Canadian Labour Economics Forum

WORKING PAPER SERIES

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

Guangbin Hong (University of Toronto and University of Chicago)

CLEF WP #71

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

Guangbin Hong†

University of Toronto

February 20, 2024

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. I estimate the model using Canadian matched employer-employee data and decompose the urban earnings premium, finding that worker and firm sorting account for 67% and 27% of this premium, respectively. The decentralized 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, leading to a 6% increase in social welfare. Model counterfactuals underscore the importance of two-sided sorting when evaluating distributional and aggregate outcomes of place-based policies.

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

JEL codes: E25, R12, R13.

*I am deeply indebted to Nathaniel Baum-Snow, Joseph B. Steinberg, Kevin Lim, Serdar Ozkan, and Kory Kroft for their continuous guidance and support. I have benefited significantly from discussions with Gabriel Ahlfeldt, Treb Allen, Costas Arkolakis, Patrick Bayer, David Berger, Adrien Bilal, Mons Chan, Arnaud Costinot, Sebastian Dyrda, Maximiliano Dvorkin, 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, Immo Schott, Gregor Schubert, William Strange, Felix Tintelnot, Matthew Turner, Xincheng Qiu, and Yue Yu, and from presentations at Bank of Canada GSPA Workshop, the RUSH Browbag, the UEA North American Meeting, Federal Reserve Bank of St. Louis, the Asian Meeting of the Econometric Society, the Labor, Firms, and Macro Workshop, the Canadian Economic Association Annual Meeting, the UEA Summer School, and the Midwest Macroeconomic Meeting, as well as numerous seminars at the University of Toronto. 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 sincerely appreciate Lady Bolongaita, Angela Prencipe, and Bassirou Gueye for helping me with the use of the Canadian Employer-Employee Dynamics Database, and 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.

†Department of Economics, University of Toronto. Email: g.hong@mail.utoronto.ca.

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.[3]

In this paper, I develop a new two-sided sorting general equilibrium model and estimate it with Canadian matched employer-employee data. I find that worker sorting, firm sorting, and city characteristics contribute to 67%, 27%, and 6% of the earnings premium of larger cities, respectively. Production complementarity between worker skill and firm productivity is key to explaining the co-location of high-skilled workers and high-productivity firms. Novel firm sorting externalities arise from the interplay of spatial sorting and local matching. Low-productivity firms overvalue locating in larger cities, where they compete for higher-skilled workers against high-productivity firms. I evaluate the impacts of place-based policies by accounting for two-sided sorting in general equilibrium. 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, leading to a 6% increase in social welfare. I also find that commonly considered place-based subsidies may have unintended deleterious distributional or aggregate consequences. The unifying theme throughout my analysis underscores the importance of two-sided sorting in understanding spatial inequality and evaluating place-based policies.

I begin with some descriptive evidence to motivate my analysis. First, city mean earnings and within-city earnings variances significantly increase with city population in Canada, which is consistent with existing findings in other countries (e.g. Baum-Snow and Pavan (2013) for the U.S. and Dauth et al. (2022) for Germany). Second, to investigate the role of worker and firm heterogeneity, I decompose individual earnings into worker and firm fixed effects following Abowd et al. (1999), henceforth AKM, and perform a set of decomposition exercises. Higher worker and firm fixed effects in larger cities explain about three-quarters and one-quarter of the urban earnings premium, respectively. Third, stronger positive assortative matching in larger cities contributes to about one-fifth of the city-size gradient of within-city earnings variance.[4] In sum, the evidence sheds light on the role of spatial sorting and local matching of heterogeneous workers and firms in spatial earnings inequality. However, the reduced-form evidence

[1]See Combes et al. (2008), Eeckhout et al. (2014), and Diamond (2016) for evidence of worker sorting, Gaubert (2018), Bilal (2023), and Lindenlaub et al. (2022) for evidence of firm sorting, and Bauluz et al. (2023) for cross-country evidence of spatial inequality.

[2]For example, many U.S. states compete for high-skilled workers by offering low taxes (Moretti and Wilson, 2017). Many local governments offer substantial subsidies to attract productive firms to locate in their jurisdictions (Greenstone et al., 2010), including the recent bidding war for Amazon HQ2. For reviews of the literature on place-based policies, see Glaeser and Gottlieb (2008); Neumark and Simpson (2015); Austin et al. (2018); Duranton and Venables (2018); Ehrlich and Overman (2020); Bartik (2020); Slattery and Zidar (2020).

[3]See Behrens and Robert-Nicoud (2015) and Diamond and Gaubert (2022) for reviews of the literature on spatial sorting.

[4]Using the same methodology, Dauth et al. (2022) document that larger cities have higher-quality workers and firms, and greater extents of assortative matching than smaller cities in Germany. Card et al. (2023) show that these patterns are also present across U.S. cities. However, neither of these two papers aims to rationalize the spatial allocation of workers and firms using a general equilibrium model or study the welfare implications of spatial sorting.

can neither explain why workers and firms choose locations in such ways nor be used to conduct policy counterfactuals.

Motivated by the descriptive evidence, I build a general equilibrium model in which heterogeneous workers and firms sort across cities and match within cities. The model incorporates the new monopsony model (Card et al., 2018; Manning, 2021; Lamadon et al., 2022) into the Rosen-Roback model (Rosen, 1974; Roback, 1982), thereby establishing a link between two-sided spatial sorting and local matching in a general equilibrium context. In the model, workers differ in skills, while firms differ in their production technologies and skill-specific amenities. The model allows for production complementarity between worker skill and firm productivity, which generates greater match returns for higher-skilled workers and higher-productivity firms. Cities differ in exogenous productivity, skill-specific amenities, and the amount of land, which is used by local developers to produce housing for residential and production purposes. City productivity is also endogenously related to the population through agglomeration spillovers.

Workers and firms choose locations to maximize utilities and profits, respectively.[5] Within a city, each firm posts skill-specific wages and each worker chooses which firm in the city to work for based on wages, non-wage amenities, and idiosyncratic preferences that are unobserved by firms. Local matching affects city-specific utilities for workers and profits for firms, which in turn influence their spatial sorting choices. One key implication is that an increase in the number of firms in a city benefits local workers by offering a wider array of employment options, yet it decreases local firms' profits by congesting the local labor market. The equilibrium interaction of these forces determines the spatial sorting and matching of workers and firms, resulting in the spatial wage structure.

To study the welfare implications of two-sided sorting, I then solve the social planner's problem and characterize sorting inefficiencies in the laissez-faire equilibrium. I find two novel types of firm sorting externalities, namely the labor market stealing and love-of-variety externalities, that cause spatial misallocation.[6] Specifically, when a firm chooses to locate in a city, it competes for workers from other firms either in the city or in other cities, which impacts the profits of affected firms and aggregate output. In addition, this firm generates welfare gains for local workers through a love-of-variety effect, since workers view local firms as imperfectly substitutable workplaces. Neither of these two effects is internalized by firms when they make location choices.

The externalities are a result of spatially segmented and imperfectly competitive local labor markets. Workers' idiosyncratic preferences give rise to firms' local labor market power. However, the idiosyncratic preferences are unobserved to firms, which makes them unable to offer individual-specific wages to account for the preference heterogeneity. This information asymmetry results in a fundamental market failure in local labor markets, generating distortions in the *lassez-faire* equilibrium through firms' inefficient location choices. The cost of spatial misallocation is greater when there exists worker-firm production complementarity, amplifying negative labor market stealing externalities caused by low-productivity firms when they compete for high-skilled workers.Therefore, a utilitarian planner would reallocate low-productivity firms towards smaller, lower-skilled cities to achieve the efficient spatial allocation. In addition, the planner would redistribute income from larger, high-wage cities towards smaller, low-wage cities for

[5]The location choices are also affected by workers' and firm owners' idiosyncratic preferences.

[6]As widely studied in the literature (e.g. Fajgelbaum and Gaubert (2020)), worker sorting is inefficient as workers do not internalize the agglomeration spillovers, which benefit both local workers and firms.

equity considerations.

I structurally estimate the model using Canadian matched employer-employee data. Despite its complexity, the model yields estimating equations that facilitate transparent identification of the model parameters. Most parameters, including the sorting elasticity parameters, the amenity parameters, and the parameters related to housing demand and supply, can be identified from the estimating equations following established approaches in the literature. A key challenge is the separate identification of city, firm, and worker productivity parameters, which are crucial for quantifying the importance of two-sided sorting.In particular, the model gives rise to an empirical earnings equation that is additive in city productivity, firm productivity, and worker-firm interaction, where the last term is a product of the corresponding firm complementarity parameter and worker skill. Applying a movers design, which is used to identify the two-way fixed effects AKM equation, necessitates both exogenous worker movements across firms and exogenous firm movements across cities.[7] However, firm movements are highly selective, endogenous, and poorly measured in the data, making them unreliable for the purpose of identification.

I address this challenge by utilizing worker movements across firms and revealed location choices of firms. First, I follow Bonhomme et al. (2019a) to identify worker skills and the firm complementarity parameters from changes in earnings when workers move across firms.[8] Then, I utilize the insight from the Rosen-Roback model to identify city productivities from revealed firm location decisions. The intuition is that unobserved city productivities can be estimated as compensating differentials to explain the observed firm sorting shares, after controlling for other determinants of firm profits (e.g. commercial rents and worker composition). For example, a city is inferred to be productive if it has high commercial rents and a tight labor market, but a disproportionately large share of firms choose to locate there. In summary, my approach leverages both the structure of my model and the richness of the matched employer-employee data. This approach allows for the separate identification of city, firm, and worker components in their contribution to earnings differentials, which is novel in the literature.

I use the estimated model to decompose spatial earnings inequality, which stems from both location heterogeneity and two-sided sorting of heterogeneous workers and firms. Quantitatively, I find that worker, firm, and city heterogeneity explain 66.8%, 26.8%, and 6.6% of the urban earnings premium, respectively.[9] Two additional insights are worth highlighting: First, I find a significant degree of production complementarity between worker skill and firm productivity, which is crucial for explaining the sorting pattern and urban premium. The interplay between production complementarity and the spatial variation in location fundamentals leads to the systematic sorting of high-skilled workers and high-productivity firms into larger, more productive cities. Model counterfactuals show that the urban earnings premium would decrease by 40.6% if production complementarity was absent. Second, the estimated agglomeration spillovers appear to be rather modest when accounting for worker and firm heterogeneity across cities. Therefore, spatial frameworks with limited heterogeneity may conflate the contributions of sorting and

[7]Although I observe between-city worker movers and within-city between-firm worker movers, it is not enough to separately identify firm and city productivities using worker movers. There will be a city-level normalization for firm productivity estimates, which prohibits the comparison of firm productivity between cities.

[8]To mitigate limited mobility bias (Andrews et al., 2008), I follow Bonhomme et al. (2019a) to group firms into clusters based on firms' 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 5 for an iteration algorithm that I develop to overcome this challenge.

[9]These three shares do not sum up to 100%. The residual -0.2% is explained by the iteration component, which captures the mean earnings premium from high-skilled workers matched with high-skill-complementary firms in each city.

agglomeration spillovers, potentially leading to an overestimation of the agglomeration elasticity.

I use the model to conduct four counterfactual exercises: 1) the optimal spatial policy, 2) local wage subsidies to productive firms in Toronto, 3) local wage subsidies in low-paying cities, and 4) the rise of remote work. The results of all four exercises underscore the importance of the sorting interdependence of workers and firms in evaluating the outcomes of spatial policies and spatially impactful changes.

I show that the optimal spatial policy, implemented with type-location-specific taxes, would strengthen the co-location of high-skilled workers and high-productivity firms in high-productivity cities. By harnessing production complementarity and better location fundamentals, this reallocation leads to a 9.6% increase in total output. Interestingly, to minimize the negative labor market stealing effects, the optimal policy tends to decrease the number of low-productivity firms in higher-skilled, larger cities, while increasing the number of high-productivity firms there. This result synthesizes the findings of Gaubert (2018) and Bilal (2023) on the optimal spatial allocation of heterogeneous firms. The former finds it is optimal to reallocate more firms to larger cities to gain from the productivity advantage and agglomeration spillovers, and the latter suggests moving unproductive employers to lower-wage cities to decrease labor market congestion in higher-wage ones. On top of that, two-sided sorting causes workers to move in a way that strengthens the co-location of better workers and firms. Furthermore, the optimal policy implements spatial transfers from high-wage cities to low-wage cities. Overall, the induced spatial reallocation and transfers give rise to a consumption-equivalent welfare gain of 6.5%.

For the counterfactual analysis of place-based subsidies and the rise of remote work, I show that the location choice interdependence of workers and firms significantly magnifies their aggregate and distributional effects. To show this, I compare the counterfactual results of the full model with model scenarios in which I limit mobility to either workers or firms. For example, I show that a 5% wage subsidy to the firms with productivity in the top five percentiles, if they choose to locate in Toronto, would increase total output and the city Gini index by 0.8% and 19.7% respectively in the full model. However, the increases are much smaller with only one-sided resorting. Total output and the city Gini index increase by only 0.3% and 3.2% respectively with only workers re-sorting, while these changes are 0.7% and 11.7% respectively with only firms re-sorting. In the model counterfactual of remote work, I show that if 25% of workers who can conduct their job from home switch to remote work, the total population and the number of firms in the largest five Canadian cities would decrease by 2.7% and 8.0% respectively.[10] Total output would increase by 9.4% because the remote work arrangement loosens the co-location constraint and thus facilitates better worker-firm matches. The spatial reallocation of workers and the change in total output are both smaller if firms' locations are held fixed.

This paper is closely related to three strands of literature. First, I contribute to the literature on spatial inequality (Combes et al., 2008; Glaeser, 2008; Moretti, 2013; Baum-Snow and Pavan, 2012; Behrens et al., 2014; Behrens and Robert-Nicoud, 2015; Davis and Dingel, 2020; Gaubert et al., 2021a; Porcher et al., 2023), which has shown that larger cities exhibit higher average wages, high-skilled worker shares, skill premia, and housing costs. There has been a growing body of literature that studies spatial sorting of heterogeneous agents, including worker sorting (Gould, 2007; Baum-Snow and Pavan, 2013; Diamond, 2016; Davis and Dingel, 2019; Diamond and Gaubert, 2022) and firm sorting (Forslid and Okubo, 2014; Gaubert, 2018; Bilal,

[10]Following the methodology developed by Dingel and Neiman (2020), Deng et al. (2020) estimate that in Canada, 60% of college-educated workers and 29% of non-college-educated workers can carry out their work from home.

2023; Lindenlaub et al., 2022), and quantifies the contribution to the urban wage premium. However, these works have studied worker and firm sorting in isolation. I contribute to this literature by developing a two-sided sorting model where the location decisions of workers and firms are interdependent. A key takeaway from my findings is that the interdependence of worker and firm sorting plays a crucial role in shaping spatial inequality and outcomes of place-based policies.

Of particular relevance to my paper is Hoelzlein (2023), who studies how two-sided sorting within a city affects the welfare impacts of place-based policies that target specific neighborhoods. In addition to focusing instead on spatial sorting across cities, my paper (1) models the matching of heterogeneous workers and firms in a local labor market, (2) separately identifies city, worker, and firm productivities using matched employer-employee data and structurally decomposes the spatial earnings differentials, and (3) formalizes the sorting externalities and designs the optimal policy. Nonetheless, I view these two papers as complementary to each other in studying the interaction of firms' and workers' location choices.

Second, I contribute to the literature on optimal spatial policies (Glaeser, 2008; Kline and Moretti, 2013; Fajgelbaum and Gaubert, 2020; Gaubert et al., 2021b; Henkel et al., 2021; Rossi-Hansberg et al., 2019), which has mainly focused on correcting inefficient sorting in the presence of local productivity and amenity spillovers. I show in my two-sided sorting framework that there is also inefficient firm sorting due to labor market stealing and love-of-variety externalities. The labor market stealing effect is similar to the one of the pooling externality studied by Bilal (2023).[11] That is, low-productive firms compete for workers with high-productive ones in large cities, but it is further magnified by the presence of firm-worker production complementarity in my framework. I also distinguish local and national labor market stealing since workers also spatially follow firms from other cities. The two sources of externality that I find are a local labor market analog of firm entry inefficiency studied by Dixit and Stiglitz (1977) and Mankiw and Whinston (1986), who show that firm entrants do not internalize business stealing from other incumbent firms and households' love-of-variety preference for more product varieties. To my knowledge, this is the first paper that formalizes such firm sorting externalities across local labor markets.

Lastly, my work relates to studies on earnings inequality that decompose earnings following AKM and subsequent papers that build on the AKM approach (Abowd et al., 1999; Card et al., 2013; Song et al., 2019; Bonhomme et al., 2019a; Lamadon et al., 2022). I contribute to this literature by incorporating geography into the earnings equation to reflect the impact of location characteristics on earnings. Some recent studies have applied the AKM approach to decompose the spatial earnings disparity into local worker and firm/industry components (Dauth et al., 2022; Card et al., 2023). In particular, Dauth et al. (2022) show that the rise in co-location of high-quality workers and high-quality plants is a key factor in explaining the growth of spatial inequality in Germany. However, the reduced-form approach cannot separately identify the city effect from the firm/industry effect. I achieve such identification with the aid of the structural model and show that the city effect explains a small yet nontrivial share of the urban earnings premium.

The rest of the paper is organized as follows. I describe the data and present the descriptive facts in Section 2. Motivated by these facts, I formulate the two-sided sorting spatial equilibrium model in Section

[11] As discussed in Bilal (2023), this externality is a spatial analog of Acemoglu (2001) in a frictional labor market with random search, where wages do not price vacancy filling rates. In my framework, the externality arises because firms' wage posting does not price its effect on labor market congestion.

3, followed by discussions on its equilibrium properties on efficiency and uniqueness in Section 4. I present how I estimate the model in Section 5, and then the quantitative analysis in Section 6. Finally, Section 7 concludes.

2 Descriptive Facts

In this section, I present descriptive facts on earnings inequality between and within Canadian cities. I show that larger cities have not only higher mean earnings but also greater earnings dispersion. I then provide suggestive evidence that such patterns are related to differences in worker and firm qualities between cities and the degrees of assortative matching within cities.

2.1 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 on firms and workers maintained by Statistics Canada. The CEEDD covers the universe of tax-paying workers and firms from 2002 to 2017. On the worker side, I observe total annual earnings from each firm in a year, residential location, as well as demographic information such as age and gender. For workers receiving earnings from multiple firms within a year, I only keep the earnings record of the firm that pays the highest earnings for that year. 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). The CEEDD does not include information on hours worked. Following Guvenen et al. (2021), I only include workers with annual earnings from the main job no less than the equivalent of working 20 hours per week for 13 weeks at the minimum hourly wage.[12] Furthermore, I exclude 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), as firms in these industries are likely to have other considerations when choosing locations and setting wages. See more details on the data and sample selection in Appendix A.

In the CEEDD, a firm is defined as an enterprise with an Enterprise ID in the Business Registry for tax and accounting purposes. For multi-location firms, indicated by a multi-location flag in the data, I only observe the location of their main headquarters. Given my research interest in decomposing earnings into firm and city components, it is important to assign each worker employed by multi-location firms to a production unit based on location. To this end, I exploit workers' residential location information to form firm-city units within multi-location firms. For example, an employee living in Ottawa and working for the Royal Bank of Canada (RBC), headquartered in Toronto, is to be assigned to an RBC - Ottawa unit. For single-location firms, one firm is equivalent to a firm-city unit. I assume each unit possesses its own production technology and makes its location choice and wage-setting decisions.[13]. I define such units as firms hereafter.

[12] Canadian provinces set their own minimum wage standards. I take the lowest of these provincial standards as the national minimum wage.

[13] The location choice of multi-plant firms is beyond the scope of this paper; see Oberfield et al. (2020), Jiang (2023) and Kleinman (2022) for a formal treatment on this topic.

A city is defined as a Census Metropolitan Area (CMA) or a Census Agglomerate (CA) delineated in the 2016 Census of Population. The concept of CMA and CAs resembles the one of commuting zones in the U.S. – they are formed by combining a population center and adjacent municipalities with a high degree of integration with the center measured by commuting flows. To only keep those with significant sizes and sufficient labor market integration, I keep CMAs and CAs with no fewer than 15,000 full-time working individuals in 2002. I 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.

To study the evolution of spatial inequality over time, I split the data into two sample periods: 2002-2009 and 2010-2017. In addition, I construct two sub-samples for each period for estimation purposes: the stayers sample and the movers sample. The selection procedure follows Lamadon et al. (2022). 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 each sample period. Following Kline et al. (2020), I restrict the movers sample to firms with at least two movers, which helps mitigate limited mobility bias. I present summary statistics for these samples in Table I.1.

2.2 Descriptive facts

2.2.1 Earnings disparities across 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, for the 2010-2017 period.[14] I show in Table 1 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. The urban earnings premium elasticity (0.023) is smaller than those estimated in other developed countries (e.g. 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)).[15] Nonetheless, spatial inequality is still sizeable in Canada and often prompts policy debates.[16]

Table I.2 shows the results of the same regressions for the 2002-2009 period. During this period, a

[14]I show the descriptive facts in the main text using the 2010-2017 baseline sample, which period I use for structural estimation to exclude the Great Recession episode. See Appendix I for the reproduction of these facts using the 2002-2009 data.

[15]Note that these estimates are obtained by controlling for different sets of worker characteristics, so they are not directly comparable. Albouy et al. (2019) controls for education (Table 1, Column (2)), Dauth et al. (2022) controls for education, education-specific age, and time effects (Figure 5 (a)), Combes et al. (2008) uses log wage directly (Table 1, Column (2)), De La Roca and Puga (2017) controls for education, occupation, experiences and firm tenure (Table 1, Column (2)). Nonetheless, they are all greater than the estimate I get using the Canadian data with fewer observables controlled for. Albouy et al. (2019) compare the urban wage premium in Canada and the U.S. They find that the earnings-population elasticity is similar when using two subsets of Canadian and American cities that overlap in the city population distribution. On the other hand, the elasticity estimate of within-city earnings variance to city population is similar to the U.S. Albouy et al. (2019) finds that in the U.S. a 1 log point increase in city population is associated with a 0.46% increase in the 90-10 gap, which is similar to 0.47% using my data by adding up 0.33% (Column (3)) and 0.14% (Column (4)) of Table 1.

[16]See here for discussions on growing concerns of regional inequality in Ontario and policy recommendations of the Opportunity Zone program.

Table 1: City-size regressions of mean and dispersion of log earnings: 2010-2017

	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 in 2010-2017. 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 in 2010-2017. All regressions are weighted by population. *p<0.1; **p<0.05; ***p<0.01.

10 log-point increase in city population is associated with a 0.35 log-point increase in mean earnings. The comparison with the estimate using the later period suggests that earnings between different-sized Canadian cities have been converging over time.

I show in Figure I.2 that the above findings are robust to controlling for industry fixed effects and time spent working in large cities; I also show in Figure I.3 that they are robust to using after-tax earnings and excluding individuals with positive business income. The discussions above can be summarized into the following fact.

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

2.2.2 Co-location of high-quality workers and high-quality 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):[17]

$$\log w_{it} = z_{j(i,t)} + a_i + \epsilon_{it} \tag{2.2}$$

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.[18] I interpret the worker and firm fixed effects as their

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

$$\log w_{it} = z_{j(i,t)} + \theta_{j(i,t)} a_i + \epsilon_{it} \tag{2.1}$$

which is more consistent with the general equilibrium model to be described in Section 3. I present the city mean log earnings decomposition using this specification in Table I.4. The results are qualitatively similar to the one estimated using equation 2.2.

[18] I follow Card et al. (2013) in restricting the earnings-age profile to be flat at the age of 40.

qualities for now. In the spatial equilibrium model that I describe later, I micro-found these fixed effects as worker skills and firm productivities. Also, note that this decomposition does not explicitly contain a city fixed effect. One can rightfully think that it is absorbed by the firm effect z as they cannot be separately identified yet.[19]

There are two well-known issues when estimating the AKM equation. First is limited mobility bias. Identification of the firm fixed effect relies on earnings changes from workers' movement across firms. As pointed out by Andrews et al. (2008), limited worker mobility makes it hard to precisely estimate the firm premium. To deal with this issue, I follow Bonhomme et al. (2019a) and group firms with similar earnings distributions into $k = 10$ clusters using the k-means clustering algorithm. The clustering approach mitigates the bias as there are many more worker movements between these clusters than between individual firms. Then, I estimate 10 cluster fixed effects rather than a large number of individual firm fixed effects. For robustness, I also perform the estimation with $k = 20, 30, 40, 50$ clusters, and the results are very similar. Despite its parsimony, the 10 (50) cluster fixed effects account for 87 (90) percent of the between-firm variance in earnings.

Second is endogenous mobility bias. The key identifying assumption for the AKM equation is that workers' selection into firms is orthogonal to the earnings residual ϵ_{it}. For example, if workers receive negative earnings shocks prior to their moves (i.e. the Ashenfelter dip), then the destination firms' premia will be overestimated. However, selection based on firm effects z, worker effects a or the controlled covariates X does not violate the identification assumption. To visualize potential selection on unobservable shocks, I follow Card et al. (2013) and plot an event-study figure of workers who move between firms and group these movers into origin-destination cluster pairs. As shown in Figure I.5, workers who move to different firm clusters from the same cluster exhibit parallel earnings trends prior to the move. In addition, workers who move to higher-paying firms on average earn more than those who move to lower-paying firms (except for those who move from cluster 5 to 1), which is consistent with the findings in Card et al. (2013).

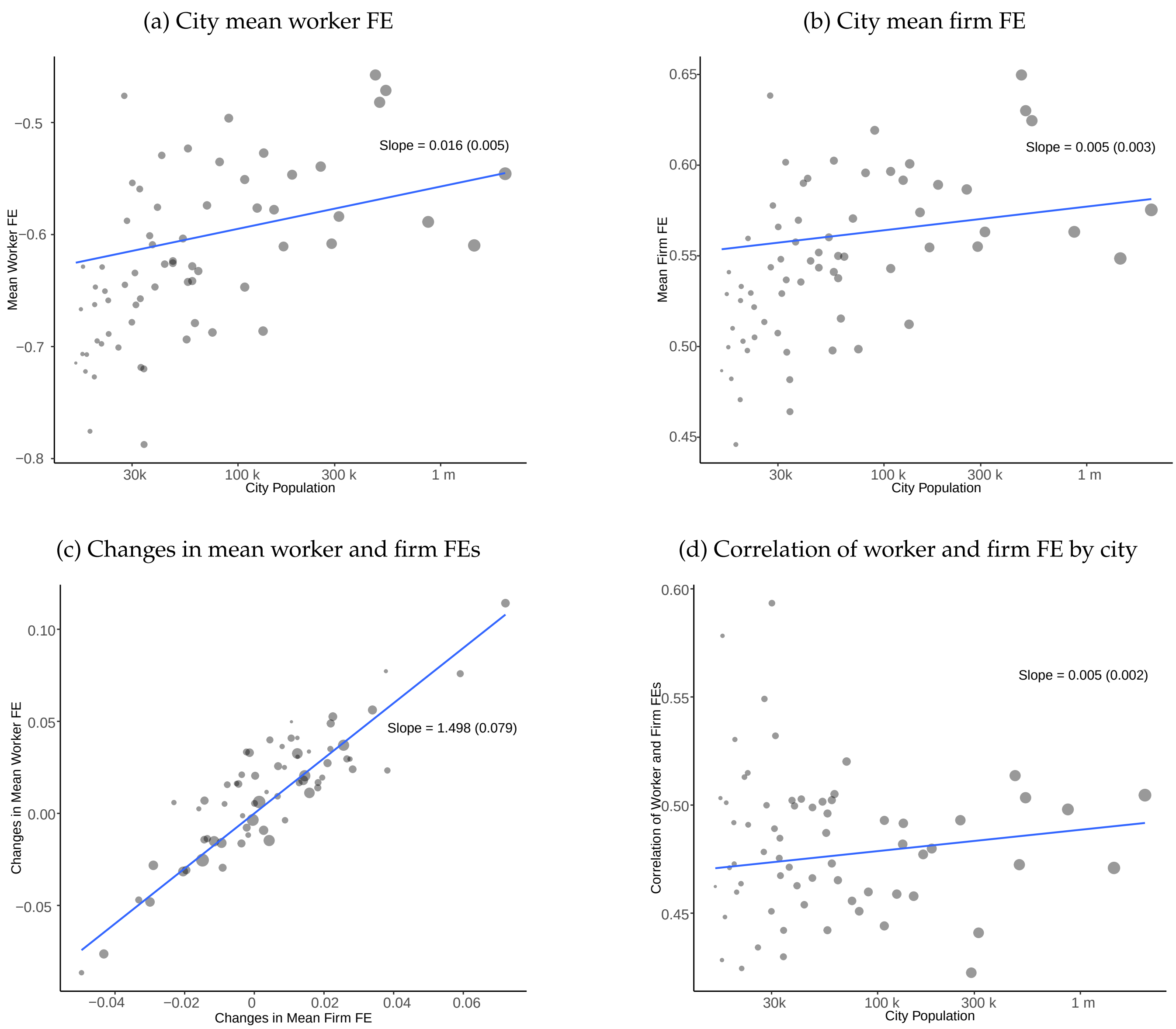
With the estimates of the AKM equation, I study how firm and worker qualities vary between cities. I plot the average firm and worker effects against city population in Figure 1 and present the results of the city-size regressions in Table 2. The results reveal significant spatial differences in worker and firm qualities. 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 urban earnings premium estimated.

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

$$\operatorname{Var}\left(\mathbb{E}_c\left[\log w_{it}\right]\right) = \underbrace{\operatorname{Var}\left(\overline{z}_c\right)}_{\text{Mean firm effect}} + \underbrace{\operatorname{Var}\left(\overline{a}_c\right)}_{\text{Mean worker effect}} + \underbrace{2\operatorname{Cov}\left(\overline{z}_c, \overline{a}_c\right)}_{\text{Co-location}} + \operatorname{Var}\left(\overline{\epsilon}_c\right), \tag{2.3}$$

[19] It is tempting to use either firm movers or worker movers within multi-location firms to separately identify the city and firm effects. For firm movers, I argue that are highly selective and poorly measured in the data. For multi-location firms, Kleinman (2022) finds that establishments within multi-location (services) firms differ significantly in their roles in the firm and productivities.

Figure 1: Descrpitive facts on worker and firm FEs: 2010-2017



Note: Panels (a) and (b) plot city mean worker and firm FEs, estimated by equation (2.2), 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.

where $\overline{z}_c$ and $\overline{a}_c$ represent city-level averages. The decomposition result is shown in Table I.3. 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 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 although the matching of individual workers and firms occurs more at random, 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.

Table 2: Decomposition of city-size regressions of mean and variance of log earnings: 2010-2017

	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 (2.2). 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 (2.4). 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.

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

Having shown the cross-sectional patterns, I now investigate how the spatial allocation of heterogeneous workers and firms evolves over time. To this end, I estimate equation (2.2) separately for 2002-2009 and 2010-2017 and then correlate changes in city mean firm effects with changes in city mean worker effects. If better workers spatially follow better firms and vice versa, then we should see a positive correlation between the two. This is borne out in the data, as shown in Panel (c) of Figure 1.[20] This result indicates strong location choice interdependence of workers and firms. Furthermore, smaller cities on average experience increases in both mean worker and firm qualities, in contrast to larger ones, which results in a reduction of the urban earnings premium over time.[21] 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.*

2.2.3 Assortative matching within cities

Large and thick labor markets have long been hypothesized to facilitate better worker-firm matches (Diamond, 1982; 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. The result is shown in Panel (d) of Figure 1, which confirms that the degree of assortative matching is higher in larger cities.[22]

[20] I have also tried using only new workers in each city, both workers entering into the labor force and those who migrate into the city, and new firms in each city, and I relate the city mean FEs of these new workers and firms. I also find a strong positive relationship between the two, as shown in Figure I.7.

[21] I show that the variance of changes in mean worker effects, the variance of changes in mean firm effects, and the covariance of the two changes each explains 43.1%, 13.6%, and 43.3% of the variance of changes in city mean log earnings.

[22] Correction of limited mobility bias is important here as failure to do so will bias downward the correlation of worker and firm effects. Moreover, the magnitude of such a bias depends on the probability of workers changing firms. Therefore, without

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

$$\operatorname{Var}_c\left(\log w_{it}\right) = \operatorname{Var}_c\left(z_{j(i,t)}\right) + \operatorname{Var}_c\left(a_i\right) + 2\operatorname{Cov}_c\left(z_{j(i,t)}, a_i\right) + \operatorname{Var}_c\left(\epsilon_{it}\right), \tag{2.4}$$

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 2. 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 qualities that contribute to greater within-city inequality, as documented in Eeckhout et al. (2014) and Combes et al. (2012). These results can be summarized into the following fact.

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

2.3 Discussion

I have shown that there is co-location of better workers and firms and a higher degree of positive assortative matching in larger cities. These two patterns should be collectively taken into consideration: spatial sorting leads to different compositions and matching patterns across cities, and local assortative matching generates different benefits across agents and thus affects their spatial sorting choices. Furthermore, changes in local worker and firm qualities over time strongly correlate with each other, which is crucial for regional rise and decline. These facts align with Moretti (2012)'s view on the regional divergence in the U.S. that "a handful of cities with [...] a solid base of human capital keep attracting good employers and offering high wages, while those at the other extreme, cities with [...] a limited human capital base, are stuck with dead-end jobs and low average wages". Therefore, place-based policies designed to mitigate spatial inequality should take the interdependence of heterogeneous workers' and firms' location choices into account.

The descriptive facts are informative about the sources of spatial inequality. However, they cannot shed light on the mechanisms underlying the observed spatial allocation, nor can it be used to evaluate the overall efficiency and specific place-based policies. In addition, it has been long posited that large cities provide better exogenous production amenities and endogenous agglomeration benefits since Marshall (1890). However, the reduced-form approach does not permit separating the true firm effects from location-specific productive advantages, thus overstating the importance of firm heterogeneity. To overcome these challenges, I will next build a model that fully specifies firms' and workers' sorting and matching problems in a system of cities.

3 Model

In this section, I build a two-sided sorting spatial equilibrium model. The model adds rich heterogeneity to the standard Rosen-Roback model (Rosen, 1974; Roback, 1982). It aims to characterize the Nash Equilibrium of location choices of heterogeneous agents, that is, in equilibrium each agent's location decision is the best response to all others' choices.

correcting the bias, higher estimated correlations in large cities may simply be due to smaller downward biases caused by more worker mobility.

3.1 Environment

Consider an economy populated by heterogeneous workers and firms. Workers, indexed by i, differ in their skill $a_i \in (\underline{a}, \bar{a})$. They obtain utility from consuming the final good and residential floor space. Firms, indexed by j, differ in their production technology summarized by (z_j, θ_j) and skill-specific amenities $\{G_j(a)\}_a$.[23] Each firm uses labor and commercial floor space to produce a homogeneous tradable final good, which is the numeraire of the economy. The measure of workers of each skill, $L(a)$, and the set of firms, $\mathcal{J}$, in the economy are exogenously given.

There are C cities in the economy. Each city, indexed by c, differs in its exogenous productivity A_c, skill-specific amenities $\{R_c(a)\}_a$, inverse housing supply elasticity γ_c and amount of land $\overline{H}_c$. Cities are local labor markets where workers and firms match with each other. There is a housing developer in each city who supplies floor spaces demanded by workers and firms. Furthermore, a city's productivity increases with its population L_c through endogenous agglomeration spillovers, with the elasticity represented as μ. I denote $\mathcal{L}_c$ and $\mathcal{J}_c$ as the sets of workers and firms located in the city c.

3.2 Worker's problem

3.2.1 Preferences

Following Lamadon et al. (2022) and Diamond and Gaubert (2022), I assume that the utility of a worker i with skill a, in city c, working for firm $j \in \mathcal{J}_c$ which offers her a wage rate $W_{jc}(a)$ is given by:

$$u_i(j, c; a) = \log\left(\frac{R_c(a)}{r_c^{\eta}}\right) + \log \tau W_{jc}(a) + \log G_{jc}(a) + \beta_w^{-1}\epsilon_{ij}, \tag{3.1}$$

where r_c denotes the housing rent of city c, η denotes the expenditure share on housing, τ represents the rebate of total land rents that is proportional to labor earnings ($\tau > 1$), and ϵ_{ij} is worker i's idiosyncratic taste for firm j.[24] The dispersion of the idiosyncratic is controlled by parameter β_w: a higher value of β_w makes the taste draws less dispersed, which implies a smaller role of the preference idiosyncrasy in determining the workers' preferred firm and location. Furthermore, I assume that the idiosyncratic taste ϵ_{ij} is drawn from a Type-I Extreme Value distribution:

$$F(\overrightarrow{\epsilon_i}) = exp\left[-\sum_c\left(\sum_{j\in\mathcal{J}_c} exp\left(-\frac{\epsilon_{ij}}{\rho_w}\right)\right)^{\rho_w}\right], \tag{3.2}$$

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 higher value of ρ_w makes the tastes of firms in the same city

[23] I will describe the production technology in detail later. The introduction of firm amenities is motivated by empirical evidence that non-pay characteristics of firms, i.e. compensating differentials, are important for understanding workers' employer choices and earnings inequality (e.g. Sorkin (2018)).

[24] Note that this is an indirect utility expression. The utility maximization problem can be written as

$$U_i(q, h; j, c) = \max_{q,h,j,c}\left\{(1-\eta)\log q + \eta \log h + \log\left[G_{jc}(a)\cdot R_c(a)\right] + \beta_w{}^{-1}\epsilon_{ij}\right\}$$

s.t.

$$q + r_c h = \tau W_{jc}(a)$$

where q represents the final good, and h represents housing.

more independent from each other. This utility function specification adds preferences for city amenities and housing rents to the one of Lamadon et al. (2022) and preferences for firm amenities and wages within cities to the one of Diamond and Gaubert (2022).[25] It combines vertical differentiation, which is differences in amenity levels, and horizontal differentiation, which comes from idiosyncratic tastes, in determining workers' choices of cities and firms. The introduction of idiosyncratic tastes also gives rise to a love-of-variety preference for more local firms, as workers view them as imperfectly substitutable workplace options. This aligns with the insight of Helsley and Strange (1990) that a larger set of horizontally differentiated firms in a local labor market facilitates better matches.

3.2.2 Labor supply and spatial sorting

Workers observe each city's amenity $R_c(a)$, housing rents r_c, the set of firms $\mathcal{J}_c$ and their wage offers $\mathbf{W}_c(a) = \{W_{jc}(a)\}_{j\in\mathcal{J}_c}$. 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 worker with skill a choosing firm j is given by:

$$S_{jc}(a; W_{jc}(a)) = L_c(a) \cdot \frac{(W_{jc}(a)G_{jc}(a))^{\frac{\beta_w}{\rho_w}}}{\mathbb{W}_c(a)}, \tag{3.3}$$

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 defined as the skill-specific wage index in city c. This equation characterizes the labor supply curve each firm faces when choosing the wage offer, with the firm-level labor supply elasticity β_w/ρ_w. Analogously, the measure of a skill a worker located in city c is then given by:

$$L_c(a) = L(a) \cdot \frac{U_c(a)^{\beta_w}}{\sum_{c'\in C} U_{c'}(a)^{\beta_w}}. \tag{3.4}$$

where $U_c(a) \equiv R_c(a)r_c^{-\eta} \cdot \mathbb{W}_c(a)^{\frac{\rho_w}{\beta_w}}$ is the expected utility of skill-a workers located in city c. Workers trade off job qualities (wage and non-wage) and city amenities versus the housing rents when making location choices, with the city-level labor supply elasticity β_w.

3.3 Firm's problem

3.3.1 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 commercial floor space, h_{jc}, with a Cobb-Douglas production technology:

$$Y_{jc}\left(\mathbf{D}_{jc}, h_{jc}\right) = \left[\int_{\underline{a}}^{\bar{a}} A_c L_c^{\mu} \cdot z_j \cdot a^{\theta_j} \cdot D_{jc}(a)\, da\right]^{1-\alpha} \cdot {h_{jc}}^{\alpha} \tag{3.5}$$

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

[25] In Appendix F.1, I have a model extension where I allow skill-specific city amenities to respond to the local skill mix (Diamond, 2016; Almagro and Dominguez-Iino, 2022). These endogenous amenity effects would further encourage worker sorting as high-skilled workers benefit from co-locating with other high-skilled workers.

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 has been proposed by Bonhomme et al. (2019a) and adopted by Lamadon et al. (2022), Gaubert et al. (2021b), Setzler and Tintelnot (2021) and Huneeus et al. (2021).

If the skill-augmenting productivity is monotonically increasing in the common productivity, then worker skill and firm productivity are complementary to each other. Mathematically speaking, this implies the matched output of a worker-firm pair is log-supermodular in worker skill and firm productivity, which leads to positive assortative matching (PAM) in local labor markets. This is consistent with Costinot (2009) and Costinot and Vogel (2015) in that output per pair must be log-supermodular to generate PAM when firm products are not perfectly substitutable. In this case, firms are imperfectly substitutable workplaces for workers. I will empirically estimate the two productivity terms later in the next section.

3.3.2 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, given one location c, it chooses optimal wage offers $\mathbf{W}_{jc} = \{W_{jc}(a)\}_a$, labor inputs $\mathbf{D}_{jc} = \{D_{jc}(a)\}_a$ and the housing input h_{jc}, which results in expected profits. Second, they choose which city to locate in based on these expected profits. The first step is specified as a standard profit maximization problem:

$$\max_{\mathbf{W}_{jc},\mathbf{D}_{jc},h_{jc}} \left\{ \left[\int_{\underline{a}}^{\bar{a}} A_c L_c^{\mu} \cdot z_j \cdot a^{\theta_j} \cdot D_{jc}(a)\, da \right]^{1-\alpha} \cdot {h_{jc}}^{\alpha} - r_c h_{jc} - \int_{\underline{a}}^{\bar{a}} W_{jc}(a) D_{jc}(a)\, da \right\} \tag{3.6}$$

subject to the labor supply curves specified by equation (3.3):

$$D_{jc}(a) = S_{jc}(a; W_{jc}(a)) = L_c(a) \cdot \frac{(W_{jc}(a) G_{jc}(a))^{\frac{\beta_w}{\rho_w}}}{\mathbb{W}_c(a)}, \forall a. \tag{3.7}$$

Following Card et al. (2018) and Lamadon et al. (2022), I assume that firms are infinitesimal, so any firm's wage offer does not affect the labor supply curve for other firms.

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

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

With this assumption, I can solve for the optimal wage offers as:[26]

$$W_{jc}(a) = \chi \cdot A_c L_c^{\mu} r_c^{\frac{\alpha}{\alpha-1}} \cdot z_j \cdot a^{\theta_j} \tag{3.8}$$

[26] This assumption abstracts from the strategic interaction of firms' wage-setting decisions, which is studied by Berger et al. (2022a). It is challenging to incorporate this strategic interaction in my context as it generates variable markdown and complicates the location decisions of both firms and workers.

where I define a constant $\chi \equiv \frac{\beta_w/\rho_w}{1+\beta_w/\rho_w} \cdot (1-\alpha)\alpha^{\frac{\alpha}{1-\alpha}}$, as the product of the wage markdown, $\frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}$, and a function of firms' housing expenditure share, $(1-\alpha)\alpha^{\frac{1}{1-\alpha}}$. Wage markdown, which equals the workers' take-home share for their marginal revenue product, increases with the firm-level labor supply elasticity β_w/ρ_w. A more elastic labor supply means that firms are more substitutable from the workers' perspective, which decreases firms' monopsony power. Setting the optimal wages at a constant share of marginal revenue product of labor is a standard result in the class of "new monopsony models", as summarized by Manning (2021). The local housing rent r_c shows up in the optimal wage offer as it affects the optimal housing input, which I solve for in Appendix B.1, and thus affects the marginal revenue product of labor. With wage setting and input choices, I then derive firm j's optimal profits in city c as:

$$\pi_c(j) = \Psi \cdot \left(A_c L_c^{\mu} r_c^{\frac{a}{\alpha-1}}\right) \cdot \left(z_j\right)^{1+\beta_w/\rho_w} \cdot \phi_{jc} \tag{3.9}$$

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

$$\phi_{jc} = \int_a \left(a^{\theta_j}\right)^{1+\beta_w/\rho_w} L_c(a) \cdot \frac{G_{jc}(a)^{\frac{\beta_w}{\rho_w}}}{\sum_{j' \in \mathcal{J}_c} \left(z_{j'} a^{\theta_{j'}} \cdot G_{j'c}(a)\right)^{\frac{\beta_w}{\rho_w}}} da. \tag{3.10}$$

Firm j's optimal profits are affected by city c's adjusted productivity $A_c L_c^{\mu} r_c^{\frac{\alpha}{\alpha-1}}$, firm j's own productivity z_j, and a labor composite term ϕ_{jc}. This last term relates to the total efficiency units of labor that firm j can employ in city c by synthesizing local labor supply and demand conditions. Note that city c's productivity does not enter the formulation of ϕ_{jc} as it does not affect within-city worker-firm matching. All else equal, a firm would earn more profits in a city where it can hire more workers, and where it complements the workers that it can hire. Importantly, variations in firms' skill-augmenting productivity θ_j and amenities $\{G_j(a)\}_a$ affect a city's heterogeneous returns for different firms, giving rise to firm spatial sorting.

3.3.3 Firm spatial sorting

There is a fixed set of firms, denoted by $\mathcal{J}$, in the economy.[27] Each firm j is owned by an entrepreneur, who spends profits only on the final good. The entrepreneur's preference is specified as:

$$u_c(j) = \ln \pi_c(j) + {\beta_f}^{-1} \nu_{jc}. \tag{3.11}$$

Symmetric to the worker's problem, each entrepreneur draws city-specific idiosyncratic preference tastes ν_{jc} from the Type-I Extreme Value distribution, whose dispersion is governed by β_f. With the optimal profits in each city given by equation (3.9), the probability of an entrepreneur j choosing city c can then be

[27]In Appendix F.2, I also specify a model extension to allow for free entry of firms at the national level, in which each entrepreneur pays a fixed cost c_e to draw their amenity and production characteristics. However, sorting shares of heterogeneous workers and firms are scale-invariant to the total measure of firms, and so are equilibrium outcomes such as housing rents, the spatial wage structure, and total output. Changes in the number of firms only affect welfare through workers' love-of-variety preference.

obtained as:[28]

$$p_c(j) = \frac{\left(\left(A_c L_c^{\mu} \bar{r}_c^{\frac{\alpha}{\alpha-1}}\right) \cdot \phi_{jc}\right)^{\beta_f}}{\sum_{c'} \left(\left(A_{c'} L_{c'}^{\mu} \bar{r}_{c'}^{\frac{\alpha}{\alpha-1}}\right) \cdot \phi_{jc'}\right)^{\beta_f}}. \tag{3.12}$$

Note that a firm's own common productivity z_j does not affect its location choice since it is carried with the firm regardless of its location. Equation (3.12) thus enables me to separately identify city productivity from firm productivity when I take the model to data. The intuition is that revealed firm location choices in the data are informative of the city fundamentals. More productive cities are able to attract larger shares of firms, holding all other profit determinants constant.

3.4 Housing supply

Housing in city c is supplied elastically by a housing developer, who combines the final good Y and land $\overline{H}_c$ to produce floor spaces:

$$H_c = \overline{H}_c \cdot Y^{\frac{1}{1+\gamma_c}} \tag{3.13}$$

where γ_c governs the return to scale of the housing production function. Floor spaces are rented to workers and firms at the rental rate r_c. The housing 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:

$$H_c^S(r_c) = \overline{H}_c^0 \cdot r_c^{\frac{1}{\gamma_c}} \tag{3.14}$$

where $\overline{H}_c^0$ 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 3.1.

3.5 Spatial equilibrium

I am now in a position to formally define the spatial equilibrium.

Definition 1. *Given city characteristics $\{A_c, R_c(a), \overline{H}_c, \gamma_c\}_{\forall c,a}$, worker measures $\{L(a)\}_{\forall a}$, the set of firm $\mathcal{J}$ with their productivites $\{z_j, \theta_j\}_{\forall j}$ 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 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:*

1. *Workers choose cities and firms to maximize their utility by equations (3.3) and (3.4);*
2. *Firms choose cities optimally by equation (3.12), and choose wage offers and housing inputs by equations (3.8) and (B.6);*
3. *The housing market clears in each city such that housing demand equals supply, which are given by equations (B.14) and (B.13);*
4. *The final good market clears by equation (B.17);*

[28] Alternatively, one can assume that entrepreneurs have city-specific idiosyncratic productivity shocks. Such shocks would enter the earnings equation, yet it is challenging to identify them from firms' fundamental productivity. Furthermore, this alternative specification would also induce a negative selection margin in that more profitable markets will attract firms with lower idiosyncratic productivity shocks. See Suárez Serrato and Zidar (2023) for more discussions on this matter.

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

4 Equilibrium Properties

4.1 Efficiency

I study how two-sided sorting affects overall efficiency and the optimal policy design in this section. Being atomistic, workers do not internalize the impact of their location decisions on local productivity through agglomeration spillovers μ, and firms do not internalize the impact of their location decisions on labor market congestion through wage indices, $\mathbb{W}_c(a)$. These impacts affect other agents' optimal choices, leading to inefficient sorting in the decentralized economy. To characterize the inefficiencies, I solve the social planner's problem and contrast the planner's solution with the laissez-faire allocation. I then propose an optimal spatial policy that implements the efficient allocation.

I start by defining the social welfare function that the planner aims to maximize. For ease of exposition, suppose there are N^w types of workers, indexed by a, and N^f types of firms (entrepreneurs), indexed by k, though this can be easily generalized to infinite types.[29] Consider a planning problem with Pareto weights $\varphi^w(a)$ for each worker type and $\varphi^f(k)$ for each firm (entrepreneur) type. The social welfare function is given by

$$\mathcal{W} = \sum_{a=1}^{N^w} \varphi^w(a) L(a) U(a) + \sum_{k=1}^{N^f} \varphi^f(k) J(k) \Pi(k), \tag{4.1}$$

where $U(a)$ and $\Pi(k)$ are the expected utility of type-a workers and type-k entrepreneurs. The planner maximizes equation (4.1) 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 final good for consumption by workers and entrepreneurs, $\{c_{kc}(a)\}$ and $\{c_c(k)\}$, and for production for housing, $\{Y_c\}$, and housing across local workers and firms, $\{h_{kc}(a)\}$ and $\{h_c(k)\}$, subject to spatial mobility and local matching constraints of workers and firms, as well as resource constraints of the final good and housing.[30] This planning problem extends the ones in Fajgelbaum and Gaubert (2020) and Rossi-Hansberg et al. (2019) by incorporating decisions on the spatial allocation of heterogeneous firms and local labor market matching.

4.1.1 Sorting externalities

To characterize the sources of sorting inefficiency in the laissez-faire equilibrium, I first discuss the differences in social and private marginal values of workers and firms, in the following proposition.

Proposition 1. *The social marginal product of a type-a worker working for a type-k firm in city c is*

$$\tilde{W}_{kc}(a) = \frac{1 + \beta_w/\rho_w}{\beta_w/\rho_w} \cdot \Big[W_{kc}(a) + \underbrace{\mu \bar{W}_c}_{\substack{\text{agglomeration} \\ \text{spillovers}}} \Big] \tag{4.2}$$

where $W_{kc}(a)$ is the competitive equilibrium wage following (3.8) *and $\bar{W}_c$ is the average wage of city c. The social*

[29] With a slight abuse of notation, a refers to both worker type and worker skill.

[30] I denote $D_{kc}(a)$ as the total measure of workers with type a employed by type k firms in city c. The number of skill a workers that each type k firm employs in city c is thus $D_{kc}(a)/J_c(k)$

marginal value of a type-k firm in city c is

$$\begin{aligned}\tilde{\pi}_c(k) = & \underbrace{\sum_a W^*_{kc}(a)\frac{D_{kc}(a)}{J_c(k)}}_{\text{firm surplus}} \underbrace{-(1-\rho_w)\sum_a \bar{W}^*_c(a)\frac{D_{kc}(a)}{J_c(k)}}_{\text{local labor stealing}} \\ & \underbrace{-\rho_w \sum_a \bar{W}^*(a)\frac{D_{kc}(a)}{J_c(k)}}_{\text{national labor stealing}} + \underbrace{\frac{\rho_w}{\beta_w}\sum_a \varphi^w(a)\frac{D_{kc}(a)}{J_c(k)}}_{\text{love-of-variety preference}}\end{aligned} \tag{4.3}$$

where $W^*_{kc}(a) = \tilde{W}_{kc}(a) - W_{kc}(a)$, and $\bar{W}^*_c(a)$ and $\bar{W}^*(a)$ are the skill-specific mean values at the city level and the national levels.

The differences in social and private values of workers and firms make clear the location choice externalities in the laissez-faire equilibrium. For workers, they do not internalize their agglomeration spillovers for other agents in the city, which is captured by $\mu\bar{W}_c$ in equation (4.2).[31] The magnitude of such spillovers is independent of the firm choice the worker makes within a city. One may suspect that workers do not internalize the impact on firm profits when making location choices. This does not present a source of externality in this economy due to the constant wage markdown - namely, workers and firms split fixed shares of the surplus of the match. The fixed rent sharing between firms and workers induces workers to indirectly internalize the impact of their workplace decisions on their employers' profits.

For firms, they only take into account the profits they can earn in a city by charging markdowns on the workers they match with, which appears as the firm surplus part in equation (4.3).[32] However, they do not internalize the following. First is the labor market stealing externality, which includes competing for workers with other firms in the same city, namely local labor stealing, and attracting workers from other cities, namely national labor stealing. The relative extent of local versus national labor stealing effects is governed by parameter ρ_w, which controls the correlation of idiosyncratic preferences between firms within a city. When $\rho = 0$, meaning that workers' idiosyncratic preferences are perfectly correlated within a city, there is no worker mobility across cities and thus no national labor stealing effect.

The reason for the labor stealing effects is that firms do not consider their location decisions' impacts on local wage indices, $\mathbb{W}_c(a)$, which ultimately affect the labor supply curves faced by every firm in the economy. This is because firms do not observe workers' idiosyncratic preferences, so they cannot set wages for each worker conditional on this preference idiosyncrasy. In other words, one firm competes with other firms for all workers through the wage index, even though some workers may particularly prefer to work for this firm. Instead, if firms could observe workers' preferences, they would optimally set wages to price this information. The information asymmetry gives rise to the fundamental market failure in the local labor market. On the one hand, an additional firm congests the local labor market, making it harder for other local firms to hire local workers. This congestion effect can be seen in equation (3.3). On the other hand, this additional firm expands local employment opportunities, which attract more workers to move

[31] Both Fajgelbaum and Gaubert (2020) and Rossi-Hansberg et al. (2019) emphasize heterogeneous productivity spillovers across different types of workers in proximity, which I abstract from. I experimented with heterogeneous agglomeration elasticities by worker skill when estimating the model, which aims to capture the heterogeneous spillovers in a reduced-form approach. However, I do not find significant differences after taking firm heterogeneity into account.

[32] One thing to note is that the surplus term also contains agglomeration spillovers.

into the city from elsewhere. This expansion effect can be seen in equation (3.4). Both effects are costly from the planner's perspective.

Second is the welfare improvement of local workers through their love-of-variety preference for more firms. As discussed before, this is because workers have idiosyncratic preferences for each local firm, thus viewing them as imperfectly substitutable workplace options. Firms do not internalize this welfare effect. Equation (4.3) is the local labor market analog of the inefficient firm entry result in Mankiw and Whinston (1986), who show that firm entrants do not consider business stealing from other firms and consumers' love-of-varietypreference when the product market features imperfect competition.[33] To my knowledge, this is the first paper that formalizes such firm sorting externalities from the labor market perspective. In this context, the inefficiency is a result of labor market segmentation and imperfect labor market competition. Firm sorting in the laissez-faire equilibrium is efficient if any of these two features is absent.

It is instructional to discuss what role the heterogeneity of cities, firms, and workers each plays in shaping firm sorting externality. I provide the discussion in the following lemma.

Lemma 1. *The following hold in the spatial economy:*

1. *When there is no city, firm, or worker heterogeneity, the social marginal value of a firm in city c is*

$$\tilde{\pi}_c = \frac{D_c}{J_c} \cdot \frac{\rho_w}{\beta_w} \varphi^w. \tag{4.4}$$

2. *When there is neither firm nor worker heterogeneity, but there is city heterogeneity, the social marginal value of a firm in city c is*

$$\tilde{\pi}_c = \frac{D_c}{J_c} \left(\rho_w \left(\bar{W}_c^* - \bar{W}^* \right) + \frac{\rho_w}{\beta_w} \varphi^w \right). \tag{4.5}$$

3. *When there is no worker heterogeneity, but there is firm and city heterogeneity, the social marginal value of a type-k firm in city c is*

$$\tilde{\pi}_c\left(k\right) = \frac{D_{kc}}{J_c\left(k\right)} \left(\left(W_{kc}^* - \bar{W}_c^*\right) + \rho_w \left(\bar{W}_c^* - \bar{W}^* \right) + \frac{\rho_w}{\beta_w} \varphi^w \right). \tag{4.6}$$

Starting from equation (4.4), when there is no heterogeneity, the social value of firms is only the love-of-variety preference part. This is because, without firm and city productivity heterogeneity, there is no difference in the marginal product of labor for a worker working in any firm in a city, i.e. $\bar{W}_c^* = \bar{W}^*$, so the planner is not concerned about the relocation of workers caused by firms' location choices. The optimal firm allocation is the one that equalizes firm sizes everywhere, which aligns with the laissez-faire equilibrium. The national labor stealing effect shows up when there is city productivity heterogeneity, as

[33] The business stealing effect is not present in this model. This is because all firms produce the same homogeneous goods and are price takers in the product market. Therefore, changes in output caused by firm location decisions have no net social value. This is consistent with Corollary 1 in Mankiw and Whinston (1986). Moreover, equation (4.3) can be re-written as

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

to explicitly include wage markdown in the expression. The residual terms account for the changes in agglomeration spillovers by re-locating workers.

shown in equation (4.5). In this case, the planner would encourage more firms to locate in more productive cities, which attract more workers there and increase total output. This result is consistent with Gaubert (2018), who argues there is insufficient firm sorting into productive larger cities.

There is an additional local labor stealing effect when firm heterogeneity is introduced, as shown in equation (4.6). The planner now internalizes the fact that heterogeneous firms' location choice also affects local matching and that the marginal product of labor differs across firms. In this case, the planner does not want firms to locate in cities where they can offer below-city-average wages. In other words, the planner would discourage low-productivity firms from co-locating with high-productivity firms. This externality is similar to the labor market pooling externality of firm sorting studied by Bilal (2023), who argues that low-productivity firms overvalue co-locating with high-productivity firms for favorable hiring conditions. As discussed in Bilal (2023), this externality is a spatial analog of Acemoglu (2001) in a frictional labor market with random search, where wages do not price vacancy filling rates. In my framework, the externality arises because firms' wage posting does not price its effect on labor market congestion. Finally, back to equation (4.3), the introduction of worker heterogeneity and worker-firm complementarity exacerbates the labor stealing externalities, as it is even more costly when a high-skilled worker matches with a low-productivity firm.

I will show in Section 6.2 that the utilitarian planner would further strengthen the co-location of higher-skilled workers and high-productivity firms compared to the laissez-faire allocation. The planner achieves so by reallocating better workers and firms to high-productivity cities and reallocating worse ones to low-productivity cities. Without worker-firm complementarity, the co-location motive for the planner is much weaker.

4.1.2 Efficiency conditions and the optimal policy

With the distinction between social and private values, I now move on to characterizing the efficiency conditions. As in Fajgelbaum and Gaubert (2020), I obtain conditions over the expenditure distribution, $\{x_{kc}(a)\}$ and $\{x_c(k)\}$, that must hold in any efficient allocation, where I define $x_{kc}(a) \equiv c_{kc}(a) + r_c h_{kc}(a)$ and $x_c(k) \equiv c_c(k)$. In Appendix C, I compare the competitive equilibrium with the expenditure distribution $\{x_{kc}(a)\}$ and $\{x_c(k)\}$ to the outcomes of the social planner's problem, which leads to the following proposition.

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

$$\underbrace{x_{kc}(a)}_{\substack{\text{private}\\ \text{consumption cost}}} = \frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}\left[\underbrace{\tilde{W}_{kc}(a)}_{\substack{\text{social marginal}\\ \text{product of labor}}} - \underbrace{O_c^w(a)}_{\text{opportunity cost}}\right] + \frac{1}{1+\beta_w/\rho_w}\underbrace{\varphi^w(a)}_{\text{Pareto weight}} \tag{4.7}$$

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

$$\underbrace{x_c(k)}_{\substack{\text{private}\\ \text{consumption cost}}} = \frac{\beta_f}{1+\beta_f}\left[\underbrace{\tilde{\pi}_c(k)}_{\substack{\text{social marginal}\\ \text{value of firm}}} - \underbrace{O^f(k)}_{\text{opportunity cost}}\right] + \frac{1}{1+\beta_f}\underbrace{\varphi^f(k)}_{\text{Pareto weight}} \tag{4.8}$$

for $J_c(k) > 0\ \forall k, c$. The opportunity cost terms, $\{O_c^w(a)\}$ and $\{O_c^f(k)\}$, are related to the multipliers on the resource *constraint of each worker and firm type in the planner's allocation. With given Pareto weights $\{\varphi^w(a)\}$ and $\{\varphi^f(k)\}$, if the planner's problem is globally concave and (4.7) and (4.8) hold, then the associated competitive equilibrium is efficient.*

The two conditions in the proposition above define the relationship between private consumption and the spatial allocation of workers and entrepreneurs that must hold in any efficient allocation. Private consumption in the efficiency conditions for workers and firms differs from their consumption in the laissez-faire equilibrium in several ways. First, the efficient consumption allocations in (4.7) and (4.8) depend on the social marginal value rather than the private counterparts to correct sorting externalities. Second, efficient consumption allocations take into account the opportunity costs of allocating workers and firms elsewhere, which are related to the Lagrange multipliers of the spatial allocation and local matching constraints. Workers' opportunity costs, O_c^w, vary by the city that workers are located in, whereas firms' opportunity costs, O^f, do not vary spatially. The former variation is driven by correlated idiosyncratic preferences between firms within a city and different matching opportunities between cities.[34] Third, the efficient consumption allocation increases less than one-for-one with the social marginal value for both workers and firms. This is optimal for the planner due to the distributional concerns. Within one agent type whom the planner values equally, the planner has incentives to redistribute towards those earning lower income, who have higher marginal utility of consumption. Lastly, the efficient consumption allocation also admits the Pareto weight, so higher-weighted agents will consume more.

Equipped with the efficiency conditions, I now design a set of tax instruments to implement the efficient spatial allocation, which is stated in the following proposition.

Proposition 3. *The efficient allocation can be implemented by a set of proportional taxes, including worker income taxes, $\{t_{kc}^w(a)\}_{\forall a,k,c}$, and firm profit taxes, $\{t_c^f(k)\}_{\forall k,c}$, specified as:*

$$t_{kc}^{w}(a) = -\frac{\frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}\tilde{W}_{kc}(a) + T_c^w(a) - W_{kc}(a)}{W_{kc}(a)}, \quad \forall a,k,c \tag{4.9}$$

$$t_c^f(k) = -\frac{\frac{\beta_f}{1+\beta_f}\tilde{\pi}_c(k) + T^f(k) - \pi_c(k)}{\pi_c(k)}, \quad \forall k,c \tag{4.10}$$

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

The optimal taxes take care of both inefficiencies due to sorting externalities and distributional concerns. I assume that firms take worker income taxes as given when choosing wages and that the land rents are rebated to workers proportionally to the after-tax worker income, $(1 - t_{kc}^w(a)) \cdot W_{kc}(a)$. As a result, the optimal wage offer does not change with the optimal policy, yet a worker's total income changes.[35]

[34] See Appendix C for the expressions of the opportunity cost terms. One can perceive the opportunity cost as the price of congestion for allocating a marginal worker or firm. The planner understands that workers and entrepreneurs have idiosyncratic preferences. Therefore, allocating more workers or firms to a city is associated with worse idiosyncratic preferences for the marginal worker or firm, which is viewed as a cost from the planner's perspective.

[35] To put in context, Fajgelbaum and Gaubert (2020) use optimal transfers to achieve the efficient allocation. I instead recast the transfers as proportional taxes taken as given by firms. The reason is that the transfer will affect the labor supply curves in a way that changes firms' optimal wage setting when they have the monopsony power.

4.2 Equilibrium uniqueness

One may be concerned that the model with two-sided sorting and local agglomeration may generate multiple equilibria. Although formal proof is still a work in progress, I provide a heuristic discussion here on how equilibrium uniqueness is affected by different model parameters. As common in spatial equilibrium models, multiple equilibria arise if agglomeration forces are stronger than dispersion ones (e.g. Bayer and Timmins (2005) and Allen and Arkolakis (2014)). Various dispersion forces in the model, including heterogeneous location fundamentals, (A_c, R_c), housing market congestions (governed by γ_c, η, α), and location-specific idiosyncratic preferences (governed by β_w and β_f), alleviate such concern. In addition, as will be shown later, the agglomeration elasticity is empirically estimated to be very small, limiting the incentive for agents to concentrate geographically.

However, there are two other reasons for multiplicity that are unique to this two-sided sorting framework. First, given production complementarity (governed by the covariance of θ and z) and local matching, high-skilled workers tend to follow where high-productivity firms locate, and vice versa for high-productivity firms. Second, workers have love-of-variety preference for more local firms (governed by β_w/ρ_w). Hence, all workers tend to locate in cities with more firms, which expands labor supply and then attracts more firms. Both mechanisms can lead to "mutual chasing" between firms and workers and thus multiple equilibria.

I conduct a numerical simulation exercise in Appendix D showing that multiple equilibria only occur when β_w/ρ_w is small and θ for high-z firms is large. The intuition is as follows. The β_w/ρ_w parameter governs both the love-of-variety preference and the extent of worker-firm assortative matching. Therefore, a large value of β_w/ρ_w implies that all workers benefit less from more local firms and high-skilled workers are more likely to be matched with the most productive firm, regardless of the number of such firms in the city. Both aspects diminish the tendency of workers to follow firms. Parameter θ for high-z firms governs the complementarity of high workers and productive firms, of which a smaller value lowers the return of matching and of co-location with each other. The estimation results of a large β_w/ρ_w estimate around 6 and moderate θ for high-z firms imply a low likelihood for multiple equilibria. I also experiment with solving the estimated model with different initial guesses - they all converge to the same equilibrium.

5 Empirical Implementation

5.1 Overview

The structural parameters to be estimated can be grouped into four blocks. First is the sorting elasticity parameters block, including labor supply elasticity parameters $\{\beta_w, \rho_w\}$ and the firm sorting elasticity parameter β_f. Second is the productivity parameters block, including worker skill a, firm productivity parameters $\{z, \theta\}$, city exogenous productivity A, and the agglomeration elasticity μ. Third is the amenity parameters block, including skill-specific firm and city amenity parameters $\{R_c(a), G_j(a)\}$. Last is the housing supply parameters block, including the housing supply elasticity γ_c, the amount of land $\bar{H}_c^0$, and housing expenditure shares for workers and firms, $\{\eta, \alpha\}$.

The procedure for estimating the parameters is summarized as follows. The sorting elasticity parameters block is estimated using the passthrough design following Lamadon et al. (2022). For the productivity parameters, I use between-firm worker movers to identify worker skills from firm and city productivity, and firms' revealed location choices to identify the city productivity. I recover amenities by matching worker sorting shares across firms and cities, controlling for (real) wage differences. I use the exogenous variation of city population from immigration to identify the agglomeration elasticity. Finally, I estimate the housing supply elasticity by tracing rent changes due to exogenous housing demand shocks and estimate city-specific land supply by matching the levels of city rents. I calibrate the worker share of expenditure on housing $\eta = 0.24$ following Davis and Ortalo-Magné (2011),[36] and the firm share on housing $\alpha = 0.06$ using information on firms' housing and wage bill expenditures.[37]

For the rest of the section, I first introduce additional assumptions that are useful for identification. These assumptions include discretizing firms into clusters and adding time-varying productivity and preference shocks onto the static model in Section 3. I then discuss the identification strategy of each parameter block. I conclude this section by presenting the estimation results and discussing the model fit.

5.2 Additional assumptions

5.2.1 Discretization

I follow Bonhomme et al. (2019a) and restrict firm production technology (z_j, θ_j) and skill-specific amenities $\mathbf{G}_j = \{G_j(a)\}_a$ 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 $k \in \{1, 2, 3, ..., K\}$. This classification of firms into clusters helps mitigate limited mobility bias, which is prevalent in the AKM estimation using movers (Andrews et al., 2008; Bonhomme et al., 2023).

5.2.2 Stochastic processes

I introduce time-varying shocks to the static model to utilize the panel feature of the matched employer-employee dataset for identification. I restate key model variables with these shocks in Appendix E.1. I introduce first workers' idiosyncratic preference shocks and skill shocks, and then firm and city productivity shocks, and lastly measurement errors on firm wage bills. I also impose an orthogonality condition on the time-varying skill and productivity shocks, which is useful for identification.

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

$$F(\overrightarrow{\epsilon_{it}}) = exp\left[-\sum_{c}\left(\sum_{j\in\mathcal{J}_{ct}} exp\left(-\frac{\epsilon_{ijt}}{\rho_w}\right)\right)^{\rho_w}\right]. \tag{5.1}$$

This assumption follows recent structural labor literature on worker-firm sorting (Card et al., 2018; Lamadon et al., 2022). Changes in workers' preferences can lead to their movements across firms and cities. Note that although equation (5.1) adds the time dimension to equation (3.2), it does not restrict the time-series properties of ϵ_{ijt}.

[36] I also experiment with heterogeneous housing expenditure shares when conducting counterfactual analysis. In these cases, I follow Eeckhout et al. (2014) to calibrate $\eta = 0.35$ for low-skilled workers and $\eta = 0.22$ for high-skilled workers.

[37] Specifically, I calculate the value of annual housing services as $total\_housing\_services = value\_of\_buildings \times (depreciation\_rate + property\_tax\_rate + mortgage\_rate - capital\_gain\_rate) + real\_estate\_rental$. Then, I calculate α =*total_housing_services/(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.

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 firms' skill-augmenting productivities, that is*

$$\theta_j \log a_{it} = \theta_j \log a_i + \log \hat{a}_{it} \tag{5.2}$$

and workers' preference for amenities only depends on the permanent worker skill a_i, that is

$$G_j(a_{it}) = G_j(a_i), \; R_c(a_{it}) = R_c(a_i). \tag{5.3}$$

The introduction of transient worker skill shocks follows Lamadon et al. (2022) and is important for identifying the effect of worker-firm interaction on earnings. Importantly, the transient skill shock does not interact with the skill-augmenting productivity term or affect workers' preference for non-wage amenities. These restrictions imply the transient shocks generate earnings changes but do not affect workers' sorting decisions to cities and firms.

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

$$\log z_{jt} = \log z_j + \log \hat{z}_{jt} \tag{5.4}$$

$$\log A_{ct} = \log A_c + \log \hat{A}_{ct} \tag{5.5}$$

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.

I allow for time-varying productivity shocks at both firm and city levels. Such dynamic processes of firm productivity follow the well-known works on production function estimation in the industrial organization literature (e.g. Olley and Pakes (1996) and Doraszelski and Jaumandreu (2018)). These time-varying productivity shocks generate wage bill shocks that are important for the passthrough design to be discussed in Section 5.3. It is necessary to have both firm and city productivity shocks in order to identify labor supply elasticities at both levels. Furthermore, I assume that the time-varying worker skill shocks and firm and city productivity shocks are independent, as in Lamadon et al. (2022).

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 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}:*

$$\log E_{jt} = \log \dot{E}_{jt} + e_{jt} \tag{5.6}$$

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.

5.3 Estimation Strategy

Step 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 Lamadon et al. (2022). This empirical design exploits wage bill shocks at firm and city levels and identifies the elasticity parameters using earnings changes of the staying workers.

I define $\bar{\dot{E}}_{ct}$ 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. As I show in Appendix E.2, changes in the mean variables isolate city-level shocks, and changes in residualized variables isolate firm-level shocks. I define worker passthrough parameters as $\delta_w \equiv \frac{1}{1+\beta_w/\rho_w}$ and $\delta_c \equiv \frac{1}{1+\beta_w}$. Parameter δ_w is identified from the passthrough of residualized firm wage bill shocks to changes in residualized log earnings for the stayers:

$$\Delta \log \hat{W}_{ijt} = \delta_w \Delta \log \hat{\dot{E}}_{jt} + \Delta \hat{a}_{it} + \delta_w \left(\Delta e_{jt} - \Delta \log \frac{\hat{\phi}_{jct}}{\hat{\phi}_{ct}} \right). \tag{5.7}$$

The extent of passthrough is controlled by the firm-level labor supply elasticity β_w/ρ_w. A larger labor supply elasticity implies that firms can adjust employment by smaller changes in wage offers, thus less rent sharing for stayers from wage bill shocks. Moreover, the passthrough equation (5.7) requires using the stayers sample to estimate, because the movers' earnings changes are caused by both rent sharing and movement across firms and/or cities.[38]

There are three residual terms in the net passthrough equation (5.7), which are the $i.i.d.$ skill transient shock Δa_{it}, 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 is caused by changes in the set of firms in the city c, which affects the labor supply curves through the wage index, $\mathbb{W}_c(a)$. The potential correlation of the 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 instrument the net wage bill shock $\log \Delta \hat{\dot{E}}_{jt}$ with its lags before year $t-q-1$, as in Lamadon et al. (2022). These lag 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 use the control function approach following Doraszelski and Jaumandreu (2018). Assuming that $\Delta \log(\hat{\phi}_{jct}/\hat{\phi}_{ct})$ follows a first-order Markov process, I can use control for it using one-period lagged variables, namely $\Delta \log \hat{W}_{ijt-1}$ and $\Delta \log \hat{\dot{E}}_{jt-1}$. In practice, I control for a cubic polynomial of these lags. See Appendix E.2 for more details.

Analogously, parameter δ_c is estimated from the passthrough of city mean real wage bill shocks to

[38] Recall from Section 2.1 that the stayers sample include workers who are associated with the same firm for at least 7 years and firms with at least 10 worker stayers.

changes in mean real log earnings, also for the stayers:

$$\Delta \log \left(\frac{\bar{W}_{ct}}{r_{ct}^{\eta}} \right) = \delta_c \Delta \log \left(\frac{\bar{\bar{E}}_{ct}}{r_{ct}^{\eta}} \right) + \delta_c \left(\Delta e_{ct} - \Delta \log \hat{\phi}_{ct} \right) \tag{5.8}$$

where e_{ct} is introduced as the measurement error of city-level mean firm wage bills, which is assumed to follow a $MA(q)$ process as e_{jt}. It is worth pointing out that workers' location choices are affected by real wage, rather than nominal, changes. City-level productivity shocks are partly capitalized into housing rents, which dampens population responses. Hence, estimating equation (5.8) using nominal terms, the empirical specification of Lamadon et al. (2022), will upward bias the estimate of the city-level labor supply elasticity β_w.

I apply the same instrumental variable strategy as the firm-level passthrough equation (5.7) to address endogeneity concerns that come from Δe_{ct} and $\log \hat{\phi}_{ct}$. One may be additionally concerned that the city-level *real* wage bill shocks are associated with amenity changes through local skill share changes (e.g. Diamond (2016)). In one specification that I run, I non-parametrically control for changes in shares of high-skilled workers to proxy potential amenity changes.

Firm sorting elasticity parameter

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

$$\Delta \log \bar{E}_{ct} = \delta_f \Delta \log \dot{E}_{ct} + \delta_f \left(\Delta e_{ct} - \Delta \log \frac{\hat{\phi}_{ct}}{\Phi_{ct}} \right) \tag{5.9}$$

where I define the firm passthrough parameter $\delta_f = \frac{1}{1+\beta_f}$. This passthrough equation for firms is estimated using a sample of firms that stay in the same city for at least 7 years and employ at least 10 workers each year. As with the passthrough regressions for workers, I also use lags of the total wage bill shocks before year $t-q-1$ as instruments to account for potential contemporaneous correlation of the total wage bill shock and the measurement error, and the control function approach to account for endogenous effects from changes in the set of local firms, captured by $\Delta \log(\hat{\phi}_{ct}/\Phi_{ct})$. See Appendix E.3 for more details.

This firm passthrough equation requires the number of firms to respond to local productivity shocks $\hat{A}_{ct}$ in each year. However, it is not necessary to have firm movers as we need for the worker passthrough equations. Suppose that there are random firm exits in all cities every year, which are replaced by new entrant firms with the same productivity distribution who then make location decisions. The underlying assumption for the firm passthrough equation is that there exists sufficient firm entry and exit such that the share of firms in each city satisfies equation (3.12) in all years.

Step 2: Productivity and amenity parameters

I now present the estimation strategy for the productivity and amenity parameters. At the heart of this strategy is a combination of the movers design of the AKM model (Abowd et al., 1999) and the idea of compensating differentials in the Rosen-Roback model (Rosen, 1974; Roback, 1982). The former identifies worker skills from the city and firm components in the earnings equation; the latter identifies the amenity and city productivity parameters that match the spatial distribution and matching patterns of heterogeneous workers and firms. I utilize the clustering approach by Bonhomme et al. (2019a) to mitigate limited mobility bias that is associated with the movers design.

The standard approach is to classify firms directly based on the empirical earnings distribution using a statistical clustering algorithm, e.g. k-means clustering. The inclusion of city productivity and the interest in identifying city and firm productivity separately complicates the clustering procedure. In this context, firm productivity and location fundamentals both contribute to workers' earnings. Hence, two firms that are located in different cities and with the same earnings schedule may differ in firm productivity. A correct classification that captures true firm heterogeneity thus requires controlling for the local factors in firms' earnings distributions. To make progress, I design an iterative procedure that relies on guessing and updating the city composite productivity, defined as $\mathbb{A}_c \equiv A_c \bar{L}_c^{\mu}$, where $\bar{L}_c$ is the average city population of the sample period. In what follows, I describe each step of this procedure.

Worker and firm fixed effects

In Appendix E.4, I construct a measure of adjusted log earnings, denoted by $\log \bar{W}_{ijt}$, by partialling out time-varying firm-level and city-level shocks and the city-level earnings determinants. It can be shown that the adjusted log earnings $\log \bar{W}_{ijt}$ is affected only by permanent worker skills a, firm productivity parameters (z, θ), and the time-varying skill shock $\hat{a}_{i,t}$:

$$\log \bar{W}_{ijt} = \log \chi + \underbrace{\log z_j}_{\text{firm productivity}} + \underbrace{\theta_j \log a_i}_{\text{worker-firm interaction}} + \hat{a}_{i,t}. \tag{5.10}$$

This equation follows the same empirical specification as in Bonhomme et al. (2019a). I classify firms into K clusters using the k-means clustering algorithm, taking firms' empirical distributions of the adjusted log earnings $\log \bar{W}_{ijt}$ as inputs. Having grouped the firms into clusters, the moment condition for identifying cluster-specific productivity parameters $\{z_k, \theta_k\}_{\forall k}$ is:

$$\mathbb{E}\left[\left(\frac{\log \bar{W}_{ij(t+1)}}{\theta_{k'}} - \frac{\log z_{k'}}{\theta_{k'}}\right) - \left(\frac{\log \bar{W}_{ijt}}{\theta_k} - \frac{\log z_k}{\theta_k}\right) \mid k \neq k'\right] = 0. \tag{5.11}$$

Equation (5.11) indicates that z_k and θ_k are identified with wage changes of between-cluster job movers, regardless of whether the move is within or between cities. The identification assumption of equation (5.11) is that the worker transient skill shock, $\hat{a}_{it}$, is orthogonal to the permanent skill a_i; and this shock does not interact with the skill-augmenting productivity θ. Equation (5.11) gives $K \times K$ moments to identify $2K$ parameters, two for each cluster. As discussed in Bonhomme et al. (2019a), identification of the skill-augmenting productivity θ exploits differences in earnings changes of job movers in opposite directions, i.e. mover from k to k' and from k' to k, provided that they differ in skills:

$$\mathbb{E}_{kk'}(a) \neq \mathbb{E}_{k'k}(a).$$

I show in Figure I.8 that this asymmetry is empirically supported by data.[39] With z and θ identified,

[39] Bonhomme et al. (2019b) clarify that documenting symmetric earnings gains and losses is not sufficient to reject the existence

worker permanent skill a_i can be retrieved from equation (5.11).

Firm and city amenities

The amenity terms, $G_k(a)$ and $R_c(a)$, are identified as compensating differentials: holding wages (real wages) fixed, firms (cities) with better amenities are able to attract more workers.[40] Hence, the amenities can be inferred from observed wages (real wages) and the composition of workers across firms (cities). Assuming that the amenities do not change over time, I show in Appendix E.5 that:

$$R_c(a) \cdot G_k(a) = \frac{\bar{r}_c^{\eta}}{\mathbb{A}_c r_c^{\frac{\alpha}{\alpha-1}} \cdot z_k a^{\theta_k}} \cdot \bar{J}_c(k)^{\frac{\beta_w}{\rho_w}} \cdot \Lambda_{kc}(a)^{\frac{\rho_w}{\beta_w}} \Lambda_c(a)^{\frac{1}{\beta_w}} \tag{5.12}$$

where $\Lambda_{kc}(a)$ is the average share of skill a workers in cluster k firms during the sample period, conditional on living in city c; $\Lambda_c(a)$ is the average share of skill a workers in city c; and $\bar{J}_c(k)$ is the average number of cluster k firms in city c. All RHS terms of the equation can be calculated from data. Equation (5.12) thus gives $K \times C$ moments to identify $K + C$ amenity parameters, up to one normalization for each skill a.

City composite productivity

The next step is to estimate the city composite productivity $\mathbb{A}_c$. To do so, I exploit the revealed firm location choices. Recall that entrepreneurs make location choices based on the optimal profits in all cities, which we know from equation (3.9) can be written as:

$$\bar{\pi}_c(k) = \Psi \cdot \left(\mathbb{A}_c \bar{r}_c^{\frac{\alpha}{\alpha-1}}\right) \cdot (z_k)^{1+\beta_w/\rho_w} \cdot \phi_{kc}. \tag{5.13}$$

I show in Appendix E.6 that with the estimates of labor supply parameters $\{\beta_w, \rho_w\}$, worker skill a, firm production parameters $\{z, \theta\}$ and amenity parameters $G_k(a)$, I can construct ϕ_{kc} for all city-cluster pairs. Having clustered all firms, I can construct the share of firms in each cluster k located in each city c, denoted as $p_c(k)$, from the data. It follows that city composite productivity $\mathbb{A}_c$ can be identified from the firm sorting equation below

$$p_c(k) = \frac{\left(\left(\mathbb{A}_c \bar{r}_c^{\frac{\alpha}{\alpha-1}}\right) \cdot \phi_{kc}\right)^{\beta_f}}{\sum_{c'} \left(\left(\mathbb{A}_{c'} \bar{r}_{c'}^{\frac{\alpha}{\alpha-1}}\right) \cdot \phi_{kc'}\right)^{\beta_f}}. \tag{5.14}$$

The intuition is that controlling for the labor composite, a city with higher rent-adjusted composite productivity should attract a larger share of firms. Therefore, holding all else constant, a city that attracts more firms is revealed to be more productive. Note that firm productivity z does not enter the equation above since a firm can carry its productivity to whichever city it chooses. Therefore, this equation gives $C \times K$ moments in order to identify C city composite productivity parameters, up to one normalization.

Discussions on the iterative procedure

It helps to discuss what economic forces in the model facilitate the convergence of the iterative procedure. The logic is the following. Consider that in one iteration of the procedure, the city composite productivity

of complementarity. See their Online Supplement Section S2 for more details.

[40] Recall that I assume firms in the same cluster share the same skill-specific amenity, namely $G_j(a) = G_{k(j)}(a)$.

of city c is guessed to be greater than its true value. With this guess, firm productivity z_j for local firms would be estimated to be smaller than their true values, which can be seen from equation (5.10). The labor market in city c would seem to be less competitive, resulting in larger $\hat{\phi}_{kc}$ in equation (5.13), and thus seem to be more profitable. However, the shares of firms in city c, $p_c(k)$, are fixed in data, despite small adjustments in firm clustering with the guess. To fit the firm sorting shares in equation (5.14), the city composite productivity will be updated to be smaller at the end of this iteration. In sum, the iterative procedure achieves convergence for a system of cities via proper allocation of earnings to city and firm components, which is disciplined by revealed firms' location choices in the data.

Note that separate identification of city and firm productivities does not depend on the discretization assumption. In the absence of limited mobility bias, I would not need to either group firms into clusters or apply the iterative procedure.

Due to prevalent measurement errors in the data, it is challenging to estimate each city's productivity precisely, and even more so for smaller cities, using this procedure, which renders convergence of the iterative algorithm difficult. To deal with this statistical issue, I smooth city composite productivity estimates against log city population using a third-order polynomial after each iteration, and iterate until the smoothed estimates converge. This does not affect the model's ability to account for the systematic contribution of the city fundamental to the urban earnings premium, but undermines its ability to conduct a structural variance decomposition exercise as smoothing city productivity against population mechanically decreases its variance.

Agglomeration elasticity and city productivity

I estimate the agglomeration elasticity μ by relating converged city composite productivity to city population. Specifically, I run the following regression:

$$\log \mathbb{A}_c = A_0 + \mu \log \bar{L}_c + \epsilon_c^A \tag{5.15}$$

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

Step 3: Housing supply parameters

Lastly, I show how I estimate the housing supply parameters, with more details in Appendix E.8. The (inverse) housing supply elasticity can be estimated by relating changes in the housing rent to changes in the housing expenditure:

$$\Delta \log r_{ct} = \Gamma_c \Delta \log EH_{ct} + \Delta e_{ct}^r \tag{5.16}$$

where $\Gamma_c \equiv \frac{\gamma_c}{1+\gamma_c}$, 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.[41] The slope coefficient Γ_c informs the housing supply elasticity: a high Γ_c means that the housing rent changes significantly to housing demand, which implies a low housing elasticity. Following

[41] I measure the average housing rent using the average monthly rent of two-bedroom units. See Appendix A for more details.

Saiz (2010), I allow the housing supply elasticity to be affected by the share of land unavailable for housing development in each city, denoted by $UNAVAL_c$, and parameterize Γ_c as a function of this share, i.e. $\Gamma_c = \Gamma + \Gamma_L UNAVAL_c$. The definition of undevelopable land follows Saiz (2010), that is, land with a slope higher than 15 degrees and land covered by water. I estimate the housing supply equation using five-year city-level changes between 2002 and 2007. I choose this period to evade the 2008-2009 housing collapse during the Great Recession and the subsequent surge in foreign investment in the Canadian real estate market, both of which stemmed from factors beyond local housing demand shocks.

An OLS estimate of the above equation may yield biased estimates of Γ_c as the changes in the housing expenditure are correlated with time-varying local characteristics which also affect changes in the housing rent. To deal with this endogeneity concern, I follow Diamond (2016) and instrument $\Delta \log EH_c$ using a shift-share Bartik instrument. Then, I choose $\overline{H}_c^0$ to fit the average housing rent of each city.

5.4 Estimation results

Sorting elasticity and housing supply parameters

I discuss here the estimation results of the sorting elasticity and housing supply elasticity parameters. The estimated parameter values using the preferred specifications and calibrated housing parameters are summarized in Table 3.

I report the passthrough estimates for workers in Table I.6. 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). Controlling for lagged changes in log earnings and wage bill as a proxy for changes in labor demand competition (i.e. $\Delta \log(\hat{\phi}_{jct}/\hat{\phi}_{ct})$), reported in Column (3), increases the estimate 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); Lamadon et al. (2022)).[42] 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.[43] Controlling for changes in labor competition, reported in Column (6), reduces the estimate to 0.25. This is because changes in log mean real wage bills are negatively correlated with the changes in labor market competition ($\Delta \log \hat{\phi}_{ct}$), which enters negatively into the structural error term. Further controlling for changes in shares of high-skilled workers, reported in Column (7), increases the estimate to 0.32. The reason is that unobserved amenity changes are positively correlated with real wage bill changes and are negatively correlated with real wage workers need to be paid to live in a city. The city-level labor supply elasticity implied by the estimated 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 (2018); Notowidigdo (2020); Bilal (2023)).

I report the passthrough estimate for firms in Table I.8. The IV estimate (0.21) in Column (2) is much smaller than the OLS estimate (1.06) in Column (1), which suggests changes in total city wage bills are strongly and positively correlated with their measurement errors. In Column (3), I control for changes

[42] Note that the sources of variation are different from mine in these two papers. Card et al. (2018) review estimates using value-added per worker and Lamadon et al. (2022) used changes in value-added, whereas I use changes in wage bills. I also estimated a restricted specification that imposes $\rho_w = 1$, which implies that workers' preference draws are uncorrelated within cities. The estimated passthrough elasticity is then 0.15, which is also consistent with the existing literature.

[43] I have also tried running city-level passthrough regressions using nominal changes in mean wages and mean wage bills. The OLS estimate (0.23) and the IV estimate (0.26) are both smaller than the ones using real changes, which implies that using nominal measures for the city-passthrough equation over-estimates the city-level labor supply elasticity.

Table 3: Summary of sorting elasticity and housing-related parameter estimates

Parameter	Description	Value	Method
β_w/ρ_w	Firm-level labor supply elasticity	6.1	Equation (5.7)
β_w	City-level labor supply elasticity	2.1	Equation (5.8)
β_f	Firm sorting elasticity	5.6	Equation (5.9)
α	Share of housing in production function	0.06	Value of housing services/VA
η	Share of expenditure on housing	0.24	Davis and Ortalo-Magné (2011)
γ_c	Inverse housing supply elasticity	0.81	Equation (5.16)

in labor competition, and the estimate further decreases the estimate to 0.15. The inferred firm sorting elasticity of 5.6 from Column (3) is within the range of the estimates Gaubert (2018) reports for firms in different tradable industries, while slightly higher than the corresponding estimates in Suárez Serrato and Zidar (2016) and Fajgelbaum et al. (2019).

The estimation results of the housing supply elasticity Γ_c are reported in Table I.11. The OLS estimate in Column (1) is 0.53 and the IV estimate in Column (2) slightly decreases to 0.45, but it is no longer statistically significant. This IV estimate of Γ_c translates to a housing supply elasticity $1/\gamma_c$ of 1.25, which is between the average elasticity of 0.5 reported in Baum-Snow and Han (2023) and 1.5 reported in Saiz (2010). I further interact changes in log housing expenditure with the share of undevelopable land in Column (3) but find a null effect from topography differences.

Productivity parameters

I report the estimates of the two firm productivity parameters by cluster in Table I.9. When estimating the earnings equation (5.10), I impose normalization that $\log z = 0$ and $\theta = 1$ for the lowest-paying firm cluster. The estimation result reveals considerable variation of common productivity z and skill-augmenting productivity θ across firm clusters. The two productivity parameters almost perfectly correlate with each other, with a correlation coefficient of 0.97, consistent with the empirical results in Lamadon et al. (2022). It indicates that more productive firms are also more efficient at harnessing worker skills, which confirms the existence of worker-firm production complementarity.

The estimation procedure generates a continuous distribution of worker skills. In practice, I rank workers by their continuous skill estimates, bin them into 100 equal-size worker groups, and calculate the average worker skill for each group. To better account for the spatial sorting of top-skilled workers, which is quantitatively important for the urban earnings premium, I further divide the top percentile group into equal-sized 21 worker groups, resulting in a total of 120 worker groups.

I plot the estimates of city productivity A on city population in Figure I.9. I find that larger cities have limited exogenous productivity advantages on average.[44] To discover the determinants of the

[44] The city productivity-population elasticity is estimated to be 0.001 (s.e. 0.014). One interesting finding is that the exogenous city productivity has a "U-shape" relationship with city population. Many productive cities are in the northern part of Canada, which have low natural amenities due to severe climate conditions.

local productive advantage, I correlate the city exogenous productivity estimates with a set of location characteristics and report the results in Table I.12. The result suggests that places with warmer summers in the east are more productive, and places in the north are more productive controlling for adverse climate conditions. As reported in Table I.10, the agglomeration elasticity μ is estimated to be very small (0.0004) and statistically insignificant, after controlling for worker and firm heterogeneity.[45]

Amenity parameters

Lastly, I plot the firm and city amenity estimates, $\{G_k(a), R_c(a)\}$, in Figure I.10. I bin both cities and workers into ten groups based on city population and worker skills. Panel (a) shows firm amenities, where I normalize the amenity of firm cluster 1 to be 1 for each worker group. It shows that the high-productivity firms have lower amenities, especially for high-skilled workers. It may be due to longer work hours, inflexible schedules, and high work stress that come with higher firm wage premia. Panel (b) shows city amenities, where I normalize the amenities to be 1 of the largest city group for each worker group. Larger cities have higher amenities for all workers and are more appreciated by higher-skilled workers, which is consistent with Diamond (2016). I also correlated city amenities with the location characteristics in Table I.12, which shows that workers appreciate places with warm winters, breezy summers, safe streets, and mountainous topography.

5.5 Model fit

I now check the model fit. The model cannot fit the data perfectly in two respects. First, it cannot match the worker sorting shares for each city-cluster bin, $\Gamma_{kc}(a)$. Second, it cannot match the firm sorting shares for each city, $p_c(k)$. To see this, note that the model is over-identified as I only use $N_c + N_k$ number of amenity parameters $\{R_c(a), G_k(a)\}$ to fit $N_c \times N_k$ sorting shares per worker skill a, and only N_c city productivity parameters A_c to fit $N_c \times N_k$ firm sorting shares. As worker and firm sorting interact closely in the model, failing to match sorting shares of either side could undermine the model's ability to explain the relationship between two-sided sorting and spatial inequality. Hence, it is important to check how the model solution matches key sorting and inequality measures in the data.

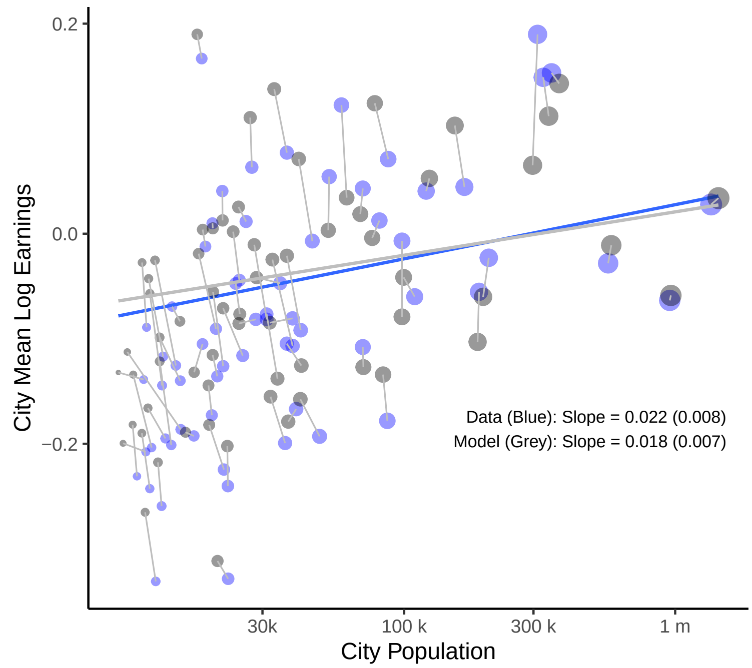
I first check the model fit of worker and firm sorting shares across cities in Figure I.11. To visualize this, I plot the data shares and the model-fitted sorting shares of workers in each skill decile and of firms in each cluster across five city groups.[46] The five city groups are the five largest cities (Top 5) group and four equal-number city groups ranked by city population. It can be seen that the model can replicate sorting shares in the data well. One caveat is that the model slightly over-predicts the share of the top cluster firms in the largest 5 cities. This is driven by within-cluster variation in firm productivity. In the data, the top-cluster firms in the largest cities are in fact larger and more productive than the firms belonging to the same cluster in other cities. In order to fit the cluster-city employment shares, the model has to overstate the share of top-cluster firms located in the largest cities.

I further examine the model fit of worker-firm match shares within cities in Figure I.12, where I plot the shares of workers in each skill deciles matched to each firm cluster in the largest five cities and smaller cities.

[45] The agglomeration elasticity should be interpreted as the static benefit of being in larger cities, controlling for city fundamental and firm heterogeneity. The small estimate does not rule out learning benefits in larger cities, which play out over the life cycle.

[46] I also regress the model-fitted sorting shares on the data sorting shares, $s_{kc}(a)$ and $s_c(k)$. The R^2 are 0.994 and 0.988 respectively for workers and firms.

Figure 2: Model fit of the urban earnings premium



Note: This figure compares the relationship of the city mean log earnings and city population generated by the model solution against the data. I plot the weighted linear regression fit of the two versions and report the slope estimates with standard errors. The gray dots and the line represent model fit and the blue ones represent the data. I connect the two dots of the same city to facilitate comparison.

The comparison shows that the model can fit the within-city match shares well. It slightly overstates the degree of assortative matching in the sense that it generates larger match shares of the most skilled workers to the most productive firms and the larger shares of the least skilled workers to the least productive firms compared to data.

Lastly, I check the model fit of the urban earnings premium – the key relationship of interest. In Figure 2, I plot the model version and the data version of the city mean log earnings versus city population. This comparison again confirms the model's ability to fit the data: it generates an urban earnings premium of 0.018, which is marginally smaller than the data estimate of 0.022. It can also be seen that the model fits large cities better than smaller cities.

5.6 Discussions

Understanding worker skill: sorting and learning

With a static model, I do not take a stand on the origin of worker skill, which could originate from different worker innate abilities, education levels, and human capital accumulation from work experiences. Baum-Snow and Pavan (2013) and De La Roca and Puga (2017) argue that the city-size wage premium is primarily driven by the faster human capital accumulation of workers in big cities, which is more beneficial for highly educated workers.[47] In fact, both papers find non-significant worker sorting within education groups according to initial worker abilities. The goal of this paper, however, is not to distinguish between

[47]The Mincer-type regression that I run to construct residualized log earnings can only control for average returns from experience, but not heterogeneous returns from different city sizes.

static sorting and dynamic learning. Instead, it aims to characterize worker and firm heterogeneity across cities and evaluate spatial policies while holding the skill and productivity distributions fixed.

As a validation, in Figure I.13 I plot the model-implied city high-skilled worker shares against college-educated worker shares. I define high-skilled workers as the ones with skill in the top three deciles of the skill distribution, which roughly matches the college-educated share of 28.5% in the 2016 Census. The figure reveals a significant correlation between the two shares, which validates the worker skill estimates using the fixed effect approach.

Endogenous labor market power

Recent studies suggest that large cities have more competitive labor markets, which results in smaller wage markdown and higher worker earnings (Hirsch et al., 2022; Bamford, 2021). For tractability purposes, I assume away endogenous labor market power in the model. Having this feature will complicate the location choices of workers and firms, as they are additionally affected by endogenous markdowns. Quantitatively, the contribution of average city markdowns on the urban premium is captured by city productivity term A_c - all else equal, less competitive labor markets should attract more firms. As I will soon show, the quantitative importance of local characteristics is small relative to worker and firm heterogeneity, implying a limited role of variable wage markdown. However, I cannot rule out the possibility that firms of different sizes enjoy different labor market power, which could potentially bias the firm productivity estimates. In addition, I estimate the firm-level passthrough parameter separately for the largest five cities and smaller cities. As shown in Table I.7, the difference in this passthrough parameter estimates by city size is not statistically significant. The table also includes passthrough estimates separately for high- and low-skilled workers, which are not significantly different either.

Multi-location firms

There has been a rise in interest in understanding multi-location firms' location choices (e.g. Oberfield et al. (2020); Jiang (2023); Kleinman (2022)), which is beyond the scope of this paper. One way to interpret multi-location firms in this model is that they are firms of multiple managers who differ in their individual abilities and have comparative advantages in different locations. Each manager chooses her own location to operate based on her characteristics. One advantage of this approach is that I can recover each establishment's productivity non-parametrically with no restrictions on the relationship of different establishments' productivity within the same firm, which allows me to decompose spatial wage inequality more flexibly. With the estimated model, I decompose the between-establishment productivity variance into a between-firm variance component and a within-firm variance component. The between-firm part accounts for 89.7% of the total variance, with the rest explained by the within-firm part. This decomposition result implies that the productivities of different production units within a firm differ but share common firm-level productivity components.

6 Quantitative Analysis

I implement the following quantitative exercises using the estimated model in this section. First, I perform a structural decomposition of the urban earnings premium and quantify the role of relevant model parameters in generating equilibrium sorting. Second, I implement the optimal spatial policy presented

Table 4: 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 displays the decomposition results of city-size regressions of city mean city log earnings, based on equation (6.1). The productivity parameters are estimated using the iterative estimation procedure described in Section 5.3. All regressions are weighted by population. *p<0.1; **p<0.05; ***p<0.01.

in Section 4. Third, I evaluate place-based subsidies that are often considered by policymakers. Lastly, I analyze how the rise of remote work, which relaxes the co-location constraint on firm-worker matches, can potentially affect the location choices of workers and firms and the spatial wage structure. I conduct more counterfactual analyses in Appendix H, which include loosening housing regulations in high-earning Canadian cities and implementing universal basic income (UBI).[48]

6.1 Decomposition of the urban earnings premium

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

$$\log \mathbb{E}_c(W_{ijt}) = \log \chi + \underbrace{\log\left(A_c \cdot L_{ct}^{\mu} \cdot r_{ct}^{\alpha/(\alpha-1)}\right)}_{\text{city characteristics}} + \underbrace{\bar{\theta} \cdot \mathbb{E}_c\left(\log a_i - \log \bar{a}\right)}_{\text{worker sorting}} + \underbrace{\mathbb{E}_c\left(\log z_j + \theta_j \log \bar{a}\right)}_{\text{firm sorting}}$$
$$+ \underbrace{\mathbb{E}_c\left[\left(\theta_{j(i,t)} - \bar{\theta}\right) \cdot \left(\log a_i - \log \bar{a}\right)\right]}_{\text{interaction effect}}. \tag{6.1}$$

The presence of skill-augmenting productivity gives rise to an interaction component in the decomposition equation, which quantifies the contribution of assortative matching through the skill-augmenting productivity to the city mean log earnings. Correspondingly, the worker sorting component is expressed as mean city worker skill (relative to the national average) multiplied by the average skill-augmenting productivity, and the firm part includes firm skill-augmenting productivity multiplied by the national average worker skill.

In order to quantify the contribution of each component to the urban earnings premium, I regress the city mean log earnings and each component of the RHS on log city population, with the results shown in

[48] In the counterfactual of loosening housing regulations, I show that the increase of total output is much larger if I ignore two-sided heterogeneity, in which case high-earnings cities are estimated to have high exogenous productivities. In the counterfactual of UBI, I show that workers will move to cheaper yet less productive cities, especially for low-skilled workers.

Table 5: 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	$cov(\bar{a}_c, \bar{z}_c)$
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:	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 6.1.1 for a detailed description of the method.

Table 4.[49] This result is the first of its kind to provide a full decomposition of the urban earnings premium. The table shows that worker and firm sorting explain 66.8% and 26.8% of the urban earnings premium, respectively.[50] The exogenous productive advantage and agglomeration benefits of larger cities are partly capitalized into higher housing rents; as a result, city characteristics altogether explain about 6.6%. The interaction component contributes none to the urban premium, which seems to suggest that worker-firm complementarity plays no role in spatial inequality.[51] However, I will show soon that such complementarity is crucial for the co-location phenomenon and the resulting spatial wage structure. I also present the results of a variance decomposition for within-city earnings inequality in Table I.13. It shows that greater earnings dispersions within larger cities are primarily associated with greater worker skill variances and worker-firm match covariances, consistent with the reduced-form results reported in Table I.5.

6.1.1 Understanding spatial sorting using a Shapley value approach

The statistical decomposition that I present in the previous section is performed using equilibrium sorting outcomes. It does not, however, inform the reason why heterogeneous workers and firms choose locations in the way shown in the data. To shed light on this, I now study how much different model parameters quantitatively contribute to spatial sorting and the urban earnings premium.

[49]The city-size regressions of three of the four components are statistically insignificant. This is partly caused by noisy estimates of the city productivity parameters and the small number of cities in the sample.

[50]In principle, both structural and reduced-form estimation of the earnings equation should have the same worker skill estimates, which are identified using worker movers in both cases. Therefore, the contribution of worker sorting to the urban earnings premium should be the same. However, the statistical discrepancies caused by a different clustering procedure in structural estimation (see section 5.3) generate a slightly smaller worker sorting share in the structural decomposition.

[51]Using the same empirical specification, Bonhomme et al. (2019a) and Lamadon et al. (2022) both find that the interaction effect explains little of the variance and mean of individual log earning. This result is surprising given the coexistence of strong positive assortative matching and a positive correlation of z and θ. I illustrate the reason behind the coexistence in Figure I.14. As shown in Panel (a), z governs the level of the earnings profile, and θ governs the curvature of the earnings profile. With a high labor supply elasticity of 6.1, the variation in θ can generate considerable positive assortative matching, However, by comparing Panels (a) and (b), one can see that the magnitude of the interaction effect is much smaller than the level effect.

The sorting and urban premium outcomes in the model are generated by variations in the following parameters. First are the parameters that enter through the city component in the earnings equation, including city productivity A_c, the agglomeration elasticity μ, and the share of housing input in production function α. These parameters affect the spatial distributions of both earnings and population. Second are the skill-specific amenity parameters, including city amenities $R_c(a)$ and firm amenities $G_k(a)$. The amenities affect utility for workers and labor supply for firms in each city. Third are the firm skill-augmenting parameters θ_k, which affect the incentive for heterogeneous workers and firms to match with each other and to co-locate in the same cities. Removing the variations in all these parameters would completely erase city-size earnings and sorting gradients, though there still exists a variation in city sizes caused by the difference in exogenous land $\overline{H}_c$. I do not choose to remove the variation in exogenous city land endowment in order to preserve a meaningful city-size distribution. Also, given my interest in quantifying how different parameters contribute to spatial outcomes through their impacts on the sorting of heterogeneous workers and firms, I do not include worker skill a and firm heterogeneity z in the set of parameters.

The aforementioned parameters all contribute to the equilibrium outcomes interdependently. Eliminating the variation in one set of parameters ω also eliminates the covariance of this parameter with other parameters. Therefore, the corresponding change in the equilibrium outcome that arises from eliminating ω cannot be attributed to ω alone. To quantify the contribution of each parameter to the equilibrium outcomes, I apply the Shapley value approach. Specifically, I simulate counterfactual economies by eliminating all combinations of the sources of variation listed above and then compute the Shapley value for each source of variation in terms of its effect on the inequality outcomes. Intuitively, this provides a measure of an average change in the equilibrium outcome when a source of variation is eliminated, under all combinations of the remaining sources of variation. I consider four equilibrium outcomes, including the urban earnings premium β^w, the worker and firm sorting gradients β^a and β^z, and the population-weighted covariance of city mean worker skill and mean firm productivity $cov(\bar{a}_c, \bar{z}_c)$.[52] More details on implementing the Shapley value approach can be found in Appendix G.

I report the results in Table 5. It shows that 17.0% of the urban earnings premium β^w is explained by city characteristics, which is close to the statistical decomposition result in Table 2. City exogenous productivity A_c and the agglomeration elasticity μ each contribute positively by 13.2% and 25.9%, whereas the housing share parameter α contributes negatively by -22.1%. The negative contribution of α means that β^w will increase by 22.1% if $\alpha = 0$. This is because firms in larger yet more expensive cities tend to rent less housing to operate, which decreases the marginal product of labor and thus earnings. The two sets of amenity parameters collectively explain 42.3% of the urban premium, where city amenity $R_c(a)$ explains 71.4% and firm amenity $G_k(a)$ explains -29.1%. As discussed above, larger cities offer relatively greater amenities for high-skilled workers, but more productive firms have relatively worse amenities for high-skilled workers. These heterogeneous amenities affect sorting and thus the urban premium. Lastly, heterogeneity in skill-augmenting productivity θ contributes another 40.6% of the urban earnings premium, by affecting the co-location incentive of high-skilled workers and high-productivity firms. Related to the structural decomposition result above, although the production complementarity does not contribute to the urban earnings premium via the interaction part, it matters crucially for how heterogeneous workers and firms sort across cities.

[52]More precisely, β^w, β^a and β^z are the regression coefficients of city mean log earnings, mean worker skill, and mean firm productivity on log city population.

For the spatial sorting outcomes, the worker skill gradient β^a is almost entirely explained by skill-biased amenities in larger cities. This is consistent with prior empirical findings that real wages do not increase city sizes more for high-skilled workers (e.g. Diamond (2016) and Card et al. (2023)), so worker sorting heavily depends on compensating differentials. The variation in θ explains little. In contrast, the variation in θ contributes significantly to the firm sorting gradient β^z, which is offset by the negative contribution of firm amenities $G_k(a)$. The city characteristics parameters do not affect spatial sorting *per se*, but they affect the sorting gradients through their impact on the city population distribution. It can be concluded that the extent to which firms' spatial sorting is affected by workers is much greater than the extent to which workers' spatial sorting is affected by firms. Lastly, the decomposition result of the co-location outcome $cov(\bar{a}_c, \bar{z}_c)$, of which 60.4% is explained by amenities and 44.6% is explained by production complementarity, again confirms the intuition discussed above.

6.2 Optimal spatial policy

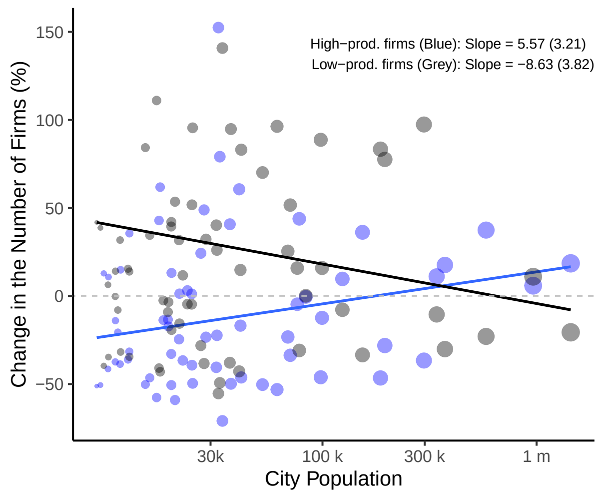
In this Section, I present the counterfactual result of implementing the optimal spatial policy. Recall from Proposition 3 that the optimal spatial policy comprises a set of tax instruments, $\{t^w_{kc}(a)\}_{\forall k,c,a}$ for workers and $\{t^f_c(k)\}_{\forall k,c}$ for firms. The aim is to maximize social welfare by 1) taxing or subsidizing agents for them to internalize the sorting externalities, 2) redistributing within types to reduce gaps in the marginal utility of consumption, and 3) redistributing between types depending on welfare weights. The sorting externalities include the labor market stealing and love-of-variety spillovers for firms and the agglomeration spillovers for workers, though the latter is of limited quantitative importance given a tiny estimate of the agglomeration elasticity μ. In what follows, I present the results of a utilitarian government that sets $\varphi^w(a) = 1, \forall a$ and $\varphi^f(k) = 1, \forall k$. I discuss the results using alternative welfare weights later. Moreover, this should be considered as a first-best long-run optimal policy as I assume that the government can fully observe worker skills and firm production technologies, and I abstract from short-run moving costs of workers and firms. 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 treatment of equilibrium uniqueness of optimal spatial policies.

Sorting efficiency and spatial reallocation. In Figure 3, I plot the changes in the number of firms and workers against city population of the efficient equilibrium relative to the laissez-faire equilibrium. To illustrate changes in the sorting pattern, I separately plot the changes for high- and low-productivity firms in Panel (a), and the changes for high- and low-skilled workers in Panel (c). Panels (b) and (d) show the changes in total number of firms and workers. I define workers with skill in the top three deciles as high-skilled workers and the remaining ones as low-skilled workers. Analogously, I define firms in the top three firm clusters, ranked by z, as high-productivity firms and the remaining ones as low-productivity firms.

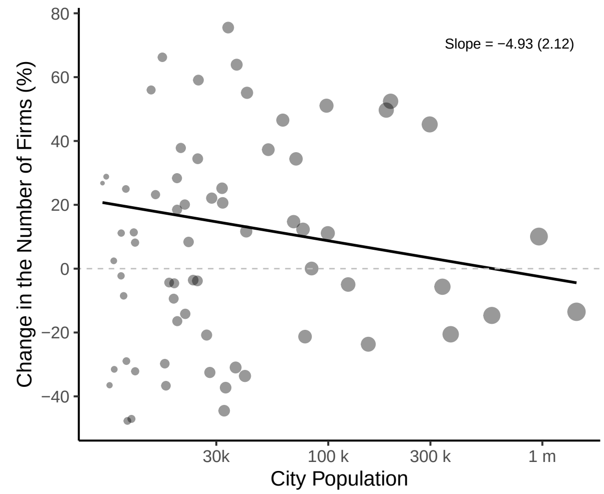
As shown in Panel (a) of Figure 3, the most salient outcome of the optimal policy is that it allocates more high-productivity firms into larger and more skilled cities, while moving low-productivity firms away from these places. With the optimal profit taxes, firms internalize their negative labor market stealing and positive love-of-variety externalities. For low-productivity firms, the negative labor market stealing externality is greater in large cities as they compete for high-skilled workers with high-productivity

Figure 3: Changes in the spatial allocation of workers and firms: optimal policy versus laissez-faire equilibrium

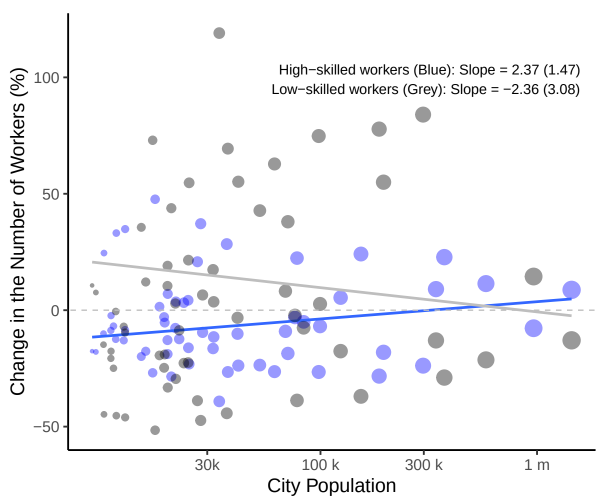
(a) Number of high-prod. and low-prod. firms



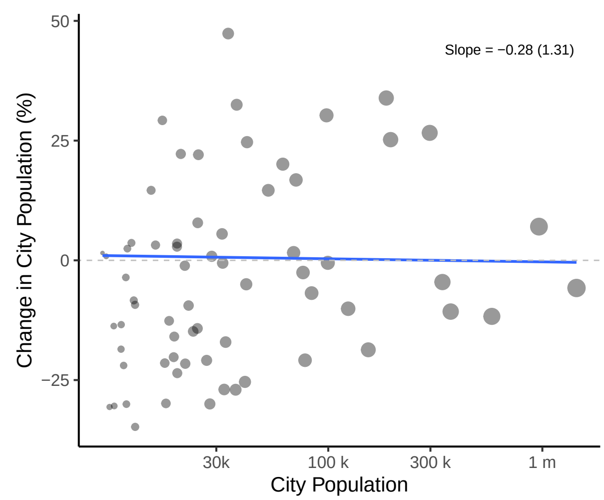
(b) Total number of firms



(c) High-skilled and low-skilled population



(d) Total population



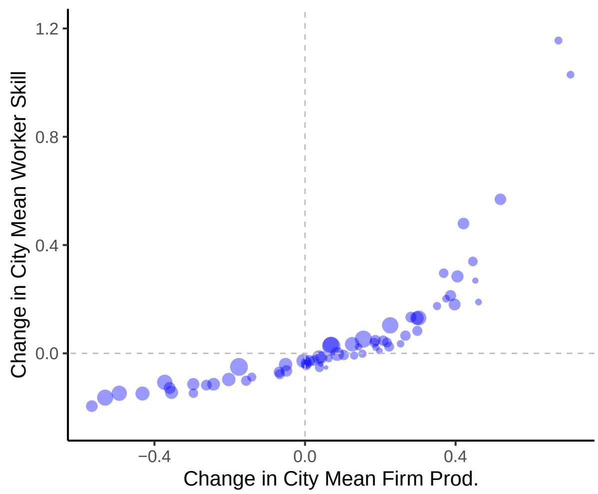
Note: These figures display changes in the spatial allocation of firms and workers caused by the optimal policy against city population in the laissez-faire equilibrium. I define firms in the top three firm clusters as high-productive firms and the remaining ones as low-productive firms, and workers with skill in the top three deciles as high-skilled workers and the remaining ones as low-skilled workers. Population-weighted OLS regression coefficients and standard errors are reported.

firms, and the positive love-of-variety is smaller in large cities as they employ fewer workers.[53] For the high-productivity firms, they also incur negative labor market stealing externality in large cities. However, such social costs are smaller compared to their social benefits as they complement high-skilled workers. The high-productivity firms also employ more workers and generate greater love-of-variety benefits in larger cities. Moreover, the increase of high-productivity firms in larger cities can attract more high-skilled workers from other places, which is efficient from the planner's perspective. In sum, the total number of firms in larger cities declines as a result of the optimal policy, yet average firm productivity increases. For

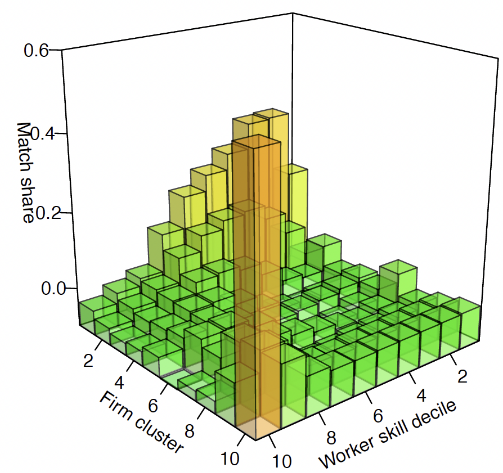
[53]It can be seen from equation (4.3) that the love-of-variety externality is proportional to a firm's employment size when all the welfare weights are equal to one.

Figure 4: Changes in city mean worker and firm FEs and match shares of between workers and firms: optimal policy versus laissez-faire equilibrium

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



(b) Changes in worker-firm match shares



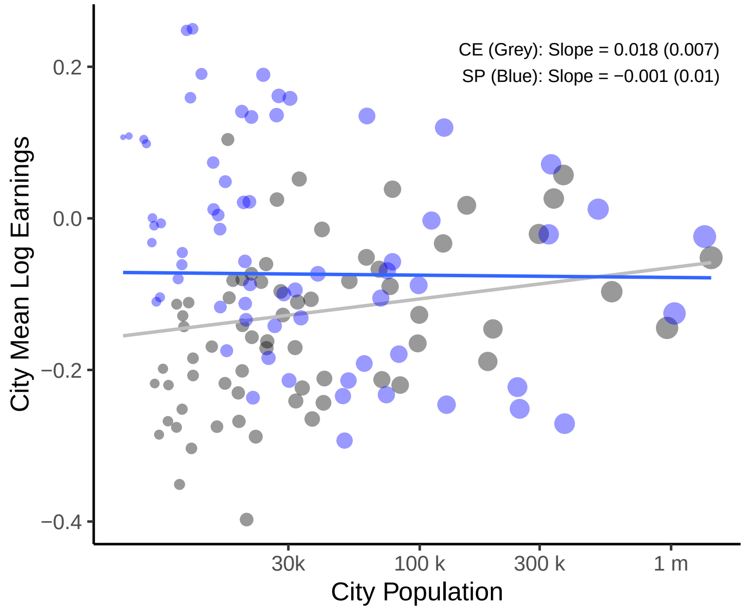
Note: Panel (a) displays the relationship between city-level changes in mean worker skill and mean firm productivity after implementing the optimal policy. Panel (b) displayed changes in the share of workers of each skill decile matched with each firm cluster.

smaller cities, the optimal policy instead allocates more but low-productivity firms there. This result is consistent with Bilal (2023) in that it is optimal to reallocate low-productivity firms to smaller cities, and is consistent with Gaubert (2018) in that the optimal policy strengthens firm spatial sorting.

Panel (c) of Figure 3 shows no significant spatial reallocation of workers for either skill towards larger cities. As a result, population and high-skilled worker share changes do not strongly correlate with original city sizes. This null correlation is expected as the optimal policy makes workers internalize their agglomeration spillovers, but this spillover elasticity μ is estimated to be very small. As I will show soon, the government implements spatial transfers from larger cities to smaller ones, which further weakens worker sorting into large cities. However, there is still significant reallocation between cities of different sizes: mid-size cities tend to grow significantly, whereas the largest and the smallest ones shrink on average. Moreover, there is also a large variation in changes in high-skilled worker shares across cities.

In Figure 4, I investigate changes in assortative co-location and matching caused by the optimal policy. Panel (a) of Figure 4 relates changes in the city mean worker skill, $\bar{a}_c$, to changes in the city mean firm productivity, $\bar{z}_c$. It is evident that the optimal policy greatly enhances the co-location of high-productivity firms and high-skilled workers. Together with Figure I.16, I show that the policy increases both mean worker skill and firm productivity for high-productivity cities and decreases both for low-productivity places. The population-weighted covariance of $\bar{a}_c$ and $\bar{z}_c$ increases by 64.9% and the rank correlation of $\Delta\bar{a}_c$ and $\Delta\bar{z}_c$ is 0.88. Panel (b) of Figure 4 shows the changes in match shares of heterogeneous firms and workers of the economy, where workers are binned into ten skill decile groups. This figure shows the change of the two. By incentivizing the co-location of high-productivity firms and high-skilled workers, the overall assortativeness of worker-firm matching increases considerably. In particular, the share of top decile workers matched to the most productive firms increases by 50 percentage points. Meanwhile, low-skilled

Figure 5: Mean city log earnings: optimal policy versus laissez-faire equilibrium



Note: This figure displays two scatterplots of mean city log earnings versus city population in the efficient equilibrium and in the laissez-faire equilibrium. Under utilitarian welfare weights, the government implements transfers from firms to workers, which I take out proportionally from worker earnings in constructing the figure. The urban earnings premia estimates and their standard errors are reported.

workers are more likely to be matched to low-productivity firms.

To sum up, Figures 3 and 4 reveal that the optimal policy significantly increases assortative co-location and matching of workers and firms. Such spatial allocation decreases the inefficient labor market stealing effect and exploits production complementarity to a much greater extent relative to the laissez-faire economy. As a result, the total output of the economy increases by 9.6%. The output gain of this first-best policy is of a similar magnitude to the gain from the counterfactual policy in Hsieh and Moretti (2019). They show that U.S. GDP would increase by 8.9% if the housing supply regulation only in New York, San Jose, and San Francisco were at the level of the median U.S. city. Without further actions, however, intensified spatial sorting would exacerbate spatial inequality and have negative impacts on social welfare. I show next that the optimal policy mitigates this issue via spatial transfers.

Spatial equality and transfers. I visualize spatial redistribution of the optimal policy in Figure I.15, where I plot net transfers as a share of total income against city population for workers and firms. It shows that the optimal policy redistributes resources from larger cities to smaller ones. In addition, the magnitude of the transfers is much greater for workers than for firms. Some cities receive worker transfers that are more than 100% of the total labor earnings and some give out transfers by more than 50%. Lastly, I compare the urban earnings premium of the two allocations in Figure 5. The urban earnings premium is entirely eliminated with the optimal spatial policy, despite a greater extent of spatial sorting. National-wide mean earnings undoubtedly increase, which comes from more efficient sorting.[54]

Social welfare. The optimal policy increases consumption-equivalent welfare by 47.4%. The large

[54] The utilitarian government also implements transfers from entrepreneurs to workers. I take out such transfers proportionally from worker earnings in Figure 5 to focus on the impact of changes in sorting and transfers between workers.

welfare increase is largely driven by the equity motive as the utilitarian government weighs all agents equally. In the laissez-faire equilibrium, the marginal utility of consumption for low-income is much higher than for high-income agents. Even holding total output fixed, transfers toward low-income agents generate considerable welfare gains. Such massive reallocation may not be realistic. To isolate the efficiency gains, I follow Berger et al. (2022b) to redesign lump-sum transfers such that all workers and entrepreneurs enjoy equal welfare gains while keeping sorting and matching allocations unchanged. This exercise gives a consumption-equivalent welfare increase of 6.5%, which is of a similar magnitude to the 4.0% welfare gain from the optimal spatial policy in Fajgelbaum and Gaubert (2020).[55]

As we have seen, the optimal policy redistributes towards workers in lower-income places. Davis and Gregory (2021) suggest removing the within-type redistribution motive of optimal spatial policies in spatial equilibrium models, as the degree of such redistribution relies heavily on the parametric assumption of the preference shock distribution. I implement a version of the optimal policy without this motive. Specifically, I design type-city-specific taxes so that workers and firms are compensated according to their social values in Proposition 2. Implementing this policy results in a welfare increase of 4.9%. The qualitative patterns of spatial reallocation of firms and workers are similar to the results discussed before, as shown in Figure I.17.

6.2.1 Discussion

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. As a result, the total output increases by only 5.1%, about half of the increase in the benchmark economy. The consumption-equivalent total welfare gain also drops from 47.4% to 45.7%, with the efficiency gains reduced from 6.5% to 4.8%.

The role of two-sided heterogeneity. It is important to account for both worker and firm heterogeneity for the optimal policy design. To show this, I solve for two optimal equilibria when the government overlooks either worker or firm heterogeneity. There are several important differences. First, the total welfare gains coming from the optimal policy considering only one-sided heterogeneity are much smaller. The welfare gains are 17.4% in the case with only worker heterogeneity and 6.7% with only firm heterogeneity. Total output increases by 0.5% in the case with only worker heterogeneity and decreases by 3.1% with only firm heterogeneity.

Second, the optimal spatial allocations of workers and firms vary dramatically, as shown in Figure I.18. When the government does not consider firm heterogeneity, it would reallocate more firms into large cities. This is optimal because it would help attract more workers to these productive locations, as implied by equation (4.5). In contrast, the optimal policy considering two-sided heterogeneity decreases the total number of firms in large cities but increases the number of high-productivity firms there. When

[55] The 6.5% welfare gain includes gains from correcting the three types of sorting externalities and from within-type redistribution. I show next that shutting down within-type redistribution will lead to a welfare increase of 4.9%, implying a 1.6% gain comes from the redistribution channel. I also show that counterfactually changing the number of firms in each city in the laissez-faire equilibrium generates a welfare gain of 0.3%, which arises from the love-of-variety effect. These results suggest that correcting the labor market stealing externality plays the largest role in the welfare gain. However, attributing the total gains to each source is challenging since they are intertwined and cannot be distinctly separated.

the government does not consider worker heterogeneity, the optimal policy still increases the number of high-productivity firms in larger cities. But the benefit from doing so is much smaller in the absence of the gains from assortative worker-firm matching. As a result, workers relocate from larger to smaller cities and total output decreases.

Alternative welfare weights. The extent to which the government redistributes among different agents depends crucially on the welfare weights. As a robustness check, I implement the optimal spatial policy using two alternative sets of welfare weights. First, I calibrate welfare weights so that all agents enjoy equal welfare gains under the optimal policy, which I term as the equal-gain weights. This involves iteratively solving the efficient allocation given a set of weights and updating the weights towards equalizing welfare gains by different agents.[56] Second, I also use Negishi weights where the weights are set to the average income of each agent type in the competitive equilibrium. The welfare gains are 5.7% and 5.2% respectively for the equal-gain weights and the Negishi weights. This suggests that the efficiency gains from the optimal spatial policy are robust to alternative welfare weight specifications.

6.3 Place-based subsidies

I use the model to evaluate two types of place-based policies that are often considered by policymakers: the first is a local wage subsidy to high-productivity firms; the second is a wage subsidy to low-wage places to decrease spatial inequality. In both cases, I assume the subsidies are financed by a flat proportional labor income tax on all workers in the economy.

6.3.1 High-productivity firm subsidies in Toronto

I simulate an economy where firms in the most productive cluster receive a 5% wage subsidy if they locate in Toronto. This counterfactual policy mimics the bid for Amazon HQ2 made by the city of Toronto government. The amount of the subsidy constitutes 0.3% of the total GDP. Moreover, I conduct this policy under different model scenarios to highlight the importance of considering heterogeneity and mobility of both workers and firms for policy evaluation. The comparison of these counterfactual economies and the benchmark economy is shown in Table 6, where Column (1) is the full model, Column (2) is the model with no worker heterogeneity, Columns (3)-(5) restrict the mobility of either firms or workers or both.

Starting from the full model in Column (1), the subsidy generates significant distributional effects. On the one hand, it increases the number of high-productivity firms in Toronto by 30.1% and high-skilled workers in Toronto by 0.6%. On the other hand, the numbers of low-productivity firms and low-skilled workers decrease by 7.7% and 0.5%, respectively. This is because low-skilled workers benefit little from more high-productivity firms locally, and low-productivity firms shrink in employment as the subsidy expands the employment of high-productivity ones. Meanwhile, they all have to face a 2.8% increase in the housing rent in Toronto, which induces them to leave. As a result, the total population and number of firms in Toronto decrease by 0.8% and 2.3% respectively. By attracting more productive agents to the same city, the total output of the economy increases by 0.8%. However, this output gain comes at a cost of an increase in spatial inequality, measured by the variance of the city mean log earnings and the city Gini index, and a decrease in total welfare, which is more severe for low-skilled workers.[57]

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

[57] Qian and Tan (2021) find that the entry of skill-intensive firms benefits local high-skilled workers but hurts low-skilled

Table 6: 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%
City Gini index	19.7%	0.5%	2.6%	11.7%	3.2%
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%	–
Housing rent in Toronto	2.8%	0.8%	0.7%	2.7%	0.9%

In the model with no worker heterogeneity, both aggregate and distributional effects of the subsidy are much smaller. The subsidy has an effect of attracting more firms and workers into Toronto, reflected in the increases in the number of firms by 2.0% and in population by 0.4%. Compared to the full model, the increase in high-productivity firms is much smaller, and some low-productivity firms also follow workers to locate in Toronto. This expansion drives up the housing rent in Toronto by 0.8%. However, absent worker heterogeneity and thus firm-worker complementarity, the production benefit of attracting high-productivity firms into large cities is much smaller. The total output increases by only 0.2%. Spatial inequality and total welfare both change little.

Results of the limited re-sorting model scenarios highlight the interdependence of firm and worker sorting. Without spatial re-sorting, shown in Column (3), the subsidy only serves to expand the employment of subsidized high-productivity firms in Toronto at the cost of other local firms, while other cities are unaffected. Hence, the economy-wide impact in this scenario is limited. With only one-sided re-sorting, shown in Columns (4) and (5), the mobility response of firms and workers are both weaker than the full model, and so are the changes in total output, welfare, and spatial inequality. In conclusion, two-sided sorting strengthens the co-location of high-productivity firms and high-skilled workers and magnifies the aggregate and distributional effects of the subsidy policy.

I perform several robustness checks of the counterfactual results by incorporating endogenous amenities, free entry, and heterogeneous housing expenditure shares into the full model. The results are shown in Columns (1)-(3) of Table I.15. Endogenous amenities increase the extent to which high-skilled workers sort to Toronto after the subsidy. Free entry increases the total number of firms, which is beneficial to workers through the love-of-variety channel. Introducing non-homothetic preference for housing also

workers.

Table 7: Counterfactual analysis: subsidizing low-wage cities

% Changes in	Full Model (1)	Limited Heterogeneity: No firm het. (2)	Limited Heterogeneity: No worker het. (3)	Limited Re-sorting: No-resorting (4)	Limited Re-sorting: Only firm (5)	Limited Re-sorting: Only worker (6)
Var. city log earnings	-28.0%	-22.2%	-16.3%	-30.1%	-30.1%	-28.8%
City Gini index	-11.6%	-12.2%	-9.6%	-13.0%	-13.0%	-12.2%
Total output	-0.3%	-0.5%	-0.3%	0.0%	0.0%	-0.2%
Total welfare	-0.1%	-0.2%	-0.1%	0.2%	0.2%	0.1%
High-skilled welfare	-0.1%	-0.2%	–	0.2%	0.2%	0.0%
Low-skilled welfare	0.0%	-0.1%	–	0.3%	0.3%	0.1%
Pop. in subsidized cities	20.9%	20.0%	21.0%	–	–	14.0%
High-skilled pop.	21.0%	20.0%	–	–	–	14.1%
Low-skilled pop.	20.8%	20.0%	–	–	–	13.9%
# Firms in subsidized cities	24.3%	22.6%	24.7%	–	6.6%	–
# high-prod. firms	25.0%	–	24.5%	–	6.8%	–
# low-prod. firms	24.2%	–	24.8%	–	6.6%	–

increases skill sorting into Toronto as high-skilled workers are less sensitive to housing price increases. The increase in the total output is higher in this case accordingly. Nonetheless, these additional model ingredients change the quantitative result by little.

6.3.2 Low-wage city subsidies

In Table 7, I show counterfactual results of a 10% wage subsidy in the 25 lowest-paying cities, accounting for 12% of the total population. As before, I simulate the policy under different model scenarios, where Column (1) shows the full model, Columns (2)-(3) show the limited heterogeneity models, and Columns (4)-(6) show limited-resorting models. I first focus on the full model, shown in Column (1). 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.

The trade-off between equity gains and efficiency costs of subsidizing distressed places has been discussed by Gaubert et al. (2021c). The takeaway from the two-sided sorting model is that the efficiency costs of such subsidies may be higher with interdependent location choices of workers and firms. This can be seen by comparing Columns (4)-(6) with Column (1). All the limited-resorting model scenarios generate smaller output decreases and positive welfare changes in contrast to the full model. This is because without fully accounting for the location choice interaction, the policy that decreases spatial inequality would seem to induce less movement to low-wage cities, which results in smaller efficiency costs to total output.

I perform the robustness checks and present the results in Columns (4)-(6) in Table I.15. As before, the quantitative results with the model extensions are similar. The two marked differences are the increase in the total number of firms with free entry and a smaller increase in the low-skilled population in low-wage cities with heterogeneous housing expenditure shares.

6.4 The rise of remote work

Lastly, I apply the model to quantitatively investigate the spatial impact of the rise of remote work. Althoff et al. (2022) show that the transition into remote work in the U.S. since the COVID-19 pandemic has shifted the workforce from bigger cities to smaller ones, which has significant distributional welfare consequences across workers. Deng et al. (2020) estimate that 39% of Canadian workers hold jobs that can be done at home, using 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. Essentially, remote work de-couples the workplace and residence place of workers. Workers can choose to live in one city while working for a firm in another city. In the longer run, firms may also respond by changing locations, as labor supply is no longer constrained to the city they live in. Therefore, this rising work arrangement has the potential to weaken the co-location motive and to spatially disperse economic activities.

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. These workers can choose to work for firms either in the city where they reside or in other places. I assume that workers contribute to the agglomeration spillovers in the cities where they live rather than the cities where their firms are located. The remote workers only take into consideration city amenities and living costs when choosing where to reside. Accordingly, firms can hire both onsite workers from the local labor market and remote workers from elsewhere. See more details in Appendix F.3. I analyze this phenomenon with two model scenarios. The first is a short-run model where the firms' locations are fixed; the second is a long-run where firms can also adjust their locations in response to changes in labor supply conditions.

I show the counterfactual results in Table 8. The most striking result is that workers move from large cities to smaller ones with the possibility of remote work. In the short run, the population in the largest five cities decreases by 2.2%, with a larger 3.8% decrease in the high-skilled population. Remote workers move towards small and cheaper cities. On the one hand, this lowers population concentration in large cities. On the other hand, the move-in of higher-skilled workers increases mean earnings in smaller cities. As a result, the variance of city mean log earnings drops significantly by 19.2%. Another benefit is an 8.1% increase in the total output. This is because remote work reallocates workers from low-productivity firms towards high-productivity ones. The population decreases in larger cities also lower housing rents, which allow firms in these places to rent more commercial space and expand production. In the long run, firms also respond by moving out from larger cities to smaller ones, which causes a greater population decline of

Table 8: 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%

2.7% in larger cities. However, spatial inequality decreases less in the long run by 15.4%. Co-movements of firms and on-site workers, in the long run, give rise to new high-paying mid-size cities, like Austin in the U.S. Total output increases more by 9.4% in the long run, as firms optimize their location choices. The comparison of short-run and long-run results again underscores the importance of the location choice interdependence between workers and firms.[58] I also show in Table I.14 that with the rise of remote work, cities with low housing rents and better amenities for high-skilled workers tend to see growth in both population and average earnings.

7 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 to help places that have fallen behind.

In this paper, I build a spatial equilibrium model with a system of cities and mobile heterogeneous workers and firms. I structurally 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. I use the model to evaluate widely considered place-based policies. By comparing the full model to limited re-sorting model scenarios, I emphasize that accounting for location choice interdependence of workers and firms is important for place-based policy evaluation.

The model informs novel sources of sorting externality, that is, firms do not internalize their labor market stealing from other firms and love-of-variety preference for more local firms by workers. I design an optimal spatial policy to achieve efficient sorting and equitable spatial redistribution. I show that the optimal policy can substantially increase social welfare and the total output of the economy. This is done via strengthened co-location of high-skilled workers and high-productivity firms, and spatial redistribution

[58] The short-run and long-run changes in population and the number of firms for five groups of cities ranked by population are plotted in Figure I.19.

towards low-paying cities.

Finally, there are notable limitations to the static equilibrium approach. On the worker side, better learning opportunities are an important advantage of moving to big cities in the early life stage. The quantitative assessment of counterfactual policies does not consider the effects on the dynamic skill formation that is tied to individual location history. On the firm side, many entrepreneurs develop their ideas in big cities and then move to smaller ones for cheaper production inputs. Therefore, the model understates the role of big cities as learning centers and innovation hubs in the spatial economy. The static framework also assumes no moving costs, which is important for evaluating the transition dynamics of spatial policies. Future work can extend the static framework to incorporate these important dimensions.

References

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

Acemoglu, D. (2001): "Good Jobs versus Bad Jobs," *Journal of Labor Economics*, 19, 1–21.

Albouy, D., A. Chernoff, C. Lutz, and C. Warman (2019): "Local Labor Markets in Canada and the United States," *Journal of Labor Economics*, 37, S533–S594.

Allen, T. and C. Arkolakis (2014): "Trade and the Topography of the Spatial Economy," *Quarterly Journal of Economics*, 129, 1085–1140.

Almagro, M. and T. Dominguez-Iino (2022): "Location Sorting and Endogenous Amenities: Evidence from Amsterdam," *Available at SSRN 4279562*.

Althoff, L., F. Eckert, S. Ganapati, and C. Walsh (2022): "The Geography of Remote Work," *Regional Science and Urban Economics*, 93, 103770.

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.

Austin, B., E. Glaeser, and L. Summers (2018): "Jobs for the Heartland: Place-Based Policies in 21st-Century America," *Brookings Papers on Economic Activity*, 151–232.

Bamford, I. (2021): "Monopsony Power, Spatial Equilibrium, and Minimum Wages," *Working Paper*.

Bartik, T. J. (2020): "Using Place-Based Jobs Policies to Help Distressed Communities," *Journal of Economic Perspectives*, 34, 99–127.

Bauluz, L., P. Bukowski, M. Fransham, A. Lee, M. López Forero, F. Novokmet, S. Breau, N. Lee, C. Malgouyres, M. Schularick, et al. (2023): "Spatial Wage Inequality in North America and Western Europe: Changes Between and Within Local Labour Markets 1975-2019," Kiel working paper.

Baum-Snow, N. and L. Han (2023): "The Microgeography of Housing Supply," *Working paper*.

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.

Bayer, P. and C. Timmins (2005): "On the Equilibrium Properties of Locational Sorting Models," *Journal of Urban Economics*, 57, 462–477.

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

Behrens, K. and F. Robert-Nicoud (2015): "Agglomeration Theory with Heterogeneous Agents," *Handbook of Regional and Urban Economics*, 5, 171–245.

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

Berger, D. W., K. F. Herkenhoff, and S. Mongey (2022b): "Minimum Wages, Efficiency and Welfare," Working Paper 29662, National Bureau of Economic Research.

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

Bonhomme, S., K. Holzheu, T. Lamadon, E. Manresa, M. Mogstad, and B. Setzler (2023): "How Much Should We Trust Estimates of Firm Effects and Worker Sorting?" *Journal of Labor Economics*, 41, 291–322.

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

——— (2019b): "Supplement to "A Distributional Framework for Matched Employer Employee Data"," *Econometrica*, 87, 699–739.

Borusyak, K., P. Hull, and X. Jaravel (2022): "Quasi-Experimental Shift-Share Research Designs," *Review of Economic Studies*, 89, 181–213.

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 (2023): "Location, Location, Location," Working Paper 31587, National Bureau of Economic Research.

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.

Costinot, A. (2009): "An Elementary Theory of Comparative Advantage," *Econometrica*, 77, 1165–1192.

Costinot, A. and J. Vogel (2015): "Beyond Ricardo: Assignment Models in International Trade," *Annual Reviews of Economics*, 7, 31–62.

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, P. A. (1982): "Aggregate Demand Management in Search Equilibrium," *Journal of Political Economy*, 90, 881–894.

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.

Doraszelski, U. and J. Jaumandreu (2018): "Measuring the Bias of Technological Change," *Journal of Political Economy*, 126, 1027–1084.

Duranton, G. and D. Puga (2004): "Micro-Foundations of Urban Agglomeration Economies," *Handbook of Regional and Urban Economics*, 4, 2063–2117.

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.

Ehrlich, M. V. and H. G. Overman (2020): "Place-Based Policies and Spatial Disparities across European Cities," *Journal of Economic Perspectives*, 34, 128–149.

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

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

Forslid, R. and T. Okubo (2014): "Spatial Sorting with Heterogeneous Firms and Heterogeneous Sectors," *Regional Science and Urban Economics*, 46, 42–56.

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

Gaubert, C., P. Kline, D. Vergara, and D. Yagan (2021a): "Trends in US Spatial Inequality: Concentrating Affluence and a Democratization of Poverty," *AEA Papers and Proceedings*, 111, 520–25.

Gaubert, C., P. M. Kline, and D. Yagan (2021b): "Place-Based Redistribution," Working Paper 28337, National Bureau of Economic Research.

——— (2021c): "Place-Based Redistribution," Working Paper 28337, National Bureau of Economic Research.

Glaeser, E. L. (2008): *Cities, Agglomeration, and Spatial Equilibrium*, Oxford University Press.

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

Gould, E. D. (2007): "Cities, Workers, and Wages: A Structural Analysis of the Urban Wage Premium," *Review of Economic Studies*, 74, 477–506.

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.

Guvenen, F., F. Karahan, S. Ozkan, and J. Song (2021): "What Do Data on Millions of US Workers Reveal about Lifecycle Earnings Dynamics?" *Econometrica*, 89, 2303–2339.

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.

Henkel, M., T. Seidel, and J. Suedekum (2021): "Fiscal Transfers in the Spatial Economy," *American Economic Journal: Economic Policy*, 13, 433–68.

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 (2018): "Who Benefits From Productivity Growth? Direct and Indirect Effects of Local TFP Growth on Wages, Rents, and Inequality," Working Paper 24661, National Bureau of Economic Research.

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

Huneeus, F., K. Kroft, and K. Lim (2021): "Earnings Inequality in Production Networks," Working Paper 28424, National Bureau of Economic Research.

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

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

Kline, P. and E. Moretti (2013): "Place Based Policies with Unemployment," *American Economic Review*, 103, 238–43.

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.

Manning, A. (2021): "Monopsony in labor markets: A review," *ILR Review*, 74, 3–26.

Marshall, A. (1890): *Principles of Economics*, London: MacMillan Press.

Moretti, E. (2012): *The New Geography of Jobs*, Houghton Mifflin Harcourt.

——— (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.

Neumark, D. and H. Simpson (2015): "Place-Based Policies," *Handbook of Regional and Urban Economics*, 5, 1197–1287.

Notowidigdo, M. J. (2020): "The Incidence of Local Labor Demand Shocks," *Journal of Labor Economics*, 38, 687–725.

Oberfield, E., E. Rossi-Hansberg, P.-D. Sarte, and N. Trachter (2020): "Plants in Space," Working Paper 27303, National Bureau of Economic Research.

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

Porcher, C., H. Rubinton, and C. Santamaría (2023): "JUE insight: The Role of Establishment Size in the City-Size Earnings Premium," *Journal of Urban Economics*, 136, 103556.

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).

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.

Setzler, B. and F. Tintelnot (2021): "The Effects of Foreign Multinationals on Workers and Firms in the United States," *Quarterly Journal of Economics*, 136, 1943–1991.

Slattery, C. and O. Zidar (2020): "Evaluating state and local business incentives," *Journal of Economic Perspectives*, 34, 90–118.

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

Sorkin, I. (2018): "Ranking Firms Using Revealed Preference," *Quarterly Journal of Economics*, 133, 1331–1393.

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.

Suárez Serrato, J. C. and O. M. Zidar (2023): "Who Benefits from State Corporate Tax Cuts? A Local Labor Market Approach with Heterogeneous Firms: Further Results," Working Paper 31206, National Bureau of Economic Research.

Appendix A Data Appendix

In this Section, I describe the data sources, variable construction, and sample selection rules 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 is required to submit for all Canadians annually. Specifically, I observe each individual's age, gender, marital status, and residence location from T1-PMF, and the number of children from T1-FF. Each individual has a unique longitudinal identifier based on the Social Insurance Number (SIN). There is an important caveat in using the location information in CEEDD. In T1-PMF, the location information in the file of year t is based on the reported postal code in year $t+1$. This is because individuals file their taxes with a one-year lag, which also applies to firms. To address this, I use each individual's location information in the T1-PMF file of year $t-1$ as her location in year t. This implies that I do not have location information for year 2001, the first year of the coverage. Hence, I only use data from 2002 to 2017.

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).

As discussed in the main text, the NALMF dataset only records the headquarters location for firms that operate in more than one locations and includes a multi-location flag. 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 also treat firms who have moved across cities as different firms before and after the move, as such relocations are often associated with changes in the production process, labor demand, and wage policies as discussed in Duranton and Puga (2004).

T4 files. Annual earnings information and the employee-employer linkage are from the Statements of Renumeration 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 can have more than one T4 records in a year if she works for more than one firms. For multiple job holders, I keep the job that offers the largest earnings of the year and call it the main job. In addition, I drop workers with annual earnings from the main job lower than a threshold, which is the equivalence of working 40 hours per week for 13 weeks paid at half of 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. I use the IMDB files to construct the immigration-based labor supply instrument (Card, 2001).

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 convert 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.

In addition, I also consider land areas covered by Ontario's Greenbelt Plan as undevelopable land. The Greenbelt Act, which became law in Feb 2005, was designed to curb urban sprawl into environmentally sensitive areas in Ontario. It prevents the re-zoning of agricultural land, heritage sites, and sensitive ecological features for use in urban development. The boundary file for the Greenbelt designation is obtained from Ontario GeoHub. I show the topography maps for the largest four Canadian cities in Figure I.1.

Share of college graduates by city. I calculate the share of individuals with a college or a higher degree for each city (CMA/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 Model Appendix

B.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:

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

subject to

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

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. I assume that firms are infinitesimal so any firm's wage-setting decision cannot affect the local wage index. Taking the derivative with respect to $\{W_{jc}(a)\}_{\forall a}$ and h_{jc}, we have the following set of FOCs:

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

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

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:

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

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

where $\chi \equiv \frac{\beta_w/\rho_w}{1+\beta_w/\rho_w} \cdot (1-\alpha)\alpha^{\frac{\alpha}{1-\alpha}}$. Equation (B.6) still contains the endogenous term Q_{jc}. Plugging equation (B.5) into the definition of Q_{jc} yields:

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

where

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

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

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

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

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

where $\Psi \equiv \frac{1}{1+\beta_w/\rho_w} \cdot (1-\alpha)\alpha^{\frac{\alpha}{1-\alpha}}$. It is obvious from equations (B.9) and (B.10) that the firm's optimal profits $\pi_c(j)$ is proportional to its total efficiency unit of labor Q_{jc} and also to its total wage bill E_{jc}.

B.2 The housing market

A housing developer of a city combines land $\overline{H}_c$, which is exogenously given for each city, with the final good Y_c to produce housing. The developer chooses final good inputs to maximize profits:

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

The FOC of this profit-maximizing problem is:

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

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

$$H_c^S(r_c) = \overline{H}_c^0 \cdot r_c^{\frac{1}{\gamma_c}} \tag{B.13}$$

where $\overline{H}_c^0 = (1+\gamma_c)^{-1/\gamma_c} \cdot \overline{H}_c{}^{(1+\gamma_c)/\gamma_c}$. Housing is demanded by both workers and firms in the city. By the properties of workers' preference and firms' production technology, workers spend a constant share η of their total income τW on housing, and firms expense a constant share α of their production costs on housing. Total housing demand for a city can thus be derived as a constant share of the total wage bill:

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

where I define E_c as the total wage bill of city c. The housing market clears for city c when demand equals supply, namely

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

I assume that the housing developer owns the land and the housing rent is priced competitively. The housing developers' profits are aggregated to a national portfolio and then rebated to workers. The amount of the rebate each worker obtains is proportional to her wages. Specifically, the rebate follows the expression below

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

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 market clearing condition is given by

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

Appendix C Efficiency

C.1 Social planner's problem

Assume the planner aims to maximize the social welfare function of the economy, which is specified as

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

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

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

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

where

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

and $\bar{C^w}$ and $\bar{C^f}$ are unrecoverable constants.

The planner chooses the following to maximize equation (C.1): 1) the amount of the final good and housing allocated to skill a workers working for a 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 recognizes that workers and firms have idiosyncratic shocks but cannot directly observe these shocks. For workers, there are within-city and between-city allocation constraints

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

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

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

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

For firms, the spatial mobility constraints are:

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

Second, there are worker and firm allocation constraints:

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

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

which are automatically satisfied given (C.7) and (C.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:

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

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

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

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

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

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

C.2 Solving the social planner's problem

The Lagrange function of the social planner's problem is

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

where $W^*_{kc}(a)$ is the multiplier for the worker allocation constraint (C.7); $\pi^*_c(k)$ is the multiplier for the firm allocation constraint (C.8); $P^*$is the multiplier for the resource constraint of the final good (C.11); R^*_c is the multiplier for local housing constraints (C.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

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

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

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

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

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

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

$$\partial D_{kc}(a): W^*_{kc}(a) + R^*_c h_{kc}(a) + P^* c_{kc}(a) - P^*(1-\alpha) \cdot \alpha^{\frac{\alpha}{1-\alpha}} R_c^{*\frac{\alpha}{\alpha-1}} \left[f_c(k,a) + \frac{\mu}{L_c} \sum_{k'} \sum_{a'} D_{k'c}(a') f_c(k',a')\right] = 0 \tag{C.20}$$

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

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

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

$$\partial Y_c: R^*_c \overline{H}_c \frac{1}{1+\gamma_c} Y_c^{\frac{\gamma_c}{1+\gamma_c}} = P^*, \forall c. \tag{C.22}$$

C.3 Characterization of the social planner's solution

Starting from equation (C.16), we can re-arrange this equation as

$$\varphi^{w}(a) \cdot (1-\eta) = P^* c_{kc}(a) - \frac{\beta_w}{\rho_w}(1-\eta) W^*_{kc}(a) + \left(\frac{\beta_w}{\rho_w} - \beta_w\right) \sum_{k'} (1-\eta) W^*_{k'c}(a) \frac{D_{k'c}(a)}{L_c(a)} + \beta_w \sum_{c'} \sum_{k'} (1-\eta) W^*_{k'c'}(a) \frac{D_{k'c'}(a)}{L(a)}. \tag{C.23}$$

Similarly, we can rewrite equation (C.17) as

$$\varphi^{w}(a)\cdot\eta = R_c^*h_{kc}(a) - \frac{\beta_w}{\rho_w}\eta W_{kc}^*(a) + \left(\frac{\beta_w}{\rho_w} - \beta_w\right)\sum_j \eta W_{k'c}^*(a)\frac{D_{k'c}(a)}{L_c(a)} + \beta_w\sum_{c'}\sum_{k'}\eta W_{k'c'}^*(a)\frac{D_{k'c'}(a)}{L(a)}. \quad \text{(C.24)}$$

Combining the two equations above, we obtain

$$\varphi^{w}(a) - \left(\frac{\beta_w}{\rho_w} - \beta_w\right)\sum_{k'} W_{k'c}^*(a)\frac{D_{k'c}(a)}{L_c(a)} - \beta_w\sum_{c'}\sum_{k'} W_{k'c'}^*(a)\frac{D_{k'c'}(a)}{L(a)} = P^*c_{kc}(a) + R_c^*h_{kc}(a) - \frac{\beta_w}{\rho_w}W_{kc}^*(a). \quad \text{(C.25)}$$

and

$$\eta P^*c_{kc}(a) = (1-\eta)R_c^*h_{kc}(a). \quad \text{(C.26)}$$

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

$$W_{kc}^*(a) = -R_c^*h_{kc}(a) - P^*c_{kc}(a) + P^*(1-\alpha)\cdot\alpha^{\frac{\alpha}{1-\alpha}}R_c^{*\frac{\alpha}{\alpha-1}}\left[f_c(k,a) + \frac{\mu}{L_c}\sum_{k'}\sum_{a'}D_{k'c}(a')f_c(k',a')\right]. \quad \text{(C.27)}$$

Plugging $W_{kc}^*(a)$ into equation (C.25) then leads to the condition for socially optimal worker expenditures

$$P^*c_{kc}(a) + R_c^*h_{kc}(a) = \frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}\left(\tilde{W}_{kc}(a) - O_c^w(a)\right) + \frac{1}{1+\beta_w/\rho_w}\varphi^w(a) \quad \text{(C.28)}$$

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; $O_c^w(a) = (1-\rho_w)\sum_{k'}W_{k'c}^*(a)\frac{D_{k'c}(a)}{L_c(a)} + \rho_w\sum_{c'}\sum_{k'}W_{k'c'}^*(a)\frac{D_{k'c'}(a)}{L(a)}$ 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 $T_c^w(a) \equiv -\frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}O_c^w(a) + \frac{1}{1+\beta_w/\rho_w}\varphi^w(a)$ and rewrite the optimal condition as

$$P^*c_{kc}(a) + R_c^*h_{kc}(a) = \frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}\tilde{W}_{kc}(a) + T_c^w(a). \quad \text{(C.29)}$$

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

$$\varphi^f(k) = P^*c_c(k) - \beta_f\pi_c^*(k) + \beta_f\sum_{c'}\pi_{c'}^*(k)\frac{J_{c'}(k)}{J(k)}. \quad \text{(C.30)}$$

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

$$\sum_a \varphi^w(a) L(a) \frac{\partial U(a)}{\partial J_c(k)} + \sum_a W_{kc}^*(a) D_{kc}(a) \frac{1}{J_c(k)} - (1-\rho_w) \sum_{k'} \sum_a W_{k'c}^*(a) D_{k'c}(a) \frac{D_{kc}(a)}{L_c(a)} \frac{1}{J_c(k)} - \rho_w \sum_{c'} \sum_{k'} \sum_a W_{k'c'}^*(a) D_{k'c'}(a) \cdot \frac{D_{kc}(a)}{L_c(a)} \frac{1}{J_c(k)} - \pi_c^*(k) - P^* c_c(k) = 0. \tag{C.31}$$

Further,

$$(1+\beta_f) P^* c_c(k) + \beta_f \sum_{c'} \pi_{c'}^*(k) \frac{J_{c'}(k)}{J(k)} - \varphi^f(k) = \beta_f \Bigg[\sum_a \varphi^w(a) L(a) \frac{\partial U(a)}{\partial J_c(k)} + \sum_a W_{kc}^*(a) D_{kc}(a) \frac{1}{J_c(k)} - (1-\beta_c \cdot \rho_w/\beta_w) \sum_{k'} \sum_a W_{k'c}^*(a) D_{k'c}(a) \frac{D_{kc}(a)}{L_c(a)} \frac{1}{J_c(k)} - \beta_c \cdot \rho_w/\beta_w \sum_{c'} \sum_{k'} \sum_a W_{k'c'}^*(a) D_{k'c'}(a) \cdot \frac{D_{kc}(a)}{L_c(a)} \frac{1}{J_c(k)} \Bigg] \tag{C.32}$$

which, after some algebra, can be simplified as

$$P^* c_c(k) = \frac{\beta_f}{1+\beta_f} \left(\tilde{\pi}_c(k) - O^f(k) \right) + \frac{1}{1+\beta_f} \varphi^f(k). \tag{C.33}$$

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

$$\tilde{\pi}_c(k) = \underbrace{\sum_a W_{kc}^*(a) \frac{D_{kc}(a)}{J_c(k)}}_{\text{firm surplus}} - (1-\rho_w) \underbrace{\sum_a \bar{W}_c^*(a) \frac{D_{kc}(a)}{J_c(k)}}_{\text{local labor stealing}} - \rho_w \underbrace{\sum_a \bar{W}^*(a) \frac{D_{kc}(a)}{J_c(k)}}_{\text{national labor stealing}} + \underbrace{\frac{\rho_w}{\beta_w} \sum_a \varphi^w(a) \frac{D_{kc}(a)}{J_c(k)}}_{\text{love-of-variety preference}} \tag{C.34}$$

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.

The expression of $\tilde{\pi}_c(k)$ makes explicit the externalities that firms do not consider when making a location decision: first, it competes for workers with other firms in the same city; second, it attracts workers from other cities; third, it increases the utility of workers in the city via their love-of-variety preference for more workplace choices. As before, I define $T^f(k) \equiv -\frac{\beta_f}{1+\beta_f} O^f(k) + \frac{1}{1+\beta_f} \varphi^f(k)$ and rewrite the firm optimality condition as

$$P^* c_c(k) = \frac{\beta_f}{1+\beta_f} \tilde{\pi}_c(k) + T^f(k). \tag{C.35}$$

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 (C.16)-(C.22), the spatial mobility constraints (C.5)-(C.8), and resources constraints (C.11)-(C.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 $\{x_{kc}^*(a), x_c^*(k)\}$,[59] equations (C.16)-(C.17) coincide with the utility maximization condition implied by (3.1), equation (C.19) coincide with the optimal housing input condition (B.6), equations (C.5)-(C.8) coincide with optimality location and matching conditions (3.3)-(3.4) and (3.12), equation (C.22) coincides with the optimality condition of the developer (B.12), equations (C.11)-(C.12) coincide with market clearing conditions (B.15)-(B.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 (C.28) and $x_c(k) = x_c^*(k)$ where $x_c^*(k)$ is given by (C.33). Conversely, if $x_{kc}(a) = x_{kc}^*(a)$ and $x_c(k) = x_c^*(k)$ for $\{x_{kc}(a), x_c(k)\}$ defined in (4.7) and (4.8) 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.

C.4 Optimal spatial policies

After characterizing the social planner's solution and recognizing externalities that come from workers' and firms' location choices, I now design a set of optimal spatial policies to restore efficient spatial allocation. The optimal spatial policies consist of (1) a set of labor 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 firm profit taxes $\{t_c^f(k)\}_{\forall k,c}$ that are specific to firm type k and city c. To restore efficient allocation, the labor and profit taxes have to reflect sorting externalities and redistribution motives. Specifically, set the labor taxes 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 set the profit taxes 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, and $\{\tilde{W}_{kc}(a)\}_{\forall a,k,c}$ is given by

$$\tilde{W}_{kc}(a) = \frac{1+\beta_w/\rho_w}{\beta_w/\rho_w}\left(W_{kc}(a) + \mu\bar{W}_c\right), \quad \forall a, k, c$$

and $\{\tilde{\pi}_c(k)\}_{\forall k,c}$ is given by

$$\tilde{\pi}_c(k) = \frac{\beta_f}{1+\beta_f}\sum_a \frac{D_{kc}(a)}{J_c(k)}\left(W_{kc}^*(a) - (1-\rho_w)\bar{W}_c^*(a) - \rho_w\bar{W}^*(a) + \frac{\rho_w}{\beta_w}\varphi^w(a)\right), \quad \forall k, c$$

[59] The price of the final good, P, is normalized to be one in the competitive equilibrium.

and $\{W^*_{kc}(a)\}_{\forall a,k,c}$ and $\{R^w_c(a)\}_{\forall a,c}$ are determined by the following system of equations

$$W^*_{kc}(a) = \frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}\tilde{W}_{kc}(a) - T^w_c(a), \quad \forall a,k,c$$
$$T^w_c(a) = -\frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}\left((1-\rho_w)\bar{W}^*_c(a) + \rho_w\bar{W}^*(a)\right) + \frac{1}{1+\beta_w/\rho_w}\varphi^w(a), \quad \forall a,c$$

and $\{\pi^*_c(k)\}_{\forall k,c}$ and $\{R^f(k)\}_{\forall k}$ are determined by the following system of equations:

$$\pi^*_c(k) = \frac{1}{1+\beta_f}\tilde{\pi}_c(k) - T^f(k), \quad \forall k,c \tag{C.36}$$
$$T^f(k) = -\frac{\beta_f}{1+\beta_f}\bar{\pi}^*(k) + \frac{1}{1+\beta_f}\varphi^f(k), \quad \forall k. \tag{C.37}$$

I assume that the housing developers' profits are rebated to workers, which is proportional to the after-tax labor income $(1-\tau^w_{kc}(a))\,W_{kc}(a)$. 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

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

Then, the budget constraint for such a worker is

$$Pc_{kc}(a) + r_c h_{kc}(a) = \tau\left(\frac{\beta_w/\rho_w}{1+\beta_w/\rho_w}\tilde{W}_{kc}(a) + T^w_c(a)\right). \tag{C.38}$$

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

$$I^f_c(k) = \left(1-t^f_c(k)\right)\pi_{kc}(a) = \frac{\beta_f}{1+\beta_f}\tilde{\pi}_c(k) + T^f(k) \tag{C.39}$$

and the budget constraint for such an entrepreneur is

$$Pc_c(k) = \left(1-t^f_c(k)\right)\pi_{kc}(a) = \frac{\beta_f}{1+\beta_f}\tilde{\pi}_c(k) + T^f(k). \tag{C.40}$$

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

$$-\sum_c\sum_k\sum_a D_{kc}(a)\left(t^w_{kc}(a)\,W_{kc}(a)\right) - \sum_c\sum_k J_c(k)\left(t^f_c(k)\,\pi_c(k)\right) = 0 \tag{C.41}$$

or equivalently

$$\sum_c\sum_k\sum_a D_{kc}(a)\,T^w_c(a) + \sum_c\sum_k J_c(k)\,T^f(k) = \sum_c\sum_k\sum_a D_{kc}(a)\left(W_{kc}(a) - \tilde{W}_{kc}(a)\right) + \sum_c\sum_k J_c(k)\left(\pi_c(k) - \tilde{\pi}_c(k)\right) \tag{C.42}$$

To show this, we can rewrite the transfer terms $T_c^w(a)$ as:

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

and $T^f(k)$ as

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

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

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

Hence, I have proved the government budget balance given by equation (C.42). Then, if the planner's problem is globally concave, the equilibrium with the specified taxes implements the efficient allocation, which is given by the following: The after-tax budget constraints for workers (C.38) are identical to the ones in the social planner's problem (C.29). 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 (3.1), are identical to the social planner's conditions (C.16) and (C.17). The between-city location choice and within-city firm choice conditions of workers, with the after-tax worker income (3.4) and (3.3) are identical to the optimal worker allocation conditions in the social planner's problem (C.6) and (C.5). The after-tax budget constraints for entrepreneurs (C.39) are identical to the ones in the social planner's problem (C.35). The location choice condition for firms, given after-tax entrepreneur income, implied by (3.12), is identical to the optimal firm allocation conditions in the social planner's problem (C.8). The first order condition for final good inputs used by the housing developers (B.12) is identical to the one of the social planner (C.22). The local labor market clearing conditions are automatically satisfied given (C.5). The market clearing conditions for the final good and housing (B.17) and (B.15) are the same as the resource constraints in the planner's problem (C.11) and (C.12).

Appendix D Uniqueness

To illustrate the role of β_w/ρ_w and θ on multiplicity, I conduct the following simulation exercise. Consider a simple economy with two cities, $c \in \{1, 2\}$, sharing the same exogenous productivity A_c, amenities R_c, and housing supply parameters $\{\gamma_c, H_0^c\}$. I set $\beta_f = 3$, $\alpha = 0$, $\eta = 0.3$ and $\mu = 0$. Furthermore, suppose there are two types of workers varying in skill $\mathbf{a} = (0.2, 1)$ and two types of firms varying in productivity $\mathbf{z} = (0, 0.5)$ and the complementarity parameter $\Theta = (1, \theta)$. I perform 3×3 simulation experiments in which I set $\beta_w/\rho_w \in \{0.6, 1.1, 2\}$ and $\theta \in \{1, 1.5, 3\}$. The "chasing mechanism" is most relevant for the pair of type 2 workers and type 2 firms. To isolate this relationship, I fix the location of type 1 workers and type 1 firms to be equally distributed across the two cities and do not allow them to move.

Figure I.6 displays the result. In the figure, red lines are the best response curves of the share of type-2 workers in city 1 given the share of type-2 firms in city 1, and blue lines are the best response curves of the share of type-2 firms in city 1 given the share of type-2 workers in city 1. The interaction of the red and blue curves thus are the Nash Equilibria. Two patterns can be drawn from the simulation results: multiple equilibria are more likely to arise when 1) β_w/ρ_w is small and 2) θ is large. These simulation results are in line with the discussion in the main text. Given my empirically estimated a large β_w/ρ_w and a moderate range of θ, the possibility for multiple equilibria is limited with other dispersion forces introduced in the model. I further experiment with solving the estimated model with different initial guesses and they all converge to the same solution.

Appendix E Identification Appendix

E.1 Time-varying terms

Here, I add the time-varying shocks and functional form assumptions specified in Section 5.2 to key terms in the model. First, worker i's wage in firm j, city c and time t becomes

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

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

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

and its optimal profit is

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

where the labor composite term is

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

E.2 Worker sorting elasticities

Here, I show how to obtain the two wage terms $\tilde{W}_{it}$, $\bar{W}_{ct}$ and two wage bill terms $\tilde{E}_{jt}$ and $\bar{E}_{ct}$ for constructing the passthrough equations (5.7) and (5.8). Start with the time-varying wage equation given by equation (E.1):

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

As can be easily seen, 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 $\tilde{W}_{it}$ to isolate firm-level shocks. The mean and residualized wage bill terms are constructed for the same purposes. Formally, with equation (E.5) the average wage for stayers is given by

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

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:

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

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 (E.2):

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

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

$$\begin{aligned}\hat{\kappa}_{jct}(a) &= L(a) \cdot \frac{\left[R_c(a) r_{ct}^{-\eta} \left(\sum_{j \in \mathcal{J}_{ct}} \left(W_{jct}(a) G_{jc}(a)\right)^{\beta_w/\rho_w}\right)^{\rho_w/\beta_w}\right]^{\beta_w}}{\sum_{c'} \left[R_{c'}(a) r_{c't}^{-\eta} \left(\sum_{j' \in \mathcal{J}_{c't}} \left(W_{j'c't}(a) G_{j'c}(a)\right)^{\beta_w/\rho_w}\right)^{\rho_w/\beta_w}\right]^{\beta_w}} \cdot \frac{\left(G_{jc}(a)\right)^{\frac{\beta_w}{\rho_w}}}{\sum_{j' \in \mathcal{J}_{ct}} \left(W_{j'ct}(a) G_{j'c}(a)\right)^{\frac{\beta_w}{\rho_w}}} \\ &= \left(A_c \hat{A}_{ct} L_{ct}^{\mu} r_{ct}{}^{\frac{\alpha}{\alpha-1}}\right)^{\beta_w - \beta_w/\rho_w} \cdot \left(r_{ct}\right)^{-\eta\beta_w} \left(G_{jc}(a)\right)^{\frac{\beta_w}{\rho_w}} \cdot \hat{\kappa}_{ct}(a)\end{aligned} \tag{E.9}$$

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

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

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:

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

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:

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

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:

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

I can now derive the passthrough equations (5.7) and (5.8) using (E.6), (E.7), (E.12) and (E.13). Taking log and first differences of (E.7) and (E.13) and plugging in the measurement error e_{jt}, we have the net passthrough equation as:

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

where $\delta_w \equiv \frac{1}{1+\beta_w/\rho_w}$. There are three residual terms in the net passthrough equation (E.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 $\log \Delta \hat{\tilde{E}}_{jt}$ 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_{jt}^e 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 following Doraszelski and Jaumandreu (2018). Specifically, I assume $\log(\hat{\phi}_{jct}/\hat{\phi}_{ct}) = F^N \left(\log(\hat{\phi}_{jct-1}/\hat{\phi}_{ct-1})\right) + \epsilon_{jct}^N = H^N(\log \hat{W}_{ijt-1}, \log \hat{\tilde{E}}_{jt-1}) + \epsilon_{jct}^N$, where ϵ_{jct}^N 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 $\log \hat{\tilde{E}}_{jt-1}$, using equation (E.14) at $t-1$. When $F^N(\cdot)$ is a linear function, the first difference of the relative shifter can be written as $\Delta \log(\hat{\phi}_{jct}/\hat{\phi}_{ct}) = F^N \left(\Delta \log(\hat{\phi}_{jct-1}/\hat{\phi}_{ct-1})\right) + \Delta\epsilon_{jct}^N = H^N(\Delta \log \hat{W}_{ijt-1}, \Delta \log \hat{\tilde{E}}_{jt-1}) + \Delta\epsilon_{jct}^N$. In practice, I control for $H^N(\cdot, \cdot)$ using a cubic polynomial of the two lagged first-difference terms.

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

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

which can be further re-arranged as:

$$\Delta \log \left(\frac{\bar{W}_{ct}}{r_{ct}^{\eta}}\right) = \delta_c \Delta \log \left(\frac{\bar{\tilde{E}}_{ct}}{r_{ct}^{\eta}}\right) + \delta_c \left(\Delta e_{ct} - \Delta \log \hat{\phi}_{ct}\right). \tag{E.16}$$

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 $\dot{\bar{E}}_{ct}$. 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 $\Delta \log \left(\dot{\bar{E}}_{ct}/r_{ct}^{\eta}\right)$ using its lags before year $t-q-1$ and control for the two lagged first-differenced terms, $\Delta \log \left(\bar{W}_{ct-1}/r_{ct-1}^{\eta}\right)$ and $\Delta \log \left(\dot{\bar{E}}_{ct-1}/r_{ct-1}^{\eta}\right)$.

E.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 (E.12) as:

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

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

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

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 (E.3) and (E.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 (E.17) and (E.18) and introducing a measurement error of $\bar{E}_{ct}$, I can obtain the following passthrough equation for firms:

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

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 $\Delta \log \dot{E}_{ct-1}$.

E.4 Worker and firm fixed effects

I show here how I construct the adjusted log earnings $\log \bar{W}_{ijt}(a)$ 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:[60]

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

and

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

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

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

E.5 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 (3.3) and (3.4), we can obtain

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

[60] 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.

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

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 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

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

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

$$\begin{aligned}\Lambda_{kc}(a)^{\frac{\rho_w}{\beta_w}}\Lambda_c(a)^{\frac{1}{\beta_w}} &= R_c(a)\bar{r}_c^{-\eta}\cdot \mathbb{A}_c r_c^{\frac{\alpha}{\alpha-1}}\left(\sum_{j\in\mathcal{J}_c}\left(z_k a^{\theta_k} G_k(a)\right)^{\frac{\beta_w}{\rho_w}}\right)^{\rho_w/\beta_w}\cdot\frac{\bar{J}_c(k)^{\frac{\beta_w}{\rho_w}}\left(z_k a^{\theta_k}G_k(a)\right)}{\left(\sum_{j\in\mathcal{J}_c}\left(z_k a^{\theta_k} G_k(a)\right)^{\frac{\beta_w}{\rho_w}}\right)^{\rho_w/\beta_w}} \\ &= \bar{r}_c^{-\eta}\cdot \mathbb{A}_c r_c^{\frac{\alpha}{\alpha-1}} z_k a^{\theta_k}\cdot \bar{J}_c(k)^{\frac{\beta_w}{\rho_w}}\cdot R_c(a)G_k(a)\end{aligned} \tag{E.26}$$

and it can be rewritten as

$$R_c(a)\cdot G_k(a) = \frac{\bar{r}_c^{\eta}}{\mathbb{A}_c r_c^{\frac{\alpha}{\alpha-1}} z_k a^{\theta_k}}\cdot \bar{J}_c(k)^{\frac{\beta_w}{\rho_w}}\cdot\Lambda_{kc}(a)^{\frac{\rho_w}{\beta_w}}\Lambda_c(a)^{\frac{1}{\beta_w}}. \tag{E.27}$$

E.6 City composite productivity

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

$$\bar{\pi}_c(k) = \Psi\cdot\left(\mathbb{A}_c\bar{r}_c^{\frac{\alpha}{\alpha-1}}\right)\cdot (z_k)^{1+\beta_w/\rho_w}\cdot\phi_{kc} \tag{E.28}$$

where

$$\phi_{kc} = \int_a \left(a^{\theta_k}\right)^{1+\beta_w/\rho_w} L_c(a)\cdot\frac{G_k(a)^{\beta_w/\rho_w}}{\sum_{j'\in\mathcal{J}_c}\left(z_{j'}a^{\theta_{j'}}\cdot G_{j'c}(a)\right)^{\frac{\beta_w}{\rho_w}}}da. \tag{E.29}$$

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. I can then plug equations (E.28) and (E.29) into (3.12), and 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}, which is given by:

$$\begin{aligned} p_c(k) &= \frac{\bar{\pi}_c(k)^{\beta_f}}{\sum_{c'} \bar{\pi}_{c'}(k)^{\beta_f}} \\ &= \frac{\left(\left(\mathbb{A}_c \bar{r}_c{}^{\frac{\alpha}{\alpha-1}}\right) \cdot \phi_{kc}\right)^{\beta_f}}{\sum_{c'} \left(\left(\mathbb{A}_{c'} \bar{r}_{c'}{}^{\frac{\alpha}{\alpha-1}}\right) \cdot \phi_{kc'}\right)^{\beta_f}}. \end{aligned} \tag{E.30}$$

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 E.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.

E.7 Agglomeration elasticity

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

$$\log \mathbb{A}_c = A_0 + \mu \log \bar{L}_c + \epsilon_c^A \tag{E.31}$$

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:

$$IV_c^L = \sum_o \frac{\sum_{t=1980}^{t=2002} IMM_{o,c,t}}{L_{c,2002}} \cdot \log I\hat{M}M_o \tag{E.32}$$

where o denotes the origin country of immigrants, $IMM_{o,c,t}$ is the number of immigrants from origin country o, landing in city c in year t, and $I\hat{M}M_o$ is the total number of immigrants of origin country o arriving in Canada since 2002. This is a shift-share style instrument. Ideally, I would construct the share term as the city population share of immigrants from each origin country at a base year. However, the IMDB data, as part of CEEDD, only has information on immigrants who landed in Canada since 1980. Nevertheless, the ratio of accumulated counts of arriving immigrants since 1980 to city population in 2002 should be strongly correlated with the share of residing immigrants in each city and thus reflects each city's exposure to future immigration. This instrument is correlated with the endogenous variable as migrant inflows increase city population. The identification assumption is that the total number of immigrants by origin country $I\hat{M}M_o$, namely the shift terms in the shift-share instrument, is orthogonal to city-level exogenous productivity (Borusyak et al., 2022).

E.8 Housing supply parameters

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

$$H_{ct}^{S}\left(r_{ct}\right) = \overline{H}_c^0 \cdot r_{ct}^{\frac{1}{\gamma_c}} = \left(\tau\eta + \frac{\alpha}{1-\alpha}\frac{1+\beta_w/\rho_w}{\beta_w/\rho_w}\right)\frac{E_{ct}}{r_{ct}} = H_c^D\left(r_{ct}\right). \tag{E.33}$$

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

$$r_{ct}^{\frac{1+\gamma_c}{\gamma_c}} = \frac{1}{\overline{H}_c^0} \cdot \left(\tau\eta + \frac{\alpha}{1-\alpha}\frac{1+\beta_w/\rho_w}{\beta_w/\rho_w}\right)E_{ct} = \frac{1}{\overline{H}_c^0} \cdot EH_{ct} \tag{E.34}$$

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:

$$\Delta\log r_{ct} = \Gamma_c \Delta \log EH_{ct} + \Delta e_{ct}^{r} \tag{E.35}$$

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

$$\Delta\log r_{ct} = \left(\Gamma + \Gamma_L UNAVAL_c\right)\Delta\log EH_{ct} + \Delta e_{ct}^{r}. \tag{E.36}$$

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:

$$IV_c^{EH} = \sum_{ind}\frac{L_{c,ind,2002}}{L_{c,2002}} \times \Delta\log wage_{ind} \tag{E.37}$$

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 $\overline{H}_c^0$ to fit the average housing rent of each city, according to equation (E.34). 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 F Model extensions

F.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:

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

where I define H and L as the high- and low-skilled groups depending on worker skill a, $L_c(H)$ and $L_c(L)$ are the numbers of high- and low-skilled workers in city c, and the four γ parameters, $\gamma_{gg'}$, are the amenity spillover elasticities from skill group g to g'. In the quantitative model with this extension, I define high-skilled workers as those with skill a in the top three deciles, low-skilled workers as those with skill a in the bottom seven deciles, and calibrate these parameters following Fajgelbaum and Gaubert (2020) as $\gamma_{HH} = 0.77$, $\gamma_{LH} = -1.24$, $\gamma_{HL} = 0.18$, $\gamma_{LL} = -0.43$.

F.2 Free entry

I assume in the baseline model that there is a fixed measure of firms that make location decisions. 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 equating the expected utility of an entrepreneur if she enters the economy to the utility of consuming c_e units of the final good:

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

where $\bar{C}^f$ is an unrecoverable constant from aggregating the Type-I Extreme Value distributed preference shocks.

F.3 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

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

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

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

where $L^O_c(a)$ 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,

$$
D_{jc}(W, a) = S^R_{jc}(W, a) + S^O_{jc}(W, a)
$$

I assume that 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 (3.4). Remote workers choose locations in only based on local amenities and costs of living

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

In the short run, I assume that firms' locations are fixed as in the benchmark equilibrium. In the long run, firms can adjust their locations in response to the shifts in the labor market.

Appendix G Shapley value

In the counterfactual exercises studied in Section 6.1.1, 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 $\mathcal{N} = \{1, 2, \dots, N\}$ 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 parameter ω_n to outcome X is $\frac{X_n}{X(\Omega)}$, which by construction sum to one across 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. $\hat{\theta}_k = \hat{\theta} = \frac{\sum_k J(k)\theta_k}{\sum_k J(k)}$.

Appendix H Additional Counterfactual Analysis

H.1 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)). The result is shown in I.16. I evaluate this policy in four models, including the full model (Column (1)), the model with no worker heterogeneity (Column (2)), the model with no firm heterogeneity (Column (3)), and the model with neither worker nor firm heterogeneity (Column (4)). The three limited or no heterogeneity models are re-estimated to fit the data, with the no heterogeneity model akin to the assumptions made in Hsieh and Moretti (2019).

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 commercial 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 4. Hence, the benefits from spatial reallocation are smaller than the benefits predicted by the simplified model with one-sided or no heterogeneity. The results of welfare changes align with this intuition, as well. This comparison, once again, emphasizes the importance of accounting for the heterogeneity of workers and firms when evaluating place-based policies.

H.2 Universal Basic Income (UBI)

There has been heated debate in the Canadian public policy sphere on the Universal Basic Income (UBI) transfer program. It has been advocated as an effective policy tool to reduce poverty and income inequality. There have been pilot programs implemented in various provinces, including Ontario and Quebec. I use the model to study the spatial consequences of a UBI transfer of $1,000 Canadian dollars to all individuals, which is about 2% of average earnings during the sample period. I assume that the transfers are financed by flat labor income taxes of all workers. The results are shown in Table I.17.

Two stark results stand out in Column (1). First, the UBI program goes a long way towards improving social welfare, by transferring from high-skill workers to low-income workers. Second, an unintended consequence of the program is that it will exacerbate spatial inequality with mobile workers and firms. This is because the unconditional transfer increases the purchasing power by more in low-rent, yet low-productive, cities. This income effect is stronger for low-skilled than high-skilled workers, thus incentivizing more low-skilled workers to relocate to these places. It can be seen from the changes in low- and high-skilled populations in low-wage cities in the table. Moreover, low-productive firms spatially follow low-skilled workers, which further worsens their employment opportunities. Such relocation causes more spatial skill segregation and income inequality. The aggregate and distributional effects are much smaller in the model without worker heterogeneity, as shown in Column (2).

The spatial consequences of UBI are much muted in limited sorting model scenarios, as shown in Columns (3)-(4). Without worker mobility, the policy has little effect on spatial inequality, while delivering its goal of decreasing income inequality. With only worker re-sorting, the spatial outcomes are also smaller compared to the full model. Fewer workers re-locate to low-wage cities and spatial inequality increases by less. The additional modeling features, shown in Columns (6)-(8), do not change the results by much.

Appendix I Additional Tables and Figures

Table I.1: Summary statistics

	2002-2009			2010-2017		
Sample	Baseline	Stayers	Movers	Baseline	Stayers	Movers
Number of Observations						
Worker-Years (in 1,000)	47815	4693	27502	51475	7504	23572
Workers (in 1,000)	9862	587	4698	10655	938	4127
Firms (in 1,000)	1479	12	1204	1588	16	1250
Cities	66	66	66	66	66	66
Worker Characteristics						
Mean Log. Earnings	10.38	10.79	10.32	10.44	10.85	10.29
Mean Age	41.17	43.50	40.25	41.68	43.84	39.96
Percent Male	56.5%	61.6%	56.2%	56.7%	56.8%	58.9%
Percent in Largest 5 cites	56.1%	58.8%	56.9%	56.7%	62.6%	56.8%
Firm characteristics						
Mean Log Wage bill per Worker	10.19	10.66	10.18	10.21	10.71	10.22
Mean Log Wage bill	11.10	14.74	11.33	11.07	14.85	11.35
Log Firm Size	0.91	4.08	1.15	0.86	4.14	1.13

Note: This table displays the summary statistics for the baseline sample, stayers sample, and movers sample for 2002-2009 and 2010–2017. See Section 2.1 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 I.2: City-size regressions of mean and dispersion of log earnings: 2002-2009

	Dependent variable:			
	Mean Log Earnings	Var. Log Earnings	90-50 Gap	50-10 Gap
	(1)	(2)	(3)	(4)
Log Population	0.035***	0.022***	0.033***	0.011**
	(0.008)	(0.005)	(0.004)	(0.005)
Constant	−0.448***	0.256***	0.319***	1.076***
	(0.107)	(0.052)	(0.048)	(0.064)
Observations	66	66	66	66
R^2	0.217	0.238	0.541	0.054

Note: This table displays the results of city-size regressions of city mean and dispersion measures of log earnings for 2002-2009. Log earnings are residualized by a cubic polynomial of age, gender, marital status, and the number of children using a Miner-type regression. I assume that the earnings profile is flag at age 40, following Card et al. (2013). The 90-50 gap is the difference between the 90^{th} and the 50^{th} percentile of log earnings in the city, and the 50-50 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 in 2002-2009. All regressions are weighted by population. *p<0.1; **p<0.05; ***p<0.01.

Table I.3: Earnings variance decomposition at individual and city levels: 2010-2017

	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 (2.2), with $k = 10$ firm clusters. The variance of city mean log earnings is decomposed according to: $\text{Var}(E_c[w_{it}]) = \text{Var}(\overline{a}_c) + \text{Var}(\overline{z}_c) + 2 \times \text{Cov}(\overline{z}_c, \overline{z}_c)$. 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 I.4: Alternative decomposition of urban earnings premium: 2010-2017

	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 in 2010–2017. The worker and firm FEs are estimated using the empirical specification with worker-firm interaction: $\log w_{it} = z_{j(i,t)} + \theta_{j(i,t)} a_i + \epsilon_{it}$. 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 for 2010-2017. All regressions are weighted by population. *p<0.1; **p<0.05; ***p<0.01.

Table I.5: Alternative decomposition of the city-size gradient of within-city earnings variance: 2010-2017

	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 for 2010-2017. All regressions are weighted by population. *p<0.1; **p<0.05; ***p<0.01.

Table I.6: 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 5.7 and (5.8). 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 I.7: 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. 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 I.6 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 I.8: Passthrough of wage bill shocks to firms

	OLS	IV	IV
	(1)	(2)	(3)
City total wage bill shock	1.06***	0.21***	0.15***
	(0.00)	(0.01)	(0.02)
Control 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 (5.9). 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 I.9: 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 I.10: 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 (5.15). 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 E.7. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table I.11: 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 (5.16). 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 E.8. All regressions are weighted by city population. *p<0.1; **p<0.05; ***p<0.01.

Table I.12: 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 I.13: 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 I.14: 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 I.15: Robustness checks on counterfactual analysis

	Counterfactual 1			Counterfactual 2		
	End. amen.	Free entry	Het. η	End. amen.	Free entry	Het. η
% 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: Robustness checks of the counterfactual analysis results. Counterfactual 1 refers to the subsidy to the productive firms in Toronto in Section 6.3.1; Counterfactual 2 refers to the low-wage city subsidy in Section 6.3.2. The details on the incorporation of endogenous amenities and free entry are discussed in Appendix F.1 and F.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 I.16: 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 highest-earnings cities.

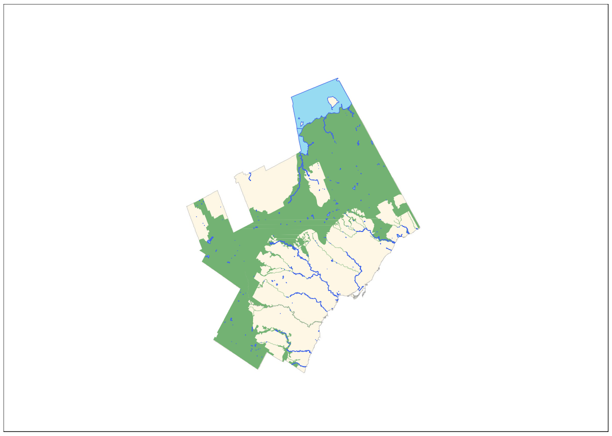
Table I.17: Counterfactual analysis: universal basic income

			Limited Re-sorting					
	Full model	No worker het.	No-resorting	Only firm	Only worker	End. amen.	Free entry	Het. η
% Changes in	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Var. city log earnings	24.8%	0.0%	-0.1%	-0.1%	16.5%	22.4%	24.9%	25.1%
City Gini index	7.5%	0.0%	-0.3%	-0.3%	4.2%	6.2%	7.6%	7.5%
Total output	-0.1%	0.0%	0.0%	0.0%	-0.1%	-0.1%	-0.1%	-0.1%
Total welfare	3.9%	0.8%	3.8%	3.8%	3.8%	4.1%	3.9%	3.8%
High-skill welfare	-1.4%	0.8%	-1.4%	-1.4%	-1.4%	-0.8%	-1.4%	-1.4%
Low-skill welfare	6.7%	0.8%	6.7%	6.7%	6.7%	6.7%	6.7%	6.6%
Pop. in low-wage cities	1.3%	0.6%	–	–	1.0%	1.2%	1.3%	1.2%
High-skill pop.	0.1%	0.6%	–	–	0.0%	-0.2%	0.1%	0.1%
Low-skill pop.	1.9%	0.6%	–	–	1.5%	1.9%	1.9%	1.8%
# Firms in low-wage cities	1.4%	0.5%	–	0.0%	–	1.5%	1.4%	1.4%
# high-prod. firms	-0.5%	0.5%	–	0.0%	–	-0.8%	-0.7%	-0.6%
# low-prod. firms	1.8%	0.5%	–	0.0%	–	1.9%	1.8%	1.7%

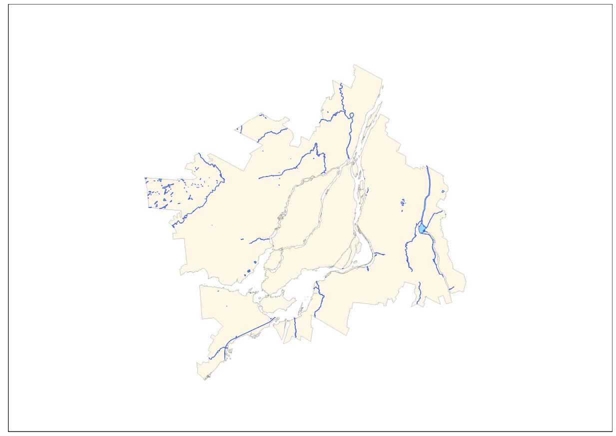
Note: Low-wage cities refer to the 25 lowest-paying cities, which account for 12% of the urban population.

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

(a) Toronto

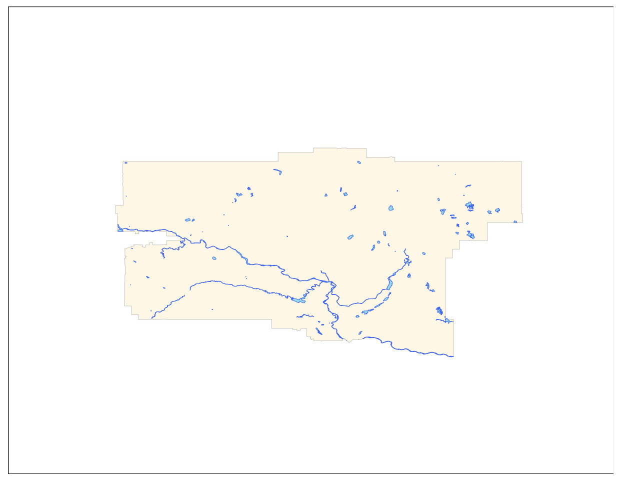


(b) Montreal



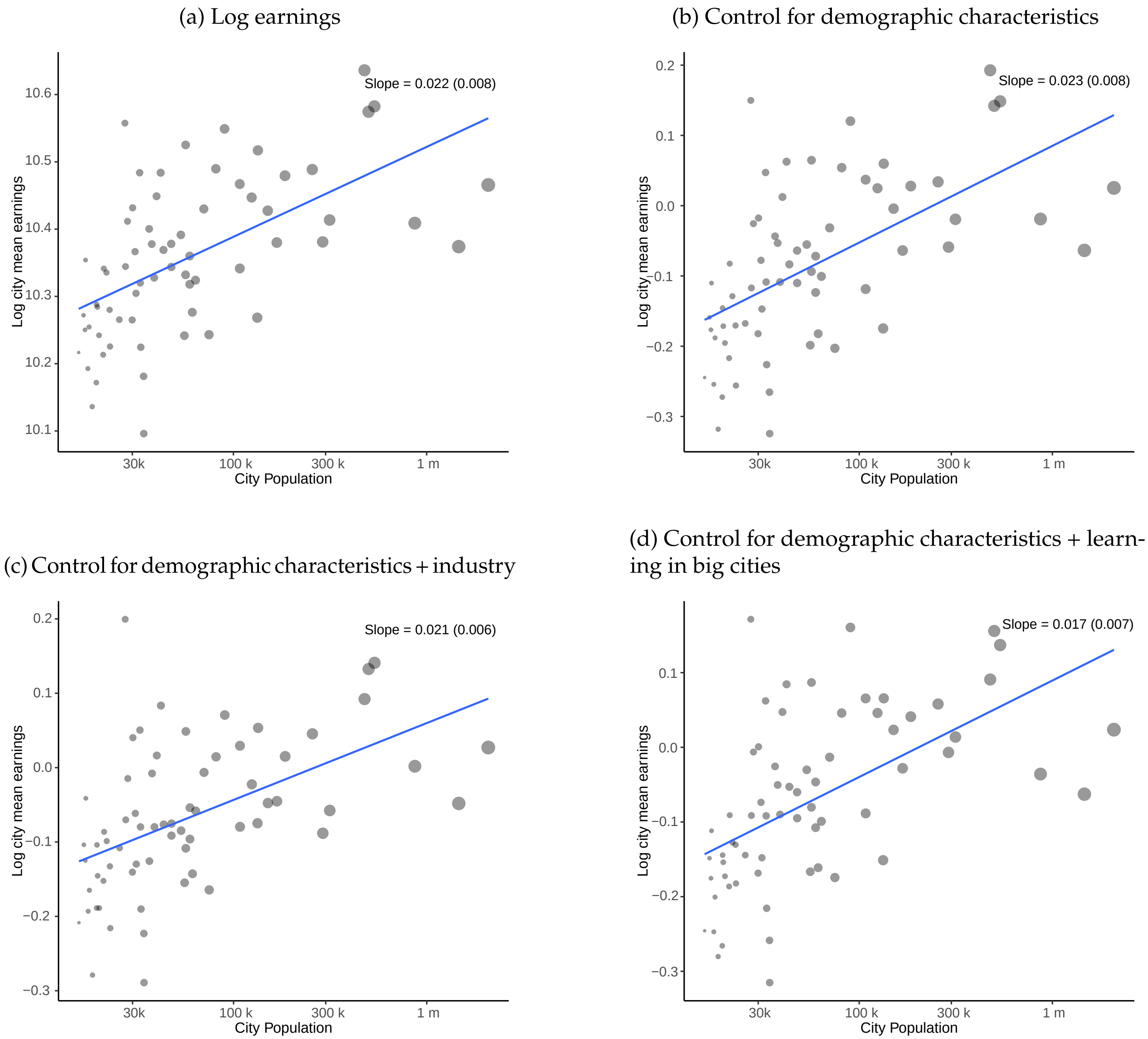
(c) Vancouver

(d) Calgary



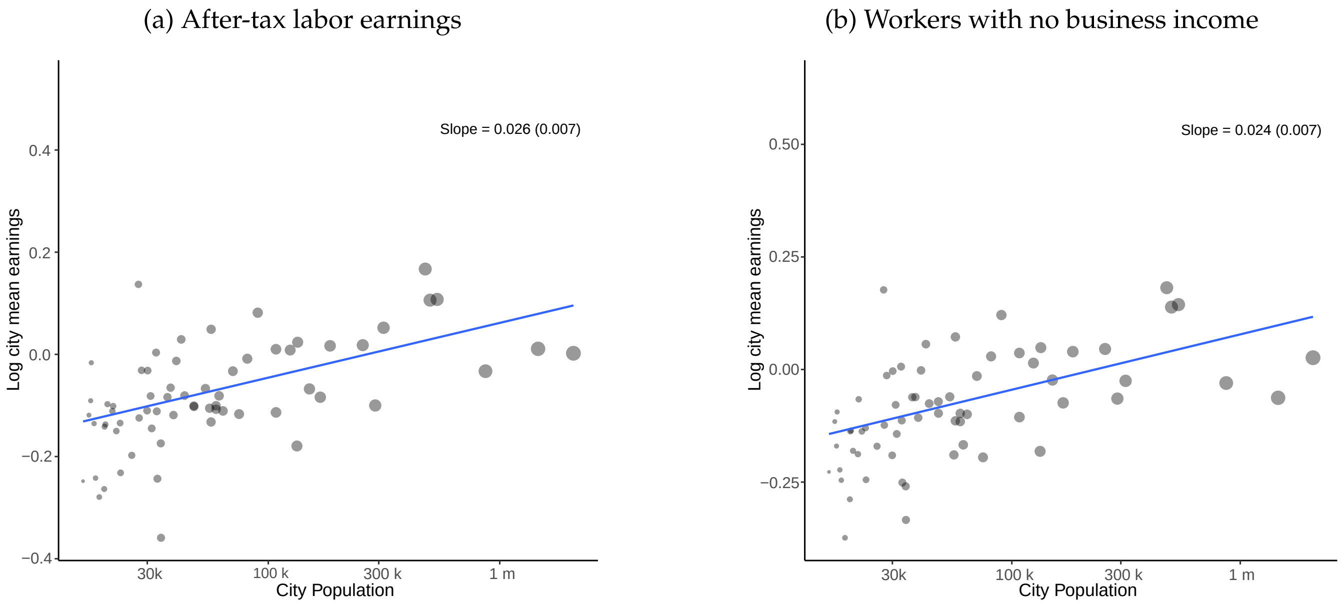
Note: Topography maps for the four largest Canadian cities. Blue areas are land areas covered by water, brown areas are land areas with slopes greater than 15 degrees, and green areas are land areas under the Greenbelt plan.

Figure I.2: City mean log earnings versus city population: 2010-2017



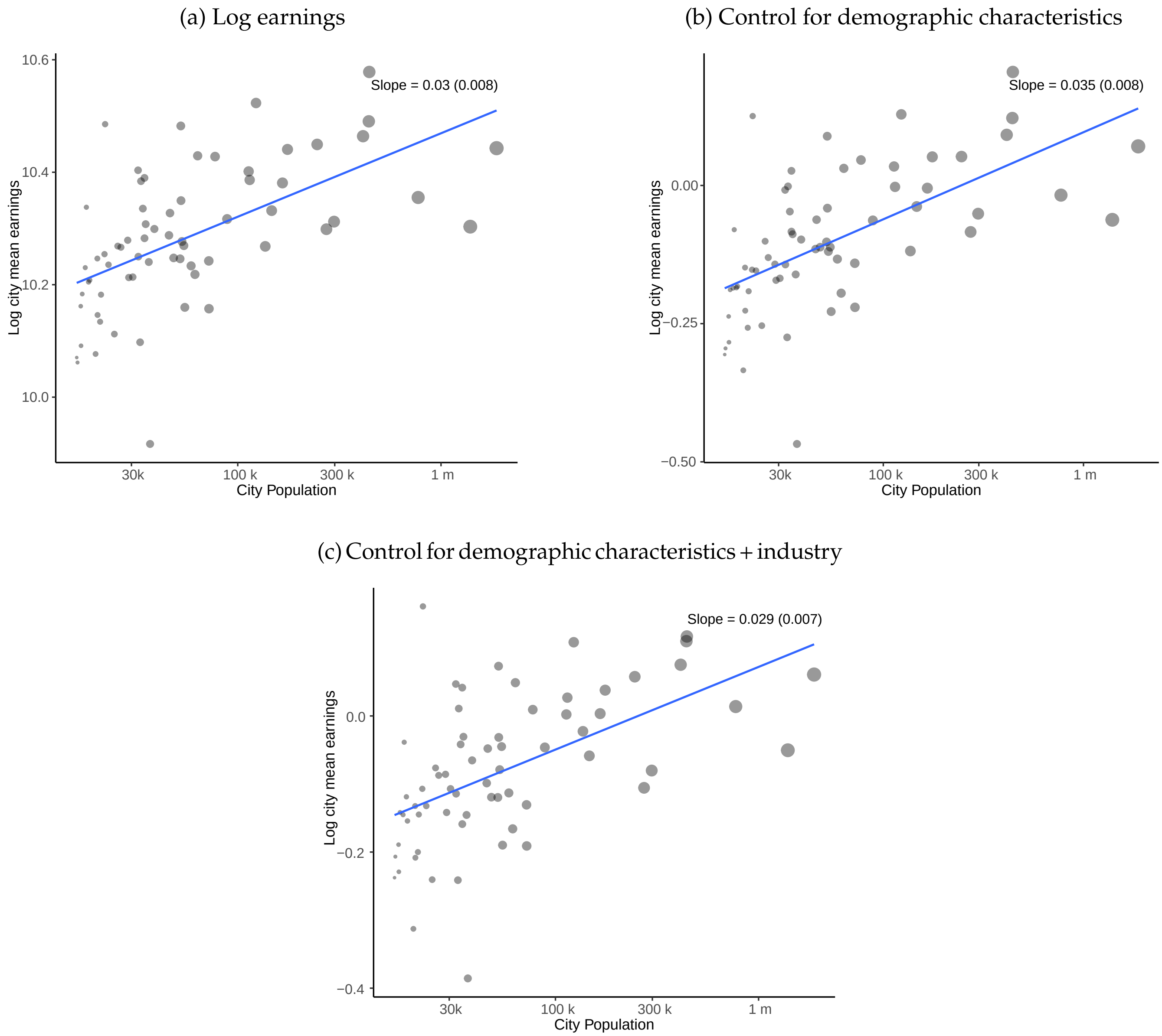
Note: These figures display results of city-size regressions of city mean log earnings for 2010-2017. Panel (a) uses raw log earnings; panel (b) controls for demographic characteristics including age profile, gender, marital status, and the number of children; panel (c) additionally controls for NAICS-4 industry dummies; panel (d) additionally controls for big-city work experiences since 2002 interacted with the current city being a big city. 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 I.3: City mean log earnings versus city population: 2010-2017



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.

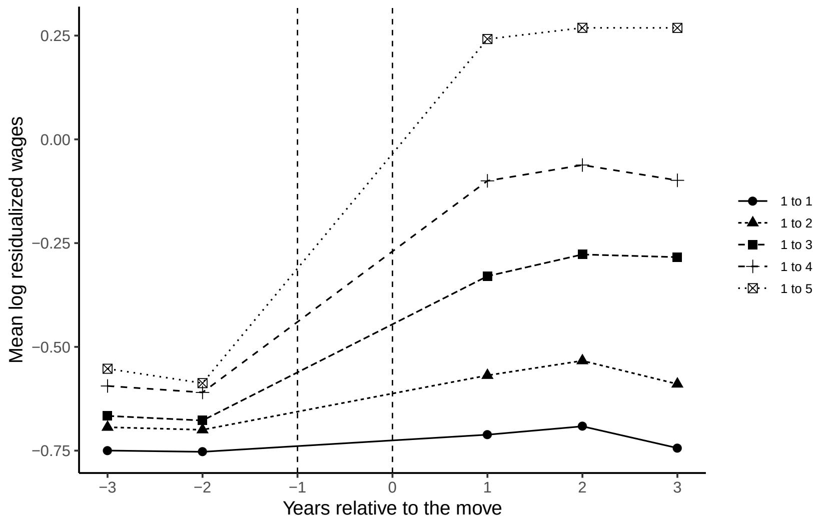
Figure I.4: City mean log earnings versus city population: 2002-2009



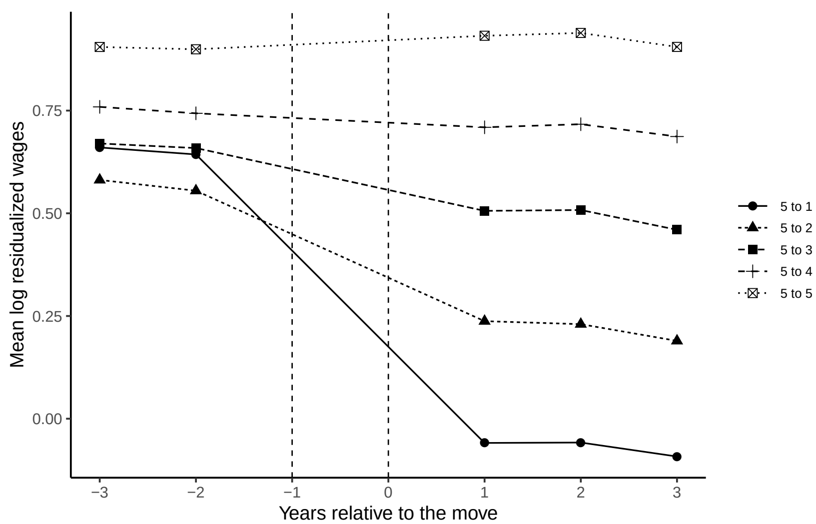
Note: These figures display results of city-size regressions of city mean log earnings for 2002-2009. Panel (a) uses raw log earnings; panel (b) controls for demographic characteristics including age profile, gender, marital status, and the number of children; panel (c) additionally controls for NAICS-4 industry dummies. Population is measured as the average number of full-time working individuals in each city in 2002-2009. Population-weighted regression coefficients and standard errors are reported.

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

(a) Movers from the bottom cluster

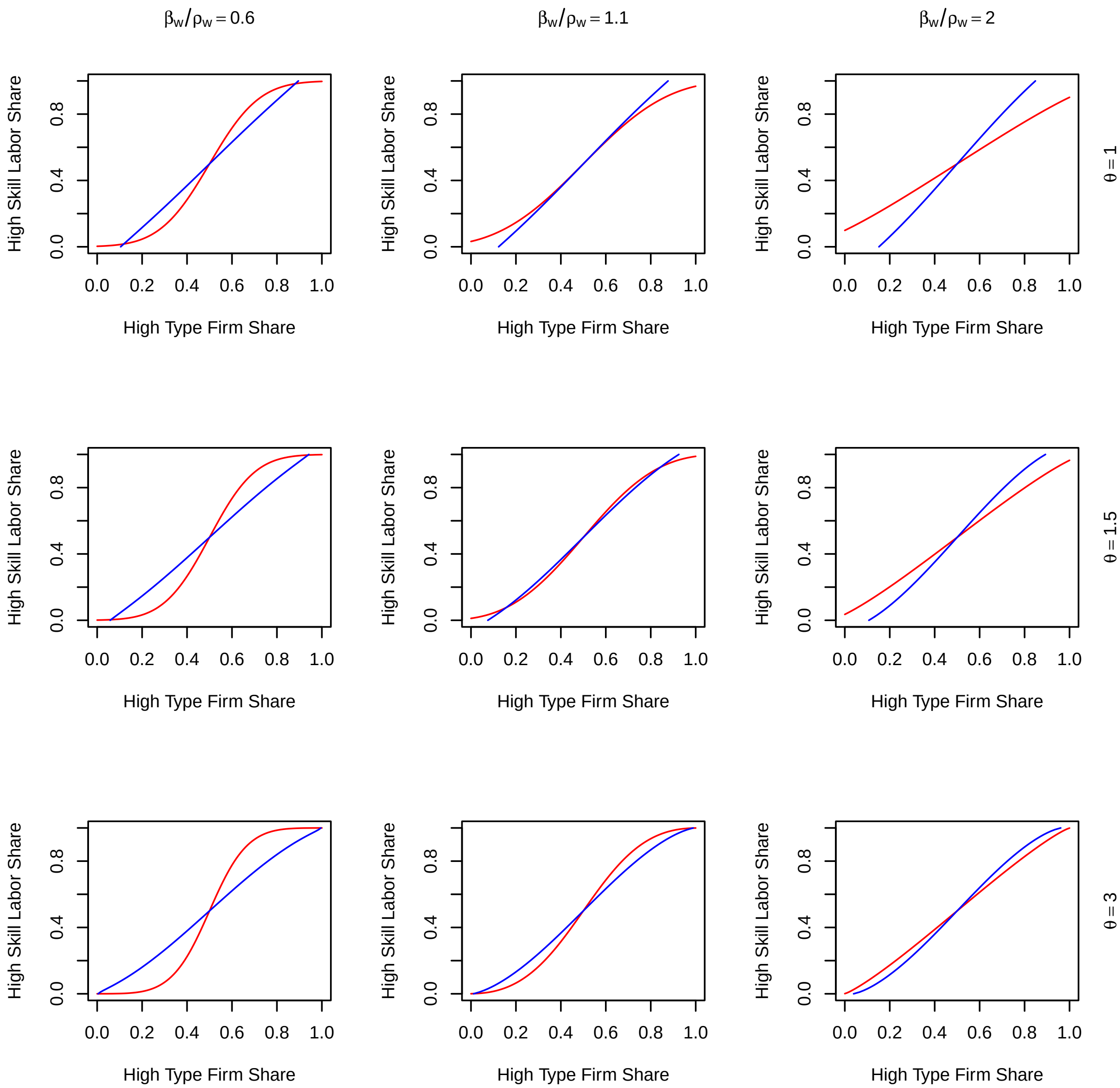


(b) Movers from the top cluster



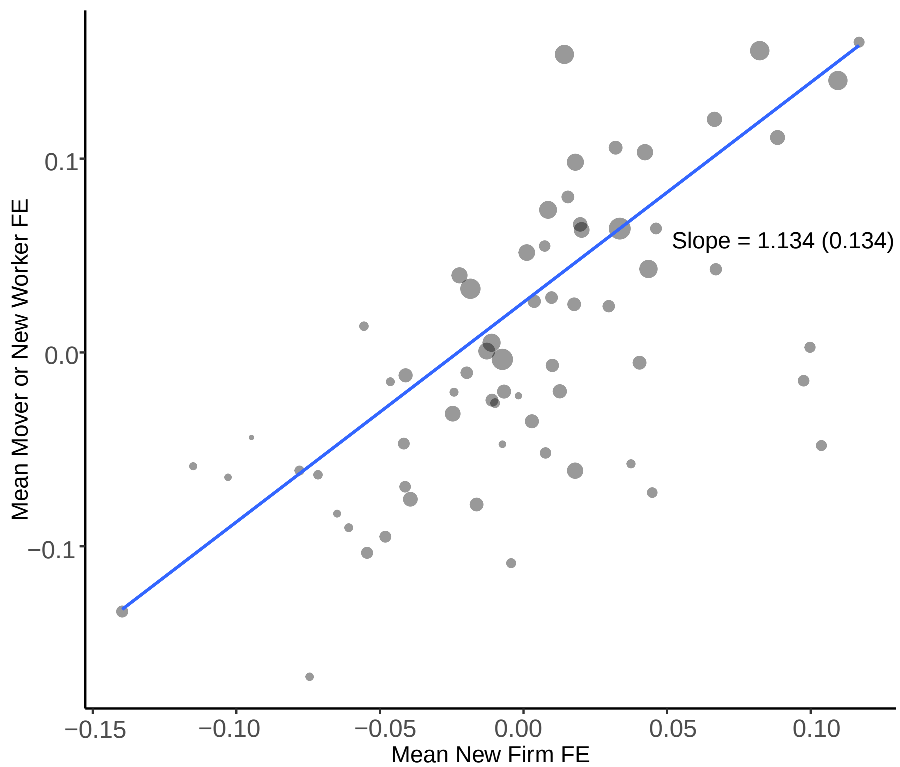
Note: These two figures display the earnings profile of between-firm worker movers before and after the move. The sample includes workers who have 1) moved between firms, 2) stayed in the previous firm for more than 4 years prior to the move and 2) stayed in the new firm for more than 3 years. The movement 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 I.6: Model simulation of two cities with various values of β_w/ρ_w and θ



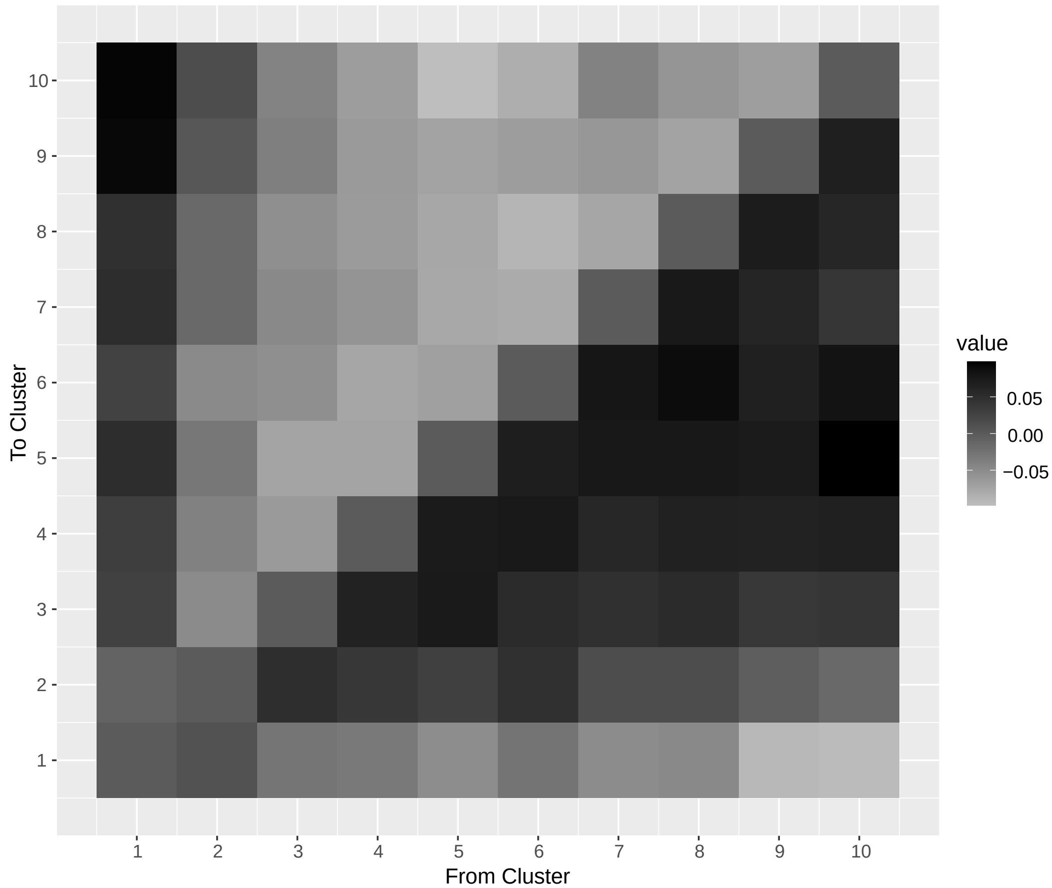
Note: Simulation results described in Section D. Red lines are the best response curves of the share of type-2 workers in city 1 given the share of type-2 firms in city 1. Blue lines are the best response curves of the share of type-2 firms in city 1 given the share of type-2 workers in city 1. The interactions of the red and blue curves are the Nash Equilibria.

Figure I.7: Mean new worker and firm FEs by city



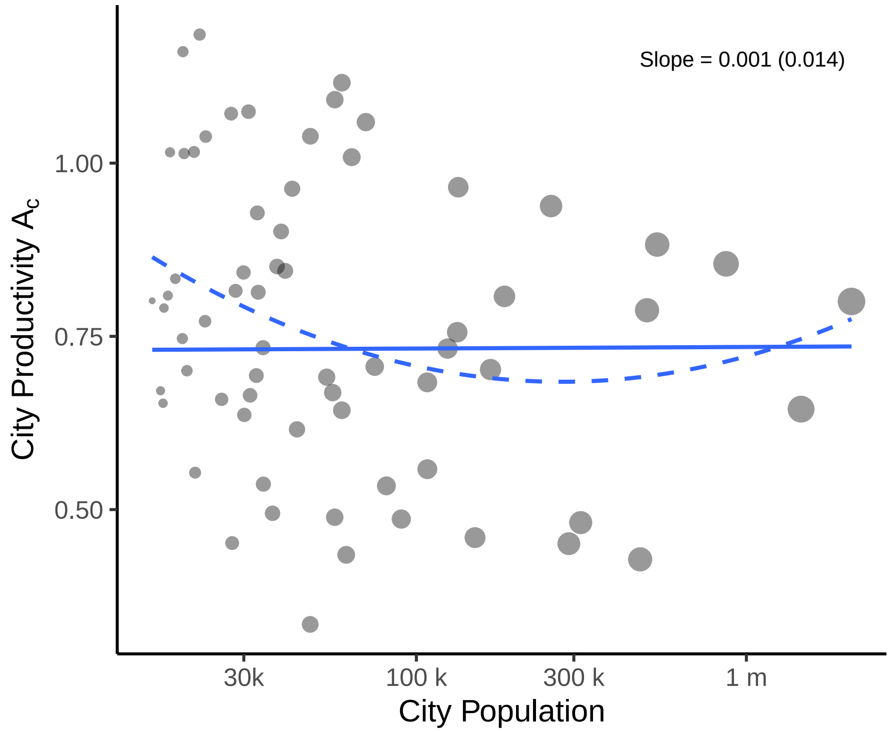
Note: This figure displays a scatterplot of the mean FEs of new or in-migration workers and of new firms for each city, in the 2010–2017 period. The worker and firm FEs are estimated using equation (2.2). Firms are grouped into $k = 10$ clusters. The population-weighted regression coefficient and the standard error are reported.

Figure I.8: 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 I.9: 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 I.10: 3D plots of firm and city amenities for different skilled workers

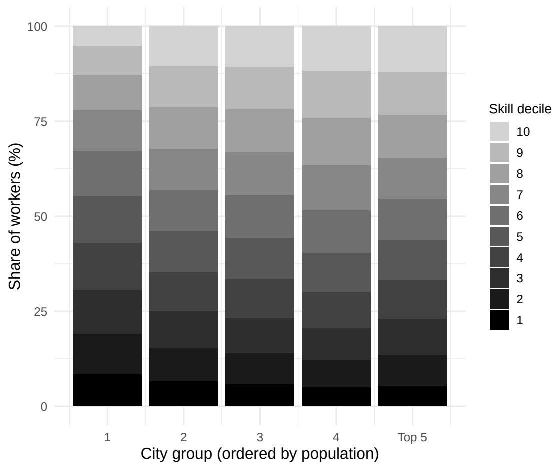
(a) Firm amenities

(b) City amenities

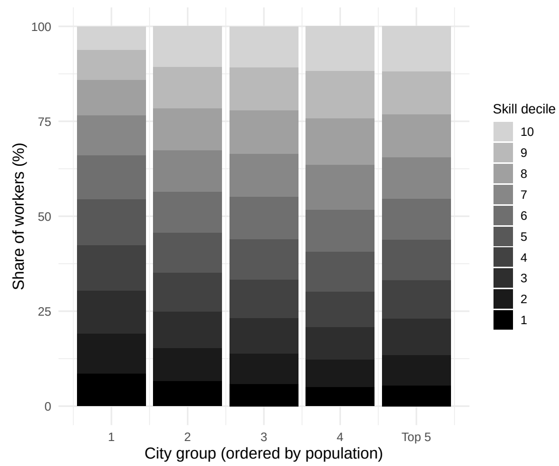
Note: This set of figures displays skill-specific firm and city amenities. Workers are binned into ten skill decile groups, and cities are ranked by population and binned into ten equal-number city groups. The amenities of firm cluster 1 and of city group 10 are normalized to 1, for all worker skill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

Figure I.11: Model fit of worker and firm sorting shares across cities

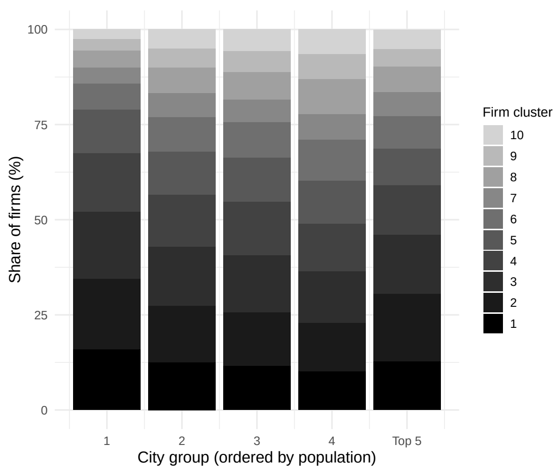
(a) Baseline Estimates: Worker



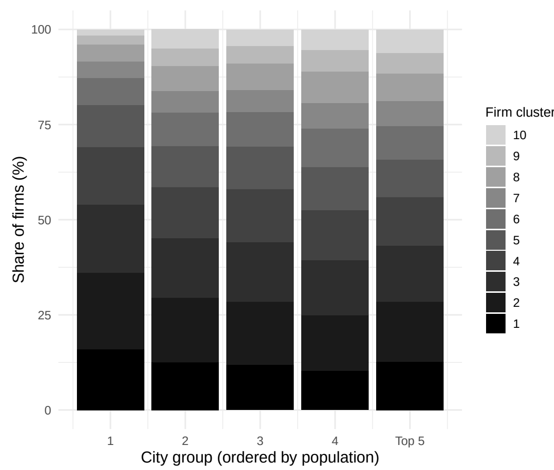
(b) Model Solution: Worker



(c) Baseline Estimates: Firm



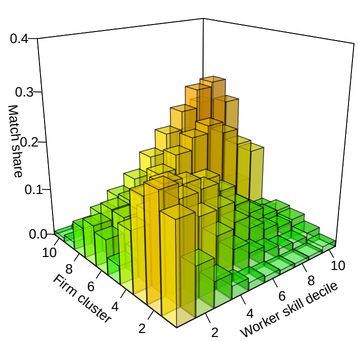
(d) Model Solution: Firm



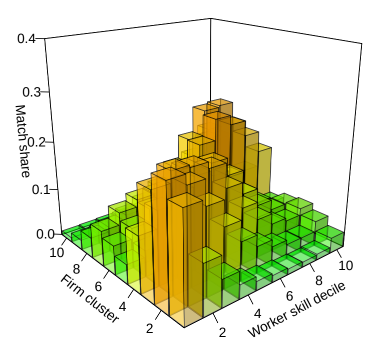
Note: The series of figures compare sorting shares of workers and firms generated by the model solution and the shares in the data: panels (a) and (b) compare worker sorting shares, and panels (c) and (d) compare firm sorting shares. Cities are binned into 5 groups: one is the largest five-cities group and the other four are equal numbers of cities ranked by population. I classify workers into ten deciles based on their skill and rank firm clusters based on common productivity z.

Figure I.12: Model fit of worker and firm match shares within cities

(a) Baseline Estimates: Largest 5 Cities



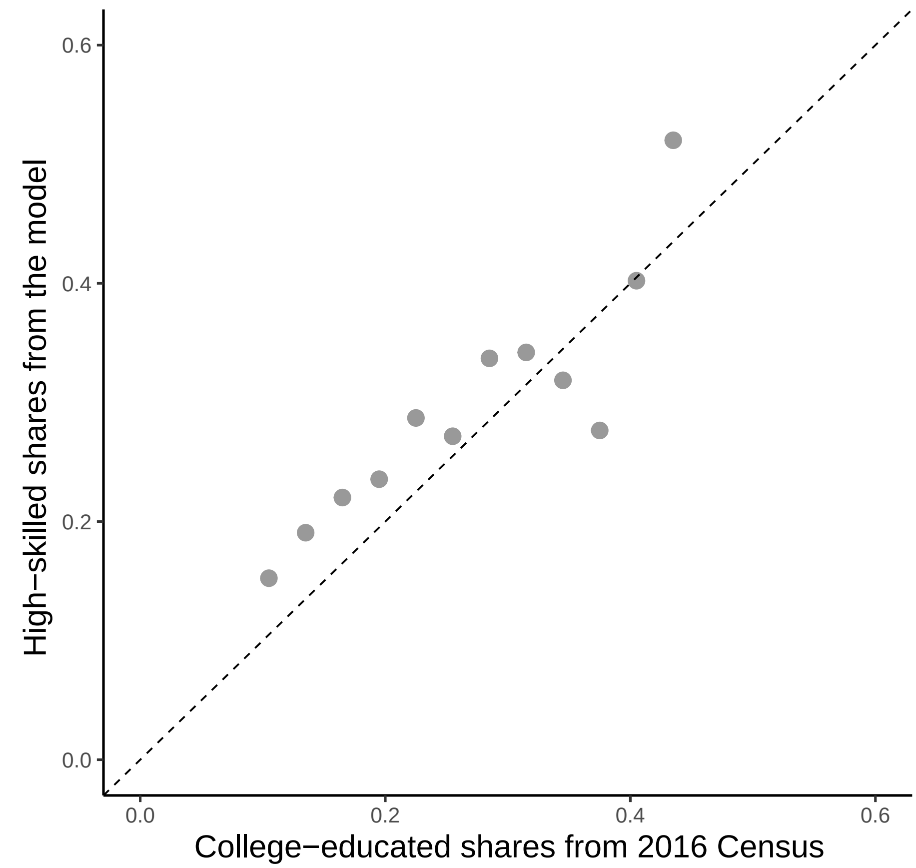
(b) Model Solution: Largest 5 Cities



(c) Baseline Estimates: Smaller Cities

(d) Model Solution: Smaller Cities

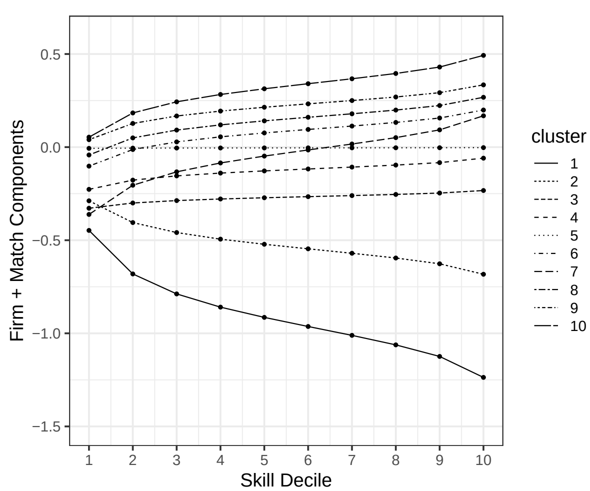
Figure I.13: High-skilled worker shares in the model versus college-educated shares from 2016 Census



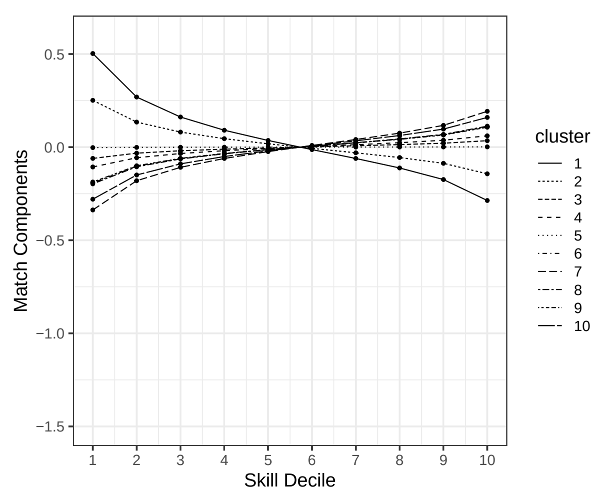
Note: This figure is a binscatter plot of the model-estimated city high-skilled worker shares versus the city college-educated shares constructed using 2016 Population Census. High-skilled workers are defined as workers with skills in the top three deciles of the skill distribution. The 45-degree line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

Figure I.14: Understanding sorting and worker-firm complementarity

(a) Firm+Interaction Components

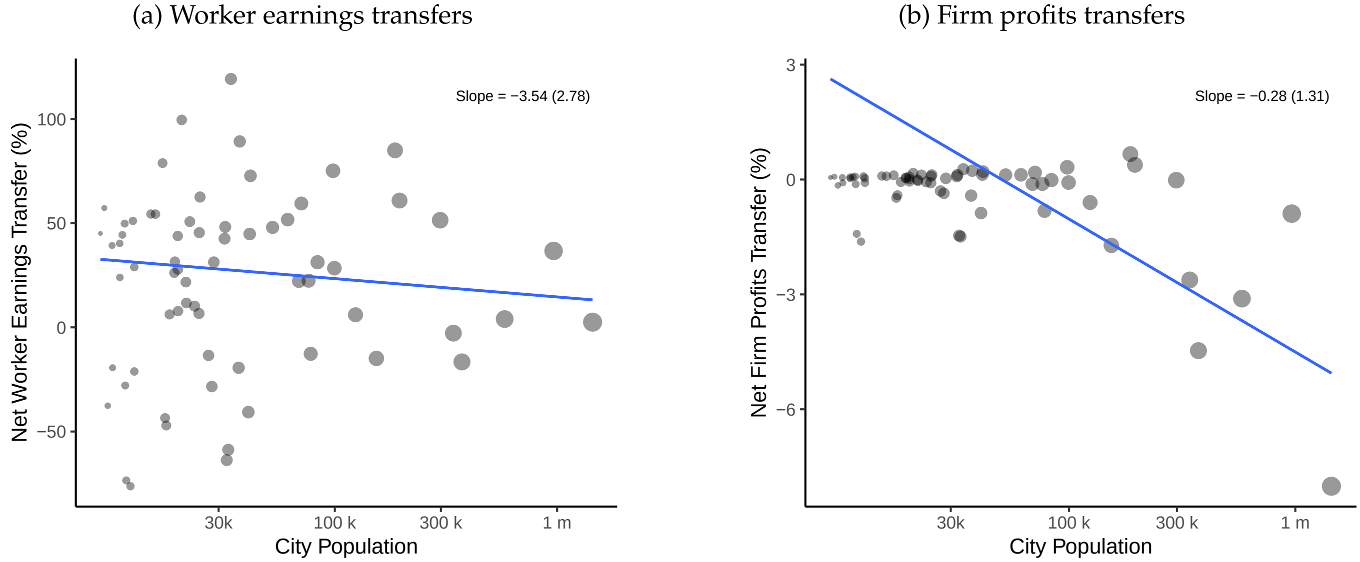


(b) Interaction Components



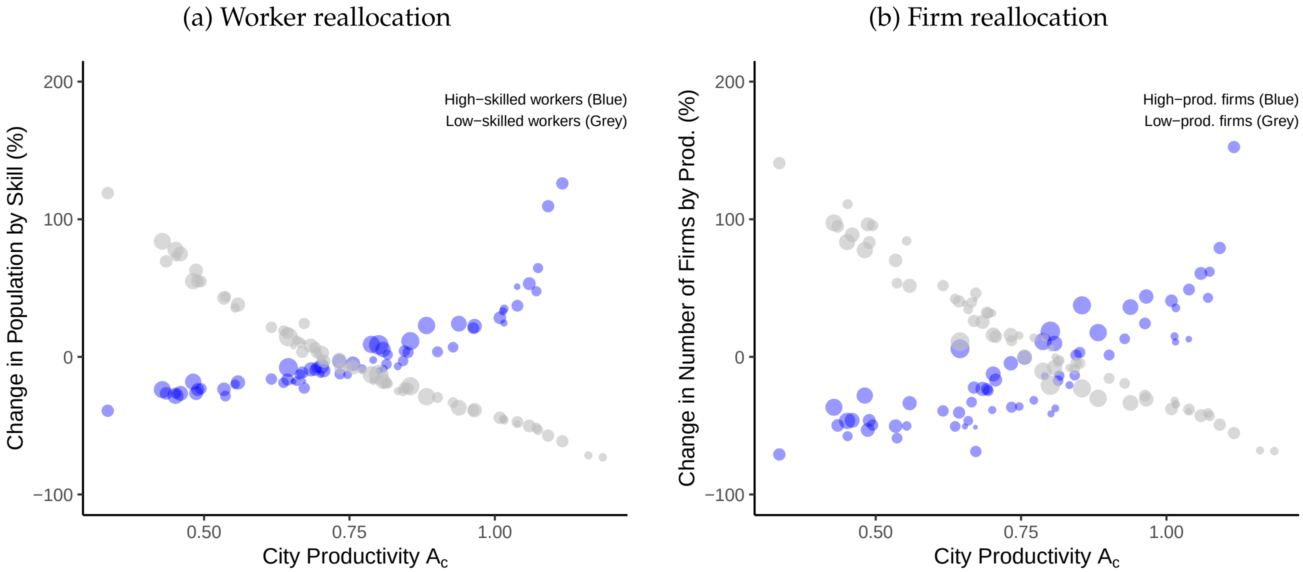
Note: The two figures plot 1) the firm+interaction components and 2) the interaction component in equation (6.1) for ten firm clusters along different skill deciles. Firm clusters are ranked by productivity z.

Figure I.15: Net transfers versus city population



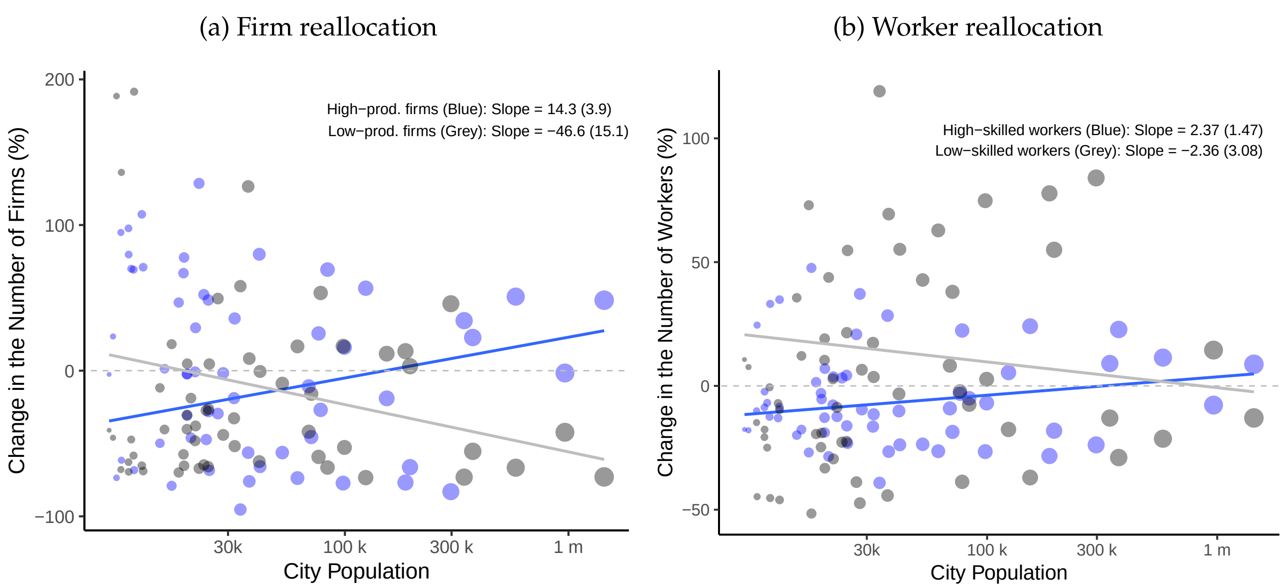
Note: These figures relate net earnings transfers by city to city population. For each city, the net worker earnings transfer is measured as the ratio of the net worker transfers received to total worker earnings; the net firm profits transfer is measured as the ratio of the net firm transfers received to total firm profits. Population-weighted OLS regression coefficients and standard errors are reported.

Figure I.16: Spatial reallocation of workers and firms of the optimal policies between cities with different exogenous productivities



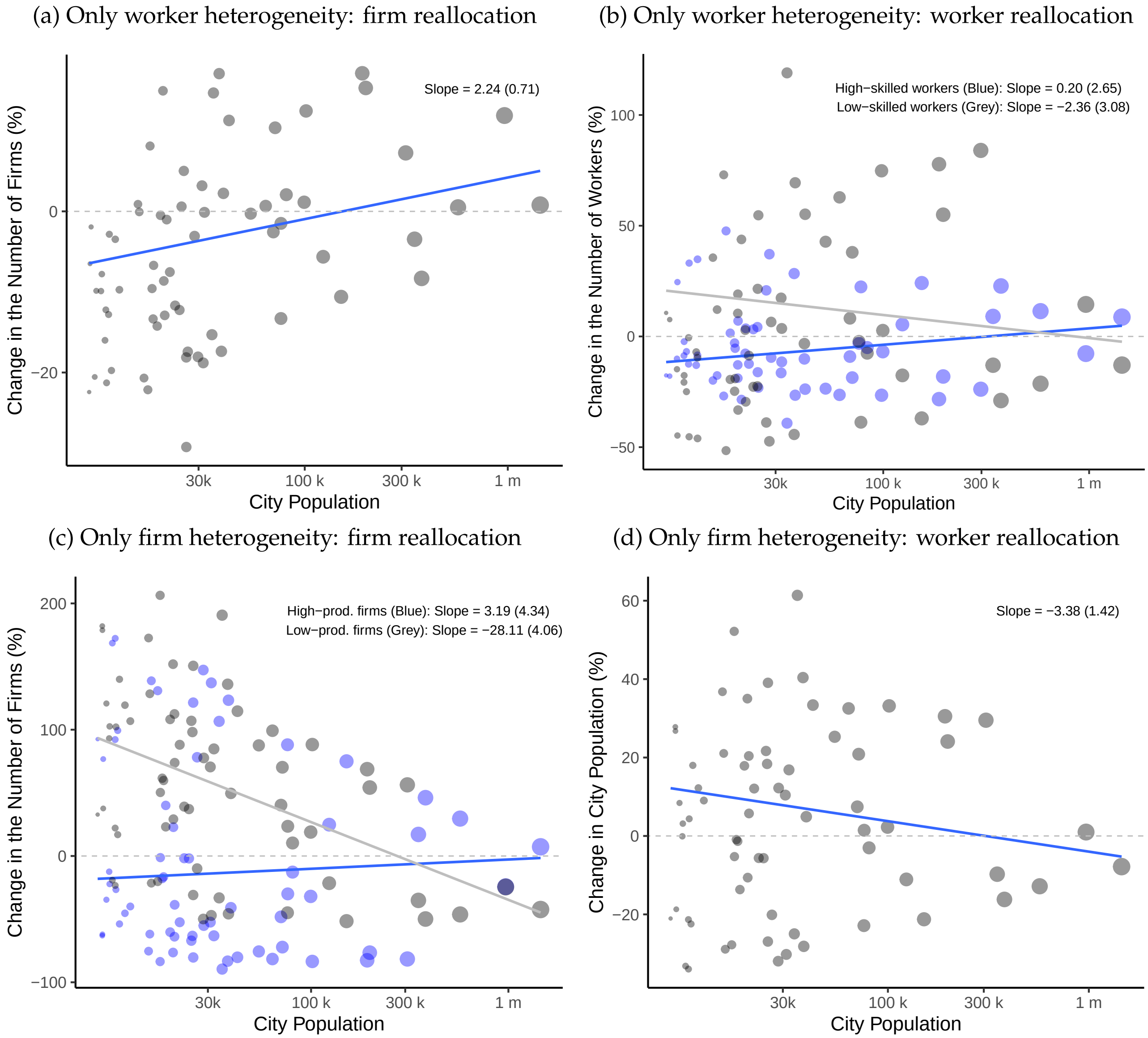
Note: The series of figures displays the changes in the number of workers and firms induced by the optimal policy versus city exogenous productivity A_c. In Panel (a), blue dots represent high-skilled workers and grey dots represent low-skilled workers; in Panel (b) blue dots represent high-productivity firms and grey dots represent low-productivity firms.

Figure I.17: Spatial reallocation of workers and firms of the optimal policies with no within-type redistribution



Note: The series of figures displays the reallocation of workers and firms induced by the optimal policy with no within-type redistribution. Specifically, I design type-city-specific proportional taxes to equate the income of all agents to their social values, which are state in Proposition 2.

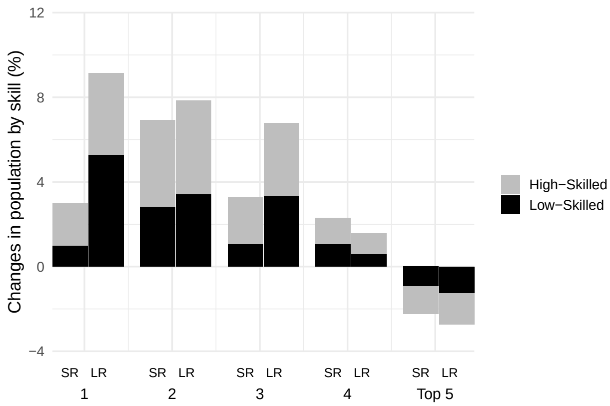
Figure I.18: Spatial reallocation of workers and firms of the optimal policies considering only one-sided heterogeneity



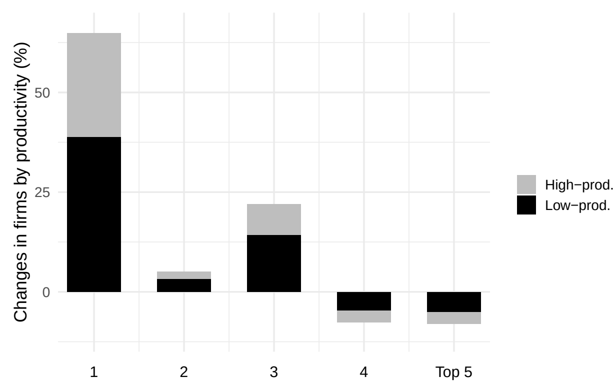
Note: The series of figures displays the reallocation of workers and firms induced by optimal policies that consider only one-sided heterogeneity.

Figure I.19: Changes in the spatial distributions of workers and firms with remote work

(a) Number of workers



(b) Number of firms



Note: The two figures display the changes in the number of workers and firms when 10% of workers can work remotely. Cities are binned into 5 groups: one is the largest five-cities group and the other four are equal numbers of cities ranked by population. In panel (a), SR refers to the short-run model where only workers can change locations; LR refers to the long-run model where both firms and workers can change locations. Panel (b) only has the results for the long-run model as firms do not move in the short-run model. I show the results separately for high- and low-skilled workers, and high- and low-productive firms.