NBER WORKING PAPER SERIES

OPTIMAL SPATIAL POLICIES, GEOGRAPHY AND SORTING

Pablo Fajgelbaum
Cecile Gaubert

Working Paper 24632
http://www.nber.org/papers/w24632

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2018, Revised June 2019

We thank Arnaud Costinot, Rebecca Diamond, Jonathan Dingel, Robert Staiger, Costas Arkolakis, and Adrien Bilal for their conference discussions. We also thank David Atkin, Lorenzo Caliendo, Stephen Redding and Frederic Robert-Nicoud for helpful comments. We thank Sam Leone and Wan Zhang for excellent research assistance. Cecile Gaubert thanks the Clausen Center for International Business and Policy and the Fisher Center for Real Estate and Urban Economics at UC Berkeley for financial support. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Optimal Spatial Policies, Geography and Sorting
Pablo Fajgelbaum and Cecile Gaubert
NBER Working Paper No. 24632
May 2018, Revised June 2019
JEL No. F12,H21,H71,R13

ABSTRACT

We study optimal spatial policies in a quantitative trade and geography framework with spillovers and spatial sorting of heterogeneous workers. We characterize the spatial transfers that must hold in efficient allocations, as well as labor subsidies that can implement them. There exists scope for welfare-enhancing spatial policies even when spillovers are common across locations. Using data on U.S. cities and existing estimates of the spillover elasticities, we find that the U.S. economy would benefit from a reallocation of workers to currently low-wage cities and from a greater mixing of high and low skill workers in these locations. Inefficient sorting may lead to substantial welfare costs.

Pablo Fajgelbaum
Department of Economics
University of California, Los Angeles
Bunche Hall 8283
315 Portola Plaza
Los Angeles, CA 90095
and NBER
pfajgelbaum@gmail.com

Cecile Gaubert
Department of Economics
University of California, Berkeley
687 Evans Hall
Berkeley, CA 94720
and NBER
cecile.gaubert@berkeley.edu

1 Introduction

A long tradition in economics argues that the concentration of economic activity leads to spillovers. For instance, dense cities are more productive thanks to agglomeration economies, but are also more congested. These spillovers shape the distribution of city size and productivity. Groups of workers with different skills arguably vary in how much they contribute to these spillovers and in how much they are impacted by them, so that these forces also shape how heterogeneous workers sort across cities. Being external in nature, spillovers likely lead to inefficient spatial outcomes. In this paper, we ask: is the observed spatial distribution of economic activity inefficient? If so, what policies would restore efficiency and what would be their welfare impact? Would an optimal spatial distribution feature stronger, or weaker, spatial disparities and sorting by skill than what is observed?

To answer these questions, we develop and implement a new approach. Our framework nests two recent strands of general-equilibrium spatial research with spillovers: location choice models in the tradition of Rosen (1979)-Roback (1982) with sorting of heterogeneous workers as in Diamond (2016), and economic geography models in the tradition of Helpman (1998) applied to quantitative setups as in Allen and Arkolakis (2014) and Redding (2016). Crucially, we generalize these models to allow for arbitrary transfers across agents and regions. We characterize the set of transfers needed to attain first-best allocations, alongside the labor income subsidies that would implement them. We then combine the framework with data across metropolitan statistical areas (MSAs) in the United States, and evaluate quantitatively the impact of implementing optimal spatial policies on sorting by skill, wage inequality, and welfare. Under existing estimates of the spillover elasticities, the results suggest that inefficient sorting may lead to substantial welfare costs, and that spatial efficiency calls for more redistribution to low-wage cities and more skill mixing in these locations.

The framework incorporates many key determinants of the spatial distribution of economic activity. Firms produce differentiated tradeable commodities and non-tradeables using labor, intermediate inputs, and land. Locations may differ in fundamental components of productivity and amenities, bilateral trade frictions, and housing supply elasticities. Productivity and amenities are endogenous through agglomeration and congestion spillovers that may depend on the composition of the workforce.[1] Different types of workers vary in how productive they are in each location, in their ownership of fixed factors such as land, in their preference for each location, and in the efficiency and amenity spillovers they generate on other workers. In the market allocation, government policies may redistribute income across agents and regions.[2]

[1]As summarized by Duranton and Puga (2004), efficiency spillovers may result from several forces such as knowledge externalities, labor market pooling, or scale economies in the production of tradeable commodities. Amenity spillovers may result from congestion through traffic or environmental factors such as noise or pollution; availability of public services such as education, health, and public transport; availability of public amenities such as parks and recreation; or specialization thanks to scale effects in the provision of urban amenities such as restaurants or entertainment.

[2]A wide range of government policies lead to spatial transfers. Some of these are explicit "place-based policies", such as tax relief schemes targeted at distressed areas (e.g. New Markets Tax Credit, or Enterprise Zones) or direct public investment in specific areas (e.g. Tennessee Valley Authority). Other policies are not explicitly spatial, but

In the model, the spillovers have complex general-equilibrium ramifications through factor mobility and trade linkages. However, in the spirit of the "principle of targeting" pointed out by Dixit (1985), the first-best allocation can be implemented by policies acting only upon inefficient margins. Here, these margins consist of labor supply and demand decisions: workers do not internalize the impact of their location choice on city-level amenities, and firms do not internalize the impact of their hiring decisions on city-level productivity. We derive a necessary efficiency condition on the joint distribution of expenditures, wages and employment across worker types and regions. Using this condition we then characterize the transfers that must hold in an efficient allocation. Furthermore, we identify a condition on the distributions of spillover and housing supply elasticities under which these optimal transfers are also sufficient to implement the efficient allocation.

This characterization generalizes the standard efficiency requirement from non-spatial environments such as Hsieh and Klenow (2009), whereby the marginal product of labor should be equalized across productive units. Here, the optimal spatial allocation balances the net benefit of spillovers (in production or amenities) against the opportunity cost of attracting workers to each location. Because the location and consumption decisions are not separable, these opportunity costs are measured in terms of local consumption expenditures, and they vary across locations due to the compensating differentials born of geographic forces (congestion in housing, amenities, trade costs, and non-traded goods). Therefore, determining whether an observed allocation is efficient and whether specific cities are too large requires information about expenditure per capita across locations, in addition to the standard requirement of observing wages and employment.

We characterize the policies that lead to optimal transfers in special cases. We first apply the results to a case where the elasticities of spillovers (in both amenities and productivity) are constant with respect to population and identical across cities. Studies of place-based policies such as Glaeser and Gottlieb (2008) and Kline and Moretti (2014a) suggest that, in this environment, there are no gains from implementing policies that reallocate workers.[3] We show that this prevailing view relies on assuming away policies that redistribute income across space. When transfers are allowed, the *laissez-faire* allocation without transfers is inefficient even under constant-elasticity spillovers that are identical across locations, as long as there are compensating differentials across regions (such as differences in amenities). Intuitively, starting from an equilibrium without transfers, differences in marginal utility of consumption lead to gains from transferring tradeable goods. These transfers in turn incentivize workers to move, leading to gains from reallocation. Under these assumptions, labor income subsidies that are constant over space restore efficiency.

We apply our results to establish the normative properties of well-known economic geography models corresponding to special cases of our framework with inelastic housing supply, a single worker type, constant elasticity spillovers and no intermediate inputs. In this context, global efficiency is characterized by the distribution of trade imbalances between regions. This distribution can be

end up redistributing income to specific places (e.g., nominal income taxes and credits, state and local tax deductions, or sectoral subsidies). Neumark et al. (2015) and Austin et al. (2018) review some of these policies.

[3]This view is echoed in literature reviews of the place-based policy literature, such as Kline and Moretti (2014b), Neumark et al. (2015) and Duranton and Venables (2018).

implemented by a simple transfer rule that is independent from the distribution of fundamentals or trade costs. We show that, because these models make different assumptions about transfers in the *laissez-faire* allocation, they have different implications for whether the optimal government intervention should redistribute income from high- to low-wage regions, or the reverse.

In the more general case with asymmetric spillovers, allocations without transfers are still generically inefficient. In addition to the forces described in the case with homogeneous workers, there are also gains from reallocating workers that generate positive spillovers to places where they are more scarce. Thus, inefficient sorting creates an additional rationale for spatial transfers and reallocation, and spatial variation in policies is needed to attain efficient sorting. For example, if low skill workers benefit in terms of productivity from high skill workers, the decentralized pattern of sorting by skill may be too strong. The optimal subsidies then increase the degree of mixing across locations relative to the competitive allocation.

Our theoretical analysis complements a body of research on optimal city sizes following Henderson (1974) that typically assumes homogeneous workers and limited heterogeneity across locations.[4] Helsley and Strange (2014) characterize properties of the optimal sorting with heterogeneous workers and spillovers, under the assumptions of homogeneous locations. We make progress by studying an optimal allocation with transfers of tradeable output, in an environment with several dimensions of spatial heterogeneity and various sources of spillovers across heterogeneous workers. A key feature of our approach is to provide a simple characterization of efficiency in terms of the expenditure distribution. Being only function of observable variables and elasticities, this conditions allows us to characterize optimal policies despite the generality of the underlying framework, and to determine the set of statistics in the data that suffices to numerically compute the optimal allocation.

We also show how to extend this approach to settings with richer spillovers, such as environments with cross-location spillovers in the spirit of Desmet and Rossi-Hansberg (2014) or with commuting as in Monte et al. (2018). In the latter, individuals decide both where to work (subject to productivity spillovers) and where to live (subject to amenity spillovers). We find that with only constant-elasticity productivity spillovers, optimal policies are identical to our benchmark case without commuting. When both amenity and productivity spillovers are present, the first-best policies combine two location-specific transfers, one varying by residence and the other by workplace.

We quantify the model using data on the distribution of economic activity across MSAs in the United States. A key motivation for our application is the well known empirical evidence that larger cities in the U.S. feature higher wages, higher share of skilled workers, and higher skill premium, as documented among others by Behrens and Robert-Nicoud (2015). Moretti (2012) points out a "great divergence" in these outcomes over the last decades. We ask whether, in the presence

[4]Flatters et al. (1974), and Helpman and Pines (1980) are early studies of optimal city sizes in models with heterogeneous cities in either amenity or productivity. See Abdel-Rahman and Anas (2004) for a review. More recent studies include Albouy et al. (2019) and Eeckhout and Guner (2017). A focus in some of these papers is to study the extensive margin of city creation. We abstract from studying this margin, and take the number of potentially populated locations as given.

of spillovers, these observed patterns of spatial disparities in the U.S. are too strong from the perspective of spatial efficiency.[5]

In our benchmark analysis we allow for two skills groups, high skill (college) and low skill (non college) workers. We combine data on labor and non-labor income, taxes and transfers at the city level from the BEA, with Census data that allows us to break down these MSA-level totals by skill group within cities. To parametrize the spillover elasticities we rely on existing estimates in the U.S. based on spillover equations that are consistent with our model. We draw the amenity spillovers and the heterogeneity in spillovers across workers from Diamond (2016), and the city-level elasticity of labor productivity with respect to employment density from Ciccone and Hall (1996).

The quantification yields welfare gains of roughly 2% to 6% across a range of specifications of the spillover elasticities. With homogeneous workers the welfare gains are negligible, suggesting that inefficient sorting drives the welfare costs. We find similar welfare gains across alternative quantifications that incorporate three groups of skill, migration frictions based on worker's region of birth, and land regulations. We find that the distortions caused by land regulations may be quantitatively as important as those caused by inefficient sorting due to spillovers.

These welfare gains are achieved by increasing income redistribution towards low-wage cities. The optimal transfers can be implemented via higher labor income taxes in high-wage cities. In the case of low skill workers, the higher taxes in high-wage cities arise because these workers generate congestion and small productivity spillovers. In contrast, for high skill workers, they arise because these workers generate positive spillovers onto low skill workers, who are more prevalent in low-wage cities. This second force offsets the strong positive spillovers that high skill workers generate among each other, which would call for a subsidy in high-wage cities.

Due to the reallocations induced by these transfers, low-wage and less skill intensive cities grow and see an increase in the share of high-skill workers. At the other end of the city-size distribution, the largest and more skill intensive cities tend to shrink. However, these cities also increase their skill share while experiencing a reduction in wage inequality. Hence, the optimal sorting of high skill workers into smaller cities exploits their cross-productivity spillovers on low-skill workers, while their concentration in some of the largest cities reflects the positive amenity spillovers they generate on their own group.

To further identify the key spillovers driving these results, we assume that the observed equilibrium is efficient and use our optimal-transfers formulas to infer the spillover elasticities that best rationalize the data. This procedure yields negative amenity spillovers of similar magnitude for both skill groups, whereas the existing estimates used in the calibration imply that low-skill and high-skill workers generate spillovers of opposite signs. In this sense, we identify a key role for the heterogeneous amenity spillovers across skill types.[6]

[5] Recent papers such as Eeckhout et al. (2014), Behrens et al. (2014), and Davis and Dingel (2012) include spatial sorting of heterogeneous individuals to rationalize some of these patterns.

[6] In our parametrization, these spillovers rely on numbers from Diamond (2016), who estimates a positive response of an urban amenity index (including congestion in transport, crime, environmental indicators, supply per capita of different public services, and variety of retail stores) to the relative supply of college workers, as well as a higher marginal valuation for these amenities for college than for non-college workers.

The rest of the paper is structured as follows. Section 2 presents a stylized model to drive intuition, then presents the general environment. Section 3 characterizes the optimal policies, teases out their implications in specific cases of the theory corresponding to the models from the literature, and determines the data that suffices to implement the model. Section 4 describes the data and the calibration. Section 5 presents the quantitative implementation and Section 6 concludes. Proofs, additional derivations and data construction are detailed in the appendix.

2 Economic Geography Model with Worker Sorting and Spillovers

2.1 A Simple Example with Homogeneous Workers

We start with a simple case nested in the environment we detail next. We use this case to show that, starting from a market allocation without policies, there are gains from reallocating workers across space. This is true even under identical and constant elasticity spillovers across space.

Suppose that workers are homogeneous and that utility per worker in a location j equals $u_j = a_j c_j$, where a_j is city-level amenities and c_j is consumption of tradeable output. Amenities take the form $a_j = A_j L_j^{\gamma_A}$, where A_j is exogenous and $L_j^{\gamma_A}$ is a spillover that depends on the population L_j of j with constant elasticity γ_A. Similarly, output per worker $z_j = Z_j L_j^{\gamma_P}$ depends on exogenous productivity Z_j and on agglomeration economies governed by the constant elasticity γ_P.

An approach in the placed-based policy literature, such as Glaeser and Gottlieb (2008) and Kline and Moretti (2014a), is to characterize efficiency assuming that $c_j = z_j$; i.e., per capita consumption of traded goods equals output in every location. Utility per worker in j becomes $v_j(L_j) = A_j Z_j L_j^{\gamma_A+\gamma_P}$, and it is equalized across locations in equilibrium because workers are perfectly mobile. In turn, the solution to a planner's problem who chooses L_j subject to the same no-transfers restriction also delivers equalization of utility.[7] Given this formulation of the planner's problem, the market allocation is efficient. This result follows from the fact that, as long as consumption equals output and there are constant elasticity spillovers, welfare is a constant-elasticity function of city size. Then, equalization of marginal returns (the planner's efficiency condition) is equivalent to equalization of average returns (the market allocation). This result is often described by saying that there are no gains to reallocation because the marginal productivity gain in one location is exactly offset elsewhere.[8]

This analysis is made under a strong restriction in the planning problem, namely that each

[7]If the planner maximizes $u \equiv \sum_j L_j v_j(L_j)$, the marginal return to adding a worker in j is $\frac{du}{dL_j} = \left(1+\gamma^A+\gamma^P\right) v_j$. Using a different notation, Proposition 1 of Glaeser and Gottlieb (2008) solves this planner problem, which implies equalization of $\frac{du}{dL_j}$, leading to equalization of v_j. Kline and Moretti (2014a) make the point that if dL workers are reallocated from i to j, then the change in the planner's objective function is $du = \left(1+\gamma^A+\gamma^P\right)(v_i - v_j)\, dL$, so that $du = 0$ (no gains from reallocations) starting from any market allocation with free mobility (because $v_i = v_j$ in those allocations).

[8]For instance, Duranton and Venables (2018) write: "When cluster expansion occurs because of labour relocation from other areas, agglomeration gains in the targeted area will come at the expense of agglomeration losses elsewhere. In the specific case where the agglomeration elasticity is constant, the gains in the targeted area will be exactly offset by the losses elsewhere."

region must consume the same amount of traded output it produces. This restriction rules out government policies that tax and redistribute income across locations. When transfers of resources between locations are allowed, the result and intuition described above no longer hold, as welfare is no longer a constant elasticity function of city size.

Starting from the same market allocation without transfers we described above, we now assume that the government can implement spatial transfers. When a transfer is implemented, the distribution of consumption per capita c_j changes and workers move following subsidies to equalize utility in space. As shown in Appendix A.1, the common level of utility across workers changes according to

$$\frac{du}{u} = \left(\gamma^P + \gamma^A\right) \sum_j \left(\frac{Y_j}{Y}\right) \frac{dL_j}{L_j}, \tag{1}$$

where dx is the infinitesimal change in x and Y_j/Y is the share of city j in output. Therefore, a transfer leading to a reallocation of dL workers from j to location i yields

$$\frac{du}{u} = \left(\gamma^P + \gamma^A\right) (z_i - z_j) \frac{dL}{Y}. \tag{2}$$

From (2), there are gains from implementing a reallocation whenever the market allocation without transfers yields differences in output per worker ($z_i \neq z_j$). In turn, this will be the case whenever there are differences in amenities ($a_i \neq a_j$), as the initial allocation without transfers equalizes utility ($a_i z_i = a_j z_j$).

This analysis shows that the *laissez-faire* allocation is inefficient even when spillovers have constant elasticity, as long as there is dispersion amenities, a_j.[9] This is true regardless of whether the source of the spillovers is amenities, productivities, or both. If, for instance, congestion forces dominate ($\gamma^P + \gamma^A < 0$) then it is optimal to implement transfers that reallocate workers to places with low output per worker and high marginal utility of consumption. With this intuition at hand, we now set out to characterize first-best spatial policies in the context of a more general spatial equilibrium model.

2.2 Environment

We consider a closed economy with a discrete number J of locations indexed by j or i. Each worker belongs to one of Θ different types. The type indexes each worker's preference and productivity in each location, as well as each worker's capacity to generate and absorb productivity and amenity spillovers. Workers are free to choose where to live. National labor market clearing is

$$\sum_j L_j^\theta = L^\theta, \tag{3}$$

[9]In a more general model with costly trade or local non-traded goods, the allocation is inefficient even with no dispersion in amenities. What matters is that amenities, non-traded goods, or trade frictions lead to compensating differentials between cities.

where L^θ is the fixed aggregate supply of group θ. The utility of a worker of type θ in location j is

$$u_j^\theta = a_j^\theta \left(L_j^1, .., L_j^\Theta\right) U\left(c_j^\theta, h_j^\theta\right). \tag{4}$$

The function $a_j^\theta(\cdot)$ captures the valuation of a worker of type θ for location j's amenities. Workers may vary in how much they value amenities associated with exogenous features of each location, and also in how much they value amenity spillovers created by each type. For example, a demographic group may prefer living in locations with higher density of their own demographic group, or may value urban amenities generated or congested by specific groups. To capture this feature, $a_j^\theta(\cdot)$ depends on the distribution of workers of different types living in j. Workers also derive utility from a bundle of differentiated tradeable commodities (c_j^θ) and from non-tradeable services including housing (h_j^θ). The utility function $U(c,h)$ is homogeneous of degree 1.

Every location produces traded and non-traded goods. Tradeable output uses an aggregate technology $Y_j\left(N_j^Y, I_j^Y\right)$ requiring services of labor N_j^Y and intermediates I_j^Y. Similarly, production in the non-traded sector is $H_j\left(N_j^H, I_j^H\right)$. The functions Y_j and H_j may be city-specific and feature constant or decreasing returns to scale, due to the use of fixed factors such as land. Therefore, the framework allows for heterogeneous housing supply elasticities across cities through the city specific decreasing returns to scale in $H_j(\cdot)$. The feasibility constraint in the non-traded sector in j is:

$$H_j\left(N_j^H, I_j^H\right) = \sum_\theta L_j^\theta h_j^\theta. \tag{5}$$

Goods in the traded sector can be shipped domestically or to other locations. The country's geography is captured by iceberg trade frictions $d_{ji} \geq 1$. These frictions mean that $d_{ji}Q_{ji}$ units must be shipped from location j to i for Q_{ji} units to arrive. The feasibility constraint of traded goods dictates:

$$Y_j\left(N_j^Y, I_j^Y\right) = \sum_i d_{ji}Q_{ji}. \tag{6}$$

Traded goods may be differentiated by origin, reflecting either industrial specialization at the regional level or variety specialization at the plant level. Specifically, the traded goods arriving in i are combined through the homothetic and concave aggregator $Q(Q_{1i}, .., Q_{Ji})$. This bundle of traded commodities may be used for final consumption or as an intermediate input in local production:

$$Q(Q_{1i}, .., Q_{Ji}) = \sum_\theta L_i^\theta c_i^\theta + I_i^Y + I_i^H. \tag{7}$$

The standard assumptions in Rosen (1979)-Roback (1982) models is that products are perfect substitutes, which implies $Q(Q_{1i}, .., Q_{Ji}) = \sum_i Q_{1i}$. Economic geography models assume differentiation by origin using constant-elasticity of substitution (CES) functional forms. For now, we do not impose these restrictions.

All workers supply one unit of labor with efficiency that may vary by worker type and location.

Each type-θ worker in location j supplies

$$z_j^\theta = z_j^\theta \left(L_j^1, .., L_j^\Theta\right) \tag{8}$$

efficiency units. The function z_j^θ captures exogenous differences in productivity between locations and skill groups, as well as productivity spillovers across workers. Spillovers take place outside the firm at the level of the city. For instance, the concentration of activity in a city gives rise to thick local labor markets that allows better matches between firms and workers, as well as knowledge spillovers –workers learn from each other through social interactions (see e.g. Duranton and Puga (2004)). As with amenities, these spillovers may depend on the distribution of types. For example, high-skill workers may benefit more than low-skill workers from being employed in the same city as other high-skill workers, or in more densely populated areas. In both traded and non-traded sectors, the services $z_j^\theta L_j^\theta$ of the various types of labor are combined through the possibly non-homothetic aggregator $N\left(z_j^1 L_j^1, .., z_j^\Theta L_j^\Theta\right)$. This aggregator also captures imperfect substitution across workers. Feasibility in the use of labor services then implies

$$N_j^Y + N_j^H = N\left(z_j^1 L_j^1, .., z_j^\Theta L_j^\Theta\right). \tag{9}$$

We highlight two key features relative to an otherwise standard neoclassical environment with a representative worker-consumer. First, the location of a worker drives both her marginal product (because productivity is place specific) and her marginal utility of consumption (through local amenities). Therefore, production and consumption decisions are not separable. Second, the framework features two potential sources of non-convexities through the amenity and productivity spillover functions. The utility of each agent may increase with the number of other agents in the same location through a_j^θ and the labor aggregator $N(\cdot)$ may feature increasing returns to the number of workers in a particular group through $z_j^\theta\left(L_j^1, .., L_j^\Theta\right) L_j^\theta$.

At this stage, it is convenient to define the productivity and the amenity spillover elasticities:

$$\gamma_{\theta,\theta'}^{P,j} \equiv \frac{L_j^\theta}{z_j^{\theta'}} \frac{\partial z_j^{\theta'}}{\partial L_j^\theta}, \quad \text{and} \quad \gamma_{\theta,\theta'}^{A,j} \equiv \frac{L_j^\theta}{a_j^{\theta'}} \frac{\partial a_j^{\theta'}}{\partial L_j^\theta}. \tag{10}$$

These elasticities capture the marginal spillover of a type θ worker on the efficiency and utility of each type θ' worker in city j. The case without spillovers corresponds to $\gamma_{\theta,\theta'}^{P,j} = \gamma_{\theta,\theta'}^{A,j} = 0$. So far we have not imposed functional forms, so that these elasticities can be variable.

2.3 Competitive Allocation

In the decentralized equilibrium each worker chooses location and consumption to maximize utility, while competitive producers hire labor and buy intermediate inputs to maximize profits. Being atomistic, these agents do not take into account the impact of their choices on the spillover functions $a_j^\theta\left(L_j^1, .., L_j^\Theta\right)$ and $z_j^\theta\left(L_j^1, .., L_j^\Theta\right)$.

Workers Conditional on living in j, a type-θ worker with expenditure level x_j^θ solves

$$\max_{c_j^\theta, h_j^\theta} U\left(c_j^\theta, h_j^\theta\right) \quad s.t. \quad P_j c_j^\theta + R_j h_j^\theta = x_j^\theta, \tag{11}$$

where P_j is the price of the bundle of traded goods and R_j is the unit price in the non-traded sector. As a result, utility per worker is

$$u_j^\theta = a_j^\theta \left(L_j^1, .., L_j^\Theta\right) \frac{x_j^\theta}{\psi\left(P_j, R_j\right)}, \tag{12}$$

where $\psi\left(P, R\right)$ is the price index associated with U. In equilibrium, all type-θ workers attain the same utility u^θ. Workers' location choice implies that

$$u_j^\theta \leq u^\theta, \tag{13}$$

with equality if $L_j^\theta > 0$.

Firms Producers of traded and non-traded commodities maximize profits. Given a distribution of wages per worker $\left\{w_j^\theta\right\}$, the wage of type-$\theta$ workers in location j equals the value of its marginal product taking as given the efficiency distribution $\left\{z_j^\theta\right\}$:

$$w_j^\theta = W_j \frac{\partial N\left(z_j^1 L_j^1, .., z_j^\Theta L_j^\Theta\right)}{\partial L_j^\theta}, \tag{14}$$

where W_j is the wage per efficiency unit of labor. We assume a no-arbitrage condition, so that the price in location i of the traded good from j equals $d_{ji} p_j$. Free entry of intermediaries who can buy and resell goods between regions ensures this condition holds. Given these prices, the trade flows are:

$$P_i \frac{\partial Q\left(Q_{1i}, .., Q_{Ji}\right)}{\partial Q_{ji}} = d_{ji} p_j, \tag{15}$$

where p_j is the domestic price of the tradeable commodity produced in j. In the competitive equilibrium the prices of final goods, P_j and R_j, adjust so that the corresponding goods markets clear.

Expenditure Per Worker The only component of the competitive allocation left to define is the per capita expenditure for a type-θ worker who lives in j, x_j^θ. Each type-θ worker in location j earns the wage w_j^θ and owns a fraction b^θ of the national returns to fixed factors Π. Individuals hold the same portfolio regardless of where they locate. In addition, we allow for government policies that tax and transfer income across locations. As a result, expenditure per capita is

$$x_j^\theta = w_j^\theta + b^\theta \Pi + t_j^\theta, \tag{16}$$

where t_j^θ is the net government transfer to a type-θ worker living in j. Using balanced budget for the government, expenditure equals net income:

$$\sum_j \sum_\theta L_j^\theta x_j^\theta = \sum_j \sum_\theta L_j^\theta w_j^\theta + \Pi. \tag{17}$$

Definition 1. *A competitive allocation consists of quantities* $c_j^\theta, h_j^\theta, L_j^\theta, Q_{ij}, N_j^Y, I_j^Y, N_j^H, I_j^H$, *utility levels* u^θ, *prices* P_j, R_j, p_j, *wages per efficiency unit* W_j, *and wages per worker* w_j^θ *such that*

(i) the consumption choices c_j^θ, h_j^θ *are a solution to* (11) *for expenditures* x_i^θ *satisfying* (16), *and employment* L_j^θ *is consistent with the spatial mobility constraint* (13);

(ii) the labor and intermediate input choices $N_j^Y, I_j^Y, N_j^H, I_j^H$ *are such that producers maximize profits, labor demand is given by* (14), *and trade flows* Q_{ji} *are given by* (15);

(iii) the government budget constraint is satisfied; i.e. (17) *holds, and*

(iv) all markets clear, i.e. (3) *to* (9) *hold.*

2.4 Planning Problem

We characterize a planning problem where the planner chooses the distribution of workers over locations and types $\left\{L_j^\theta\right\}$, consumption of traded and non-traded commodities $\left\{c_j^\theta, h_j^\theta\right\}$, trade flows $\{Q_{ij}\}$, and the allocation of efficiency units and intermediate inputs, $\left\{N_j^Y, I_j^Y, N_j^H, I_j^H\right\}$. The planner implements policies that treat in the same way all individuals within a type θ, and is bound by the spatial mobility constraint (13). Along with that constraint, the market clearing conditions (3) to (9) define a set $\mathcal{U}$ of attainable utility levels $\left\{u^\theta\right\}$. The optimal planning problem is

$$\begin{aligned} &\max u^\theta \\ &\text{s.t.: } u^{\theta'} = \underline{u}^{\theta'} \quad \text{for } \theta' \neq \theta \\ &\qquad u^{\theta'} \in \mathcal{U} \quad \text{for all } \theta' \end{aligned}$$

The set of solutions of this problem given an arbitrary θ for all feasible values of $\underline{u}^{\theta'} \in \mathcal{U}$ for $\theta' \neq \theta$ defines the utility frontier. Existence is guaranteed, since the planner optimizes a continuous objective function over the compact nonempty set defined by the feasibility constraints. Competitive equilibria according to Definition 1 may not correspond to a point on the frontier due to spatial inefficiencies: workers do not internalize the impact of their location choice on amenities through a_j^θ and firms do not internalize the impact of their hiring decisions on efficiency through z_j^θ. We turn next to the solution and implementation of this planning problem.

3 Optimal Transfers

3.1 Simple Example with Heterogeneous Workers

We return to the simplified setup of Section 2.1, now augmented with several worker types.[10] We examine the effect of implementing small spatial transfers, starting from a market allocation without transfers, such that the the welfare of every group but one (θ_0) is kept constant. As shown in Appendix A.2, the utility of this group changes according to:

$$\frac{du^{\theta_0}}{u^{\theta_0}} = \frac{1}{Y^{\theta_0}} \sum_j \sum_{\theta \in \Theta} \left(\sum_{\theta' \in \Theta} \left(\gamma^P_{\theta,\theta'} + \gamma^A_{\theta,\theta'} \right) w^{\theta'}_j \frac{L^{\theta'}_j}{L^{\theta}_j} \right) dL^{\theta}_j, \tag{18}$$

where dL^{θ}_j is the population change triggered by the transfers, w^{θ}_j is the wage of workers of type θ in j, and Y^{θ_0} are the aggregate wages of θ_0 workers.

Naturally, it is better to reallocate workers into cities where they generate larger spillovers. If type θ generates positive spillovers on type θ' ($\gamma^P_{\theta,\theta'} + \gamma^A_{\theta,\theta'} > 0$), it is desirable to reallocate type θ into cities where type θ' is more productive (i.e., where $w^{\theta'}_j$ is high), much as in (1) in the one-group case. Hence, as in the case with homogeneous workers from Section 2.1, the allocation without transfers is generically inefficient even with constant-elasticity spillovers.

Furthermore, it is profitable to reallocate workers that generate positive spillovers into locations where they are relatively scarce (i.e., where $L^{\theta'}_j / L^{\theta}_j$ is low), reflecting that sorting in the undistorted equilibrium can be inefficient. This gain from reallocation happens even without compensating differentials through amenities, which were necessary to obtain gains in the homogeneous-worker case. In this sense, inefficient sorting creates an additional rationale for gains from spatial transfers.

3.2 Efficiency Condition and Optimal Transfers

We now characterize efficiency in the general model. It is useful to note that the competitive allocation can be determined given an arbitrarily chosen expenditure distribution $\left\{ x^{\theta}_j \right\}$ over types and locations. We can then choose the transfers t^{θ}_j to implement the arbitrarily chosen x^{θ}_j using (16). Therefore, we can obtain a condition over the expenditure distribution x^{θ}_j that must hold in any efficient allocation, regardless of what particular policy tools are used to achieve it. Comparing an allocation with expenditures x^{θ}_j to the outcomes of the planning problem, detailed in Definition 2 of Appendix A.3, we obtain the following result.

Proposition 1. *If a competitive equilibrium is efficient, then*

$$W_j \frac{dN_j}{dL^{\theta}_j} + \sum_{\theta'} \frac{x^{\theta'}_j L^{\theta'}_j}{a^{\theta'}_j} \frac{\partial a^{\theta'}_j}{\partial L^{\theta}_j} = x^{\theta}_j + E^{\theta} \quad \textit{if } L^{\theta}_j > 0, \tag{19}$$

[10] Compared to the full framework, we assume that only tradeable output is valued in consumption ($u^{\theta}_j = a^{\theta}_j c^{\theta}_j$), labor is the only factor of production, goods are perfect substitutes across origins and traded without frictions, and the spillover elasticities defined in (10) are constant, $\gamma^{P,j}_{\theta,\theta'} = \gamma^P_{\theta,\theta'}$ and $\gamma^{A,j}_{\theta,\theta'} = \gamma^A_{\theta,\theta'}$.

for all j and θ and some constants $\{E^{\theta}\}$. If the planner's problem is globally concave and (19) *holds for some specific $\{E^{\theta}\}$, then the competitive equilibrium is efficient.*

Condition (19) defines a relationship between expenditure per capita and the labor allocation that must hold in any efficient allocation. This condition shows the equalization of the marginal benefits and costs of type-θ workers across inhabited locations. The first term on the left is the value of the marginal product of labor, including both the direct output effect and the productivity spillovers.[11] The second term is the marginal benefit (or costs if negative) through amenity spillovers on each group of workers living in j, measured in expenditure equivalent terms.

These marginal benefits from allocating a type θ worker to location j are equated to the marginal costs on the right. The first term, x_j^{θ}, results from the non-separability between a worker's location and consumption: each type-θ worker in j requires x_j^{θ} units of expenditures in that particular location. From a social planning perspective this is a cost, because each additional worker in j translates into lower consumption of traded and non-traded commodities for other workers in that location. The last term, E^{θ}, is the multiplier of the national market clearing constraint (3) in the planner's problem and measures the opportunity cost of employing a type-θ worker elsewhere.

We can draw several useful implications from this result. First, asking whether the spatial allocation is efficient is equivalent to asking whether the expenditure distribution in the market allocation lines up with (19), because the set of equations defining the competitive allocation coincides with the set defining the planner's allocation, except potentially for the expenditure distribution. Therefore, despite the multiple general-equilibrium ramifications of the spillovers, market inefficiencies can be fully tackled through policies acting on x_j^{θ}. This compartmentalization of the inefficiencies reflects a broader "principle of targeting" noted by Bhagwati and Johnson (1960) in trade-policy contexts and by Sandmo (1975) and Dixit (1985) in economies with external effects. We construct policies that implement the optimal expenditure distribution in the next section.

Second, Proposition 1 extends a familiar efficiency condition from the misallocation literature to spatial environments. In our economy, "space" enters through trade costs, non-traded goods, congestion and amenities. In the absence of these forces, there would be no compensating differentials across locations and, as a result, the equilibrium would exhibit the same expenditure per capita x_j^{θ} for each type θ across locations. In that case, the model would be equivalent to one describing the allocation of workers across firms, and (19) would collapse to the familiar condition that the marginal value-product of labor is equalized across locations.

Third, information about the distribution of expenditure per capita x_j^{θ} is needed to assess the economy's efficiency. In studies of misallocation across firms (Hsieh and Klenow, 2009), the absence of compensating differentials justifies the practice of inferring allocative inefficiencies from differences in income per worker. In our spatial environment with compensating differentials, the non-separability of consumption and production means that the net marginal benefit of reallocating

[11] $\frac{dN_j}{dL_j^{\theta}}$ denotes the total differential of $N_j = N\left(z_j^1\left(L_j^1,..,L_j^{\Theta}\right)L_j^1,..,z_j^{\Theta}\left(L_j^1,..,L_j^{\Theta}\right)L_j^{\Theta}\right)$ with respect to L_j^{θ}.

a worker includes the local expenditure of that worker. As a result, assessing the efficiency of the allocation requires data on the distribution of expenditure per capita x_j^θ. Given knowledge of this distribution, further information on how the returns to fixed factors Π are distributed is not necessary to assess efficiency.[12]

Finally, we note that (19) is a necessary but not sufficient condition for efficiency. Even if this condition holds, inefficient market equilibria could exist. However, the inefficient allocations consistent with (19) can be ruled out if the planner's problem is globally concave, as in that case only one allocation that satisfies the first order conditions of the planner. In Section 3.6 we introduce conditions for global concavity of the planner's problem.[13]

We now move on to deriving transfers that implement the previous efficiency conditions. Combining (16) and the definitions of the spillover elasticities (10) with Proposition 1 and labor demand (14), we obtain the optimal transfers.

Proposition 2. *The optimal allocation can be implemented by the transfers*

$$t_j^{\theta*} = \sum_{\theta'} \left(\gamma_{\theta,\theta'}^{P,j} w_j^{\theta'*} + \gamma_{\theta,\theta'}^{A,j} x_j^{\theta'*} \right) \frac{L_j^{\theta'*}}{L_j^{\theta*}} - \left(b^\theta \Pi^* + E^\theta \right) \quad \text{if } L_j^{\theta*} > 0, \tag{20}$$

where the terms $\left(x_j^{\theta*}, w_j^{\theta*}, L_j^{\theta*}, \Pi^* \right)$ *are the outcomes at the efficient allocation, and* $\left\{ E^\theta \right\}$ *are constants equal to the multipliers on the resource constraint of each type in the planner's allocation.*

The optimal transfers $t_j^{\theta*}$ take care of inefficiencies due to spillovers as well as of distributional concerns.[14] In the absence of spillovers we would still have $t_j^{\theta*} = -\left(b^\theta \Pi^* + E^\theta \right)$, so that the transfers would take care of redistribution across types, as implied by the second welfare theorem. The burden of dealing with the spatial inefficiencies falls on the spatial component of the optimal transfers, corresponding to the first term in (20).

In Section 3 we use conditions (16) and (20) for two separate quantitative goals. First, given the spillover elasticities, we use them to determine the efficiency of the observed allocation from data on wages, expenditures, and employment. Second, under the assumption that the observed allocation is efficient, we use the condition to recover the spillover elasticities $\left\{ \gamma_{\theta,\theta'}^{P,j}, \gamma_{\theta,\theta'}^{A,j} \right\}$ from the observed data.

[12] As it was noted early on in studies of optimal city size, assumptions about ownership of fixed factors are relevant to determine the efficiency of the market allocation (Pines and Sadka, 1986). The expenditure distribution implied by (19) that implements the efficient allocation is invariant to assumptions about ownership of fixed factors. A different rule to distribute Π from that assumed in (16) would imply a different set of optimal transfers t_j^θ to implement the optimal expenditure distribution, but would not affect (19).

[13] At the current level of generality, it is possible that a market allocation does not exist or exhibits multiplicity for an arbitrarily chosen distribution of expenditures. However, a market allocation exists if (19) holds.

[14] These optimal transfers apply to populated locations. The planner could choose not to allocate some types to some locations or to leave some locations empty. Implementing this extensive margin entails taxing away all the income of those types.

3.3 Optimal Subsidies with Constant Elasticity Spillovers

From now on we adopt constant-elasticity forms for the functions: $\gamma^{P,j}_{\theta,\theta'} = \gamma^{P}_{\theta,\theta'}$ and $\gamma^{A,j}_{\theta',\theta} = \gamma^{A}_{\theta',\theta}$. Under these assumptions, the optimal transfers in (20) simplify to $t^\theta_j = s^\theta_j w_j - T^\theta$, where

$$s^\theta_j = \frac{\gamma^P_{\theta,\theta} + \gamma^A_{\theta,\theta}}{1-\gamma^A_{\theta,\theta}} + \sum_{\theta'\neq\theta} \frac{\gamma^P_{\theta,\theta'} w^{\theta'}_j + \gamma^A_{\theta,\theta'} x^{\theta'}_j}{1-\gamma^A_{\theta,\theta}} \frac{L^{\theta'}_j}{w^\theta_j L^\theta_j} \tag{21}$$

and

$$T^\theta = b^\theta \Pi + \frac{E^\theta}{1-\gamma^A_{\theta,\theta}}. \tag{22}$$

This representation readily implies that the optimal transfers can be implemented by labor income subsidies s^θ_j coupled with lump-sum tax T^θ. The labor income subsidy s^θ_j is a function of wages, expenditures and population. The labor subsidies tackle spatial inefficiencies due to spillovers, while the lump-sum transfers take care of distributional concerns. We now draw the implications of this formula in special cases.

No Spillover Across Types We consider first a case with several worker types, but with $\gamma^P_{\theta',\theta} = \gamma^A_{\theta',\theta} = 0$ for $\theta' \neq \theta$, so that there are no spillovers across types. The optimal subsidy (21) becomes:

$$s^\theta = \frac{\gamma^P_{\theta,\theta} + \gamma^A_{\theta,\theta}}{1-\gamma^A_{\theta,\theta}}. \tag{23}$$

In the special case of a single worker type, the policy is further simplified to (s, T) with $s = \frac{\gamma^P + \gamma^A}{1-\gamma^A}$. This formula has a simple interpretation. Under negative congestion spillovers for type θ ($\gamma^A_{\theta,\theta} < 0$), if the agglomeration spillover of that type is not too strong ($\gamma^P_{\theta,\theta} < -\gamma^A_{\theta,\theta}$), then all workers of type θ should pay as tax the same fraction of their income everywhere (a negative subsidy, $s^\theta < 0$). In this case, the net transfer t^θ_j received by type-θ workers is smaller, and potentially negative, in cities where their wage is higher.

The presence of compensating differentials is the key reason why, even with constant elasticity spillovers, the *laissez-faire* allocation is generically inefficient. We made this point in Section 2.1 in a special case starting at an equilibrium without transfers. We have now shown that the global optimum is obtained using a constant subsidy-cum-lump sum transfer scheme $\left(s^\theta, T^\theta\right)$ that does not vary across space. To see why this policy distorts the spatial allocation despite being space-independent, we must again consider the role of the compensating differentials. From the mobility constraint (13), indifference across populated locations j and j' implies:

$$\frac{\psi\left(P_{j'}, R_{j'}\right)/a^\theta_{j'}\left(L_{j'}\right)}{\psi\left(P_j, R_j\right)/a^\theta_j\left(L_j\right)} = \frac{\left(1+s^\theta\right) W_{j'} z^\theta_{j'}\left(L_{j'}\right) + T^\theta + b^\theta \Pi}{\left(1+s^\theta\right) W_j z^\theta_j\left(L_j\right) + T^\theta + b^\theta \Pi}. \tag{24}$$

The left hand side is the relative compensating differential (amenity-adjusted cost of living) and the right hand side is the relative expenditure (equal to relative after-tax income) between locations j' and j for type θ. In the presence of amenities, non-traded goods or trade costs, the relative compensating differentials vary across space. As a result, changes to the policy scheme (s^θ, T^θ) lead to changes in the employment distribution of type θ. In the absence of these compensating differentials, the indifference condition would collapse to $W_j z_j (L_j) = W_{j'} z_{j'} \left(L_{j'}\right)$ for any $\left(s^\theta, T^\theta\right)$, and these policies would cease to impact the spatial allocation.

Spillovers Across Types We already saw in the example at the beginning of this section that inefficient sorting creates a rationale for transfers. To see how the optimal subsidies look like, consider a polar case without amenity spillovers and without efficiency spillover on the same type. Assume, furthermore, that there are only two types, $\theta = U, S$ for unskilled and skilled. Then, the optimal subsidy to type-θ workers located in j simplifies to

$$s_j^\theta = \gamma_{\theta,\theta'}^P \left(\frac{w_j^{\theta'} L_j^{\theta'}}{w_j^\theta L_j^\theta} \right). \tag{25}$$

In this special case, the optimal subsidy for workers in group θ varies across locations according to the distribution of relative wage bills, $w_j^\theta L_j^\theta$. A positive cross efficiency spillover implies a higher marginal gain from attracting a given worker type to locations where the economic size of the other type is relatively larger. The result is a higher optimal subsidy for the types that generate spillovers where they are more scarce. Relative to a *laissez-faire* equilibrium, this policy tempers the degree of sorting across cities. Condition (25) also implies $\frac{ds_j^S}{ds_j^U} < 0 \longleftrightarrow \gamma_{S,U}^P \gamma_{U,S}^P > 0$, so that subsidies of both types are negatively correlated across cities if both types generate positive efficiency spillovers. These basic intuitions will help us rationalize the quantitative findings about the spatial efficiency of the current transfer scheme in the U.S. economy.

Link to Henry George Theorem We discussed above an implementation of the optimal transfers (20) with labor income subsidies (21) and lump-sum taxes (22). However, other implementations are possible. Is it possible, in our context, to tax only the returns to fixed factors Π (instead of raising lump-sum taxes) in order to finance place-specific subsidies to mobile factors? This question is motivated by the Henry George Theorem, which says that, in some environments, land taxes raise just enough revenue to finance efficient government expenditures.[15] This question is only meaningful when the optimal labor income subsidies are positive, as otherwise the tax system necessarily entails taxing mobile factors. Then, under some regularity conditions, our model implies

[15] See Arnott (2004) for a review. In systems-of-cities models following Henderson (1974), if public goods are the source of agglomeration then it is efficient to tax land rents and use the proceeds to finance public expenditures. With increasing returns to scale in production, the theorem is cast as an equality between land rents and the value of output times the degree of returns to scale at the level of a city (see Section III of Arnott, 2004). These results hold at the city level, and are derived in models with homogeneous workers, identical locations, no spatial interactions among cities, and free entry of cities.

that the returns to the fixed factors Π add up to more than the total lump-sum taxes in (22).[16] In this case, the tax system implementing optimal subsidies may feature aggregate redistribution from fixed factors to mobile factors.

3.4 Economic Geography Frameworks

The environment laid out in Section 2 nests standard economic geography models, such as Helpman (1998), Allen and Arkolakis (2014) and Redding (2016).[17] These models are the basis of a growing body of quantitative research studying the spatial implications of regional shocks, summarized by Redding and Turner (2015) and Redding and Rossi-Hansberg (2017). However, their normative implications have barely been explored.[18] We now apply the previous results to shed light on optimal policies in these environments.

To specialize our setup to these models we assume a single worker type, Cobb-Douglas preferences with weight α_C on traded goods, and a constant amenity spillover elasticity γ^A. Utility per worker in location j then is

$$u_j = A_j L_j^{1+\gamma^A} c_j^{\alpha_C} h_j^{1-\alpha_C} \tag{26}$$

Production only uses labor and the efficiency spillover has a constant elasticity γ^P, so that tradeable output in region j is

$$Y_j = Z_j L_j^{1+\gamma^P}. \tag{27}$$

Supply of non-traded goods in location j is inelastic and equal to H_j. In a competitive allocation, workers in j receive a wage w_j equal to tradeable output per worker.

Applying Proposition 1 under these assumptions, we find that a linear relationship between expenditure and wages implements the efficient allocation

$$x_j = (1-\eta) w_j + \eta \bar{w}, \tag{28}$$

where $\bar{w}$ is the average wage in the economy and $\eta \equiv 1 - \frac{\alpha_C(1+\gamma^P)}{1-\gamma^A}$ combines the spillover elasticities and the expenditure share in traded goods. The corresponding optimal transfers are linear in wages:

[16] Using (22) we obtain: $\sum_\theta L^\theta T^\theta = \Pi + \sum_\theta \frac{L^\theta E^\theta}{1-\gamma^A_{\theta,\theta}}$. Hence if the planning problem is convex (implying $E^\theta < 0$), and own-congestion spillovers are not too strong ($\gamma^A_{\theta,\theta} < 1$), we get $\Pi > \sum_\theta L^\theta T^\theta$.

[17] Our presentation so far has assumed that each location sells a different product under perfect competition. In Appendix A.4 we show that the analysis would be the same assuming free entry of producers of differentiated varieties under monopolistic competition as in the standard Krugman (1980) model. The key reason why this equivalence holds is that under CES preferences the number of producers M_j and the bilateral trade flows are efficient given the allocation of labor L_j. Therefore, the labor allocation remains the only inefficient margin and our propositions and results from Section 3.4 go through. These properties would not go through under monopolistic competition outside of CES. In that case, the entry and bilateral pricing decisions would be inefficient (Zhelobodko et al., 2012).

[18] In his review of the policy implications of empirical economic-geography studies, Combes (2011) notes the lack of a general-equilibrium analysis of the optimal allocation of employment in a model of regional trade allowing for geographic inter-dependencies. Other recent papers studying spatial policies in geography models include Allen et al. (2015) who consider zoning restrictions within a city, Fajgelbaum and Schaal (2017) who consider transport network investment, and Gaubert (2015) who characterizes the optimal allocation in a model heterogeneous firms and a complementarity between city size and firm productivity.

$t_j = \eta(\bar{w} - w_j)$. Barring knife-edge cases on the parameters ($\eta = 0$) or the fundamentals (such that $w_j = w$), the efficient allocation generically features trade imbalances. In particular, under the empirically consistent case of $\eta < 0$, efficiency requires net trade deficits in high-wage regions.

Should the optimal policy that implements (28) redistribute towards or away from high-wage locations? The answer depends on the distribution of non-labor income (the returns to land H_j). To answer this question, we can assume like Caliendo et al. (2018) that a fraction ω of the returns to fixed factors is distributed locally to the L_j workers in j and the remainder is evenly split across all workers. The optimal policy can again be expressed as a constant labor subsidy s that is common across locations and equal to

$$s = \frac{1+\gamma^P}{1-\gamma^A}\left[1 - (1-\alpha_C)\,\omega\right] - 1, \tag{29}$$

with lump-sum transfer equal $T = -s\bar{w}$. Even in the absence of spillovers, the equilibrium is inefficient as long as there is some local ownership ($\omega > 0$). In this case, we obtain a non-zero subsidy that corrects the distortion introduced by local ownership. With spillovers, the optimal policy redistributes income away from low-wage regions when $s > 0$, and into low-wage regions under a labor tax ($s < 0$). Assuming common ownership of the national portfolio ($\omega = 0$) as in Helpman (1998), and continuing to assume that $\eta < 0$, spatial efficiency requires income redistribution to regions with above-average wage ($s > 0$). In contrast, assuming away trade imbalances as in Allen and Arkolakis (2014) and Redding (2016), the optimal policy redistributes income to low-wage regions ($s < 0$).

In sum, the details of the microeconomic structure and the country's economic geography (represented by bilateral trade costs) do not impact the relationship between optimal trade imbalances and wages, nor the policies that implement them, whereas the ownership of fixed factors determines whether the optimal policies should redistribute income towards or away from high-wage regions.

3.5 Additional Forces

Preference Draws within Types To incorporate that workers may have idiosyncratic preferences for location, we extend the model to assume that a worker l of type θ derives utility $u_j^\theta \epsilon_j^l$ from living in location j, where ϵ_j^l captures idiosyncratic preferences that are i.i.d. and distributed Fréchet, $\Pr\left(\epsilon_j^l < x\right) = e^{-x^{-1/\sigma_\theta}}$. The preference draws are eliminated when $\sigma_\theta = 0$, in which case we return to the original formulation of the model. Every other aspect of the model remains the same except for the spatial mobility constraint (13), which is is now replaced with the following labor-supply equation:

$$\frac{L_j^\theta}{L^\theta} = \left(\frac{u_j^\theta}{u^\theta}\right)^{1/\sigma_\theta}. \tag{30}$$

Taking into account this difference, we can compute the optimal allocation and define optimal transfers using the same definition of the planner's problem as in 2.4. Then, Propositions 1 and 2 go through with only one modification: instead of $\gamma_{\theta,\theta}^{A,j}$, the relevant amenity spillover elasticity on the own type becomes $\tilde{\gamma}_{\theta,\theta}^{A,j} \equiv \gamma_{\theta,\theta}^{A,j} - \sigma_\theta$. Hence, without spillovers we obtain a (negative) labor subsidy $s^\theta = -\frac{\sigma_\theta}{1+\sigma_\theta}$. These subsidies tackle distributional concerns rather than inefficiencies. The incentives for redistribution arise from the combination of two reasons: i) different individuals l within a group θ receive the same planner's weight; and ii) the planner conditions outcomes on location j and type θ, but not on individual preference draws ε_j^θ. Because on average individuals have higher preference draws conditional on having sorted into lower-wage locations, the planner has incentives to redistribute towards those locations.

Commuting We apply the analysis to a framework with commuting in the style of Ahlfeldt et al. (2015) and Monte et al. (2018). . We assume only one type of agent. The difference with our benchmark model is that now an individual l chooses the commuting pattern ji; i.e., a residence location j and a workplace i. The amenity spillovers depend on the number of residents L_j^R, and the productivity spillovers depend on the number of workers L_i^W. The productivity of a commuter from j to i is $z_i\left(L_i^W\right)$, and the common component of utility (4) is $u_{ji} = a_i\left(L_j^R\right)U_{ji}\left(c_{ji}, h_{ji}\right)$, where the function U_{ji} may vary by ij to capture disutility from commuting travel time. We also allow for an idiosyncratic worker-level shock ϵ_{ij}^l according to a Fréchet distribution, $\Pr\left(\epsilon_{ij}^l < x\right) = e^{-x^{-1/\sigma}}$, so that the utility of a commuter l from j to i is $u_{ji}\epsilon_{ji}^l$. As a result, the flow of commuters of j to i is $L_{ji} = L\left(\frac{u_{ji}}{u}\right)^{1/\sigma}$. In the market, each of these commuter makes total expenditures x_{ji} at j. Every other aspect of the model is the same as in the benchmark.

We show in Appendix A.6 that the optimal transfers received by a consumer-worker can be decomposed as the sum of two types of transfers. The first component of transfer depends on where she works,

$$t_i^W = \frac{\gamma_i^P - \sigma}{1+\sigma} w_i^*, \tag{31}$$

and the second component depends on where she lives,

$$t_j^R = \frac{\gamma_j^A}{1+\sigma} \sum_{i'} \frac{L_{ji'}^* x_{ji'}^*}{L_j^R}, \tag{32}$$

so that the optimal transfer are $t_{ji}^* = t_i^W + t_j^R - T$.[19] The workplace policy t_i^W is the Pigouvian tax fixing the inefficiency in production, while the residence policy t_j^R isolates the role of amenity spillovers. The two policies are separable. That is, even with commuting, the optimal transfer varies only by place rather than by bilateral commuting pattern. Absent spillovers at the place of residence ($\gamma^A = 0$), the workplace transfer t_i^W is the only one active and takes the same form as in the benchmark model without commuting.

Spillovers Across Locations Recent studies such as Lucas and Rossi-Hansberg (2002) and Rossi-Hansberg (2005) emphasize that economic activity in one location may generate spillovers in other locations. We now derive the optimal transfers in this case. To simplify the exposition, we

[19] These expressions assume that the returns to fixed factors Π are evenly distributed in the population. Lump-sum transfers adjust so that the government budget is balanced.

consider a special case of our model with homogeneous workers and constant-elasticity spillovers in amenities. However, we now extend our model to allow for the efficiency of location j to be an arbitrary function of the number of workers in every location: $z_j = z_j\left(\left\{L_{j'}\right\}\right)$. This formulation accommodates a commonly used specification where spillovers decay with distance between spatial units.[20] We define the efficiency spillover elasticity across locations,

$$\gamma^{P,j,j'} = \frac{\partial z_{j'}}{\partial L_j}\frac{L_j}{z_{j'}}, \tag{33}$$

as the elasticity of the efficiency of workers at j' with respect to the number of workers located in j. Following similar steps to propositions 1 and 2, the optimal transfers now are:

$$t_j = \frac{\gamma^{P,j,j} + \gamma^A}{1-\gamma^A} w_j + \sum_{j' \neq j} \frac{\gamma^{P,j,j'}}{1-\gamma^A}\frac{L_{j'} w_{j'}}{L_j} + T. \tag{34}$$

We find as before that the optimal transfers can be characterized as a function of spillover elasticities and outcomes such as wages and employment, regardless of micro heterogeneity in fundamentals. In particular, non-localized spillovers lead to the intuitive implication that the optimal transfers should be higher in places that generate strong spillovers to larger locations, as measured by their total wage bill.

Fiscal Competition We have only inspected policies set by a national government. We now ask what set of transfers may arise in a Nash equilibrium where local governments levy taxes on the fixed factors within their boundaries and rebate the revenue to its residents. Under some conditions such that spillover elasticities are not too strong, in a special case of our model with homogeneous workers discussed in Appendix A.10, local governments who maximize the utility per worker of their residents will tax away all the fixed factor income and distribute the revenue to their residents. Intuitively, local governments do not have tools to transfer value from elsewhere to their residents, nor will they want to give away net income to other locations. The resulting Nash equilibrium features no imbalances in equilibrium, so that consumption per capita of tradeable output equals the wage. As we have discussed in Section 3.4, such an outcome is inefficient and can be corrected via the optimal subsidy (29) evaluated at $\omega = 1$.[21]

[20] E.g., Ahlfeldt et al. (2015) assume $z_{j'} = \left(\sum_j L_j e^{-\delta t_{jj'}}\right)^{\alpha}$ where $t_{jj'}$ is travel time between j and j' and δ is a decay parameter.

[21] Canonical frameworks of fiscal competition, such as Wilson (1986) and Zodrow and Mieszkowski (1986), include features that are not present in our analysis such as mobile capital across regions and local financing of public goods that are valued by individuals or firms. In the appendix A.8 we develop an extension to include public goods assuming that they are optimally financed by a federation. The optimal policies that we consider in the benchmark model to deal with spillovers go through in that case. It could be interesting for future work to further include some of the forces of the fiscal competition literature into our model.

3.6 Quantitative Implementation

We now return to the main model and show how to bring it to the data. First, we impose the functional-form assumptions that will be used in the quantitative implementation. Second, with these functional forms, we identify conditions that guarantee concavity of the planner's problem and, in the light of Proposition 1, uniqueness of the competitive allocation under the optimal spatial policies. Then, we identify a set of data that suffices to identify the fundamentals of the model and compute the allocation under the optimal spatial policies.

Functional Forms On the demand side, we assume that preferences for traded and non-traded goods are Cobb-Douglas:

$$U(c,h) = c^{\alpha_C} h^{1-\alpha_C}, \tag{35}$$

while the aggregator of traded commodities is CES,

$$Q(Q_{1i},..,Q_{Ji}) = \left(\sum_i Q_{ji}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \tag{36}$$

where $\sigma > 0$ is the elasticity of substitution across products from different origins. On the supply side, the production functions of traded and non-traded goods are

$$Y_j\left(N_j^Y, I_j^Y\right) = z_j^Y \left(N_j^Y\right)^{1-b_{Y,j}^I} \left(I_j^Y\right)^{b_{Y,j}^I}, \tag{37}$$

$$H_j\left(N_j^H, I_j^H\right) = z_j^H \left(\left(N_j^H\right)^{1-b_{H,j}^I} \left(I_j^H\right)^{b_{H,j}^I} \right)^{\frac{1}{1+d_{H,j}}}, \tag{38}$$

where $d_{H,j} \geq 0$ and $\left\{z_j^Y, z_j^H\right\}$ are TFP shifters. Traded goods are produced under constant returns to scale, but we allow for decreasing returns in the housing sector. The coefficient $d_{H,j}$ is the inverse housing supply elasticity of location j in the market allocation, which may vary across regions. The aggregator of labor types is CES,

$$N_j = \sum_{i=1}^{I} \left[\sum_{\theta \in \Theta_i} \left(z_j^\theta L_j^\theta \right)^{\rho_i} \right]^{\frac{1}{\rho_i}}, \tag{39}$$

where $\frac{1}{1-\rho_i} > 0$ is the elasticity of substitution across types of workers. Finally, we impose constant-elasticity forms for the spillovers:

$$z_j^\theta\left(L_j^1,..,L_j^\Theta\right) = Z_j^\theta \prod_{\theta'} \left(L_j^{\theta'}\right)^{\gamma_{\theta',\theta}^P}, \tag{40}$$

$$a_j^\theta\left(L_j^1,..,L_j^\Theta\right) \equiv A_j^\theta \prod_{\theta'} \left(L_j^{\theta'}\right)^{\gamma_{\theta',\theta}^A}. \tag{41}$$

These functional forms are consistent with studies that estimate spillover elasticities, allowing us to draw from existing estimates. The Z_j^θ capture exogenous comparative advantages in production across types and A_j^θ capture preferences for location across types. We refer to $\left\{Z_j^\theta, A_j^\theta\right\}$ as fundamental components of productivity or amenities. Together with the assumptions on production technologies, these functional forms impose Inada conditions, which imply that all locations are populated in the optimal allocation if the planner's problem is convex.

Concavity Condition To ease the notation, we introduce the following composite elasticities of efficiency and congestion spillovers:

$$\Gamma^P = \max_\theta \left\{ \sum_{\theta'} \gamma^P_{\theta',\theta} \right\}, \text{ and } \Gamma^A = \min_\theta \left\{ -\sum_{\theta'} \gamma^A_{\theta',\theta} \right\}.$$

Also, we let $D = \min_j \{d_{H,j}\}$ be the lowest inverse elasticity of housing supply. Under the functional form assumptions (35) to (41) we have the following property.

Proposition 3. *The planning problem is concave if*

$$\Gamma^A > \Gamma^P, \tag{42}$$

$\Gamma^A \geq 0$ *and* $\gamma^A_{\theta,\theta'} > 0$ *for* $\theta \neq \theta'$. *Under a single worker type* ($\Theta = 1$)*, the planning problem is quasi-concave if:*

$$1 - \gamma^A > \left(1 + \gamma^P\right) \left(\frac{1 - \alpha_C}{1 + D} + \alpha_C \right). \tag{43}$$

Condition (42) ensures that congestion forces are at least as large as agglomeration forces. Specifically, the congestion from the type that generates the weakest congestion, measured by Γ^A, dominates the agglomeration from the type that generates the strongest agglomeration, measured by Γ^P. Condition (42) may also apply when there is a single type, in which case it simplifies to $\gamma^P + \gamma^A < 0$. With a single type, further assuming Cobb-Douglas preferences over traded and non-traded goods we obtain the weaker restriction (43), which allows for spillovers to be net agglomerative. [22]

Proposition 3 establishes conditions under which the market allocation is unique given the optimal spatial policies. It extends existing uniqueness results in two dimensions. First, it complements results that characterize uniqueness of the spatial equilibrium under no policy intervention and trade balance (Allen et al., 2014). Second, it holds in a context with heterogeneous workers and cross-groups spillovers. We note that our uniqueness condition applies at the optimal expenditure distribution. Multiplicity is still possible for sub-optimal policies or no policy intervention, but this poses no limitation for our approach.

[22] The CES restriction (36) on the aggregator of trade flows $Q(\cdot)$ is not needed for any of these results. Therefore, these condition holds regardless of product differentiation across locations. We also note that these conditions are sufficient but not necessary for uniqueness. The planner's problem can be concave outside of these strong parameter restrictions. Numerical simulations confirm the intuition that the amount of product differentiation between regions governed by the aggregator $Q(.)$ helps make the planner's problem concave.

Implementation in Changes and Data Requirements We bring the model to the data by assuming that the observed allocation is generated by a decentralized equilibrium consistent with Definition 1, subject to the functional form assumptions (35) to (41). To compute the optimal allocation, we solve for a planner's problem that optimizes over the the expenditure of each type of worker $\hat{x_j^\theta}$ and over the remaining endogenous variables $\left\{\hat{P}_i, \hat{p}_i, \hat{Y}_i, \hat{W}_i, \hat{N}_j, \hat{L_j^\theta}, \hat{R}_i, \hat{u^\theta}\right\}$ relative to the observed equilibrium. The planner chooses these variables to maximize the welfare gains of one group $\hat{u^\theta}$, for arbitrarily chosen welfare changes of the remaining groups. We then vary the welfare changes of the other groups to trace the utility frontier.

The solution to this problem in changes is equivalent to the solution to the problem in levels from Section (2.4), assuming that the same parameters and fundamentals underlie the observed equilibrium and the planning problem in levels. In the spirit of Dekle et al. (2008), as shown in Section A.9, this strategy allows us to identify the data that is sufficient to back out fundamentals and solve for the optimal allocation, as stated in the following proposition.

Proposition 4. *Assume that the observed data is generated by a competitive equilibrium consistent with Definition 1 under the functional forms* (35) *to* (41)*. Then, relative to the initial equilibrium, the optimal allocation can be fully characterized as function of:*

i) the distributions of wages, employment and expenditures across labor types and locations;
ii) the distribution of bilateral import and export shares across locations;
iii) the utility and production function parameters $\left\{\alpha_C, \sigma, \rho, b^I_{Y,j}, b^I_{H,j}, d_{H,j}\right\}$*; and*
iv) the spillover elasticities $\left\{\gamma^A_{\theta',\theta}, \gamma^P_{\theta',\theta}\right\}$.

The proposition establishes data requirements that are sufficient, in the light of our model, to characterize the optimal allocation. In particular, it is sufficient to observe the data in i) and ii) and the elasticities in iii) and iv).

Importantly, this implementation does not impose restrictions on the distributional policies across locations in the observed equilibrium. The net transfers that generate the expenditure distribution x_j^θ exactly match those in the data. In particular, they are not constrained to match a specific tax rule. Nor do we impose that the observed allocation is inefficient: the efficiency of the observed allocation depends on whether the distribution of expenditures lines up with condition (19) in Proposition 1. It could be that the transfers in place are such that the empirical relationship between expenditures, wages and employment is not far from that relationship, in which case our implementation of the planner's problem would predict small welfare gains from implementing optimal policies.

We also highlight that, by following this approach, we ensure that the model is parametrized in a way that exactly matches the observed outcomes enumerated in items i) and ii) of the proposition. The procedure does not need to separately identify the levels of the parameters of the model (e.g. fundamentals $\left\{Z_j^\theta, A_j^\theta\right\}$, TFP shifters $\left\{z_j^Y, z_j^H\right\}$, and bilateral trade costs $\{d_{ij}\}$). It implicitly pins down a combination of the parameters that rationalizes the data in points i) and ii) as an equilibrium outcome of the model from Definition 1. Matching these data is sufficient to perform counterfactuals, together with the knowledge of the elasticities enumerated in points iii) and iv) that govern the responses of the endogenous variables to a change in transfers.

4 Data and Calibration

To take the model to the data, we use as an empirical setting the distribution of economic activity across Metropolitan Statistical Areas (MSAs) in the United States in the year 2007. We identify worker types θ with observable skill groups. Specifically, following Diamond (2016), our benchmark analysis studies the spatial allocation of two skill groups, high skill (college) and low skill (non college) workers. Because of data limitations, our analysis abstracts from more detailed definitions of skill types.[23]

4.1 Data

As established in point i) of Proposition 4, we need data on income and expenditures by group and MSA. To that end, we rely on the BEA's Regional Accounts, which report labor income, capital income and welfare transfers by MSA. A complementary BEA dataset for the years 2000 to 2007 reports total taxes paid by individuals and MSA (Dunbar, 2009). Taken together, these sources give us a dataset at the MSA level. We then apportion each of these MSA-level totals into two labor groups: high skill, defined as workers who have completed at least four years of college, and low skill, defined as every other working age individual. To implement this apportionment, we use shares of labor income, capital income transfers corresponding to each group in each MSA from the American Community Survey (IPUMS-ACS, Ruggles et al. (2017)) collected by the Census, and use shares of taxes for each group in each MSA from the March supplement of the Current Population Survey (IPUMS-CPS, Flood et al., 2017) . Our dataset covers 209 MSAs for which we have both BEA and Census information.[24]

An important concern when measuring these variables is that the model does not include heterogeneity across individuals within each group of skill θ, whereas in reality these groups are heterogeneous across cities. If we did not control for this heterogeneity, our procedure to implement the model would interpret the observed variation in net individual transfers across MSAs within a group as place-based transfers, when they reflect, in part, differences in the types of workers within each group across MSAs. In principle, this concern can be mitigated by allowing for several θ groups corresponding to the fine individual characteristics observed in the ACS. While potentially feasible, such an approach would increase the dimension of the problem and the number of elasticities to calibrate. Alternatively, we choose to purge the observed measures of income, expenditure, taxes and transfers by skill and MSA from compositional effects using a set of socio-demographic

[23] See Baum-Snow and Pavan (2013) and Roca and Puga (2017) for evidence on the role of heterogeneity within observable types in accounting for wage dispersion and sorting.

[24] These areas correspond to 95% of the population and 96% of income of all US metropolitan areas. Metropolitan areas in the US in turn cover 78% of the population, and 83% of personal income.

controls at the MSA-group level built from individual level Census data (IPUMS) on age, educational attainment, sector of activity, race, and labor force participation status of individuals in a given MSA-group. In the quantification we then use measures of income, expenditures, taxes and transfers that are net of variation in socio-demographic composition within groups across MSAs. We discuss the details of this step in Appendix B.

We use the variables above to construct expenditure per capita, x_i^θ, using its definition (16) as labor plus capital income net of taxes and transfers, which also corresponds to the BEA's definition of disposable income. In the model we assume no variation in capital income across cities for each type. Therefore, we use a group-specific measure of capital income consistent with the fact that 52% of non-labor income is owned by high skill workers according to the BEA/ACS data.[25]

As implied by ii) of Proposition 4, quantifying the model also requires data on trade flows between MSAs. The Commodity Flow Survey (CFS) reports the flow of manufacturing goods shipped between CFS zones in the US every five years. The CFS zones correspond to larger geographic units than our unit of observation, the MSA. To overcome this data limitation, we adapt the approach in Allen and Arkolakis (2014), who use estimates of trade frictions as function of geography to project CFS-level flows to the MSA level. In our context, we use the gravity equation predicted by the model to find the unique estimates of trade flows between MSAs that are consistent with actual distance between MSAs, existing estimates of trade frictions with respect to distance, and observed trade imbalances, computed as the difference between income in the traded sector and expenditure on traded goods (for both final and intermediate use) in each MSA.

Finally, to calibrate the labor shares in production in part iii) of Proposition 4, we use ACS data on employment in traded and non-traded sectors by MSA.[26] We also adjust this measure to remove variation from compositional effects following a similar approach to the one described above for income, expenditure, taxes and transfers.

4.2 Calibration

Our model is consistent with Diamond (2016) and generates similar estimating equations to those used in her analysis. We use the same definition of geographic units (MSA) and skill groups (College and Non College), and we rely on similar data sources for quantification. Therefore, her estimates constitute a natural benchmark to parametrize the model. In what follows, we discuss these elasticities and several alternative specifications that are also used in the quantitative section.

Utility and Production Function Parameters $\left\{\alpha_C, \sigma, \rho, b^I_{Y,j}, b^I_{H,j}, d_{H,j}\right\}$ We use the Diamond (2016) estimate of the Cobb-Douglas share of traded goods in expenditure ($\alpha_C = 0.38$), of the inverse housing supply elasticity ($d_{H,j}$ in (38)) for each MSA,[27] and of the elasticity of

[25] This step involves setting a national share of profits in GDP consistent with the general equilibrium of the model. See Appendix B for details.

[26] We define employment in the following NAICS sectors as corresponding to the non-traded sector in the model: retail, real estate, construction, education, health, entertainment, hotels and restaurants.

[27] For MSAs that we cannot match to Diamond (2016) we use the average housing supply elasticity across MSAs.

substitution between high and low skill, estimated at 1.6 and implying $\rho = 0.392$.

We calibrate the Cobb-Douglas share of intermediates in traded good production ($b^{I}_{Y,j} = 0.468$ for all j in (37)) using the share of material intermediates in all private good industries production in 2007 from the U.S. KLEMS data. Having calibrated the previous parameters, the Cobb-Douglas share of labor in non-traded production in each city ($1 - b_{H,j}$ in (38)) can be chosen to match the share of workers in the non-traded sector of each MSA, as detailed in Section B.2. We assume an elasticity of substitution σ among traded varieties in (36) equal to 5, corresponding to a central value of the estimates reported by Head and Mayer (2014).

Efficiency Spillovers $\left\{\gamma^{P}_{\theta',\theta}\right\}$ Previous empirical studies, such as Ciccone and Hall (1996), Combes et al. (2008), and Kline and Moretti (2014a), estimate elasticities of labor productivity with respect to employment density. Across specifications, these studies find elasticities in the range of $(0.02, 0.2)$.[28] Hence, we set a a properly weighted average of the elasticities $\gamma^{P}_{\theta',\theta}$, corresponding to what the empirical specifications of these previous studies would recover in data generated by our model, to match the benchmark value for the U.S. economy of 0.06 from Ciccone and Hall (1996). In addition, Diamond (2016) estimates an elasticity of MSA wages with respect to population by skill group. As detailed in Appendix B.2, under the previous normalization, these estimates can be mapped to the relative values of our $\gamma^{P}_{\theta,\theta'}$ parameters using the wage equation (14) and the elasticity of substitution between skilled and unskilled workers ρ.

As a result we obtain $\left(\gamma^{P}_{UU}, \gamma^{P}_{SU}, \gamma^{P}_{US}, \gamma^{P}_{SS}\right) = (.003,\ .044,\ .02,\ .053)$. This approach preserves an aggregate elasticity of labor productivity with respect to density that is consistent with standard estimates. It is also consistent with the cross-spillover elasticities implied by Diamond (2016), who recovers there cross-spillovers from the elasticity of city-level wages by skill group with respect to the supply of workers of each skill. These parameters imply stronger efficiency spillovers generated by high skill workers, and close to zero spillovers from low skill workers.[29]

Amenity Spillovers $\left\{\gamma^{A}_{\theta',\theta}\right\}$ Diamond (2016) estimates elasticities of labor supply by skill group with respect to an MSA-level amenity index that includes congestion in transport, crime, environmental indicators, supply per capita of different public services, and variety of retail stores. She estimates a higher marginal valuation for these amenities for college than for non-college workers. In addition, she estimates a positive elasticity for the supply of this MSA-level amenity index with respect to the relative supply of college workers. As detailed in Appendix B.2, we can combine these estimates and map them our amenity spillovers $\gamma^{A}_{\theta',\theta}$ using the labor-supply equation implied by the spatial mobility constraint (13). As a result we obtain $\left(\gamma^{A}_{UU}, \gamma^{A}_{SU}, \gamma^{A}_{US}, \gamma^{A}_{SS}\right) = (-.43,\ .18,\ -1.24,\ .77)$. These parameters imply positive amenity spillovers generated by high skill workers, and negative spillovers generated by low skill workers.[30]

[28] Most of the studies reviewed by Combes and Gobillon (2015) and Melo et al. (2009) also fall in this range.

[29] Micro studies of peer effects note that policies designed to implement an optimal mixing of heterogeneous workers may deliver undesired outcomes due to endogenous group formation decisions after the policy is implemented (e.g., Carrell et al., 2013). Our city-level analysis abstract from these considerations.

[30] At these values, all but one of the concavity conditions implied by Proposition 3 are satisfied. Specifically,

Alternative Parametrizations of the Spillover Elasticities We implement all our counterfactuals under different parametrizations of the spillover elasticities. The alternatives deviate from the benchmark described so far in terms of the efficiency or amenity spillover elasticities. In particular, we implement the model under: i) a more conservative parametrization that scales down the amenity spillover elasticities $\gamma^A_{\theta,\theta'}$ by 50% (referred to as the "Low amenity spillover" parametrization); ii) mappings of the amenity spillovers $\gamma^A_{\theta,\theta'}$ assuming values of the elasticity of city amenities to the share of college workers that are either one standard deviation above or below Diamond (2016) point estimates (referred to "High cross amenity spillover" and "Low cross amenity spillover" parametrizations, respectively); iii) a less conservative parametrization that scales up the efficiency spillover elasticities $\gamma^P_{\theta,\theta'}$ to 0.12, i.e., twice the benchmark of 0.06 from Ciccone and Hall (1996) (referred to "High efficiency spillover" parametrization); iv) a more conservative parametrization that scales down the efficiency spillover elasticities by a factor 2 (referred to "Low efficiency spillover" parametrization); and v) parametrizations of efficiency spillovers that correspond to alternative values of the complementarity parameter ρ, as detailed in the Online Appendix C.5. The values of these alternative parametrizations are reported in Appendix B.2.

4.3 Stylized Facts

Figure 1 revisits standard stylized facts on spatial disparities and sorting in the data, as well as a relatively less known fact on the spatial structure of net transfers between cities. These facts will serve as a benchmark to evaluate the impact of optimal spatial policies.

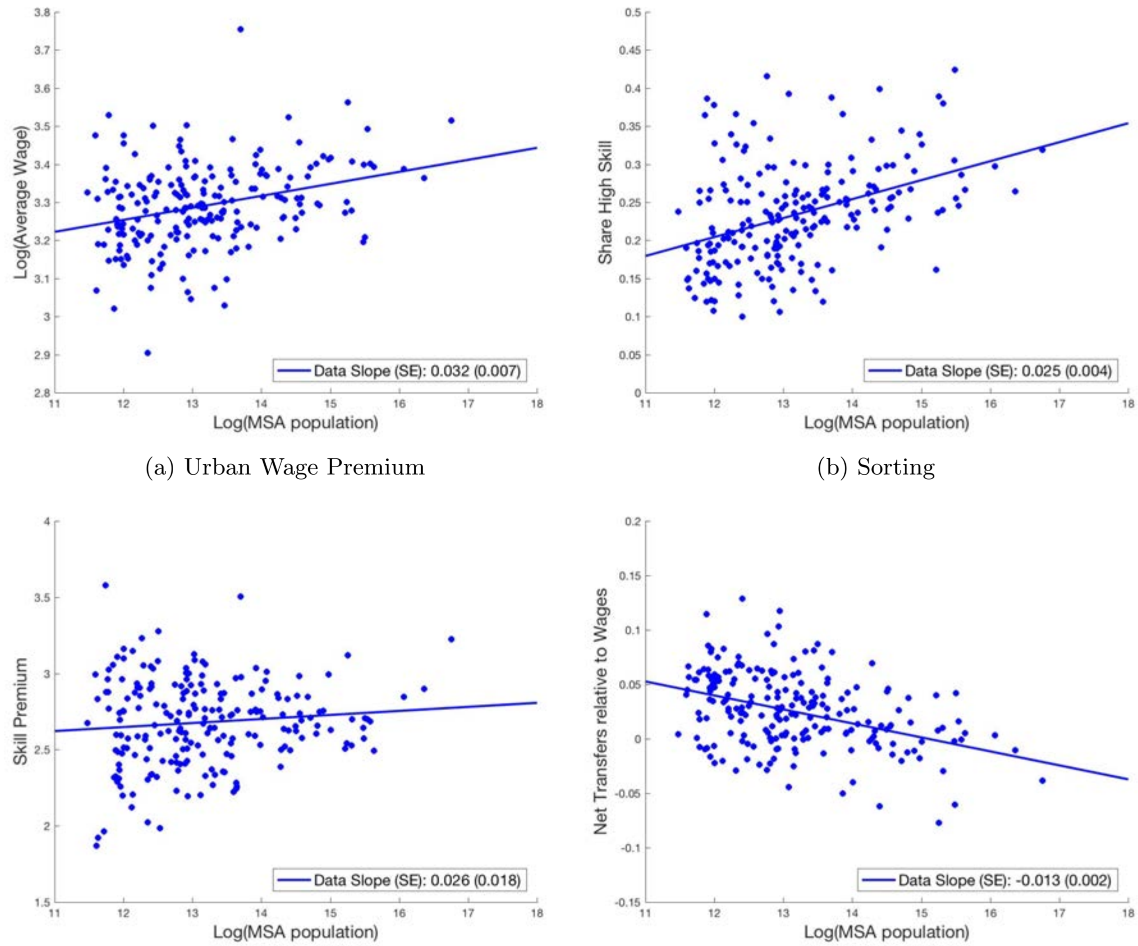
Panels A to C show the standard facts about spatial disparities and sorting as function of city size, or "urban premia". Panel A documents the urban wage premium, defined as the increase in average nominal wages with city size. The elasticity of wages to city size is 3.2%. Panel B shows spatial sorting, in terms of the share of high-skill workers. The semi-elasticity of the share of high skill workers with respect to city size is 2.5%. I.e., doubling population increases the skill share by 2.5 percentage points. Panel C shows the urban skill premium, defined as the increase in the ratio of high- to low-skill wage as city size increases. The slope of 0.03 means that larger cities feature a more unequal nominal wage distribution. The first fact suggests differences in productivity and cost of living across cities, while the last two suggest complementarities between city size and skill.

Panel D shows a somewhat less known fact, the relationship between city size and net imbalances.[31] For each city we construct the net imbalance as the difference between expenditures and total income (from labor and non-labor sources). The graph shows net imbalance relative to labor income at the MSA level across MSAs. Given our construction of the expenditure variable, these differences in imbalances across cities result purely from the government policies that we

the conditions that $\Gamma^A > \Gamma^P$, $\Gamma^A > 0$, and $\gamma^P_{\theta,\theta'} > 0$ for $\theta \neq \theta'$ are all satisfied, as well as the condition that $\gamma^A_{SU} > 0$. However, our parametrization sets $\gamma^A_{US} < 0$. In principle, therefore, concavity of the planner's problem is not guaranteed. However, in the quantitative exercise we check for the possibility of multiple local maxima by repeating the welfare maximization algorithm starting from 100 spatial allocations taken at random. Reassuringly, we fail to find any alternative local maximum.

[31] Albouy (2009) shows that high wage cities contribute disproportionally to the federal tax burden in the U.S.

Figure 1: Urban Premia



(a) Urban Wage Premium

(b) Sorting

(c) Urban Skill Premium

(d) Net Transfers

Note: each figure shows data across MSAs. All the city level outcomes reported on the vertical axis are adjusted by socio-demographic characteristics of each city, as detailed in Appendix B.1.

measure (taxes and transfers). The negative slope reflects that government policies redistribute income from larger, high wage, high skill cities to smaller, low wage, low skill cities. These transfers are net of compositional effects according to detailed demographic characteristics in IPUMS, as mentioned above. Therefore, distributive government policies that vary with these characteristics across individuals do not underlie these patterns across cities.

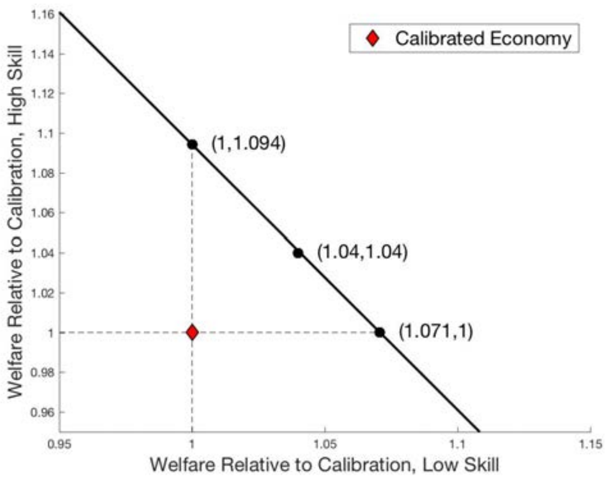
5 Optimal Spatial Policies in the U.S. Economy

5.1 Optimal Transfers, Reallocations, and Welfare Gains

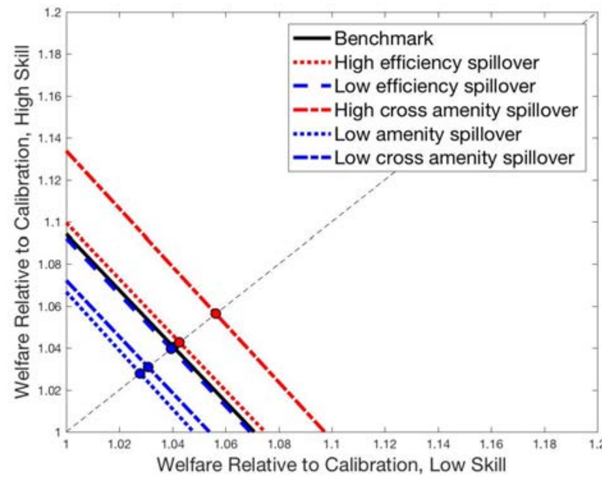
We now explore the implications of optimal spatial policies using the calibrated model. We solve the planner's problem in changes relative to the observed equilibrium described in Section

Figure 2: Utility Frontier of the U.S. Economy between High and Low Skill Workers

(a) Benchmark



(b) Alternative Parametrizations



The figure shows the optimal welfare changes $\left(\hat{u^L}, \hat{u^H}\right)$ between the optimal and observed allocation, corresponding to the solution of the planner's problem in relative changes described in Appendix A.9. Each point corresponds to a maximization of $\hat{u^H}$ subject to a different lower bound on $\hat{u^L}$. The benchmark parametrization on the left panel corresponds to the black line on the right panel. The circles in the right panel represent intersections with the 45 degree line where the welfare of skilled and unskilled workers increase by the same amount.

3.6. We maximize over the change in utility of skilled workers, $\hat{u^S}$, subject to a lower bound for the change in utility of unskilled worker, $\hat{u^U}$. Varying this lower bound traces the Pareto frontier.

Aggregate Welfare Gains The left panel of Figure 2 shows the utility frontier of the U.S. economy in the benchmark parametrization, expressed in changes relative to the observed equilibrium. The point (1,1) represented with a red diamond corresponds to allocations where the welfare of skilled and unskilled workers is unchanged compared to the calibrated equilibrium. When the welfare gain of unskilled and skilled workers is restricted to be the same, optimal transfers lead to a 4% welfare gain for both types of workers. When only the welfare of one group is maximized subject to a constant level of welfare for the other group, we find gains of 9.4% for high skill workers and of 7.1% for low skill workers.

The right panel of Figure 2 shows the utility frontier for the benchmark and for each of the alternative parametrizations discussed in Section 4.2. The frontier shifts up and down with little change in slope. The welfare gains from implementing optimal policies are larger in the two frontiers in red, corresponding to high efficiency and amenity spillovers. The gains are lower with low amenity spillovers. Table 1 shows the welfare gains corresponding to the intersection between these frontiers and the 45 degree line, such that skilled and unskilled workers gain the same. Across these specifications, the common welfare gains range from roughly 2% to 6%. Lowering the amenity spillover by 50% brings the common welfare gain down to 2.8%, while multiplying the efficiency spillovers by 2 increases the gain to 4.3%.

Hence, we find sizable welfare gains from the optimal spatial reallocation. Inefficiencies in sorting are a key driver of this magnitude. With homogeneous workers, the welfare gains from implementing the optimal allocation are negligible at 0.06%. Similarly, implementing the analysis on counterfactual data without differences across skill groups (with no spatial sorting by skill, no urban skill premium, and no relative differences in expenditures), the welfare gains fall to 0.25%.[32] Accounting for skill heterogeneity is therefore important for the aggregate welfare effects of spatial policies. Our results also suggest significantly higher welfare gains compared to estimates of removing dispersion in spatial polices or other spatial wedges in the U.S.[33]

Table 1: Welfare gains under different levels of the spillovers

	Spillovers	Welfare Gain (%)
(1)	Benchmark	4.0
(2)	High efficiency spillover	4.3
(3)	Low efficiency spillover	3.9
(4)	Low amenity spillover	2.8
(5)	High cross-amenity spillover	5.6
(6)	Low cross-amenity spillover	3.1
(7)	Smaller ρ	2.4-3.9

The table reports the common welfare gains for skilled and unskilled workers under alternative parametrizations described in Section 4.2. Row (2) corresponds to $\gamma^P_{\theta'\theta}$ that are twice as large as in the benchmark. Row (3) corresponds to $\gamma^P_{\theta'\theta}$ 50% lower than the benchmark. Row (4) corresponds to $\gamma^A_{\theta',\theta}$ 50% lower than the benchmark. Rows (5) and (6) are configurations assuming higher or lower cross-amenity spillovers corresponding to the plus or less one standard deviation of the estimates in Diamond (2016). See Appendix B.2 for details on these parametrizations. Row (7) corresponds to efficiency spillovers calibrated using different values of the production function parameter ρ, as detailed in Online Appendix C.5.

Actual versus Optimal Transfers How does the optimal spatial income redistribution compare to the data? Let t^θ_j be the optimal transfers received by type θ according to (20) in Proposition 2. Figure 3 shows the net transfers per capita relative to wages t^θ_j/w^θ_j by MSA and worker type on the vertical axis, against the wage w^θ_j of each MSA in both the data (blue circles) and the optimal allocation (red diamonds), for low skill workers (hollow markers) and high skill workers (solid markers). We represent the optimal allocation corresponding to the point on the Pareto frontier in the left panel of Figure 2 where welfare gains are equal for both types of workers.[34]

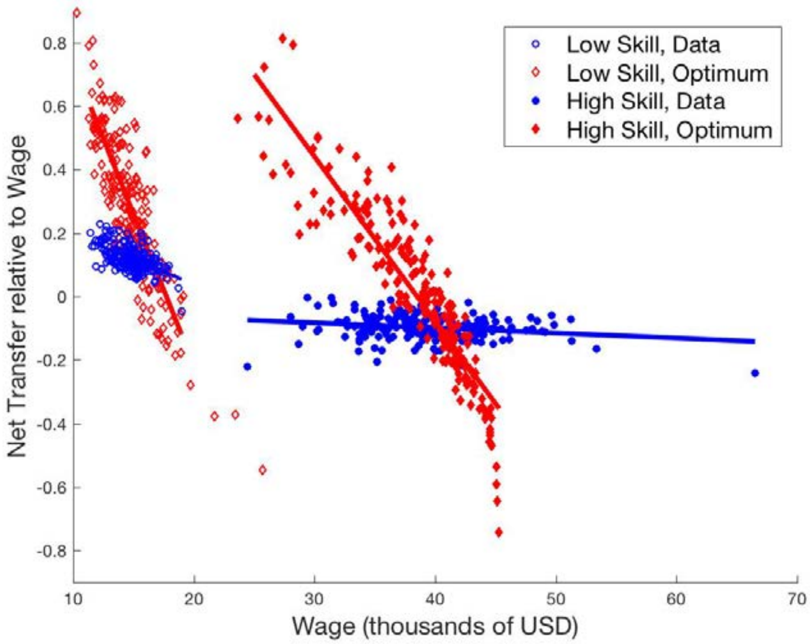
The transfers in the data present a clear pattern of redistribution from high skill workers and high-wage cities towards low skill workers and low-wage cities. Net average transfers are positive for

[32]Figure B.1 in Appendix C.1 shows that, assuming homogeneous workers, the observed transfers across MSAs in the optimal allocation are quite close the data. Figure B.2 shows that the welfare gains can be substantial under counterfactual data with high wage dispersion. Section C.1 in the Online Appendix describes the details of the calibration with homogeneous workers.

[33]Desmet and Rossi-Hansberg (2013) find welfare gains of 0.9% from eliminating frictions across U.S. cities, Albouy (2009) finds losses of 0.2% from the tax dispersion created by federal income taxes, and Fajgelbaum et al. (2018) find gains of of 0.6% from harmonizing state taxes. The small welfare gains to optimal reallocation without worker heterogeneity are in line with results in Eeckhout and Guner (2017) and Ossa (2018).

[34]The main impact of a different Pareto weight is to shift the transfer schedules up and down depending on the Planner's preference for each group, without changing the qualitative patterns we discuss.

Figure 3: Per Capita Transfers by Skill Level and MSA, Data and Optimal Allocation



Note: each point in the figure corresponds to an MSA-skill group combination. The vertical axis shows the difference between the average transfer relative to wage and the horizontal axis shows the average wage. For details of how the data is constructed see Appendix B. The slopes of each linear fit (with SE) are: Low Skill, Data: -0.02 (0.001); Low Skill, Optimum: -0.095 (0.004); High Skill, Data: -0.002 (0.001); High Skill, Optimum: -0.05 (0.002). The figure corresponds to planner's weights such that both types of workers experience the same welfare gain in Figure 2.

low skill workers and negative for high skill workers in most MSAs. Within skill groups, net transfers decrease with the wage of the MSA.[35] The observations in red show the efficient allocation, which satisfies the optimality condition from Proposition 1. Across cities, the optimal transfers relative to labor income decrease more steeply with wages than in the data for both labor types, implying a stronger redistribution towards low-wage cities than what is observed empirically.

To understand what drives these optimal transfers, we return to the expression for optimal subsidies (21). The first term of (21) is driven by own spillovers, while the second term is shaped by cross spillovers. In our parametrization of spillovers for low skill workers, both of these terms are negative. The negative cross-spillovers through amenities lead to the higher tax of low skill workers in large, high-wage cities where a larger share of expenditures accrues to high skill workers. The logic that rationalizes a higher labor tax in high-wage cities is different for high skill workers. In our parametrization, high skill workers generate positive own spillovers. According to the first term in (21), these positive spillovers would call for a labor income *subsidy*. However, this force is more than offset by strong positive cross spillovers onto low skill workers, which calls for more

[35]On average across MSAs, they equal 1.8 thousand dollars for low-skill workers, or 12% of their average wage. For high skill workers, the corresponding numbers are -3.8 thousand dollars or -10% of the average wage. In cities where high skill workers earn on average more than $50k per year, net transfers of high skill workers are -8.9 thousand dollars or -15% of wages.

mixing of high-skill workers with low-skill workers. A higher tax in high-wage cities directs skilled workers into small, low-wage cities that are relatively abundant in low skill workers.

While both low and high skill workers are on average reallocated towards lower-wage cities, it is a priori ambiguous for which group this effect is stronger. We examine the question of optimal sorting below.

Optimal Reallocation and Sorting The optimal transfers change the spatial distribution of economic activity compared to the data. By changing the location incentives of workers, they affect spatial sorting and the city size distribution. These reallocations in turn impact labor productivity and wages through agglomeration spillovers, and the distribution of urban amenities through amenity spillovers. These effects feed back to location choices, changing the spatial pattern of skill premia and inequality. We now describe the spatial equilibrium resulting from this process.

Figure 4 shows the pattern of reallocation. First, Panel (a) shows the initial total population of each MSA on the horizontal axis and the change in population implied by the optimal allocation relative to the initial allocation on the vertical axis, defined as $\hat{L_j} - 1$. The stronger redistribution to low-wage locations discussed in the previous section implies that, on average, there is reallocation from large to small cities. However, there is also considerable heterogeneity in growth rates over the size distribution, including middle- and small-MSAs that shrink alongside large MSA's that grow, so that initial city size is a poor predictor of whether a city is too large or too small in the observed allocation (the R^2 of the linear regression is 15%).[36]

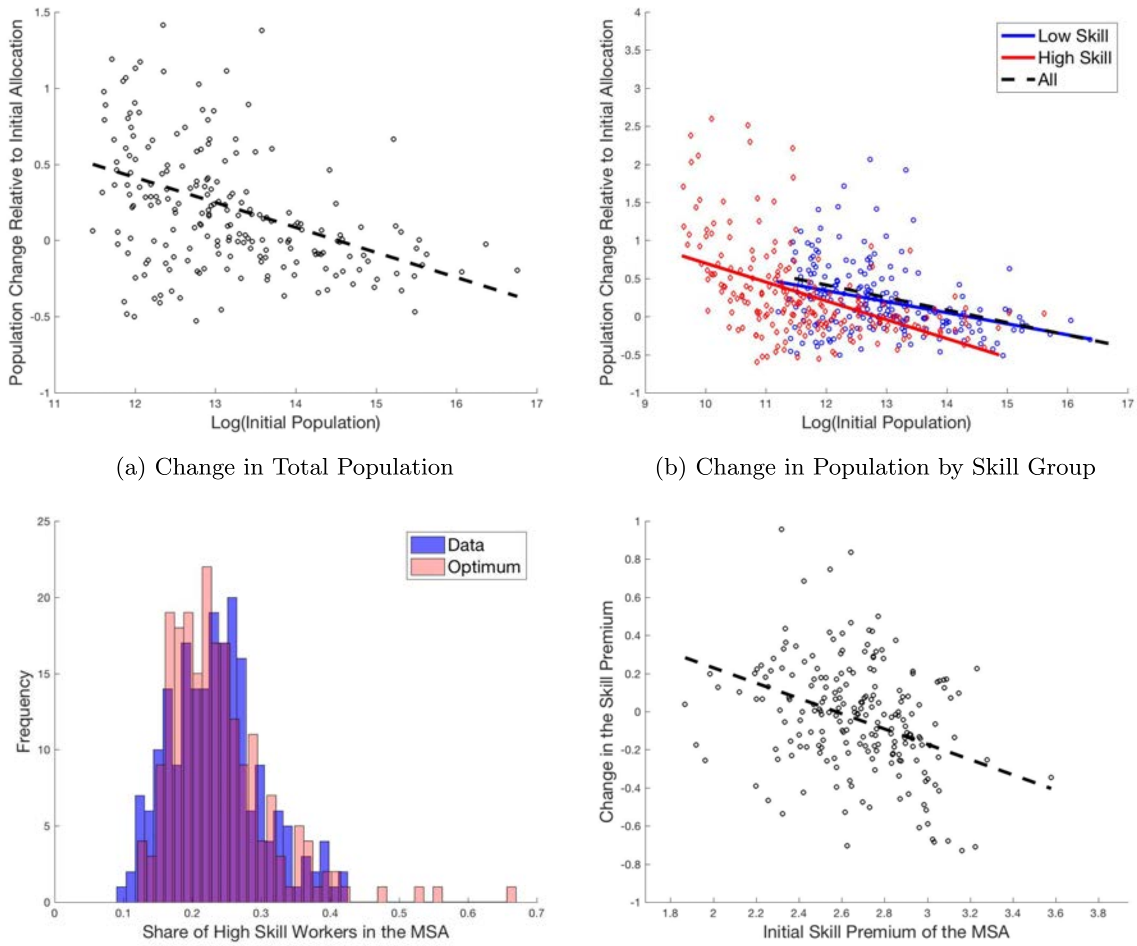
Second, panels (b) and (c) illustrate changes in sorting patterns. Panel (b) shows changes in population by skill, alongside the linear fit from panel (a), while panel (c) shows the histogram of skill shares across MSAs in the initial and optimal allocation. On average, reallocations towards initially smaller places is stronger within the high-skill group. As a result, the skill share distribution becomes more compressed at the bottom of the distribution (panel (c)). However, the optimal reallocations also result in more intensively high-skilled cities at the top of the distribution. These shifts reflects that the share of high-skill workers grows both in cities with initially very low skill share and in some large cities with very high skill share.[37]

At the same time, we find in panel (d) that the skill premium tends to increase in initially less unequal cities, which tend to be smaller cities, and to decrease in initially more unequal and larger cities. Together with the sorting patterns described above, this result suggests that two different mechanisms drive the optimal sorting by skill. At the bottom of the city size distribution, optimal sorting is dominated by the positive cross-spillovers generated by high-skill workers on low-skill workers. At the top, optimal sorting is driven by positive amenity spillovers generated by high-skill workers on their own group. This force leads to higher skill concentration in those locations, but also to a lower skill premium.

[36]Albouy et al. (2019) and Eeckhout and Guner (2017) argue that large cities are too small in models with homogeneous workers, one-dimensional heterogeneity and spillover elasticities only.

[37]This pattern is illustrated in Figure A.1 in Appendix B.2. Weighting by initial population MSA, the relationship between initial skill share and optimal growth in the skill share is U-shaped.

Figure 4: Changes in Population, Skill Shares, and Skill Premium across MSAs



(a) Change in Total Population

(b) Change in Population by Skill Group

(c) Histogram of High-Skill Shares across MSAs

(d) Change in Skill Premium

Note: Panel (a) shows the change in population between the optimal allocation and the initially observed equilibrium and the linear fit. Slope (SE): -0.16 (0.03); R^2=0.15. Panel (b) displays the same outcomes for high and low skill workers. Slopes (with SE): High Skill: -0.25 (0.03); Low Skill: -0.15 (0.03). Panel (d) displays in the vertical axis the difference in the skill premium between the optimal and initial allocation. Slope (SE): -0.4 (0.07). The figures correspond to planner's weights such that both types of workers experience the same welfare gain in Figure 2.

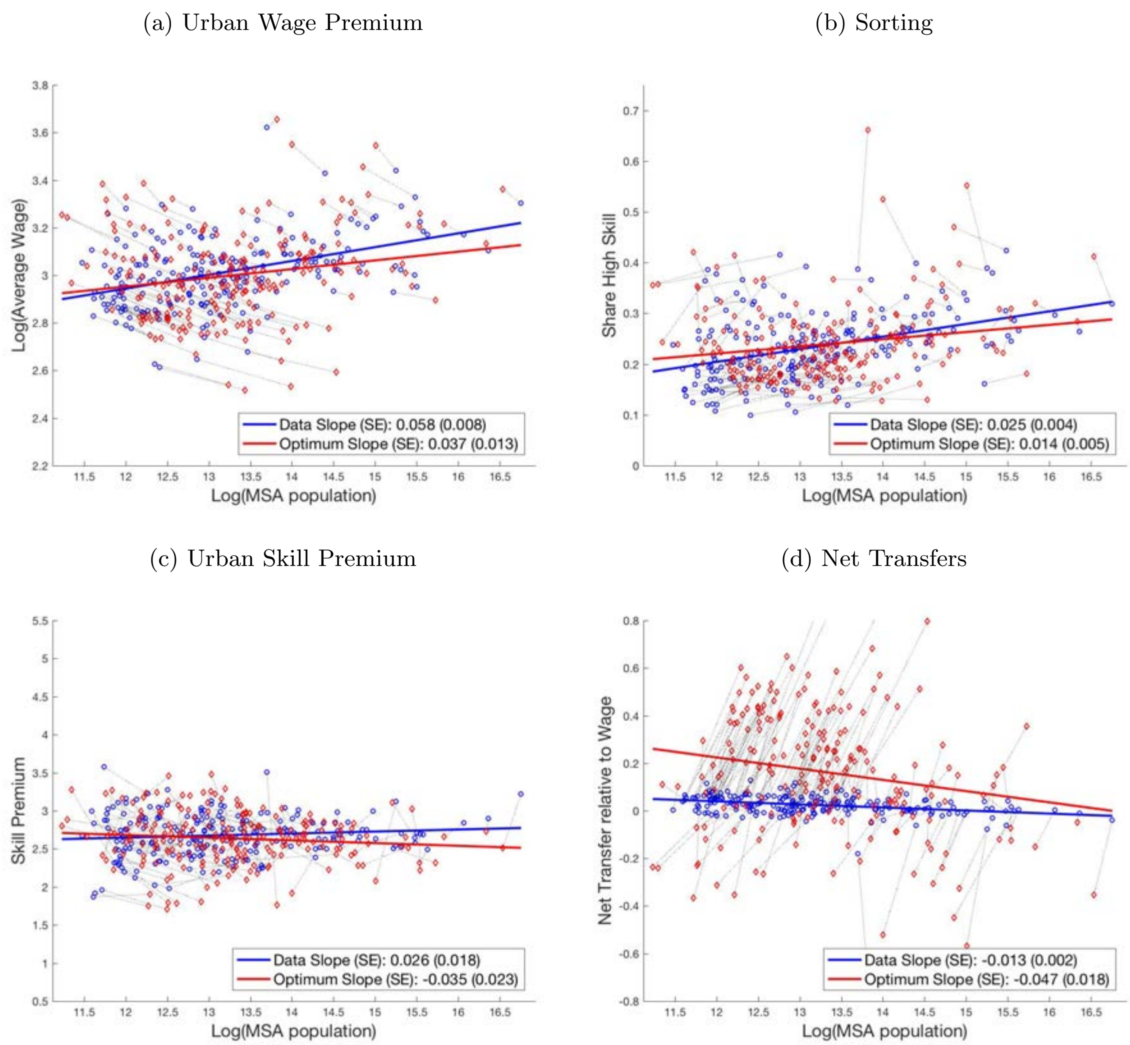
The Urban Premia in the Optimal Allocation We now revisit the stylized facts about urban premia presented in Section 4. Figure 5 reproduces Figure 1. Each pair of linked observations corresponds to the same MSA in the data and in the optimal allocation.[38] The optimal allocation features a higher absolute value of the imbalances at the city level (panel (d)), since redistribution to smaller MSAs is stronger in the optimal allocation. As discussed above, the optimal allocation features more mixing of high and low skills on average (panel (b)), but at the same time the initially largest MSAs become more skill-intensive. Specifically, 8 of the 10 largest cities increase their skill

[38]Here, we compare the data to an optimal allocation corresponding to the same welfare gains to all workers. The patterns of urban premia are almost identical as we move to extreme points of the utility frontier, because these points are implemented through lump-sum transfers across types which have small effects on the urban premia. These patterns are also similar under alternative parametrizations of the spillovers from Table 1.

share.[39] The urban skill premium vanishes (panel (c)), implying that the sorting pattern from panel (b) ends up being detached from the urban skill premium. Instead, it is driven by stronger preferences for urban amenities among high skill workers. As seen in panel (a), the wage premium in the large cities is still noticeable, but lower than in the data. It is driven by an average productivity advantage across both skill groups in larger cities, rather than by a relatively higher productivity of high-skill workers in these places.

Regional Patterns Figure 6 shows the growth in population (left panel) and skill shares (right). Cities are weighted by initial population, with darker red circles representing more positive growth.

Figure 5: Urban Premia, Data and Optimal Allocation



Note: each panel reports outcomes across MSAs in the data and in the optimal allocation. Each linked pair of observations corresponds to the same MSA.

[39]If the top 10 cities are excluded, the relationship between the share of high-skill workers and MSA population in the optimal allocation becomes flat.

As the economy moves to the optimal allocation, population tends to be reallocated away from coastal regions. For example, in California cities like Los Angeles and San Francisco loose population while smaller cities inland next to them grow. In terms of the skill shares, consistent with the observation in panel (b) of Figure 5, we observe that the largest urban centers such as Los Angeles, New York and San Francisco become more skill intensive despite loosing population, reflecting the higher preferences of high-skill workers for those locations. However, many other small cities also grow in their skill share, driving down the average urban skill share in Panel (b) of Figure 5.

5.2 Inferring the Spillover Elasticities assuming Efficiency in the Data

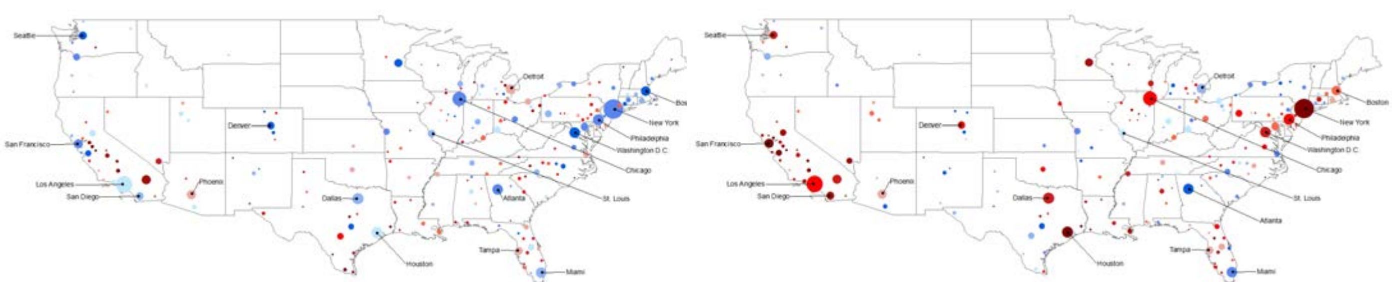
Our logic so far was to discipline the model with existing estimates of the spillover elasticities, and then use it to compute the efficient allocation. We now invert this logic, and instead ask: what spillover elasticities would be consistent with assuming that the observed spatial allocation is efficient? By comparing these inferred spillover elasticities with those used in the calibration, this exercise allows us to identify the key elasticities behind our results.

Proposition 4 establishes that any observed allocation can be rationalized as an equilibrium from the model. However, nothing guarantees that an observed allocation can be rationalized as an efficient equilibrium for some set of spillover elasticities. Therefore, for this exercise, we have to make further assumptions. First, we assume that there is measurement error in the data. Second, we assume that the elasticities are constant. Assuming that the observed allocation is optimal, the condition on optimal transfers (20) must hold. Combined with the definition of expenditure per worker in (16), we obtain the following optimal relationship between transfers, wages, expenditures,

Figure 6: Optimal Reallocations

(a) Population

(b) Skill Share



The maps show the growth in population (left panel) and share of college workers (right panel) from the observed to the optimal allocation. Cities are weighted by initial population. Red means positive growth and blue is negative growth.

and employment:

$$t_j^\theta = a_0^\theta + a_1^\theta w_j^\theta + a_2^\theta \left(\frac{w_j^{\theta' \neq \theta} L_j^{\theta' \neq \theta}}{L_j^\theta} \right) + a_3^\theta \left(\frac{x_j^{\theta' \neq \theta} L_j^{\theta' \neq \theta}}{L_j^\theta} \right) + \varepsilon_j^\theta, \tag{44}$$

for $\theta \in \{U, S\}$, where ε_j^θ is a measurement error term, and the reduced-form parameters have the following structural interpretations: $a_0^\theta \equiv -b^\theta \Pi^* - \frac{E^\theta}{1-\gamma_{\theta,\theta}^A}$, $a_1^\theta \equiv \frac{\gamma_{\theta,\theta}^P + \gamma_{\theta,\theta}^A}{1-\gamma_{\theta,\theta}^A}$, $a_2^\theta \equiv \frac{\gamma_{\theta,\theta'}^P}{1-\gamma_{\theta,\theta}^A}$, and $a_3^\theta = \frac{\gamma_{\theta,\theta'}^A}{1-\gamma_{\theta,\theta}^A}$. We estimate the parameters $\left\{a_i^\theta\right\}$ by running (44) as a regression in the cross-section, and then infer the spillover elasticities $\left\{\gamma_{\theta,\theta'}^A, \gamma_{\theta,\theta'}^P\right\}$ up to a normalization for each type.[40] We normalize the own-spillover elasticity for productivity to the benchmark level for the U.S. used in Section 4.2.

This exercise yields $\left(\gamma_{UU}^A, \gamma_{SU}^A, \gamma_{US}^A, \gamma_{SS}^A\right) = (-.09,\ -.16,\ .06,\ -.32)$ and $\left(\gamma_{UU}^P, \gamma_{SU}^P, \gamma_{US}^P, \gamma_{SS}^P\right) = (.003,\ .20,\ .-08,\ .053)$.[41] The average level of both types of spillovers are similar to the parameters implied by the empirical estimates used in the calibration. In both these inferred elasticities and the calibrated ones, the amenity spillovers are larger than the agglomeration spillovers, and high-skill workers generate stronger efficiency spillovers than low-skill workers. However, the assumption that the observed allocation is optimal implies negative amenity spillovers both across and within skill groups, whereas the calibrated elasticities imply positive amenity spillovers generated by high skilled workers. Therefore, we find that heterogeneity in spillovers across groups is key for spatial policies, as previously suggested by the contrast between the quantified model under homogeneous and heterogeneous workers.

5.3 Alternative Specifications

To gauge the sensitivity of our findings, we now turn to implementing the calibration and counterfactuals for alternative specifications. Each of these cases formally extend our benchmark quantification. We re-calibrate the model each time, compute the welfare gain common to all workers on the utility frontier, and compare it to the benchmark case. We defer the details of the implementation to the online appendix.

Land Use Regulations Several papers (Bunten, 2017; Herkenhoff et al., 2018; Hsieh and Moretti, 2019; Parkhomenko, 2018) argue that local land use regulations create spatial distortions by lowering the housing supply elasticity. In our benchmark procedure, we have interpreted the housing supply elasticity as a technological restriction in the planner's problem. We now extend the model

[40] This normalization is needed because from (44) the own-spillover elasticities for productivity and amenities are not separately identified. Assuming values for $\gamma_{\theta,\theta}^P$ we can then infer the remaining elasticities as follows: $\gamma_{\theta,\theta}^A = \frac{a_1^\theta - \gamma_{\theta,\theta}^P}{1+a_1^\theta}$, $\gamma_{\theta,\theta'}^P = a_2^\theta\left(1-\gamma_{\theta,\theta}^A\right)$, and $\gamma_{\theta,\theta'}^A = a_3^\theta\left(1-\gamma_{\theta,\theta}^A\right)$.

[41] The regressions have an R-squared of 0.32 for high skill and of 0.15 for low skill. However, when we use the revealed-optimal elasticities to compute the efficient allocation relative to the observed allocation, the procedure confirms that such an allocation is very close to optimal. In that case, we obtain negligible welfare gains of 0.07%. If the R-squared of the procedure had been 1, there would have been no further welfare gains to achieve.

to capture the notion that the housing supply elasticity can be endogenous to local regulations, and to allow the federal planner to change these regulations. We model land use regulations as a local tax rate imposed on the sales of non-traded goods in each city j:

$$1 - \frac{1}{1-\tau_{H,j}} \left(R_j H_j\right)^{-\tau_{H,j}} \tag{45}$$

As a result, the housing supply elasticity becomes:

$$\frac{\partial \ln H_j}{\partial \ln R_j} = \frac{1-\tau_{H,j}}{d_{H,j}+\tau_{H,j}}. \tag{46}$$

This specification microfounds a housing supply elasticity that includes both a technology constraint $d_{H,j}$ due geographic characteristics as in Saiz (2010) as well as land regulations $\tau_{H,j}$ as in the previous papers. The higher the parameter $\tau_{H,j}$, the lower the housing supply elasticity compared to its undistorted level. Our benchmark parametrization is nested when $\tau_{H,j} = 0$ for all locations, in which case there is a zero tax rate.

We evaluate the welfare effects of two policy exercises: (i) implementing optimal transfers while keeping local taxes $\tau_{H,j}$ unchanged ($\hat{\tau}_{H,j} = 1$); and (ii) implementing optimal transfers while at the same time removing distortions ($\hat{\tau}_{H,j} = 0$). The first exercise asks whether accounting for wedges in the initial allocation due to land regulations matters for the welfare gains from implementing optimal transfers designed to deal with spillovers. In turn, by construction, the second exercise must deliver greater gains than implementing optimal transfers alone.

The results are presented in rows (2) and (3) of Table 2. Implementing optimal transfers while keeping the initial distortions lowers the welfare gains to 3.7% from 4.0%. Hence, accounting for land regulations does not fundamentally affect the gains from optimal redistribution. However, row (3) shows that removing land distortions on top of implementing optimal transfers more than doubles the welfare gains compared to leaving local regulations unchanged. This result suggests that both margins (optimal redistribution, and land use regulations) are roughly equally important sources of misallocation.

Table 2: Welfare gains of Implementing Optimal Transfers under alternative specifications

	Cases	Welfare Gain (%)
(1)	Benchmark	4.0
(2)	Land Regulations, keeping distortions	3.7
(3)	Land Regulations, removing distortions	8.6
(4)	Three skill groups	3.9
(5)	Imperfect Mobility	4.3

Note: The table shows the welfare gains from implementing the optimal transfers in different parametrizations. We report the common welfare gains to all workers on the utility frontier. See online Appendix for details.

Multiple Skills with Non-Homothetic Production The benchmark calibration features two skill groups (college and non-college graduates). We now implement an extension with three skill groups. Instead of the aggregator (39) applied to unskilled and skilled workers, we model three skill groups indexed by their ability, $\theta = \{L, M, H\}$ standing for low-, medium-, and high-skill workers. Their output is aggregated to the city level according to:

$$N_j = \left(\left(z_j^L L_j^L \right)^\rho + \left(z_j^H L_j^H \right)^\rho \right)^\lambda + \left(z_j^M L_j^M \right)^\rho . \tag{47}$$

This production function follows Eeckhout et al. (2014), who propose this nesting to capture that larger cities disproportionally attract both high- and low-skill workers, while smaller cities feature relatively more medium-skill workers. Assuming $\lambda > 1$, this production function is non-homothetic between the medium-skill workers and the nest of low and high-skill workers. Hence, as production increases, the relative demand for the second group increases. Empirically, we define high skilled workers H in the same way as the skilled workers in our two-groups case, but split our previous group of unskilled workers (without complete college) into those with some college education (M) and those with no college education (L). We continue to assume the same structure for the spillovers as in our benchmark case, on the basis of $U = \{L, M\}$ and $S = \{H\}$ types. As shown in row (4) of Table 2 the welfare effects are very similar to the benchmark case, while Figure B.3 in the online appendix shows that the patterns of transfers and reallocation are also similar. The optimal transfers on average reallocate workers to smaller cities but even more so for skilled workers, without a strong difference between the reallocation patterns of low- and medium-skilled workers.

Imperfect Mobility Our benchmark case assumed that workers are perfectly mobile across regions. We now incorporate two forces to account for imperfect mobility. First, we redefine a type θ to include not only a worker's skill but also her region of origin $o \in \mathcal{O}$. Workers from different origins may vary in their preference for locations and productivity. Specifically, to account for migration frictions, we assume that a worker may face a disutility cost from living in a place different from her region of origin. This additional margin of heterogeneity allows the model to capture a salient fact from the data, namely that that place of birth is a strong predictor of region of residence. In production, we assume that workers with the same skill level are perfect substitutes regardless of origin. Second, following our discussion in Section (3.5), we also incorporate preference draws within types according to a Fréchet distribution with parameter σ_θ.[42] Turning to the quantification, we classify workers as being born in one of 5 different Census regions, and compute the welfare gains of implementing optimal transfers taking into account heterogeneous preferences for location of workers of different origins. As shown in Table 2, we find welfare gains across all groups of 4.3%, close to the 4% from the baseline case. Furthermore, once aggregated by skill across origins, the reallocation patterns are also similar to the baseline case. We conclude that the main takeaways of

[42] This formulation nests our benchmark specification in the case of a single origin of workers and $\sigma_\theta \to 0$. Because we have assumed that workers are perfect substitutes in production regardless of origin, the curvature introduced by these draws allows us to pin down the number of workers from each origin living in a given destination. Formally, these draws introduce a notion of congestion at bilateral level. An alternative assumption leading to a similar property would have been assume that workers of different origins are imperfect substitutes in production. Our current specification with extreme-value draws is closer to static models capturing migration frictions such as Bryan and Morten (2015) and Diamond (2016).

the benchmark analysis are robust to incorporating this form of mobility frictions.

Other specifications We have also implemented the analysis under additional alternative assumptions. First, our theoretical results imply that matching the observed expenditures distribution is relevant. Indeed, when we ignore the transfers in the data and set worker expenditures equal to income, the welfare gains increase to 6.3% from 4% in the baseline.[43] Second, we re-do the quantification assuming that the returns to fixed factors are locally distributed to residents of each location.[44] Our theoretical discussion from Section 3.4 shows that this assumption entails an additional distortion. Consistent with this result, we find that the common welfare gains of implementing optimal expenditures increases to 4.9% relative to 4% in the baseline. Finally, the welfare results are quantitatively very close to the baseline if we assume away trade costs. In this case, we use counterfactual data in which expenditure shares are equally distributed across cities of origin, rather than relying on bilateral trade shares that decay with distance as in our baseline quantification. The reason why the welfare implications of both quantifications are very similar is that the procedure fully recalibrates the model (including amenities and productivity), so that wages, transfers and employment are perfectly matched in all cities in both cases. These moments play a key role in pinning down the potential welfare gains of moving to an efficient allocation.

6 Conclusion

We study optimal policies in a spatial framework with spillovers and sorting of heterogeneous workers. The framework accommodates many key determinants of the spatial distribution of economic activity such as geographic frictions and asymmetric amenity and productivity spillovers across workers.

We derive the set of optimal transfers across workers and regions. There exists scope for welfare-enhancing spatial policies even when spillovers are common across locations. In that case, constant labor income subsidies and lump-sum transfers over space implement the efficient allocation, regardless of micro heterogeneity in fundamentals. When workers are heterogeneous and there are spillovers across different types of workers, spatial efficiency requires place-specific subsidies to attain optimal sorting.

We then show how to use the framework to assess the efficiency of the observed spatial allocation. The distributions of fundamentals needed to compute the optimal allocation can be backed out from data on wages, employment, and expenditures across worker types and regions. We apply the model to the distribution of economic activity across MSAs in the U.S. using existing estimates of the

[43]Because the transfers tend to be negative in larger cities, ignoring transfers leads to an under-estimation of the amenity levels implied by the model in larger cities.

[44]The weak correlation between capital income in the data and a proxy for housing profits across cities computed as $\gamma_j/(\gamma_j+1)X_j$, where X_j is total expenditure in the city from the data and γ_j is the housing supply elasticity in city j, suggests that the assumption of common ownership is a reasonable benchmark. Other assumptions on the distribution of profits with some degree of local ownership generate an inefficiency. Results are formally equivalent under local ownership and in a model with absentee landlords where the planner maximizes welfare of workers.

spillover elasticities. The results suggest that inefficient sorting may lead to substantial welfare costs. Spatial efficiency calls for more redistribution to low-wage cities and more skill mixing in these locations. It also results in the largest cities becoming more skill intensive, but with lower wage inequality.

Overall, we find that accounting for skill heterogeneity and spillovers across different types of workers is important for the design and aggregate welfare effects of spatial policies. The results suggest that nudging current U.S. policies towards generating a greater mixing of high and low skill workers in low-wage cities could be socially desirable.

References

Abdel-Rahman, H. and M. Fujita (1990). Product variety, marshallian externalities, and city sizes. *Journal of regional science 30*(2), 165–183.

Abdel-Rahman, H. M. and A. Anas (2004). Theories of systems of cities. *Handbook of regional and urban economics 4*, 2293–2339.

Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf (2015). The economics of density: Evidence from the berlin wall. *Econometrica 83*(6), 2127–2189.

Albouy, D. (2009). The unequal geographic burden of federal taxation. *Journal of Political Economy 117*(4), 635–667.

Albouy, D., K. Behrens, F. Robert-Nicoud, and N. Seegert (2019). The optimal distribution of population across cities. *Journal of Urban Economics 110*, 102–113.

Allen, T. and C. Arkolakis (2014). Trade and the topography of the spatial economy. *Quarterly Journal of Economics 1085*, 1139.

Allen, T., C. Arkolakis, and X. Li (2015). Optimal city structure. *Yale University, mimeograph*.

Allen, T., C. Arkolakis, and Y. Takahashi (2014). Universal gravity. Technical report, National Bureau of Economic Research.

Arnott, R. (2004). Does the henry george theorem provide a practical guide to optimal city size? *American Journal of Economics and Sociology 63*(5), 1057–1090.

Austin, B., E. Glaeser, and L. Summers (2018). Saving the heartland: Place-based policies in 21st century america. *Brookings Papers on Economic Activity Conference Draft*.

Baum-Snow, N. and R. Pavan (2013). Inequality and city size. *Review of Economics and Statistics 95*(5), 1535–1548.

Behrens, K., G. Duranton, and F. Robert-Nicoud (2014). Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy 122*(3), 507–553.

Behrens, K. and F. Robert-Nicoud (2015). Agglomeration theory with heterogeneous agents. In *Handbook of regional and urban economics*, Volume 5, pp. 171–245. Elsevier.

Bhagwati, J. and H. G. Johnson (1960). Notes on some controversies in the theory of international trade. *The Economic Journal 70*(277), 74–93.

Bryan, G. and M. Morten (2015). Economic development and the spatial allocation of labor: Evidence from indonesia. *Manuscript, London School of Economics and Stanford University*, 1671–1748.

Bunten, D. (2017). Is the rent too high? aggregate implications of local land-use regulation.

Caliendo, L., F. Parro, E. Rossi-Hansberg, and P.-D. Sarte (2018). The impact of regional and sectoral productivity changes on the us economy. *Review of Economic Studies 85*, 2042–2096.

Carrell, S. E., B. I. Sacerdote, and J. E. West (2013). From natural variation to optimal policy? the importance of endogenous peer group formation. *Econometrica 81*(3), 855–882.

Ciccone, A. and R. E. Hall (1996). Productivity and the density of economic activity. *The American Economic Review*, 54–70.

Combes, P.-P. (2011). The empirics of economic geography: how to draw policy implications? *Review of World Economics 147*(3), 567–592.

Combes, P.-P., G. Duranton, and L. Gobillon (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics 63*(2), 723–742.

Combes, P.-P. and L. Gobillon (2015). The empirics of agglomeration economies. In *Handbook of regional and urban economics*, Volume 5, pp. 247–348. Elsevier.

Davis, D. R. and J. I. Dingel (2012). A spatial knowledge economy. Technical report, National Bureau of Economic Research.

Dekle, R., J. Eaton, and S. Kortum (2008). Global rebalancing with gravity: Measuring the burden of adjustment. Technical Report 3, International Monetary Fund.

Desmet, K. and E. Rossi-Hansberg (2013). Urban accounting and welfare. *American Economic Review 103*(6), 2296–2327.

Desmet, K. and E. Rossi-Hansberg (2014). Spatial development. *American Economic Review 104*(4), 1211–43.

Diamond, R. (2016). The determinants and welfare implications of us workers' diverging location choices by skill: 1980–2000. *The American Economic Review 106*(3), 479–524.

Dixit, A. (1985). Tax policy in open economies. *Handbook of public economics 1*, 313–374.

Dunbar, A. E. (2009). Metropolitan area disposable personal income: Methodology and results for 2001-2007.

Duranton, G. and D. Puga (2004). Micro-foundations of urban agglomeration economies. In *Handbook of regional and urban economics*, Volume 4, pp. 2063–2117. Elsevier.

Duranton, G. and A. J. Venables (2018). Pace-based policies for development. Technical report, National Bureau of Economic Research.

Eeckhout, J. and N. Guner (2017). Optimal spatial taxation: Are big cities too small?

Eeckhout, J., R. Pinheiro, and K. Schmidheiny (2014). Spatial sorting. *Journal of Political Economy 122*(3), 554–620.

Fajgelbaum, P. D., E. Morales, J. C. Suárez Serrato, and O. Zidar (2018). State taxes and spatial misallocation. *The Review of Economic Studies 86*(1), 333–376.

Fajgelbaum, P. D. and E. Schaal (2017). Optimal transport networks in spatial equilibrium. Technical report, National Bureau of Economic Research.

Flatters, F., V. Henderson, and P. Mieszkowski (1974). Public goods, efficiency, and regional fiscal equalization. *Journal of Public Economics 3*(2), 99–112.

Flood, S., M. King, S. Ruggles, and J. R. Warren (2017). Integrated public use microdata series, current population survey: Version 5.0.[dataset]. minneapolis: University of minnesota, 2017.

Gaubert, C. (2015). Firm sorting and agglomeration. *University of California, Berkeley*.

Glaeser, E. L. and J. D. Gottlieb (2008). The economics of place-making policies. *Brookings Papers on Economic Activity 39*(1 (Spring)), 155–253.

Head, K. and T. Mayer (2014). Gravity equations: Workhorse, toolkit, and cookbook. *Handbook of International Economics, Vol. 4*.

Helpman, E. (1998). The size of regions: transport and housing as factors in agglomeration. In D. Pines, E. Sadka, and I. Zilcha (Eds.), *Topics in Public Economics*, pp. 33–54. Cambridge University Press Cambridge.

Helpman, E. and D. Pines (1980). Optimal public investment and dispersion policy in a system of open cities. *The American Economic Review 70*(3), 507–514.

Helsley, R. W. and W. C. Strange (2014). Coagglomeration, clusters, and the scale and composition of cities. *Journal of Political Economy 122*(5), 1064–1093.

Henderson, J. V. (1974). The sizes and types of cities. *The American Economic Review*, 640–656.

Herkenhoff, K. F., L. E. Ohanian, and E. C. Prescott (2018). Tarnishing the golden and empire states: Land-use restrictions and the us economic slowdown. *Journal of Monetary Economics 93*, 89–109.

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and manufacturing TFP in China and India. *Quarterly Journal of Economics 124*(4), 1403–1448.

Hsieh, C.-T. and E. Moretti (2019). Housing constraints and spatial misallocation. *American Economic Journal: Macroeconomics 11*(2), 1–39.

Khajavirad, A., J. J. Michalek, and N. V. Sahinidis (2014). Relaxations of factorable functions with convex-transformable intermediates. *Mathematical Programming 144*(1-2), 107–140.

Kline, P. and E. Moretti (2014a). Local economic development, agglomeration economies and the big push: 100 years of evidence from the Tennessee Valley Authority. *Quarterly Journal of Economics*.

Kline, P. and E. Moretti (2014b). People, places, and public policy: Some simple welfare economics of local economic development programs. *Annual Review of Economics 6*(1), 629–662.

Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. *American Economic Review*, 950–959.

Lucas, R. E. and E. Rossi-Hansberg (2002). On the internal structure of cities. *Econometrica 70*(4), 1445–1476.

Melo, P. C., D. J. Graham, and R. B. Noland (2009). A meta-analysis of estimates of urban agglomeration economies. *Regional science and urban Economics 39*(3), 332–342.

Meyer, B., W. Mok, and J. Sullivan (2009). The under-reporting of transfers in household surveys: Its nature and consequences. *National Bureau of Economic Research, Inc, NBER Working Papers*.

Monte, F., S. J. Redding, and E. Rossi-Hansberg (2018). Commuting, migration, and local employment elasticities. *American Economic Review 108*(12), 3855–90.

Moretti, E. (2012). *The new geography of jobs.* Houghton Mifflin Harcourt.

Neumark, D., H. Simpson, et al. (2015). Place-based policies. *Handbook of Regional and Urban Economics 5*, 1197–1287.

Ossa, R. (2018). A quantitative analysis of subsidy competition in the us. Technical report, National Bureau of Economic Research.

Parkhomenko, A. (2018). The rise of housing supply regulation in the us: Local causes and aggregate implications. *University of Southern California*.

Pines, D. and E. Sadka (1986). Comparative statics analysis of a fully closed city. *Journal of Urban Economics 20*(1), 1–20.

Redding, S. J. (2016). Goods trade, factor mobility and welfare. *Journal of International Economics 101*, 148–167.

Redding, S. J. and E. A. Rossi-Hansberg (2017). Quantitative spatial economics. *Annual Review of Economics 9*(1).

Redding, S. J. and M. A. Turner (2015). Transportation costs and the spatial organization of economic activity. *Handbook of Regional and Urban Economics 5*, 1339–1398.

Roback, J. (1982). Wages, rents, and the quality of life. *Journal of Political Economy*, 1257–1278.

Roca, J. D. L. and D. Puga (2017). Learning by working in big cities. *The Review of Economic Studies 84* (1), 106–142.

Rosen, S. (1979). Wage-based indexes of urban quality of life. *Current issues in urban economics 3*, 324–345.

Rossi-Hansberg, E. (2005). A spatial theory of trade. *American Economic Review 95* (5), 1464–1491.

Ruggles, S., S. Flood, R. Goeken, J. Grover, E. Meyer, J. Pacas, and M. Sobek (2017). Ipums usa: Version 8.0 [dataset]. minneapolis, mn.

Saiz, A. (2010). The geographic determinants of housing supply. *The Quarterly Journal of Economics 125* (3), 1253–1296.

Sandmo, A. (1975). Optimal taxation in the presence of externalities. *The Swedish Journal of Economics*, 86–98.

Wildasin, D. E. (1980). Locational efficiency in a federal system. *Regional Science and Urban Economics 10* (4), 453–471.

Wilson, J. D. (1986). A theory of interregional tax competition. *Journal of urban Economics 19* (3), 296–315.

Zhelobodko, E., S. Kokovin, M. Parenti, and J.-F. Thisse (2012). Monopolistic competition: Beyond the constant elasticity of substitution. *Econometrica 80* (6), 2765–2784.

Zodrow, G. R. and P. Mieszkowski (1986). Pigou, tiebout, property taxation, and the underprovision of local public goods. *Journal of urban economics 19* (3), 356–370.

A Appendix to Sections 2 and 3

A.1 Appendix to Section 2.1

We show that (1) holds. The market allocation in the case considered in this section is defined by the following conditions:

$$u = a_j\left(L_j\right)c_j, \tag{A.1}$$

$$\sum_j L_j c_j = \sum_j L_j z_j, \tag{A.2}$$

$$\sum_j L_j = L. \tag{A.3}$$

The first condition says that utility is equalized, the second condition is goods market clearing, and the last condition is labor market clearing. Solving for c_j from the first condition and replacing in (A.2) we obtain the following expression for utility:

$$u = \frac{\sum_j L_j z_j\left(L_j\right)}{\sum_j \frac{L_j}{a_j\left(L_j\right)}}. \tag{A.4}$$

The planner maximizes this term subject to (A.3). Totally differentiating this expression with respect to employment, after a few manipulations we obtain:

$$\hat{u} = \left(1+\gamma^P\right)\frac{\sum_j z_j dL_j}{\sum_j L_j z_j} - \left(1-\gamma^A\right)\frac{\sum_j \frac{1}{a_j} dL_j}{\sum_j \frac{L_j}{a_j}}$$

Further using (A.1) and (A.2) we obtain:

$$\hat{u} = \frac{\gamma^P \sum_j z_j dL_j + \gamma^A \sum_j c_j dL_j - \sum_j t_j dL_j}{Y}, \tag{A.5}$$

where $t_j \equiv c_j - z_j$ is the transfer received by j in the market allocation and Y is aggregate output. The no-transfers equilibrium implies $t_j = c_j - z_j = 0$, which using the definition of output as $Y_j = z_j L_j$ gives (1).

A.2 Appendix to Section 2.1

We derive (18). The market allocation in a case with multiple worker types is the solution to the following conditions

$$u^\theta = a_j^\theta \left(L_j^U, L_j^S\right) c_j^\theta, \tag{A.6}$$

$$\sum_{\theta=U,S} \sum_j L_j^\theta \left(c_j^\theta - z_j^\theta \left(L_j^U, L_j^S\right)\right) \leq 0, \tag{A.7}$$

$$\sum_j L_j^\theta = L^\theta. \tag{A.8}$$

Combining the first two conditions and following similar steps to Section (A.1), utility of unskilled workers can be written:

$$u^U = \frac{\sum_{\theta=U,S} \sum_j L_j^\theta z_j^\theta - \sum_{\theta' \neq \theta} \underline{u}^{\theta'} \sum_j \frac{L_j^{\theta'}}{a_j^{\theta'}}}{\sum_j \frac{L_j^U}{a_j^U}} \tag{A.9}$$

Taking a first order approximation to this expression while keeping $\underline{u}^{\theta'}$ constant and further using the mobility constraints (A.6) we obtain:

$$\frac{du}{u} = \frac{\sum_\theta \sum_j L_j^\theta z_j^\theta \left(\frac{dL_j^\theta}{L_j^\theta} + \sum_{\theta'} \gamma_{\theta',\theta}^P \frac{dL_j^{\theta'}}{L_j^{\theta'}}\right) - \sum_\theta \sum_j c_j^\theta L_j^\theta \left(\frac{dL_j^\theta}{L_j^\theta} - \sum_{\theta'} \gamma_{\theta',\theta}^A \frac{dL_j^{\theta'}}{L_j^{\theta'}}\right)}{\sum_j c_j^U L_j^U}. \tag{A.10}$$

which, after some manipulations, becomes:

$$\frac{du}{u} = \frac{\sum_\theta \sum_j \left[-\left(c_j - z_j\right) L_j^\theta + \sum_{\theta'} \left(\gamma_{\theta,\theta'}^P L_j^{\theta'} z_j^{\theta'} + \gamma_{\theta,\theta'}^A c_j^{\theta'} L_j^{\theta'}\right)\right] \frac{dL_j^\theta}{L_j^\theta}}{\sum_j c_j^U L_j^U}. \tag{A.11}$$

Imposing no transfers ($c_j = z_j$) and using that $z_j^\theta = w_j^\theta$ in a market allocation gives the result.

A.3 Planning Problem and Proofs of Propositions 1 to 3

The planning problem can be described as follows.

Definition 2. *The planning problem is*

$$\max L^\theta u^\theta$$

subject to (i) the spatial mobility constraints

$$L_j^\theta u^\theta \leq L_j^\theta a_j^\theta \left(L_j^1, .., L_j^\Theta\right) U \left(c_j^\theta, h_j^\theta\right) \text{ for all } j;$$

$$L_j^{\theta'} \underline{u}^{\theta'} \leq L_j^{\theta'} a_j^\theta \left(L_j^1, .., L_j^\Theta\right) U \left(c_j^\theta, h_j^\theta\right) \text{ for all } j \text{ and } \theta' \neq \theta;$$

(ii) the tradable and non-tradable goods feasibility constraints

$$\sum_i d_{ji} Q_{ji} \leq Y_j \left(N_j^Y, I_j^Y\right) \text{ for all } j, i;$$

$$\sum_\theta L_j^\theta c_j^\theta + I_j^Y + I_j^H \leq Q\left(Q_{1j}, .., Q_{Jj}\right) \text{ for all } j;$$

$$\sum_\theta L_j^\theta h_j^\theta \leq H_j \left(N_j^H, I_j^H\right) \text{ for all } j;$$

(iii) local and national labor-market clearing,

$$N_j^Y + N_j^H = N\left(z_1^\theta \left(L_j^1, .., L_j^\Theta\right) L_j^1, .., z_j^\Theta \left(L_j^1, .., L_j^\Theta\right) L_j^\Theta\right) \text{ for all } j;$$

$$\sum_j L_j^\theta = L^\theta \text{ for all } \theta; \text{ and}$$

(iv) non-negativity constraints on consumption, trade flows, intermediate inputs, and labor.

Proposition 1. *If a competitive equilibrium is efficient, then*

$$W_j \frac{dN_j}{dL_j^\theta} + \sum_{\theta'} \frac{x_j^{\theta'} L_j^{\theta'}}{a_j^{\theta'}} \frac{\partial a_j^{\theta'}}{\partial L_j^\theta} = x_j^\theta + E^\theta \qquad \text{if } L_j^\theta > 0, \tag{A.12}$$

for all j and θ and some constants $\left\{E^\theta\right\}$. If the planner's problem is globally concave and (A.12) *holds for some specific $\left\{E^\theta\right\}$, then the competitive equilibrium is efficient.*

Proof. First we present the system of necessary first order conditions in the planner's problem. Then we contrast it with the market allocation. The Lagrangian of the planning problem is:

$$\begin{aligned}
\mathscr{L} = {} & u^\theta \\
& - \sum_j \omega_j^\theta L_j^{\theta'} \left(u^\theta - a_j^{\theta'} \left(L_j^1, .., L_j^\Theta\right) U\left(c_j^{\theta'}, h_j^{\theta'}\right)\right) \\
& - \sum_{\theta' \neq \theta} \sum_j \omega_j^{\theta'} L_j^{\theta'} \left(\underline{u}^{\theta'} - a_j^{\theta'} \left(L_j^1, .., L_j^\Theta\right) U\left(c_j^{\theta'}, h_j^{\theta'}\right)\right) \\
& - \sum_j p_j^* \left(\sum_i d_{ji} Q_{ji} - Y_j \left(N_j^Y, I_j^Y\right)\right) \\
& - \sum_j P_j^* \left(\sum_\theta L_j^\theta c_j^\theta + I_j^Y + I_j^H - Q\left(Q_{1j}, .., Q_{Jj}\right)\right) - \sum_j R_j^* \left(\sum_\theta L_j^\theta h_j^\theta - H_j \left(N_j^H, I_j^H\right)\right) \\
& - \sum_j W_j^* \left(N_j^Y + N_j^H - N\left(z_j^1 \left(L_j^1, .., L_j^\Theta\right) L_j^1, .., z_j^\Theta \left(L_j^1, .., L_j^\Theta\right) L_j^\Theta\right)\right) \\
& - \sum_\theta E^\theta \left(\sum_j L_j^\theta - L^\theta\right) + ...
\end{aligned} \tag{A.13}$$

where we omit notation for the non-negativity constraints. The first-order conditions with respect to trade flows, labor services and intermediate inputs are:

$$[Q_{ji}] \qquad P_i^* \frac{\partial Q\left(Q_{1i}, .., Q_{Ji}\right)}{\partial Q_{ji}} \leq p_j^* \tau_{ji}, \tag{A.14}$$

$$\left[N_j^Y, N_j^H\right] \qquad p_j^* \frac{\partial Y_j}{\partial N_j^Y} \leq W_j^*; R_j^* \frac{\partial H_j}{\partial N_j^H} \leq W_j^*, \tag{A.15}$$

$$\left[I_j^Y, I_j^H\right] \qquad p_j^* \frac{\partial Y_j}{\partial I_j^Y} \leq P_j^*; R_j^* \frac{\partial H_j}{\partial I_j^H} \leq P_j^*, \tag{A.16}$$

each holding with equality in an interior solution. The first-order conditions with respect to individual consumption of traded and non-traded goods can be written:

$$\left[c_j^\theta\right] \qquad \omega_j^\theta a_j^\theta \frac{\partial U\left(c_j^\theta, h_j^\theta\right)}{\partial c_j^\theta} c_j^\theta = P_j^* c_j^\theta$$

$$\left[h_j^\theta\right] \qquad \omega_j^\theta a_j^\theta \frac{\partial U\left(c_j^\theta, h_j^\theta\right)}{\partial h_j^\theta} h_j^\theta = R_j^* h_j^\theta$$

Adding up the last two expressions and using degree-1 homogeneity of U gives

$$\omega_j^\theta a_j^\theta U\left(c_j^\theta, h_j^\theta\right) = x_j^{\theta *}, \tag{A.17}$$

where

$$x_j^{\theta *} \equiv R_j^* h_j^\theta + P_j^* c_j^\theta. \tag{A.18}$$

Therefore, we can write

$$\left[c_j^\theta\right] \qquad c_j^\theta = \frac{\alpha_C\left(c_j^\theta, h_j^\theta\right)}{P_j^*} x_j^{\theta *} \tag{A.19}$$

$$\left[h_j^\theta\right] \qquad h_j^\theta = \frac{1-\alpha_C\left(c_j^\theta, h_j^\theta\right)}{R_j^*} x_j^{\theta *} \tag{A.20}$$

where $\alpha_C\left(c,h\right) \equiv \frac{\partial U(c,h)}{\partial c} \frac{c}{U(c,h)}$ is the elasticity of U with respect to c.

Using (A.18) and the slackness condition on the spatial mobility constraint, the first-order condition of the planning problem with respect to L_j^θ is:

$$\sum_{\theta'} \omega_j^{\theta'} L_j^{\theta'} \frac{\partial a_j^{\theta'}\left(L_j^1, .., L_j^\Theta\right)}{\partial L_j^\theta} U\left(c_j^{\theta'}, h_j^{\theta'}\right) + W_j^* \frac{dN_j}{dL_j^\theta} \leq x_j^{\theta *} + E^\theta, \tag{A.21}$$

with equality if $L_j^\theta > 0$. Further using (A.17), if $L_j^\theta > 0$ then:

$$W_j^* \frac{dN_j}{dL_j^\theta} + \sum_{\theta'} \frac{\left(x_j^{\theta *}\right)' L_j^{\theta'}}{a_j^{\theta'}} \frac{\partial a_j^{\theta'}}{\partial L_j^\theta} = x_j^{\theta *} + E^\theta. \tag{A.22}$$

In locations with $L_j^\theta = 0$ then $c_j^\theta = h_j^\theta = x_j^{\theta *} = 0$. Therefore, $L_j^\theta = 0$ for all locations such that:

$$W_j^* \frac{dN_j}{dL_j^\theta} + \sum_{\theta' \neq \theta} \frac{\left(x_j^{\theta *}\right)' L_j^{\theta'}}{a_j^{\theta'}} \frac{\partial a_j^{\theta'}}{\partial L_j^\theta} \leq E^\theta. \tag{A.23}$$

An optimal allocation is given by quantities $\left\{Q_{ji}, N_j^Y, N_j^H, I_j^Y, I_j^H, c_j^\theta, h_j^\theta, L_j^\theta, u^\theta\right\}$ and multipliers $\left\{P_j^*, p_j^*, R_j^*, W_j^*, \omega_j^\theta\right\}$ such that the first-order conditions (A.14)-(A.22) and the constraints enumerated in (i) to (iii) in Definition 2 hold.

It is straightforward to show that (A.14) to (A.16), (A.19) and (A.20) coincide with the optimality conditions of producers and consumers (i) and (ii) in the competitive equilibrium from Definition 1 given competitive prices $\{P_j, p_j, R_j, W_j\}$ equal to the multipliers $\left\{P_j^*, p_j^*, R_j^*, W_j^*\right\}$ and decentralized expenditure x_j^θ equal to $x_j^{\theta *}$. In addition, the restrictions (i) to (iii) from definition 2 of the planning problem are the same as restriction (iii) from the competitive equilibrium. Therefore, the system characterizing the competitive solution for $\left\{Q_{ji}, N_j^Y, N_j^H, I_j^Y, I_j^H, c_j^\theta, h_j^\theta, L_j^\theta\right\}$ given the prices $\{P_j, p_j, R_j, W_j\}$ and the expenditure x_j^θ is the same as the system characterizing the planner allocation for those same quantities given the multipliers $\left\{P_j^*, p_j^*, R_j^*, W_j^*\right\}$ and $x_j^{\theta *}$. As a result, if the competitive allocation is efficient, then $x_j^\theta = x_j^{\theta *}$ where $x_j^{\theta *}$ is given by (A.22). Conversely, if $x_j^\theta = x_j^{\theta *}$ for $x_j^{\theta *}$ defined in (A.12) given the W^θ that solves the planner's problem, there is a solution for the competitive allocation such that $\{P_j, p_j, R_j, W_j\} = \left\{P_j^*, p_j^*, R_j^*, W_j^*\right\}$. If the planning problem is concave then there is a unique solution to the system characterizing the planner's allocation, in which case $\{P_j, p_j, R_j, W_j\} = \left\{P_j^*, p_j^*, R_j^*, W_j^*\right\}$ is the only competitive equilibrium.

□

Proposition 2. *The optimal allocation can be implemented by the transfers*

$$t_j^{\theta *} = \sum_{\theta'} \left(\gamma_{\theta,\theta'}^{P,j} w_j^{\theta' *} + \gamma_{\theta,\theta'}^{A,j} x_j^{\theta' *} \right) \frac{L_j^{\theta' *}}{L_j^{\theta *}} - \left(b^\theta \Pi^* + E^\theta \right), \tag{A.24}$$

where the terms $\left(x_j^{\theta *}, w_j^{\theta *}, L_j^{\theta *}, \Pi^*\right)$ *are the outcomes at the efficient allocation, and* $\left\{E^\theta\right\}$ *are constants equal to the multipliers on the resource constraint of each type in the planner's allocation.*

Proof. The result follows from combining Proposition 1 with the expressions for labor demand (14) and expenditure per capita (16). Using the labor demand condition (14) we obtain that the value of the marginal product of labor can be written as function of wages, employment and elasticities:

$$W_j \frac{dN_j}{dL_j^\theta} = w_j^\theta \left(1 + \gamma_{\theta,\theta}^{P,j}\right) + \sum_{\theta' \neq \theta} w_j^{\theta'} \left(\frac{L_j^{\theta'}}{L_j^\theta} \right) \gamma_{\theta,\theta'}^{P,j}. \tag{A.25}$$

Combining this expression with the definition of $\gamma_{\theta,\theta'}^{A,j}$ in (10), we can re-write (19) as follows:

$$w_j^\theta - x_j^\theta + \sum_{\forall \theta'} \left(\gamma_{\theta,\theta'}^{P,j} w_j^{\theta'} + \gamma_{\theta,\theta'}^{A,j} x_j^{\theta'} \right) \frac{L_j^{\theta'}}{L_j^\theta} = E^\theta. \tag{A.26}$$

Combining this last expression with (16) gives the result.

□

Proposition 3. *The planning problem is concave if* $\Gamma^A > \Gamma^P$, $\Gamma^A \geq 0$ *and* $\gamma_{\theta,\theta'}^A > 0$ *for* $\theta \neq \theta'$. *Under a single worker type* ($\Theta = 1$)*, the planning problem is quasi-concave if* $1 + \gamma^A > \left(1 + \gamma^P\right) \left[\frac{1-\alpha_C}{1+D} + \alpha_C \right]$.

Proof. We consider the following planning problem defined in section 2.4:

$$\begin{aligned} &\max \, u^\theta \\ &\text{s.t.: } u^{\theta'} = \underline{u}^{\theta'} \quad \text{for } \theta' \neq \theta \\ &\qquad u^{\theta'} \in \mathcal{U} \quad \text{for all } \theta' \end{aligned}$$

where θ is a given type, $\mathcal{U}$ is the set of attainable utility levels $\left\{u^\theta\right\}$ and $\underline{u}^{\theta'}$ for $\theta' \neq \theta$ is an arbitrary attainable utility level for group θ'. $\mathcal{U}$ is characterized by a set of feasibility constraints which are defined in the main text, and which we come back to below. We show here that this problem, noted $\mathcal{P}$, can be recast as a concave problem, under the condition stated in proposition 2. Therefore, a local maximum of $\mathcal{P}$ is necessarily its unique global maximum. The planning problem $\mathcal{P}$ can be recast as the following equivalent problem $\mathcal{P}'$, after simple algebraic manipulations:

$$\max_{\left\{v^\theta, U_j^\theta, C_j^\theta, H_j^\theta, \widetilde{L}_j^\theta, \widetilde{N}_j^k, I_j^k, Q_{ij}, M_j, S_j\right\}} v^\theta \tag{A.27}$$

subject to the set of constraints $\mathcal{C}$:

$$\underline{v}^{\theta'} - \mathcal{F} \left(\frac{U_j^{\theta'} \prod_{\theta'' \neq \theta'} \left(\widetilde{L}_j^{\theta''}\right)^{\frac{\gamma_{\theta'',\theta'}^A}{1+\Gamma^P}}}{\left(\widetilde{L}_j^{\theta'}\right)^{\frac{1-\gamma_{\theta',\theta'}^A}{1+\Gamma^P}}} \right) \leq 0 \text{ for all } j \text{ and } \theta'; \tag{A.28}$$

$$U_j^\theta - U\left(C_j^\theta, H_j^\theta\right) \leq 0 \tag{A.29}$$

$$\sum_i d_{ji} Q_{ji} - \left(b_Y^N \left(N_j^Y\right)^{\beta_Y} + b_Y^I \left(I_j^Y\right)^{\beta_Y} \right)^{\frac{1}{\beta_Y}} \leq 0 \text{ for all } j, i; \tag{A.30}$$

$$\sum_\theta C_j^\theta + \left(I_j^Y\right) + \left(I_j^H\right) - Q\left(Q_{1j}, .., Q_{Jj}\right) \leq 0 \text{ for all } j; \tag{A.31}$$

$$\sum_{\theta} H_j^{\theta} - \left(b_H^N \left(N_j^H\right)^{\beta_H} + b_H^I \left(I_j^H\right)^{\beta_H}\right)^{\frac{1}{\beta_H}} \leq 0 \tag{A.32}$$

$$M_j - \left[\sum_{\theta} \left(Z_j^{\theta} \prod_{\theta'} \left(\widetilde{L}_j^{\theta'}\right)^{\frac{\gamma^P_{\theta',\theta}}{1+\Gamma^P}} \left(\widetilde{L}_j^{\theta}\right)^{\frac{1}{1+\Gamma^P}}.\right)^{\rho}\right]^{\frac{1}{\rho}} \leq 0 \text{ for all } j; \tag{A.33}$$

$$N_j^Y + N_j^H - M_j \leq 0 \tag{A.34}$$

$$\sum_j \left(\widetilde{L}_j^{\theta}\right)^{\frac{1}{1+\Gamma^P}} - L^{\theta} = 0 \text{ for all } \theta \tag{A.35}$$

To reach these expressions, we have introduced the auxiliary variables M_j, S_j, and U_j^{θ} we have used the following change of variables

$$\begin{aligned} v^{\theta} &= \mathcal{F}\left(u^{\theta}\right) \\ H_j^{\theta} &= L_j^{\theta} h_j^{\theta} \\ C_j^{\theta} &= L_j^{\theta} c_j^{\theta} \\ \widetilde{L}_j^{\theta} &= \left(L_j^{\theta}\right)^{1+\Gamma^P} \quad \text{for all } j \in 1,...,N \text{ and } \theta, \end{aligned}$$

where the function $\mathcal{F}(.)$ is defined by $\mathcal{F}(x) = -x^b$ for $b = \frac{1+\Gamma^P}{\Gamma^P - \Gamma^A}$. Problems $\mathcal{P}$ and $\mathcal{P}'$ are equivalent: any solution to $\mathcal{P}'$ is a solution to $\mathcal{P}$ and vice-versa. We then consider the relaxed problem $\mathcal{P}''$ that is identical to $\mathcal{P}'$ except that the last constraint of $\mathcal{P}'$ is relaxed into an inequality constraint:

$$L^{\theta} - \sum_j \left(\widetilde{L}_j^{\theta}\right)^{\frac{1}{1+\Gamma^P}} \leq 0 \text{ for all } \theta. \tag{A.36}$$

We now show that problem $\mathcal{P}''$ has a concave objective and convex constraints under the assumptions of proposition 2. To that end, we show that under these assumptions, each constraint of $\mathcal{P}''$ is convex.

Consider first the constraint (A.28), and examine specifically the expression:

$$f_j^{\theta}(U_j^{\theta}, \left\{L^{\theta}\right\}, \left\{L^{\theta'}\right\}) = \frac{U_j^{\theta} \prod_{\theta' \neq \theta} \left(\widetilde{L}_j^{\theta'}\right)^{\frac{\gamma^A_{\theta',\theta}}{1+\Gamma^P}}}{\left(\widetilde{L}_j^{\theta}\right)^{\frac{1-\gamma^A_{\theta,\theta}}{1+\Gamma^P}}} \tag{A.37}$$

This expression is a multivariate function of the form $f(y,z) = \frac{\prod_{i=1}^k y_i^{a_i}}{z^b}$ where $a_i > 0$, $b > 0$ and $\sum_{i=1}^k a_i < b$. By proposition 11 of Khajavirad et al. (2014), such functions are G-concave, meaning that the function $G(f(y,z))$ is concave in (y,z), for functions $G(x)$ that are concave transforms of $-x^{\frac{1}{\sum a_i - b}}$. Assumptions made on parameter values in Proposition 3 ensure that $\gamma^A_{\theta',\theta} \geq 0$ for all $\theta' \neq \theta$ and $1 + \frac{\gamma^A_{\theta',\theta}}{1+\Gamma^P} < \frac{1-\gamma^A_{\theta,\theta}}{1+\Gamma^P}$, which follows from $\Gamma^A > \Gamma^P$. Therefore, by Proposition 11 of Khajavirad et al. (2014), the transformation $G_{\theta}(x) = -x^{\frac{1+\Gamma^P}{\Gamma^P - \left(\sigma_{\theta} + \sum_{\theta'} \gamma^A_{\theta',\theta}\right)}}$ ensures that $G_{\theta}(f_j^{\theta}(.))$ is concave. Finally, given the definition of Γ^A, $\mathcal{F}(.)$ is a concave transform of $G_{\theta}(.)$. Therefore, (A.28) is convex for all θ'.

Second, functions of the form $f(x_1,...,x_n) = \left[\sum a_i x_i^{\beta}\right]^{\rho}$ are concave whenever $\beta \in (0,1)$ and $\rho\beta \leq 1$. Therefore, constraints (A.30), (A.32) and (A.36) are convex.

The constraint (A.29) is convex because $U(.)$ is concave. The constraint (A.31) is convex because the aggregator $Q(.)$ is concave.

Next, consider the constraint (A.33). The second term is the negative of a composition of an increasing CES function with exponent $\rho \leq 1$, which is concave, and a series of functions of the form

$$f(x_1,...,x_{\Theta}) = \prod_{\theta'} \left(x^{\theta'}\right)^{\frac{\gamma^P_{\theta',\theta}}{1+\Gamma^P}} \left(x^{\theta}\right)^{\frac{1}{1+\Gamma^P}}.$$

As concave transforms of a geometric mean, these functions are concave, whenever $\frac{1+\sum_{\theta'}\gamma^P_{\theta',\theta}}{1+\Gamma^P}\in(0,1)$. This restriction holds by definition of Γ^P. We finally invoke that the vector composition of a concave function that is increasing in all its elements with a concave function is concave. Therefore, constraint (A.33) is convex. Finally, constraint (A.34) is linear hence convex.

It follows that the relaxed problem $\mathcal{P}''$ is a maximization problem with concave objective and convex inequality constraints. It admits at most one global maximum, and a vector satisfying its first order conditions is necessarily the global maximum. If at this unique optimal point for $\mathcal{P}''$ the relaxed constraint (A.36) binds, so that (A.35) holds, we guarantee that the solution to $\mathcal{P}''$ is also the unique global maximizer of $\mathcal{P}$' and the unique global maximizer of the equivalent problem $\mathcal{P}$.[45]

We now specialize to the case of a single type of workers ($\Theta=1$) where the decreasing returns to scale in the production of housing help make the problem concave. The relaxed planner's problem $\mathcal{P}''$ can be further simplified in this case to:

$$\max_{\left\{v^\theta,U_j^\theta,C_j^\theta,H_j^\theta,\widetilde{L}_j^\theta,\widetilde{N}_j^k,I_j^k,Q_{ij},M_j,S_j\right\}}\min_j\frac{\left(C_j^\theta\right)^{\alpha_C}\left(\widetilde{H}_j^\theta\right)^{\frac{1-\alpha_C}{1+d_{H,j}}}}{\left(\widetilde{L}_j^\theta\right)^{\frac{1-\gamma^A_{\theta,\theta}}{1+\Gamma^P}}}$$

subject to the constraints (A.30), (A.31), (A.33), (A.34) and (A.36), which are unchanged except that they now hold for only one group. The modified constraint for housing production is:

$$\widetilde{H}_j^\theta-\left(b_H^N\left(\widetilde{N_j^H}\right)^{\beta_H(1+D)}+b_H^I\left(\widetilde{I_j^H}\right)^{(1+D)\beta_H}\right)^{\frac{1}{\beta_H}\frac{1}{1+D}}\leq 0 \tag{A.38}$$

where we have used the following change of variable $\widetilde{H}_j^\theta=\left(H_j^\theta\right)^{1+d'_{H,j}}$. The modified housing market constraint (A.38) is convex. The objective of the planner is quasi-concave as the minimum of a ratio of a concave and a convex function, as long as $(1-\alpha_C)\frac{1}{1+d'_{H,j}}+\alpha_C\leq\frac{1-\gamma^A_{\theta,\theta}}{1+\Gamma^P}$ in each city. The constraints are convex. Therefore, the problem is a quasi-concave maximization problem as long as the parameter restriction in (ii) holds.

□

A.4 Equivalence with Monopolistic Competition

Consider the economic geography environment from Section 3.4. As a reminder, that environment starts from the general model from Section 2 and imposes only one labor type, inelastic housing supply ($H_j\left(N_j^H,I_j^H\right)=H_j$ is a constant), and only labor used in production of traded goods ($Y_j\left(N_j^Y,I_j^Y\right)=N_j^Y=N_j=z_j\left(L_j\right)L_j$). Now suppose that, in addition, the production structure in the traded sector is the same as in Krugman (1980): in each location j, M_j homogeneous plants produce differentiated varieties with constant elasticity of substitution κ across them, and setting up a plant in location j requires F_j units of labor. The resulting environment corresponds to Redding (2016) or Helpman (1998) in the absence of individual preference shocks ($\sigma=0$).

We now show that the competitive allocation of such an extended model, as well as their normative implications, are equivalent to the model with homogeneous products analyzed in Section 3.4 under an aggregate production function equal to:

$$\widetilde{Y_j}\left(L_j\right)=K_j\left(z_j\left(L_j\right)L_j\right)^{\frac{\kappa}{\kappa-1}}, \tag{A.39}$$

where $K_j\equiv\frac{\kappa-1}{\kappa}\left(\kappa F_j\right)^{\frac{1}{1-\kappa}}$ is a constant. Therefore, a monopolistic competition model with no productivity spillovers is equivalent to a homogeneous-product model with perfect competition and spillover elasticity equal to $\gamma^P=\frac{1}{\kappa-1}$. This property relates to the result, dating back to at least Abdel-Rahman and Fujita (1990) and also shown by Allen and Arkolakis (2014) within their model, that CES product differentiation with monopolistic competition has the same aggregate implications as a constant-elasticity aggregate production function with increasing returns. In our

[45] We have not proven that (A.36) necessarily binds at the optimal solution for $\mathcal{P}''$. Therefore, we verify that this is indeed the case in the solution to $\mathcal{P}''$ in the implementation.

context, we must also demonstrate that the equivalence extends to the welfare implications summarized in Proposition 1.

Environment We start by describing how the physical environment of this model differs from the environments from Section 2. Now, the input to the aggregator $Q(\{Q_{ji}\})$ is $Q_{ji} = M_j^{\frac{\kappa}{\kappa-1}} q_{ji}$, where M_j is the number of plants in j and q_{ji} is the quantity exported by each of these from j to i. The feasibility constraint for traded goods (6) becomes $z_j(L_j) L_j = M_j \left(\sum_i \tau_{ji} q_{ji} + F_j\right)$ to account for the use of labor in setting up plants. Combining these two expressions, that constraint can be further expressed:

$$M_j^{\frac{1}{\kappa-1}} \left(z_j(L_j) L_j - F_j M_j\right) = \sum_i \tau_{ji} Q_{ji}. \tag{A.40}$$

Competitive Equilibrium Now we describe how the market allocation differs from the baseline environments. First, the producers' profit maximization condition is now:

$$\max \sum_i \left(\tilde{p_{ji}} - \tau_{ji} W_j\right) q_{ji} \tag{A.41}$$

subject to $q_{ji} = Q_{ji} \left(\frac{\tilde{p_{ji}}}{p_{ji}}\right)^{-\kappa}$, where $p_{ji} = M_j^{\frac{1}{1-\kappa}} \tilde{p_{ji}}$ is the price index corresponding to the exports from j to i and $\tilde{p_{ji}}$ is the price at which each firm from j sells in i. The solution to this problem yields the standard constant markup rule, $\tilde{p_{ji}} = \tau_{ji} \frac{\kappa}{\kappa-1} W_j$. We have as before that the price in location i of the aggregate traded good from j, p_{ji}, can be expressed according to the "mill pricing" rule as $\tau_{ji} p_j$, where now the price index corresponding to the domestic sales of traded goods in j is:

$$p_j \equiv M_j^{\frac{1}{1-\kappa}} \frac{\kappa}{\kappa-1} W_j. \tag{A.42}$$

As a result, condition (15) still determines the flows in the competitive equilibrium. Combining these pricing rules with (A.41), imposing zero profits and using (6) we obtain the number of producers in a competitive allocation:

$$M_j = \frac{z_j(L_j) L_j}{\kappa F_j}. \tag{A.43}$$

And further combining with (A.40), we can write

$$\widetilde{Y_j}(L_j) = \sum_j \tau_{ij} Q_{ij} \tag{A.44}$$

for $\widetilde{Y_j}$ given in (A.39).

We conclude that the competitive allocation can be represented as in the model without product differentiation from Definition 1 under the restrictions from Section 3.4 and assuming the aggregate production function $\widetilde{Y_j}(L_j)$. I.e., it is given by quantities $\{c_j, h_j, L_j, Q_{ij}, L_j\}$ and prices P_j, R_j, p_j, such that: (i) consumers optimize (i.e., c_j, h_j are a solution to (11) given expenditures x_i); (ii) trade flows are given by (15); (iii) employment L_j is consistent with the spatial mobility constraint (13); and (iv) all markets clear, i.e. (3), (5) and (A.44) hold.[46]

[46] Note that the definition of the competitive allocation can dispense with the wage W_j, which can be determined residually from (A.42).

Planning Problem In turn, the planning problem from Definition (2) is now associated with the Lagrangian

$$\begin{aligned}\mathscr{L} = u - \sum_j \omega_j \left(u - a_j \left(L_j\right) U \left(c_j, h_j\right) \left(\frac{L_j}{L}\right)^{-\sigma} \right) \\ - \sum_j p_j^* \left(\sum_i d_{ji} Q_{ji} - M_j^{\frac{1}{\kappa - 1}} \left(z_j \left(L_j\right) L_j - F_j M_j\right) \right) \\ - \sum_j P_j^* \left(L_j c_j - Q \left(Q_{1j}, .., Q_{Jj}\right)\right) - \sum_j R_j^* \left(L_j h_j - H_j\right) - W \left(\sum_j L_j - L \right) + ... \end{aligned} \tag{A.45}$$

Relative to Definition 2, now the planner also chooses the number of firms M_j in each location and faces the constraint (A.40) instead of (6). We readily see that entry is efficient by noting that the first-order condition with respect to M_j implies (A.43). As a result, the market clearing constraint in the second line of (A.45) can be replaced by (A.44). The resulting planning problem is equivalent to Definition 2 applied to the economic geography model in Section 3.4 under the production function $\widetilde{Y_j} \left(L_j\right)$.

A.5 Preference Draws within Types

The Lagrangian of planning problem described in the extension to preference draws in Section 3.5 is a special case of (A.13), except that now the spillover function $a_j^{\theta'} \left(L_j^1, .., L_j^\Theta\right)$ is replaced by $a_i^{\theta'} \left(L_i^\theta\right)^{-\sigma_\theta}$. Following the same steps as in the proof of Proposition 1, we find that condition (19) is extended to

$$W_j \frac{dN_j}{dL_j^\theta} + \sum_{\theta'} \frac{x_j^{\theta'} L_j^{\theta'}}{a_j^{\theta'}} \frac{\partial a_j^{\theta'}}{\partial L_j^\theta} = x_j^\theta \left(1 + \sigma_\theta\right) + E^\theta \quad \text{if } L_j^\theta > 0. \tag{A.46}$$

Following the same steps as in the proof of Proposition 2, we find that (20) is extended to

$$t_j^\theta = \gamma_{\theta,\theta}^{P,j} + \left(\gamma_{\theta,\theta}^{A,j} - \sigma_\theta\right) + \sum_{\theta' \neq \theta} \left(\gamma_{\theta,\theta'}^{P,j} w_j^{\theta' *} + \gamma_{\theta,\theta'}^{A,j} x_j^{\theta' *}\right) \frac{L_j^{\theta' *}}{L_j^{\theta *}} - \left(b^\theta \Pi^* + E^\theta\right). \tag{A.47}$$

The general-equilibrium structure underlying propositions 3 and 4 under the assumptions of the quantitative model can be expressed exactly as in the proof of Proposition 3 and as in the planning problem in relative changes from Section A.9 below, the only modification being that the term $\gamma_{\theta,\theta}^A$ is replaced by $\gamma_{\theta,\theta}^A - \sigma_\theta$.

A.6 Commuting

The Lagrangian of the planning problem described in the extension to spillovers across locations in Section 3.5 is

$$\begin{aligned}
\mathscr{L} = u &- \sum_i \sum_j \omega_{ji} \left(u - L_{ji}^{-\sigma} L^{\sigma} a_j \left(L_j^R\right) U_{ji}\left(c_{ji}, h_{ji}\right)\right) \\
&- \sum_j p_j^* \left(\sum_i d_{ji} Q_{ji} - Y_j\left(N_j^Y, I_j^Y\right)\right) \\
&- \sum_j P_j^* \left(\sum_i L_{ji} c_{ji} + I_j^Y + I_j^H - Q\left(Q_{1j}, .., Q_{Jj}\right)\right) \\
&- \sum_j W_j^* \left(N_j^I + N_j^H - z_j\left(L_j^W\right) L_j^W\right) \\
&- \sum_j R_j^* \left(\sum_i L_{ji} h_{ji} - H_j\left(N_j^H, I_j^H\right)\right) \\
&- E\left(\sum_i \sum_j L_{ij} - L\right) + ... \qquad \text{(A.48)}
\end{aligned}$$

where the residents and workers at j and i are, respectively:

$$\begin{aligned}
L_j^R &= \sum_{i'} L_{ji'}, \\
L_i^W &= \sum_{j'} L_{j'i}.
\end{aligned}$$

Now the planner optimizes over the bilateral flows L_{ji} from place of residence j to place of work i, the consumption of tradeables and non-tradeables c_{ji} and h_{ji} of each of these commuters$i$, and the same remaining margins as in the benchmark model (trade flows Q_{ji} and allocation of inputs into production of tradeables and non-tradeables). The first-order condition with respect to L_{ji} is:

$$[L_{ji}] : -\sigma\omega_{ji} L_{ji}^{-\sigma-1} a_j\left(L_j^R\right) U_{ji}\left(c_{ji}, h_{ji}\right) + \sum_{i'} \omega_{i'j} L_{ji'}^{-\sigma} a_j'\left(L_j^R\right) U_{ji}\left(c_{ji'}, h_{ji'}\right) + P_i^*\left(z_i'\left(L_i^W\right) L_i^W + z_i\left(L_i^W\right)\right) = x_{ji}^* + E \qquad \text{(A.49)}$$

In addition, the first order conditions over c_{ji} and h_{ji} and homogeneity of degree 1 of U_{ji} imply $\omega_{ji} L_{ji}^{-\sigma-1} a_j\left(L_j^R\right) U_{ji}\left(c_{ji}, h_{ji}\right) = x_{ji}^*$. Combining this expression with (2), using the definition of spillover elasticities $\gamma_i^P = \frac{z_i'\left(L_i^W\right)}{z_i\left(L_i^W\right)} L_i^W$ and $\gamma_j^A = \frac{a_j'\left(L_j^R\right)}{a_j\left(L_j^R\right)} L_j^R$ and re-arranging we get:

$$x_{ji}^* = \frac{\gamma_j^A}{1+\sigma} \sum_{i'} \frac{L_{ji'} x_{ji'}^*}{L_j^R} + \left(\frac{\gamma_i^P + 1}{1+\sigma}\right) P_i^* z_i\left(L_i^W\right) - \frac{E}{1+\sigma}. \qquad \text{(A.50)}$$

To reach (31) we further use that the wage received by a commuter who works in i is $w_i^* = P_i^* z\left(L_i^W\right)$, and the definition of expenditures $x_{ji}^* = w_i^* + \frac{\Pi^*}{L} + t_{ji}^*$.

A.7 Spillovers Across Locations

The Lagrangian of the planning problem described in the extension to spillovers across locations in Section 3.5 is a special case of (A.13), except that now the the supply of efficiency units in j is $N_j\left(\{L_{j'}\}\right) = z_j\left(\{L_{j'}\}\right) L_j$. Compared to our derivation of Proposition 1, the only difference is the first-order condition with respect to employment. Now, instead of (A.22) we reach:

$$\sum_{j'} W^*_{j'} \frac{dN_{j'}}{dL_j} + x^*_j \frac{L_j}{a_j} \frac{\partial a_j}{\partial L_j} = x^*_j + E. \tag{A.51}$$

In addition, instead of (A.25) we now have:

$$W_j \frac{dN_j}{dL_j} = \begin{cases} w_{j'} \frac{L_{j'}}{L_j} \gamma^{P,j,j'} & \text{if } j' \neq j, \\ w_j \left(\gamma^{P,j,j} + 1\right) & \text{if } j' = j. \end{cases} \tag{A.52}$$

Combining the last two expressions with (16) gives (34).

A.8 Public Spending

In the main text focused on pure transfers without including considerations for spending in public goods. Flatters et al. (1974) and Wildasin (1980) have shown that the optimal financing of local public goods may involve a fiscal union with inter-governmental grants to correct distortions caused by local taxes. We now accommodate these forces and show that they do not impact the optimal policies designed to deal with the spillovers. We allow for government spending valued by workers (G^U_j) to enter as a shifter in the amenity valuation for location j, $a^\theta_j = a^\theta_j \left(G^U_j, L^1_j, .., L^\Theta_j\right)$, and for government spending in public goods that impact productivity (G^Y_j) to enter as a shifter in the efficiency units of labor, $N_j = N_j \left(G^Y_j, z^1_j L^1_j, .., z^\Theta_j L^\Theta_j\right)$. We also allow for a production function of public goods $G_j = G_j \left(I^G_j, H^G_j\right)$ that takes as inputs the bundle of traded commodities (I^G_j) and non-traded services (H^G_j).

The Lagrangian of the planning problem is now

$$\begin{aligned}
\mathscr{L} = u &- \sum_j \omega_j L_j \left(u - a_j \left(G^U_j, L_j\right) U\left(c_j, h_j\right)\right) - \sum_j p^*_j \left(\sum_i d_{ji} Q_{ji} - Y_j \left(N^Y_j, I^Y_j\right)\right) \\
&- \sum_j P^*_j \left(I^G_j + L_j c_j + I^Y_j + I^H_j - Q\left(Q_{1j}, .., Q_{Jj}\right)\right) - \sum_j R^*_j \left(H^G_j + L_j h_j - H_j \left(N^H_j, I^H_j\right)\right) \\
&- \sum_j P^{G*}_j \left(G^Y_j + G^U_j - G_j \left(I^G_j, H^G_j\right)\right) - \sum_j W^*_j \left(N^I_j + N^H_j - N_j \left(G^Y_j, z_j\left(L_j\right) L_j\right)\right) \\
&- E \left(\sum_j L_j - L\right) + ...
\end{aligned} \tag{A.53}$$

Letting $x^P_j \equiv P^*_j c_j + R^*_j h_j$ be private expenditure, following the same steps as in the proof of Proposition 1 we find

$$x^P_j = \frac{1+\gamma^P}{1-\gamma^A} w_j - \frac{E}{1-\gamma^A}. \tag{A.54}$$

Combining the first order condition over c_j with optimization over government spending gives

$$\begin{aligned}
\left[G^U_j\right]: & \qquad L_j x^P_j \gamma^{A,G} = P^{G*}_j G^U_j \\
\left[G^Y_j\right]: & \qquad W^*_j N_j \varepsilon_{N,G} = P^{G*}_j G^Y_j
\end{aligned}$$

where $\gamma^{A,G} \equiv \frac{\partial a_j}{\partial G_j} \frac{G_j}{a_j}$ and $\varepsilon_{N,G} = \frac{\partial N}{\partial G_j} \frac{G_j}{N_j}$ are elasticities of amenities and labor efficiency to government spending.

Adding up the two previous equations, using the first-order conditions over I^G_j and H^G_j, and applying homogeneity of the the production function of government spending $G_j \left(I^G_j, H^G_j\right)$ we reach:

$$x_j = \left[\left(1+\gamma^{A,G}\right) \frac{1+\gamma^P}{1+\gamma^A} + \frac{\varepsilon_{N,G}}{\varepsilon_{N,L}}\right] w_j - \frac{1+\gamma^{A,G}}{1+\gamma^A} E, \tag{A.55}$$

where $x_j \equiv x^P_j + x^G_j$ is the sum of public and private expenditure per capita. We conclude that, in addition to the standard spillover elasticities, (A.55) now includes the elasticities of labor efficiency N_j with respect to government spending ($\varepsilon_{N,G}$) and the services of labor ($\varepsilon_{N,L}$), as well as the elasticity of amenities with respect to spending ($\gamma^{A,G}$). We again obtain that a simple linear relationship between expenditure per capita and wages must hold in an efficient allocation.

In addition to this necessary condition, efficiency now also requires an optimal breakdown of total expenditure into its private and public components. Further combining (A.56) with (A.54) gives

$$x_j^G = \gamma^{A,G} x_j^P + \frac{\varepsilon_{N,G}}{\varepsilon_{N,L}} w_j. \tag{A.56}$$

These efficiency conditions can be implemented through the a policy scheme (s, T) as before, now coupled with inter-governmental transfers. Importantly, the optimal subsidy s remains the same as in the absence of public spending, as implied by (23). Simultaneously, the optimal government expenditure in j is attained by an inter-governmental grant equal to $x_j^G L_j$. Hence, as long as government transfers across regions are flexible and efficient, bringing in government spending to the picture does not change the optimal spatial policy designed to deal with the spillovers.

A.9 Planning Problem in Relative Changes and Proof of Proposition 4

We show how to express the solution for the competitive allocation under an optimal new policy relative to an initial equilibrium consistent with Definition 1, and then define the planning problem that optimizes over the policy space.

Preliminaries We adopt the functional forms from Section 3.6. From the profit maximization problem of producers and market clearing in the housing market we obtain the following sectoral labor demand conditions:

$$W_i N_i^Y = \left(1 - b_{Y,i}^I\right) p_i Y_i, \tag{A.57}$$

$$W_i N_i^H = \frac{1 - b_{H,i}^I}{1 + d_{H,i}} \left(1 - \alpha_C\right) X_i. \tag{A.58}$$

These terms imply the non-traded labor share, $\frac{N_i^H}{N_i}$, as function of the share of gross expenditures over tradeable income $\frac{X_i}{p_i Y_i}$:

$$\frac{N_i^H}{N_i} = \frac{\frac{1-b_{H,i}^I}{1+d_{H,i}} \frac{1-\alpha_C}{1-b_{Y,i}^I} \left(\frac{X_i}{p_i Y_i}\right)}{\frac{1-b_H^I}{1+d_{H,i}} \frac{1-\alpha_C}{1-b_{Y,i}^I} \left(\frac{X_i}{p_i Y_i}\right) + 1}. \tag{A.59}$$

Using (A.57) and (A.58) along with labor-market clearing (A.15), we can further express final consumption expenditures over tradeable income as a function of the shares of wages in expenditures:

$$\frac{X_i}{p_i Y_i} = \frac{1 - b_{Y,i}^I}{\frac{W_i N_i}{X_i} - \frac{1-b_{H,i}^I}{1+d_{H,i}} \left(1 - \alpha_C\right)}. \tag{A.60}$$

We now re-formulate some of the equilibrium from Definition 1 conditions to include prices. Consider first the market clearing condition (7). Multiplying both sides by the price of the traded bundle P_j, letting $E_j^Y \equiv P_j Q_j$ be the gross expenditures in tradeable goods in j (used both as intermediate and for final consumption), and using equilibrium in the housing market and the optimality condition for the choice of intermediate inputs in the traded sector, we can re-write that condition as

$$E_j^Y = \left(\alpha_C + \left(1 - \alpha_C\right) \frac{b_{H,j}^I}{d_{H,j} + 1}\right) X_j + b_Y^I \left(p_j Y_j\right), \tag{A.61}$$

where $X_j = \sum_{\theta'} L_j^{\theta'} x_j^{\theta'}$ are the aggregate expenditures of workers in region j. This condition says that aggregate expenditures in traded goods results from the aggregation of expenditures by consumers and final producers. Second, consider the market condition (6) for traded commodities. Multiplying both sides by the price of traded commodities at j, p_j, this condition is equivalent to

$$\sum_i s_{ji}^X = 1, \tag{A.62}$$

where $s_{ji}^X \equiv \left(\frac{E_i}{p_j Y_j}\right) s_{ji}^M$ is region i's share of j's sales of tradeable goods (i.e., the export share of i in j) and $s_{ji}^M \equiv \frac{p_{ji} Q_{ji}}{E_i}$ is region $j's$ share of i's purchases of tradeable goods (i.e., the import share of region j in i). Finally, aggregating the budget constraints of individual consumers gives

$$\sum_j s_{ji}^M \equiv 1. \tag{A.63}$$

Equilibrium in Relative Changes We now express the solution for the competitive allocation from Definition 1 under the new policy relative to an initial equilibrium. Consider a policy change that affects the equilibrium expenditure distribution $\{x_i^\theta\}$. We now show that the outcomes in the new equilibrium relative to the initial equilibrium are given by a set of changes in prices $\left\{\hat{P}_i, \hat{p}_i, \hat{R}_i\right\}$, wages $\left\{\hat{W}_i\right\}$, employment by group $\left\{\hat{L_i^\theta}\right\}$, supply of efficiency units $\left\{\hat{N}_i\right\}$, production of tradeable goods $\left\{\hat{Y}_i\right\}$, and utility levels $\left\{\hat{u^\theta}\right\}$ that satisfy a set of conditions given the change in expenditure per capita by group and location $\left\{\hat{x_i^\theta}\right\}$. The planner's problem in relative changes will then choose the optimal $\left\{\hat{x_i^\theta}\right\}$.

From the previous expressions we obtain the following system in relative changes:

$$\sum_j s_{ij}^X \left(\frac{\hat{p}_i}{\hat{P}_j}\right)^{1-\sigma} \hat{E_j^Y} = \hat{p}_i \hat{Y}_i \text{ for all } i, \tag{A.64}$$

$$\sum_j s_{ji}^M \left(\frac{\hat{p}_j}{\hat{P}_i}\right)^{1-\sigma} = 1 \text{ for all } i, \tag{A.65}$$

$$\left(1 - \frac{N_i^H}{N_i}\right) \hat{p}_i \hat{Y}_i + \frac{N_i^H}{N_i} \hat{X}_i = \hat{W}_i \hat{N}_i \text{ for all } i, \tag{A.66}$$

$$\hat{W}_i^{1-b_{Y,i}^I} \hat{P}_i^{b_{Y,i}^I} = \hat{p}_i \text{ for all } i, \tag{A.67}$$

where $\hat{X}_j = \sum_\theta s_j^{X,\theta} \hat{x_j^\theta} \hat{L_j^\theta}$ is the change in aggregate expenditures by region and $s_j^{X,\theta}$ is group θ's share in the consumer expenditures in j in the initial equilibrium. Equations (A.64) and (A.65) follow from expressing (A.62) and (A.63) in relative changes and using the CES functional form (36). In condition (A.64), using (A.61) implies that the change in expenditures in tradeable commodities is:

$$\hat{E_j^Y} = \left(1 - b_{Y,j}^{\tilde{I}}\right) \hat{X}_j + b_{Y,j}^{\tilde{I}} \hat{p}_j \hat{Y}_j. \tag{A.68}$$

where

$$b_{Y,j}^{\tilde{I}} \equiv b_Y^I \frac{p_j Y_j}{E_j^Y} = \frac{b_Y^I}{\left(\alpha_C + (1-\alpha_C) \frac{b_{H,j}^I}{d_{H,j}+1}\right) \frac{X_j}{p_j Y_j} + b_Y^I} \tag{A.69}$$

Condition (A.66) follows from expressing labor-market clearing (9) in relative changes together with (A.57) and (A.58), where the non-traded labor share $\frac{N_i^H}{N_i}$ is defined in (A.59). Condition (A.67) follows from optimization of producers of tradeable commodities.

The system (A.64) to (A.67) defines a solution for $\left\{\hat{P}_j, \hat{p}_j, \hat{Y}_j, \hat{W}_j\right\}$ given the change in the number of efficiency units $\hat{N}_i$ and expenditures in each region $\hat{X}_i$, and independently from heterogeneity across groups or spillovers. Heterogeneous groups and spillovers enter through $\hat{N}_i$. To reach an explicitly expression for $\hat{N}_i$, we first note that the labor demand expression in the market allocation (14) allows us to back out the efficiency of each group:

$$z_i^\theta = \frac{w_i^\theta}{W_i} \left(\frac{L_i^\theta}{N_i}\right)^{\frac{1}{\rho}}, \tag{A.70}$$

Expressing the CES functional form for the aggregation of labor types in (39) in relative changes and using (A.70) we obtain:

$$\hat{N}_i = \left(\sum_{\theta} s_i^{W,\theta} \left(\hat{z_i^\theta} \hat{L_i^\theta} \right)^{\rho} \right)^{\frac{1}{\rho}}, \tag{A.71}$$

where

$$\hat{z_i^\theta} = \prod_{\theta'} \left(\hat{L_i^{\theta'}} \right)^{\gamma^P_{\theta',\theta}} \tag{A.72}$$

and where $s_j^{W,\theta} = \frac{w_j^\theta L_j^\theta}{\sum_{\theta'} w_j^{\theta'} L_j^{\theta'}}$ is group θ share of wages in city j. Expression (A.71) relates the total change in efficiency units in a location to the distribution of wage bills in the observed allocation, the changes in employment by group, and the production function and spillover elasticity parameters.

The change in the number of workers $\left\{ \hat{L_i^\theta} \right\}$ of each type in every location that is initially populated must also be consistent with the spatial mobility constraint, (13),

$$\hat{u^\theta} = \hat{a_i^\theta} \frac{\hat{x_i^\theta}}{\hat{P_i}^{\alpha_C} \hat{R_i}^{1-\alpha_C}}, \tag{A.73}$$

where

$$\hat{a_i^\theta} = \prod_{\theta'} \left(\hat{L_i^{\theta'}} \right)^{\gamma^A_{\theta',\theta}} \tag{A.74}$$

and where $\hat{R}_i$ is the change in the price of non-traded goods in location i. This relative price can be expressed as solely a function of the changes in the price of the own traded good, the price index of traded commodities, and the aggregate expenditures in i:

$$\hat{R}_i = \left(\hat{p_i}^{\frac{1-b^I_{H,i}}{1-b^I_{Y,i}}} \hat{P}_i^{b^I_{H,i} - b^I_{Y,i} \frac{1-b^I_{H,i}}{1-b^I_{Y,i}}} \hat{X}_i^{d_{H,i}} \right)^{\frac{1}{1+d_{H,i}}}. \tag{A.75}$$

To obtain this expression, we first solved for the rental rate R_i from the equilibrium in the housing market, used the zero-profit condition in the traded sector and expressed the resulting expression in relative changes.

Finally, the national labor market must clear for each labor type is

$$\sum_j s_j^{L,\theta} \hat{L_j^\theta} = 1 \text{ for all } \theta, \tag{A.76}$$

where $s_j^{L,\theta} = \frac{L_j^\theta}{\sum_{\theta'} L_j^{\theta'}}$ is group θ's share of employment in city j.

In sum, the system of equilibrium equations can be broken into two distinct blocks. The system (A.64) to (A.67) defines a solution for $\left\{ \hat{P}_j, \hat{p}_j, \hat{Y}_j, \hat{W}_j \right\}$ given the change in the number of efficiency units $\hat{N}_i$ and expenditures in each region $\hat{X}_i$ independently from heterogeneity across groups or spillovers. In turn, the system (A.71) to (A.76) defines a solution for $\left\{ \hat{N}_j, \hat{L}_j^\theta, \hat{u^\theta} \right\}$ given $\left\{ \hat{p_i}, \hat{P}_i, \left\{ \hat{x_i^\theta} \right\}, \hat{X}_i \right\}$. As a result, an equilibrium in changes given a change in expenditure per capita $\left\{ \hat{x_j^\theta} \right\}$ consists of $\left\{ \hat{P}_i, \hat{p}_i, \hat{Y}_i, \hat{W}_i, \hat{N}_j, \hat{L_j^\theta}, \hat{R}_i, \hat{u^\theta} \right\}$ such that equations (A.64) to (A.76) hold. These equations conform a system of $5J + \Theta J + \Theta$ equations in equal number of unknowns, where J is the number of locations and Θ is the number of types.

Planner's Problem in Relative Changes In the implementation, we solve an optimization over $\left\{ \hat{x_j^\theta} \right\}$ subject to $\left\{ \hat{P}_i, \hat{p}_i, \hat{Y}_i, \hat{W}_i, \hat{N}_j, \hat{L_j^\theta}, \hat{R}_i, \hat{u^\theta} \right\}$ consistent with (A.64) to (A.76) in order to maximize the utility of a given group θ, $\hat{u^\theta}$, subject to a lower bound for the change in utility of the other groups ($\hat{u^{\theta'}} \geq \underline{\hat{u^{\theta'}}}$ for $\theta' \neq \theta$). This problem (call it $\mathcal{P}_2''$) differs formally from the baseline problem in Definition 2 (call it $\mathcal{P}_2$) for two reasons. First, it features prices, expenditures and incomes rather than being expressed in terms of quantities alone, as in conditions (A.57) to (A.63). We denote by $\mathcal{P}_2'$ an intermediary problem expressed in terms of income and expenditure rather than quantities, but still in levels. Second, $\mathcal{P}_2''$ is expressed in changes relative to an initial equilibrium rather than in levels. We show here that the two problems are nevertheless equivalent. Therefore, the problem that we implement has a unique maximizer under the conditions of Proposition 2.

To see that the two problems have the same solutions, we first focus on the first order conditions of problem $\mathcal{P}_2$ and compare them to the problem in levels $\mathcal{P}_2'$ expressed in income and expenditures terms rather than in quantities. Conditions (A.14) and (A.16) define the Lagrange multipliers corresponding to good and factor prices for $\mathcal{P}_2$. They are identical to the price index definition constraint of problem $\mathcal{P}_2'$. Furthermore, manipulating these equations together with the constraints expressed in quantities leads to the constraints expressed in terms of income and expenditure. Therefore, a vector satisfies the first order conditions for $\mathcal{P}_2$ if and only if it satisfies the first order conditions for $\mathcal{P}_2$'. Then, note that the problem in relative changes stated here is simply the problem $\mathcal{P}_2'$ modified through the changes of variable $x \to x_o \widehat{x}$ for all variables, where x_o is a constant corresponding to the observed data and $\widehat{x}$ the optimization variable in $\mathcal{P}_2'$. The problem in relative changes considered here and the problem $\mathcal{P}_2'$, and in turn problem $\mathcal{P}_2$, have therefore the same solutions, subject to the appropriate change of variables. In particular, a point that satisfies the first order conditions under the conditions of Proposition 3 is the (unique) global maximizer for both problems.

Proof of Proposition 4 Proposition 4 follows from inspecting (A.64) to (A.76) under the planner's problem in relative changes defined above. Note that, given the elasticities $\left\{\alpha_C, \rho, b^I_{Y,j}, b^I_{H,j}, d_{H,j}\right\}$, and as long as $b^I_Y > 0$, computing the change in tradeable expenditures requires information about gross expenditures over tradeable income, $\frac{X_j}{p_j Y_j}$. This information is also needed to compute the non-traded labor share $\frac{N_i^H}{N_i}$ in (A.66). However, as shown in (A.59) and (A.60), $\frac{X_j}{p_j Y_j}$ can be constructed from the elasticities $\left\{\alpha_C, b^I_Y, b^I_H, d_{H,j}\right\}$ and the share of wages in gross expenditures, $\frac{W_i N_i}{X_i}$.

A.10 Fiscal Competition

Consider a version of our model with homogeneous workers, homogeneous and perfectly traded products, and exogenous housing supply. Assume, in addition, that workers face idiosyncratic preference draws, as in Section 3.5, and that the spillovers have constant elasticities. Assuming a tax τ_j^H on the returns to land, when the tax revenue is rebated back to workers then expenditure per worker is $x_j = \left(w_j + \frac{\Pi}{L}\right) + t_j^H \frac{R_j H_j}{L_j}$, where the returns to land net of taxes are $\frac{\Pi}{L} = \frac{\sum_{j'} \left(1-\tau_{j'}^H\right) R_{j'} H_{j'}}{L}$, implying:

$$x_j = Z_j L_j^{\gamma_P} + \sum_{j' \neq j} \left(1 - \tau_{j'}^H\right) R_{j'} H_{j'} + \left(1 + \tau_j^H \frac{1 - L_j}{L_j}\right) R_j H_j, \tag{A.77}$$

where total population was normalized to 1. Further using equilibrium in housing markets, we obtain:

$$x_j = x_j \left(\tau_j^H, u_j\right) \equiv \frac{Z_j L_j \left(u_j\right)^{\gamma_P} + \frac{1}{L} \sum_{j' \neq j} \left(1 - \tau_{j'}^H\right) R_{j'} H_{j'}}{1 - \alpha \left(L_j \left(u_j\right) + \tau_j^H \left(1 - L_j \left(u_j\right)\right)\right)}, \tag{A.78}$$

where labor supply to region j is $L_j\left(u_j\right) = \left(\frac{u_j}{u}\right)^{1/\sigma}$. Given the Fréchet shocks assumption, u_j is the average utility of the residents of location j. Furthermore, using the definition of u_j in (35), under the current assumptions we obtain $u_j = \left(K_j x_j \left(\tau_j^H, u_j\right)^{\alpha}\right)^{\frac{\sigma}{\sigma+1-\alpha_C-\gamma^A}}$, where $K_j = \left(H_j/\alpha_C\right)^{1-\alpha_C} A_j u^{-\left(\gamma^A+\alpha_C-1\right)/\sigma}$ is a constant. Assuming the local government takes country-level utility u as given, we get $\frac{du_j}{d\tau_j^H} > 0$ if $\gamma^A < \sigma + 1 - \alpha_C$ and $\frac{\partial K_j x_j \left(\tau_j^H, L_j\left(u_j\right)\right)^{\frac{\alpha}{1-\left(\gamma^A+\alpha-1\right)/\sigma}}}{\partial u_j} < 1$, which can be ensured by properly normalizing the constant K_j. Under these conditions, a local government who maximizes u_j will maximize the tax rate τ_j^H and therefore set it equal to 1. These policies yield $x_j = w_j + \frac{R_j H_j}{L_j}$, which further implies $c_j = w_j$, or no equilibrium imbalances.

B Data Appendix

We detail the construction of the variables used to implement the counterfactuals. We rely on four primary data sources: i) BEA regional economic accounts, CA4 Personal Income and Employment by Major Component (https://www.bea.gov/regional/downloadzip.cfm); ii) estimates of disposable income by MSA from Dunbar (2009) based on BEA regional economic accounts;[47] iii) March CPS based on the IPUMS-CPS, ASEC 2007-2012 samples and iv) IPUMS-ACS, 2007-2012 samples.

B.1 Appendix to Section 4.1 (Data)

MSA-Level Outcomes We first extract from Dunbar (2009) the following information: *population*, *personal income*, and *personal taxes* paid by MSA, in 2007. To split personal income by source of income, we merge this data with the BEA Regional Economic Accounts. We compute the share of personal income corresponding to each possible source: labor income, capital income, and transfers. Specifically, we measure labor income as BEA's *earning by place of work*;[48] capital income as the sum of *proprietor's income*, and *dividends, interests and rents*; and transfers as *current transfer receipts*.[49] Combining these shares with the total personal income and taxes by MSA from Dunbar (2009) provides us with a measure of labor income, capital income, transfers and taxes at the MSA level.

Break-Down By Skill Group We split these totals at the MSA level into two groups, high skill and low skill. To that end, we use the ACS data, part of the Integrated Public Use Microdata Series (Flood et al., 2017), for the years 2007-2012.The ACS reports, at the individual level, labor income, capital income, government transfers,, MSA of residence and level of education. Consistent with Diamond (2016), we define as high skill those workers who have completed 4 years of college, or more; and as low skill those who have completed less than 4 years of college, or have not gone to college. We aggregate individual level data from the ACS at the MSA-group level, to get an MSA-level estimate of capital income, labor income and transfers by group, as well as the population of both groups.[50] To compute taxes paid by group and by city, we follow a similar procedure, using taxes reported in the March CPS.[51] We do not use this information directly, as the MSA aggregates from individual-level data might be noisy, in particular for smaller MSAs. Instead, we use this information to construct the shares of the MSA-level outcomes from the BEA described above corresponding to each group of workers. That is, for each MSA i, we compute $s_i^L = \frac{\tilde{X}_i^L}{\tilde{X}_i^L + \tilde{X}_i^H}$ where $\widetilde{X}_i^\theta$ denotes total MSA i-group θ level capital income, labor income, transfers, taxes or population in the census data. We use this share s_i^L, together with the MSA-level dataset for income described above, to build our measure $X_i^\theta = X_i s_i^\theta$ of MSA-group level population, labor income, capital income, transfers and taxes. We also compute the corresponding per-capita measures for each MSA-group: $x_i^\theta = \frac{X_i^\theta}{L_i^\theta}$.

[47] https://www.bea.gov/papers/xls/dpi_msa_working_paper_2001_2007_results.xls

[48] The BEA's earning by place of work is comprised of: wages and salaries, supplements to wages and salaries, proprietor's income, net of contributions for government social insurance, plus adjustment for residence.

[49] Current transfer receipts is defined as the sum of government social benefits and net current transfer receipts from business (https://www.bea.gov/glossary/glossary_p.htm).

[50] One may be worried that the ACS transfers measure suffer from systematic under-reporting issues (Meyer et al., 2009). An alternative way to compute transfers is to use an accounting approach. If one allocates social security (old age) to 65+, in the proportion of labor earnings, Medicare in the proportion of +65 individuals, and the rest of transfers (Medicaid, UI, VA, etc.) to low skill only, the correlation between the transfers per capita received by the low skill in the two methods is .96.

[51] Specifically, we aggregate the following categories to measure capital income: income from interest, from dividends, from rents. We aggregate the following categories to measure labor income: wage and salary income, non-farm business income, farm income, income from worker's compensation, alimony and child support. We aggregate the following categories to measure transfers: welfare income, social security income, income from SSI, income from unemployment benefits, income from veteran's, survivor's, disability benefit, income from educational assistance. We aggregate the following categories to measure taxes paid: federal income tax liability, after all credits, and state income tax liability, after all credits.

Controlling for Heterogeneity within Groups Before applying it in the quantification we purge the raw data described above from compositional effects across MSAs. We use the ACS data to obtain the share of individuals with the following characteristics for each MSA-skill group: age by bins: <20, 20-40, 40-60, >60; detailed level of educational attainment: less that 8th grade, grade 9-12, some college (those are relevant for the low skill group) and bachelor, masters or professional degree (for the high skill group); share black; share male ; share unemployed, share out of the labor force; and share working in manufacturing, services, agriculture. We also use hours worked per capita as another control. We then proceed as follows: denoting by x_i^θ the per-capita measure $\frac{X_i^\theta}{L_i^\theta}$ (where X_i^θ can stand for labor income, capital income, transfers or taxes) in MSA i and group θ, we run the following MSA level regression, separately for each group θ:

$$x_i^\theta = x_0^\theta + \sum_j \beta_j^\theta DEM_{ij}^\theta + \varepsilon_i^\theta, \tag{A.79}$$

where DEM_{ij}^θ is the demographic variable j enumerated above in MSA i and for group θ. The coefficients β_j^θ measures how demographic characteristic j correlates with x_i^θ across cities within group θ. We then adjust the observed x_i^θ from compositional differences across cities measured as deviations from the population mean:

$$\hat{x}_i^\theta \equiv x_i^\theta - \sum_j \hat{\beta}_j^\theta \left(DEM_{ij}^\theta - \overline{DEM_j^\theta} \right) \tag{A.80}$$

where $\hat{\beta}_j^\theta$ is the estimate from (A.79) and $\overline{DEM_j^\theta} \equiv \frac{1}{I}\sum_i DEM_{ij}^\theta$.[52] The corresponding MSA-level variable is $\hat{X}_i^\theta = \hat{x}_i^\theta L_i^\theta$. The resulting data $(\hat{X}_i^\theta, \hat{x}_i^\theta, L_i^\theta)$ is our MSA-group level dataset, where X stands for labor income, capital income, transfers and taxes.

Expenditure per Capita We construct expenditure by group and by MSA, x_i^θ in the model, as disposable income by group. Disposable income is

$$x_i^\theta = w_i^\theta - \tau_i^\theta + T_i^\theta + b^\theta \Pi^H. \tag{A.81}$$

The variables $\left\{w_i^\theta, \tau_i^\theta, T_i^\theta\right\}$, respectively labor income per capita, tax paid per capita, and transfer received per capita, are directly taken from the BEA/ACS dataset constructed above. We measure b^θ as the average fraction of national capital income owned by each type θ worker in BEA/ACS dataset. This step gives $b^S L^S = 0.52$ and $b^U L^U = 0.48$. Finally, we set a value for national profits and returns to land Π^H that is consistent with the general equilibrium of the model. Using profit maximization in the housing sector and market clearing in the non-tradeable sector we obtain the following expression for Π^H as function of calibrated elasticities and observable outcomes:

$$\Pi^H = \frac{(1-\alpha_C)\sum_i \frac{d_{H,i}}{d_{H,i}+1} \sum_\theta L_i^\theta \left(w_i^\theta - \tau_i^\theta + T_i^\theta\right)}{1-(1-\alpha_C)\sum_i \frac{d_{H,i}}{d_{H,i}+1}\sum_\theta b^\theta L_i^\theta}. \tag{A.82}$$

Using x_i^θ we then construct X_i (aggregate expenditure by MSA) as $X_i = \sum L_i^\theta x_i^\theta$ and $s_j^{X,\theta}$ (share of expenditures by type within MSA). Following these adjustments, we need to ensure that the sum of transfers paid by the government equal the sum of taxes levied. To that end, we scale all transfers uniformly so that they add up to the sum of taxes. This ensures that the government budget constraint holds.[53]

Traded and Non-Traded Sectors We need data on the relative size of the non-traded sector in each city to calibrate the labor shares by sector. The ACS data also reports the sector of activity of workers. We measure at the MSA level the share of workers who work in the non traded sector by counting all workers in the following NAICS

[52] I.e., we define $\hat{x}_i^\theta \equiv \hat{x}_0^\theta + \sum_j \hat{\beta}_j^\theta \overline{DEM_j^\theta} + \hat{\varepsilon}_i^\theta$, where $\hat{\varepsilon}_i^\theta$ is the estimated residual from (A.79).

[53] This step implies that transfers are uniformly scaled down by 35%. The fact that total taxes and transfers do not match in our dataset comes in part from having removed heterogeneity that is not place-specific from the data and from our treatment of capital to be consistent with the sources of capital income (profits from housing rents), which scales down its share in income relative to the data.

sectors: retail, real estate, construction, education, health, entertainment, hotels and restaurants. This measure is not group-specific. To remove unmodeled heterogeneity in this measure, we compute a series of MSA-level socio-demographic characteristics, as above, and regress the share of workers in the non-traded sector on these demographic characteristics. We compute, as above, the predicted share of workers in the non-traded sector in each city, assuming that demographic characteristics of the city are at the nationwide mean.

Trade Shares We need data on trade shares between MSAs, s_{ij}^M and s_{ij}^X (import and export shares). These flows are observed in the CFS data, but not at the finer geographic level that we consider here (MSA). Therefore, we adapt the procedure in Allen and Arkolakis (2014), whereby the import shares from the CFS data are used to parametrize the elasticity of trade with respect to distance. In particular, the model implies the following expression for share of location i's imports originating from j:

$$s_{ji}^M = \left(\frac{d_{ji}}{P_i} \frac{W_j^{1-b_Y^I} P_j^{b_Y^I}}{z_j} \right)^{1-\sigma} \equiv \left(d_{ji} \delta_j^D \delta_i^O \right)^{1-\sigma}, \tag{A.83}$$

where δ_i^O and δ_j^D are origin and destination fixed effects. We assume that trade costs have the form $\ln d_{ji} = \psi \ln dist_{ji} + e_{ji}$, where $dist_{ji}$ is the great circle distance between MSAs j and i. We the use Allen and Arkolakis (2014) estimate for ψ and set trade costs to $d_{ji} = dist_{ji}^{\psi}$. We then construct the smoothed import shares s_{ji}^M between MSAs using (A.83). To that end we must obtain the values of $\{\delta_j^D, \delta_i^O\}$, which are uniquely pinned down, up to a normalization, by considering the identity that sales equals income,

$$p_j Y_j = \sum_i s_{ji}^M E_i, \tag{A.84}$$

together with equation (A.83) and the definition of the price index, leading to:

$$\left(\delta_i^O \right)^{\sigma-1} = \sum_j \left(d_{ji} \delta_j^D \right)^{1-\sigma}. \tag{A.85}$$

Plugging (A.83) and (A.85) in (A.84), we get a system N equations in N unknowns, which we solve to recover $\{\delta_j^D, \delta_i^O\}$ and in turn s_{ji}^M. The export shares are then constructed using $s_{ji}^X \equiv \left(\frac{E_i}{p_j Y_j} \right) s_{ji}^M$, where spending E_i and traded income $p_j Y_j$.

B.2 Appendix to Section 4.2 (Calibration)

Intermediate Input Shares We provide details about the calibration of the intermediate input share in non-traded goods. We use the following equilibrium relationship from the market clearing condition in the non traded sector in city j:

$$1 - b_{H,j}^I = \frac{W_j N_j^H}{(1-\alpha_C) X_j} \left(1 + d_{H,j} \right). \tag{A.86}$$

We compute this expression using the observed wage bill of workers in non-traded sectors $W_j N_j^H$ and total expenditure X_j described in the previous subsection, and our calibrated values for α_C and $d_{H,j}$ described in Section 4.2.

Efficiency Spillover Elasticities The standard estimate of city-level spillovers reviewed by Combes and Gobillon (2015) are obtained from a regression of average city wages w_j on city population L_j. In log-changes, such an equation would take the form: $\hat{w}_j = \gamma^P \hat{L}_j + \psi_j$, where ψ_j is a city effect and γ^P is the city-level spillover elasticity. In our environment, city-level wages are $w_j L_j = N_j W_j$. Under the assumptions of the quantitative model, applying (A.71), an exogenous shift in the total population of city j keeping its composition across groups constant would then imply:

$$\hat{w}_j = \left[s_j^{W,S} \left(\gamma_{S,S}^P + \gamma_{U,S}^P \right) + \left(1 - s_j^{W,S} \right) \left(\gamma_{S,U}^P + \gamma_{U,U}^P \right) \right] \hat{L}_j + \hat{W}_j, \tag{A.87}$$

where $s_j^{W,S}$ is the share of skilled workers in wages in city j. Hence, through the lens of our model, the coefficient γ^P estimated at the city level in the empirical literature would correspond to $\overline{s^{W,S}}\left(\gamma_{S,S}^P+\gamma_{U,S}^P\right)+\left(1-\overline{s^{W,S}}\right)\left(\gamma_{S,U}^P+\gamma_{U,U}^P\right)$, where $\overline{s^{W,S}}$ is the average skilled worker share across cities. Therefore, we uniformly normalize the distribution of the $\gamma_{\theta,\theta'}^P$ coefficients such that, under their scaled values, $\overline{s^{W,S}}\left(\gamma_{S,S}^P+\gamma_{U,S}^P\right)+\left(1-\overline{s^{W,S}}\right)\left(\gamma_{S,U}^P+\gamma_{U,U}^P\right)=\gamma^P$. We set $\gamma^P=0.06$, which is consistent with the standard estimate for the U.S. from Ciccone and Hall (1996), and $\overline{s^{W,S}}=0.49$ as observed in our data.

Having chosen the level of the $\gamma_{\theta,\theta'}^P$ coefficients, we must still choose their distribution. Under the assumptions of the quantitative model, the labor demand condition (14) gives the following expression for the log wage of type-θ worker:

$$\ln w_j^\theta=\left[\rho\left(1+\gamma_{\theta,\theta}^P\right)-1\right]\ln\left(L_j^\theta\right)+\rho\gamma_{\theta',\theta}^P\ln\left(L_j^{\theta'}\right)+\ln W_j-(\rho-1)\ln N_j+\ln\varepsilon_j^\theta, \tag{A.88}$$

where $\ln\varepsilon_j^\theta=\rho\ln Z_j^\theta$ captures productivity shocks at the worker-city level. In data generated by this model and expressed in differences over time, we would have

$$\Delta\ln w_j^\theta=\left[\rho\left(1+\gamma_{\theta,\theta}^P\right)-1\right]\Delta\ln\left(L_j^\theta\right)+\rho\gamma_{\theta',\theta}^P\Delta\ln\left(L_j^{\theta'}\right)+\Delta\kappa_j+\Delta\ln\varepsilon_j^\theta, \tag{A.89}$$

where $\Delta\kappa_j=\Delta\ln W_j-(\rho-1)\Delta\ln N_j$ is a city effect. We can use (A.89) to map estimates from Diamond (2016). Specifically, she estimates equations (27) and (28) in her paper using Bartik shocks as instruments. The only difference between these equations in her paper and (A.89) is the fixed effect $\Delta\kappa_j$ here. Assuming that the inclusion of the fixed effect $\Delta\kappa_j$ would not alter Diamond (2016) estimates, we can directly map her estimates from Column 3 of Table 5, i.e. $\rho\left(1+\gamma_{S,S}^P\right)-1=0.229$, $\rho\gamma_{U,S}^P=0.312$, $\rho\left(1+\gamma_{U,U}^P\right)-1=-0.552$, $\rho\gamma_{S,U}^P=0.697$.

The elasticities resulting from this procedure are reported in the first row of Table A.1. The second row reports the coefficients from an alternative parametrization used in the quantitative section where we target $\gamma^P=0.12$ instead of $\gamma^P=0.06$.

Parametrization	γ_{UU}^P	γ_{SU}^P	γ_{US}^P	γ_{SS}^P
Benchmark	0.003	0.044	0.020	0.053
High Efficiency Spillover	0.007	0.087	0.039	0.106

Table A.1: Alternative Parametrizations of Efficiency Spillovers

Amenity Spillover Elasticities Diamond (2016) reports estimates for equation (31) in her paper, which (using our notation for the variables in common with her analysis) has the form:

$$\Delta\ln L_j^\theta=a_0^\theta\Delta\ln\left(\frac{w_j^\theta}{P_j}\right)+a_1^\theta\Delta\ln\left(\frac{R_j}{P_j}\right)+a_2^\theta\Delta\ln\left(a_j^D\right)+\Delta\xi_j^\theta, \tag{A.90}$$

where $a_j^D\equiv\left(L_j^S/L_j^U\right)^{\gamma^a}$ is the endogenous component of amenities in her analysis[54] and (a_0,a_1,a_2) are estimated coefficients. Column (3) of Table 5 of Diamond (2016) reports the following estimates: $\left(a_0^U,a_0^S,a_2^U,a_2^S,\gamma^a\right)=(4.026,2.116,0.274,1.012,2.6)$. We generate equation (A.90) in our setup and match the coefficients from our model to these estimates. For generality, we do so allowing for idiosyncratic preference draws within each type as in Section 3.5 (i.e., assuming $\sigma_\theta>0$). The labor-supply equation implied by (30) is

$$\sigma_\theta\ln L_j^\theta=\ln\left(\frac{x_j^\theta}{P_j}\right)-(1-\alpha_C)\ln\left(\frac{R_j}{P_j}\right)+\ln\left(a_j^\theta\right)+\left(\sigma_\theta\ln L^\theta-\ln u^\theta\right). \tag{A.91}$$

Let $\zeta^{A,S}=\gamma^a$ and $\zeta^{A,U}=-\gamma^a$, and then redefine our amenity index a_j^θ for $\theta=U,S$ in (41) as a function of

[54]This index captures congestion in transport, crime, environmental indicators, supply per capita of different public services, and variety of retail stores. See Table 4 of Diamond (2016).

the amenity index a_j^D from Diamond (2016) as follows: $a_j^\theta = A_j^\theta \left(L_j^\theta\right)^{\gamma_{\theta,\theta}^A - \beta^{a,\theta}\zeta^{A,\theta}} a_j^D$, where $\beta^{a,\theta} \equiv \frac{\gamma_{\theta',\theta}^A}{\zeta^{A,\theta'}}$ is by construction constant over θ'. Using this equivalence in (A.91), re-arranging and expressing that equation in changes we obtain

$$\Delta \ln L_j^\theta = \frac{1}{\left(\sigma_\theta - \gamma_{\theta,\theta}^A\right) + \beta^{a,\theta}\zeta^{A,\theta}} \Delta \ln\left(\frac{x_j^\theta}{P_j}\right) - \frac{1-\alpha_C}{\left(\sigma_\theta - \gamma_{\theta,\theta}^A\right) + \beta^{a,\theta}\zeta^{A,\theta}} \Delta \ln\left(\frac{R_j}{P_j}\right)$$
$$+ \frac{\beta^{a,\theta}}{\left(\sigma_\theta - \gamma_{\theta,\theta}^A\right) + \beta^{a,\theta}\zeta^{A,\theta}} \Delta \ln\left(a_j^D\right) + \Delta\xi_j^\theta, \tag{A.92}$$

where $\Delta\xi_j^\theta \equiv \frac{1}{\left(\sigma_\theta - \gamma_{\theta,\theta}^A\right) + \beta^{a,\theta}\zeta^{A,\theta}} \left(\ln A_j^\theta + \sigma_\theta \ln L^\theta - \ln u^\theta\right)$. Comparing (A.90) with (A.92) readily allows us to map Diamond (2016) estimates to our parameters as follows:

$$\gamma_{\theta,\theta}^A - \sigma_\theta = \frac{a_2^\theta}{a_0^\theta}\zeta^{A,\theta} - \frac{1}{a_0^\theta}, \tag{A.93}$$

$$\gamma_{\theta',\theta}^A = \frac{a_2^\theta}{a_0^\theta}\zeta^{A,\theta'} \tag{A.94}$$

for $\theta = U, S$. Conditional the estimates of $\left(a_0^U, a_0^S, a_2^U, a_2^S, \gamma^a\right)$, we back out the value of $\gamma_{\theta,\theta}^A - \sigma_\theta$ but are unable to distinguish $\gamma_{\theta,\theta}^A$ from $-\sigma_\theta$. Our benchmark model is presented assuming $\sigma_\theta = 0$. However, as discussed in Section 3.5, $\gamma_{\theta,\theta}^A - \sigma_\theta$ is the relevant combination of parameters to characterize optimal allocations and policies under the definition of the planner problem with idiosyncratic preference draws defined in that section.

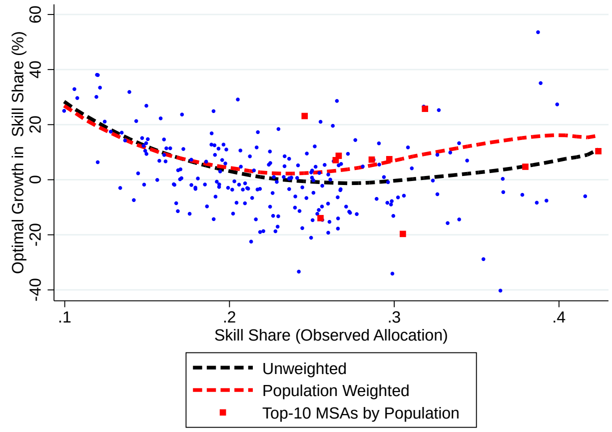
The resulting numbers are reported in the first row of Table A.2. The second row reports the coefficients from an alternative parametrization used in the quantitative section where we scale all amenity spillovers down by 50% relative to the benchmark. The third and fourth rows report parametrizations that, instead the coefficient $\gamma^a = 2.6$ reported in Column (3) of Table 5 of Diamond (2016), use that point estimate plus or minus the standard deviation reported in that table, respectively.

	γ_{UU}^A	γ_{SU}^A	γ_{US}^A	γ_{SS}^A
Benchmark	-0.43	0.18	-1.24	0.77
Low amenity spillover	-0.21	0.09	-0.62	0.38
High cross-amenity spillover	-0.46	0.22	-1.51	1.04
Low cross-amenity spillover	-0.39	0.14	-0.97	0.50

Table A.2: Alternative Parametrizations of Amenity Spillovers

Additional Optimal Sorting Figures

Figure A.1: Change in Skill Share between Data and Optimal Allocation



Note: each point in the figure corresponds to an MSA. The figure shows unweighted and initial population-weighted non-parametric curves. The 10 largest cities in the initial allocation are shown as red squares.

C Online Appendix: Alternative Model Specifications

For each alternative specification we first discuss how the system (A.64) to (A.76) in Appendix Section A.9 used to solve for the optimal allocation is modified. In each case, we only refer to the equations that are modified compared to the baseline. We then describe for each case the details of the calibration.

C.1 Homogeneous Workers

Model The system (A.64) to (A.76) remains the same but is applied for the case of only one skill type.

Calibration We use the same aggregate MSA-level variables constructed for the case with heterogeneous workers. To determine the spillover elasticities, we set one-group elasticities $\left(\gamma^A, \gamma^P\right)$ to the value that would be estimated through the lens of the labor supply and demand equations of the single-group model, if one were to use an MSA-level dataset generated by the model with heterogeneous groups and elasticities $\{\gamma^P_{\theta',\theta}\}$ and $\{\gamma^P_{\theta',\theta}\}$ calibrated above. This procedure by construction delivers $\gamma^P = 0.06$, equal to the value drawn from Ciccone and Hall (1996). To set γ^A we note that under a single worker type, the labor-supply equation implied by (13) expressed in time differences becomes

$$\Delta \ln L_j = -\frac{1}{\gamma^A}\left(\Delta \ln\left(\frac{x_j}{P_j}\right) - (1-\alpha_C)\,\Delta \ln\left(\frac{R_j}{P_j}\right)\right) + \Delta\xi_j \tag{B.1}$$

where $\Delta\xi_j$ includes changes in aggregate labor supply and exogenous components of amenities, A_j. In turn, under multiple worker types, the labor supply equation at the city level results from aggregating the supply of multiple workers:

$$\Delta \ln L^\theta_j = -\sum_\theta \frac{s^{L,\theta}_j}{\gamma^A_{\theta,\theta}}\left(\Delta \ln\left(\frac{x^\theta_j}{P_j}\right) - (1-\alpha_C)\,\Delta \ln\left(\frac{R_j}{P_j}\right)\right) - \sum_\theta s^{L,\theta}_j \sum_{\theta'\neq\theta} \frac{\gamma^A_{\theta',\theta}}{\gamma^A_{\theta,\theta}}\Delta \ln L^{\theta'}_j + \Delta\xi^\theta_j \tag{B.2}$$

where $\Delta\xi^\theta_j$ includes changes in the labor supply of type-θ workers and in the exogenous component of amenities, A^θ_j. We can draw an equivalence between the aggregate elasticity that would be estimated assuming homogeneous workers (i.e., using (B.1)) when the true model includes heterogeneous workers, so that the data is generated by (B.2). In the latter, assuming a shock that exogenously changes population and expenditure per capita in the same proportion for every worker, aggregating the labor supplies by skill we obtain:

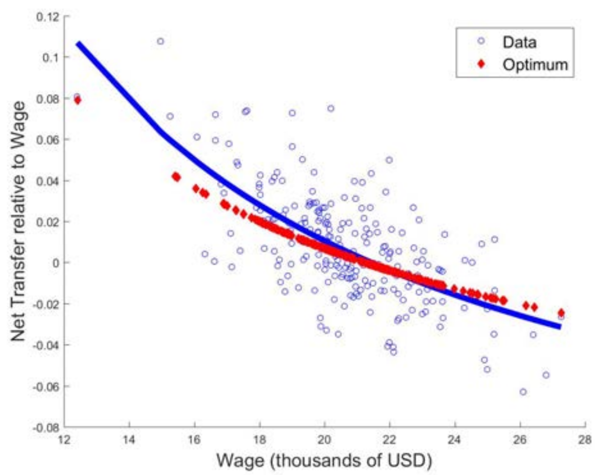
$$\hat{L_j} = \left(- \frac{\sum_\theta \frac{s_j^{L,\theta}}{\gamma_{\theta,\theta}^A}}{1 + \sum_\theta \sum_{\theta' \neq \theta} \frac{s_j^{L,\theta} \gamma_{\theta',\theta}^A}{\gamma_{\theta,\theta}^A}} \right) \left(\hat{x_j} - \hat{P_j} - (1 - \alpha_C) \left(\hat{R_j} - \hat{P_j} \right) \right) + \Delta \overline{\xi_j}, \tag{B.3}$$

where $s_j^{L,\theta}$ is the share of type θ workers in j and $\Delta\overline{\xi_j} \equiv \sum_\theta s_j^{L,\theta} \Delta \xi_j^\theta$. Comparing (B.1) with (B.3), we obtain that, at the average share of type-θ workers in the economy $\overline{s^{L,\theta}} = \frac{1}{J} \sum_j s_j^{L,\theta}$, the coefficient that would be recovered is:

$$\gamma^A = \frac{1 + \sum_\theta \sum_{\theta' \neq \theta} \frac{s_j^{L,\theta} \gamma_{\theta',\theta}^A}{\gamma_{\theta,\theta}^A}}{\sum_\theta \frac{s_j^{L,\theta}}{\gamma_{\theta,\theta}^A}}. \tag{B.4}$$

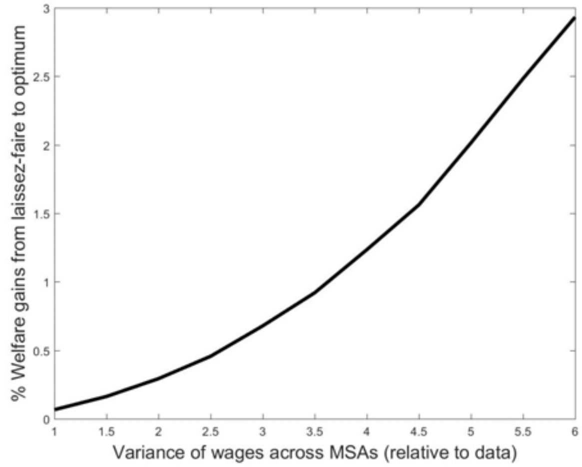
When implementing the model with a single worker type we use this expression to determine γ^A. This procedure delivers an aggregate amenity elasticity of $\gamma^A = -0.19$.

Figure B.1: Optimal Transfers and Reallocation under Homogeneous Workers



Note: This figure shows the transfer per worker relative to the wage in the optimal allocation and in the data. As implied by Section 3.3, the optimal net transfer relative to the wage takes the form $\frac{t_j}{w_j} = s + \frac{T}{w_j}$ for $s = \frac{\gamma^P + \gamma^A}{1 - \gamma^A}$. The solid lines shows the relationship $\frac{t_j}{w_j} = a + b \frac{1}{w_j}$ under parameters a and b that correspond to the best fit in an OLS regression.

Figure B.2: Gains from Optimal Policies given Different Initial Equilibria under Homogeneous Workers



Note: We simulate *laissez-faire* equilibria with no government transfers under different fundamentals such that the joint distribution of wages and city sizes differs from the data in terms of the variance of the wage distribution across MSAs and the correlation between wages and city sizes across MSAs. In all the equilibria the distribution of city sizes has the same variance as in the data. Correlation and variances are reported in relative terms compared to the data. For each variance-correlation combination we draw 400 random distributions of wages and city sizes, and report the mean welfare gains from implementing optimal policies across these simulations.

C.2 Land Regulations

Model The system changes as a function of the distortion in the initial equilibrium,τ_j^H and its change in a counterfactual $\hat{\tau_j^H}$. Equation A.66 becomes

$$\frac{N_j^H}{N_j}\frac{\left(\hat{X}_j\right)^{1-\tau_j^H\hat{\tau_j^H}}}{((1-\alpha_C)X_j)^{\tau_j^H\left(\hat{\tau_j^H}-1\right)}}+\left(1-\frac{N_j^H}{N_j}\right)\hat{W}_j\hat{N_j^Y}=\hat{W}_j\hat{N}_j \text{ for all } i.$$

Equations (A.68) and (A.69) become

$$\hat{E_j^Y}=\left(\frac{\alpha_C+\frac{b^I_{H,j}}{1+d_{H,j}}(1-\alpha_C)\left((1-\alpha_C)X_j\hat{X}_j\right)^{-\tau_j^H\hat{\tau_j^H}}}{\alpha_C+\frac{b^I_{H,j}}{1+d_{H,j}}(1-\alpha_C)((1-\alpha_C)X_j)^{-\tau_j^H}}\right)\left(1-\tilde{b^I_Y}\right)\hat{X}_j+\tilde{b^I_Y}\left(\hat{p_j}\hat{Y_j}\right) \tag{B.5}$$

and

$$\tilde{b^I_{Y,j}}=\frac{b^I_Y}{\left(\alpha_C+\frac{b^I_{H,j}}{1+d_{H,j}}(1-\alpha_C)((1-\alpha_C)X_j)^{-\tau_j^H}\right)\frac{X_j}{p_jY_j}+b^I_Y}. \tag{B.6}$$

Finally, (A.75) becomes:

$$\hat{R}_i=\left(\hat{p_i}^{\frac{1-b^I_{H,i}}{1-b^I_{Y,i}}}\hat{P}_i^{b^I_{H,i}-b^I_{Y,i}\frac{1-b^I_{H,i}}{1-b^I_{Y,i}}}\hat{X}_i^{d_{H,i}}\right)^{\frac{1}{1+d_{H,i}}}. \tag{B.7}$$

Calibration Diamond (2016) decomposes the housing supply elasticity between a part driven by geography γ_j^{geo} and a part driven by regulation γ_j^{reg} for each city j. The mapping to our model is: $\gamma_j^{geo}+\gamma_j^{reg}=\frac{d_{H,j}+\tau_j^H}{1-\tau_j^H}$, so that we set:

$$\tau_j^H = \frac{\gamma_j^{reg}}{1+\gamma_j^{reg}},$$
$$d_{H,j} = \gamma_j^{geo}\left(1-\tau_j^H\right).$$

The tax rate on sales $R_j H_h$ paid by non tradable producers is $1 - \frac{1}{1-\tau_{H,j}}\left(R_j H_j\right)^{-\tau_{H,j}}$. To calibrate scale of the tax, we normalize the scale of $R_j H_h$ so that the tax share of housing expenditures equals 10%. We have checked that results are fairly insensitive to the specific value of this re-scaling. We assume revenues from the tax on housing are rebated to firms. This assumption implies that the tax rate only distorts housing supply without distorting any additional margin. The rest of the model is calibrated following the same steps as in the benchmark except for a few steps. Specifically, we must recompute the total profits made by firms Π^H:

$$\Pi^H = \sum_j \left(1-\alpha_C\right) X_j \left[1 - \frac{1}{\left(d_j^H+1\right)\left(\left(1-\alpha_C\right) X_j\right)^{\tau_j^H}}\right],$$

where

$$X_j = \sum w_j^\theta L_j^\theta + \Pi^H + \sum \left(\tau_j^\theta - T_j^\theta\right) L_j^\theta.$$

The values of X_j and Π^H are calibrated so that these equations hold. In addition, the calibration of the non traded shares is amended to:

$$1-\eta_{H,I}^i = \frac{1+d_{H,j}}{1-\alpha_C}\left(\frac{W_j N_j^{NT}}{X_j}\right)\left(\left(1-\alpha_C\right) X_j\right)^{\tau_j^H}.$$

The rest of the calibration is unaffected.

C.3 Production with 3 Skill Groups

Model We continue to assume the same structure for the spillovers as in our benchmark case, on the basis of $U = \{L, M\}$ and $S = \{H\}$ types, so that (40) and (41) now become:

$$z_j^\theta = Z_j^\theta \left(L_j^U + L_j^M\right)^{\gamma_{U,\theta}^P} \left(L_j^H\right)^{\gamma_{S,\theta}^P}, \tag{B.8}$$

$$a_j^\theta \equiv A_j^\theta \left(L_j^U + L_j^M\right)^{\gamma_{U,\theta}^A} \left(L_j^H\right)^{\gamma_{S,\theta}^A}, \tag{B.9}$$

where we have noted, for $j = P, A$, $\gamma_{U,\theta}^j = \gamma_{U,U}^j$ and $\gamma_{S,\theta}^j = \gamma_{S,U}^j$ for $\theta = \{L, M\}$, $\gamma_{U,\theta}^j = \gamma_{U,S}^j$ and $\gamma_{S,\theta}^j = \gamma_{S,S}^j$ for $\theta = \{H\}$. Following similar steps as in the benchmark model, the total number of efficiency units (A.71) becomes

$$\hat{N}_i = \frac{\sum_{\theta\in\{U,H\}}\left(w_i^\theta L_i^\theta\right)/\delta}{\sum_{\theta\in\{U,H\}}\left(w_i^\theta L_i^\theta\right)/\delta + w_i^M L_i^M}\left(N_i^{\hat{U}H}\right)^\delta + \frac{w_i^M L_i^M}{\sum_{\theta\in\{U,H\}}\left(w_i^\theta L_i^\theta\right)/\delta + w_i^M L_i^M} z_i^{\hat{M}} L_i^{\hat{M}}, \tag{B.10}$$

where $N_i^{\hat{U}H}$ is the change in the efficiency units supplied by low and high skill workers:

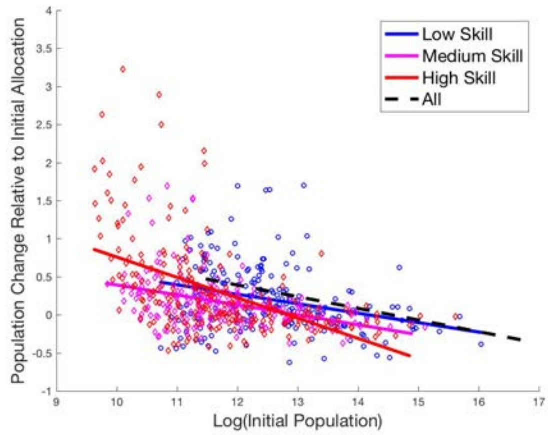
$$N_i^{\hat{U}H} = \left[\frac{w_i^U L_i^U}{\sum_{\theta'\in\{U,H\}} w_i^{\theta'} L_i^{\theta'}}\left(z_i^{\hat{U}} L_i^{\hat{U}}\right)^\rho + \frac{w_i^M L_i^M}{\sum_{\theta'\in\{U,H\}} w_i^{\theta'} L_i^{\theta'}}\left(z_i^{\hat{M}} L_i^{\hat{M}}\right)^\rho\right]^{\frac{1}{\rho}}. \tag{B.11}$$

In turn, the spillover functions (A.72) and (A.74) become:

$$\hat{z_i^\theta} = \left(\hat{L_j^U}\right)^{\gamma_{U,\theta}^P}\left(\frac{L_j^M}{L_j^S}\hat{L_j^M} + \frac{L_j^H}{L_j^S}\hat{L_j^M}\right)^{\gamma_{S,\theta}^P}, \tag{B.12}$$

$$\hat{a_i^\theta} = \left(\hat{L_j^U}\right)^{\gamma_{U,\theta}^A}\left(\frac{L_j^M}{L_j^S}\hat{L_j^M} + \frac{L_j^H}{L_j^S}\hat{L_j^M}\right)^{\gamma_{S,\theta}^A}. \tag{B.13}$$

Figure B.3: Change in Population by Skill Group (3 skills)



Calibration To calibrate this version of the model, we extend our dataset to 3 skill groups. Using the same procedure as described in the main text, we build a Census/BEA dataset for three skill group. We define L as low-skill workers, with no college education; M as medium-skill workers, with some college education; and H and high-skill workers, with 4 years of college or more. To calibrate the production function parameter, we follow Eeckhout et al. (2014). We use the same value of ρ as in our benchmark calibration ($\rho = 0.392$) and back out δ using the same formula as in Eeckhout et al. (2014), which gives $\delta = 1.124$.[55] The rest of the calibration is unchanged.

C.4 Imperfect Mobility

Model In this case, the type $\theta = (s, o)$ indexes both skill and origin. City amenity and productivity are now not only skill- but also origin-specific:

$$z_j^{s,o} = Z_j^{s,o} \prod_{s' \in \{U,S\}} \left(\sum_{o \in \mathcal{O}} L_j^{s,o} \right)^{\gamma^P_{s',s}} \tag{B.14}$$

$$a_j^{s,o} = A_j^{s,o} \prod_{s' \in \{U,S\}} \left(\sum_{o \in \mathcal{O}} L_j^{s,o} \right)^{\gamma^A_{s',s}} \tag{B.15}$$

In production, we further assume that workers from the same origin are perfect substitutes in production. Specifically, rather than (39) we now impose

$$N_j = \left[\sum_{s \in \{U,S\}} \left(\sum_{o \in \mathcal{O}} z_j^{s,o} L_j^{s,o} \right)^{\rho} \right]^{\frac{1}{\rho}}.$$

Following similar steps as in the benchmark model, the total number of efficiency units (A.71) becomes

$$\hat{N}_i = \left[\sum_{s \in \{U,S\}} \left(\frac{\sum_{o \in O} w_i^{s,o} L_i^{s,o}}{W_i N_i} \right) \left(\hat{N}_i^s \right)^{\rho} \right]^{\frac{1}{\rho}} \tag{B.16}$$

[55] See Eeckhout et al. (2014), section VIII. Quantifying the Production Technology. Given a value for ρ (noted γ in Eeckhout et al. (2014)), equation A27 of their Appendix A gives the expression for λ,as a function of $\rho$and of summary statistics from the data on wages and population by skill group.

where the change in the efficiency units supplied by workers with skill s is

$$\hat{N}_i^s = \hat{z}_i^s \sum_{o \in O} \left(\frac{w_i^{s,o} L_i^{s,o}}{\sum_{o'} w_i^{s,o'} L_i^{s,o'}} \right) L_i^{\hat{s},o}. \tag{B.17}$$

The spillover functions (A.72) and (A.74) take the same form as before, where now the change in the number of workers in skill group s is:

$$\hat{L_j^s} = \sum_{o \in \mathcal{O}} \left(\frac{L_j^{s,o}}{L_j^s} \right) L_j^{\hat{s},o}. \tag{B.18}$$

Finally, (A.73) becomes:

$$u^{\hat{s},o} = \left(L_j^{\hat{s},o} \right)^{-\sigma_s} \hat{a_j^s} \frac{x_j^{\hat{s},o}}{\hat{P_j}^{\alpha_C} \hat{R_j}^{1-\alpha_C}}. \tag{B.19}$$

Calibration The ACS reports the state of birth. To limit computational burden, we use as origin the region of birth corresponding to one of five Census regions (NW,SW,NE,SE and foreign-born). For each MSA, we compute the share of workers born in each of these 5 regions, and the corresponding share of total wages. We then split the total population and wage bill for each skill group and MSA (as calibrated in the benchmark) into these 5 regions of origins using these shares. We assume that total disposable income for each skill and MSA, as calibrated in the benchmark exercise, is split into recipients from these 5 regions according to their share of the wage bill. To calibrate the Frechet parameter that governs idiosyncratic preferences for location we use a value of $\sigma = 1/3$, which corresponds to a median value across existing estimates reported in Fajgelbaum et al. (2018). The rest of the calibration is unchanged.

C.5 Other specifications

Expenditure vs wage The calibration that ignores the transfers in the data and sets worker expenditures equal to income simply sets $x_j^\theta = w_j^\theta$ and $t_j^\theta = 0$.

Local ownership of fixed factors Under the assumption that land ownership is local, we construct expenditure by group and by MSA, x_i^θ in the model, similarly to Equation A.81, except that now profits are city-specific:

$$x_i^\theta = w_i^\theta - \tau_i^\theta + T_i^\theta + b^\theta \Pi_i^H. \tag{B.20}$$

The local returns to land Π_i^H that are consistent with the general equilibrium of the model are:

$$\Pi_j^H = \left(\frac{\gamma_i^H}{\gamma_i^H + 1} \right) (1 - \alpha_C) X_i,$$

where $X_i = \sum_\theta x_i^\theta$ is total final expenditure in the city. We combine these expressions, to calibrate $X_i = \frac{\sum_\theta \left(w_i^\theta - \tau_i^\theta + T_i^\theta \right) L_i^\theta}{1 - \left(\frac{\gamma_i^H}{\gamma_i^H + 1} \right) (1-\alpha_C)}$, where $\left\{ w_i^\theta, \tau_i^\theta, T_i^\theta, L_i^\theta \right\}$ are taken from the data. The rest of the procedure is unchanged.

Assuming away trade costs Absent trade costs, the price of tradables is the same in all cities. All destination cities buy the same share of output coming from various origin cities. In particular, the share of location i's imports originating from j is proportional to total output of j Y_j, so that:

$$s_{ji}^M = \frac{Y_j}{\sum_k Y_k}, \tag{B.21}$$

The export shares are then constructed using $s_{ji}^X \equiv \left(\frac{E_i}{p_j Y_j} \right) s_{ji}^M$, where spending E_i and traded income $p_j Y_j$.

Complementarity vs spillovers In the two skill calibration, we also explore results for alternative values for the complementarity between H and L. The weaker the complementarity parameter, the stronger the calibrated values of the cross-productivity spillovers. Table B.1 shows the welfare gains corresponding to different values of ρ, recalibrating the productivity spillovers each time. The first row is the baseline. The second row takes a complementarity parameter twice as small as in the baseline. The third row takes an elasticity of substitution twice as small as in the baseline. The last row take a very low value for the complementarity parameter, proxying for the limit case $\rho = -\infty$. The stronger the productivity spillovers, the less congestion there is to correct for in the economy. As a result, welfare gains decrease when productivity spillovers get stronger.

Table B.1: Complementarity in Production and Spillovers

Specification	Elasticity of substitution	Welfare Gain (%)
$\rho = 0.392$	1.65	4.0
$\rho = 0.392/2$	1.25	3.9
$\rho = -0.216$	1.65/2	3.7
$\rho = -10$	0.09	2.4