in cluster-k firms located in city c during the sample period, conditional on residing

[19]BLM clarify that documenting symmetric earnings gains and losses is not sufficient to reject the existence of complementarity. See their Online Supplement Section S2 for more details.

[20]Borovičková and Shimer (2024) demonstrate that, in a dynamic framework with search frictions, selective hiring and worker mobility can lead to assortative matching and endogenous mobility, the latter of which cannot be detected by an event-study test.

in city c; $\Lambda_c(a)$ is the share of skill-a workers in city c. Intuitively, a firm (or city) is inferred to have higher amenities if it attracts more workers despite offering lower nominal (real) earnings. Since the amenities are assumed to be time-invariant, all the empirical measures on the right-hand side are averaged over the sample period. Equation (4.12) thus provides $K \times C$ moments to identify $K + C$ amenity parameters for each skill level a, subject to one normalization per skill.

City composite productivity

Next, I leverage the revealed firm location choices to estimate city composite productivity $\mathbb{A}_c$, using the empirical counterpart of the firm sorting equation (2.12):

$$\bar{p}_c(k) = \frac{\left(\mathbb{A}_c \bar{r}_c^{\frac{\alpha}{\alpha-1}} \cdot \phi_{kc}\right)^{\beta_f}}{\sum_{c'} \left(\mathbb{A}_{c'} \bar{r}_{c'}^{\frac{\alpha}{\alpha-1}} \cdot \phi_{kc'}\right)^{\beta_f}}, \ \forall k, c \tag{4.13}$$

After clustering firms in the previous step, I can construct $\bar{p}_c(k)$, the average share of cluster-k firms located in city c, in the data. The labor composite term ϕ_{kc} can be constructed 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 $\{G_k(a)\}_{k,a}$, with the details shown in Appendix F.6. As discussed before, constructing ϕ_{kc} does not require knowing the city productivity. The labor composite term ϕ_{kc} captures the labor supply and competition conditions faced by a cluster-k firm in city c. Due to the variations in skill-augmenting productivities and skill-specific amenities across clusters, ϕ_{kc} varies across firm clusters for the city. All else equal, firms whose technology complements high-skilled workers and those who have greater skill-biased amenities will have higher ϕ in cities with more high-skilled workers, making these cities relatively more profitable locations for such firms.

Equation (4.13) thus provides $C \times K$ moments to identify C city composite productivity parameters, subject to one normalization. The identification assumption is that all types of firms perceive city productivity as identical. Consequently, the model cannot perfectly match the cluster-specific sorting shares $\bar{p}_c(k)$. Nonetheless, I will later show that the model matches the average firm productivity in each city reasonably well, supporting the identification assumption.

Discussions on the iterative procedure

Several points about the iterative procedure merit discussion. First, the separate identification of city and firm productivity parameters does not rely on the discretization assumption. If there were no limited mobility bias, neither grouping firms into clusters nor applying the iterative procedure would be necessary.

Second, the convergence of the iterative procedure is disciplined by the firm sorting shares observed in the data. Suppose that in one iteration, the city composite productivity of city c s guessed to be greater than its true value. As a result, the implied firm productivity z_j for firms in c would be smaller than their true values, indicating a less competitive labor market in city c and leading to higher $\{\phi_{kc}\}_k$. This overestimation makes city c appear more profitable for all firms, implying that it should attract more firms than observed in the data. The city composite productivity is subsequently adjusted downward to align with the observed firm sorting shares. The entire vector of city productivity is iteratively updated in this manner until convergence, when the spatial distribution of firms is rationalized.

Third, prevalent measurement errors in the data, particularly in smaller cities, complicate the estimation of city productivities and impede the convergence of the iterative estimation procedure. To address this, I smooth productivity estimates against city population using a third-order polynomial after each iteration until convergence. While this preserves the model's ability to match the spatial earnings structure, it artificially reduces the variance of the city productivity estimates, making AKM-style earnings variance decomposition unreliable.

Agglomeration elasticity and city productivity

I estimate the agglomeration elasticity μ using the converged city composite productivity:

$$\log \mathbb{A}_c = A_0 + \mu \log \bar{L}_c + \epsilon_c^A \tag{4.14}$$

where A_0 is the intercept reflecting the normalization, $\bar{L}_c$ is the average city-c population, and ϵ_c^A is the error term representing the city exogenous productivity. An OLS estimate of the parameter μ is biased as city population can be correlated with the unobserved city exogenous productivity. Thus, I use an immigration-based population shock as an instrument following Card (2001). See more details in Appendix F.7. City exogenous productivity can then be recovered as $A_c = \mathbb{A}_c \cdot \bar{L}_c^{-\mu}$.

4.3.3 Housing supply and demand parameters

Lastly, I estimate the housing supply elasticity by relating changes in the housing rent to changes in the housing expenditure:

$$\Delta\log r_{ct} = \Gamma \Delta \log EH_{ct} + \Delta e_{ct}^r \tag{4.15}$$

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. A large Γ indicates that housing rent is highly responsive to changes in housing demand, reflecting low housing supply elasticity. Equation

(4.15) is estimated using five-year city-level changes between 2002 and 2007. I choose this period to avoid the 2008-2009 housing bust and the subsequent surge in foreign investment in the Canadian real estate market.

Changes in housing expenditure may be correlated with uobsered local shocks that also affect housing rent. To address the endogeneity concerns, I follow Diamond (2016) and use a shift-share bartik instrument to instrument for $\Delta \log EH_{ct}$. Then, I calibrate $\bar{H}_c^0$ to match the average housing rent of each city. Following Saiz (2010), I include an interaction between $\Delta \log EH_{ct}$ and the share of undevelopable land to allow γ to vary across cities. See more details in Appendix F.8.

For the parameters governing housing demand, I calibrate the worker share of expenditure on housing $\eta = 0.24$ following Davis and Ortalo-Magné (2011),[21] and the firm share on housing $\alpha = 0.06$ using information on firms' housing and wage bill expenditures.[22]

4.4 Estimation results

Sorting elasticity and housing supply parameters

Here I discuss the estimation results of the sorting elasticity and housing supply elasticity parameters. The estimated parameter values using the preferred specifications and the calibrated housing parameters are summarized in Table 1.

I report the passthrough estimates for workers in Table J.8. For the firm-level passthrough parameter, the IV estimate in column (2) is 0.13, which is about half of the OLS estimate in column (1). When controlling the lagged variables to proxy for changes in the labor demand condition (i.e., $\Delta \log(\hat{\phi}_{jct}/\hat{\phi}_{ct})$), as shown in column (3), the estimate increases slightly to 0.14. This passthrough elasticity indicates a firm-level labor supply elasticity of 6.1, within the range of estimates reported in the literature (e.g. Card et al. (2018) and LMS).[23] For the city-level passthrough parameter, the OLS estimate of 0.28 in column (1) and the IV estimate of 0.30 in column (2) are quite similar. Controlling for changes in the labor demand condition (i.e. $\Delta \log \hat{\phi}_{ct}$), reported in column (6), reduces the estimate to 0.25. This occurs because changes in average real wage bill are negatively correlated with $\Delta \log \hat{\phi}_{ct}$, which

[21] I have also calibrated heterogeneous housing expenditure shares for the counterfactual analysis. In these cases, I follow Eeckhout et al. (2014) and calibrate $\eta = 0.35$ for low-skilled workers and $\eta = 0.22$ for high-skilled workers.

[22] 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). Only the largest firms report the value of buildings and the real estate rental expenses in the NALMF data, so I use the subset of firms with non-zero total housing services to calibrate α.

[23] In a restricted specification that imposes $\rho_w = 1$, that is to assume uncorrelated preference draws within cities, the estimated passthrough elasticity is 0.15. This aligns with labor supply elasticity estimates that do not explicitly account for the spatial dimension of labor supply (e.g. Card et al. (2018)).

Table 1: Sorting elasticity and housing-related parameter estimates

Parameter	Description	Value (S.E.)	Method
β_w/ρ_w	Firm-level labor supply elasticity	6.1 (1.0)	Equation (4.7)
β_w	City-level labor supply elasticity	2.1 (0.03)	Equation (4.8)
β_f	Firm sorting elasticity	5.6 (1.0)	Equation (4.9)
$1/\gamma$	Housing supply elasticity	1.2 (2.0)	Equation (4.15)
α	Production elasticity of housing	0.06	Value of housing services/VA
η	Expenditure share on housing	0.24	Davis and Ortalo-Magné (2011)

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

enters the structural error term with a negative sign. Further controlling for changes in shares of high-skilled workers, reported in column (7), increases the estimate to 0.32. This is because unobserved amenity changes are positively correlated with real wage bill changes, higher amenities lower the real wages required to attract and retain workers in a city. The city-level labor supply elasticity implied by the estimated parameter in column (7) is 2.1, which is within the range of 0.5–5 reported in the literature (e.g. Suárez Serrato and Zidar (2016); Hornbeck and Moretti (2024); Bilal (2023)).

I present the passthrough estimates for firms in Table J.10. The IV estimate in column (2) is 0.21, substantially smaller than the OLS estimate of 1.06 in column (1), indicating the presence of substantial measurement errors that are strongly and positively correlated with measured city-level total wage bill. In column (3), after accounting for changes in labor competition, the passthrough estimate further declines to 0.15. The inferred firm sorting elasticity of 5.6 from column (3) falls within the range reported by Gaubert (2018) for different tradable industries, although slightly higher than the estimates in Suárez Serrato and Zidar (2016) and Fajgelbaum et al. (2019).

I present the estimates of housing supply elasticity in Table J.13. The OLS estimate in column (1) is 0.53, whereas the IV estimate in column (2) decreases slightly to 0.45 and loses statistical significance, likely due to the instrument's limited statistical power (F statistics is 24.7 in the first stage). This IV estimate of Γ implies a housing supply elasticity $1/\gamma = 1.25$, which close to the average elasticity 1.5 estimated by Saiz (2010). In column (3), I allow the elasticity to vary with the share of undevelopable land but find no significant effect of topography differences. Consequently, I will use a homogeneous housing supply elasticity across Canadian cities in the counterfactual analysis.

Productivity parameters

I report the estimates of the two firm productivity parameters by cluster in Table J.11. I impose a normalization such that $z = \theta = 1$ for the firm cluster with the lowest average earnings. The estimation results reveal significant variation in common productivity (z) and skill-augmenting productivity (θ) across clusters. The two productivity parameters exhibit an almost perfect correlation, with a coefficient of 0.97, aligning with the findings in LMS. This strong correlation between the common and skill-augmenting productivity parameters provides strong evidence of worker-firm production complementarity.

The estimation procedure generates a continuous distribution of worker skills. In practice, I rank workers by skill and divide them into 100 equal-sized groups. To better capture the spatial sorting of top-skilled workers—a key driver of the urban earnings premium—I further split the top percentile into 21 equal-sized groups, resulting in a total of 120 worker groups. To validate the skill estimates, Figure J.11 shows that the model-implied shares of high-skilled workers in each city are strongly correlated with the shares of college-educated workers from the 2016 Census data.

I plot city productivity estimates against log city population in Figure J.8. The results indicate that larger cities, on average, have limited exogenous productivity advantages over smaller ones. I correlate estimated city productivity with the location characteristics in Table J.14, finding that cities with warmer summers tend to be more productive, and those in eastern and northern Canada tend to have greater productivity advantages.

As reported in Table J.12, the agglomeration elasticity μ is estimated to be minimal and statistically insignificant. These results suggest that local productivity determinants play a minor role in explaining the urban earnings premium after accounting for worker and firm heterogeneity.

Amenity parameters

I present the firm and city amenity estimates, $\{G_k(a), R_c(a)\}$, in Figure J.9. Panel (a) shows that high-productivity 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). I correlate estimated city amenities with the location characteristics in Table J.14, revealing that Canadian workers prefer warm winters, breezy summers, safe streets, and mountainous landscapes.

4.5 Model fit

I now assess the 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 as observed 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. Therefore, it is crucial to assess how well the model solution aligns with key sorting and inequality measures in the data.

Figure J.10 confirms the model's ability to match key patterns in the data. Panels (a) and (b) demonstrate that the model closely reproduces the city-level mean worker skill and firm common productivity, though the fit for mean firm productivity is slightly noisier, as expected. Panel (c) shows that the model captures the within-city matching pattern well, measured as the share of the top 30% of skilled workers matched to 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, which is slightly below the data estimate of 0.021. It also shows that the model does a better job at matching average earnings and population for larger cities than smaller ones.

4.6 Discussions

Skill sorting and learning. The model is agnostic about the origin of worker skill, which may be from innate abilities, education, and work experience. Baum-Snow and Pavan (2013) 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.[24] This paper does not seek to distinguish between static sorting and dynamic learning. Rather, the objective is to characterize cross-sectional worker and firm heterogeneity across cities and to evaluate the impact of place-based policies given the spatial distributions of heterogeneous workers and firms. Accordingly, the agglomeration elasticity μ only captures the static economies of scale. Given that higher-skilled workers and higher-productivity firms benefit more from learning spillovers (Davis and Dingel, 2019; Baum-Snow et al., 2024), an optimal spatial policy that accounts for these dynamic gains would promote even stronger spatial sorting.

Endogenous labor market power. Recent work suggests that large cities have more competitive labor markets, resulting in smaller wage markdown and higher earnings (e.g. Hirsch et al. (2022)). The model abstract from endogenous labor market power, which

[24] Controlling for workers' big city work experience only slightly decreases the estimate of the urban earnings premium, which is shown in Figure J.3.

would further complicate the location choices of workers and firms. As a robustness check, I separately estimate the firm-level passthrough parameter for the largest five cities and the smaller cities in Table J.9. The results show no significant differences in the degree of passthrough between the two city groups, suggesting that spatial variation in monopsony power has a limited impact on spatial inequality in the Canadian context.

Multi-location firms. There is growing interest in the location choices of multi-location firms (e.g., Oberfield et al. (2024); Kleinman (2022)), which is beyond the scope of this paper. Since establishments within a firm share non-rival inputs like managerial expertise and intangible capital, their productivities should be highly correlated. As a validation of my estimation strategy, which makes no assumptions about such correlations, I decompose productivity variance into between- and within-firm components for multi-city firms and find that the between-firm component accounts for 89.7% of the total variance. This decomposition result confirms the strong within-firm correlation while highlighting the model's ability to capture within-firm, between-establishment heterogeneity.

5 Quantitative Analysis

In this section, I conduct three quantitative analyses using the estimated model: (1) a structural decomposition of the urban earnings premium, (2) implementation of the optimal spatial policy, and (3) evaluation of place-based subsidies.

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:

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

The decomposition takes the spatial allocations of workers and firms, as well as the earnings distribution, as given; thus, it does not require solving the model. The interaction component captures the contribution of assortative matching via skill-augmenting productivity to city mean log earnings. Accordingly, the worker component is evaluated at the average skill-augmenting productivity, while the firm component's skill-augmenting part is evaluated at the national average worker skill. To evaluate the contributions of each component to the urban earnings premium, I regress city mean log earnings and each right-hand-side compo-

Table 2: 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	66	66	66	66
R^2	0.100	0.071	0.023	0.010	0.000

Note: 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.

nent on log city population, with the results reported in Table 2.[25]

Table 2 is the first in the literature to decompose the spatial earnings differential into location characteristics, two-sided heterogeneity, and local matching components. The results indicate that worker and firm sorting collectively account for over 90% of the urban earnings premium, contributing 66.8% and 26.8%, respectively. This highlights the key role of two-sided sorting in explaining spatial earnings inequality. The greater contribution of worker sorting compared to firm sorting is consistent with prior studies applying AKM regressions to decompose the urban earnings premium (Dauth et al., 2022; Card et al., 2025). The city component accounts for only 6.6% of the premium. Not only are the productivity advantages of larger cities limited, but these benefits are also capitalized into higher housing rents, which in turn reduces the marginal product of labor. Perhaps surprisingly, the interaction component shows no correlation with city population.[26] However, I will demonstrate next that worker-firm complementarity is crucial for driving the co-location pattern and the spatial earnings structure.

Admittedly, three of the four coefficients in Table 2 are statistically insignificant. This

[25] I also perform a decomposition of within-city inequality and correlate the variance components with population. The results, shown in Table J.6, reveal that higher earnings dispersion in larger cities is primarily driven by greater worker skill variance and stronger covariances between worker skill and firm common productivity.

[26] This result is consistent with the findings of BLM and LMS, who also show that the interaction effect has a minimal impact on earnings variance. As illustrated in Figure J.12, the variation in the earnings profile driven by skill-augmenting productivity (θ) is much smaller than that driven by firm productivity (z). While the variation in θ drives the assortative matching with elastic labor supply, the interaction component remains negligible and does not vary across cities.

is likely due to the measurement errors and the limited number of cities in the sample. Moreover, the urban earnings premium in Canada is estimated to be 0.021, which is relatively modest compared to other developed economies, including 0.067 in the U.S. (Albouy et al., 2019), 0.037 in Germany (Dauth et al., 2022), 0.049 in France (Combes et al., 2008), and 0.045 in Spain (De La Roca and Puga, 2017). This limited variation makes it challenging to precisely identify the contributions of different components.

5.1.1 Understanding spatial sorting using a Shapley value approach

The statistical decomposition in the previous section is based on the spatial allocation of workers and firms in equilibrium. I now employ the Shapley value approach to shed light on the underlying drivers for the observed sorting and spatial inequality patterns.

In the model, spatial sorting and earnings disparities are driven by variations in three key parameter groups. First are A_c, μ, and α that determine each city's real productivity. Second are the skill-specific amenity parameters $R_c(a)$ and $G_k(a)$ that affect worker sorting through compensating differentials. Third are the firm skill-augmenting parameters θ_k that determine the benefits of assortative matching. Removing variation in these parameters would eliminate spatial sorting and earnings differentials.[27]

These parameters interact to jointly shape equilibrium outcomes. Eliminating the variation in any parameter also involves removing its covariance with other parameters, meaning the resulting change in equilibrium outcomes cannot be attributed solely to that parameter. To disentangle these interaction effects, I apply the Shapley value approach, which involves simulating counterfactual economies by eliminating all combinations of parameter variations. Intuitively, the Shapley value measures the average change in an equilibrium outcome when a specific variation is removed, considering all combinations of remaining variations. More details on the Shapley value approach are discussed in Appendix H. I analyze four equilibrium outcomes: the urban earnings premium (β^w), the worker and firm sorting gradients (β^a and β^z), and the covariance of city mean worker skill and firm productivity ($\text{Cov}(\bar{a}_c, \bar{z}_c)$). Table 3 presents the results, detailing the percentage contribution of each parameter to these outcomes.

Column (1) shows that 17.0% of the urban earnings premium β^w is explained by city characteristics. Among these, city exogenous productivity A_c and the agglomeration elasticity μ contribute positively, by 13.2% and 25.9%, respectively, while the housing share parameter α contributes negatively, at -22.1%. This negative value means that if production did not require housing input ($\alpha = 0$), the urban earnings premium would increase by

[27] Such an equilibrium still preserves a meaningful city-size distribution by retaining variation in $\bar{H}_c$. In addition, there is no between-city variation in μ and α. I set them to zero to compute their respective Shapley values.

Table 3: Shapley decomposition of the urban earnings premium and spatial sorting measures

	Urban Premium	Spatial Sorting		Co-location
		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%

Note: This table displays the Shapley value decomposition results of four equilibrium outcomes. The shares in the row named *city prod. characteristics* sum up the shares of the three parameters below, and the shares in the row named *amenities* sum up the shares of the parameters below. See Section 5.1.1 for a detailed description of the method.

22.1%. Amenity parameters collectively account for 42.3% of the premium, with city amenities $R_c(a)$ contributing 71.4% and firm amenities $G_k(a)$ contributing −29.1%. Larger cities provide better amenities for high-skilled workers, while more productive firms offer worse amenities for the same group, influencing sorting patterns and the urban premium. Heterogeneity in skill-augmenting productivity θ_k account for 40.6% of the premium, primarily by driving the co-location of high-skilled workers and high-productivity firms in larger cities, as I demonstrate next.

Columns (2)–(3) reveal that the worker skill gradient β^a is primarily driven by urban amenities in larger cities. Meanwhile, skill-augmenting productivity θ_k has little effect on β^a but significantly impacts the firm sorting gradient β^z. Eliminating the variation in city productivity characteristics increases sorting gradients by making the city population distribution less dispersed. Finally, column (4) shows that amenities account for 60.4% of the co-location covariance $\text{Cov}(\bar{a}_c, \bar{z}_c)$, while production complementarity contributes 44.6%.

These decomposition results provide two key insights. 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 2. 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 economic patterns.

5.2 Optimal spatial policy

In this section, I present 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 minimize disparities in the marginal utility of consumption, and 3) redistributing between types depending on welfare weights.

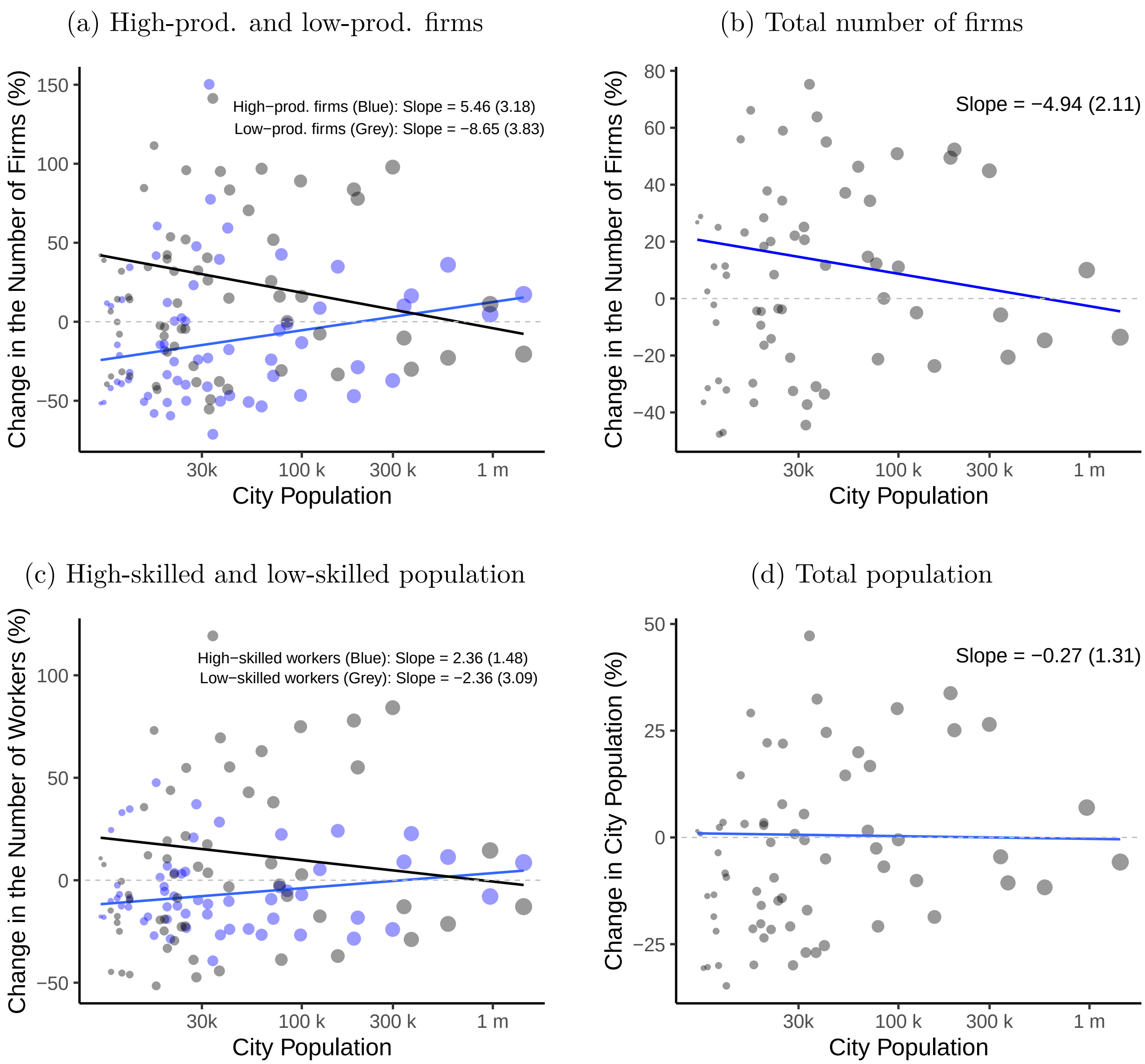
In what follows, I first present the results of a utilitarian government that sets $\varphi^w(a) = 1, \forall a$ and $\varphi^f(k) = 1, \forall k$, followed by results using alternative welfare weights.[28] The results should be viewed as a first-best, long-run optimal policy: I assume that the government can fully observe the full distribution of worker and firm characteristics and has access to the flexible tax instruments, and I abstract from short-term moving frictions.

Sorting efficiency and spatial reallocation. Figure 1 plots the changes in the number of firms and workers resulting from the optimal policy against the initial city population. Panel (a) differentiates between high- and low-productivity firms, classifying the three firm clusters with the highest z as high-productivity. Panel (c) distinguishes high- and low-skilled workers, defining those in the top three skill deciles as high-skilled. Panels (b) and (d) show the changes in the total number of firms and workers in each city, respectively.

As shown in Panel (a) of Figure 1, the most salient outcome of the optimal policy is that it incentivizes more high-productivity firms into larger and more skilled cities, while allocating low-productivity firms away from these places. With the optimal profit taxes, firms internalize the labor market stealing and love-of-variety externalities. For low-productivity firms, the negative labor market stealing externality is greater (in the absolute term) in large cities as they compete for high-skilled workers with high-productivity firms, whereas the positive love-of-variety is smaller in large cities as they employ fewer workers. High-productivity firms in large cities also incur labor market stealing externalities, but these costs are outweighed by their social benefits. They complement high-skilled workers, employ more people, and generate substantial love-of-variety benefits. Furthermore, their increased presence in larger cities attracts additional high-skilled workers, aligning with the planner's efficiency goals as shown in Equation (3.4). In sum, the optimal policy reduces the total number of firms yet increases average firm productivity in larger cities. These findings align with Bilal (2023), who finds it optimal to reallocate low-productivity firms towards smaller cities, and with Gaubert (2018), who demonstrates that the optimal policy enhances firm spatial sorting. I

[28] I assume that the social welfare function is sufficiently concave so that there is a unique equilibrium when implementing the policy. See Fajgelbaum and Gaubert (2020) for a formal analysis of equilibrium uniqueness under optimal spatial policies.

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



Note: 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. See the text for the definition of high-productivity firms and high-skilled workers. Population-weighted regression coefficients and standard errors are reported.

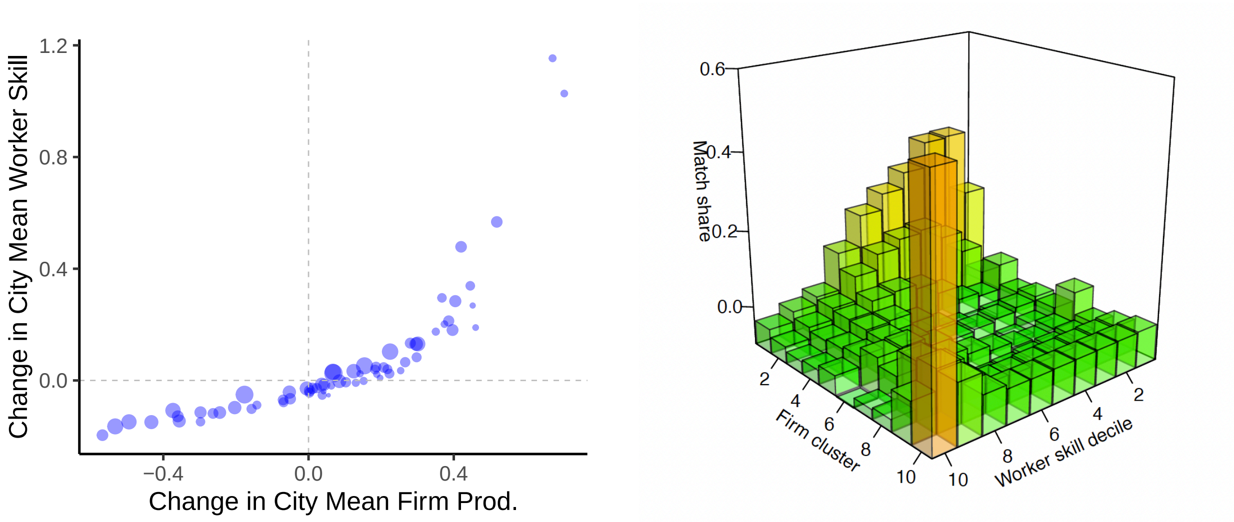
contribute to these studies by considering the role of two-sided heterogeneity.

Panel (c) of Figure 1 shows an increase in skill sorting, though not as pronounced as the change in firm sorting. Panel (d) shows that population in larger cities tend to shrink slightly. This is expected, given that agglomeration spillovers are estimated to be small. As a result, the optimal policy does little to incentivize the relocation of workers to higher-wage, larger

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

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

(b) Changes in worker-firm match shares



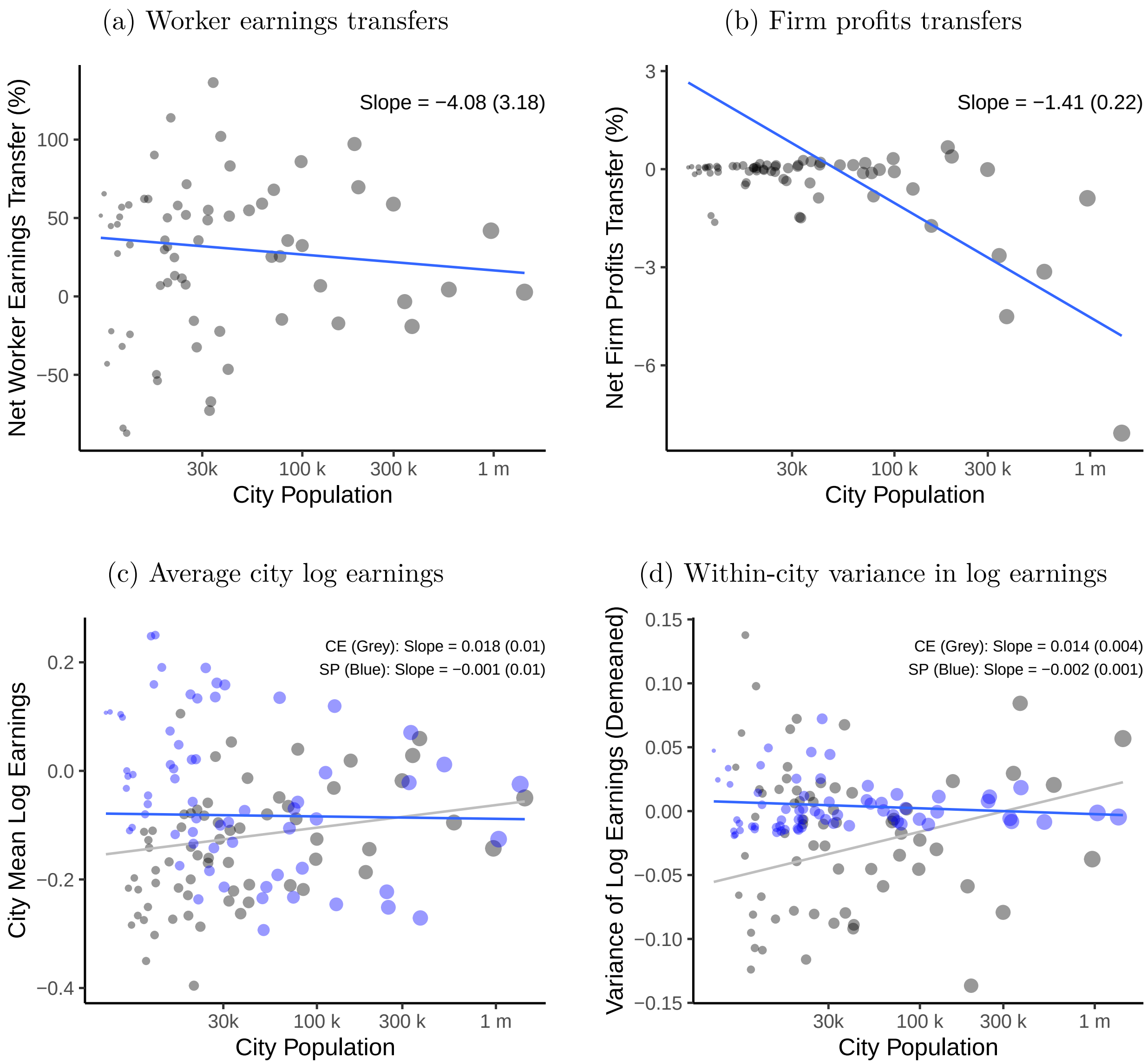
Note: These figures show how the optimal policy changes between-city spatial sorting and within-city assortative matching. Panel (a) displays the relationship between city-level changes in mean worker skill and mean firm productivity. Panel (b) displayed changes in the share of workers of each skill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

cities. Moreover, as I will soon show, the government implements spatial transfers from larger to smaller cities, further reducing the incentives for workers to sort into large cities. Although population changes are not strongly correlated with city size, there is significant worker reallocation across cities: seven cities see their populations grow by more than 25%, while nine cities experience declines of more more than 25%.

In Figure 2, I investigate how the optimal policy affects spatial sorting and worker-firm matching. Panel (a) highlights an increased co-location of high-productivity firms and high-skilled workers. The population-weighted covariance of $\bar{a}_c$ and $\bar{z}_c$ rises by 68.3% compared with the laissez-faire equilibrium, with a rank correlation of 0.98 between their changes. Together with Figure J.13, these results show that the optimal policy increases the sorting of high-skilled workers and high-productivity firms into high-productivity cities. Panel (b) shows a substantial increase in the assortativeness of worker-firm matching. Notably, the share of top-decile workers matched with the most productive cluster increases by roughly 50 percentage points, while low-skilled workers are more likely to match with low-productivity firms.

Overall, Figures 1 and 2 indicate that the optimal policy significantly increases assortative co-location and matching of workers and firms. By better leveraging production

Figure 3: Spatial transfers and changes in spatial inequality



Note: 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.

complementarities, this spatial reallocation increases total economic output by 9.6%. However, without additional interventions, intensified spatial sorting would exacerbate spatial inequality, negatively impacting social welfare by increasing disparities in the marginal utility of consumption. I show next that the optimal policy addresses this issue using spatial transfers.

Transfers and changes in spatial inequality. In panels (a)–(b) of Figure 3, I plot the share of net transfers relative to total income against city population, and I do so for workers and firms separately. 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. I find that the optimal policy significantly redistributes from larger, high-wage cities to smaller, low-wage ones, thereby promoting equity across 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.[29] 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.

Social welfare. The optimal policy results in a 47.4% increase in consumption-equivalent welfare. This substantial improvement is primarily driven by the equity motive. In the laissez-faire equilibrium, the marginal utility of consumption is considerably higher for low-income agents than for high-income agents. Consequently, 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 Appendix D.5 for details). This exercise results in a welfare gain of 4.9% for all types of agents in the economy, a magnitude comparable to the 4.0% welfare gain from the optimal spatial policy in Fajgelbaum and Gaubert (2020).[30]

Given the minimal estimated agglomeration spillover elasticity μ, the welfare improvements 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 welfare gain of 0.3% via this channel (0.6% for workers). The remaining welfare improvements are attributable to more efficient worker–firm matching achieved by correcting labor market stealing externalities.[31]

[29]The utilitarian government 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-skill transfers. To examine its varying impact across cities, I demean within-city log earnings variances in both equilibrium scenarios.

[30]Note that this welfare gain in Fajgelbaum and Gaubert (2020) contains the gain from within-type redistribution. 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 such a concern.

[31]It is challenging to fully disentangle the gains from correcting labor market stealing and love-of-variety externalities, as both are linked to the distribution of firm sizes, as shown in equation (3.4). The love-of-variety gain of 0.6% for workers 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.

5.2.1 Discussions

Alternative welfare weights. The degree of government redistribution among different types of workers and firms depends on the specified Pareto weights. As a robustness check, I implement the optimal spatial policy using two alternative sets of weights that limit between-type redistribution. First, I calibrate weights so that all agents experience equal welfare gains under the optimal policy.[32] Second, I employ Negishi weights, which are set as each agent type's average income in the laissez-faire equilibrium. The resulting welfare gains using these two sets of welfare weights are 5.7% and 5.2%, respectively, indicating 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 $\theta_j = \bar{\theta}, \forall j$. This removes the positive correlation between firm productivity z and skill-augmenting productivity θ, thus shutting down the firm-worker production complementarity. 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). As a result, the total welfare gain decreases from 47.4% to 45.7%.

The role of two-sided heterogeneity. Lastly, I show that it is important to account for both worker and firm heterogeneity for the optimal policy design. To show this, I implement the optimal policy under re-estimated models with only worker or firm heterogeneity. The results are shown in Figure J.14. When firms are assumed to be homogeneous, the optimal 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 optimal policy leads to a qualitatively similar reallocation. Quantitatively, however, the reallocation of firms is much stronger when workers are deemed to be homogeneous. This is because larger cities are estimated to be substantially more productive when worker heterogeneity is not taken into account, prompting greater spatial reallocation.

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 for 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 incorporating worker

[32] Calibrating the equal-gain weights for all 120 worker groups and firm clusters is computationally infeasible. In practice, I group workers into 20 categories and calibrate weights for 20 worker groups and 10 firm clusters.

Table 4: Counterfactual analysis: subsidizing productive firms in Toronto

			Limited Re-sorting		
	Full model	No worker het.	No-resorting	Only firm	Only worker
% Changes in	(1)	(2)	(3)	(4)	(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-skilled welfare	−0.2%	–	−0.1%	−0.2%	−0.1%
Low-skilled welfare	−0.5%	–	−0.2%	−0.5%	−0.2%
Pop. in Toronto	−2.3%	0.4%	–	–	0.0%
High-skilled pop.	0.6%	–	–	–	0.5%
Low-skilled 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%

and firm heterogeneity and mobility. The results are presented in Table 4: column (1) uses the full model, column (2) excludes worker heterogeneity, and columns (3)–(5) limit the mobility of firms, workers, or both.

In the full model (column (1)), the subsidy, which amounts to 0.3% of total GDP, has significant effects on the distributional and aggregate outcomes of the economy. In Toronto, the policy attracts more high-productivity firms, which grow by 30.1%, and more high-skilled workers, who grow by 0.6%. However, the resulting increase in labor market competition and a 2.8% rise in housing rents adversely impact local low-productivity firms and low-skilled workers, leading to their outflows of 7.7% and 3.8%, respectively. These counterfactual results align with the empirical findings in Qian and Tan (2021), who show that the entry of skill-intensive firms benefits local high-skilled workers but hurts low-skilled ones. As a result, Toronto's total population falls by 0.8%, and the number of firms falls by 2.3%. Nationally, aggregate output rises by 0.8%, which comes at a cost of greater spatial inequality and reduced welfare, particularly for low-skilled workers.[33]

Without worker heterogeneity (column (2)), the effects of the policy are much smaller. The subsidy would attract fewer high-productivity firms to Toronto and result in a smaller

[33]These counterfactual results are robust to incorporating endogenous amenities, free entry, and heterogeneous housing expenditure shares, which are shown in Table J.18.

increase in housing rent. The aggregate gains in total output are also limited, while spatial inequality and total welfare remain largely unchanged. Without spatial re-sorting (column (3)), the subsidy expands the employment of subsidized firms in Toronto, with no effect on other cities other than the taxes collected to finance the subsidy and thus minimal aggregate effects. When there is only one-sided re-sorting (columns (4)–(5)), the mobility responses of firms and workers are weaker than in the full model, leading to smaller changes in total output, welfare, and spatial inequality. Compared to the full model, these results highlight that accounting for two-sided heterogeneity and mobility is important for evaluating the effects of place-based subsidies

6 Conclusion

The economic fortune of a city is tightly linked to the types of workers and firms that it can attract. In many countries, high-skilled workers and high-productivity firms have been increasingly concentrated in a handful of successful cities, exacerbating the economic inequality between regions. Uncovering the interplay of workers' and firms' location decisions is thus important for understanding spatial inequality and designing place-based policies.

In this paper, I build a spatial equilibrium model with a system of cities and mobile heterogeneous workers and firms. I estimate the model using Canadian matched employer-employee data. I show that worker and firm sorting both play important roles in shaping spatial inequality and that production complementarity is crucial to explain the co-location of high-skilled workers and productive firms.

The model informs novel sources of sorting externality as a result of imperfect labor market competition and two-sided sorting. I design an optimal spatial policy that achieves efficient sorting and equitable spatial redistribution. I show 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.

I use the model to evaluate place-based policies, highlighting the crucial role of two-sided sorting in policy assessment. In conduct more policy counterfactuals in Appendix I, including spatial transfers to low-income cities and loosening housing supply constraints in productive cities. Beyond place-based interventions, the model can be used to study the spatial impact of economy-wide policies and technology changes. I show that the rise of remote work can substantially reduce spatial inequality, and that a universal basic income (UBI) policy encourages the relocation of workers and firms to smaller, more affordable cities.

The static equilibrium approach is a first step in understanding the cause and consequences of two-sided spatial sorting. However, it abstracts from dynamic considerations that significantly influence spatial sorting. For example, larger cities provide learning opportunities that shape individuals' lifetime location decisions and earnings trajectories. Moving costs influence mobility and short-term responses to local shocks. Wealth accumulation through homeownership can also play a crucial role in location choices. 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.

Albouy, D., A. Chernoff, C. Lutz, and C. Warman (2019): "Local Labor Markets in Canada and the United States," *Journal of Labor Economics*, 37, 533–594.

Allen, T., C. Arkolakis, and X. Li (2024): "On the equilibrium properties of spatial models," *American Economic Review: Insights*, 6, 472–489.

Almagro, M. and T. Dominguez-Iino (2024): "Location Sorting and Endogenous Amenities: Evidence from Amsterdam," Working Paper 2022-162, Becker Friedman Institute for Economics.

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.

Baum-Snow, N., N. Gendron-Carrier, and R. Pavan (2024): "Local productivity spillovers," *American Economic Review*, 114, 1030–1069.

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.

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.

Borovičková, 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.

Combes, P.-P., G. Duranton, and L. Gobillon (2008): "Spatial Wage Disparities: Sorting Matters!" *Journal of Urban Economics*, 63, 723–742.

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.

——— (2020): "The Comparative Advantage of Cities," *Journal of International Economics*, 123, 103291.

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.

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.

——— (2025): "Optimal Spatial Policies," Working Paper 33493, National Bureau of Economic Research.

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.*

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.

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. (2013): "Real Wage Inequality," *American Economic Journal: Applied Economics*, 5, 65–103.

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): "he 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.

Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. Von Wachter (2019): "Firming Up Inequality," *Quarterly Journal of Economics*, 134, 1–50.

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.

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

Appendix A Data Appendix

In this Section, I describe the data sources, variable construction, and sample selection for the empirical analysis of the paper.

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

The main data source for the analysis is the Canadian Employer-Employee Dynamic 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). These datasets cover the period from 2001–2017.

T1 files. Individual characteristics are obtained from T1–PMF and T1–FF. These two files are based on information reported in the Income Tax and Benefit Return form (T1), which all Canadians are required to submit annually. Specifically, I observe each individual's age, gender, marital status, and residence location in T1–PMF, and the number of children in T1–FF. Each individual has a unique longitudinal identifier based on the Social Insurance Number (SIN). In T1–PMF, the location information for year t is derived from the postal code reported in year *t+1* , as taxe forms are filed in the following year. Accordingly, I use the location information of each individual in the T1-PMF file of year *t-1* as her location in year *t*.

NALMF. Firm information is obtained from NALMF, which is a longitudinal database of Canadian enterprises linking the Business Register (BR), the Statements of Remuneration Paid (T4), the Payroll Account Deductions (PD7), and the Corporate Income Tax Returns (T2) for incorporated firms and the Financial Declaration and Business Declaration form (T1FD–BD) for unincorporated firms. 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. I drop firms with missing industry information and exclude from the baseline sample firms in industries including agriculture (NAICS 11), mining (NAICS 21), utilities (NAICS 22), education (NAICS 61), hospitals (NAICS 62), non-profit organizations (NAICS 813) and public administrations (NAICS 92).

The NALMF dataset records only the headquarters location for firms operating in multiple locations and includes a multi-location indicator. For the multi-location firms, I partition them into firm-city units based on their employee's residence location and refer to them as firms throughout the empirical analysis. I treat firms that relocate across cities as distinct entities before and after the move.

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. A worker may have multiple T4 records in a given year if employed by more than one firm. For individuals holding multiple jobs, I retain the job with the highest annual earnings, referring to it as the main job. Additionally, I exclude workers whose annual earnings from their main job fall below a threshold equivalent to working 40 hours per week for 13 weeks at half the minimum hourly wage.

IMDB. The IMDB files record information on immigrants from 1980. For each immigrant, it records gender, country of birth, the landing year in Canada, the landing age, and the destination location.

A.2 Other Data Sources

Other data sources used in the analysis are introduced as follows.

CPI. I use the Consumer Price Index of all items from StatsCan (Table 18-10-0006-01) to denominate all monetary values in 2002 Canadian dollars.

Housing rent. I obtain the housing rent data from the Canada Mortgage and Housing Corporation (StatsCan Table 34-10-0133-01), which covers average monthly rents for CMAs and CAs with no fewer than a population of 10,000 by each type of unit (including bachelor units and one-bedroom to three-bedroom units). I use the average monthly rent of two-bedroom units as the measure of housing cost. Moretti (2013) uses the average monthly rent of two or three-bedroom units. I only use two-bedroom units as 1) I don't have data on the number of units rented by each type and 2) the rent data has better coverage of two-bedroom units than three-bedroom ones.

Minimum wage. In Canada, the minimum wage for employees not in federal administrations is set by each province and territory. I obtain a history of minimum wage data from each provincial-level government's website. Then, I define and construct the annual national minimum wage by picking the lowest minimum wage of the provinces and territories for each year. I use this to calculate the minimum earnings threshold for the baseline sample.

The share of undevelopable land. I follow Saiz (2010) to define undevelopable land as land with a slope over 15 degrees and land covered by water. To calculate land slope, I obtain elevation data from the CanVec Elevation Features product, which is maintained by Natural Resources Canada (NRCan). I use the finest 1:50K version of the product to calculate the land slope at the highest resolution. I obtain geographic boundaries of cities and water bodies (lake and river polygons) from StatsCan's Census 2016 boundary files.

Additionally, I classify land areas designated under Ontario's Greenbelt Plan as undevelopable. Enacted in February 2005, the Greenbelt Act was introduced to limit urban sprawl into environmentally sensitive regions of Ontario. It restricts the rezoning of agricultural land, heritage sites, and ecologically significant areas for urban development. The Greenbelt boundary file is obtained from Ontario GeoHub. Figure J.1 presents topographic maps for the four largest Canadian cities.

Share of college graduates by city. I calculate the share of individuals with a college or a higher degree for each city (CMA and CA) using the 2016 Population Census.

Other city characteristics. I collect other city characteristics that may be correlated with city fundamentals (productivity and amenities). These include geographic longitude and latitude from Google Maps, weather information such as from average January and July temperature and average wind speed and the air quality index (AQI) from Environment Canada, the provincial education quality index from the Conference Board of Canada, and the share of land covered by road and the crime severity index, both from StatsCan.

Appendix B Descriptive Facts

In this section, I present descriptive facts on earnings inequality between and within Canadian cities. I then provide suggestive evidence that such patterns are associated with systematic differences in worker and firm between cities and the degrees of assortative matching within cities.

B.1 Earnings disparities between and within Canadian cities

Larger cities have been shown to have higher average earnings and greater earnings dispersion (e.g. Baum-Snow and Pavan (2013) and De La Roca and Puga (2017)). I first confirm that these patterns are also present in Canadian data. I show in Table J.2 city-size regressions of city-level mean log earnings and dispersion measures, including the variance, 90-50 difference, and 50-10 difference. Log earnings are residualized by a third-order age polynomial, gender, marital status, and the number of children using a Mincer-type regression. The results indicate that a 10 log-point increase in city population is associated with a 0.23 log-point increase in mean earnings and a 0.24 log-point increase in the variance, with the latter mainly driven by higher 90-50 gaps in larger cities.

I show in Figure J.3 that the above findings are robust to controlling for industry fixed effects and time spent working in large cities; I also show in Figure J.4 that they are robust to using after-tax earnings and excluding individuals with positive business income. These results can be summarized into the following fact.

Fact 1. *Larger cities have higher mean earnings and greater earnings dispersion than smaller cities.*

B.2 Co-location of high-earning workers and high-paying firms

To uncover the contribution of worker and firm heterogeneity towards spatial earnings disparities, I decompose earnings into firm and worker components following Abowd et al. (1999):[34]

$$\log w_{it} = z_{j(i,t)} + a_i + \epsilon_{it} \tag{B.2}$$

where i indexes an individual, j indexes a firm, and t indexes a year; $\log w_{it}$ is the residual of a Mincer-type regression of log earnings on a set of year dummies and a cubic polynomial in age, z and a are firm and worker fixed effects, and ϵ_{it} is the earnings residual. I follow Bonhomme et al. (2019) and group firms with similar earnings distributions into $k = 10$ clusters using the k-means clustering algorithm.

With the estimates of the AKM equation, I study how firm and worker fixed effects vary between cities. I plot the average firm and worker effects against city population in Figure J.2 and present the results of the city-size regressions in Table J.3. The results reveal significant spatial differences in worker and firm fixed effects. Specifically, a 10 log-point increase in city population is associated with a 0.16 log-point increase in average worker effects and a 0.06 log-point increase in average firm effects, corresponding to 73.3% and 26.7% of the estimated urban earnings premium.

Following Dauth et al. (2022), I also decompose the variance of mean city log earnings:

$$\mathrm{Var}\left(\mathbb{E}_c\left[\log w_{it}\right]\right) = \underbrace{\mathrm{Var}\left(\bar{z}_c\right)}_{\text{Mean firm effect}} + \underbrace{\mathrm{Var}\left(\bar{a}_c\right)}_{\text{Mean worker effect}} + \underbrace{2\,\mathrm{Cov}\left(\bar{z}_c, \bar{a}_c\right)}_{\text{Co-location}} + \mathrm{Var}\left(\bar{\epsilon}_c\right), \tag{B.3}$$

where $\bar{z}_c$ and $\bar{a}_c$ represent city-level averages. The decomposition result is shown in Table J.5. The variance of mean worker effects explains 41.1% of the total between-city variation, the variance of mean firm effects explains about 13.1%, and the covariance of the two explains about 43.9%. The large proportion attributed to the covariance term underscores the

[34] I also implement a specification with the skill-augmenting effect following Bonhomme et al. (2019):

$$\log w_{it} = z_{j(i,t)} + \theta_{j(i,t)} a_i + \epsilon_{it} \tag{B.1}$$

which is more consistent with the general equilibrium model described in Section 2. I present the city mean log earnings decomposition using this specification in Table J.4. The results are qualitatively similar to the one estimated using equation B.2.

importance of co-location in shaping spatial earnings disparities. The covariance between worker and firm fixed effects accounts for a significantly larger share of the variance between cities (43.9%) than the variance between individuals (16.3%). This comparison suggests that similarly ranked workers and firms are much more likely to co-locate in the same city. The evidence can be summarized as the following fact.

Fact 2. *Larger cities have higher-earning workers and high-productivity firms, both of which contribute significantly to between-city earnings inequality.*

I now examine how the spatial distribution of heterogeneous workers and firms evolves over time. I estimate equation (B.2) separately for 2002–2009 and 2010–2017 and correlate changes in city-level mean firm and worker effects. A positive correlation would suggest that higher-quality workers follow higher-quality firms and vice versa, which is confirmed in Panel (c) of Figure J.2.[35] This finding underscores the strong interdependence between worker and firm location choices.[36] I summarize this finding as follows.

Fact 3. *Over time, city-level changes in worker quality strongly and positively correlate with changes in firm quality.*

B.3 Assortative matching within cities

Large and thick labor markets have long been hypothesized to facilitate better worker-firm matches (Helsley and Strange, 1990). To study this, I calculate the correlation of AKM worker and firm effects within each city and regress it on city size. Panel (d) of Figure J.2 confirms that the degree of assortative matching is higher in larger cities.

To examine how much of the greater earnings dispersion in larger cities is due to a higher degree of assortative matching, I decompose within-city earnings variance according to

$$\mathrm{Var}_c\left(\log w_{it}\right) = \mathrm{Var}_c\left(z_{j(i,t)}\right) + \mathrm{Var}_c\left(a_i\right) + 2\,\mathrm{Cov}_c\left(z_{j(i,t)}, a_i\right) + \mathrm{Var}_c\left(\epsilon_{it}\right), \tag{B.4}$$

with which I can use to calculate the variance and covariance components by city and regress them on city population. I present the results in Table J.3. The result shows that the covariance component explains about 20% of the city-size gradient of within-city variance. Lager cities also have greater variations in worker and firm fixed effects that contribute to greater within-city inequality, consistent with evidence documented by Eeckhout et al. (2014) and Combes et al. (2012). These results can be summarized into the following fact.

[35] The result holds when focusing only on new workers—both labor market entrants and migrants—and new firms, as shown in Figure J.6.

[36] The variance of changes in mean worker effects, mean firm effects, and their covariance account for 43.1%, 13.6%, and 43.3% of the variance in city mean log earnings, respectively.

Fact 4. *Larger cities have higher degrees of positive assortative matching between workers and firms, which contribute to greater within-city inequality.*

Appendix C Model Appendix

C.1 Firm's problem

Firm j in city c chooses wage offers $\mathbf{W}_{jc} = \{W_{jc}(a)\}_{\forall a}$ and the housing input h_{jc} to maximize profits:

$$\max_{\mathbf{W}_{jc},h_{jc}} \left(\int_{\underline{a}}^{\bar{a}} A_c L_c^{\mu} \cdot z_j \cdot a^{\theta_j} \cdot D_{jc}(a) da \right)^{1-\alpha} \cdot {h_{jc}}^{\alpha} - r_c h_{jc} - \int_{\underline{a}}^{\bar{a}} W_{jc}(a) D_{jc}(a)\ da \tag{C.1}$$

subject to

$$D_{jc}(a) = L_c(a) \cdot \frac{(W_{jc}(a) G_{jc}(a))^{\frac{\beta_w}{\rho_w}}}{\mathbb{W}_c(a)} = W_{jc}(a)^{\frac{\beta_w}{\rho_w}} \cdot \kappa_{jc}(a) \tag{C.2}$$

where $\kappa_{jc}(a) \equiv L_c(a) \cdot G_{jc}(a)^{\frac{\beta_w}{\rho_w}} \cdot \mathbb{W}_c^{-1}(a)$ is defined as a firm-specific labor supply shifter of each skill a. Taking the derivative with respect to $\{W_{jc}(a)\}_{\forall a}$ and h_{jc} under assumption 1, we have the following set of FOCs:

$$W_{jc}(a) : (1-\alpha) \frac{\beta_w}{\rho_w} \kappa_{jc}(a) {Q_{jc}}^{-\alpha} W_{jc}(a)^{\frac{\beta_w}{\rho_w}-1} \cdot A_c L_c^{\mu} z_j a^{\theta_j} \cdot {h_{jc}}^{\alpha} - \left(1 + \frac{\beta_w}{\rho_w}\right) \kappa_{jc}(a) \cdot W_{jc}(a)^{\beta_w/\rho_w} = 0 \tag{C.3}$$

$$h_{jc} : \alpha Q_{jc}^{1-\alpha} {h_{jc}}^{\alpha-1} - r_c = 0 \tag{C.4}$$

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. We could obtain the optimal wage offers and housing inputs from the FOCs as:

$$W_{jc}(a) = \chi \cdot A_c L_c^{\mu} r_c^{\frac{\alpha}{\alpha-1}} \cdot z_j \cdot a^{\theta_j} \tag{C.5}$$

$$h_{jc} = \alpha^{\frac{1}{1-\alpha}} r_c^{\frac{1}{\alpha-1}} \cdot Q_{jc} \tag{C.6}$$

where $\chi \equiv \frac{\beta_w/\rho_w}{1+\beta_w/\rho_w} \cdot (1-\alpha)\alpha^{\frac{\alpha}{1-\alpha}}$. Plugging equation (C.5) into Q_{jc} yields:

$$\begin{aligned} Q_{jc} &= \int_{\underline{a}}^{\bar{a}} \left[A_c L_c^{\mu} \cdot z_j a^{\theta_j} \right]^{1+\beta_w/\rho_w} \left(\chi {r_c}^{\frac{\alpha}{\alpha-1}} \right)^{\beta_w/\rho_w} \kappa_{jc}(a)\ da \\ &= (A_c L_c^{\mu}) \cdot \left(z_j\right)^{1+\beta_w/\rho_w} \cdot \phi_{jc} \end{aligned} \tag{C.7}$$

where

$$\phi_{jc} = \int_a \left(a^{\theta_j}\right)^{1+\beta_w/\rho_w} L_c(a) \cdot \frac{G_j(a)^{\frac{\beta_w}{\rho_w}}}{\sum_{j'\in\mathcal{J}_c}\left(z_{j'}a^{\theta_{j'}} \cdot G_{j'}(a)\right)^{\frac{\beta_w}{\rho_w}}} da \tag{C.8}$$

is a firm-specific term summarizing local labor market competitiveness and worker-firm complementarity. Firm j's wage bill is the sum of wages to all workers it employs

$$E_{jc} = \int_{\underline{a}}^{\bar{a}} W_{jc}(a)D_{jc}(a)\ da = \chi\left(A_c L_c^{\mu} r_c^{\frac{a}{a-1}}\right) \cdot \left(z_j\right)^{1+\beta_w/\rho_w} \cdot \phi_{jc}. \tag{C.9}$$

Then, plugging equations (C.6) and (C.9) back to (C.1) yields firm j's optimal profits in city c:

$$\begin{aligned}\pi_c(j) &= {Q_{jc}}^{1-\alpha}{h_{jc}}^{\alpha} - r_c h_{jc} - E_{jc} \\ &= \Psi \cdot \left(A_c {L_c}^{\mu} r_c^{\frac{a}{\alpha-1}}\right) \cdot \left(z_j\right)^{1+\beta_w/\rho_w} \cdot \phi_{jc}\end{aligned} \tag{C.10}$$

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

C.2 The housing market

The housing developer of each city combines land $\bar{H}_c$, which is exogenously given, and the final good Y_c to produce housing. The developer chooses the amount of final good to maximize profits:

$$\max_{Y_c} r_c \cdot \bar{H}_c {Y_c}^{\frac{1}{1+\gamma_c}} - Y_c. \tag{C.11}$$

The FOC of this profit-maximizing problem is:

$$\frac{1}{1+\gamma_c} r_c \cdot \bar{H}_c {Y_c}^{-\frac{\gamma_c}{1+\gamma_c}} - 1 = 0, \tag{C.12}$$

from which we can solve for $Y_c = \left(\frac{1}{1+\gamma_c} r_c \bar{H}_c\right)^{\frac{1+\gamma_c}{\gamma_c}}$. Then, the housing supply curve is:

$$H_c^S\left(r_c\right) = \bar{H}_c^0 \cdot r_c^{\frac{1}{\gamma_c}} \tag{C.13}$$

where $\bar{H}_c^0 = (1+\gamma_c)^{-1/\gamma_c} \cdot {\bar{H}_c}^{(1+\gamma_c)/\gamma_c}$. Total housing demand in city c is given by

$$\begin{aligned}H_c^D(r_c) &= \frac{\eta}{r_c}\int_{\underline{a}}^{\bar{a}} \sum_{j\in\mathcal{J}_c} \tau W_{jc}(a)D_{jc}(a)\ da + \frac{\alpha}{1-\alpha}\frac{1+\beta_w/\rho_w}{\beta_w/\rho_w}\frac{1}{r_c}\int_{\underline{a}}^{\bar{a}} \sum_{j\in\mathcal{J}_c} W_{jc}(a)D_{jc}(a)\ da \\ &= \left(\tau\eta + \frac{\alpha}{1-\alpha}\frac{1+\beta_w/\rho_w}{\beta_w/\rho_w}\right)\frac{E_c}{r_c},\end{aligned} \tag{C.14}$$

where I define E_c as the total wage bill of city c. The housing market clearing condition is

$$H_c^D(r_c) = H_c^S(r_c). \tag{C.15}$$

I assume that housing developers own the land and set rental prices competitively. Their profits are pooled into a national portfolio and redistributed to workers, with each worker's rebate proportional to their wages. Specifically, the rebate is given by

$$(\tau - 1)\sum_i W_i = \sum_c (r_c H_c - Y_c) = \sum_c \frac{\gamma_c}{1+\gamma_c} r_c H_c. \tag{C.16}$$

It is also useful to specify the final good market clearing condition here. The homogeneous final good is supplied by firms and demanded by workers and entrepreneurs for consumption, and housing developers for housing production. The final good market clearing condition is thus

$$\sum_c \sum_j \int_a (1-\eta) \cdot D_{jc}(a)\, \tau W_{jc}(a)\, da + \sum_c \sum_j \pi_c(j) + \sum_c Y_c = \sum_c \sum_j Q_{jc}^{1-\alpha} h_{jc}^{\alpha}. \tag{C.17}$$

Appendix D Efficiency

D.1 Social planner's problem

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

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

where $\varphi^w(a)$ and $\varphi^f(k)$ are the Pareto weights for skill a workers and type k entrepreneurs. I assume there are discrete types of workers and firms in formulating the social planner's problem. Utility of each type of agent is

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

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

where

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

$\bar{C}^w$ 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 (D.1): 1) the amount of the final good and housing allocated to skill a workers working for a cluster k firm and city c, $c_{kc}(a)$ and $h_{kc}(a)$; 2) the amount of the final good allocated to type k entrepreneurs in city c, $c_c(k)$; 3) the amount of the housing allocated to cluster k firms in city c for production, $h_c(k)$; 4) the total number of skill a workers in cluster k firms located in city c, $D_{kc}(a)$; 5) the number of cluster k firms located in city c, $J_c(k)$ and 6) the amount of final good used to produce housing for each city c, I_c.

The planner is subject to the following constraints. First, there are the spatial mobility and local labor matching constraints of workers and firms. I assume that the planner does not observe each worker's and firm's idiosyncratic preferences. For workers, the planner is subject to both within-city and between-city allocation constraints

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

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

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

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

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

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

Second, there are worker and firm allocation constraints:

$$\sum_c \sum_k D_{kc}(a) = L(a) \tag{D.9}$$

$$\sum_c J_c(k) = J(k) \tag{D.10}$$

which are automatically satisfied given (D.7) and (D.8).

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

$$\sum_c \sum_k \sum_a D_{kc}(a)\, c_{kc}(a) + \sum_c \sum_k J_c(k)\, c_c(k) + \sum_c Y_c \leq \sum_c \sum_k J_c(k) \left(\sum_a \frac{D_{kc}(a)}{J_c(k)} f_c(k,a) \right)^{1-\alpha} h_c(k)^{\alpha} \tag{D.11}$$

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

$$\sum_k \sum_a D_{kc}(a)\, h_{kc}(a) + \sum_k J_c(k)\, h_c(k) \leq \bar{H}_c {Y_c}^{1/(1+\gamma_c)}, \ \forall c. \tag{D.12}$$

Lastly, the planner faces the non-negativity constraints for all consumption and housing allocations to workers and firms:

$$c_{kc}(a) \geq 0, \ h_{kc}(a) \geq 0 \tag{D.13}$$

$$c_c(k) \geq 0, \ h_c(k) \geq 0. \tag{D.14}$$

D.2 Solving the social planner's problem

The Lagrange function of the social planner's problem is

$$\begin{aligned}
\mathcal{L} = & \sum_a \varphi^w(a)\, L(a)\, U(a) + \sum_k \varphi^f(k)\, J(k)\, \Pi(k) \\
& - \sum_c \sum_k \sum_a W^*_{kc}(a)\, D_{kc}(a) \Bigg[\log\left(\frac{D_{kc}(a)}{L(a)} \right) \\
& \qquad - \log\left(J_c(k) \cdot \left(c_{kc}(a)^{1-\eta}\, h_{kc}(a)^{\eta} \right)^{\beta_w/\rho_w} \cdot U_c(a)^{\beta_w - \beta_w/\rho_w} \cdot \left(\sum_c U_c(a)^{\beta_w} \right)^{-1} \right) \Bigg] \\
& - \sum_c \sum_k \pi^*_c(k)\, J_c(k) \left[\log\left(\frac{J_c(k)}{J(k)} \right) - \log\left(c_c(k)^{\beta_f} \cdot \left(\sum_c c_c(k)^{\beta_f} \right)^{-1} \right) \right] \\
& - \sum_c R^*_c \left[\sum_k \sum_a D_{kc}(a) h_{kc}(a) + \sum_k J_c(k) h_c(k) - \bar{H}_c {I_c}^{1/(1+\gamma_c)} \right] \\
& - P^* \left[\sum_c \left(\sum_k \sum_a D_{kc}(a) c_{kc}(a) + \sum_k J_c(k) c_c(k) + Y_c \right) - \sum_c \sum_k J_c(k) \left(\sum_a \frac{D_{kc}(a)}{J_c(k)} f_c(k,a) \right)^{1-\alpha} h_c(k)^{\alpha} \right] \\
& - \ldots
\end{aligned} \tag{D.15}$$

where $W^*_{kc}(a)$ is the multiplier for the worker allocation constraint (D.7); $\pi^*_c(k)$ is the multiplier for the firm allocation constraint (D.8); $P^*$is the multiplier for the resource constraint of the final good (D.11); R^*_c is the multiplier for local housing constraints (D.12) for each city c. I omit the terms for the non-negativity constraints in the Lagrange function and focus on the internal solution of the problem. The first-order conditions of the planner's problem are derived below.

First, for the ones associated with the final good and housing allocated to workers, $c_{kc}(a)$ and $h_{kc}(a)$, we have

$$\varphi^w(a)\,L(a)\,\frac{\partial U(a)}{\partial c_{kc}(a)} + \frac{\beta_w}{\rho_w}\frac{(1-\eta)W^*_{kc}(a)\,D_{kc}(a)}{c_{kc}(a)} + \frac{\beta_w(\rho_w-1)}{\rho_w}\sum_{k'} W^*_{k'c}(a)\,D_{k'c}(a)\cdot\frac{\partial \log U_c(a)}{\partial c_{kc}(a)} + \sum_{c'}\sum_{k'} W^*_{k'c'}(a)\,D_{k'c'}(a)\cdot\frac{\partial \log\left(\sum_c U_c(a)^{\beta_w}\right)^{-1}}{\partial c_{kc}(a)} - P^* D_{kc}(a) = 0 \tag{D.16}$$

and

$$\varphi^w(a)\,L(a)\,\frac{\partial U(a)}{\partial h_{kc}(a)} + \frac{\beta_w}{\rho_w}\frac{\eta W^*_{kc}(a)\,D_{kc}(a)}{h_{kc}(a)} + \frac{\beta_w(\rho_w-1)}{\rho_w}\sum_{k'} W^*_{k'c}(a)\,D_{k'c}(a)\cdot\frac{\partial \log U_c(a)}{\partial h_{kc}(a)} + \sum_{c'}\sum_{k'} W^*_{k'c'}(a)\,D_{k'c'}(a)\cdot\frac{\partial \log\left(\sum_c U_c(a)^{\beta_w}\right)^{-1}}{\partial h_{kc}(a)} - R^*_c D_{kc}(a) = 0. \tag{D.17}$$

Then, for the ones associated with the final good and floor space allocated to firms, $c_c(k)$ and $h_{kc}(k)$, we have

$$\varphi^f(k)\,J(k)\,\frac{\partial \Pi(k)}{\partial c_c(k)} + \beta_f \pi^*_c(k)\,J_c(k)\,\frac{1}{c_c(k)} + \sum_{c'} \pi^*_{c'}(k)\,J_{c'}(k)\,\frac{\partial \log\left(\sum_c c_c(k)^{\beta_f}\right)^{-1}}{\partial c_c(k)} - P^* J_c(k) = 0 \tag{D.18}$$

and

$$R^*_c - P^*\alpha\left(\sum_a \frac{D_{kc}(a)}{J_c(k)} f_c(k,a)\right)^{1-\alpha} h_c(k)^{\alpha-1} = 0. \tag{D.19}$$

For the ones with the worker allocation $D_{kc}(a)$, we have

$$W^*_{kc}(a) + R^*_c h_{kc}(a) + P^* c_{kc}(a) = P^*(1-\alpha)\cdot\alpha^{\frac{\alpha}{1-\alpha}} R_c^{*\,\frac{\alpha}{\alpha-1}}\left[f_c(k,a) + \frac{\mu}{L_c}\sum_{k'}\sum_{a'} D_{k'c}(a')\,f_c(k',a')\right] \tag{D.20}$$

and the ones with the firm allocation $J_c(k)$, we have

$$\sum_a \varphi^w(a) L(a) \frac{\partial U(a)}{\partial J_c(k)} + \sum_a W^*_{kc}(a) \frac{D_{kc}(a)}{J_c(k)} + (\beta_w - \beta_w/\rho_w) \sum_{k'} \sum_a W^*_{k'c}(a) D_{k'c}(a) \frac{\partial \log U_c(a)}{\partial J_c(k)} - \sum_{c'} \sum_{k'} \sum_a W^*_{k'c'}(a) D_{k'c'}(a) \frac{\partial \log \sum_c U_c(a)^{\beta_w}}{\partial J_c(k)} - \pi^*_c(k) - P^* c_c(k) = 0. \tag{D.21}$$

Finally, the first-order condition with respect to the final good used to produce housing in city c, I_c, is:

$$R^*_c \bar{H}_c \frac{1}{1+\gamma_c} Y_c^{\frac{\gamma_c}{1+\gamma_c}} = P^*, \forall c. \tag{D.22}$$

D.3 Characterization of the social planner's solution

Combining equations (D.16) and (D.17), we obtain

$$\varphi^w(a) - \left(\frac{\beta_w}{\rho_w} - \beta_w\right) \sum_{k'} W^*_{k'c}(a) \frac{D_{k'c}(a)}{L_c(a)} - \beta_w \sum_{c'} \sum_{k'} W^*_{k'c'}(a) \frac{D_{k'c'}(a)}{L(a)} = P^* c_{kc}(a) + R^*_c h_{kc}(a) - \frac{\beta_w}{\rho_w} W^*_{kc}(a) \tag{D.23}$$

and

$$\eta P^* c_{kc}(a) = (1-\eta) R^*_c h_{kc}(a). \tag{D.24}$$

Then, from equation (D.20), we have

$$W^*_{kc}(a) = -R^*_c h_{kc}(a) - P^* c_{kc}(a) + P^*(1-\alpha) \cdot \alpha^{\frac{\alpha}{1-\alpha}} R_c^{*\frac{\alpha}{\alpha-1}} \left[f_c(k,a) + \frac{\mu}{L_c} \sum_{k'} \sum_{a'} D_{k'c}(a') f_c(k',a') \right]. \tag{D.25}$$

With equation (D.23), we can now derive the condition for socially optimal worker consumptions

$$P^* c_{kc}(a) + R^*_c h_{kc}(a) = \frac{\beta_w/\rho_w}{1+\beta_w/\rho_w} \left(\tilde{W}_{kc}(a) - \mathcal{O}^w_c(a) \right) + \frac{1}{1+\beta_w/\rho_w} \varphi^w(a) \tag{D.26}$$

where I define $\tilde{W}_{kc}(a) = P^*(1-\alpha) \cdot \alpha^{\frac{\alpha}{1-\alpha}} R_c^{*\frac{\alpha}{\alpha-1}} \left[f_c(k,a) + \frac{\mu}{L_c} \sum_k \sum_a D_{kc}(a) f_c(k,a) \right]$ as the social value of skill a worker in firm type k and city c, which consist of her own marginal product of labor and the agglomeration spillover for all workers in the city; $\mathcal{O}^w_c(a) = (1-\rho_w) \sum_{k'} W^*_{k'c}(a) \frac{D_{k'c}(a)}{L_c(a)} + \rho_w \sum_{c'} \sum_{k'} W^*_{k'c'}(a) \frac{D_{k'c'}(a)}{L(a)}$ 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 all cities, $\sum_{c'} \sum_{k'} W^*_{k'c'}(a) \frac{D_{k'c'}(a)}{L(a)}$. As common in the urban literature, workers do not internalize the agglomeration spillovers affected by their location choices, which is part of the social value. 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

$$P^* c_{kc}(a) + R_c^* h_{kc}(a) = \frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}\tilde{W}_{kc}(a) + T_c^w(a). \tag{D.27}$$

We now obtain the same condition for firms. Start from equation (D.18), we have

$$\varphi^f(k) = P^* c_c(k) - \beta_f \pi_c^*(k) + \beta_f \sum_{c'} \pi_{c'}^*(k) \frac{J_{c'}(k)}{J(k)}. \tag{D.28}$$

Then, plugging $\pi_c^*(k)$ from equation (D.21) into the previous equation yields

$$\sum_a \varphi^w(a) L(a) \frac{\partial U(a)}{\partial J_c(k)} + \sum_a W_{kc}^*(a) D_{kc}(a) \frac{1}{J_c(k)} - (1-\rho_w) \sum_{k'} \sum_a W_{k'c}^*(a) D_{k'c}(a) \frac{D_{kc}(a)}{L_c(a)} \frac{1}{J_c(k)} - \rho_w \sum_{c'} \sum_{k'} \sum_a W_{k'c'}^*(a) D_{k'c'}(a) \cdot \frac{D_{kc}(a)}{L_c(a)} \frac{1}{J_c(k)} - \pi_c^*(k) - P^* c_c(k) = 0. \tag{D.29}$$

We can then derive the socially optimal entrepreneur consumptions as

$$P^* c_c(k) = \frac{\beta_f}{1+\beta_f}\left(\tilde{\pi}_c(k) - \mathcal{O}^f(k)\right) + \frac{1}{1+\beta_f}\varphi^f(k). \tag{D.30}$$

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

$$\begin{aligned}\tilde{\pi}_c(k) = & \underbrace{\sum_a W_{kc}^*(a) \frac{D_{kc}(a)}{J_c(k)}}_{\text{firm surplus}} - \underbrace{(1-\rho_w)\sum_a \bar{W}_c^*(a) \frac{D_{kc}(a)}{J_c(k)}}_{\text{local labor stealing}} \\ & - \underbrace{\rho_w \sum_a \bar{W}^*(a) \frac{D_{kc}(a)}{J_c(k)}}_{\text{national labor stealing}} + \underbrace{\frac{\rho_w}{\beta_w} \sum_a \varphi^w(a) \frac{D_{kc}(a)}{J_c(k)}}_{\text{love-of-variety preference}}\end{aligned} \tag{D.31}$$

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 all cities. As before, I define $T^f(k) \equiv -\frac{\beta_f}{1+\beta_f}\mathcal{O}^f(k) + \frac{1}{1+\beta_f}\varphi^f(k)$ and rewrite the firm optimality condition as

$$P^* c_c(k) = \frac{\beta_f}{1+\beta_f}\tilde{\pi}_c(k) + T^f(k). \tag{D.32}$$

An efficient allocation is given by allocations $\{c_{kc}(a), h_{kc}(a), c_c(k), D_{kc}(a), L_c(a), J_c(k), Y_c\}$ and multipliers $\{P^*, R_c^*, W_{kc}^*(a), \pi_c^*(k)\}$ such that the first-order conditions (D.16)-(D.22), the spatial mobility constraints (D.5)-(D.8), and resources constraints (D.11)-(D.12) hold.

Given competitive prices $\{P, r_c, W_{kc}(a), \pi_c(k)\}$ equal to multipliers $\{P^*, R_c^*, W_{kc}^*(a), \pi_c^*(k)\}$ and decentralized expenditure $\{x_{kc}(a), x_c(k)\}$ equal to $\{x_{kc}^*(a), x_c^*(k)\}$, equations (D.16)–(D.17) coincide with the utility maximization condition implied by (2.1), equation (D.19) coincide with the optimal housing input condition (C.6), equations (D.5)–(D.8) coincide with optimality location and matching conditions (2.3)–(2.4) and (2.12), equation (D.22) coincides with the optimality condition of the developer (C.12), equations (D.11)–(D.12) coincide with market clearing conditions (C.15)-(C.17). Therefore, the system characterizing the competitive solution for $\{c_{kc}(a), h_{kc}(a), c_c(k), D_{kc}(a), L_c(a), J_c(k), Y_c\}$ given the prices $\{P, r_c, W_{kc}(a), \pi_c(k)\}$ and the expenditures $\{x_{kc}(a), x_c(k)\}$ is the same as the system characterizing the planner's allocation for those same quantities given the multipliers $\{P^*, R_c^*, W_{kc}^*(a), \pi_c^*(k)\}$ and expenditures $\{x_{kc}(a), x_c(k)\}$. As a result, if the competitive equilibrium is efficient, when $x_{kc}(a) = x_{kc}^*(a)$ where $x_{kc}^*(a)$ is given by (D.26) and $x_c(k) = x_c^*(k)$ where $x_c^*(k)$ is given by (D.30). Conversely, if $x_{kc}(a) = x_{kc}^*(a)$ and $x_c(k) = x_c^*(k)$ for $\{x_{kc}(a), x_c(k)\}$ defined in (3.5) and (3.6) given the multipliers that solve the planner's problem, there is a solution for the competitive allocation such that $\{P, R_c, W_{kc}(a), \pi_c(k)\} = \{P^*, R_c^*, W_{kc}^*(a), \pi_c^*(k)\}$. If the planning problem is concave then there is a unique solution to the system characterizing the planner's solution, in which case $\{P, R_c, W_{kc}(a), \pi_c(k)\} = \{P^*, R_c^*, W_{kc}^*(a), \pi_c^*(k)\}$ is the only competitive equilibrium.

D.4 Optimal spatial policy

After characterizing the social planner's solution, I now design the optimal spatial policy to restore efficient spatial allocation. The optimal spatial policy consist of (1) a set of labor income taxes $\{t_{kc}^w(a)\}_{\forall a,k,c}$ that are specific to worker skill a, firm type k and city c, and (2) a set of corporate profit taxes $\{t_c^f(k)\}_{\forall k,c}$ that are specific to firm type k and city c. Specifically, the labor income taxes are set as

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

,and the corporate income taxes are set as

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

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.

I assume that the housing developers' profits are rebated to workers, which is proportional to the after-tax labor income $(1-\tau^w_{kc}(a))\, W_{kc}(a)$. Hence, the total after-tax income of workers are firms are $I^w_{kc}(a) = \tau\left(1-t^w_{kc}(a)\right) W_{kc}(a)$ and $I^f_c(k) = \left(1-t^f_c(k)\right)\pi_{kc}(a)$, respectively. With the instruments and the rebate rule, the total after-tax income of a worker with skill a working for firm k in city c is

$$I^w_{kc}(a) = \tau\left(1-t^w_{kc}(a)\right) W_{kc}(a) = \tau\left(\frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}\tilde{W}_{kc}(a) + T^w_c(a)\right),$$

and the budget constraint for such a worker is

$$Pc_{kc}(a) + r_c h_{kc}(a) = \tau\left(\frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}\tilde{W}_{kc}(a) + T^w_c(a)\right). \tag{D.33}$$

The after-tax income of an entrepreneur operating firm k in city c is

$$I^f_c(k) = \left(1-t^f_c(k)\right)\pi_{kc}(a) = \frac{\beta_f}{1+\beta_f}\tilde{\pi}_c(k) + T^f(k), \tag{D.34}$$

and the budget constraint for such an entrepreneur is

$$Pc_c(k) = \left(1-t^f_c(k)\right)\pi_{kc}(a) = \frac{\beta_f}{1+\beta_f}\tilde{\pi}_c(k) + T^f(k). \tag{D.35}$$

I now show that the government's budget balances with the specified taxes, that is

$$-\sum_c\sum_k\sum_a D_{kc}(a)\left(t^w_{kc}(a)\, W_{kc}(a)\right) - \sum_c\sum_k J_c(k)\left(t^f_c(k)\,\pi_c(k)\right) = 0 \tag{D.36}$$

or equivalently

$$\sum_c\sum_k\left(\sum_a D_{kc}(a)\, T^w_c(a) + J_c(k)\, T^f(k)\right) = \sum_c\sum_k\Bigg(\sum_a D_{kc}(a)\left(W_{kc}(a) - \tilde{W}_{kc}(a)\right) + J_c(k)\left(\pi_c(k) - \tilde{\pi}_c(k)\right)\Bigg) \tag{D.37}$$

To show this, we can rewrite $T_c^w(a)$ and $T^f(k)$ as

$$T_c^w(a) = -\left((1-\rho_w)\tilde{W}_c(a) + \rho_w \tilde{W}(a)\right) + \frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}\left((1-\rho_w)\bar{I}_c^w(a) + \rho_w \bar{I}^w(a)\right) + \frac{1}{1+\beta_w/\rho_w}\bar{I}^w(a)$$

$$T^f(k) = -\frac{\beta_f}{1+\beta_f}\sum_c \left(\frac{1+\beta_f}{\beta_f}\tilde{\pi}_c(k) - I_c^f(k)\right)\frac{J_c(k)}{J(k)} + \frac{1}{1+\beta_f}\sum_c I_c^f(k)\frac{J_c(k)}{J(k)}.$$

Now, note that

$$\sum_c\sum_k\sum_a \frac{1}{\tau} I_{kc}^w(a) D_{kc}(a) + \sum_c\sum_k I_c^f(k) J_c(k) = \sum_c\sum_k\sum_a W_{kc}(a) D_{kc}(a) + \sum_c\sum_k \pi_c(k) J_c(k).$$

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

$$\begin{aligned}
\mathcal{T} =& \sum_c\sum_k\sum_a D_{kc}(a)\Bigg[-\left((1-\rho_w)\tilde{W}_c(a) + \rho_w\tilde{W}(a)\right) + \frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}\left((1-\rho_w)\frac{\bar{I}_c^w(a)}{\tau} + \rho_w\frac{\bar{I}^w(a)}{\tau}\right) \\
&+ \frac{1}{1+\beta_w/\rho_w}\frac{\bar{I}^w(a)}{\tau}\Bigg] \\
&- \frac{\beta_f}{1+\beta_f}\sum_c\sum_k J_c(k)\left[\sum_{c'}\sum_{k'}\left(\frac{1+\beta_f}{\beta_f}\tilde{\pi}_{c'}(k') - I_c^f(k)\right)\frac{J_{c'}(k')}{J(k')} + \frac{1}{1+\beta_f}\bar{I}^f(k)\right] \\
=& -\sum_c\sum_k\left(\sum_a \tilde{W}_{kc}(a) D_{kc}(a) + \tilde{\pi}_c(k) J_c(k)\right) + \sum_c\sum_k\left(\sum_a W_{kc}(a) D_{kc}(a) + \pi_c(k) J_c(k)\right) \\
=& \sum_c\sum_k\sum_a\left(W_{kc}(a) - \tilde{W}_{kc}(a)\right) D_{kc}(a) + \sum_{c'}\sum_{k'}\left(\pi_c(k) - \tilde{\pi}_c(k)\right) J_c(k).
\end{aligned}$$

Hence, I have proved the government budget balance given by equation (D.37). Then, if the planner's problem is globally concave, the equilibrium with the specified taxes implements the efficient allocation. Concretely, the after-tax budget constraints for workers (D.33) are identical to the ones in the social planner's problem (D.27). Land rents are redistributed proportional to after-tax income and do not affect workers' location or firm choices. The consumption choices of workers on the final good and housing, implied by (2.1), are identical to the social planner's conditions (D.16) and (D.17). The between-city location choice and within-city firm choice conditions of workers, with the after-tax worker income (2.4) and (2.3) are identical to the optimal worker allocation conditions in the social planner's problem (D.6) and (D.5). The after-tax budget constraints for entrepreneurs (D.34) are identical to the ones in the social planner's problem (D.32). The location choice condition for firms, given after-tax entrepreneur income, implied by (2.12), is identical to the optimal firm allocation conditions in the social planner's problem (D.8). The first order condition for final good inputs used by the housing developers (C.12) is identical to the one of the social planner (D.22). The local labor market clearing conditions are automatically satisfied given (D.5). The market clearing conditions for the final good and housing (C.17) and (C.15) are the same as the resource constraints in the planner's problem (D.11) and (D.12).

D.5 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, following a scheme similar to Berger et al. (2025). The transfer design consistent of two steps.

First, remove the optimal taxes $\{t^w_{kc}(a)\}k, c, a$ and $\{t^f_c(k)\}c, k$ while maintaining the efficient spatial allocation. Allow workers and firms to rematch within each city,[37] 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. These transfers are financed by total land rent and are spent exclusively on the final good.

Appendix E 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 clarity and simplicity of notation, 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. $\sigma = \left[\{\log L_c(i)\}_{\forall c,i}, \{\log J_c(K)\}_{\forall c,k}\right]$, where $L_c(i)$ denote the measure of type-i workers in city c, and $J_c(k)$ denote the measure

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

of type-k firms in city c. 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

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

for workers and

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

for firms, where

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

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 $L_c(i), \forall c, i$:

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

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

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

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

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

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 can be rewritten as

$$2\rho_w < 1$$
$$\rho_w < \frac{1}{2}.$$

For the conditions with respect to firms, it requires

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

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

$$1 + \beta_f \cdot s_c(k) > \sum_{k' \neq k} \left| \frac{\partial \log J_c(k)}{\partial \log J_c(k')} \right| + \sum_{c' \neq c} \sum_{k'} \left| \frac{\partial \log J_c(k)}{\partial \log J_{c'}(k')} \right| + \sum_{c'} \sum_{i} \left| \frac{\partial \log J_c(k)}{\partial \log L_{c'}(i)} \right|, \forall c, k.$$
$$1 + \beta_f \cdot s_c(k) > \beta_f (1 - s_c(k)) + \beta_f + 2\beta_f, \forall c, k.$$
$$1 > 4\beta_f - 2\beta_f s_c(k), \forall c, k.$$

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

Several remarks follow. First, the proposition provides a sufficient but not necessary condition for equilibrium uniqueness. As noted in Allen, Arkolakis, and Li (2024), Jacobian-based conditions for uniqueness are often overly stringent. Second, the condition requires only that the ratio of city-level to firm-level labor supply elasticity (β_w/ρ_w) be small, without restricting β_w. A larger ratio strengthens workers' tendency to follow firms spatially, meaning the condition ensures between-city dispersion in idiosyncratic preferences exceeds within-city dispersion. Third, equilibrium uniqueness is unaffected by the strength of production complementarities if idiosyncratic preferences are sufficiently dispersed. Fourth, there also exist other congestion forces in the model, including exogenous location differences and housing market congestion, that help with uniqueness. Lastly, although the estimated parameters do not satisfy the proposition's condition, solving the model from different initial guesses consistently yields the same solution.

Appendix F Identification Appendix

F.1 Time-varying terms

Here, I add the time-varying shocks and functional form assumptions specified in Section 4.2 to key terms in the model. First, worker i's wage in firm j, city c and time t becomes

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

Firm $j \in \mathcal{J}_c$'s wage bill is

$$E_{jct} = \chi \cdot \left(A_c \hat{A}_{ct} L_{ct}^{\mu} r_{ct}^{\frac{a}{\alpha-1}} \right) \cdot (z_j \hat{z}_{jt})^{1+\beta_w/\rho_w} \cdot \phi_{jct} \tag{F.2}$$

and its optimal profit is

$$\pi_{ct}(j) = \Psi \cdot \left(A_c \hat{A}_{ct} L_{ct}^{\mu} r_{ct}^{\frac{a}{\alpha-1}} \right) \cdot (z_j \hat{z}_{jt})^{1+\beta_w/\rho_w} \cdot \phi_{jct} \tag{F.3}$$

where

$$\phi_{jct} = \int_a \left(a^{\theta_j}\right)^{1+\beta_w/\rho_w} L_{ct}(a) \cdot \frac{G_{jc}(a)^{\beta_w/\rho_w}}{\sum_{j' \in \mathcal{I}_c} \left(z_{j'} \hat{z}_{j't} a^{\theta_{j'}} \cdot G_{j'c}(a) \right)^{\frac{\beta_w}{\rho_w}}} da. \tag{F.4}$$

F.2 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 (4.7) and (4.8). Start with the time-varying wage equation given by equation (F.1):

$$W_{ijt}(a) = \chi \cdot \underbrace{r_{ct}^{\frac{\alpha}{\alpha-1}} \cdot A_c \hat{A}_{ct} L_{ct}^{\mu}}_{\text{city component}} \cdot \underbrace{z_j \hat{z}_{jt}}_{\text{firm component}} \cdot \underbrace{a_i^{\theta_j} \hat{a}_{it}}_{\text{worker component}}. \tag{F.5}$$

Wage changes for the worker stayers 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. Recall that worker skill transient shocks are assumed to be *i.i.d.* and uncorrelated with city and firm shocks. In what follows, I will use the average city wage $\bar{W}_{ct}$ of stayers to isolate the city-level shocks and the residualized wage $\hat{W}_{it}$ to isolate firm-level shocks. The mean and residualized wage bill terms are constructed for the same purposes. Formally, with equation (F.5) the average wage for stayers is given by

$$\begin{aligned} \bar{W}_{ct} &= \mathbb{E}_{S(i)=1}\left[\chi \cdot r_{ct}^{\frac{\alpha}{\alpha-1}} \cdot A_c \hat{A}_{ct} L_{ct}^{\mu} \cdot z_j \hat{z}_{jt} \cdot a_i^{\theta_j} \hat{a}_{it}\right] \\ &= \chi \cdot r_{ct}^{\frac{\alpha}{\alpha-1}} \cdot A_c \hat{A}_{ct} L_{ct}^{\mu} \cdot \bar{\Xi}_c \end{aligned} \tag{F.6}$$

where $\bar{\Xi}_c \equiv \mathbb{E}_{S(i)=1}\left[z_j \cdot a_i^{\theta_j}\right]$ is defined as the city wage index of the stayers. It does not change over time because the set of workers and firms in the stayer sample is fixed, and the firm productivity and worker transient skill shocks average out cross-sectionally within a city, for which reason I omit these shocks in the definition of $\bar{\Xi}$ and other city-level aggregate variables from now on. The residualized wage is thus constructed as:

$$\hat{W}_{ijt} = W_{ijt}/\bar{W}_{ct} = (\bar{\Xi}_c)^{-1} \cdot z_j \hat{z}_{jt} \cdot a_i^{\theta_j} \hat{a}_{it}. \tag{F.7}$$

This completes the construction of two wage terms $\tilde{W}_{it}$, $\bar{W}_{ct}$. For the two wage bill terms, start with firm wage bill from equations (F.2):

$$\begin{aligned} E_{jct} &= \chi \left(A_c \hat{A}_{ct} L_{ct}^{\mu} r_{ct}^{\frac{\alpha}{\alpha-1}}\right) \cdot (z_j \hat{z}_{jt})^{1+\beta_w/\rho_w} \cdot \phi_{jct} \\ &= \left(\chi r_{ct}^{\frac{\alpha}{\alpha-1}} A_c \hat{A}_{ct} L_{ct}^{\mu} \cdot z_j \hat{z}_{jt}\right)^{1+\beta_w/\rho_w} \int_a \left(a^{\theta_j}\right)^{1+\beta_w/\rho_w} \hat{a}_{it} \hat{\kappa}_{jct}(a) da \end{aligned} \tag{F.8}$$

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

$$\hat{\kappa}_{jct}(a) = \left(A_c \hat{A}_{ct} L_{ct}^{\mu} r_{ct}^{\frac{\alpha}{\alpha-1}}\right)^{\beta_w - \beta_w/\rho_w} \cdot (r_{ct})^{-\eta\beta_w} \left(G_{jc}(a)\right)^{\frac{\beta_w}{\rho_w}} \cdot \hat{\kappa}_{ct}(a) \tag{F.9}$$

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

$$\hat{\kappa}_{ct}(a) = \frac{L(a) \cdot \left[R_c(a)\left(\sum_{j\in\mathcal{J}_{ct}}\left(\Xi_c \cdot z_j a^{\theta_j} \cdot G_{jc}(a)\right)^{\beta_w/\rho_w}\right)^{\rho_w/\beta_w}\right]^{\beta_w}}{\sum_{c'}\left[R_{c'}(a) r_{c't}^{-\eta}\left(\sum_{j'\in\mathcal{J}_{c't}}\left(W_{j'c't}(a) G_{j'c}(a)\right)^{\beta_w/\rho_w}\right)^{\rho_w/\beta_w}\right]^{\beta_w} \cdot \sum_{j'\in\mathcal{J}_{ct}}\left(\Xi_c \cdot z_{j'} a^{\theta_{j'}} \cdot G_{j'c}(a)\right)^{\frac{\beta_w}{\rho_w}}}. \tag{F.10}$$

This labor supply shifter deviates from the counterpart in Lamadon et al. (2022) in the sense that the set of firms in each city, $\mathcal{J}_{ct}$, varies over time. As I will show soon, the changes in the composition of firms in the city correlate with the city productivity shock and may cause bias of the passthrough parameter. With the expression of $\hat{\kappa}_{ct}(a)$, we can rewrite firm wage bill E_{jct} as:

$$\begin{aligned} E_{jct} &= \left(\chi r_{ct}{}^{\frac{\alpha}{\alpha-1}} A_c \hat{A}_{ct} L_{ct}^{\mu} \cdot z_j \hat{z}_{jt}\right)^{1+\beta_w/\rho_w} \int_a \left(a^{\theta_j}\right)^{1+\beta_w/\rho_w} \hat{a}_{it} \kappa_{jct}(a) da \\ &= \left(r_{ct}{}^{\frac{\alpha}{\alpha-1}} A_c \hat{A}_{ct} L_{ct}^{\mu} \cdot z_j \hat{z}_{jt}\right)^{1+\beta_w} \cdot \left(\chi z_j \hat{z}_{jt}\right)^{1+\beta_w/\rho_w} \cdot \left(r_{ct}\right)^{-\eta\beta_w} \hat{\phi}_{jct} \end{aligned} \tag{F.11}$$

where I define $\hat{\phi}_{jct} \equiv \phi_{jct} \cdot \left(r_{ct}{}^{\frac{\alpha}{\alpha-1}} A_c \hat{A}_{ct} L_{ct}^{\mu}\right)^{-\beta_w} = \int_a \left(a^{\theta_j}\right)^{1+\beta_w/\rho_w} \left(G_{jc}(a)\right)^{\frac{\beta_w}{\rho_w}} \cdot \hat{\kappa}_{ct}(a) da$. Another time-varying term to derive for the passthrough equations is the time-varying city average wage bill, $\bar{E}_{ct}$, of the firms in the stayers sample, which is the sum of wage bills, E_{jct}, of these firms in city c:

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

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

$$\hat{E}_{jt} = \frac{E_{jt}}{\bar{E}_{ct}} = |\mathcal{J}_c^S| \cdot \left(z_j \hat{z}_{jt}\right)^{1+\beta_w/\rho_w} \cdot \frac{\hat{\phi}_{jct}}{\hat{\phi}_{ct}}. \tag{F.13}$$

I can now derive the passthrough equations (4.7) and (4.8) using (F.6), (F.7), (F.12) and (F.13). Taking log and first differences of (F.7) and (F.13) and plugging in the measurement error e_{jt}, we have the net passthrough equation as:

$$\Delta \log \hat{W}_{ijt} = \delta_w \Delta \log \dot{\hat{E}}_{jt} + \Delta \hat{a}_{it} + \delta_w \left(\Delta e_{jt} - \Delta \log \frac{\hat{\phi}_{jct}}{\hat{\phi}_{ct}}\right) \tag{F.14}$$

where $\delta_w \equiv \frac{1}{1+\beta_w/\rho_w}$. There are three residual terms in the net passthrough equation (F.14), that are changes in the *i.i.d.* skill transient shock Δa_{it}, changes in the firm wage bill measurement error Δe_{jt}, and changes in the relative wage bill shifter $\Delta \log(\hat{\phi}_{jct}/\hat{\phi}_{ct})$. The last term appears due to changes in the set of firms in the city c, which affects competition for workers through the wage index, $W_c(a)$, and thus the labor supply curve faced by firms. The potential correlation of the net wage bill shock with the measurement error and with the relative wage bill shifter gives 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 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 (F.14) 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, I control for $H^N(\cdot,\cdot)$ using a cubic polynomial of the two lagged first-difference terms.

Similarly, using equations (F.6) and (F.12) and defining the city passthrough parameter $\delta_c \equiv \frac{1}{1+\beta_w}$, I have the mean passthrough equation as:

$$\Delta \log \bar{W}_{ct} = \delta_c \Delta \log \bar{\dot{E}}_{ct} + \delta_c \eta \beta_w \Delta \log r_{ct} + \delta_c \left(\Delta e_{ct} - \Delta \log \hat{\phi}_{ct}\right), \tag{F.15}$$

which can be further re-arranged as:

$$\Delta \log \left(\frac{\bar{W}_{ct}}{r^\eta_{ct}}\right) = \delta_c \Delta \log \left(\frac{\bar{\dot{E}}_{ct}}{r^\eta_{ct}}\right) + \delta_c \left(\Delta e_{ct} - \Delta \log \hat{\phi}_{ct}\right). \tag{F.16}$$

Therefore, 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, as long as the measurement error of $\bar{\dot{E}}_{ct}$. I assume that e_{ct} also follows a $MA(k)$ process, the same as e_{jt}. Same as the identification of the net passthrough equation, I instrument $\Delta \log \left(\bar{\dot{E}}_{ct}/r^\eta_{ct}\right)$ using its lags before year $t-q-1$ and control for the two lagged first-differenced terms, $\Delta \log \left(\bar{W}_{ct-1}/r^\eta_{ct-1}\right)$ and $\Delta \log \left(\bar{\dot{E}}_{ct-1}/r^\eta_{ct-1}\right)$.

F.3 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 (F.12) as:

$$\bar{E}_{ct} = \frac{\chi^{1+\beta_w/\rho_w}\left(A_c\hat{A}_{ct}L_{ct}^{\mu}r_{ct}^{\frac{\alpha}{\alpha-1}}\right)^{1+\beta_w}(r_{ct})^{-\eta\beta_w}}{|\mathcal{J}_c^S|}\cdot\hat{\phi}_{ct}. \tag{F.17}$$

Then, I derive the total firm wage bill in city c, year t by summing up total firm wage bills by cluster k:

$$\begin{aligned} E_{ct} &= \sum_k J_{kct}\cdot\bar{E}_{kct} \\ &= \left(\chi\cdot r_{ct}^{\frac{\alpha}{\alpha-1}}A_c\hat{A}_{ct}L_{ct}^{\mu}\right)^{(1+\beta_f)(1+\beta_w/\rho_w)}(r_{ct})^{-\eta\beta_w(1+\beta_f)}\cdot\hat{\Phi}_{ct} \end{aligned} \tag{F.18}$$

where I define $\hat{\Phi}_{ct} \equiv \sum_k \left(\hat{\phi}_{kct}z_k^{1+\beta_w/\rho_w}\right)^{1+\beta_f} J_k\cdot\frac{1}{\sum_c \bar{E}_{kct}^{\beta_f}}$. The second line of the above equation makes uses of two properties of the model. First, firm profit is proportional to wage bill, as shown in equations (F.3) and (F.2). Hence, choosing a city based on profits is equivalent to choosing a city based on wage bills. Second, firms in the same cluster have the same probability of choosing a city.

Combining equations (F.17) and (F.18) and introducing a measurement error of $\bar{E}_{ct}$, I can obtain the following passthrough equation for firms:

$$\Delta\log\dot{\bar{E}}_{ct} = \delta_f\Delta\log\dot{E}_{ct} + \delta_f\left(\Delta e_{ct} - \Delta\log\frac{\hat{\phi}_{ct}}{\hat{\Phi}_{ct}}\right). \tag{F.19}$$

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

F.4 Worker skill and firm producitivity

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:[38]

$$\begin{aligned}\bar{E}_{kct} &= \frac{1}{|\mathcal{J}^S_{kc}|} \cdot \sum_{j \in \mathcal{J}^S_{kc}} E_{jct} \\ &= \frac{\chi^{1+\beta_w/\rho_w} \cdot \left(A_c \hat{A}_{ct} L^{\mu}_{ct} r_{ct}^{\frac{\alpha}{\alpha-1}}\right)^{1+\beta_w} (r_{ct})^{-\eta\beta_w}}{|\mathcal{J}^S_{kc}|} (z_k)^{1+\beta_w/\rho_w} \cdot \hat{\phi}_{kct}\end{aligned} \tag{F.20}$$

and

$$\hat{E}^{k(j)}_{jct} = \frac{E_{jct}}{\bar{E}_{kct}} = (\hat{z}_{jt})^{1+\beta_w/\rho_w}. \tag{F.21}$$

It can also be shown that the city-level productivity shock can be extracted from the mean city log earnings of the stayers $\bar{W}^S_{ct}$. Therefore, with equations (F.1), (F.21) and (F.6) and assuming that the firm and city productivity shocks have mean zero across years, I can show that

$$\begin{aligned}\log \check{W}_{ijt}(a) =& \log W_{ijt} - \left(\delta_w \left(\log\left(\hat{E}^k_{jct}\right) - \log\left(\bar{\hat{E}}^k_{jc}\right)\right) + \left(\log \bar{W}^S_{ct} - \log \bar{W}^S_c\right)\right) - \frac{\alpha}{\alpha-1} \log \bar{r}_c - \log \mathbb{A}_c \\ =& \log z_{j(i,t)} + \theta_{j(i,t)} \log a_i + \hat{a}_{it}\end{aligned} \tag{F.22}$$

where $\mathbb{A}_c \equiv A_c \bar{L}^{\mu}_c$.

F.5 Firm and city amenities

The amenities terms $R_c(a)$ and $G_j(a)$ do not change over time, so I use average worker shares across cities and firms and average wage indices to derive them. Using worker's sorting probabilities specified in equation (2.3) and (2.4), we can obtain

$$\Lambda_{kc}(a) = \frac{\bar{J}_c(k) \left(z_k a^{\theta_k} G_k(a)\right)^{\frac{\beta_w}{\rho_w}}}{\sum_{j' \in \mathcal{J}_c} \left(z_{j'} a^{\theta'_j} G_{k'}(a)\right)^{\frac{\beta_w}{\rho_w}}} \tag{F.23}$$

$$\Lambda_c(a) = \frac{\left[R_c(a) \bar{r}_c^{-\eta} \cdot \mathbb{A}_c r_c^{\frac{\alpha}{\alpha-1}} \left(\sum_{j \in \mathcal{J}_c} \left(z_j a^{\theta_j} G_k(a)\right)^{\frac{\beta_w}{\rho_w}}\right)^{\rho_w/\beta_w}\right]^{\beta_w}}{\sum_{c'} \left[R_{c'}(a) \bar{r}_{c'}^{-\eta} \cdot \mathbb{A}_{c'} r_{c'}^{\frac{\alpha}{\alpha-1}} \left(\sum_{j' \in \mathcal{J}_{c'}} \left(z_{j'} a^{\theta_{j'}} G_{k'}(a)\right)^{\frac{\beta_w}{\rho_w}}\right)^{\rho_w/\beta_w}\right]^{\beta_w}} \tag{F.24}$$

where $\Lambda_{kc}(a)$ is the average share of skill-a workers in cluster-k firms, conditional on living in city c; $\Lambda_c(a)$ is the average share of skill-a workers in city c; and J_{kc} is the average number

[38] For the first iteration, I will cluster firms directly using observed log earnings $\log W_{ijt}(a)$. For other iterations, I will use the classification from the previous iteration.

of cluster-k firms in city c. Within the same skill a, multiplying the amenity terms with any constant will not change workers' allocation to cities and firms. Thus, I am allowed to impose a normalization for every worker skill a, that is

$$\sum_{c}\left[R_c(a)\bar{r}_c^{-\eta}\cdot\mathbb{A}_c r_c^{\frac{\alpha}{\alpha-1}}\left(\sum_{j\in\mathcal{J}_c}\left(z_j a^{\theta_j}G_k(a)\right)^{\frac{\beta_w}{\rho_w}}\right)^{\rho_w/\beta_w}\right]^{\beta_w}=1. \tag{F.25}$$

Plugging the normalization into the two sorting shares given by equations (F.23) and (F.24) yields:

$$R_c(a)\cdot G_k(a)=\frac{\bar{r}_c^{\eta}}{\mathbb{A}_c r_c^{\frac{\alpha}{\alpha-1}}z_k a^{\theta_k}}\cdot\bar{J}_c(k)^{\frac{\beta_w}{\rho_w}}\cdot\Lambda_{kc}(a)^{\frac{\rho_w}{\beta_w}}\Lambda_c(a)^{\frac{1}{\beta_w}}. \tag{F.26}$$

F.6 City productivity

From equation (F.3), I can obtain a cluster-k firm's average profits in city c over time as

$$\bar{\pi}_c(k)=\Psi\cdot\left(\mathbb{A}_c\bar{r}_c^{\frac{\alpha}{\alpha-1}}\right)\cdot(z_k)^{1+\beta_w/\rho_w}\cdot\phi_{kc} \tag{F.27}$$

where

$$\phi_{kc}=\int_a\left(a^{\theta_k}\right)^{1+\beta_w/\rho_w}L_c(a)\cdot\frac{G_k(a)^{\beta_w/\rho_w}}{\sum_{j'\in\mathcal{J}_c}\left(z_{j'}a^{\theta_{j'}}\cdot G_{j'c}(a)\right)^{\frac{\beta_w}{\rho_w}}}da. \tag{F.28}$$

We can see that a cluster k firm's profit in city c is determined by rent-adjusted city composite productivity, $\mathbb{A}_c\bar{r}_c^{\frac{\alpha}{\alpha-1}}$, firm productivity z_k and a labor composite term ϕ_{kc}, which summarizes the skill-specific labor supply conditions and labor market competitiveness by city. This labor composite term thus informs the expected total efficiency unit of labor that the firm can hire in the city. With the estimates of labor supply parameters (β_w,ρ_w), worker skill a, firm production parameters (z,θ) and amenity parameters $G_k(a)$, I can construct ϕ_{kc} for each city-cluster pair. With equations(F.27) and (F.28) into (2.12), I can obtain a relationship between average firm sorting shares over time $p_c(k)$, city productivity terms $\mathbb{A}_c\bar{r}^{/(\alpha-1)}$, and the labor composites ϕ_{kc}:

$$p_c(k)=\frac{\left(\left(\mathbb{A}_c\bar{r}_c^{\frac{\alpha}{\alpha-1}}\right)\cdot\phi_{kc}\right)^{\beta_f}}{\sum_{c'}\left(\left(\mathbb{A}_{c'}\bar{r}_{c'}^{\frac{\alpha}{\alpha-1}}\right)\cdot\phi_{kc'}\right)^{\beta_f}}. \tag{F.29}$$

The key intuition here is that firm productivity z_k does not affect its location decision as it can be carried to all locations. I can calculate the sorting shares $p_c(k)$ from the data, once all firms are grouped into clusters. The firm sorting elasticity β_f has been estimated in Section F.3. City average housing rents $\bar{r}_c$ are observed and the housing share α is calibrated to be 0.06. Therefore, the equation above gives $K \times C$ firm sorting equations to identify C city composite productivity parameters, with one free normalization.

F.7 Agglomeration elasticity

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

$$\log \mathbb{A}_c = A_0 + \mu \log \bar{L}_c + \epsilon_c^A \tag{F.30}$$

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. An OLS estimate of the parameter μ is biased if city population is correlated with its exogenous productivity. To obtain a causal estimate, I construct the immigration-based instrument as follows:

$$IV_c^L = \sum_o \frac{\sum_{t=1980}^{t=2002} IMM_{o,c,t}}{L_{c,2002}} \cdot \log I\hat{M}M_o \tag{F.31}$$

where, o denotes the origin country of immigrants, $IMM_{o,c,t}$ is the number of immigrants from o arriving in city c in year t, and $I\hat{M}M_o$ is the total number of immigrants from o arriving in Canada since 2002. This shift-share instrument ideally uses the city population share of immigrants from each origin at a base year. However, IMDB data (part of CEEDD) only tracks immigrants since 1980. Still, the ratio of accumulated immigrant arrivals since 1980 to the 2002 city population should approximate each city's exposure to future immigration. The instrument is correlated with migrant inflows, which affect city population, and relies on the assumption that $I\hat{M}M_o$, the shift term, is orthogonal to city-level exogenous productivity.

F.8 Housing supply parameters

Equating housing supply with demand given by equations (C.13) and (C.14), we have the following relationship, with γ_c as the inverse housing supply elasticity of city c:

$$H_{ct}^S(r_{ct}) = \bar{H}_c^0 \cdot r_{ct}^{\frac{1}{\gamma_c}} = \left(\tau\eta + \frac{\alpha}{1-\alpha}\frac{1+\beta_w/\rho_w}{\beta_w/\rho_w}\right)\frac{E_{ct}}{r_{ct}} = H_c^D(r_{ct}). \tag{F.32}$$

Recall that $\bar{H}_c^0 \equiv (1+\gamma_c)^{-1/\gamma_c} \cdot \bar{H}_c{}^{(1+\gamma_c)/\gamma_c}$ is a city-specific parameter governing the level of housing supply. Re-arranging the equation, we have:

$$r_{ct}^{\frac{1+\gamma_c}{\gamma_c}} = \frac{1}{\bar{H}_c^0} \cdot \left(\tau\eta + \frac{\alpha}{1-\alpha} \frac{1+\beta_w/\rho_w}{\beta_w/\rho_w} \right) E_{ct} = \frac{1}{\bar{H}_c^0} \cdot EH_{ct} \tag{F.33}$$

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, we have an estimating equation:

$$\Delta \text{log } r_{ct} = \Gamma_c \Delta \log EH_{ct} + \Delta e_{ct}^r \tag{F.34}$$

where $\Gamma \equiv \frac{\gamma_c}{1+\gamma_c}$ is the parameter of interest and e_{ct}^r represents the measurement error of observed housing rents. This equation relates changes in the housing rent to changes in the housing expenditure. Following Saiz (2010), I assume that the housing supply elasticity is affected by the share of land unavailable for housing development in each city, denoted as $UNAVAL_c$, and parameterize Γ_c as a function of this share, i.e. $\Gamma_c = \Gamma + \Gamma_L UNAVAL_c$. The equation above then becomes

$$\Delta \text{log } r_{ct} = \left(\Gamma + \Gamma_L UNAVAL_c\right) \Delta \log EH_{ct} + \Delta e_{ct}^r. \tag{F.35}$$

City-level housing rent and housing expenditure changes may be due to unobserved labor and housing market shocks. To deal with the endogeneity concerns, I follow Diamond (2016) to instrument $\Delta \log EH_{ct}$ using a shift-share bartik IV. Specifically, the instrument is constructed as:

$$IV_c^{EH} = \sum_{ind} \frac{L_{c,ind,2002}}{L_{c,2002}} \times \Delta \log wage_{ind} \tag{F.36}$$

which is a base-year industry employment share weighted sum of national wage increases by industry. The inverse housing supply elasticity parameter can then be recovered as $\gamma_c = \frac{\Gamma_c}{1-\Gamma_c}$. I choose $\bar{H}_c^0$ to fit the average housing rent of each city, according to equation (F.33). 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.

Appendix G Model extensions

G.1 Endogenous amenity

Diamond (2016) shows that endogenous amenity is a key mechanism for the fact that high-skilled workers increasingly concentrate in high-wage, high-rent cities. I incorporate this mechanism in my model in a reduced-form way following Fajgelbaum and Gaubert (2020). Suppose 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:

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

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 the four γ parameters, $\gamma_{gg'}$, are the amenity spillover elasticities from skill group g to g'. I define high-skilled workers as those with skill a in the top three deciles, and the rest as low-skilled workers. I calibrate the spillover parameters following Fajgelbaum and Gaubert (2020) as $\gamma_{HH} = 0.77$, $\gamma_{LH} = -1.24$, $\gamma_{HL} = 0.18$, $\gamma_{LL} = -0.43$.

G.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

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

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

G.3 Remote work

The model can also be extended to incorporate remote work.

Labor market. Assuming 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'. In other words, these remote workers can be matched to firms that are outside their residence locations. The labor supply curve of such workers can be written as

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

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

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

where $L^O_c\left(a\right)$ is the measure of skill-$a$on-site workers in city c. They post skill-specific wages to hire both remote and on-site workers,

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

Firms are infinitesimal in both on-site and remote labor markets. Consequently, firms charge a constant markdown under the marginal product of labor from both remote and on-site workers. Then, firms' optimal profits can be re-written as:

$$\pi_{jc} = \Psi \cdot r_c^{\frac{\alpha}{\alpha-1}} \cdot \left(Q_{jc}^{R} + Q_{jc}^{O}\right)$$

where $Q_{jc}^{R}$and Q_{jc}^{O} are the total efficiency units of remote and on-site labor. I assume that remote workers contribute to the agglomeration spillovers in their residence city.

Location choice. On-site workers choose locations in the same way as equation (2.4). Remote workers choose locations in only based on local amenities and costs of living

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

I assume that in the short run, firm locations remain fixed, while in the long run, firms can relocate to adjust.

Appendix H Shapley value

In the counterfactual exercises studied in Section 5.1.1, I use the Shapley value approach to decompose the contribution of different sources of variation to equilibrium spatial sorting and inequality outcomes. Here, I formally discuss this approach. Let $\Omega = \cup_{n=1}^{N}\omega_n$ denote a vector of estimated model parameters, where each ω_n is a parameter that affects the equilibrium outcome of interest; let $X(\Omega)$ denote the value of the equilibrium outcome X using the parameter vector Ω. I compute values of equilibrium outcome X using counterfactual values $\hat{\omega}_n$ and use these to evaluate the contribution of each parameter ω_n. Let $\mathcal{N} = \{1, 2, , N\}$ and then let $\hat{\Omega}_S = \{\cup_{n\in S}\hat{\omega}_n\} \cup \{\cup_{n\notin S}\omega_n\}$ denote the parameter vector with counterfactual values for the parameter subset $S \subseteq N$. The Shapley value X_n for parameter ω_n in relation 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 the parameter ω_n to the outcome X is $\frac{X_n}{X(\Omega)}$, which by construction sums to one for all $n \in \mathcal{N}$.

When implementing the method, I set the counterfactual values as follows. For the city-characteristics parameters, I set $\hat{A}_c = 1$, $\hat{\mu} = 0$, and $\hat{\alpha} = 0$ to eliminate variations that come from exogenous city characteristics and endogenous population and housing rents. For amenity parameters, I set $\hat{R}_c(a) = \hat{R}_c = \frac{\sum_a L_c(a) R_c(a)}{L_c(a)}$ and $\hat{G}_k(a) = \hat{G}_k = \frac{\sum_c \sum_a D_{kc}(a) G_k(a)}{\sum_c \sum_a D_{kc}(a)}$. Intuitively, these counterfactual values eliminate the variation of city and firm amenities for different skilled workers, while preserving the average amenities of all cities and firms. For the skill-augmenting parameters, I set their counterfactual values as the weighted average of all firms, i.e. $\hat{\theta}_k = \frac{\sum_k J(k)\theta_k}{\sum_k J(k)}$.

Appendix I Additional Counterfactual Analysis

I.1 Remote work

I apply the model to quantitatively investigate the spatial impact of the rise of remote work. Deng et al. (2020) estimate that 39% of Canadian workers hold jobs that can be done at home, following the methodology developed by Dingel and Neiman (2020). The ability to work from home is biased towards high-skilled workers: the potential shares are 60% and 29% for college-educated and non-college-educated workers, respectively. I evaluate such impacts through the lens of the two-sided sorting model. I simulate an economy where 25% of workers who can potentially work from home become remote workers, that is, 15% for high-skilled workers and 7% for low-skilled workers.

The results in Table J.15 show that remote work shifts workers from large to smaller cities. In the short run, the five largest cities lose 2.2% of their population, with a 3.8% decline among high-skilled workers. These workers relocate to smaller, more affordable cities, reducing population concentration in large cities while raising mean earnings in smaller ones. Consequently, the variance of city mean log earnings drops by 19.2%. Another benefit is an increase in total output by 8.1%, as remote work allows workers in smaller cities to work for high-productivity firms that typically locate in larger cities. In the long run, firms also relocate, amplifying the population decline in large cities to 2.7%. However, spatial inequality decreases less as firm and worker co-movements create new high-paying, mid-size cities. Total output increases further to 9.4% as firms optimize their locations.

Table J.16 shows that cities with low housing costs and strong amenities for high-skilled workers benefit from remote work. These cities are predicted to attract more in-migration of high-skilled remote workers and enjoy greater local earnings growth.

I.2 Low-wage city subsidies

In Table J.17, I present counterfactual results of a 10% wage subsidy in the 25 lowest-paying cities, which account for 12% of the total population. Such placed-based transfers are often considered by governments to support workers in low-wage areas. As before, I simulate the policy under model scenarios with different assumptions on heterogeneity and mobility.

Column (1) shows that the subsidy that amounts to 1% of total output goes a long way in reducing spatial inequality: the variance of city log earnings decreases by 28.0% and the city Gini index decreases by 11.6%. The redistribution occurs at a cost of a 0.3% decrease in the total output, as the subsidy increases economic activities in low-productivity cities. The population in the subsidized cities increases by 20.9% and the number of firms increases by 24.3%. Furthermore, there is no variation in the response of workers with different skills and of firms with different productivity. The policy also hurts the welfare of high-skilled workers slightly to help the lower-skilled workers in distressed cities. The model scenarios with limited heterogeneity have both qualitatively and quantitatively similar results to the full model. This suggests that when a policy does not specifically target particular worker or firm subgroups, neglecting one-sided heterogeneity does not substantially affect policy evaluation.

I.3 Housing supply regulation

Lastly, I conduct a counterfactual experiment where I increase the housing supply elasticity of the five highest-earning cities by 25%. Such policies have been advocated by the literature as an effective approach to promote housing affordability and reduce spatial misallocation (e.g. Hsieh and Moretti (2019)). Assessment of the policy under different model scenarios are shown in J.19.

Total output rises in all four models as the policy can 1) induce movements of workers and firms into more productive cities and 2) increase the supply of floor spaces which are inputs for the good-producing firms. Comparing the four columns, however, it can be seen that the full model predicts the smallest output gains. This is because after accounting for worker and firm heterogeneity, the importance of city fundamentals in driving the productive advantage of a region is small, as can be seen in Table 2. Hence, the simplified models that do not fully consider worker and firm heterogeneity overstate the benefit.

I.4 Universal Basic Income (UBI)

The Universal Basic Income (UBI) transfer program has sparked debate in Canadian public policy. Advocates highlight its potential to reduce poverty and inequality, with pilot programs in Ontario and Quebec. I use the model to examine the spatial effects of a \$1,000

CAD transfer (about 2% of average earnings), funded by a flat labor income tax. Results are shown in Table J.20.

Column (1) reveals that UBI significantly improves social welfare by redistributing from high-skill to low-income workers. However, it unintentionally worsens spatial inequality by increasing purchasing power more in low-rent, low-productivity cities, especially for low-skilled workers. This incentivizes their relocation, drawing low-productivity firms with them and exacerbating spatial skill segregation and income inequality. The effects are much smaller in the model without worker heterogeneity (Column 2). The spatial effects are muted in limited sorting scenarios (Columns 3–4). Without worker mobility, UBI reduces income inequality without affecting spatial inequality. When only worker re-sorting occurs, fewer relocate to low-wage cities, dampening spatial inequality.

Appendix J Additional Tables and Figures

Table J.1: Summary statistics

Sample	Baseline	Stayers	Movers
Number of Observations			
Worker-Years (in 1,000)	51475	7504	23572
Workers (in 1,000)	10655	938	4127
Firms (in 1,000)	1588	16	1250
Cities	66	66	66
Worker Characteristics			
Mean Log. Earnings	10.44	10.85	10.29
Mean Age	41.68	43.84	39.96
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

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

Table J.2: City-size regressions of mean and dispersion of log earnings

	Dependent variable:			
	Mean Log Earnings	Var. Log Earnings	90-50 Gap	50-10 Gap
	(1)	(2)	(3)	(4)
Log Population	0.023**	0.024***	0.033***	0.014**
	(0.008)	(0.005)	(0.004)	(0.006)
Constant	−0.296**	0.228***	0.330***	1.013***
	(0.107)	(0.066)	(0.059)	(0.073)
Observations	66	66	66	66
R^2	0.109	0.238	0.454	0.084

Note: This table displays the results of city-size regressions of city mean and dispersion measures of log earnings. Log earnings are residualized by a cubic polynomial of age, gender, marital status, and the number of children using a Miner-type regression. I follow Card et al. (2013) to assume that the earnings profile is flat at age 40. The 90-50 gap is the difference between the 90^{th} and the 50^{th} percentile of log earnings in the city, and the 50-10 gap is the difference between the 50^{th} and the 10^{th} percentile. Population is measured as the average number of full-time working individuals in each city. All regressions are weighted by population. *p<0.1; **p<0.05; ***p<0.01.

Table J.3: Decomposition of city-size regressions of mean and variance of log earnings

	Dependent variable:						
	Mean log earnings			Variance log earnings			
	Total	Mean Worker	Mean Firm	Total	Var. Worker	Var. Firm	2× Covar.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log Population	0.022**	0.016**	0.006*	0.024***	0.016***	0.001**	0.005***
	(0.008)	(0.005)	(0.003)	(0.005)	(0.003)	(0.000)	(0.001)
% Explained	100.0%	73.3%	26.7%	100.0%	66.2%	3.4%	19.8%
Observations	66	66	66	66	66	66	66
R^2	0.106	0.141	0.051	0.238	0.278	0.067	0.154

Note: This table displays the decomposition results of city-size regressions of city mean and variance of log earnings in 2010-2017. The worker and firm FEs are estimated using equation (B.2). Columns (2) and (3) represent city mean worker and firm FEs. Columns (5)-(7) are the within-city variances of worker FEs and firm FEs, and two times the covariance of worker and firm FEs. The within-city variance decomposition follows equation (B.4). The shares explained by Columns (5)-(7) do not sum up to 1 due to the presence of the earnings residual ϵ. Population is measured as the average number of full-time working individuals in each city in 2010-2017. All regressions are weighted by population. *p<0.1; **p<0.05; ***p<0.01.

Table J.4: Alternative decomposition of urban earnings premium

	Total	Mean Worker	Mean Firm	Mean Interaction
	(1)	(2)	(3)	(4)
Log Population	0.022**	0.015**	0.006	0.001
	(0.008)	(0.005)	(0.004)	(0.001)
% Explained	100.0%	68.7%	28.7%	2.6%
Observations	66	66	66	66
R^2	0.106	0.145	0.046	0.032

Note: This table displays the decomposition results of city-size regressions of city mean log earnings. The worker and firm FEs are estimated using the empirical specification with worker-firm interaction: $\log w_{it} = z_{j(i,t)} + \theta_{j(i,t)} a_i + \epsilon_{it}$. Mean city log earnings can then be expressed as $\mathbb{E}_c(\log w_{it}) = \bar{\theta} \cdot \mathbb{E}_c(a_i - \bar{a}) + \mathbb{E}_c(z_{j(i,t)} + \theta_j \bar{a}) + \mathbb{E}_c\left[(\theta_{j(i,t)} - \bar{\theta})(a_{i,t} - \bar{a})\right]$. Columns (2)–(4) represent city mean worker, firm, and match effects, which correspond to the three RHS terms of the decomposition equation. Population is measured as the average number of full-time working individuals in each city. All regressions are weighted by population. *p<0.1; **p<0.05; ***p<0.01.

Table J.5: Earnings variance decomposition at individual and city levels

	Individual level		City level	
	Value (1)	Share (2)	Value (3)	Share (4)
Log earnings or city mean log earnings	0.698	100%	0.0107	100%
Variance Components				
Var(Worker)	0.428	61.3%	0.0044	41.1%
Var(Firm)	0.037	5.3%	0.0014	13.1%
Var(Residual)	0.119	17.0%	0.0002	1.9%
Covariance Components				
2×Cov(Worker, Firm)	0.114	16.3%	0.0047	43.9%

Note: This table displays the between-individual and between-city earnings variance decomposition results. Worker and firm FEs are estimated by equation (B.2), with $k = 10$ firm clusters. The variance of city mean log earnings is decomposed according to: $\text{Var}\left(E_c\left[w_{it}\right]\right) = \text{Var}\left(\bar{a}_c\right) + \text{Var}\left(\bar{z}_c\right) + 2 \times \text{Cov}\left(\bar{z}_c, \bar{z}_c\right)$. Columns (1) and (3) show the values of the variance terms, and columns (2) and (4) show the percentage of each term with respect to the total variance.

Table J.6: 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	66	66	66	66

Note: 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 J.7: Alternative decomposition of the city-size gradient of within-city earnings variance

	Total	Var. Worker	Var. Firm	2× Covar	Var+Covar. Interaction
	(1)	(2)	(3)	(4)	(5)
Log Population	0.024**	0.013**	0.002**	0.006***	0.001
	(0.005)	(0.003)	(0.001)	(0.002)	(0.001)
% Explained	100.0%	51.7%	8.2%	25.4%	3.2%
Observations	66	66	66	66	66
R^2	0.238	0.181	0.080	0.153	0.013

Note: This table displays the decomposition results of city-size regressions of city-level variance in log earnings in 2010–2017. The worker and firm FEs are estimated using the empirical specification with worker-firm interaction: $\log w_{it} = z_{j(i,t)} + \theta_{j(i,t)} a_i + \epsilon_{it}$. Variance city log earnings can then be expressed as $Var_c(\log w_{it}) = \bar{\theta}^2 \cdot Var_c(a_i) + Var_c(z_{j(i,t)} + \theta_j \bar{a}) + 2 \times Cov[\bar{\theta} a_i, z_{j(i,t)} + \theta_j \bar{a}] +$ Inter. Terms, where the last represents variance terms related to θ. Columns (2)–(5) represent each of the four RHS terms of the decomposition equation. Population is measured as the average number of full-time working individuals in each city. All regressions are weighted by population. *p<0.1; **p<0.05; ***p<0.01.

Table J.8: 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 wage bill shock	0.27***	0.13***	0.14***				
	(0.00)	(0.02)	(0.02)				
City real mean wage bill shock				0.28***	0.30***	0.25***	0.32***
				(0.00)	(0.00)	(0.00)	(0.00)
Control for:							
Lag changes in wage and wage bill			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	2534400	2534400	2534400	2534400	2534400	2534400
F statistic (First Stage)		4466.9	3924.5		159646.2	148297.0	102495.5

Note: This table reports estimation results of passthrough of firm and city wage bill shocks to worker earnings by equation 4.7 and (4.8). These equations 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.

Table J.9: Heterogeneous passthrough estimates

	Firm Passthrough				*City Passthrough*			
	Big city	Small city	High-skilled	Low-skilled	Big city	Small city	High-skilled	Low-skilled
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Firm wage bill shock	0.15***	0.12***	0.12***	0.08***				
	(0.02)	(0.02)	(0.02)	(0.02)				
City mean real wage bill shock					0.37***	0.21***	0.30***	0.30***
					(0.00)	(0.00)	(0.00)	(0.00)
T-statistics: test for equality	0.903		1.498		255.186		5.254	
p-value	(0.367)		(0.134)		(0.000)		(0.000)	
R^2	0.01	0.01	0.02	-0.00	0.29	0.27	0.32	0.30
Num. obs.	1517900	1016500	1014000	1520400	1517900	1016500	1014000	1520400
F statistic (First Stage)	2365.26	2231.88	1969.24	2959.43	1117666.70	85271.64	66732.96	94955.58

Note: This table reports heterogeneous estimates of firm-level and city-level worker passthrough equations, for big versus small cities and high-skilled versus low-skilled workers. Big cities refer to the largest five Canadian cities. High-skilled workers are defined as workers with skills in the top three deciles. The firm-level passthrough regressions follow the specification of Column (2) in Table J.8 and the city-level regressions follow Column (5). The t-statistics with their p-values for the test of equal passthrough estimates between the sub-groups are reported. The number of observations is rounded to the nearest hundred. *p<0.1; **p<0.05; ***p<0.01.

Table J.10: Passthrough of wage bill shocks to firms

	OLS (1)	IV (2)	IV (3)
City total wage bill shock	1.06***	0.21***	0.15***
	(0.00)	(0.01)	(0.02)
Control for:			
Lag changes in city and firm mean wage bill			Yes
R^2	0.31	-0.01	0.24
Num. obs.	94100	94100	94100
F statistic (First Stage)		2237.77	1450.79

Note: This table reports estimation results of passthrough of city total wage bill shocks to firms by equation (4.9). These are estimated using a sample of firms from the 2010-2017 baseline sample that stay in the same city for at least 7 years and employ at least 10 workers each year. 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 changes in local labor market competition by including a cubic polynomial of one-period lagged changes in log city and firm mean wage bill in the regression. The number of observations is rounded to the nearest hundred. *p<0.1; **p<0.05; ***p<0.01.

Table J.11: 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

Note: Firm productivity parameter estimates by cluster, ranked by z.

Table J.12: Agglomeration elasticity

	Dependent variable: $\log \mathbb{A}_c$	
	OLS	IV
	(1)	(2)
Log Population	0.0056	0.0044
	(0.0165)	(0.0179)
R^2	0.00	0.00
Num. obs.	66	66
F statistic (First Stage)		120.8

Note: This table reports the estimation results of the agglomeration spillover elasticity by equation (4.14). Column (1) is the OLS estimate and Column (2) is the IV estimate. The instrument is the immigration-based IV with more details discussed in Section F.7. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table J.13: Housing Supply Elasticity Results

	Dependent variable: Δr_c		
	OLS	IV	IV
	(1)	(2)	(3)
Changes in Log Housing Expenditure	0.53^{***}	0.45	0.47
	(0.13)	(0.33)	(0.36)
× Share of Undev. Land			−0.01
			(0.01)
R^2	0.30	0.29	0.29
Num. obs.	66	66	66
F statistic (First Stage)		24.74	24.74

Note: This table presents the estimation results of the housing supply elasticity by equation (4.15). Column (1) is the OLS estimate and Columns (2)-(3) are the IV estimates, of which Column (3) interacts the changes in log housing expenditure with the share of undevelopable land. The instrument is the shift-share bartik IV with more details discussed in Section F.8. All regressions are weighted by city population. *p<0.1; **p<0.05; ***p<0.01.

Table J.14: 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)
Constant	−2.189	−18.909	−15.937
	(1.492)	(16.691)	(13.077)
Observations	50	50	50
R^2	0.746	0.637	0.633

Note: 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. *p<0.1; **p<0.05; ***p<0.01.

Table J.15: Counterfactual analysis: remote work

	Short-run (1)	Long-run (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%

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

Table J.16: Changes in city mean log earnings and population with remote workers

Changes in:	Log Pop.		Mean Log Earnings	
	Short-run	Long-run	Short-run	Long-run
	(1)	(2)	(3)	(4)
Log city rent	-0.028^{*}	0.005	-0.043^{***}	-0.037^{***}
	(0.014)	(0.011)	(0.010)	(0.013)
Log city HS amenities	0.232^{***}	0.103^{***}	0.147^{***}	0.126^{***}
	(0.024)	(0.019)	(0.017)	(0.022)
Log city LS amenities	-0.282^{***}	-0.177^{***}	-0.164^{***}	-0.166^{***}
	(0.027)	(0.021)	(0.019)	(0.024)
Observations	66	66	66	66
R^2	0.713	0.807	0.654	0.628

Note: Regression of changes in log city population and mean city log earnings on log city rent in the baseline equilibrium and log city high-skilled and low-skilled amenities. Short-run refers to the model scenario; Long-run refers to the model scenario where only workers can move; Column (2) displays long-run outcomes where both workers and firms move. All regressions are weighted by benchmark city population with no remote workers. *p<0.1; **p<0.05; ***p<0.01.

Table J.17: Counterfactual analysis: subsidizing low-wage cities

		Limited Heterogeneity		Limited Re-sorting		
	Full Model	No firm het.	No worker het.	No-resorting	Only firm	Only worker
% Changes in	(1)	(2)	(3)	(4)	(5)	(6)
Var. city log earnings	14.4%	14.5%	15.9%	−25.6%	−28.0%	−31.6%
City Gini index	19.5%	19.7%	18.8%	−10.8%	−11.6%	−13.7%
Total output	0.8%	0.8%	1.0%	−0.3%	−0.3%	−0.2%
Total welfare	−0.4%	−0.4%	−0.4%	−0.1%	0.3%	0.0%
High-skilled welfare	−0.1%	−0.2%	−0.4%	−0.4%	0.3%	−0.2%
Low-skilled welfare	−0.5%	−0.5%	−0.4%	0.1%	0.3%	0.1%
Pop. in treated cities	−2.3%	−2.3%	−2.1%	20.8%	20.9%	18.4%
High-skilled pop.	0.7%	0.6%	1.1%	20.6%	21.0%	21.4%
Low-skilled pop.	−3.8%	−3.8%	−3.7%	20.9%	20.8%	16.9%
# Firms in treated cities	−0.8%	−0.7%	−0.4%	24.1%	26.6%	21.7%
# high-prod. firms	30.0%	30.2%	29.6%	24.3%	27.3%	26.9%
# low-prod. firms	−7.6%	−7.6%	−7.6%	24.0%	26.5%	20.8%

Note: Low-wage cities refer to the 25 lowest-paying cities, which account for 12% of the urban population.

Table J.18: Robustness checks on place-based subsidies

% Changes in	Counterfactual 1			Counterfactual 2		
	End. amen. (1)	Free entry (2)	Het. η (3)	End. amen. (4)	Free entry (5)	Het. η (6)
Var. city log earnings	14.4%	14.5%	15.9%	−25.6%	−28.0%	−31.6%
City Gini index	19.5%	19.7%	18.8%	−10.8%	−11.6%	−13.7%
Total output	0.8%	0.8%	1.0%	−0.3%	−0.3%	−0.2%
Total welfare	−0.4%	−0.4%	−0.4%	−0.1%	0.3%	0.0%
High-skilled welfare	−0.1%	−0.2%	−0.4%	−0.4%	0.3%	−0.2%
Low-skilled welfare	−0.5%	−0.5%	−0.4%	0.1%	0.3%	0.1%
Pop. in treated cities	−2.3%	−2.3%	−2.1%	20.8%	20.9%	18.4%
High-skilled pop.	0.7%	0.6%	1.1%	20.6%	21.0%	21.4%
Low-skilled pop.	−3.8%	−3.8%	−3.7%	20.9%	20.8%	16.9%
# Firms in treated cities	−0.8%	−0.7%	−0.4%	24.1%	26.6%	21.7%
# high-prod. firms	30.0%	30.2%	29.6%	24.3%	27.3%	26.9%
# low-prod. firms	−7.6%	−7.6%	−7.6%	24.0%	26.5%	20.8%

Note: Robustness checks of the counterfactual analysis results. Counterfactual 1 refers to the subsidy to the productive firms in Toronto in Section 5.3; Counterfactual 2 refers to the low-wage city subsidy in Section I.2. The details on the incorporation of endogenous amenities and free entry are discussed in Appendix G.1 and G.2. In the heterogeneous housing expenditure shares version, I calibrate $\eta_L = 0.35$ and $\eta_H = 0.22$ according to Eeckhout et al. (2014).

Table J.19: 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%

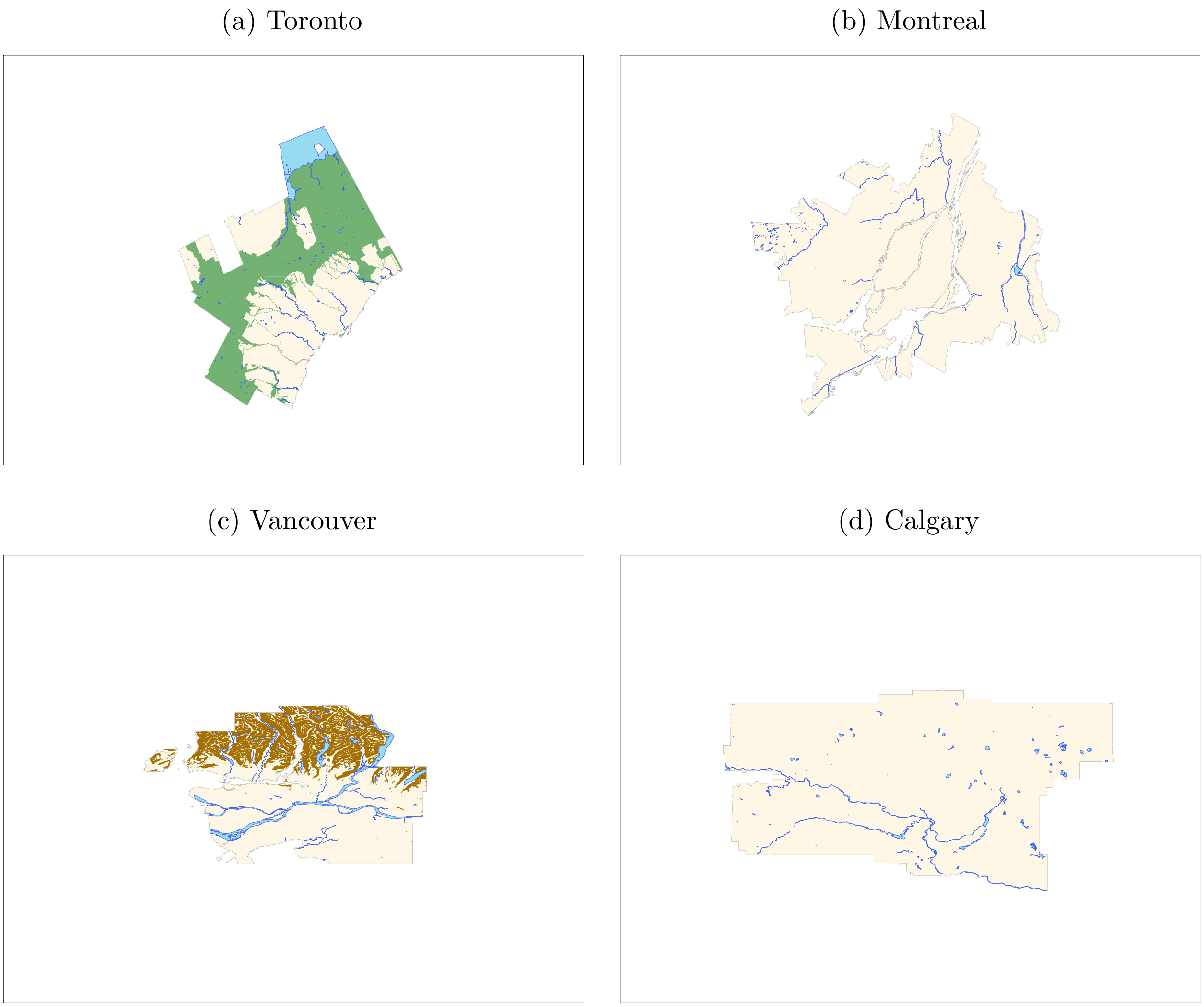
Note: The top 5 cities refer to the five cities with highest average earnings.

Table J.20: Counterfactual analysis: universal basic income

			Limited Re-sorting		
	Full model	No worker het.	No-resorting	Only firm	Only worker
% Changes in	(1)	(2)	(3)	(4)	(5)
Var. city log earnings	24.8%	0.0%	−0.1%	−0.1%	16.5%
City Gini index	7.5%	0.0%	−0.3%	−0.3%	4.2%
Total output	−0.1%	0.0%	0.0%	0.0%	−0.1%
Total welfare	3.9%	0.8%	3.8%	3.8%	3.8%
High-skill welfare	−1.4%	0.8%	−1.4%	−1.4%	−1.4%
Low-skill welfare	6.7%	0.8%	6.7%	6.7%	6.7%
Pop. in low-wage cities	1.3%	0.6%	–	–	1.0%
High-skill pop.	0.1%	0.6%	–	–	0.0%
Low-skill pop.	1.9%	0.6%	–	–	1.5%
# Firms in low-wage cities	1.4%	0.5%	–	0.0%	–
# high-prod. firms	−0.5%	0.5%	–	0.0%	–
# low-prod. firms	1.8%	0.5%	–	0.0%	–

Note: Low-wage cities refer to the 25 lowest-paying cities, which account for 12% of the urban population.

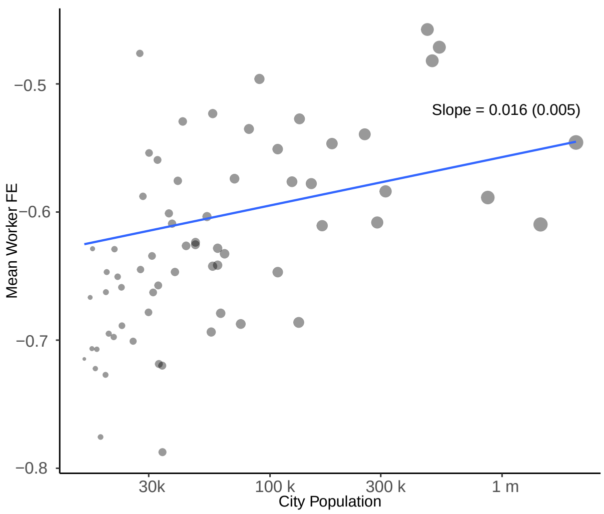
Figure J.1: Topography maps of largest four Canadian cities



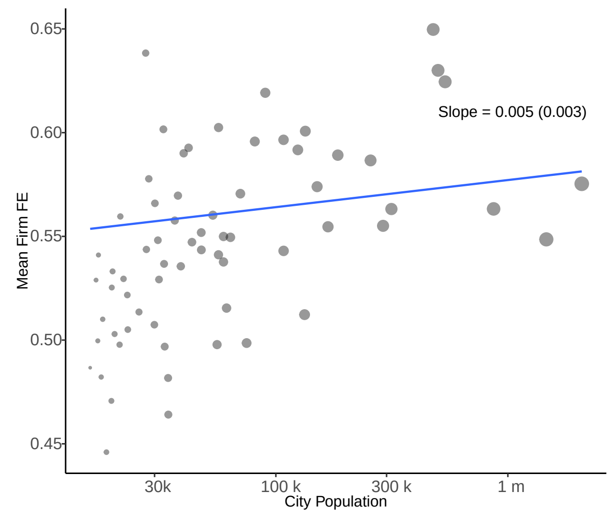
Note: Topography maps for the four largest Canadian cities. Blue areas are land areas covered by water, brown areas are land areas with slopes greater than 15 degrees, and green areas are land areas under the Greenbelt plan.

Figure J.2: Descriptive facts on worker and firm FEs

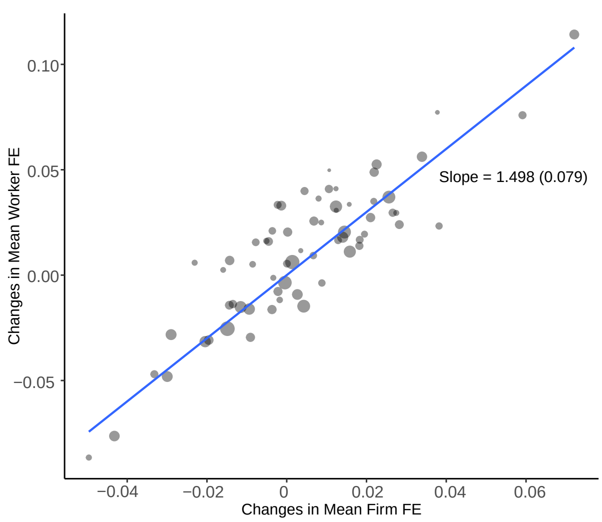
(a) City mean worker FE



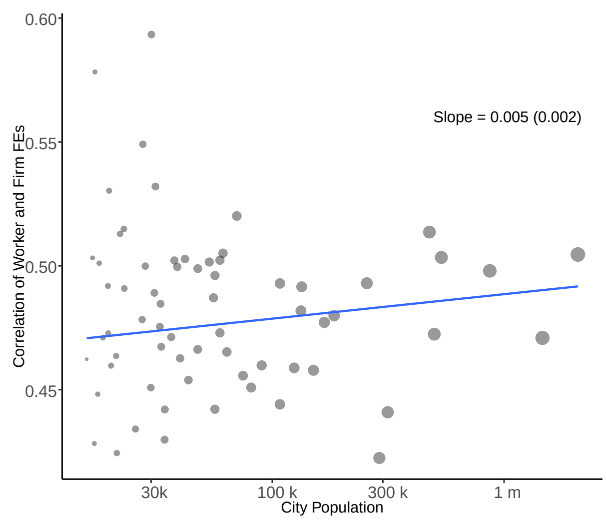
(b) City mean firm FE



(c) Changes in mean worker and firm FEs



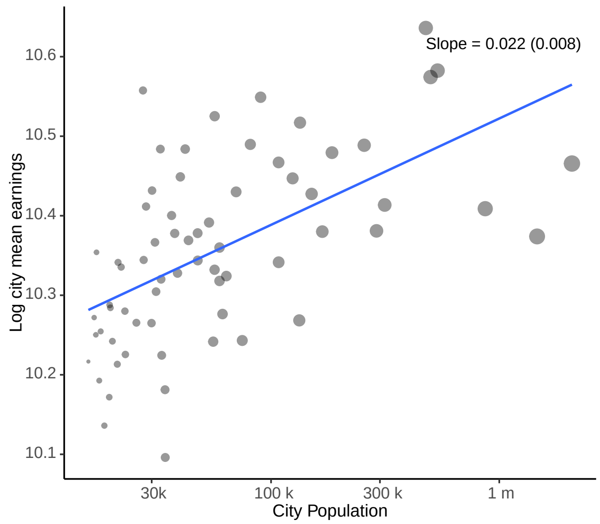
(d) Correlation of worker and firm FE by city



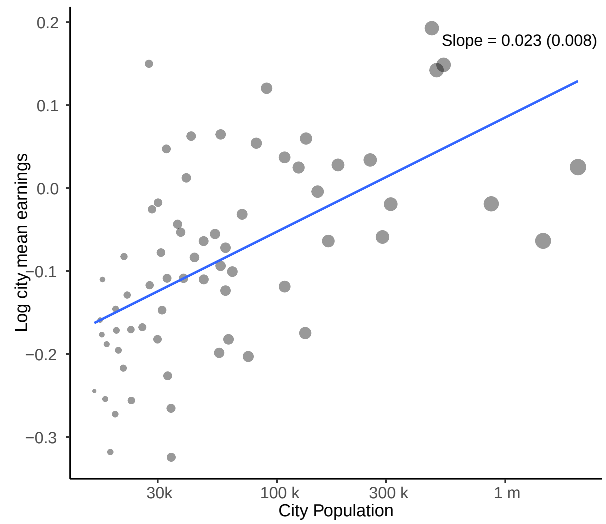
Note: Panels (a) and (b) plot city mean worker and firm FEs, estimated by equation (B.2), against city population for 2010-2017. Panel (c) plots the changes in city mean worker FEs against changes in city mean firm FE from 2002-2009 to 2010-2017, with the mean changes normalized to zero. Panel (d) plots the correlation of worker and firm FEs for each city against city population for 2010-2017. Firms are grouped into $k = 10$ clusters. Population is measured as the average number of full-time working individuals in each city in 2010-2017. Population-weighted OLS regression coefficients and standard errors are reported.

Figure J.3: City mean log earnings versus city population

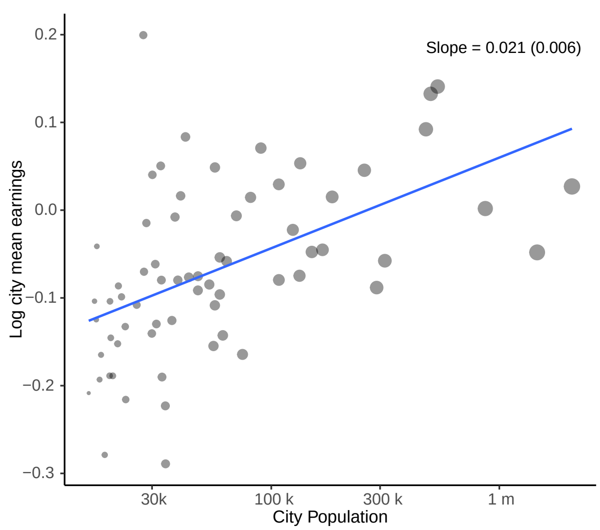
(a) Log earnings



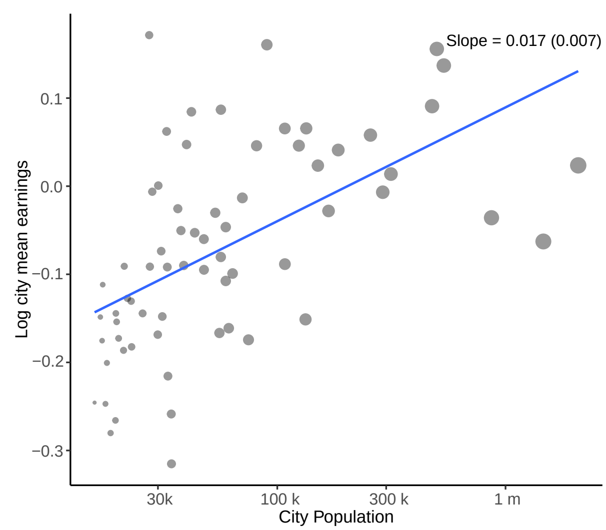
(b) Control for demographic characteristics



(c) Control for demographic characteristics + industry



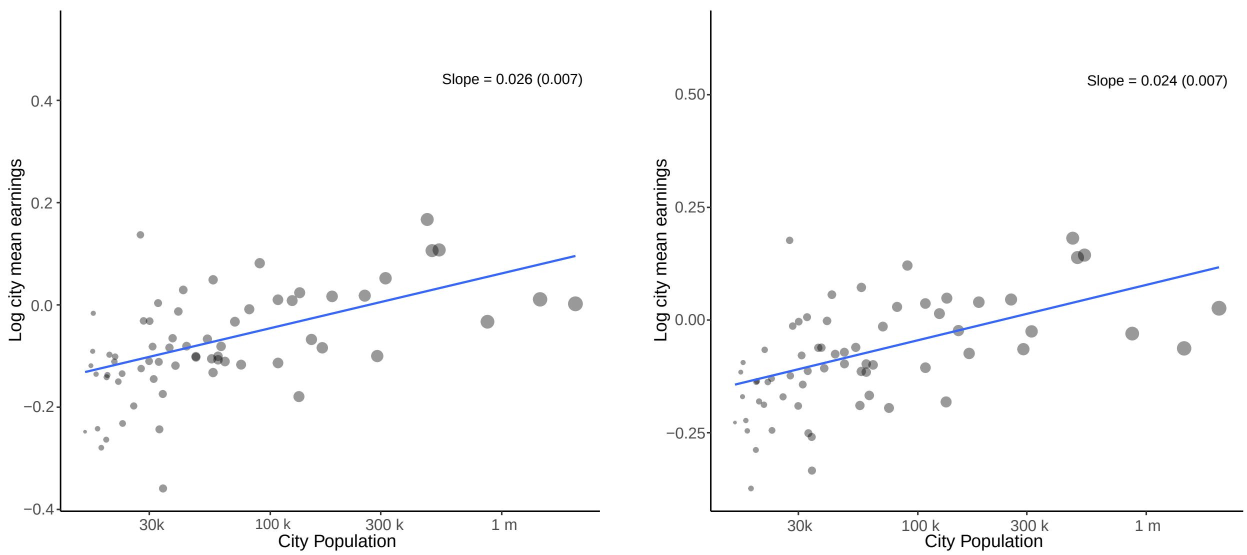
(d) Control for demographic characteristics + learning in big cities



Note: These figures display results of city-size regressions of city mean log earnings. Panel (a) uses raw log earnings; panel (b) controls for demographic characteristics including age profile, gender, marital status, and the number of children; panel (c) additionally controls for NAICS-4 industry dummies; panel (d) additionally controls for the number of years of work experiences in the big cites since 2002 interacted with the current city being a big city, where big cities refer to the largest five Canadian cities. Population is measured as the average number of full-time working individuals in each city. Population-weighted regression coefficients and standard errors are reported.

Figure J.4: City mean log earnings versus city population

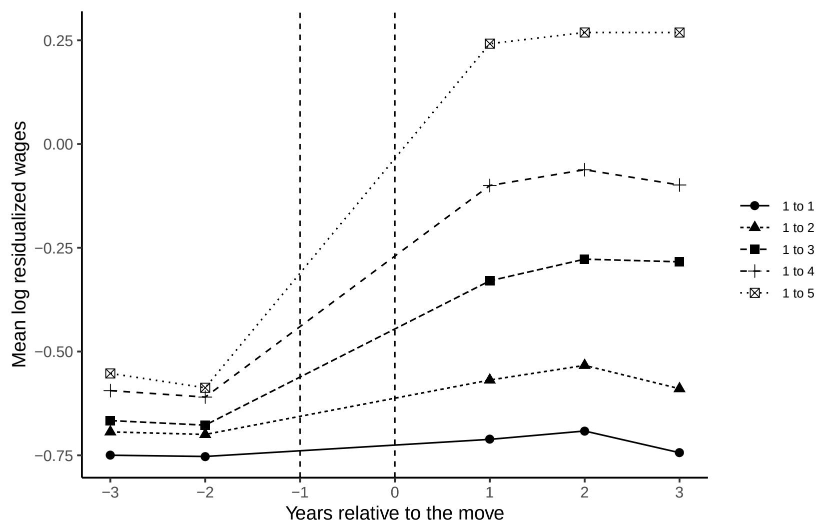
(a) After-tax labor earnings (b) Workers with no business income



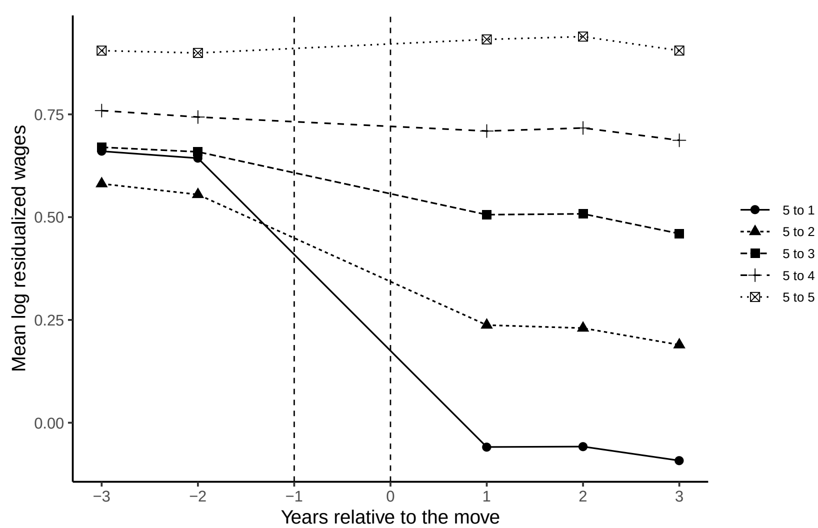
Note: These figures display results of city-size regressions of city mean log earnings for 2010-2017. Panel (a) uses after-tax annual earnings; panel (b) excludes individuals with non-zero business income. Both panels control for worker demographic characteristics. Population is measured as the average number of full-time working individuals in each city in 2010-2017. Population-weighted regression coefficients and standard errors are reported.

Figure J.5: Event study figure of between-firm worker movers

(a) Movers from the bottom cluster

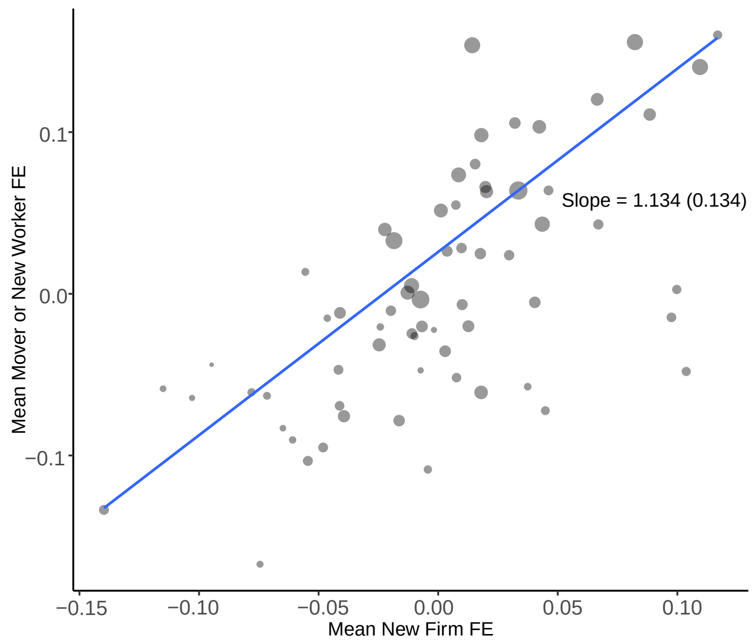


(b) Movers from the top cluster



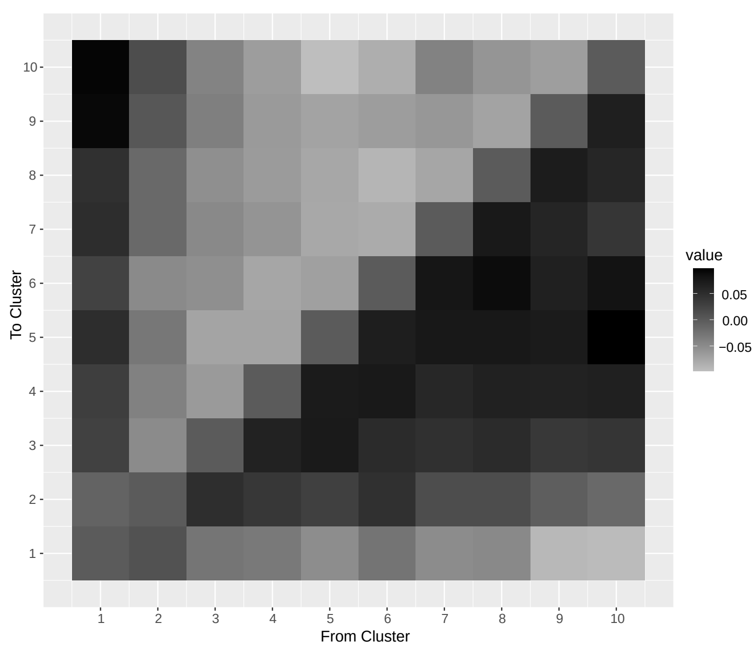
Note: These two figures display the earnings profile of between-firm worker movers before and after the move. The sample includes workers who have 1) moved between firms, 2) stayed in the previous firm for more than 4 years prior to the move and 2) stayed in the new firm for more than 3 years. The movement 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 and show workers who move from the top and bottom ones in the figures.

Figure J.6: Mean new worker and firm FEs by city



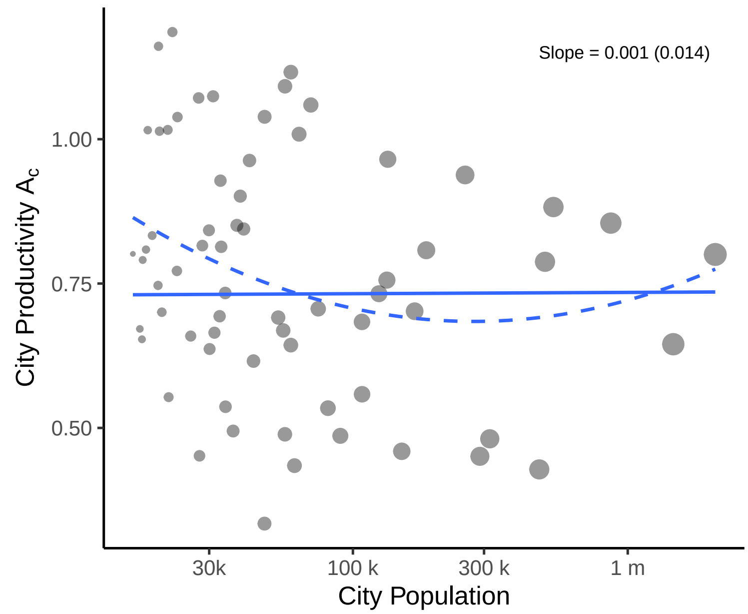
Note: This figure displays a scatterplot of the mean FEs of new or in-migration workers and of new firms for each city, in the 2010–2017 period. The worker and firm FEs are estimated using equation (B.2). Firms are grouped into $k = 10$ clusters. The population-weighted regression coefficient and the standard error are reported.

Figure J.7: Comparing mean worker skills of movers in opposite directions



Note: This figure compares mean worker skills of movers in opposite directions. The value of each $(k, k^{'})$ cell is calculated as $\mathbb{E}_{kk'}(a) - \mathbb{E}_{k'k}(a)$.

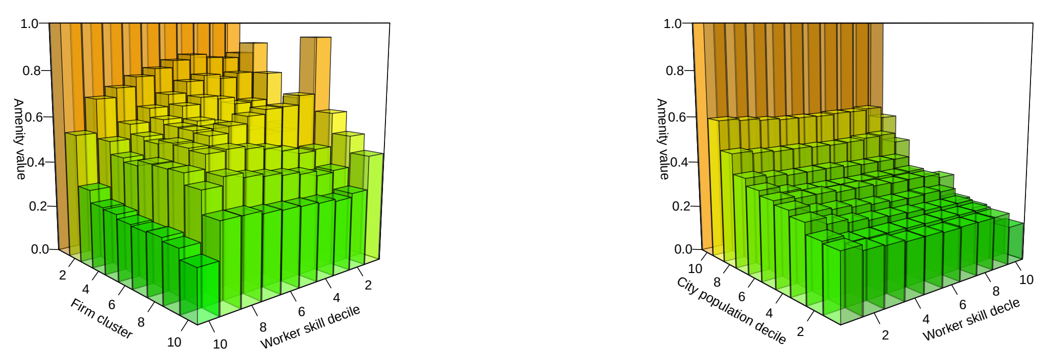
Figure J.8: City exogenous productivity estimates versus population



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

Figure J.9: 3D plots of firm and city amenities for different skilled workers

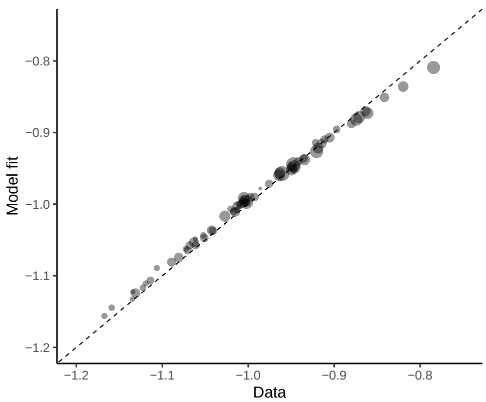
(a) Firm amenities (b) City amenities



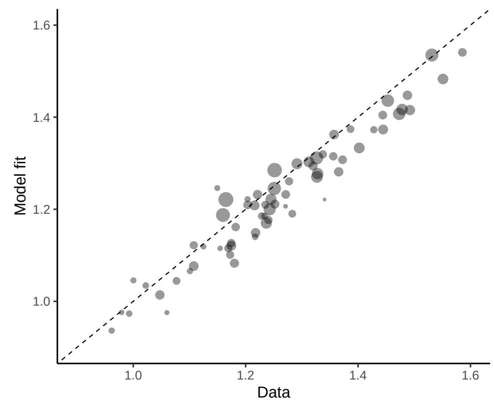
Note: This set of figures displays skill-specific firm and city amenities. Workers are binned into ten skill decile groups, and cities are ranked by population and binned into ten equal-number city groups. The amenities of firm cluster 1 and of city group 10 are normalized to 1, for all worker skill deciles.

Figure J.10: Model fit

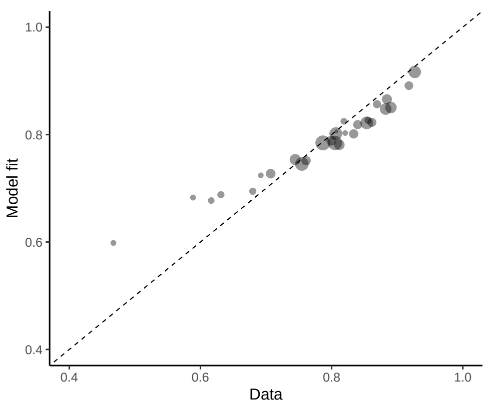
(a) Mean Worker Skill



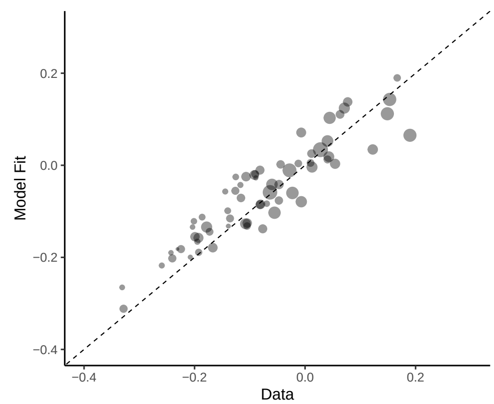
(b) Mean Firm Prod.



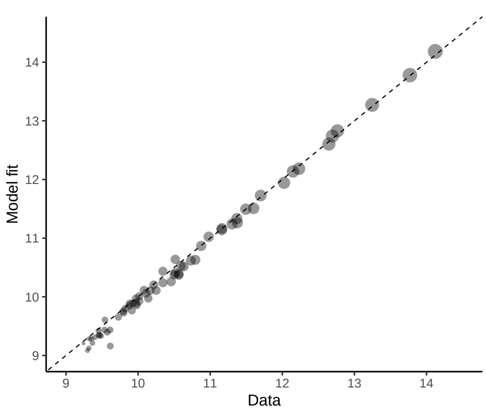
(c) Match Patterns



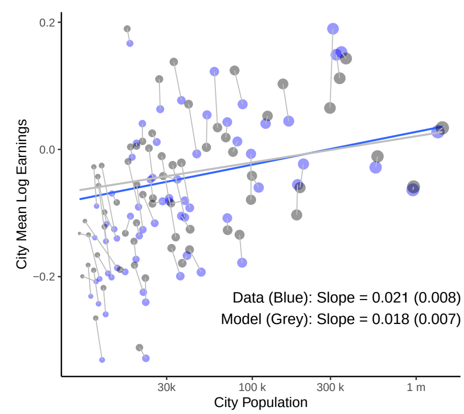
(d) Mean Log. Earnings



(e) Log. Population

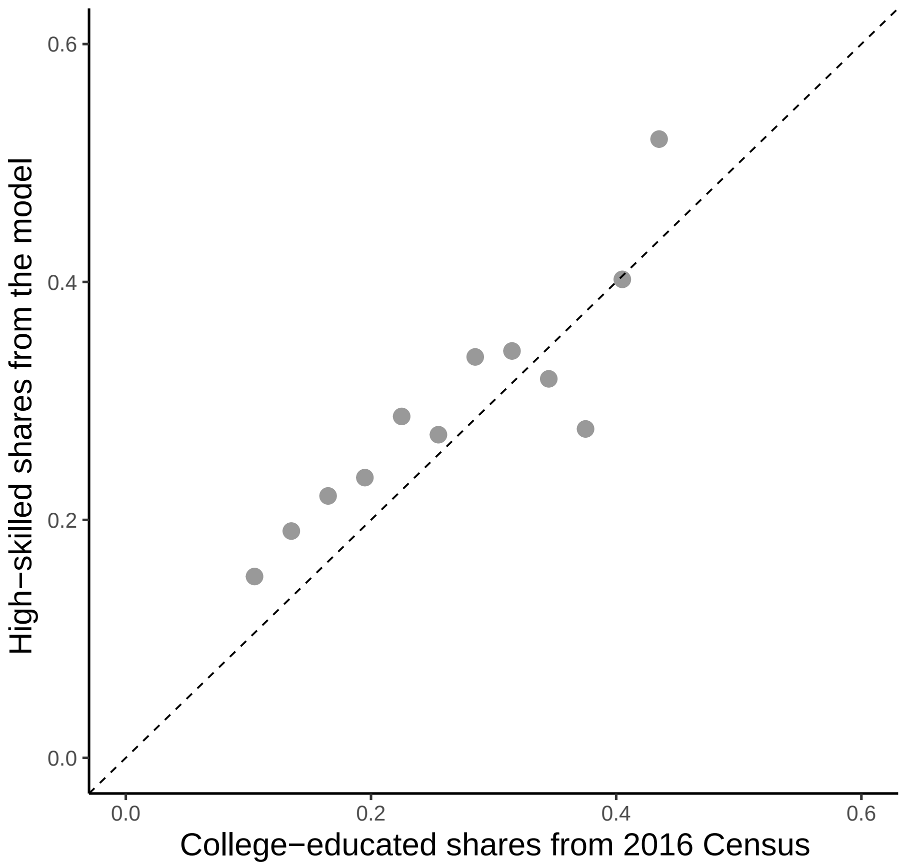


(f) Urban Earnings Premium



Note: These figures compare the city-level statistics generated by the model solution to the corresponding statistics from the data or estimates. 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.

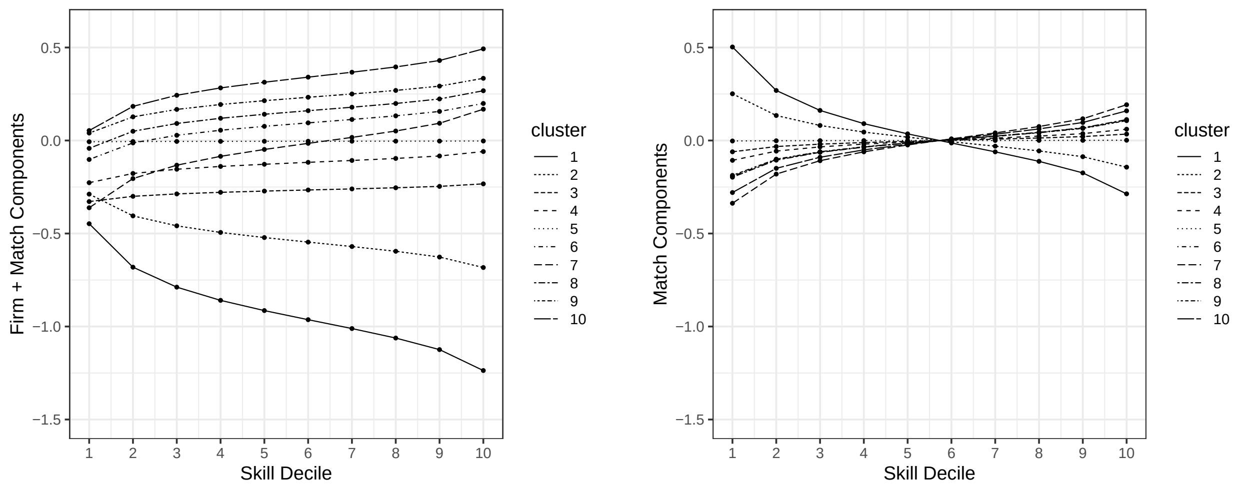
Figure J.11: High-skilled worker shares in the model versus college-educated shares from 2016 Census



Note: This figure is a binscatter plot of the model-estimated city high-skilled worker shares versus the city college-educated shares constructed using 2016 Population Census. High-skilled workers are defined as workers with skills in the top three deciles of the skill distribution. The 45-degree line is plotted in the figure.

Figure J.12: Understanding sorting and worker-firm complementarity

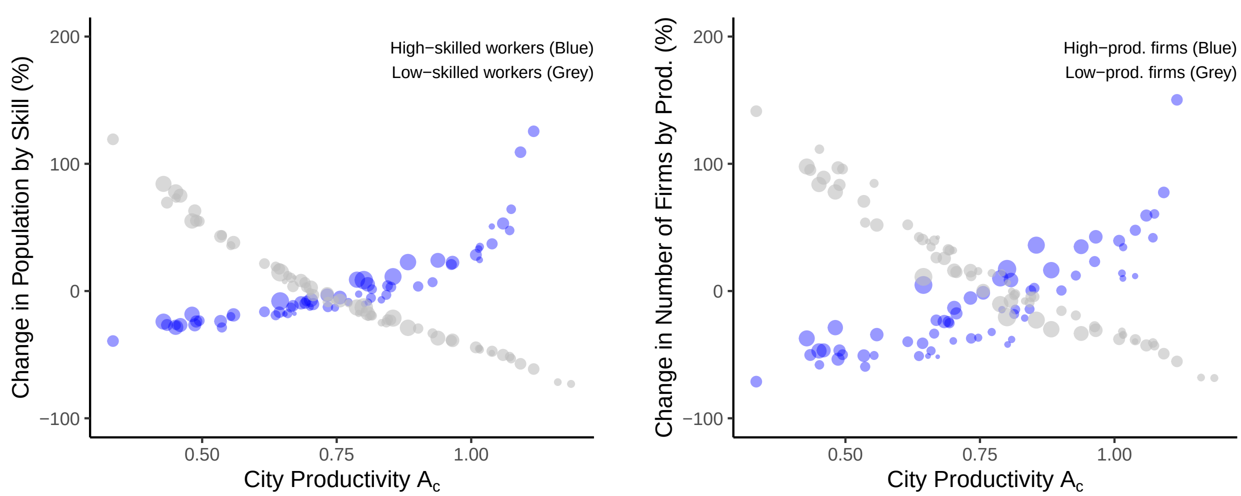
(a) Firm+Interaction Components (b) Interaction Components



Note: The two figures plot 1) the firm+interaction components and 2) the interaction component in equation (5.1) for ten firm clusters along different skill deciles. Firm clusters are ranked by productivity z.

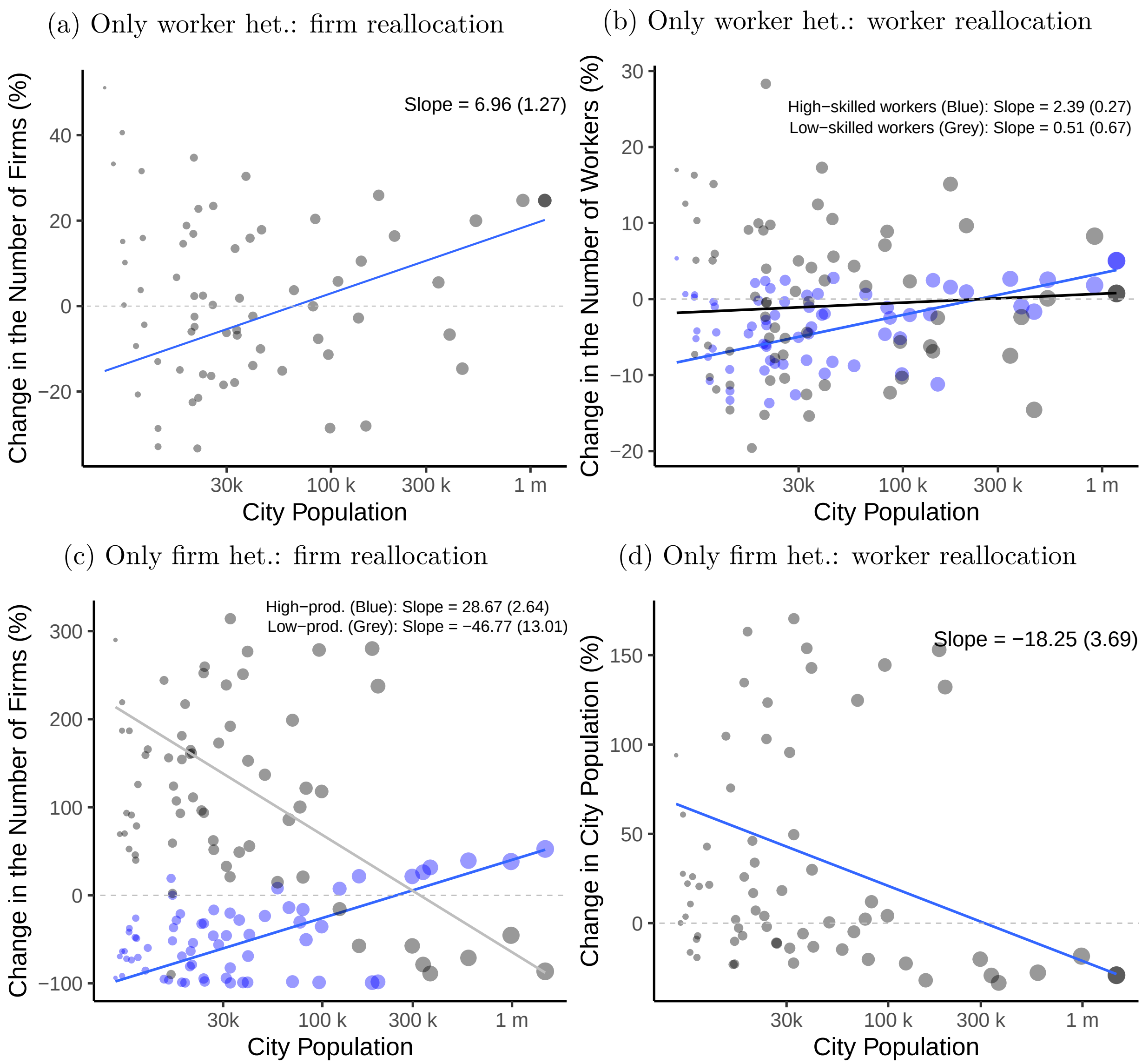
Figure J.13: Spatial reallocation of workers and firms of the optimal policy between cities with different exogenous productivities

(a) Worker reallocation (b) Firm reallocation



Note: The series of figures displays the changes in the number of workers and firms induced by the optimal policy versus city exogenous productivity A_c. 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.

Figure J.14: Spatial reallocation of workers and firms of the optimal policies with one-sided heterogeneity



Note: The series of figures displays the reallocation of workers and firms induced by optimal policies in models with only one-sided heterogeneity.