OPTIMAL SPATIAL POLICIES, GEOGRAPHY, AND SORTING*

PABLO D. FAJGELBAUM AND CECILE GAUBERT

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. The optimal allocation features a greater share of high-skill workers in smaller cities relative to the observed allocation. Inefficient sorting may lead to substantial welfare costs. *JEL* Codes: H21, H23, R12.

I. 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 they 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 affected 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 article, 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?

*We thank the editor, Pol Antràs, and five anonymous referees. 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 Quarterly Journal of Economics (2020), 959–1036. doi:10.1093/qje/qjaa001.
Advance Access publication on January 6, 2020.

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) and 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 effect 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 a greater share of high-skill workers in these locations.

The framework incorporates many key determinants of the spatial distribution of economic activity. Firms produce differentiated tradeable commodities and nontradeables 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 may 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. A key assumption of our analysis is that these effects are not internalized by the firms making hiring decisions. 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

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 on 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 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 nonspatial 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 because of the compensating differentials born of geographic forces (congestion in housing, amenities, trade costs, and nontraded 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.

areas (e.g., New Markets Tax Credit or Enterprise Zones) or direct public investment in specific areas (e.g., the Tennessee Valley Authority). Other policies are not explicitly spatial but end up redistributing income to specific places (e.g., nominal income taxes and credits, state and local tax deductions, sectoral subsidies). Neumark and Simpson (2015) review empirical literature on place-based policies and conclude that the evidence on their success at creating local jobs in the United States is mixed depending on the specific policy and area being treated. Although some local enterprise programs have been found to be unsuccessful at attracting local jobs, larger programs such as federal empowerment zones in high-poverty rate areas of the United States or the Tennessee Valley Authority have been found to have positive effects (Busso, Gregory, and Kline 2013; Kline and Moretti 2014a).

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 incentivize workers to move, leading to gains from reallocation. Under these assumptions, we derive the labor income subsidies that 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 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. For example, if low-skill workers benefit in terms of productivity from high-skill workers, the decentralized pattern

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

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 the optimal allocation of a national planner who can implement transfers across cities, in an environment with several dimensions of spatial heterogeneity and different sources of spillovers across heterogeneous workers.[5] A key feature of our approach is to provide a simple characterization of efficiency in terms of the expenditure distribution. Being only a function of observable variables and elasticities, this condition allows us to characterize optimal policies despite the generality of the underlying framework and to determine the set of statistics in the data that suffice 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, Redding, and Rossi-Hansberg (2018). In the latter, individuals decide 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.

4. Flatters, Henderson, and Mieszkowski (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 (2012), Albouy et al. (2019), and Eeckhout and Guner (2015). 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.

5. We only inspect policies set by a national government. 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.

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 on urban premia: larger cities in the United States feature higher wages, higher share of skilled workers, and higher skill premium, as documented by Behrens and Robert-Nicoud (2015), among others. Moretti (2012) points out a "great divergence" in these outcomes over the past decades. We ask whether, in the presence of spillovers, these observed patterns of spatial disparities in the United States are too strong from the perspective of spatial efficiency.[6]

In our benchmark analysis, we allow for two skills groups, high-skill (college) and low-skill (noncollege) workers. We combine data on labor and nonlabor income, taxes, and transfers at the city level from the Bureau of Economic Analysis, with Census data that allows us to break down these MSA-level totals by skill group within cities. To parameterize the spillover elasticities, we rely on existing estimates in the United States 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. In our benchmark parameterization, these gains are attained through a reallocation of 11% of the population. 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 toward low-wage cities. The optimal transfers can be implemented via higher labor income taxes in high-wage cities. In

6. Recent papers, such as Eeckhout, Pinheiro, and Schmidheiny (2014), Behrens, Duranton, and Robert-Nicoud (2014), and Davis and Dingel (2012), include spatial sorting of heterogeneous individuals to rationalize some of these patterns.

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.

The effect of these transfers is a reallocation of workers from currently large high-wage cities to small low-wage cities. In terms of skill mix, the initially less skill-intensive cities grow and see an increase in the share of high-skill workers. The largest and the most skill-intensive cities shrink, but they also increase their skill share. The resulting optimal allocation features a greater share of high-skill workers in small cities compared with the observed allocation as well as lower wage inequality in large cities, to the point that the urban skill premium (i.e., the higher return to high-skill labor in larger cities) vanishes. In sum, in the optimal allocation, the patterns of urban premia described before are all weakened: larger cities feature relatively lower wages, lower share of skilled workers, and lower skill premium compared to the observed allocation.

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.[7]

The rest of the article is structured as follows. Section II presents a stylized model to drive intuition, then presents the general environment. Section III characterizes the optimal policies, teases out their implications in specific cases of the theory

7. In our parameterization, 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 noncollege workers.

corresponding to the models from the literature, and determines the data that suffice to implement the model. Section IV describes the data and the calibration. Section V presents the quantitative implementation, and Section VI concludes. Proofs, additional derivations, and data construction are detailed in the Appendix.

II. Economic Geography Model with Worker Sorting and Spillovers

II.A. 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 place-based policy literature, such as Glaeser and Gottlieb (2008) and Kline and Moretti (2014a), is to characterize efficiency assuming that $c_j = z_j$; that is, 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.[8] 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

8. If the planner maximizes $u \equiv \sum_j L_j v_j(L_j)$, the marginal return to adding a worker in j is $(1 + \gamma^A + \gamma^P)\, v_j$. Using a different notation, Proposition 1 of Glaeser and Gottlieb (2008) solves this planner problem, which leads to equalization of marginal returns and therefore of v_j. Kline and Moretti (2014a) make the similar point that if dL workers are reallocated from i to j, there are no gains from reallocation starting from any market allocation with free mobility.

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.[9]

This analysis is made under a strong restriction in the planning problem, namely that each 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.[10]

We now assume that the government can implement spatial transfers. With transfers, the distribution of consumption per capita c_j changes and workers move to equalize utility in space. As shown in Appendix Section A, starting from transfers $t_j \equiv c_j - z_j$ received by workers in j, when a transfer is implemented the common level of utility across workers changes according to:

$$\frac{du}{u} = \frac{\gamma^P \sum_j z_j dL_j + \gamma^A \sum_j c_j dL_j - \sum_j t_j dL_j}{Y}, \tag{1}$$

where dx is the infinitesimal change in x and Y is aggregate output. The no-transfers equilibrium implies $t_j = 0$. Combined with the definition of output ($Y_j = z_j L_j$), this leads to:

$$\frac{du}{u} = \left(\gamma^P + \gamma^A\right) \sum_j \left(\frac{Y_j}{Y}\right) \frac{dL_j}{L_j}. \tag{2}$$

9. For instance, Duranton and Venables (2018, footnote 10 p.13) 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."

10. Intuitively, the no-transfer market allocation equates amenities times consumption per capita $a_j c_j$ across locations, where consumption equals output, $c_j = z_j$. With constant elasticities and no transfers, the planner equates $(1 + \gamma_A + \gamma_P) a_j c_j$ across locations, which gives the same result. But starting from this allocation, c_j may be reallocated to locations with high amenity value. So there are incentives to transfer output, which in turn leads to reallocation of workers.

Therefore, a transfer leading to a reallocation of dL workers from j to location i yields

$$\frac{du}{u} = \left(\gamma^P + \gamma^A\right)(z_i - z_j)\frac{dL}{Y}. \tag{3}$$

From equation (3), 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 in compensating differentials through amenities, a_j. In a more general model where the compensating differentials arise through costly trade or nontraded goods, the allocation is inefficient even with no dispersion in amenities. What matters is that amenities, nontraded goods, or trade frictions lead to compensating differentials between cities.[11] This result holds 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 set out to characterize first-best spatial policies in the context of a more general spatial equilibrium model.

II.B. 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. Among other things, 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

11. Our analysis assumes that the planner values the utility of ex ante identical workers in the same way, regardless of where they live. The no-transfer allocation could be efficient if the planner had different Pareto weights for identical workers who live in different locations, for a particular distribution of those weights.

clearing is:

$$\sum_j L_j^\theta = L^\theta, \tag{4}$$

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

$$u_j^\theta = a_j^\theta \left(L_j^1, .., L_j^\Theta\right) U\left(c_j^\theta, h_j^\theta\right). \tag{5}$$

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 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 nontradeable services including housing (h_j^θ). The utility function $U(c, h)$ is homogeneous of degree 1.

Every location produces traded and nontraded 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 nontraded 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 nontraded sector in j is:

$$H_j\left(N_j^H, I_j^H\right) = \sum_\theta L_j^\theta h_j^\theta. \tag{6}$$

Goods in the traded sector can be shipped domestically or to other locations. The country's geography is captured by iceberg trade frictions $d_{ji} \geqslant 1$. 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:

$$Y_j\left(N_j^Y, I_j^Y\right) = \sum_i d_{ji} Q_{ji}. \tag{7}$$

Traded goods may be differentiated by origin, reflecting either industrial specialization at the regional level or variety specialization at the plant level.[12] 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:

$$Q(Q_{1i}, .., Q_{Ji}) = \sum_\theta L_i^\theta c_i^\theta + I_i^Y + I_i^H. \tag{8}$$

The standard assumptions in the Rosen (1979)-Roback (1982) models is that products are perfect substitutes, which implies $Q(Q_{1i}, .., Q_{Ji}) = \sum_j Q_{ji}$. 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

$$z_j^\theta = z_j^\theta\left(L_j^1, .., L_j^\Theta\right) \tag{9}$$

efficiency units. The function z_j^θ captures exogenous differences in productivity between locations and skill groups and 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

12. We abstract from modeling multiple traded sectors with input-output linkages across them. Rossi-Hansberg, Sarte, and Schwartzman (2019) study optimal spatial policies in a framework with these features.

and nontraded sectors, the services $z_j^\theta L_j^\theta$ of the various types of labor are combined through the possibly nonhomothetic aggregator $N(z_j^1 L_j^1, .., z_j^\Theta L_j^\Theta)$. This aggregator also captures imperfect substitution across workers. Feasibility in the use of labor services then implies

$$N_j^Y + N_j^H = N\left(z_j^1 L_j^1, .., z_j^\Theta L_j^\Theta\right). \tag{10}$$

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.[13] Second, the framework features two potential sources of nonconvexities through the amenity and productivity spillover functions. The utility of each agent may change 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 $z_j^\theta(L_j^1, .., L_j^\Theta)L_j^\theta$.

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

$$\gamma_{\theta,\theta'}^{P,j} \equiv \frac{L_j^\theta}{z_j^{\theta'}} \frac{\partial z_j^{\theta'}}{\partial L_j^\theta} \quad \text{and} \quad \gamma_{\theta,\theta'}^{A,j} \equiv \frac{L_j^\theta}{a_j^{\theta'}} \frac{\partial a_j^{\theta'}}{\partial L_j^\theta}. \tag{11}$$

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 these elasticities can be variable.

II.C. 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

13. Allowing for commuting (as in Section III.E) makes the production and consumption locations distinct. However, they are still nonseparable, as long as commuting costs are nonzero, because the choice of workplace depends on the residential choice through commuting access.

impact of their choices on the spillover functions $a_j^\theta(L_j^1, .., L_j^\Theta)$ and $z_j^\theta(L_j^1, .., L_j^\Theta)$.

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

$$\max_{c_j^\theta, h_j^\theta} U\left(c_j^\theta, h_j^\theta\right) \quad \text{s.t.} \quad P_j c_j^\theta + R_j h_j^\theta = x_j^\theta, \tag{12}$$

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

$$u_j^\theta = a_j^\theta\left(L_j^1, .., L_j^\Theta\right) \frac{x_j^\theta}{\psi\left(P_j, R_j\right)}, \tag{13}$$

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

$$u_j^\theta \leqslant u^\theta, \tag{14}$$

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

2. Firms. Producers of traded and nontraded commodities maximize profits:

$$\Pi_j^Y = \max_{N_j^Y, I_j^Y} p_j Y_j\left(N_j^Y, I_j^Y\right) - W_j N_j^Y - P_j I_j^Y, \tag{15}$$

$$\Pi_j^H = \max_{N_j^H, I_j^H} R_j H_j\left(N_j^H, I_j^H\right) - W_j N_j^H - P_j I_j^H, \tag{16}$$

where p_j is the domestic price of the tradeable commodity produced in j and W_j is the wage per efficiency unit of labor. Workers collectively own a national portfolio of these returns, which amounts to $\Pi = \sum_j \Pi_j^Y + \Pi_j^H$.

Given a distribution of wages per worker $\{w_j^\theta\}$, the wage of type-θ workers in location j equals the value of its marginal product taking as given the efficiency distribution $\{z_j^\theta\}$:

$$w_j^\theta = W_j \frac{\partial N\left(z_j^1 L_j^1, .., z_j^\Theta L_j^\Theta\right)}{\partial L_j^\theta}. \tag{17}$$

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:

$$P_i \frac{\partial Q(Q_{1i}, .., Q_{Ji})}{\partial Q_{ji}} = d_{ji} p_j, \tag{18}$$

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.

3. 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. Workers of different types may differ in their ownership of fixed factors, but they 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

$$x_j^\theta = w_j^\theta + b^\theta \Pi + t_j^\theta, \tag{19}$$

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

$$\sum_j \sum_\theta L_j^\theta x_j^\theta = \sum_j \sum_\theta L_j^\theta w_j^\theta + \Pi. \tag{20}$$

DEFINITION 1. A competitive allocation consists of quantities c_j^θ, h_j^θ, L_j^θ, 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, returns to fixed factors Π, 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 equation (12) for expenditures x_i^θ satisfying equation (19), and employment L_j^θ is consistent with the spatial mobility constraint (14); (ii) the labor, intermediate input choices $N_j^Y, I_j^Y, N_j^H, I_j^H$ and profits Π are such that producers maximize profits, labor demand is given by equation (17), and trade flows Q_{ji} are given by equation (18); (iii) the government budget constraint is satisfied; that is, equation (20) holds, and (iv) all markets clear, that is, equations (4) to (10) hold.

II.D. Planning Problem

Our aim is to contrast this decentralized allocation with the solution to the planner's problem. We consider a planning problem where the planner chooses the distribution of workers over locations and types $\{L_j^\theta\}$, consumption of traded and nontraded commodities $\{c_j^\theta, h_j^\theta\}$, trade flows $\{Q_{ij}\}$, and the allocation of efficiency units and intermediate inputs, $\{N_j^Y, I_j^Y, N_j^H, I_j^H\}$. The planner implements policies that treat all individuals within a type in the same way, and is bound by the spatial mobility constraint (14). Along with that constraint, the market-clearing conditions (4) to (10) define a set $\mathcal{U}$ of attainable utility levels. The optimal planning problem is

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

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, because 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 effect of their hiring decisions on efficiency through z_j^θ. We turn to the solution and implementation of this planning problem.

III. Optimal Transfers

Before characterizing the optimal allocation in a general setup, we build intuition by augmenting our simple example from Section II.A with heterogeneous workers, which helps illustrate the additional role played by inefficient sorting.

III.A. Simple Example with Heterogeneous Workers

We return to the simplified setup of Section II.A, now augmented with several worker types.[14] 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 Section B, the utility of this group changes according to:

$$\frac{du^{\theta_0}}{u^{\theta_0}} = \frac{1}{Y^{\theta_0}} \sum_j \sum_{\theta \in \Theta} \left(\sum_{\theta' \in \Theta} \left(\gamma^P_{\theta,\theta'} + \gamma^A_{\theta,\theta'} \right) w^{\theta'}_j \frac{L^{\theta'}_j}{L^{\theta}_j} \right) dL^{\theta}_j, \tag{21}$$

where dL^{θ}_j is the population change triggered by the transfers, $w^{\theta'}_j$ is the wage of type-θ workers 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 equation (2) in the one-group case. Hence, as in the case with homogeneous workers from Section II.A, 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 $\frac{L^{\theta'}_j}{L^{\theta}_j}$ 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 discussed in Section II.A. Therefore,

14. 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 equation (11) are constant, $\gamma^{P,j}_{\theta,\theta'} = \gamma^P_{\theta,\theta'}$ and $\gamma^{A,j}_{\theta,\theta'} = \gamma^A_{\theta,\theta'}$.

inefficient sorting creates an additional rationale for gains from spatial transfers.

III.B. Efficiency Condition and Optimal Transfers

To characterize efficiency in the general model, it is useful to note that the competitive allocation can be determined given an arbitrarily chosen expenditure distribution $\{x_j^\theta\}$ over types and locations. We can then choose the transfers t_j^θ to implement the arbitrarily chosen x_j^θ using equation (19). Therefore, we can obtain a condition over the expenditure distribution x_j^θ that must hold in any efficient allocation, regardless of what particular policy tools are used to achieve it. Comparing an allocation with expenditures x_j^θ to the outcomes of the planning problem, detailed in Definition 2 of Appendix Section C, we obtain the following result.

PROPOSITION 1. If a competitive equilibrium is efficient, then

$$\underbrace{W_j \frac{dN_j}{dL_j^\theta}}_{\substack{\text{marginal product of labor} \\ \text{(private+spillovers)}}} + \underbrace{\sum_{\theta'} \frac{L_j^{\theta'}}{L_j^\theta} \gamma_{\theta,\theta'}^{A,j} x_j^{\theta'}}_{\substack{\text{marginal amenities} \\ \text{(spillovers)}}} = \underbrace{x_j^\theta}_{\substack{\text{consumption cost} \\ \text{(private)}}} + \underbrace{E^\theta}_{\substack{\text{opportunity cost,} \\ \text{of type } \theta}} \tag{22}$$

if $L_j^\theta > 0$, for all j and θ and some constants $\{E^\theta\}$. If the planner's problem is globally concave and condition (22) holds for some specific $\{E^\theta\}$, then the competitive equilibrium is efficient.

Condition (22) 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. Using the labor demand condition (17), we obtain that the value of the marginal product of labor can be written as a function of wages, employment, and elasticities:

$$W_j \frac{dN_j}{dL_j^\theta} = w_j^\theta \left(1 + \gamma_{\theta,\theta}^{P,j}\right) + \sum_{\theta' \neq \theta} \frac{L_j^{\theta'}}{L_j^\theta} \gamma_{\theta,\theta'}^{P,j} w_j^{\theta'}. \tag{23}$$

The second term in equation (22) is the marginal benefit (or cost 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 nonseparability 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 nontraded commodities for other workers in that location. The last term, E^θ, is the multiplier of the national market-clearing constraint (4) 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 equation (22), 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.

Second, Proposition 1 extends a familiar efficiency condition from the misallocation literature to spatial environments. In our economy, "space" enters through trade costs, nontraded 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 equation (22) 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 nonseparability of consumption and production means that the net marginal benefit of reallocating 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.[15]

Finally, we note that equation (22) 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 equation (22) can be ruled out if the planner's problem is globally concave, as in that case there is only one allocation that satisfies the first-order conditions of the planner. In Section III.F we introduce conditions for global concavity of the planner's problem.[16]

Given the efficiency conditions (22), we now derive transfers that implement them. Combining equation (19) and the definitions of the spillover elasticities equation (11) with Proposition 1 and labor demand equation (17), we obtain the following proposition:

PROPOSITION 2. The optimal allocation can be implemented by the transfers

$$t_j^{\theta *} = \sum_{\theta'} \left(\gamma_{\theta,\theta'}^{P,j} w_j^{\theta' *} + \gamma_{\theta,\theta'}^{A,j} x_j^{\theta' *} \right) \frac{L_j^{\theta' *}}{L_j^{\theta *}} - \left(b^\theta \Pi^* + E^\theta \right) \quad \text{if } L_j^{\theta *} > 0, \tag{24}$$

15. As was noted early on in studies of optimal city size, assumptions about ownership of fixed factors are relevant to determining the efficiency of the market allocation (Pines and Sadka 1986). The expenditure distribution implied by equation (22) that implements the efficient allocation is invariant to assumptions about ownership of fixed factors. A different rule to distribute Π from that assumed in equation (19) would imply a different set of optimal transfers t_j^θ to implement the optimal expenditure distribution, but would not affect equation (22).

16. 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, if a solution to the planner's problem exists, then there is a market allocation consistent with equation (22).

where the terms $(x_j^{\theta *}, w_j^{\theta *}, L_j^{\theta *}, \Pi^*)$ are the outcomes at the efficient allocation, and $\{E^\theta\}$ 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 and distributional concerns.[17] In the absence of spillovers, we would still have $t_j^{\theta *} = -(b^\theta \Pi^* + E^\theta)$, 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 equation (24).

We will use conditions (19) and (24) for two separate quantitative goals in Section V. 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 conditions to recover the spillover elasticities $\{\gamma_{\theta,\theta'}^{P,j}, \gamma_{\theta,\theta'}^{A,j}\}$ from the observed data.

III.C. Optimal Subsidies with Constant-Elasticity Spillovers

The optimal subsidies formula takes a simple form when spillovers have constant elasticities. We make this assumption from now on and write: $\gamma_{\theta,\theta'}^{P,j} = \gamma_{\theta,\theta'}^{P}$ and $\gamma_{\theta',\theta}^{A,j} = \gamma_{\theta',\theta}^{A}$. The optimal transfers in equation (24) then simplify to $t_j^\theta = s_j^\theta w_j^\theta - T^\theta$, where

$$(25)\qquad s_j^\theta = \frac{\gamma_{\theta,\theta}^P + \gamma_{\theta,\theta}^A}{1-\gamma_{\theta,\theta}^A} + \sum_{\theta' \neq \theta} \frac{\gamma_{\theta,\theta'}^P w_j^{\theta'} + \gamma_{\theta,\theta'}^A x_j^{\theta'}}{1-\gamma_{\theta,\theta}^A} \frac{L_j^{\theta'}}{w_j^\theta L_j^\theta}$$

and

$$(26)\qquad T^\theta = b^\theta \Pi + \frac{E^\theta}{1-\gamma_{\theta,\theta}^A}.$$

This representation readily implies that the optimal transfers can be implemented by labor income subsidies s_j^θ coupled with

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

lump-sum tax T^{θ}. The labor income subsidy s_j^{θ} is a function of wages, expenditures, and population. The labor subsidies tackle spatial inefficiencies due to spillovers, and the lump-sum transfers take care of distributional concerns. Differences in the holdings of the national portfolio across types affect the level of lump-sum transfers only. They do not create a rationale for spatially differentiated policies. We now draw the implications of this formula in special cases.

1. 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 equation (25) becomes:

$$s^{\theta} = \frac{\gamma^{P}_{\theta,\theta} + \gamma^{A}_{\theta,\theta}}{1 - \gamma^{A}_{\theta,\theta}}. \tag{27}$$

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_j^{θ} 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, even with constant elasticity spillovers, that the laissez-faire allocation is generically inefficient. We made this point in Section II.A 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 (s^{θ}, T^{θ}) 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 (14), indifference across populated locations j and j' implies:

$$\frac{\psi\left(P_{j'}, R_{j'}\right)}{\psi\left(P_{j}, R_{j}\right)} \frac{a_j^{\theta}\left(L_j\right)}{a_{j'}^{\theta}\left(L_{j'}\right)} = \frac{\left(1+s^{\theta}\right) W_{j'} z_{j'}^{\theta}\left(L_{j'}\right) + T^{\theta} + b^{\theta}\Pi}{\left(1+s^{\theta}\right) W_{j} z_{j}^{\theta}\left(L_{j}\right) + T^{\theta} + b^{\theta}\Pi}. \tag{28}$$

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, nontraded 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 $W_j z_j(L_j) = W_{j'} z_{j'}(L_{j'})$ for any (s^{θ}, T^{θ}), and these policies would cease to affect the spatial allocation.

2. 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 what the optimal subsidies look like, consider a polar case without amenity spillovers and without efficiency spillovers 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

$$s_j^{\theta} = \gamma_{\theta,\theta'}^{P} \left(\frac{w_j^{\theta'} L_j^{\theta'}}{w_j^{\theta} L_j^{\theta}} \right). \tag{29}$$

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 (29) 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.

3. Link to Henry George Theorem. We discussed an implementation of the optimal transfers equation (24) with labor income subsidies equation (25) and lump-sum taxes equation (26). 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.[18] This question is only meaningful when the optimal labor income subsidies are positive, because otherwise the tax system necessarily entails taxing mobile factors. Then, under some regularity conditions, our model implies that the returns to the fixed factors Π add up to more than the total lump-sum taxes in equation (26).[19] In this case, the tax system implementing optimal subsidies may feature aggregate redistribution from fixed factors to mobile factors.

III.D. Economic Geography Frameworks

The environment laid out in Section II nests standard economic geography models, such as Helpman (1998), Allen and Arkolakis (2014), and Redding (2016).[20] 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,

18. 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.

19. Using equation (26) 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 $\Pi > \sum_\theta L^\theta T^\theta$.

20. Our presentation so far has assumed that each location sells a different product under perfect competition. In Online Appendix A 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 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 III.D 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).

their normative implications have barely been explored.[21] 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

$$u_j = A_j L_j^{1+\gamma^A} c_j^{\alpha_C} h_j^{1-\alpha_C}. \tag{30}$$

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

$$Y_j = Z_j L_j^{1+\gamma^P}. \tag{31}$$

Supply of nontraded 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

$$x_j = (1-\eta)w_j + \eta\bar{w}, \tag{32}$$

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: $t_j = \eta(\bar{w} - w_j)$. 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.

21. 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 interdependencies. Other recent papers studying spatial policies in geography models include Allen, Arkolakis, and Li (2015), who consider zoning restrictions within a city; Fajgelbaum and Schaal (2017), who consider transport network investment; and Gaubert (2018), who characterizes the optimal allocation in a model with heterogeneous firms and a complementarity between city size and firm productivity.

Should the optimal policy that implements equation (32) redistribute toward or away from high-wage locations? The answer depends on the distribution of nonlabor 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

$$s = \frac{1+\gamma^P}{1-\gamma^A}[1-(1-\alpha_C)\omega]-1, \tag{33}$$

with lump-sum transfer equal to $T = -s\bar{w}$. 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 nonzero 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 affect 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 toward or away from high-wage regions.

III.E. Additional Forces

Our results on optimal transfers can be extended to economic geography environments that incorporate additional margins. We review some of these extensions that correspond to popular modeling choices in the literature.

1. 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, $\Pr(\epsilon_j^l < x) = e^{-x^{-\frac{1}{\sigma_\theta}}}$. 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 (14), which is replaced with the following labor supply equation:

$$\frac{L_j^\theta}{L^\theta} = \left(\frac{u_j^\theta}{u^\theta}\right)^{\frac{1}{\sigma_\theta}}. \tag{34}$$

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 Section II.D. 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. As a result, the planner will have incentives to redistribute to locations where individuals have a higher marginal utility of consumption of tradeables, driven by their preference draw. Because on average individuals have higher draws conditional on having sorted into lower-wage locations, the planner has incentives to redistribute towards those locations.

2. Commuting. We apply the analysis to a framework with commuting in the style of Ahlfeldt et al. (2015) and Monte, Redding, and Rossi-Hansberg (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 consisting of 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 $z_i(L_i^W)$, and the common component of utility equation (5) is $u_{ji} = a_j(L_j^R)U_{ji}(c_{ji}, h_{ji})$, where the function U_{ji} may vary by ji to capture disutility from commuting travel time. We also allow for an idiosyncratic worker-level shock ϵ^l_{ji} according to a Fréchet distribution, $\Pr(\epsilon^l_{ji} < x) = e^{-x^{-\frac{1}{\sigma}}}$, so that the utility of a commuter l from j to i is $u_{ji}\epsilon^l_{ji}$. The resulting flow of commuters from j to i is $L_{ji} = L(\frac{u_{ji}}{u})^{\frac{1}{\sigma}}$. In the market allocation, each 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 Section E that the optimal transfers can be decomposed as the sum of two types of transfers. The first component depends on the workplace,

$$t^W_i = \frac{\gamma^P_i - \sigma}{1+\sigma} w^*_i, \tag{35}$$

and the second component depends on the residence,

$$t^R_j = \frac{\gamma^A_j}{1+\sigma} \sum_{i'} \frac{L^*_{ji'} x^*_{ji'}}{L^R_j}. \tag{36}$$

The optimal transfer is $t^*_{ji} = t^W_i + t^R_j - T$, where T is a lump-sum transfer that adjusts for government budget balance.[22] The workplace policy t^W_i is the Pigouvian tax fixing the inefficiency in production, whereas the residence policy t^R_j isolates the role of amenity spillovers. The two policies are additive. That is, even with commuting, the optimal transfer still varies by place rather than by bilateral commuting pattern.[23] Absent amenity spillovers ($\gamma^A = 0$), the workplace transfer t^W_i is the only one active and takes the same form as in the benchmark model without commuting.

3. 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 consider a special case of our model with homogeneous workers and constant-elasticity spillovers in amenities. However, we now extend our

22. These expressions assume that the returns to fixed factors Π are evenly distributed in the population.

23. This result abstracts from congestion in commuting, which would provide a rationale for imposing a tax based on commuting patterns.

model to allow the efficiency of location j to be an arbitrary function of the number of workers in every location: $z_j = z_j(\{L_{j'}\})$. This formulation accommodates a commonly used specification where spillovers decay with distance between spatial units.[24] We define the efficiency spillover elasticity across locations,

$$\gamma^{P,j,j'} = \frac{\partial z_{j'}}{\partial L_j}\frac{L_j}{z_{j'}}, \tag{37}$$

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:

$$t_j = \frac{\gamma^{P,j,j} + \gamma^A}{1-\gamma^A} w_j + \sum_{j' \neq j} \frac{\gamma^{P,j,j'}}{1-\gamma^A}\frac{L_{j'} w_{j'}}{L_j} + T. \tag{38}$$

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, nonlocalized 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.

III.F. Quantitative Implementation

Having established the theoretical characterization of an optimal allocation, we now lay out a methodology to bring it to the data. Doing so requires imposing functional-form assumptions, and identifying conditions under which the quantitative methodology is well behaved—that is, conditions under which optimal spatial policies lead to a unique equilibrium that can therefore be unambiguously recovered. Finally, we identify the data requirement of the procedure. We later implement this quantitative methodology.

24. This type of spillover has been used to study economic activity at different spatial scales. For instance, Ahlfeldt et al. (2015) assume $z_{j'} = (\sum_j L_j e^{-\delta t_{jj'}})^{\alpha}$ where $t_{jj'}$ is travel time between blocks j and j' within a city and δ is a decay parameter, while Desmet, Nagy, and Rossi-Hansberg (2018) study these spillovers at a broader scale.

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

$$U(c,h) = c^{\alpha_C} h^{1-\alpha_C}, \tag{39}$$

while the aggregator of traded commodities is CES,

$$Q(Q_{1i},..,Q_{Ji}) = \left(\sum_j Q_{ji}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}, \tag{40}$$

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

$$Y_j\left(N_j^Y, I_j^Y\right) = z_j^Y \left(N_j^Y\right)^{1-b_{Y,j}^I} \left(I_j^Y\right)^{b_{Y,j}^I}, \tag{41}$$

$$H_j\left(N_j^H, I_j^H\right) = z_j^H \left(\left(N_j^H\right)^{1-b_{H,j}^I} \left(I_j^H\right)^{b_{H,j}^I}\right)^{\frac{1}{1+d_{H,j}}}, \tag{42}$$

where $d_{H,j} \geqslant 0$ and $\{z_j^Y, z_j^H\}$ are total factor productivity (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,

$$N_j = \sum_{i=1}^{I} \left[\sum_{\theta\in\Theta_i} \left(z_j^\theta L_j^\theta\right)^{\rho_i}\right]^{\frac{1}{\rho_i}}, \tag{43}$$

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:

$$z_j^\theta\left(L_j^1,..,L_j^\Theta\right) = Z_j^\theta \prod_{\theta'} \left(L_j^{\theta'}\right)^{\gamma_{\theta',\theta}^P}, \tag{44}$$

$$a_j^\theta\left(L_j^1,..,L_j^\Theta\right) \equiv A_j^\theta \prod_{\theta'} \left(L_j^{\theta'}\right)^{\gamma_{\theta',\theta}^A}. \tag{45}$$

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 $\{Z_j^\theta, A_j^\theta\}$ 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.

2. 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 (equations (39) to (45)) we have the following property.

PROPOSITION 3. The planning problem is concave if

$$\Gamma^A > \Gamma^P, \tag{46}$$

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

$$1 - \gamma^A > \left(1 + \gamma^P\right)\left(\frac{1-\alpha_C}{1+D} + \alpha_C\right). \tag{47}$$

Condition (46) 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. These conditions are sufficient but not necessary for uniqueness, as the planner's problem can be concave outside of these strong parameter restrictions. In the case of a single type, condition (46) simplifies to $\gamma^P + \gamma^A < 0$; further assuming Cobb-Douglas preferences over traded and nontraded goods we obtain a weaker restriction that allows for spillovers to be net agglomerative (equation (47)).[25]

25. The CES restriction (40) on the aggregator of trade flows $Q(\cdot)$ is not needed for these results. Therefore, these conditions hold regardless of product

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, Arkolakis, and Takahashi 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 suboptimal policies or no policy intervention, but this poses no limitation for our approach.

3. Implementation in Changes and Data Requirements. To bring the model to the data, we take the following steps. First, we assume that the observed data allocation is consistent with our model. That is, it is generated by a decentralized equilibrium consistent with Definition 1, subject to the functional form assumptions (39) to (45). Second, we solve for the planner problem described in Section II.D. We show in that section that in the spirit of the exact-hat algebra method developed by Dekle, Eaton, and Kortum (2008), this problem in levels is equivalent to a problem where the endogenous variables are expressed relative to their initial value. Letting $\hat{x} = \frac{x'}{x}$, where x is the value of a variable in the observed equilibrium and x' is the value in an alternative equilibrium, we solve for the changes in the endogenous variables $\{\hat{x_j^\theta}, \hat{P}_i, \hat{p}_i, \hat{Y}_i, \hat{W}_i, \hat{N}_j, \hat{L_j^\theta}, \hat{R}_i, \hat{u^\theta}\}$ 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 relative to the initial equilibrium. The following proposition summarizes our approach and the corresponding data requirements.

PROPOSITION 4. Assume that the observed data is generated by a competitive equilibrium consistent with Definition 1 under the functional forms equations (39) to (45). Then, relative to

differentiation across locations. Numerical simulations confirm the intuition that the amount of product differentiation between regions governed by the aggregator $Q(\cdot)$ helps make the planner's problem concave.

the initial equilibrium, the optimal allocation can be fully characterized as a 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 $\{\alpha_C, \sigma, \rho, b^I_{Y,j}, b^I_{H,j}, d_{H,j}\}$; and
iv. the spillover elasticities $\{\gamma^A_{\theta',\theta}, \gamma^P_{\theta',\theta}\}$.

This exact-hat algebra approach is convenient to take the model to the data because it sidesteps the estimation of many parameters (the city-type shifters of amenities $\{Z^\theta_j, A^\theta_j\}$, TFP shifters $\{z^Y_j, z^H_j\}$, and bilateral trade costs $\{d_{ij}\}$). These parameters turn out not to appear in the formulation of the model solution in changes relative to the observed equilibrium. It is important to point out that this approach is not without limitations. First, it assumes away measurement error. This means that the procedure implicitly calibrates a combination of the previous parameters to exactly match the data in points i and ii of Proposition 4 as an equilibrium outcome of the model from Definition 1. Second, these parameters are treated as exogenous fundamentals that are invariant between equilibria. Therefore, this approach ignores the possibility that some of these parameters could change in response to reallocation of workers.

Importantly, the quantitative implementation laid out in Proposition 4 does not impose restrictions on the distributional policies across locations in the observed equilibrium. The net transfers that generate the expenditure distribution x^θ_j 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 (22) 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.

IV. Data and Calibration

To take the model to the data, we use as an empirical setting the distribution of economic activity across MSAs in the United States in 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 (noncollege) workers. Because of data limitations, our analysis abstracts from more detailed definitions of skill types.[26]

IV.A. 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 data set for 2000 to 2007 reports total taxes paid by individuals and MSA (Dunbar 2009). Taken together, these sources give us a data set at the MSA level. We apportion 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 and 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 data set covers 209 MSAs for which we have both BEA and Census information.[27]

The model accommodates an arbitrary number of finely defined skill types θ. When going to the data, to implement the analysis we reduce the number of types to only two groups, defined by education. Having made this choice, an important concern when measuring these variables is that the model does not include heterogeneity across individuals within each group of skill θ, whereas

26. 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.

27. These areas correspond to 95% of the population and 96% of income of all U.S. metropolitan areas. Metropolitan areas in the United States in turn cover 78% of the population, and 83% of personal income.

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 sociodemographic 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 sociodemographic composition within groups across MSAs. We discuss the details of this step in Online Appendix B.

We use the variables above to construct expenditure per capita, x_i^θ, using its definition (equation 19) 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 nonlabor income is owned by high-skill workers according to the BEA/ACS data.[28]

As implied by point 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 United States 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 a 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

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

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 nontraded sectors by MSA.[29] We adjust this measure to remove variation from compositional effects following a similar approach to the one described above for income, expenditure, taxes, and transfers.

IV.B. 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 noncollege), and we rely on similar data sources for quantification. Therefore, her estimates constitute a natural benchmark to parameterize the model. In what follows, we discuss these elasticities and several alternative specifications that are also used in the quantitative section.

1. Utility and Production Function Parameters $\{\alpha_C, \sigma, \rho, b^I_{Y,j}, b^I_{H,j}, d_{H,j}\}$. 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 equation (42)) for each MSA, and of the elasticity of substitution between high and low skill, estimated at 1.6 and implying $\rho = 0.392$.[30]

We calibrate the Cobb-Douglas share of intermediates in traded-good production ($b^I_{Y,j} = 0.468$ for all j in equation (41)) 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 nontraded production in each city ($1 - b_{H,j}$ in equation (42)) can be chosen to match the share of workers in the nontraded sector of each MSA, as detailed in Online Appendix B.2. We assume an elasticity of substitution σ among traded varieties in

29. We define employment in the following NAICS sectors as corresponding to the nontraded sector in the model: retail, real estate, construction, education, health, entertainment, hotels, and restaurants.

30. For MSAs that we cannot match to Diamond (2016), we use the average housing supply elasticity across MSAs.

equation (40) equal to 5, corresponding to a central value of the estimates reported by Head and Mayer (2014).

2. Efficiency Spillovers $\{\gamma^{P}_{\theta',\theta}\}$. Previous empirical studies, such as Ciccone and Hall (1996), Combes, Duranton, and Gobillon (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).[31] Hence, we set 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 Online 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 (17) and the elasticity of substitution between skilled and unskilled workers ρ.

As a result we obtain $(\gamma^{P}_{UU}, \gamma^{P}_{SU}, \gamma^{P}_{US}, \gamma^{P}_{SS}) = (0.003,\ 0.044,\ 0.02,\ 0.053)$. 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 these 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 0 spillovers from low-skill workers.[32]

3. Amenity Spillovers $\{\gamma^{A}_{\theta',\theta}\}$. 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 noncollege workers. In addition, she estimates a positive elasticity for the supply

31. Most of the studies reviewed by Combes and Gobillon (2015) and Melo, Graham, and Noland (2009) also fall in this range.

32. 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, Sacerdote, and West 2013). Our city-level analysis abstracts from these considerations.

of this MSA-level amenity index with respect to the relative supply of college workers. As detailed in Online Appendix B.2, we can combine these estimates and map them to our amenity spillovers $\gamma^A_{\theta',\theta}$ using the labor supply equation implied by the spatial mobility constraint (14). As a result, we obtain $(\gamma^A_{UU}, \gamma^A_{SU}, \gamma^A_{US}, \gamma^A_{SS}) = (-0.43,\ 0.18,\ -1.24,\ 0.77)$. These parameters imply strong positive amenity spillovers generated by high-skill workers and negative spillovers generated by low-skill workers.[33]

4. Alternative Parameterizations of the Spillover Elasticities. We implement all our counterfactuals under different parameterizations 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 parameterization that scales down the amenity spillover elasticities $\gamma^A_{\theta,\theta'}$ by 50% (referred to as the "low amenity spillover" parameterization); (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's (2016) point estimates (referred to as the "high cross amenity spillover" and "low cross amenity spillover" parameterizations, respectively); (iii) a less conservative parameterization that scales up the efficiency spillover elasticities $\gamma^P_{\theta,\theta'}$ to 0.12, that is, twice the benchmark of 0.06 from Ciccone and Hall (1996) (referred to as the "high efficiency spillover" parameterization); (iv) a more conservative parameterization that scales down the efficiency spillover elasticities by a factor of 2 (referred to as the "low efficiency spillover" parameterization); and (v) parameterizations of efficiency spillovers that correspond to alternative values of the complementarity parameter ρ, as detailed in Online Appendix D.5. The values of these alternative parameterizations are reported in Online Appendix B.2.

33. At these values, all but one of the concavity conditions implied by Proposition 3 are satisfied. Specifically, 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 parameterization 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.

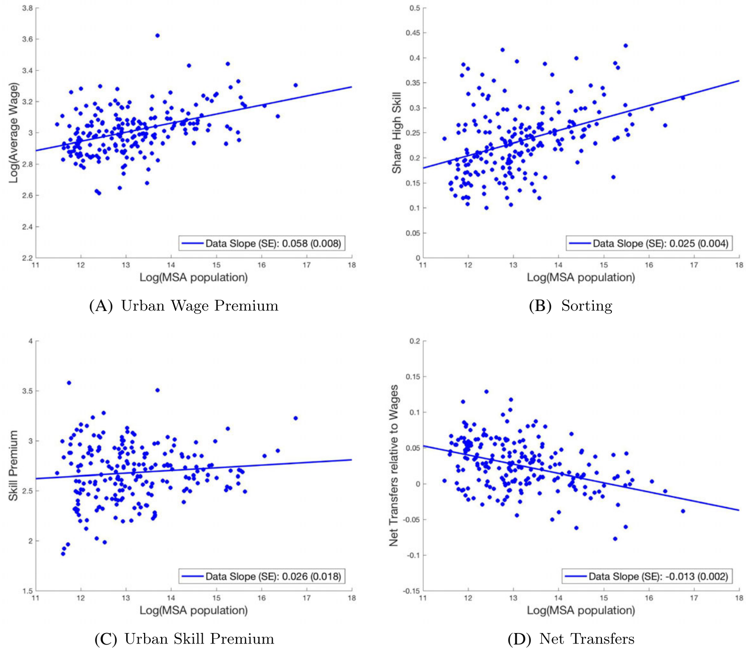


FIGURE I

Urban Premia

Each figure shows data across MSAs. All the city-level outcomes reported on the vertical axes of panels A to C are adjusted by sociodemographic characteristics of each city, as detailed in Online Appendix B.1.

IV.C. Stylized Facts

Figure I 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 a 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 5.8%.[34] Panel B shows spatial sorting,

34. This elasticity includes the composition effect due to a higher share of high-skill workers in larger cities. Controlling for composition, the elasticity is 3.2%.

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%. That is, doubling the 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, and 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. For each city we construct the net imbalance as the difference between expenditures and total income (from labor and nonlabor 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 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 already. Therefore, distributive government policies that vary with these characteristics across individuals do not underlie these patterns across cities.

V. Optimal Spatial Policies in the U.S. Economy

With these data in hand, we use the methodology laid out in Section III.F to solve numerically for optimal spatial allocations in the empirical context of the U.S. economy. We contrast these optimal allocations with the current spatial equilibrium of the United States and quantify the corresponding welfare gains.

V.A. *Optimal Transfers, Reallocations, and Welfare Gains*

To quantify an optimal allocation, we solve the planner's problem in changes relative to the observed equilibrium. 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 workers, $\hat{u^U}$. Varying this lower bound traces the Pareto frontier.

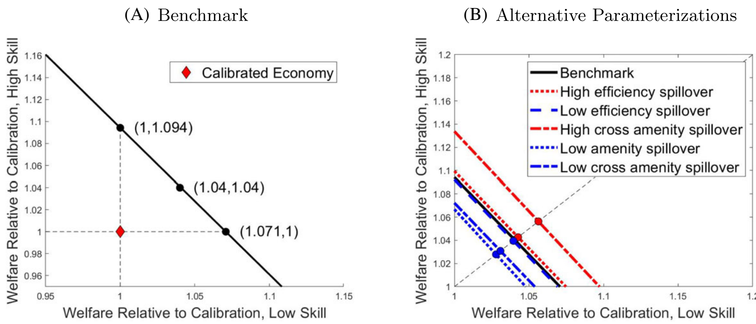


FIGURE II

Utility Frontier of the U.S. Economy between High- and Low-Skill Workers

The figure shows the optimal welfare changes $(\hat{u^L}, \hat{u^H})$ between the optimal and observed allocation, corresponding to the solution of the planner's problem in relative changes described in Appendix Section G. Each point corresponds to a maximization of $\hat{u^H}$ subject to a different lower bound on $\hat{u^L}$. The benchmark parameterization in the left panel corresponds to the black benchmark line in the right panel.

1. Aggregate Welfare Gains. The left panel of Figure II shows the utility frontier of the U.S. economy in the benchmark parameterization, expressed in changes relative to the observed equilibrium. The point (1,1) represented with a red diamond (color version online) 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 II shows the utility frontier for the benchmark in a solid line and for each of the alternative parameterizations discussed in Section IV.B. The frontier shifts up and down with little change in slope. The welfare gains from implementing optimal policies are larger in the two darkest frontiers (online, in red), corresponding to high efficiency and amenity spillovers. The gains are lower with low amenity spillovers. Table I 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

TABLE I

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)	Lower production elasticity	2.4–3.9

Notes. The table reports the common welfare gains for skilled and unskilled workers under alternative parameterizations described in Section IV.B. 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 plus or minus one standard deviation of the estimates in Diamond (2016). See Online Appendix B.2 for details about these parameterizations. Row (7) corresponds to efficiency spillovers calibrated using different values of the production function parameter ρ, as detailed in Table A.3 in Online Appendix D.5.

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%.[35] 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 United States.[36]

35. Figure A.3 in Online Appendix D.1 shows that, assuming homogeneous workers, the observed transfers across MSAs in the optimal allocation are quite close to the data. Figure A.4 shows that the welfare gains can be substantial under counterfactual data with high wage dispersion. Section D.1 in the Online Appendix describes the details of the calibration with homogeneous workers.

36. 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 0.6% from harmonizing state taxes. The small welfare gains to optimal

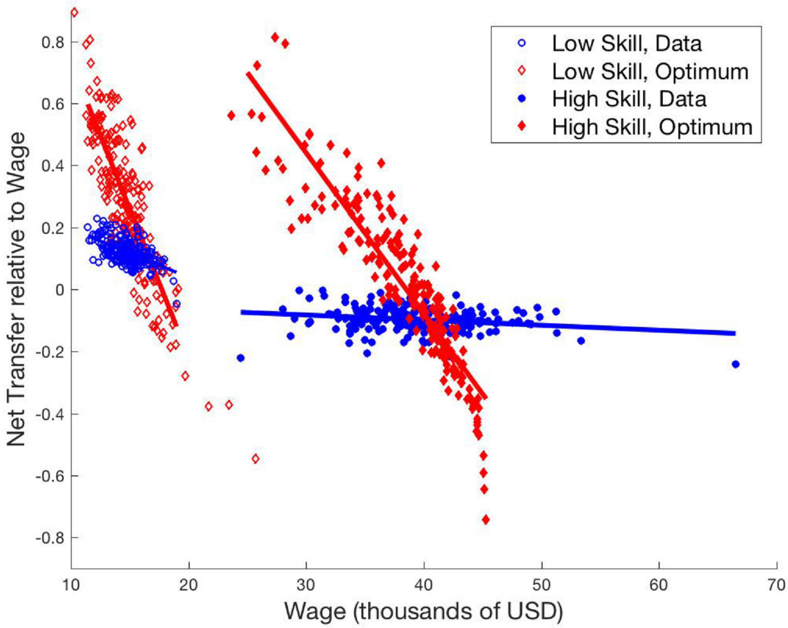


FIGURE III

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

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 Online Appendix B. The slopes of each linear fit (with std. err.) 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 II.

2. Actual versus Optimal Transfers. How does the optimal spatial income redistribution compare to the data? Let t_j^θ be the optimal transfers received by type θ according to equation (24) in Proposition 2. Figure III shows the net transfers per capita relative to wages $\frac{t_j^\theta}{w_j^\theta}$ by MSA and worker type on the vertical axis, against the wage w_j^θ of each MSA in both the data (black circles; blue online) and the optimal allocation (gray diamonds; red online), for low-skill workers (open markers) and high-skill workers

reallocation without worker heterogeneity are in line with results in Eeckhout and Guner (2015) and Ossa (2018).

(solid markers). We represent the optimal allocation corresponding to the point on the Pareto frontier in the left panel of Figure II where welfare gains are equal for both types of workers.[37]

The transfers in the data present a clear pattern of redistribution from high-skill workers and high-wage cities toward low-skill workers and low-wage cities. Net average transfers are positive for low-skill workers and negative for high-skill workers in most MSAs. Within skill groups, net transfers decrease with the wage of the MSA. On average across MSAs, they equal \$1.8 thousand for low-skill workers, or 12% of their average wage. For high-skill workers, the corresponding numbers are −\$3.8 thousand or −10% of the average wage. In cities where high-skill workers earn on average more than \$50k a year, net transfers of high-skill workers are −\$8.9 thousand or −15% of wages. 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 toward low-wage cities than what is observed empirically.[38]

To understand what drives these optimal transfers, we return to the expression for optimal subsidies, equation (25). The first term of equation (25) is driven by own spillovers, and the second term is shaped by cross-spillovers. In our parameterization of spillovers for low-skill workers, both 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 parameterization, high-skill workers generate positive own spillovers. According to the first term in equation (25), 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 mixing of high-skill workers with low-skill workers. A higher tax

37. 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.

38. Figure A.1 in Online Appendix C plots the optimal transfer scheme against labor income. It shows that income alone is an imperfect predictor of the optimal tax, suggesting that second-best policies based on income alone could not perfectly replicate it. Characterizing second-best policies in our framework is an interesting avenue left for future research.

in high-wage cities directs skilled workers into small, low-wage cities that are relatively abundant in low-skill workers.

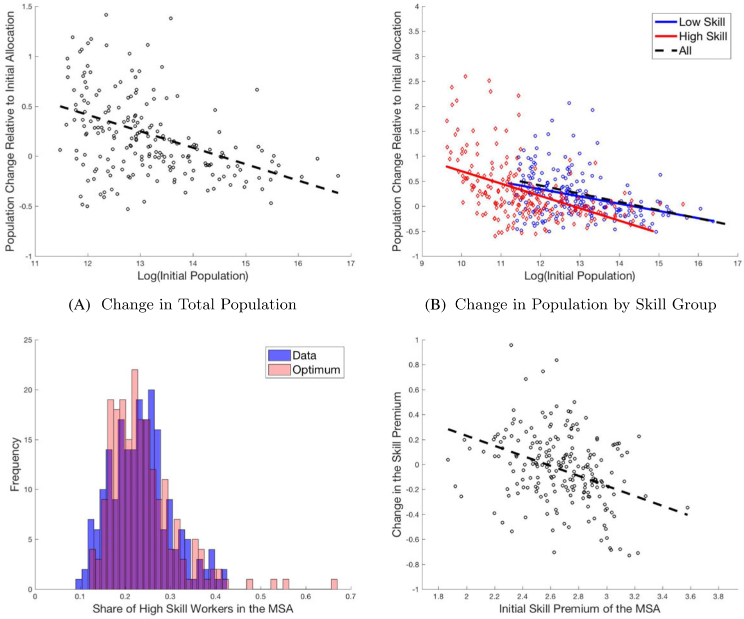
Although both low- and high-skill workers are on average reallocated toward lower-wage cities, it is a priori ambiguous for which group this effect is stronger. We examine the question of optimal sorting below.

3. Optimal Reallocation and Sorting. The optimal transfers change the spatial distribution of economic activity compared with the data. By changing the location incentives of workers, they affect spatial sorting and the city size distribution. These reallocations in turn affect 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 IV 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 mid-sized and small MSAs that shrink alongside large MSAs 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%).[39]

Even though the tax changes are large, only 11% of the population is reallocated to reach the optimum. When moving to the optimal allocation, a regression of population changes on the change in the net-of-tax rate (i.e., 1 minus the tax rate) across locations yields an elasticity of 1.2.[40]

39. Albouy et al. (2019) and Eeckhout and Guner (2015) argue that large cities are too small in models with homogeneous workers, one-dimensional heterogeneity, and spillover elasticities only.

40. This general-equilibrium elasticity of population to taxes implied by the model falls within the [0, 2] range corresponding to the quasi-experimental estimates of migration responses to taxes summarized by Kleven et al. (2019). This literature estimates an elasticity of migration to taxes that does not account for general-equilibrium outcomes. Our quantification relies in part on the labor supply elasticity estimated by Diamond (2016), who estimates an elasticity of migration to wage changes (rather than taxes) of approximately 2 and 4 for college and noncollege workers, respectively.



(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

FIGURE IV

Changes in Population, Skill Shares, and Skill Premium across MSAs

Panel A shows the change in population between the optimal allocation and the initially observed equilibrium and the linear fit. Slope (std. err.): −0.16 (0.03); R^2=0.15. Panel B displays the same outcomes for high- and low-skill workers. Slopes (with std. err.): High Skill: −0.25 (0.03); Low Skill: −0.15 (0.03). Panel C displays the histogram of high-skill shares across MSAs in the data and in the optimum allocation. Panel D displays on the vertical axis the difference in the skill premium between the optimal and initial allocation. Slope (std. err.): −0.4 (0.07). The figures correspond to planner's weights such that both types of workers experience the same welfare gain in Figure II.

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, reallocation toward 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 in cities with initially very low skill share and in some large cities with very high skill share.[41]

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.

4. The Urban Premia in the Optimal Allocation. Changes in the spatial allocation can be conveniently summarized by coming back to the urban premia from Figure I and computing them in the optimal allocation. We contrast them in Figure V: each pair of linked observations corresponds to the same MSA in the data (black; blue online) and in the optimal allocation (gray; red online).[42] 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.

The optimal allocation features a higher share of high-skill workers in smaller cities (Panel B). At the same time, the figure shows that the initially largest MSAs shrink and become more skill-intensive. Specifically, 8 of the 10 initially largest cities increase their skill share.[43] The urban skill premium vanishes (Panel C), implying that the sorting pattern from Panel B ends

41. This pattern is illustrated in Figure A.2 in Online Appendix C. Weighting by initial population MSA, the relationship between initial skill share and optimal growth in the skill share is U-shaped.

42. 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 parameterizations of the spillovers from Table I.

43. 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.

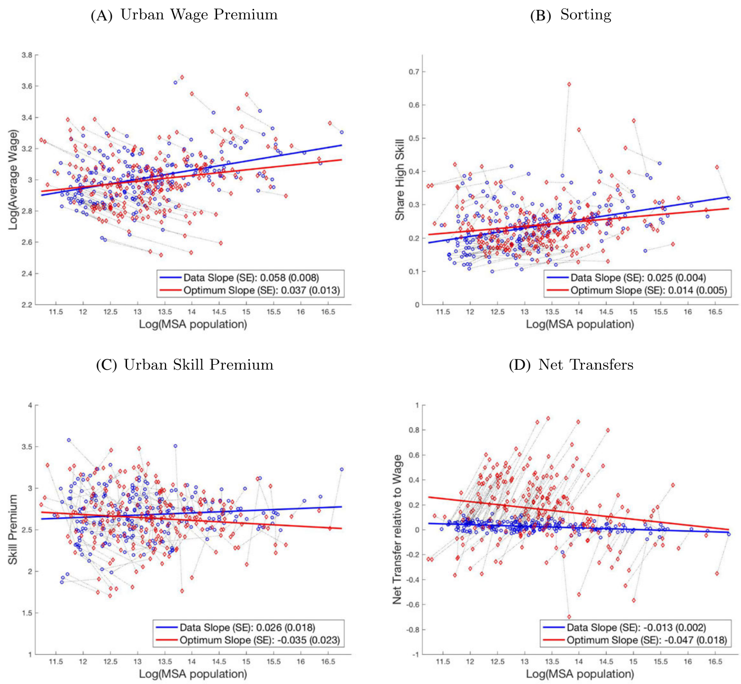


FIGURE V

Urban Premia: Data and Optimal Allocation

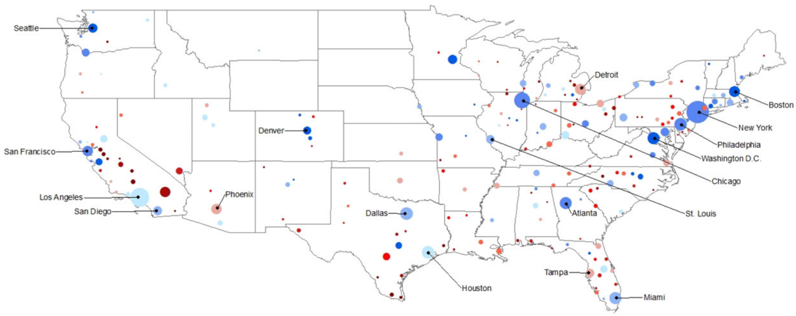
Each panel reports outcomes across MSAs in the data and in the optimal allocation. Each linked pair of observations corresponds to the same MSA.

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.

In sum, in the optimal allocation the urban premia are weakened: larger cities feature relatively lower average wages, share of skilled workers, and skill premium compared with the data.

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

(A) Population



(B) Skill Share

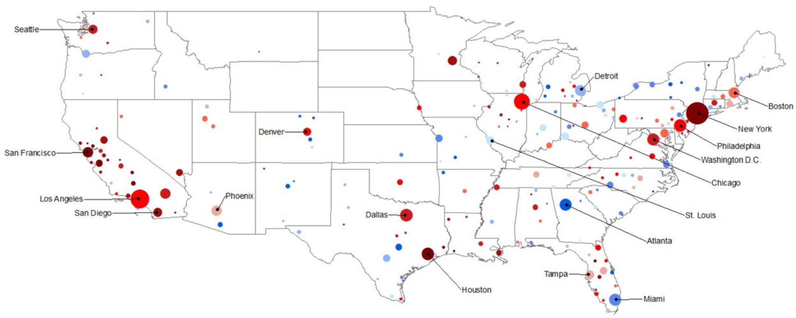


FIGURE VI

Optimal Population Reallocation and Change in Skill Share

The maps show the growth in population (top) and share of college workers (bottom) from the observed to the optimal allocation. Cities are weighted by initial population. Dark gray (red online) means positive growth and light gray (blue online) is negative growth.

growth. 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 lose population while smaller cities inland next to them grow. In terms of the skill shares, the five largest MSAs (New York, Los Angeles, Chicago, Dallas, and Philadelphia) and some other large MSAs (such as Washington, Boston, and San Francisco) become more skill intensive despite losing population. In these MSAs, the skill premium falls, reflecting the higher preferences of high-skill workers for those locations. A few large MSAs (such as Miami, Atlanta, and Detroit) shrink both in terms of overall population and the skill share. Many small cities grow in their skill share, ultimately driving down the urban skill share in Figure V, Panel B.

V.B. 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 to 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 equation (24) must hold. Combined with the definition of expenditure per worker in equation (19), we obtain the following optimal relationship between transfers, wages, expenditures, and employment:

$$t_j^\theta = a_0^\theta + a_1^\theta w_j^\theta + a_2^\theta \left(\frac{w_j^{\theta'\neq\theta} L_j^{\theta'\neq\theta}}{L_j^\theta}\right) + a_3^\theta \left(\frac{x_j^{\theta'\neq\theta} L_j^{\theta'\neq\theta}}{L_j^\theta}\right) + \varepsilon_j^\theta, \tag{48}$$

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 $\{a_i^\theta\}$ by running equation (48) as a regression in the cross-section, and then infer the spillover elasticities $\{\gamma_{\theta,\theta'}^A, \gamma_{\theta,\theta'}^P\}$ up to a normalization for each type.[44] We normalize the own-spillover elasticity for

44. This normalization is needed because from equation (48) the own-spillover elasticities for productivity and amenities are not separately identified. Assuming

productivity to the benchmark level for the United States used in Section IV.B.

This exercise yields $(\gamma^A_{UU}, \gamma^A_{SU}, \gamma^A_{US}, \gamma^A_{SS}) = (-0.09, -0.16, 0.06, -0.32)$ and $(\gamma^P_{UU}, \gamma^P_{SU}, \gamma^P_{US}, \gamma^P_{SS}) = (0.003, 0.20, -0.08, 0.053)$.[45] The average level of both types of spillovers is similar to the parameters implied by the empirical estimates used in the calibration. In 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, heterogeneity in the sign of spillovers across groups plays an important role in shaping optimal policies. This result is consistent with our previous finding that heterogeneity in spillovers between groups matters, obtained from the contrast between the quantified model under homogeneous and heterogeneous workers.

V.C. Alternative Specifications

To gauge the sensitivity of our findings, we turn to implementing the calibration and counterfactuals for alternative specifications. Each case formally extends our benchmark quantification. We recalibrate the model each time, compute the welfare gain common to all workers on the utility frontier, and compare it with the benchmark case. We defer the details of the implementation to the Online Appendix.

1. Land Use Regulations. Several papers (Bunten 2017; Herkenhoff, Ohanian, and Prescott 2018; Parkhomenko 2018; Hsieh and Moretti 2019) argue that local land use regulations

values for $\gamma^P_{\theta,\theta}$ we can then infer the remaining elasticities as follows: $\gamma^A_{\theta,\theta} = \frac{a^\theta_1 - \gamma^P_{\theta,\theta}}{1+a^\theta_1}$, $\gamma^P_{\theta,\theta'} = a^\theta_2(1-\gamma^A_{\theta,\theta})$, and $\gamma^A_{\theta,\theta'} = a^\theta_3(1-\gamma^A_{\theta,\theta})$.

45. The regressions have an R^2 of 0.32 for high skill and of 0.15 for low skill. Therefore, the first-order conditions of the planner are not exactly satisfied in the data even after choosing the revealed-optimal elasticities that best fit equation (48). However, when we use these revealed-optimal elasticities to compute the efficient allocation relative to the observed allocation, we obtain negligible welfare gains of 0.07%. Hence, the procedure confirms that under the revealed optimal elasticities, the observed allocation is very close to optimal.

create spatial distortions by lowering the housing supply elasticity. In our benchmark procedure, we interpreted the housing supply elasticity as a technological restriction in the planner's problem. We now extend the model 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 nontraded goods in each city j:

$$1 - \frac{1}{1-\tau_{H,j}} \left(R_j H_j\right)^{-\tau_{H,j}}. \tag{49}$$

As a result, the housing supply elasticity becomes:

$$\frac{\partial \ln H_j}{\partial \ln R_j} = \frac{1-\tau_{H,j}}{d_{H,j}+\tau_{H,j}}. \tag{50}$$

This specification microfounds a housing supply elasticity that includes a technology constraint $d_{H,j}$ due to geographic characteristics as in Saiz (2010) and land regulations $\tau_{H,j}$ as in the previous papers. The higher the parameter $\tau_{H,j}$, the lower the housing supply elasticity compared with its undistorted level. Our benchmark parameterization 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 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 II. 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 with leaving local regulations unchanged. This result suggests that both margins (optimal

TABLE II

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 parameterizations. We report the common welfare gains to all workers on the utility frontier. See the Online Appendix for details.

redistribution and land use regulations) are roughly equally important sources of misallocation.[46]

2. Multiple Skills with Nonhomothetic 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 equation (43) 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:

$$N_j = \left(\left(z_j^L L_j^L\right)^\rho + \left(z_j^H L_j^H\right)^\rho\right)^\lambda + \left(z_j^M L_j^M\right)^\rho . \tag{51}$$

This production function follows Eeckhout, Pinheiro, and Schmidheiny (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 nonhomothetic between the medium-skill workers and the nest of low- and high-skill workers. Hence, as production increases, the relative demand for the

46. In terms of optimal city sizes, in the counterfactual that removes the wedges in addition to implementing optimal transfers we find that larger cities grow relative to small cities, reversing the pattern from Figure IV, Panel A. Therefore, the positive impact of removing wedges on the growth of the largest cities more than offsets the negative impact of the optimal transfers. In this case, the flattening of the urban wage premium and the pattern of sorting from Figure V, Panels A and B is even stronger because of an inflow of low-skill workers to large cities. This inflow in turn leads to lower wages for low-skill workers in large cities and to an increase in the urban skill premium relative to the data.

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 II, the welfare effects are very similar to the benchmark case, and Figure A.5 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. This result suggests that our conclusions are robust to refining the substitution patterns between skills in the production function. We note that, compared with the two-groups case, this extension has only changed the production function but not the spillovers structure. It would be interesting in future work to revisit our analysis in a context with richer spillovers across extreme skill groups.

3. 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 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 III.E, we incorporate preference draws within types according to a Fréchet distribution with parameter σ_θ.[47] Turning to the quantification,

47. 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

we classify workers as being born in one of five 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 II, 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 the benchmark analysis are robust to incorporating this form of mobility frictions.

4. Other Specifications. We have 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.[48] Second, we redo the quantification assuming that the returns to fixed factors are locally distributed to residents of each location.[49] Our theoretical discussion from Section III.D shows that this assumption entails an additional distortion. Consistent with this result, we find that the common welfare gains of implementing optimal expenditures increase 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 the welfare

a bilateral level. An alternative assumption leading to a similar property would have been to 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).

48. Because the transfers tend to be negative in larger cities, ignoring transfers leads to an underestimation of the amenity levels implied by the model in larger cities.

49. The weak correlation between capital income in the data and a proxy for housing profits across cities computed as $\frac{\gamma_j}{(\gamma_j+1)X_j}$, 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.

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.

VI. 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 apply the model to the distribution of economic activity across MSAs in the United States using existing estimates of the 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 a higher share of high-skill workers in these locations. It also calls for the currently largest MSAs to shrink and to become 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. Our analysis abstracted from various margins that could be important for future work. We implemented the analysis in a closed economy, but optimal spatial policies within a country could interact with international migration and trade. We only considered first-best policies set by a national planner and abstracted from second-best policies or from fiscal competition between local jurisdictions. Finally, we only considered a static model, where each worker type is fixed regardless of location. We leave it to future work to study dynamic and long-run implications of spatial policies when worker productivity or tastes can change over time through skill formation or as a function of the skill mix in the community.

Appendix: Proofs and Additional Derivations

A. Appendix to Section II.A

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

$$u = a_j\left(L_j\right)c_j, \tag{52}$$

$$\sum_j L_j c_j = \sum_j L_j z_j, \tag{53}$$

$$\sum_j L_j = L. \tag{54}$$

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 condition (53) we obtain the following expression for utility:

$$u = \frac{\sum_j L_j z_j\left(L_j\right)}{\sum_j \frac{L_j}{a_j(L_j)}}. \tag{55}$$

The planner maximizes this term subject to condition (54). 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 conditions (52) and (53) we obtain equation (1).

B. Appendix to Section III.A

We derive equation (21). The market allocation is the solution to the following conditions:

$$u^{\theta} = a_j^{\theta} c_j^{\theta}, \tag{56}$$

$$\sum_{\theta} \sum_{j} L_j^{\theta} \left(c_j^{\theta} - z_j^{\theta} \right) \leqslant 0, \tag{57}$$

$$\sum_{j} L_j^{\theta} = L^{\theta}. \tag{58}$$

Combining the first two conditions and following similar steps to Appendix Section A, utility of group θ_0 can be written:

$$u^{\theta_0} = \frac{\sum_{\theta} \sum_{j} L_j^{\theta} z_j^{\theta} - \sum_{\theta' \neq \theta_0} \underline{u}^{\theta'} \sum_j \frac{L_j^{\theta'}}{a_j^{\theta'}}}{\sum_j \frac{L_j^{\theta_0}}{a_j^{\theta_0}}}. \tag{59}$$

Taking a first-order approximation to this expression while keeping $\underline{u}^{\theta'}$ constant and using the mobility constraints (56) we obtain:

$$\frac{du^{\theta_0}}{u^{\theta_0}} = \frac{\sum_{\theta} \sum_{j} L_j^{\theta} z_j^{\theta} \left(\frac{dL_j^{\theta}}{L_j^{\theta}} + \sum_{\theta'} \gamma_{\theta',\theta}^{P} \frac{dL_j^{\theta'}}{L_j^{\theta'}} \right) - \sum_{\theta} \sum_{j} c_j^{\theta} L_j^{\theta} \left(\frac{dL_j^{\theta}}{L_j^{\theta}} - \sum_{\theta'} \gamma_{\theta',\theta}^{A} \frac{dL_j^{\theta'}}{L_j^{\theta'}} \right)}{\sum_j c_j^{\theta_0} L_j^{\theta_0}}, \tag{60}$$

which, after some manipulations, becomes:

$$\frac{du^{\theta_0}}{u^{\theta_0}} = \frac{\sum_{\theta} \sum_{j} \left[-t_j^{\theta} L_j^{\theta} + \sum_{\theta'} \left(\gamma_{\theta_0,\theta'}^{P} L_j^{\theta'} z_j^{\theta'} + \gamma_{\theta_0,\theta'}^{A} c_j^{\theta'} L_j^{\theta'} \right) \right] \frac{dL_j^{\theta}}{L_j^{\theta}}}{\sum_j c_j^{\theta_0} L_j^{\theta_0}}, \tag{61}$$

where $t_j^{\theta} \equiv c_j^{\theta} - z_j^{\theta}$ is the transfer to group θ in j. Imposing no transfers ($c_j^{\theta} = z_j^{\theta}$) and using that $z_j^{\theta} = w_j^{\theta}$ in a market allocation gives the result equation (21).

C. Planning Problem and Proofs of Propositions 1–3

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

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

(ii) the tradeable and nontradeable goods feasibility constraints

$$\sum_i d_{ji} Q_{ji} \leqslant Y_j\left(N_j^Y, I_j^Y\right) \text{ for all } j, i;$$
$$\sum_\theta L_j^\theta c_j^\theta + I_j^Y + I_j^H \leqslant Q\left(Q_{1j}, .., Q_{Jj}\right) \text{ for all } j;$$
$$\sum_\theta L_j^\theta h_j^\theta \leqslant H_j\left(N_j^H, I_j^H\right) \text{ for all } j;$$

(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) nonnegativity constraints on consumption, trade flows, intermediate inputs, and labor.

PROPOSITION 1. If a competitive equilibrium is efficient, then

$$W_j \frac{dN_j}{dL_j^\theta} + \sum_{\theta'} \frac{x_j^{\theta'} L_j^{\theta'}}{a_j^{\theta'}} \frac{\partial a_j^{\theta'}}{\partial L_j^\theta} = x_j^\theta + E^\theta \quad \text{if } L_j^\theta > 0, \tag{62}$$

for all j and θ and some constants $\{E^\theta\}$. If the planner's problem is globally concave and (62) holds for some specific $\{E^\theta\}$, 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:

$$\begin{aligned}
\mathscr{L} = u^{\theta} &- \sum_{j} \omega_j^{\theta} L_j^{\theta'} \left(u^{\theta} - a_j^{\theta'} \left(L_j^1, .., L_j^{\Theta} \right) U \left(c_j^{\theta'}, h_j^{\theta'} \right) \right) \\
&- \sum_{\theta' \neq \theta} \sum_{j} \omega_j^{\theta'} L_j^{\theta'} \left(\underline{u}^{\theta'} - a_j^{\theta'} \left(L_j^1, .., L_j^{\Theta} \right) U \left(c_j^{\theta'}, h_j^{\theta'} \right) \right) \\
&- \sum_{j} p_j^* \left(\sum_{i} d_{ji} Q_{ji} - Y_j \left(N_j^Y, I_j^Y \right) \right) \\
&- \sum_{j} P_j^* \left(\sum_{\theta} L_j^{\theta} c_j^{\theta} + I_j^Y + I_j^H - Q \left(Q_{1j}, .., Q_{Jj} \right) \right) \\
&- \sum_{j} R_j^* \left(\sum_{\theta} L_j^{\theta} h_j^{\theta} - H_j \left(N_j^H, I_j^H \right) \right) \\
&- \sum_{j} W_j^* \left(N_j^Y + N_j^H - N \left(z_j^1 \left(L_j^1, .., L_j^{\Theta} \right) L_j^1, .., z_j^{\Theta} \left(L_j^1, .., L_j^{\Theta} \right) L_j^{\Theta} \right) \right) \\
&- \sum_{\theta} E^{\theta} \left(\sum_{j} L_j^{\theta} - L^{\theta} \right) + ... \qquad (63)
\end{aligned}$$

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

$$[Q_{ji}] \qquad P_i^* \frac{\partial Q(Q_{1i}, .., Q_{Ji})}{\partial Q_{ji}} \leqslant p_j^* \tau_{ji}, \qquad (64)$$

$$\left[N_j^Y, N_j^H \right] \qquad p_j^* \frac{\partial Y_j}{\partial N_j^Y} \leqslant W_j^*; R_j^* \frac{\partial H_j}{\partial N_j^H} \leqslant W_j^*, \qquad (65)$$

$$\left[I_j^Y, I_j^H \right] \qquad p_j^* \frac{\partial Y_j}{\partial I_j^Y} \leqslant P_j^*; R_j^* \frac{\partial H_j}{\partial I_j^H} \leqslant P_j^*, \qquad (66)$$

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

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

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

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

$$\omega_j^\theta a_j^\theta U\left(c_j^\theta, h_j^\theta\right) = x_j^{\theta *}, \tag{67}$$

where

$$x_j^{\theta *} \equiv R_j^* h_j^\theta + P_j^* c_j^\theta. \tag{68}$$

Therefore, we can write

$$[c_j^\theta] \qquad c_j^\theta = \frac{\alpha_C\left(c_j^\theta, h_j^\theta\right)}{P_j^*} x_j^{\theta *} \tag{69}$$

$$[h_j^\theta] \qquad h_j^\theta = \frac{1-\alpha_C\left(c_j^\theta, h_j^\theta\right)}{R_j^*} x_j^{\theta *}, \tag{70}$$

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

Using expression (68) and the slackness condition on the spatial mobility constraint, the first-order condition of the planning problem with respect to L_j^θ is:

$$\sum_{\theta'} \omega_j^{\theta'} L_j^{\theta'} \frac{\partial a_j^{\theta'}\left(L_j^1, .., L_j^\Theta\right)}{\partial L_j^\theta} U\left(c_j^{\theta'}, h_j^{\theta'}\right) + W_j^* \frac{dN_j}{dL_j^\theta} \leqslant x_j^{\theta *} + E^\theta, \tag{71}$$

with equality if $L_j^\theta > 0$. Further using expression (67), if $L_j^\theta > 0$ then:

$$W_j^* \frac{dN_j}{dL_j^\theta} + \sum_{\theta'} \frac{\left(x_j^{\theta *}\right)' L_j^{\theta'}}{a_j^{\theta'}} \frac{\partial a_j^{\theta'}}{\partial L_j^\theta} = x_j^{\theta *} + E^\theta. \tag{72}$$

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:

$$W_j^* \frac{dN_j}{dL_j^\theta} + \sum_{\theta' \neq \theta} \frac{\left(x_j^{\theta *}\right)' L_j^{\theta'}}{a_j^{\theta'}} \frac{\partial a_j^{\theta'}}{\partial L_j^\theta} \leqslant E^\theta. \tag{73}$$

An optimal allocation is given by quantities $\{Q_{ji}, N_j^Y, N_j^H, I_j^Y, I_j^H, c_j^\theta, h_j^\theta, L_j^\theta, u^\theta\}$ and multipliers $\{P_j^*, p_j^*, R_j^*, W_j^*, \omega_j^\theta\}$ such that the first-order conditions (64)–(72) and the constraints enumerated in (i) to (iii) in Definition 2 hold.

It is straightforward to show that conditions (64) to (66), (69), and (70) 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 $\{P_j^*, p_j^*, R_j^*, W_j^*\}$ 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 $\{Q_{ji}, N_j^Y, N_j^H, I_j^Y, I_j^H, c_j^\theta, h_j^\theta, L_j^\theta\}$ 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 $\{P_j^*, p_j^*, R_j^*, W_j^*\}$ 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 expression (72). Conversely, if $x_j^\theta = x_j^{\theta *}$ for $x_j^{\theta *}$ defined in expression (62) given the W_j that solves the planner's problem, then there is a solution for the competitive allocation such that $\{P_j, p_j, R_j, W_j\} = \{P_j^*, p_j^*, R_j^*, W_j^*\}$. If the planning problem is concave then there is a unique solution to the system characterizing the planner's allocation, in which case $\{P_j, p_j, R_j, W_j\} = \{P_j^*, p_j^*, R_j^*, W_j^*\}$ is the only competitive equilibrium. □

PROPOSITION 2. The optimal allocation can be implemented by the transfers

$$t_j^{\theta *} = \sum_{\theta'} \left(\gamma_{\theta,\theta'}^{P,j} w_j^{\theta' *} + \gamma_{\theta,\theta'}^{A,j} x_j^{\theta' *}\right) \frac{L_j^{\theta' *}}{L_j^{\theta *}} - \left(b^\theta \Pi^* + E^\theta\right), \tag{74}$$

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

Proof. Combining equation (23) and condition (22), we get:

$$w_j^\theta - x_j^\theta + \sum_{\forall\theta'}\left(\gamma_{\theta,\theta'}^{P,j} w_j^{\theta'} + \gamma_{\theta,\theta'}^{A,j} x_j^{\theta'}\right)\frac{L_j^{\theta'}}{L_j^\theta} = E^\theta. \tag{75}$$

Combining this last expression with equation (19) gives the result. □

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

Proof. We consider the following planning problem defined in Section II.D:

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

where θ is a given type, $\mathcal{U}$ is the set of attainable utility levels $\{u^\theta\}$ 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 that are defined in the main text and to which we come back 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:

$$\max_{\left\{v^\theta, U_j^\theta, C_j^\theta, H_j^\theta, \widetilde{L}_j^\theta, \widetilde{N}_j^k, I_j^k, Q_{ij}, M_j, S_j\right\}} v^\theta \tag{76}$$

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

$$\underline{v}^{\theta'} - \mathcal{F}\left(\frac{U_j^{\theta'}\prod_{\theta''\neq\theta'}\left(\widetilde{L}_j^{\theta''}\right)^{\frac{\gamma_{\theta'',\theta'}^A}{1+\Gamma^P}}}{\left(\widetilde{L}_j^{\theta'}\right)^{\frac{1-\gamma_{\theta',\theta'}^A}{1+\Gamma^P}}}\right) \leqslant 0 \text{ for all } j \text{ and } \theta'; \tag{77}$$

$$U_j^\theta - U\left(C_j^\theta, H_j^\theta\right) \leqslant 0 \tag{78}$$

$$\sum_i d_{ji} Q_{ji} - \left(b_Y^N \left(N_j^Y\right)^{\beta_Y} + b_Y^I \left(I_j^Y\right)^{\beta_Y}\right)^{\frac{1}{\beta_Y}} \leqslant 0 \text{ for all } j, i; \tag{79}$$

$$\sum_\theta C_j^\theta + \left(I_j^Y\right) + \left(I_j^H\right) - Q\left(Q_{1j}, .., Q_{Jj}\right) \leqslant 0 \text{ for all } j; \tag{80}$$

$$\sum_\theta H_j^\theta - \left(b_H^N \left(N_j^H\right)^{\beta_H} + b_H^I \left(I_j^H\right)^{\beta_H}\right)^{\frac{1}{\beta_H}} \leqslant 0 \tag{81}$$

$$M_j - \left[\sum_\theta \left(Z_j^\theta \prod_{\theta'} \left(\widetilde{L}_j^{\theta'}\right)^{\frac{\gamma_{\theta',\theta}^P}{1+\Gamma^P}} \left(\widetilde{L}_j^\theta\right)^{\frac{1}{1+\Gamma^P}}.\right)^\rho\right]^{\frac{1}{\rho}} \leqslant 0 \text{ for all } j; \tag{82}$$

$$N_j^Y + N_j^H - M_j \leqslant 0 \tag{83}$$

$$\sum_j \left(\widetilde{L}_j^\theta\right)^{\frac{1}{1+\Gamma^P}} - L^\theta = 0 \text{ for all } \theta. \tag{84}$$

To reach these expressions, we introduced the auxiliary variables M_j and U_j^θ and we used the following change of variables: $v^\theta = \mathcal{F}(u^\theta)$, $H_j^\theta = L_j^\theta h_j^\theta$, $C_j^\theta = L_j^\theta c_j^\theta$, and $\widetilde{L}_j^\theta = (L_j^\theta)^{1+\Gamma^P}$ for all j and θ, where the function $\mathcal{F}(\cdot)$ 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:

$$L^\theta - \sum_j \left(\widetilde{L}_j^\theta\right)^{\frac{1}{1+\Gamma^P}} \leqslant 0 \text{ for all } \theta. \tag{85}$$

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 (77), and examine specifically the expression:

$$f_j^\theta\left(U_j^\theta, \{L^\theta\}, \{L^{\theta'}\}\right) = U_j^\theta \prod_{\theta' \neq \theta} \left(\widetilde{L}_j^{\theta'}\right)^{\frac{\gamma_{\theta',\theta}^A}{1+\Gamma^P}} \left(\widetilde{L}_j^\theta\right)^{-\frac{1-\gamma_{\theta,\theta}^A}{1+\Gamma^P}}. \tag{86}$$

This expression is a multivariate function of the form $f(y, z) = \prod_{i=1}^{k} y_i^{a_i} z^{-b}$ where $a_i > 0$, $b > 0$ and $\sum_{i=1}^{k} a_i < b$. By Proposition 11 of Khajavirad, Michalek, and Sahinidis (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 $\gamma^A_{\theta',\theta} \geqslant 0$ 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, Michalek, and Sahinidis (2014), the transformation $G_\theta(x) = -x^{\frac{1+\Gamma^P}{\Gamma^P - (\sigma_\theta + \sum_{\theta'} \gamma^A_{\theta',\theta})}}$ ensures that $G_\theta(f_j^\theta(\cdot))$ is concave. Finally, given the definition of Γ^A, $\mathcal{F}(\cdot)$ is a concave transform of $G_\theta(\cdot)$. Therefore, constraint (77) is convex for all θ'.

Second, functions of the form $f(x_1, ..., x_n) = [\sum a_i x_i^\beta]^\rho$ are concave whenever $\beta \in (0, 1)$ and $\rho\beta \leqslant 1$. Therefore, constraints (79), (81), and (85) are convex.

The constraint (78) is convex because $U(\cdot)$ is concave. The constraint (80) is convex because the aggregator $Q(\cdot)$ is concave.

Next, consider the constraint (82). The second term is the negative of a composition of an increasing CES function with exponent $\rho \leqslant 1$, 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 (82) is convex. Finally, constraint (83) 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 (85) binds, so that constraint (84) 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 $\mathcal{P}$.[50]

We now specialize to the case of a single type of worker ($\Theta = 1$) 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:

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

subject to constraints (79), (80), (82), (83), and (85), which are unchanged except that they now hold for only one group. We have used the following change of variable $\widetilde{H}_j^\theta = (H_j^\theta)^{1+d_{H,j}}$. The modified constraint for housing production is:

$$\text{(87)} \qquad \widetilde{H}_j^\theta - \left(b_H^N \left(\widetilde{N_j^H}\right)^{\beta_H(1+D)} + b_H^I \left(\widetilde{I_j^H}\right)^{(1+D)\beta_H}\right)^{\frac{1}{\beta_H}\frac{1}{1+D}} \leqslant 0.$$

The modified housing market constraint (87) 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 $(1-\alpha_C)\frac{1}{1+d_{H,j}} + \alpha_C \leqslant \frac{1-\gamma_{\theta,\theta}^A}{1+\Gamma^P}$ 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. □

D. Preference Draws within Types

The Lagrangian of the planning problem in Section III.E is a special case of expression (63), except that now the spillover function $a_j^{\theta'}(L_j^1, .., L_j^\Theta)$ is replaced by $a_i^{\theta'}(L_i^\theta)^{-\sigma_\theta}$. Following the same steps as in the proof of Proposition 1, we find that condition (22) is extended to

$$\text{(88)} \qquad W_j \frac{dN_j}{dL_j^\theta} + \sum_{\theta'} \frac{x_j^{\theta'} L_j^{\theta'}}{a_j^{\theta'}} \frac{\partial a_j^{\theta'}}{\partial L_j^\theta} = x_j^\theta (1+\sigma_\theta) + E^\theta \quad \text{if } L_j^\theta > 0.$$

50. We have not proven that expression (85) 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.

Following the same steps as in the proof of Proposition 2, we find that equation (24) is extended to

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

$$- \left(b^\theta \Pi^* + E^\theta\right). \tag{89}$$

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 Appendix Section G below, the only modification being that the term $\gamma_{\theta,\theta}^A$ is replaced by $\gamma_{\theta,\theta}^A - \sigma_\theta$.

E. Commuting

The Lagrangian of the planning problem described in the extension to spillovers across locations in Section III.E is

$$\mathscr{L} = u - \sum_j \sum_i \omega_{ji} \left(u - L_{ji}^{-\sigma} L^\sigma a_j \left(L_j^R\right) U_{ji}\left(c_{ji}, h_{ji}\right)\right)$$

$$- \sum_j p_j^* \left(\sum_i d_{ji} Q_{ji} - Y_j \left(N_j^Y, I_j^Y\right)\right)$$

$$- \sum_j P_j^* \left(\sum_i L_{ji} c_{ji} + I_j^Y + I_j^H - Q\left(Q_{1j}, .., Q_{Jj}\right)\right)$$

$$- \sum_j W_j^* \left(N_j^I + N_j^H - z_j \left(L_j^W\right) L_j^W\right)$$

$$- \sum_j R_j^* \left(\sum_i L_{ji} h_{ji} - H_j \left(N_j^H, I_j^H\right)\right)$$

$$-E \left(\sum_j \sum_i L_{ji} - L\right) + ... \tag{90}$$

where $L_j^R = \sum_{i'} L_{ji'}$, and $L_i^W = \sum_{j'} L_{j'i}$ are the residents and workers at j and i, respectively. The planner optimizes over the bilateral flows L_{ji} from place of residence j to place of work i, the consumption of tradeables and nontradeables 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 the allocation of inputs into production of tradeables and nontradeables). The first-order condition with respect to L_{ji} is:

$$\begin{aligned}\left[L_{ji}\right]: -\sigma\omega_{ji}L_{ji}^{-\sigma-1}a_j\left(L_j^R\right)U_{ji}\left(c_{ji},h_{ji}\right)+\sum_{i'}\omega_{ji'}L_{ji'}^{-\sigma}a_j'\left(L_j^R\right)\\ \times U_{ji'}\left(c_{ji'},h_{ji'}\right)+W_i^*\left(z_i'\left(L_i^W\right)L_i^W+z_i\left(L_i^W\right)\right)\\ =P_j^*c_{ji}+R_j^*h_{ji}+E.\end{aligned}\tag{91}$$

In addition, the first-order conditions over c_{ji} and h_{ji} and homogeneity of degree 1 of U_{ji} imply $\omega_{ji}L_{ji}^{-\sigma-1}a_j(L_j^R)U_{ji}(c_{ji},h_{ji})=x_{ji}^*$. Combining this expression with condition (91), using the definition of spillover elasticities $\gamma_i^P=\frac{z_i'(L_i^W)}{z_i(L_i^W)}L_i^W$ and $\gamma_j^A=\frac{a_j'(L_j^R)}{a_j(L_j^R)}L_j^R$, and rearranging we get:

$$x_{ji}^*=\frac{\gamma_j^A}{1+\sigma}\sum_{i'}\frac{L_{ji'}x_{ji'}^*}{L_j^R}+\frac{1+\gamma_i^P}{1+\sigma}W_i^*z_i\left(L_i^W\right)-\frac{E}{1+\sigma}.\tag{92}$$

To reach equation (35) we further use that the wage received by a commuter who works in i is $w_i^*=W_i^*z_i(L_i^W)$, and the definition of expenditures $x_{ji}^*=w_i^*+\frac{\Pi^*}{L}+t_{ji}^*$.

F. Spillovers across Locations

The Lagrangian of the planning problem described in the extension to spillovers across locations in Section III.E is a special case of expression (63), except that now the supply of efficiency units in j is $N_j(\{L_{j'}\})=z_j(\{L_{j'}\})L_j$. Compared to our derivation of Proposition 1, the only difference is the first-order condition with respect to employment. Now, instead of expression (72) we reach:

$$\sum_{j'}W_{j'}^*\frac{dN_{j'}}{dL_j}+x_j^*\frac{L_j}{a_j}\frac{\partial a_j}{\partial L_j}=x_j^*+E.\tag{93}$$

In addition, we now have:

$$W_j\frac{dN_j}{dL_j}=\begin{cases}w_{j'}\frac{L_{j'}}{L_j}\gamma^{P,j,j'} & \text{if } j'\neq j,\\ w_j\left(\gamma^{P,j,j}+1\right) & \text{if } j'=j.\end{cases}\tag{94}$$

Combining the last two expressions with equation (19) gives equation (38).

G. *The 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 over the policy space.

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

$$W_i N_i^Y = \left(1 - b_{Y,i}^I\right) p_i Y_i, \tag{95}$$

$$W_i N_i^H = \frac{1 - b_{H,i}^I}{1 + d_{H,i}} \left(1 - \alpha_C\right) X_i. \tag{96}$$

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

$$\frac{N_i^H}{N_i} = \frac{\frac{1-b_{H,i}^I}{1+d_{H,i}} \frac{1-\alpha_C}{1-b_{Y,i}^I} \left(\frac{X_i}{p_i Y_i}\right)}{\frac{1-b_H^I}{1+d_{H,i}} \frac{1-\alpha_C}{1-b_{Y,i}^I} \left(\frac{X_i}{p_i Y_i}\right) + 1}. \tag{97}$$

Using equations (95) and (96) along with labor market clearing expression (65), we can further express final consumption expenditures over tradeable income as a function of the shares of wages in expenditures:

$$\frac{X_i}{p_i Y_i} = \frac{1 - b_{Y,i}^I}{\frac{W_i N_i}{X_i} - \frac{1-b_{H,i}^I}{1+d_{H,i}} \left(1 - \alpha_C\right)}. \tag{98}$$

We now reformulate some of the equilibrium conditions from Definition 1 to include prices. Consider first the market-clearing condition (8). 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 rewrite that condition as

$$E_j^Y = \left(\alpha_C + (1-\alpha_C)\frac{b_{H,j}^I}{d_{H,j}+1}\right) X_j + b_Y^I \left(p_j Y_j\right), \tag{99}$$

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 on traded goods results from the aggregation of expenditures by consumers and final producers. Second, consider the market condition (7) for traded commodities. Multiplying both sides by the price of traded commodities at *j*, p_j, this condition is equivalent to

$$\sum_i s_{ji}^X = 1, \tag{100}$$

where $s_{ji}^X \equiv (\frac{E_i}{p_j Y_j}) s_{ji}^M$ 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

$$\sum_j s_{ji}^M \equiv 1. \tag{101}$$

2. 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 $\{\hat{P}_i, \hat{p}_i, \hat{R}_i\}$, wages $\{\hat{W}_i\}$, employment by group $\{\hat{L}_i^\theta\}$, supply of efficiency units $\{\hat{N}_i\}$, production of tradeable goods $\{\hat{Y}_i\}$, and utility levels $\{\hat{u}^\theta\}$ that satisfy a set of conditions given the change in expenditure per capita by group and location $\{\hat{x}_i^\theta\}$. The planner's problem in relative changes will then choose the optimal $\{\hat{x}_i^\theta\}$.

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

$$\sum_j s_{ij}^X \left(\frac{\hat{p}_i}{\hat{P}_j}\right)^{1-\sigma} \hat{E_j^Y} = \hat{p}_i \hat{Y}_i \text{ for all } i, \tag{102}$$

$$\sum_j s_{ji}^M \left(\frac{\hat{p}_j}{\hat{P}_i}\right)^{1-\sigma} = 1 \text{ for all } i, \tag{103}$$

$$\left(1 - \frac{N_i^H}{N_i}\right) \hat{p}_i \hat{Y}_i + \frac{N_i^H}{N_i} \hat{X}_i = \hat{W}_i \hat{N}_i \text{ for all } i, \tag{104}$$

$$\hat{W}_i^{1-b_{Y,i}^I} \hat{P}_i^{b_{Y,i}^I} = \hat{p}_i \text{ for all } i, \tag{105}$$

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 (102) and (103) follow from expressing equations (100) and (101) in relative changes and using the CES functional form equation (40). In condition (102), using equation (99) implies that the change in expenditures on tradeable commodities is:

$$\hat{E_j^Y} = \left(1 - b_{Y,j}^{\tilde{I}}\right) \hat{X}_j + b_{Y,j}^{\tilde{I}} \hat{p}_j \hat{Y}_j, \tag{106}$$

where

$$b_{Y,j}^{\tilde{I}} \equiv b_Y^I \frac{p_j Y_j}{E_j^Y} = \frac{b_Y^I}{\left(\alpha_C + (1-\alpha_C) \frac{b_{H,j}^I}{d_{H,j}+1}\right) \frac{X_j}{p_j Y_j} + b_Y^I}. \tag{107}$$

Condition (104) follows from expressing labor market clearing equation (10) in relative changes together with expressions (95) and (96), where the nontraded labor share $\frac{N_i^H}{N_i}$ is defined in expression (97). Condition (105) follows from optimization by producers of tradeable commodities.

The system (102) to (105) defines a solution for $\{\hat{P}_j, \hat{p}_j, \hat{Y}_j, \hat{W}_j\}$ 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. Heterogeneous groups and spillovers enter through $\hat{N}_i$. To reach an explicit expression for $\hat{N}_i$, we first note that the labor demand expression in the market allocation equation (17) allows us to back out the efficiency of each group:

$$z_i^\theta = \frac{w_i^\theta}{W_i}\left(\frac{L_i^\theta}{N_i}\right)^{\frac{1}{\rho}}, \tag{108}$$

Expressing the CES functional form for the aggregation of labor types in equation (43) in relative changes and using equation (108) we obtain:

$$\hat{N}_i = \left(\sum_\theta s_i^{W,\theta}\left(\hat{z_i^\theta}\hat{L_i^\theta}\right)^\rho\right)^{\frac{1}{\rho}}, \tag{109}$$

where

$$\hat{z_i^\theta} = \prod_{\theta'}\left(\hat{L_i^{\theta'}}\right)^{\gamma^P_{\theta',\theta}}, \tag{110}$$

and where $s_j^{W,\theta} = \frac{w_j^\theta L_j^\theta}{\sum_{\theta'} w_j^{\theta'} L_j^{\theta'}}$ is group θ's share of wages in city j. Expression (109) 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 $\{\hat{L_i^\theta}\}$ of each type in every location that is initially populated must also be consistent with the spatial mobility constraint (14),

$$\hat{u^\theta} = \hat{a_i^\theta}\frac{\hat{x_i^\theta}}{\hat{P_i}^{\alpha_C}\hat{R_i}^{1-\alpha_C}}, \tag{111}$$

where

$$\hat{a_i^\theta} = \prod_{\theta'}\left(\hat{L_i^{\theta'}}\right)^{\gamma^A_{\theta',\theta}}, \tag{112}$$

and where $\hat{R}_i$ is the change in the price of nontraded 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:

$$\hat{R}_i = \left(\hat{p}_i^{\frac{1-b^I_{H,i}}{1-b^I_{Y,i}}} \hat{P}_i^{b^I_{H,i} - b^I_{Y,i}\frac{1-b^I_{H,i}}{1-b^I_{Y,i}}} \hat{X}_i^{d_{H,i}} \right)^{\frac{1}{1+d_{H,i}}}. \tag{113}$$

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 condition that the national labor market must clear for each labor type is

$$\sum_j s_j^{L,\theta} \hat{L}_j^\theta = 1 \text{ for all } \theta, \tag{114}$$

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 (102) to (105) defines a solution for $\{\hat{P}_j, \hat{p}_j, \hat{Y}_j, \hat{W}_j\}$ 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 (109) to (114) defines a solution for $\{\hat{N}_j, \hat{L}_j^\theta, \hat{u^\theta}\}$ given $\{\hat{p_i}, \hat{P}_i, \{\hat{x_i^\theta}\}, \hat{X}_i\}$. As a result, an equilibrium in changes given a change in expenditure per capita $\{\hat{x_j^\theta}\}$ consists of $\{\hat{P}_j, \hat{p}_j, \hat{Y}_j, \hat{W}_j, \hat{N}_j, \hat{L_j^\theta}, \hat{R}_j, \hat{u^\theta}\}$ such that equations (102) to (114) hold. These equations form a system of $5J + \Theta J + \Theta$ equations in an equal number of unknowns, where J is the number of locations and Θ is the number of types.

3. The Planner's Problem in Relative Changes. In the implementation, we solve an optimization over $\{\hat{x_j^\theta}\}$ subject to $\{\hat{P}_j, \hat{p}_j, \hat{Y}_j, \hat{W}_j, \hat{N}_j, \hat{L_j^\theta}, \hat{R}_j, \hat{u^\theta}\}$ consistent with equations (102) to (144) 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 ($\hat{u^{\theta'}} \geqslant \underline{\hat{u^{\theta'}}}$ 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 (95) to (101). 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 expenditure terms rather than in quantities. Conditions (64) and (66) 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 $\mathcal{P}_2$, therefore have 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 equations (102) to (114) under the planner's problem in relative changes defined above. Note that given the elasticities $\{\alpha_C, \rho, b^I_{Y,j}, b^I_{H,j}, d_{H,j}\}$, and as long as $b^I_Y > 0$, computing the change in tradeable expenditures requires information about gross expenditures over tradeable income, $\frac{X_j}{p_jY_j}$. This information is also needed to compute the nontraded labor share $\frac{N_i^H}{N_i}$ in equation (104). However, as shown in equations (97) and (98), $\frac{X_j}{p_jY_j}$ can be constructed from the elasticities $\{\alpha_C, b^I_Y, b^I_H, d_{H,j}\}$ and the share of wages in gross expenditures, $\frac{W_iN_i}{X_i}$. □

University of California, Los Angeles, and National Bureau of Economic Research
University of California, Berkeley, National Bureau of Economic Research, and Centre for Economic Policy Research

Supplementary Material

An Online Appendix for this article can be found at *The Quarterly Journal of Economics* online. Data and code replicating tables and figures in this article can be found in Fajgelbaum and Gaubert (2019), in the Harvard Dataverse, doi: 10.7910/DVN/KZUG3W.

References

Abdel-Rahman, Hesham M., and Alex Anas, "Theories of Systems of Cities," *Handbook of Regional and Urban Economics*, 4 (2004), 2293–2339.

Ahlfeldt, Gabriel M., Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf, "The Economics of Density: Evidence from the Berlin Wall," *Econometrica*, 83 (2015), 2127–2189.

Albouy, David, "The Unequal Geographic Burden of Federal Taxation," *Journal of Political Economy*, 117 (2009), 635–667.

———, "Evaluating the Efficiency and Equity of Federal Fiscal Equalization," *Journal of Public Economics*, 96 (2012), 824–839.

Albouy, David, Kristian Behrens, Frédéric Robert-Nicoud, and Nathan Seegert, "The Optimal Distribution of Population Across Cities," *Journal of Urban Economics*, 110 (2019), 102–113.

Allen, Treb, and Costas Arkolakis, "Trade and the Topography of the Spatial Economy," *Quarterly Journal of Economics*, 129 (2014), 1085–1139.

Allen, Treb, Costas Arkolakis, and Xiangliang Li, "Optimal City Structure," Mimeo, Yale University, 2015.

Allen, Treb, Costas Arkolakis, and Yuta Takahashi, "Universal Gravity," NBER Technical Report, 2014.

Arnott, Richard, "Does the Henry George Theorem Provide a Practical Guide to Optimal City Size?," *American Journal of Economics and Sociology*, 63 (2004), 1057–1090.

Baum-Snow, Nathaniel, and Ronni Pavan, "Inequality and City Size," *Review of Economics and Statistics*, 95 (2013), 1535–1548.

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

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

Bhagwati, Jagdish, and Harry G. Johnson, "Notes on Some Controversies in the Theory of International Trade," *Economic Journal*, 70 (1960), 74–93.

Bryan, Gharad, and Melanie Morten, "Economic Development and the Spatial Allocation of Labor: Evidence from Indonesia," Manuscript, London School of Economics and Stanford University, 2015.

Bunten, Devin, "Is the Rent Too High? Aggregate Implications of Local Land-Use Regulation," FEDS Working Paper No. 2017-064, 2017.

Busso, Matias, Jesse Gregory, and Patrick Kline, "Assessing the Incidence and Efficiency of a Prominent Place Based Policy," *American Economic Review*, 103 (2013), 897–947.
Caliendo, Lorenzo, Fernando Parro, Esteban Rossi-Hansberg, and Pierre-Daniel Sarte, "The Impact of Regional and Sectoral Productivity Changes on the US Economy," *Review of Economic Studies*, 85 (2018), 2042–2096.
Carrell, Scott E., Bruce I. Sacerdote, and James E. West, "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation," *Econometrica*, 81 (2013), 855–882.
Ciccone, Antonio, and Robert E. Hall, "Productivity and the Density of Economic Activity," *American Economic Review*, 86 (1996), 54–70.
Combes, Pierre-Philippe, "The Empirics of Economic Geography: How to Draw Policy Implications?," *Review of World Economics*, 147 (2011), 567–592.
Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon, "Spatial Wage Disparities: Sorting Matters!," *Journal of Urban Economics*, 63 (2008), 723–742.
Combes, Pierre-Philippe, and Laurent Gobillon, "The Empirics of Agglomeration Economies," *Handbook of Regional and Urban Economics*, 5 (2015), 247–348.
Davis, Donald R., and Jonathan I. Dingel, "A Spatial Knowledge Economy," NBER Technical Report, 2012.
Dekle, Robert, Jonathan Eaton, and Samuel Kortum, "Global Rebalancing with Gravity: Measuring the Burden of Adjustment," IMF Technical Report 3, 2008.
Desmet, Klaus, Dávid Krisztián Nagy, and Esteban Rossi-Hansberg, "The Geography of Development," *Journal of Political Economy*, 126 (2018), 903–983.
Desmet, Klaus, and Esteban Rossi-Hansberg, "Urban Accounting and Welfare," *American Economic Review*, 103 (2013), 2296–2327.
———, "Spatial Development," *American Economic Review*, 104 (2014), 1211–1243.
Diamond, Rebecca, "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980–2000," *American Economic Review*, 106 (2016), 479–524.
Dixit, Avinash, "Tax Policy in Open Economies," *Handbook of Public Economics*, 1 (1985), 313–374.
Dunbar, Ann E., "Metropolitan Area Disposable Personal Income: Methodology and Results for 2001–2007," BEA WP2009-05, 2009.
Duranton, Gilles, and Diego Puga, "Micro-Foundations of Urban Agglomeration Economies," *Handbook of Regional and Urban Economics*, 4 (2004), 2063–2117.
Duranton, Gilles, and Anthony J. Venables, "Pace-Based Policies for Development," NBER Technical Report, 2018.
Eeckhout, Jan, and Nezih Guner, "Optimal Spatial Taxation: Are Big Cities too Small?," CEPR Discussion Paper No. DP10352, 2015.
Eeckhout, Jan, Roberto Pinheiro, and Kurt Schmidheiny, "Spatial Sorting," *Journal of Political Economy*, 122 (2014), 554–620.
Fajgelbaum, Pablo D., and Cecile Gaubert, "Replication Data for: 'Optimal Spatial Policies, Geography, and Sorting'," (2019), Harvard Dataverse, doi: 10.7910/DVN/KZUG3W.
Fajgelbaum, Pablo D., Eduardo Morales, Juan Carlos Suárez Serrato, and Owen Zidar, "State Taxes and Spatial Misallocation," *Review of Economic Studies*, 86 (2018), 333–376.
Fajgelbaum, Pablo D., and Edouard Schaal, "Optimal Transport Networks in Spatial Equilibrium," NBER Technical Report, 2017.
Flatters, Frank, Vernon Henderson, and Peter Mieszkowski, "Public Goods, Efficiency, and Regional Fiscal Equalization," *Journal of Public Economics*, 3 (1974), 99–112.
Flood, Sarah, Miriam King, Steven Ruggles, and J. Robert Warren, "Integrated Public Use Microdata Series, Current Population Survey: Version 5.0," [data set], University of Minnesota, 2017.

Gaubert, Cecile, “Firm Sorting and Agglomeration,” *American Economic Review*, 108 (2018), 3117–3153.
Glaeser, Edward L., and Joshua D. Gottlieb, “The Economics of Place-Making Policies,” *Brookings Papers on Economic Activity*, 39 (2008), 155–253.
Head, Keith, and Thierry Mayer, “Gravity Equations: Workhorse, Toolkit, and Cookbook,” *Handbook of International Economics*, 4 (2014), 131–195.
Helpman, Elhanan, “The Size of Regions: Transport and Housing as Factors in Agglomeration,” in *Topics in Public Economics* David Pines, Efraim Sadka, and Itzhak Zilcha eds. (Cambridge: Cambridge University Press, 1998), 33–54.
Helpman, Elhanan, and David Pines, “Optimal Public Investment and Dispersion Policy in a System of Open Cities,” *The American Economic Review*, 70 (1980), 507–514.
Helsley, Robert W., and William C. Strange, “Coagglomeration, Clusters, and the Scale and Composition of Cities,” *Journal of Political Economy*, 122 (2014), 1064–1093.
Henderson, J. Vernon, “The Sizes and Types of Cities,” *American Economic Review*, 64 (1974), 640–656.
Herkenhoff, Kyle F., Lee E. Ohanian, and Edward C. Prescott, “Tarnishing the Golden and Empire States: Land-Use Restrictions and the US Economic Slowdown,” *Journal of Monetary Economics*, 93 (2018), 89–109.
Hsieh, Chang-Tai, and Peter J. Klenow, “Misallocation and Manufacturing TFP in China and India,” *Quarterly Journal of Economics*, 124 (2009), 1403–1448.
Hsieh, Chang-Tai, and Enrico Moretti, “Housing Constraints and Spatial Misallocation,” *American Economic Journal: Macroeconomics*, 11 (2019), 1–39.
Khajavirad, Aida, Jeremy J. Michalek, and Nikolaos V. Sahinidis, “Relaxations of Factorable Functions with Convex-Transformable Intermediates,” *Mathematical Programming*, 144 (2014), 107–140.
Kleven, Henrik, Camille Landais, Mathilde Muñoz, and Stefanie Stantcheva, “Taxation and Migration: Evidence and Policy Implications,” NBER Technical Report, 2019.
Kline, Patrick, and Enrico Moretti, “Local Economic Development, Agglomeration Economies and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority,” *Quarterly Journal of Economics*, 129 (2014a), 275–331.
———, “People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs,” *Annual Review of Economics*, 6 (2014b), 629–662.
Krugman, Paul, “Scale Economies, Product Differentiation, and the Pattern of Trade,” *American Economic Review*, 70 (1980), 950–959.
Lucas, Robert E., and Esteban Rossi-Hansberg, “On the Internal Structure of Cities,” *Econometrica*, 70 (2002), 1445–1476.
Melo, Patricia C., Daniel J. Graham, and Robert B. Noland, “A Meta-Analysis of Estimates of Urban Agglomeration Economies,” *Regional Science and Urban Economics*, 39 (2009), 332–342.
Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg, “Commuting, Migration, and Local Employment Elasticities,” *American Economic Review*, 108 (2018), 3855–3890.
Moretti, Enrico, *The New Geography of Jobs* (Boston: Houghton Mifflin Harcourt, 2012).
Neumark, David, and Helen Simpson, “Place-Based Policies,” *Handbook of Regional and Urban Economics*, 5 (2015), 1197–1287.
Ossa, Ralph, “A Quantitative Analysis of Subsidy Competition in the US,” NBER Technical Report, 2018.
Parkhomenko, Andrii, “The Rise of Housing Supply Regulation in the US: Local Causes and Aggregate Implications,” University of Southern California, 2018.
Pines, David, and Efraim Sadka, “Comparative Statics Analysis of a Fully Closed City,” *Journal of Urban Economics*, 20 (1986), 1–20.
Redding, Stephen J., “Goods Trade, Factor Mobility and Welfare,” *Journal of International Economics*, 101 (2016), 148–167.

Redding, Stephen J., and Esteban A. Rossi-Hansberg, “Quantitative Spatial Economics,” *Annual Review of Economics*, 9 (2017), 21–58.

Redding, Stephen J., and Matthew A. Turner, “Transportation Costs and the Spatial Organization of Economic Activity,” *Handbook of Regional and Urban Economics*, 5 (2015), 1339–1398.

Roback, Jennifer, “Wages, Rents, and the Quality of Life,” *Journal of Political Economy*, 90 (1982), 1257–1278.

Roca, Jorge De La, and Diego Puga, “Learning by Working in Big Cities,” *Review of Economic Studies*, 84 (2017), 106–142.

Rosen, Sherwin, “Wage-Based Indexes of Urban Quality of Life,” *Current Issues in Urban Economics*, 3 (1979), 324–345.

Rossi-Hansberg, Esteban, “A Spatial Theory of Trade,” *American Economic Review*, 95 (2005), 1464–1491.

Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Felipe Schwartzman, “Cognitive Hubs and Spatial Redistribution,” NBER Technical Report, 2019.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek, “IPUMS USA: Version 8.0 [data set],” Minneapolis, 2017.

Saiz, Albert, “The Geographic Determinants of Housing Supply,” *Quarterly Journal of Economics*, 125 (2010), 1253–1296.

Sandmo, Agnar, “Optimal Taxation in the Presence of Externalities,” *Swedish Journal of Economics*, 77 (1975), 86–98.

Wilson, John D., “A Theory of Interregional Tax Competition,” *Journal of Urban Economics*, 19 (1986), 296–315.

Zhelobodko, Evgeny, Sergey Kokovin, Mathieu Parenti, and Jacques-François Thisse, “Monopolistic Competition: Beyond the Constant Elasticity of Substitution,” *Econometrica*, 80 (2012), 2765–2784.

Zodrow, George R., and Peter Mieszkowski, “Pigou, Tiebout, Property Taxation, and the Underprovision of Local Public Goods,” *Journal of Urban Economics*, 19 (1986), 356–370.

CALL FOR NOMINATIONS

$300,000 Nemmers Prize in Economics

Northwestern University invites nominations for the Erwin Plein Nemmers Prize in Economics, to be awarded during the 2026–27 academic year. The prize pays the recipient $300,000. Recipients of the Nemmers Prize present lectures, participate in department seminars, and engage with Northwestern faculty and students in other scholarly activities.

Details about the prize and the nomination process can be found at **nemmers.northwestern.edu**. Candidacy for the Nemmers Prize is open to those with careers of outstanding achievement in their disciplines as demonstrated by major contributions to new knowledge or the development of significant new modes of analysis.Individuals of all nationalities and institutional affiliations are eligible except current or recent members of the Northwestern University faculty and past recipients of the Nemmers or Nobel Prize. **Nominations will be accepted until January 14, 2026.**

The Nemmers prizes are made possible by a generous gift to Northwestern University by the late Erwin Esser Nemmers and the late Frederic Esser Nemmers.

Northwestern

Nemmers Prizes • Office of the Provost • Northwestern University • Evanston, Illinois 60208
nemmers.northwestern.edu