

Optimal Spatial Policies, Geography and Sorting*

Pablo D. Fajgelbaum

UCLA and NBER

Cecile Gaubert

UC Berkeley, NBER and CEPR

First Draft: February 2018

Abstract

We study optimal spatial policies in quantitative trade and geography frameworks with spillovers and sorting of heterogeneous workers. We first characterize efficient spatial transfers and the labor subsidies that would implement them. Then, we quantify the aggregate and distributional effects of implementing these policies in the U.S. economy. Under homogeneous workers and constant-elasticity spillovers, a constant labor subsidy over space restores efficiency regardless of micro heterogeneity in fundamentals and trade costs. In that case, the quantification suggests that the observed spatial transfers in the U.S. are close to efficient. Spillovers across heterogeneous workers create an additional rationale for place-specific subsidies to attain optimal sorting. Under heterogeneous workers, the quantification suggests that optimal spatial policies may require stronger redistribution towards low-wage cities than in the data, reduce wage inequality in larger cities, weaken spatial sorting by skill, and lead to significant welfare gains. Spillovers across different types of workers are a key driving force behind these results.

*E-mail: pfajgelbaum@econ.ucla.edu, cecile.gaubert@berkeley.edu. We thank Arnaud Costinot for his discussion of the paper. We thank seminar participants at Berkeley, Dallas Fed, IIES, MIT, Richmond Fed, Rochester, and UCLA. 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.

1 Introduction

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

We develop and implement a new approach to answer these questions. We rely on a framework that nests two recent strands of quantitative spatial research with spillovers: the location choice model with sorting of heterogeneous workers in Diamond (2016), and the economic geography models in Allen and Arkolakis (2014) and Redding (2016). Crucially, we generalize these models to allow, in the market allocation, for arbitrary policies that transfer resources across agents and regions. We use the framework to characterize the spatially efficient net transfers across regions and workers, as well as the labor taxes and subsidies that would implement them. Then, we use data across MSA's in the United States to quantify the impact of optimal spatial transfers on the spatial allocation, wage inequality and welfare of low and high skill workers relative to the observed equilibrium.

The framework incorporates many key determinants of the spatial distribution of economic activity. Firms produce differentiated tradeable commodities and non-tradeables such as housing 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. Different types of workers vary in how productive they are in each location, 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 arbitrarily redistribute income across agents and regions. Unless the right policies are implemented, competitive equilibria are typically Pareto inefficient because workers do not internalize the impact of their location choice on aggregate amenities, and firms do not internalize the impact of their hiring decisions on aggregate productivity.

In the model, the multiple spatial forces lead to complex general-equilibrium ramifications of the spillovers. However, in the spirit of the “principle of targeting” pointed out by Dixit (1985), the first-best allocation can be tractably implemented by policies that act upon inefficient margins only. In particular, we derive a condition on the joint distribution of expenditures and wages across workers and regions that ensures the existence of an efficient competitive allocation. This condition generalizes a standard efficiency requirement from non-spatial environments with convex

technologies 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 agglomeration (through production or amenities) against the opportunity cost of attracting workers-consumers to each location. Because the location and consumption decisions are not separable, these opportunity costs are measured in terms of local consumption expenditures. Therefore, they vary across locations due to the compensating differentials born of geographic forces (congestion, amenities, trade costs, and non-traded goods). As a result, assessing the efficiency of an observed allocation requires information about both wages and expenditure per capita across locations. We carry this idea into the quantification, where we assess the efficiency of the observed spatial allocation based on data on wages and expenditures across MSA's.

We show that the optimal expenditure distribution ensuring spatial efficiency can be implemented by labor taxes or subsidies coupled with lump-sum transfers. Under homogeneous workers and constant-elasticity spillovers, a constant labor subsidy over space coupled with a lump sum transfer suffices to restore efficiency. Hence, somewhat paradoxically, in this case the optimal “place-based” policy turns out to be independent from space: every location is taxed in the same way. Spatial variation in optimal policies arises when there are spillovers across types. For example, if the productivity of low skill workers depends on the presence of high skill workers, spatial efficiency may require more mixing of heterogeneous workers across cities than in the competitive allocation, and the decentralized pattern of sorting by skill may be too strong. In the quantitative application, we explore whether the observed allocation features inefficient sorting.

We apply the results to clarify the normative welfare properties of well-known economic geography models from the literature, such as Allen and Arkolakis (2014), Helpman (1998) and Redding (2016). These models correspond to special cases of our framework provided we assume inelastic housing supply, a single worker type and no intermediate inputs. We find that, in these models, global efficiency is characterized by the distribution of trade imbalances between regions and can be implemented by a simple transfer rule. Specifically, efficiency is attained only if, in addition to labor income, each worker living in a location where the wage is w receives a transfer of $\eta(\bar{w} - w)$, where $\bar{w}$ is the average wage in the economy and η is a constant that depends on spillover elasticities and the non-traded expenditure share. We show that, because these models make different assumptions about imbalances in the absence of policy, they have very different implications for whether the optimal policy should redistribute income from high- to low-wage regions or vice-versa.

We then show how to use the main framework to quantify efficient outcomes given the data. First, given the optimal policy, the non-convexities implied by the spillovers may lead to multiple inefficient market equilibria alongside the optimal allocation. To tackle this issue, we identify a condition on the distributions of spillover and housing supply elasticities ensuring concavity of the planner's problem, and therefore uniqueness of the competitive allocation under the optimal spatial policies. This property generalizes existing uniqueness results to spatial environments with optimal policies and heterogeneous workers with asymmetric spillovers. Second, we identify the set of sufficient statistics in the data which, in addition to the spillover elasticities and standard production

and utility function parameters, are needed to implement the optimal allocation. In addition to standard data requirements to quantify economic geography models, such as the distributions of wages and employment across worker types and regions, backing out the fundamentals needed to compute the optimal allocation also requires information on expenditures per capita across worker types and regions.

We take the distribution of economic activity across MSA's in the U.S. in the year 2007 as our empirical setting. We allow for two skills groups, high skill (college) and low skill (non college) workers. To fulfill the data requirements enumerated above, we combine data on labor income, taxes and transfers at the city level from the BEA, with Census data that allows to break down these MSA-level totals down by demographic groups within cities and to control for within-group heterogeneity based on socio-demographic characteristics by skill group and MSA. To parametrize the spillover elasticities we rely on standard estimates from the literature, including Diamond (2016), Ciccone and Hall (1996) and Kline and Moretti (2014a), who estimate equations that are consistent with our model in the U.S. context.

We first implement the quantification through the lens of a version of our model that assumes away heterogeneity by skill. In that case, our results imply negligible welfare gains from optimal reallocation, suggesting that spatial transfers in the U.S. are close to efficient in terms of aggregate welfare. We demonstrate that this result is not a feature of our model but a feature of the data by applying our approach to counterfactual data corresponding to equilibria under alternative fundamentals and no spatial transfers. In equilibria where the variance of wages by MSA or the correlation between wage and city size is higher in absolute value than in the data, the welfare gains from optimal reallocation may be substantial.

Then, we compute spatially efficient allocations using the full model featuring both high and low skilled workers. Across cities, the optimal transfers decrease more steeply with city wages than in the data, implying a stronger redistribution towards low-wage cities. As a result, on average, there is employment reallocation from large to small cities. However, there is also considerable heterogeneity in optimal growth rates over the size distribution, and initial city size is a poor predictor of whether a city is too large or too small in the observed allocation.

When moving from the observed to the optimal equilibrium, the increase in transfers per capita from large to small cities is stronger for high-skill workers. As a result, the optimal reallocation towards initially smaller places is sharper for high skill workers. The differential patterns of reallocation by skill lead to weaker spatial sorting in the optimal allocation than in the data, pointing to too little mixing of skills in the observed equilibrium. Within-city wage inequality, in terms of the skill premium, on average decreases in initially larger and more unequal cities. As a result, the urban skill premium is tempered. In contrast to the case with homogeneous workers, the quantification with heterogeneous workers suggests potentially large welfare gains. We tease out the source of this difference by shutting down alternative channels in the full model with heterogeneous workers, and find that the spillovers across workers play a substantial role in driving the gap between the measured welfare gains under heterogeneous and homogeneous workers.

Overall, the results suggest that heterogeneity in skills is important to optimally design and measure the aggregate welfare gains from spatial policies and that spatially efficient allocations are likely to feature weaker spatial sorting and lower inequality within larger cities than what is observed empirically.

The rest of the paper is structured as follows. Section 2 connects the paper to the related literature. Section 3 presents the general environment and defines the decentralized allocation and the planner’s problem. Section 4 characterizes the optimal policies, teases out its implication in specific cases of the model corresponding the models from the literature, and determines the data that suffices to implement the model. Section 5 described the data and the calibration. Section 6 presents the quantitative implementation and Section 7 concludes. Proofs, additional derivations and details of the data construction are sent to the Appendix.

2 Relation to the Literature

The economic geography models covered by our framework introduce labor mobility and spillovers in general-equilibrium models with trade frictions such as Eaton and Kortum (2002) and Anderson and van Wincoop (2003).¹ These models are the basis of a growing body of research, summarized by Redding and Turner (2015) and Redding and Rossi-Hansberg (2017), that studies the counterfactual implications of particular shocks to economic fundamentals or trade costs in spatial settings. The standard applications impose restrictions on transfers across locations through assumptions on how the income generated by fixed factors is distributed across locations. We show that the optimal allocation must be implemented through a particular set of spatial transfers, and characterize the distribution of imbalances that ensures spatial efficiency.

Our framework also nests the model of Diamond (2016), who allows for asymmetric amenity and efficiency spillovers across workers with different skills in a Rosen (1979)-Roback (1982) location model and takes the model to the data using the U.S. as an empirical setting. Her analysis shows that these forces partly determine the spatial distribution of skill shares and skill wage premia, including the recent “great divergence” in these outcomes across cities in the U.S. pointed out by Moretti (2012), among others.² Giannone (2017) argues that productivity spillovers across workers are relevant to understand the spatial effects of skill biased technological change. In this paper, we focus on the normative implications of these forces. In doing so we are also motivated by the well known empirical evidence that larger cities feature higher wages, higher share of skilled workers, and higher skill premium, as documented among others by Combes et al. (2008). We ask quantitatively whether these observed patterns of spatial disparities are too strong from the perspective of spatial

¹These trade models belong to the class considered by Arkolakis et al. (2012), who show that the welfare implications of changes in trade costs depend on the elasticity of trade with respect to trade costs. Our theoretical analysis does not impose the constant-elasticity form for trade flows that is typically assumed in the gravity literature. The optimal spatial tax turns out to be independent from the trade elasticity.

²Baum-Snow and Pavan (2013) document that a considerable share of the increase in wage inequality in the U.S. is due to the increase in wage dispersion within larger cities after controlling for worker characteristics.

efficiency.³ Our results suggest that spatially efficient allocations would typically feature less spatial wage disparities and lower wage dispersion within larger cities than what is currently observed.

Also related to our research are recent papers such as Desmet and Rossi-Hansberg (2013), Brandt et al. (2013), Hsieh and Moretti (2015) and Behrens et al. (2017) that quantitatively study spatial misallocation through wedges, as well as others such as Fajgelbaum et al. (2015) and Ossa (2015) that study specific spatial policies such as state taxes or subsidies. These studies compare the actual allocation to counterfactual ones with different distributions of wedges or policies, but do not implement the globally optimal spatial allocation.⁴ In addition, these models usually proceed under the assumption that there is a single worker type, while a key component of our analysis is the allocation of heterogeneous workers with asymmetric spillovers. Relative to the typically small welfare impact of removing dispersion in spatial wedges under homogeneous workers, our results suggest significant welfare effects from moving to an optimal allocation in the presence of spillovers across heterogeneous skill groups. Gaubert (2015) studies a model without trade frictions, homogeneous workers, heterogeneous firms and a complementarity between city size and firm productivity, and characterizes the optimal allocation in that setup.

Starting with Henderson (1974) and Flatters et al. (1974), a large theoretical literature in urban economics studies optimal city sizes.⁵ Relative to that literature, we theoretically characterize and quantify globally efficient policies in a framework that simultaneously accounts for productivity and amenity spillovers, multiple sources of spatial heterogeneity, realistic trade frictions, and sorting of heterogeneous workers with asymmetric spillovers. Due to these multiple forces, and given that we allow for an arbitrary distribution of initial transfers to be observed from the data, there is no systematic prediction as to whether larger cities are too large or too small in a decentralized allocation. Our quantitative results suggest that an optimal policy would tend, on average, to reallocate workers towards currently smaller cities, but we also find considerable dispersion on whether cities would grow or shrink conditional on their observed initial size.

Finally, our paper also relates to analyses of place-based policies. A large empirical literature, summarized by Kline and Moretti (2014b) and Neumark et al. (2015), estimates the local impact of policies such as government investments in public goods and regional subsidies, often in the presence of agglomeration economies. For instance, Kline and Moretti (2014a) study the effect of a large regional development program. Glaeser et al. (2008) and Kline (2010) draw some of the aggregate implications of these policies. We characterize and implement first-best place-based policies in a quantitative spatial framework accounting for multiple indirect effects across economy.⁶ While we

³Recent papers including spatial sorting of heterogeneous individuals to rationalize some of these patterns include Eeckhout et al. (2014), Behrens et al. (2014), and Davis and Dingel (2012). In the first paper there are no spillovers. The latter feature an inefficient allocation but do not study optimal sorting.

⁴Ossa (2015) simulates non-cooperative and cooperative firm subsidies in a framework with homogeneous workers, home market effects and no amenity spillovers. His analysis abstracts from transfers across regions to attain the global optimum, as subsidies are financed locally and the cooperative solution features zero subsidies. Eeckhout and Guner (2015) numerically find an income tax that maximizes welfare in a model without trade frictions, homogeneous workers and efficiency spillovers only. Other recent papers studying different policies include Allen et al. (2015) who consider zoning restrictions within a city and Fajgelbaum and Schaal (2017) who consider transport network investments.

⁵For a review of this theoretical literature see sections 2 and 8 of Abdel-Rahman and Anas (2004).

⁶In his review of the policy implications of empirical economic-geography studies, Combes (2011) notes the lack

focus on place-based taxes and subsidies, we show that our approach can be easily adapted to accommodate investments in public goods.

3 Economic Geography Model with Worker Sorting and Spillovers

3.1 Environment

We start by describing the physical environment, then the competitive equilibrium and finally the planner's allocation. We consider a closed economy with a discrete number J of locations typically indexed by j or i . Workers are heterogeneous. Each worker belongs to one of Θ different types. The type indexes each worker's preference and productivity in each location, as well as each worker's capacity to generate and absorb productivity and amenity spillovers. The utility of a worker of type θ in location j is

$$u_j^\theta = a_j^\theta (L_j^1, \dots, L_j^\Theta) U(c_j^\theta, h_j^\theta). \quad (1)$$

The function $a_j^\theta(\cdot)$ captures the valuation of a worker of type θ for location j 's exogenous and endogenous amenities. Workers may vary in how much they value fundamental amenities associated with exogenous features of each location. They may also vary 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. Workers also derive utility from a bundle of differentiated tradeable commodities (c_j^θ) and from non-tradeable services including housing (h_j^θ).⁷ The utility function $U(c, h)$ is homogeneous of degree 1.⁸

In both the competitive and optimal equilibria that we will consider, the number L_j^θ of type- θ workers located at j is endogenous. National labor market clearing is

$$\sum_j L_j^\theta = L^\theta, \quad (2)$$

where L^θ is the fixed aggregate supply of group θ .

Every location produces traded and non-traded goods. Tradeable output uses the aggregate technology $Y_j(N_j^Y, I_j^Y)$ requiring services of labor N_j^Y and intermediates I_j^Y . Similarly, production in the non-traded sector is $H_j(N_j^H, I_j^H)$. The functions Y_j and H_j may be city-specific and

of a general-equilibrium analysis of the optimal allocation of employment in a model of regional trade allowing for geographic inter-dependencies.

⁷There is no heterogeneity across workers within each type. However, since the number of types is unrestricted, this is not a restriction on the theory. For example, two skilled workers with different preferences for location can be represented as two different types. A standard formulation in the empirical literature is to allow for heterogeneity in preferences for location across workers distributed according to an extreme value distribution. In Section 4.1 we show how to incorporate this margin explicitly.

⁸The analysis can be easily generalized to homothetic utility functions which are not homogeneous of degree 1 at the cost of more burdensome notation. E.g., we can accommodate functions of the form $F_0(U(c_j^\theta, h_j^\theta))$ at the cost of keeping track of the elasticity of F with respect to U . Such a formulation would capture decreasing marginal utility from the consumption bundle of each location.

feature constant or decreasing returns to scale. They can also be constants if the good is supplied inelastically. The decreasing returns arise due to the use of fixed factors in production, such as land. Therefore, the framework allows for different housing supply elasticities across cities through the city specific decreasing returns to scale in $H_j(\cdot)$.⁹

The feasibility constraint in the non-traded sector in j is

$$H_j(N_j^H, I_j^H) = \sum_{\theta} L_j^{\theta} h_j^{\theta}. \quad (3)$$

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

$$Y_j(N_j^Y, I_j^Y) = \sum_i d_{ji}Q_{ji}. \quad (4)$$

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

$$Q(Q_{1i}, \dots, Q_{Ji}) = \sum_{\theta} L_i^{\theta} c_i^{\theta} + I_i^Y + I_i^H. \quad (5)$$

We finally map the distribution of employment across types, $\{L_j^{\theta}\}$, to the labor services available in each sector. 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

$$z_j^{\theta} = z_j^{\theta}(L_j^1, \dots, L_j^{\Theta}) \quad (6)$$

efficiency units. The function z_j^{θ} captures exogenous differences in productivity between locations and skill groups, as well as productivity spillovers across workers. 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 living in the same place as other high-skill workers, or from living in more densely populated areas. In both traded and non-traded sectors, the services $z_j^{\theta}L_j^{\theta}$ of the various types of labor are combined through the constant-returns aggregator $N(z_j^1L_j^1, \dots, z_j^{\Theta}L_j^{\Theta})$. This

⁹To keep notations light, we do not feature land explicitly as a variable input and we assume that the fixed factors in production cannot be reallocated between the traded and non-traded sectors. This assumption can also be easily relaxed. If $H_j(N_j^H, I_j^H)$ is homothetic, so that it can be expressed as $H_j(N_j^H, I_j^H) = H_j^1(H_j^0(N_j^H, I_j^H))$ for some $H_j^0(N_j^H, I_j^H)$ that is homogeneous of degree 1, then in the competitive allocation the elasticity of H_j^1 is the housing supply elasticity.

aggregator also captures imperfect substitution across workers. Feasibility in the use of labor services dictates that

$$N_j^Y + N_j^H = N(z_j^1 L_j^1, \dots, z_j^\Theta L_j^\Theta). \quad (7)$$

This completes the description of the physical environment. We highlight two key features relative to an otherwise standard neoclassical framework with a representative worker/consumer. First, the location of a worker drives both her marginal product (productivity is place specific) and her marginal utility of consumption (through local amenities). Therefore, production and consumption decisions are not separable. Second, the framework features two potential sources of non-convexities through the amenity and productivity spillover functions. The utility of each agent may increase with the number of other agents in the same location through a_j^θ , and the labor aggregator $N(\cdot)$ may feature increasing returns to the number of workers in a particular group through $z_j^\theta (L_j^1, \dots, L_j^\Theta) L_j^\theta$. Both sources of non-convexities operate through the same margin, L_j^θ . The non-separability and the non-convexities in the labor allocation will play key roles in driving the policy implications of spillovers and data requirements for the quantitative implementation.

3.2 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. Agents do not take into account the impact of their choices on the spillover functions $a_j^\theta (L_j^1, \dots, L_j^\Theta)$ and $z_j^\theta (L_j^1, \dots, L_j^\Theta)$.

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

$$\max_{c_j^\theta, h_j^\theta} U(c_j^\theta, h_j^\theta) \quad s.t. \quad P_j c_j^\theta + R_j h_j^\theta = x_j^\theta, \quad (8)$$

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, the common component of utility from (1) becomes

$$u_j^\theta = a_j^\theta (L_j^1, \dots, L_j^\Theta) \frac{x_j^\theta}{\psi(P_j, R_j)}, \quad (9)$$

where $\psi(P, R)$ is the price index associated with U . Since workers are homogeneous within each type, their utility is equalized. Worker's location choice implies that

$$u_j^\theta \leq u^\theta, \quad \text{if } L_j^\theta > 0. \quad (10)$$

Firms Producers of traded and non-traded commodities solve:

$$\Pi_j^Y = \max_{N_j^Y, I_j^Y} p_j Y_j(N_j^Y, I_j^Y) - W_j N_j^Y - P_j I_j^Y Q_{ji}, \quad (11)$$

$$\Pi_j^H = \max_{N_j^H, I_j^H} R_j H_j(N_j^H, I_j^H) - W_j N_j^H - P_j I_j^H, \quad (12)$$

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. Given a distribution of wages per worker $\{w_j^\theta\}$, the wage per efficiency unit is the solution to a standard cost-minimization problem,

$$W_j = \min_{\{L_j^\theta\}} \sum_j w_j^\theta L_j^\theta \quad s.t. \quad N(z_j^1 L_j^1, \dots, z_j^\Theta L_j^\Theta) = 1.$$

Importantly, this labor hiring decision takes the efficiency z_j^θ as given, ignoring its dependence on the employment distribution. As a result, the wage of type- θ workers in location j equals the value of its marginal product given the efficiency distribution $\{z_j^\theta\}$:

$$w_j^\theta = W_j \frac{\partial N(z_j^1 L_j^1, \dots, z_j^\Theta L_j^\Theta)}{\partial L_j^\theta}. \quad (13)$$

We assume a no-arbitrage condition, so that the price in location i of the traded good from j equals $d_{ji}p_j$. This condition is ensured for instance by free entry of intermediaries who can buy and resell goods between regions. Given these prices, the trade flows are:

$$P_i \frac{\partial Q(Q_{1i}, \dots, Q_{ji})}{\partial Q_{ji}} = d_{ji}p_j. \quad (14)$$

In the competitive equilibrium the prices of consumption goods, P_j and R_j , adjust so that the corresponding goods markets clear.

Expenditure Per Worker The only component of the competitive allocation left to define is the per capita expenditure x_j^θ . By definition, x_j^θ is the sum of all sources of income for a type- θ worker who lives in j . This covers labor and non-labor income, as well as, potentially, government taxes and transfers. 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 $\Pi \equiv \sum_j \Pi_j^Y + \Pi_j^H$, so that $\sum b^\theta L^\theta = 1$. We assume that each individual's portfolio is type- but not place-specific: individuals of a given group hold the same portfolio regardless of where they locate.¹⁰ In addition, we allow for flexible government policies

¹⁰This assumption rules out cases considered in the trade and economic geography literature in which individuals are assumed to own part of the fixed factors where they choose to live, as in Caliendo et al. (2014). These cases introduce an additional distortion in the competitive allocation, therefore creating an additional role for policy besides dealing with spillovers. Empirically, that assumption has been made to help to match trade imbalances. In our model, the distribution of trade imbalances across cities can be fully explained by policies that tax and redistribute income across locations. Note though that assuming local ownership of fixed factors is straightforward to include in our

that tax and transfer income across locations, including city-specific labor income subsidies (or taxes if negative) τ_j^θ and lump-sum transfers T_j^θ . As a result, expenditure per capita is

$$x_j^\theta = w_j^\theta + b^\theta \Pi + t_j^\theta, \quad (15)$$

where the net transfer to a type- θ worker living in j is

$$t_j^\theta \equiv \tau_j^\theta w_j^\theta + T_j^\theta. \quad (16)$$

Net government transfers equal zero, implying that expenditure equals income:

$$\sum_j \sum_\theta L_j^\theta x_j^\theta = \sum_j \sum_\theta L_j^\theta w_j^\theta + \Pi. \quad (17)$$

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

- (i) the consumption choices c_j^θ, h_j^θ are a solution to (8) for expenditures x_i^θ satisfying (15), and employment L_j^θ is consistent with the spatial mobility constraint (10);
- (ii) the labor and intermediate input choices $N_j^Y, I_j^Y, N_j^H, I_j^H$ are a solution to (11) and (12), labor demand is given by (13), and trade flows Q_{ji} are given by (14);
- (iii) the government budget constraint is satisfied; i.e. (17) holds, and
- (iii) all markets clear, i.e. (2) to (7) hold.

3.3 Planning Problem

We characterize a planning problem where the planner chooses the endogenous margins of the competitive allocation: distribution of workers over locations and types $\{L_j^\theta\}$, consumption of traded and non-traded commodities $\{c_j^\theta, h_j^\theta\}$, trade flows $\{Q_{ij}\}$ and allocation of efficiency units and intermediate inputs, $\{N_j^Y, I_j^Y, N_j^H, I_j^H\}$. The planner implement policies that treat all individuals within a type θ in the same way and is bound by the spatial mobility constraint (10). Along with that constraint, the market clearing conditions (2) to (7) define a set $\mathcal{U}$ of attainable utility levels $\{u^\theta\}$. The optimal planning problem consists in reaching the frontier of that set,

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

The solutions of this problem given an arbitrary θ and all feasible values of $\underline{u}^{\theta'} \in \mathcal{U}$ for $\theta' \neq \theta$ defines the utility frontier. 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

analysis, and in fact in Section (4.1) we formally show the implications of this assumption for optimal policies.

choice on amenities through a_j^θ and firms do not internalize the impact of their hiring decisions on efficiency through z_j^θ . We turn in the next section to the solution of the planning problem.

4 Optimal Policies

To understand what policies may remedy the market inefficiencies we first characterize a necessary condition for an efficient allocation. Then we show that it can be implemented by proper choice of spatial taxes and subsidies. Formal derivations are detailed in Appendix A.1.

Optimal Expenditure Distribution We note that the competitive allocation from Definition (1) can be fully determined given an arbitrarily chosen expenditure distribution of $x_j^\theta \geq 0$ satisfying (17), without imposing the requirement that x_j^θ satisfies the rule (15). In other words, we can modify Definition (1) to treat expenditure per capita x_j^θ as primitive. Using the wages and profits at the resulting equilibrium allocation, we can then compute taxes and transfers $\{\tau_j^\theta, T_j^\theta\}$ consistent with (15) given the arbitrarily chosen x_j^θ .

This logic suggests that we can first 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. We do so now, and then move to the implementation of the optimal expenditure distribution through specific policies. Comparing a decentralized allocation given expenditures x_j^θ to the outcomes of the planning problem, detailed in Definition (2) of Appendix (A.1), we reach the following result.

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

$$W_j \frac{dN_j}{dL_j^\theta} + \sum_{\theta'} \frac{x_j^{\theta'} L_j^{\theta'}}{a_j^{\theta'}} \frac{\partial a_j^{\theta'}}{\partial L_j^\theta} = x_j^\theta + E^\theta \quad \text{if } L_j^\theta > 0, \quad (18)$$

for all j and θ and some constants $\{E^\theta\}$. If the planner's problem is globally concave and (18) holds for some specific $\{E^\theta\}$, then the competitive equilibrium is efficient.

Condition (18) 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 all inhabited locations. The first term on the left is the value of the marginal product of labor, including both the direct output effect and the productivity spillovers.¹¹ The second term is the marginal benefit (or costs if negative) through amenity spillovers on each group of workers living in j , measured in expenditure equivalent terms.

These marginal benefits from allocating a type θ worker to location j are equated to the marginal costs on the right. The first term, x_j^θ , results from the non-separability between a worker's location and consumption: each type- θ worker in j requires x_j^θ units of expenditures in that particular

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

location. From a social planning perspective this is a cost, because each additional worker in j translates into lower consumption of traded and non-traded commodities for other workers in that location. The last term, E^θ , measures the opportunity cost of employing a worker elsewhere, and is not place-specific.¹²

We can draw several useful implications from this result. First, it implies that market inefficiencies can be fully tackled through policies acting only on x_j^θ to reach (18). This is because the set of equations defining the competitive allocation coincides with the set defining the planner's allocation, except potentially for the expenditure distribution.¹³ Therefore, asking whether the spatial allocation is efficient is equivalent to asking whether the expenditure distribution in the market allocation lines up with (18). Despite the multiple general-equilibrium ramifications of the spillovers, no other margin but the expenditure distribution requires policy intervention.¹⁴ We construct policies that implement the optimal expenditure distribution in the next section.

Second, Proposition 1 extends a familiar efficiency condition from the misallocation literature to spatial environments. In our economy, “space” enters through trade costs, non-traded goods, congestion and amenities. In the absence of these forces,¹⁵ 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 selling differentiated products as in Hsieh and Klenow (2009), and (18) would collapse to the familiar condition that the marginal value-product of labor, $W_j \left(dN_j/dL_j^\theta \right)$, 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, the absence of compensating differentials justifies the common practice of inferring allocative inefficiencies from differences in income and number of workers. In our spatial environment with compensating differentials, the non-separability 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 . This property motivates an efficiency check of the observed allocation based on wage and expenditure data, and it drives the data requirements for the quantitative implementation of the model.

Finally, we note that (18) is a necessary but not sufficient condition for efficiency. Even if it

¹²The constant E^θ is the multiplier of the national market clearing constraint (2) in the planner's problem. It is the marginal welfare impact of adding an additional type- θ worker in the economy.

¹³In the planner's allocation, x_j^θ coincides with shadow value of the local goods assigned to each θ worker in j . The market allocation of efficiency units $\{N_j^H, N_j^Y\}$, intermediates $\{I_j^H, I_j^Y\}$, trade $\{Q_{ij}\}$ and relative consumption per capita $\left\{ \frac{c_j^\theta}{h_j^\theta} \right\}$ is efficient if the equilibrium prices $\{W_j, R_j, P_j, p_j\}$ equal the multipliers of the constraints on efficiency units, non-traded goods, traded goods, and exports faced by the planner. See the proof of Proposition 1.

¹⁴This compartmentalization of the inefficiencies in few margins 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.

¹⁵I.e., assuming away trade frictions ($d_{ij} = 1$, implying $P_j = P$), non-traded goods ($U(c_j^\theta, h_j^\theta) = c_j^\theta$), congestion ($\frac{\partial a_j^{\theta'}}{\partial L_j^\theta} = 0$) and preferences for location ($a_j^\theta = a_j$).

holds, there could exist inefficient market equilibria. However, the inefficient allocations consistent with (18) can be ruled out if the planner's problem is globally concave. In that case, there is only one allocation that satisfies the first order conditions of the planner's problem. Because these first order conditions correspond to the definition of a competitive equilibrium given expenditures, it must be that only one competitive equilibrium exists and is efficient. In Section 4.2 we introduce conditions for global concavity of the planner's problem.

First-Best Subsidies We move on now to considering specific policies that can implement the optimal expenditure distribution presented above. These policies will implement the optimal expenditure distribution if the market equilibrium with government transfer is unique, an issue we come back to in Section 4.2. To that end, we define the productivity spillover elasticity

$$\gamma_{\theta,\theta'}^{P,j} \equiv \frac{L_j^\theta}{z_j^{\theta'}} \frac{\partial z_j^{\theta'}}{\partial L_j^\theta}, \quad (19)$$

and the amenity spillover elasticity

$$\gamma_{\theta,\theta'}^{A,j} \equiv \frac{L_j^\theta}{a_j^{\theta'}} \frac{\partial a_j^{\theta'}}{\partial L_j^\theta}. \quad (20)$$

These elasticities capture the marginal spillover of a type θ worker on the efficiency and utility of each type θ' worker in city j . The case without spillovers corresponds to $\gamma_{\theta,\theta'}^{P,j} = \gamma_{\theta,\theta'}^{A,j} = 0$. So far we have not imposed functional forms, so that these elasticities can be variable. Combining labor demand (13) and expenditure per capita (15) in the market allocation with (1) we obtain the following result.

Proposition 2. *The optimal allocation can be implemented by labor income subsidies τ_j^θ such that*

$$\tau_j^\theta = \frac{\gamma_{\theta,\theta}^{P,j} + \gamma_{\theta,\theta}^{A,j} + \sum_{\theta' \neq \theta} \frac{\gamma_{\theta,\theta'}^{P,j} w_j^{\theta'*} + \gamma_{\theta,\theta'}^{A,j} x_j^{\theta'*}}{w_j^{\theta*}} \frac{L_j^{\theta'*}}{L_j^{\theta*}}}{1 - \gamma_{\theta,\theta}^{A,j}} \quad (21)$$

coupled with transfers

$$T_j^\theta = -b^\theta \Pi^* - \frac{E^\theta}{1 - \gamma_{\theta,\theta}^{A,j}}.$$

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

We highlight that the lump-sum transfers T_j^θ take care of distributional concerns across types in the planner's problem, and only vary over space if the own spillover elasticity $\gamma_{\theta,\theta}^{A,j}$ does. In the absence of spillovers ($\gamma_{\theta,\theta'}^{P,j} = \gamma_{\theta,\theta'}^{A,j} = 0$ for all θ, θ', j), these transfers would still be present for redistribution across types (as implied by the second welfare theorem), but the labor subsidies τ_j^θ would be equal to zero because the spatial allocation would be efficient. Therefore, the burden of dealing with the spatial inefficiencies falls on the spatial income subsidies τ_j^θ . We also note that,

even if lump-sum transfers were flexible over space, they would not always suffice to implement the first best in the absence of subsidies.¹⁶ Instead, a combination of labor subsidies with lump-sum transfers always does. We move on now to studying the implications of this optimal policy scheme in special environments.

4.1 Applications

In this section we explore the implications for optimal policy in different well-known models. So far we have not imposed any functional form, but a predominant assumption in theoretical and applied studies is that the spillover functions $z_j^\theta(\cdot)$ and $a_j^\theta(\cdot)$ adopt the constant-elasticity forms, $\gamma_{\theta,\theta'}^{P,j} = \gamma_{\theta,\theta'}^P$ and $\gamma_{\theta',\theta}^{A,j} = \gamma_{\theta',\theta}^A$. From now on we adopt these constant-elasticity formulations. We will also adopt them for the quantitative application (as defined in (37) and (38) below).

Optimal Place-Based Policies with Homogeneous Workers Consider first a case with a single worker type, $|\Theta| = 1$. This case corresponds to the majority of studies on spatial misallocation and place-based policies reviewed in the introduction. Denoting the productivity and amenity spillovers by γ^P and γ^A , respectively, and using the wage equation (13), the necessary efficiency condition over the distribution of expenditure per capita 18 simplifies to

$$x_j = \frac{1 + \gamma^P}{1 - \gamma^A} w_j - \frac{E}{1 - \gamma^A}, \quad (22)$$

whereas expression (21) implies that optimal labor subsidy becomes

$$\tau = \frac{\gamma^P + \gamma^A}{1 - \gamma^A}. \quad (23)$$

This optimal subsidy is accompanied by the lump-sum transfer $T = -\frac{1}{L}\Pi^* - \frac{E}{1 - \gamma^A}$, which ensures that condition (17) and the government budget constraint hold. Under negative congestion spillovers ($\gamma^A < 0$), if the agglomeration spillover is not too strong ($\gamma^P < -\gamma^A$), then all workers should pay the same tax (a negative subsidy, $\tau < 0$) everywhere.

We point out two important implications from this result. First, the optimal policy scheme (τ, T) does not vary over space. Somewhat paradoxically, the optimal “place-based” policy turns out to be independent from space: every location is taxed in the same way. In spite of this, the policy does impact the spatial allocation. To see why, it is key to consider the role of the compensating differentials implied by the spatial dimensions (trade costs, non-traded goods, and amenities). From the mobility constraint (10), indifference across populated locations j and j' implies:

$$\frac{\psi(P_{j'}, R_{j'}) / a_{j'}(L_{j'})}{\psi(P_j, R_j) / a_j(L_j)} = \frac{(1 + \tau) W_{j'} z_{j'}(L_{j'}) + T + \frac{\Pi}{L}}{(1 + \tau) W_j z_j(L_j) + T + \frac{\Pi}{L}}. \quad (24)$$

¹⁶In particular, T_j^θ can implement the first best if and only if $\{L_j^\theta\}$ can be inverted from the spatial mobility constraint (9). With multiple worker types, this inversion would be infeasible under no amenity spillovers ($\gamma_{\theta,\theta'}^{A,j} = 0$).

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 . In the presence of amenities, non-traded goods and trade costs, the relative compensating differentials vary across space. As a result, changes to the policy scheme (τ, T) leads to changes in the employment distribution. In contrast, in the absence of compensating differentials, the policy scheme (τ, T) ceases to impact the spatial allocation.

Second, (23) implies that the decentralized allocation without intervention is not efficient. Otherwise we would have found that $\tau = T = 0$ is optimal. But we have found that (τ, T) are different from zero, and that they generally distort the allocation. This property means that there exist aggregate effects from reallocating workers across regions, and in particular that there are gains from implementing the tax scheme (4) relative to a *laissez-faire* allocation.

This second result runs against the notion that, in an economy with a single type of worker and constant-elasticity spillovers, there can only be gains from policy interventions that reallocate workers if the elasticity in the spillovers varies across locations. This notion is expressed, for example, by Glaeser et al. (2008) and by Kline and Moretti (2014a), among others.¹⁷ Even though our framework includes margins that not included in these models (such as trade costs or intermediate inputs), these margins are not necessary to generate the result that there are gains from implementing the optimal (τ, T) scheme. For example, this result holds in a classic Rosen-Roback framework with local externalities, where, in addition to the features of the model we are currently discussing, there are no trade costs ($d_{ij} = 1$), no valuation for non-traded goods ($u(c, h) = c$), no intermediate inputs in production ($Y_j(N_j^Y, I_j^Y) = Y_j(z_j(L_j)L_j)$), and perfectly substitutable products across locations ($Q_i = \sum_j Q_{ji}$).¹⁸

Economic Geography Frameworks and Monopolistic Competition The environment laid out in Section 3 includes standard economic geography models, such as Helpman (1998), Allen and Arkolakis (2014) and Redding (2016) including either perfect competition or monopolistic competition with free entry. Our presentation so far assumed that each location j sells a j -specific homogeneous tradeable product under perfect competition. In Appendix A.2 we show that the normative implications are equivalent to those of a differentiated-products model with free entry of monopolistic producers instead of productivity spillovers.¹⁹

These models are the backbone of a growing body of quantitative research studying the spatial

¹⁷See the review of the literature on place-based policies by Neumark et al. (2015).

¹⁸In the further special case where, in addition to the previous restrictions, we impose constant-returns to scale production function Y_j , no spillover in amenities, and no differences in amenities across locations ($a_j = a(L_j)$), the decentralized allocation without taxes turns out to be efficient despite the existence of efficiency spillovers. In that case, the spatial forces are shut down and there are no compensating differentials. Efficiency requires equalization of value of marginal products of labor, which happens to be achieved in the market allocation through the spatial mobility constraint (10), because output per worker is proportional to the marginal product of labor. In that case, the policies (τ, T) do not distort the allocation.

¹⁹In particular, under a single worker type, the normative implications of a differentiated-products model with elasticity of substitution $\kappa > 1$ across varieties and no productivity spillovers are equivalent to those of a homogeneous-product model with spillover elasticity equal to $\gamma^P = \frac{1}{\kappa-1}$. Therefore, our normative analysis also encompasses the models in Helpman (1998) and Redding (2016), who assume free entry of monopolistically competitive firms.

implications of regional shocks such as infrastructure investments and regional productivity changes, summarized by Redding and Rossi-Hansberg (2017).²⁰ However, their normative implications have barely been explored. We now apply the previous results to shed light on the optimal transfers and policies in these models.

To specialize the physical environment to these models we need to impose a number of assumptions. First, we assume a single worker type with Cobb-Douglas preferences with weight α_C on the traded goods and with a constant amenity spillover elasticity γ^A . The common component of utility 1 then becomes $u_j = A_j L_j^{1+\gamma^A} c_j^{\alpha_C} h_j^{1-\alpha_C}$. Production only uses labor and the efficiency spillover has a constant elasticity γ^P , so that tradeable output in region j is $Y_j = Z_j L_j^{1+\gamma^P}$. Supply of non-traded goods in location j is inelastic and equal to H_j . In a competitive allocation, workers in j receive a gross wage w_j equal to tradeable output per worker.

Applying Proposition 1, under these assumptions we find that the expenditure of workers in j in the optimal allocation is:

$$x_j = w_j(1 - \eta) + \eta \bar{w}, \quad (25)$$

where $\bar{w} \equiv \sum \frac{L_i}{L} w_i$ is the average wage in the economy and $\eta \equiv 1 - \frac{\alpha_C(1+\gamma^P)}{1-\gamma^A}$ is a constant that combines the key elasticities. Our key finding is that, in these models, a simple linear relationship between expenditure and wages implements the efficient allocation. Micro heterogeneity in amenities, productivity, and trade costs do not affect the shape of this relationship. We note that, empirically, the corresponding optimal transfers $t_j^\theta = \eta(\bar{w} - w_j)$ show up as trade imbalances.²¹ Therefore, barring knife-edge cases on the parameters ($\eta = 0$) or the fundamentals (such that $w_j = \bar{w}$), the efficient allocation generically features trade imbalances. In particular, under the empirically consistent case of $\eta < 0$, efficiency requires net trade deficits in high-wage regions.

Should the optimal policy that implements 25 redistribute towards or away high-wage areas? The answer depends on the distribution of income in the market allocation, absent policy intervention. This distribution is shaped by a key assumption, namely how the returns to the fixed factor H_j are allocated across workers in different locations. The existing applications in this literature make different assumptions in this regard, ranging from no transfers to even ownership. To nest these cases we can assume, as in Caliendo et al. (2014), that a fraction ω of the returns to H_j is distributed locally to the L_j workers in j and the remainder is evenly split across all workers.²² The pattern of ownership assumed in (15) is equivalent in this context to $\omega = 0$. Then, proceeding as in Section 4, the optimal policy can be expressed as a labor subsidy τ coupled with a lump-sum

²⁰Some features studied in recent applications within this literature that are not included in our model include multiple traded sectors with input-output linkages as in Caliendo et al. (2014) and commuting as in Monte et al. (2015).

²¹Net transfers equal the difference between value added and expenditure per capita of a region, which in turn equal the difference between value added and expenditure in the tradeable sector.

²²After imposing these assumptions on the framework laid out in Section 3, the model becomes equivalent to Helpman (1998) when $\gamma^A = \omega = 0 < \gamma^P$, to Allen and Arkolakis (2014) when $\alpha_C = 1$, and to Redding (2016) when $\gamma^A = 0 < \gamma^P$ and $\omega = 1$. We remain more general than these models in that the trade-flows aggregator $Q(Q_{1i}, \dots, Q_{Ji})$ is not restricted to the constant-elasticity (CES) form. Therefore, we do not impose a constant trade elasticity.

payment T . The optimal subsidy is common across locations and equal to

$$\tau = \frac{1 + \gamma^P}{1 - \gamma^A} [1 - (1 - \alpha_C) \omega] - 1, \quad (26)$$

with lump-sum transfer of $T = -\tau \bar{w}$. The optimal policy redistributes income away from low-wage regions under a labor subsidy ($\tau > 0$), and into low-wage regions under a labor tax ($\tau < 0$).

We now identify the optimal policy in different models. We consider again the case when $\eta < 0$. In Helpman (1998), who assumes common ownership of the national portfolio ($\omega = 0$), the default transfers to low-income regions in the competitive equilibrium are too large from the lens of spatial efficiency; therefore, the optimal policy redistributes income to regions with above-average wage ($\tau > 0$). In contrast, in Allen and Arkolakis (2014) and Redding (2016), who assume away trade imbalances, the default transfers to low-income regions in the competitive equilibrium are too small; as result, the optimal policy redistributes income to low-wage regions ($\tau < 0$).

In sum, we have shown that several details of the microeconomic structure and the country's economic geography (represented by bilateral trade costs) do not impact the relationship between optimal trade imbalances and wages in economic geography models, nor the policies that implement them. We have also shown how the specific assumptions about ownership of fixed factors in the *laissez-faire* allocation of these models impact whether the optimal policies should redistribute income towards or away from high-wage regions.

Examples with Heterogeneous Types We now discuss some examples to gain intuition on the shape of optimal policies with heterogeneous types. First, consider a case without spillovers across types, $\gamma_{\theta, \theta'}^P = \gamma_{\theta, \theta'}^A = 0$ for all $\theta \neq \theta'$. Then, a direct implication of (21) is that the optimal labor subsidy is type-specific, but constant over space:

$$\tau^\theta = \frac{\gamma_{\theta, \theta}^P + \gamma_{\theta, \theta}^A}{1 - \gamma_{\theta, \theta}^A}. \quad (27)$$

As in the case with homogeneous workers, in this case with no cross-type spillovers the optimal place-based subsidy is not place-specific. In contrast, cross-spillovers create complementarity across types and a rationale for place-specific labor subsidies for each labor type. To see how, consider a polar case without any amenity spillover ($\gamma_{\theta, \theta'}^A = 0$ for all θ, θ') and without efficiency spillover on the own type ($\gamma_{\theta, \theta}^P = 0$). Then, (21) simplifies to

$$\tau_j^\theta = \gamma_{\theta, \theta'}^P \left(\frac{w_j^{\theta'} L_j^{\theta'}}{w_j^\theta L_j^\theta} \right). \quad (28)$$

In this special case, the subsidy for workers in group θ varies across locations according to the distribution of wage bills. A positive efficiency spillover of type θ on type θ' ($\gamma_{\theta, \theta'}^P > 0$) implies a higher marginal gain from attracting type- θ workers to locations where the economic size of the population of type θ' workers is relatively larger, as captured by their wage bill. The result is

a higher optimal subsidy for θ -workers where they are more scarce.²³ Relative to a *laissez-faire* equilibrium, this policy is likely to temper the degree of sorting across cities. In Section 6 we will return to these basic intuitions, as they will help us rationalize the quantitative findings about the spatial efficiency of the current transfer scheme across regions and workers with different levels of skill in the U.S. economy.

Extension to Public Spending So far we focused the spatial policies on pure transfers. However, place-based policies often take the form of funding of government spending in public goods that are valued by consumers and firms.²⁴ Our previous results can be easily generalized to this case. Relative to the previous framework we can allow for government spending valued by workers (G_j^U) to enter as an additional shifter in the amenity valuation for location j , $a_j^\theta = a_j^\theta(G_j^U, L_j^1, \dots, L_j^\theta)$, and government spending in public goods that impact productivity (G_j^Y) to enter as an additional shifter in the efficiency units of labor, $N_j = N_j(G_j^Y, z_j^1 L_j^1, \dots, z_j^\theta L_j^\theta)$. We can also allow for a production function of public goods $G_j = G_j(I_j^G, H_j^G)$ that takes as input the bundle of traded commodities (I_j^G) and non-traded services (H_j^G). This formulation encompasses most standard representations of public spending, including classic studies such as Flatters et al. (1974) and recent quantitative analyses such as Fajgelbaum et al. (2015). For example, whether the public good is rival or not can be captured by how the population distribution enters in the amenity function $a_j^\theta(\cdot)$.²⁵

Following the same steps as in Section 3.2, we define a decentralized allocation under the assumption that, for any given G_j , the government uses inputs efficiently. Focusing for simplicity on the case with a single type of worker, the generalized version of the necessary efficiency condition (22) is:

$$x_j = \left[(1 + \gamma^{A,G}) \frac{1 + \gamma^P}{1 + \gamma^A} + \frac{\varepsilon_{N,G}}{\varepsilon_{N,L}} \right] w_j - \frac{1 + \gamma^{A,G}}{1 + \gamma^A} E, \quad (29)$$

where $x_j \equiv x_j^P + x_j^G$ is the sum of public and private expenditure per capita. In addition to the standard spillover elasticities, (29) includes the elasticities of labor efficiency N_j with respect to government spending ($\varepsilon_{N,G}$) and the services of labor ($\varepsilon_{N,L}$), as well as the elasticity of amenities with respect to spending ($\gamma^{A,G}$), all of which are here assumed to be constant. Therefore, here too we obtain that a simple linear relationship between total expenditure per capita and wages must

²³For example, consider a case with unskilled and skilled workers, denoted by $\theta = U, S$ respectively, where only skilled workers generate positive productivity spillovers on the other type. Then, the optimal subsidies are $\tau^S = \gamma_{\theta, \theta'}^P \left(\frac{w^U L^U}{w^S L^S} \right)$ and $\tau^U = 0$. In this case $\frac{d\tau^S}{d\tau^U} < 0 \iff \gamma_{S,U}^P \gamma_{U,S}^P > 0$, so that subsidies of high and low skill workers are negatively correlated over space if both types generate positive efficiency spillovers. It is easy to show that, with a CES aggregator of skills and $\gamma_{S,U} < 1$, the optimal subsidy to high skill workers is higher where the population of low skill workers is higher.

²⁴Recent analyses of large place specific investment programs include Glaeser et al. (2008) and Kline (2010), and recent reviews of the literature on these investments include Kline and Moretti (2014b) and Neumark et al. (2015).

²⁵For example, with a single worker type, we can write $a_j(G_j^U, L_j) = A_j \left(\frac{G_j}{L_j^\chi} \right)^{\gamma^{A,G}} L_j^{\gamma_0^A}$. The coefficient χ captures the degree to which the public good is rival in consumption. In this context, using our previous notation we would have: $\gamma^A = \gamma_0^A - \gamma^{A,G}\chi$.

hold in an efficient allocation. In addition to this necessary condition, efficiency now also requires an optimal breakdown of total expenditure into its private and public components:

$$x_j^G = \gamma^{A,G} x_j^P + \frac{\varepsilon_{N,G}}{\varepsilon_{N,L}} w_j. \quad (30)$$

The efficiency conditions (29) and (30) can be implemented as before through a policy scheme of the form (τ, T) , now coupled with inter-governmental transfers to finance spending. Importantly, under any such scheme, the optimal tax τ remains the same as in the absence of public spending, as defined in (23). The reason is that the public goods do not modify the planner's incentives to assign workers according to the trade-off between marginal product and private expenditure cost described before. As a result, τ still targets the spillovers. Simultaneously, the optimal government expenditure in j is attained by an inter-governmental grant equal to $x_j^G L_j$. Hence, as long as transfers across regions are flexible, bringing in government spending to the picture does not change the optimal spatial tax we derived before.

Extension to Preference Draws within Types A standard assumption in the urban economics literature is that workers differ in their individual tastes for location. Our theory accommodates this margin, as θ may index preferences. Two workers who are identical in every dimension except for their preference for location can be represented as belonging to two different θ 's. However, for broad definitions of the types (e.g., if θ defines high and low skill workers), preference heterogeneity within types may be relevant.

This margin can be incorporated in our analysis. Letting $\mathcal{L}_\theta$ index the set of type- θ workers, the utility of a worker $l \in \mathcal{L}_\theta$ is now $u_j^\theta \epsilon_j^l$, for u_j^θ defined in (1). Now, ϵ_j^l captures the worker's idiosyncratic preference for location j . We make the standard assumption that the preference draw ϵ_j^l of a type- θ worker is i.i.d. across workers and locations according to a Fréchet distribution, $\Pr(\epsilon_j^l < x) = e^{-x^{-1/\sigma_\theta}}$. The parameter $\sigma_\theta \in [0, 1]$ indexes workers' immobility. The individual preference draws are eliminated when $\sigma_\theta = 0$, in which case we return to the original formulation of the model. Under this assumption, every aspect of the model environment and competitive allocation described in sections 3.1 and 3.2 remains unchanged except for the spatial mobility constraint (10), which is now replaced with the following labor-supply equation:

$$\frac{L_j^\theta}{L^\theta} = \left(\frac{u_j^\theta}{u^\theta} \right)^{1/\sigma_\theta}. \quad (31)$$

Now, u^θ is the expected utility of a type- θ individual before drawing the ϵ_j^l 's and the common components of utility u_j^θ may vary across populated locations in equilibrium.

We can compute the optimal allocation and construct optimal policies using the same definition of the planner's problem as in 3.3, replacing constraint (10) with (31). In this formulation, the planner still conditions outcomes on location j and type θ , but not on the individual preference draws ϵ_j^θ . I.e., the planner's objective function weighs symmetrically every individual within a

group, without knowledge of the draw received by each particular individual, Propositions 1 to 4 then go through with minor modifications. In particular, Proposition 2 takes the same form as before, with one difference: 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, preference draws that are unobserved by the planner slightly alter the optimal policy and create a reason for intervention. Even without spillovers ($\gamma_{\theta',\theta}^{A,j} = \gamma_{\theta',\theta}^{P,j} = 0$), this force leads to a labor tax, $\tau^\theta = -\frac{\sigma_\theta}{1+\sigma_\theta} < 0$ coupled with a lump-sum subsidy $T^\theta > 0$, resulting in redistribution towards low-wage locations. We must be careful in interpreting the reason for this policy intervention. Now, the planner has incentives to redistribute income across cities for reasons other than spillovers. The reason is that, in the planner's problem, every individual within a group receives the same weight. However, in equilibrium, individuals with the same θ who sort into different locations have different marginal utilities from expenditures in traded and non-traded commodities.²⁶ In particular, individual sorting into lower-wage locations receive on average higher draws, creating the incentive to redistribute income towards low-wage locations.

4.2 Quantitative Implementation

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

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

$$U(c, h) = c^{\alpha_C} h^{1-\alpha_C}, \quad (32)$$

while the aggregator of traded commodities is CES,

$$Q(Q_{1i}, \dots, Q_{Ji}) = \left(\sum_i Q_{ji}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \quad (33)$$

where $\sigma > 0$ is the elasticity of substitution across products from different origins.

On the supply side, the production functions of traded and non-traded goods are Cobb-Douglas

²⁶I.e., the conditional distributions of ϵ_j^l is not the same across locations after sorting: $\Pr[\epsilon_j^l < x \mid j = \arg \max_j u_j^\theta, \epsilon_j^l]$ will vary by j .

in labor and intermediate inputs:

$$Y_j (N_j^Y, I_j^Y) = z_j^Y (N_j^Y)^{1-b_{Y,j}^I} (I_j^Y)^{b_{Y,j}^I}, \quad (34)$$

$$H_j (N_j^H, I_j^H) = z_j^H \left((N_j^H)^{1-b_{H,j}^I} (I_j^H)^{b_{H,j}^I} \right)^{\frac{1}{1+d_{H,j}}}, \quad (35)$$

where $d_{H,j} \geq 0$ and $\{z_j^Y, z_j^H\}$ are TFP shifters. We note that traded goods are produced under constant returns to scale, but we allow for decreasing returns in the housing sector. In particular, the coefficient $d_{H,j}$ plays the role of the inverse housing supply elasticity of location j , and is allowed to vary across regions.²⁷

The aggregator of labor types is CES,

$$N_j (L_j^1, \dots, L_j^\Theta) = \left[\sum_{\theta} \left(z_j^\theta (L_j^1, \dots, L_j^\Theta) L_j^\theta \right)^\rho \right]^{\frac{1}{\rho}}, \quad (36)$$

where $\rho > 0$ is the elasticity of substitution across types of workers.

Finally, we impose constant-elasticity forms for the spillovers:

$$z_j^\theta (L_j^1, \dots, L_j^\Theta) = Z_j^\theta \prod_{\theta'} (L_j^{\theta'})^{\gamma_{\theta',\theta}^P}, \quad (37)$$

$$a_j^\theta (L_j^1, \dots, L_j^\Theta) \equiv A_j^\theta \prod_{\theta'} (L_j^{\theta'})^{\gamma_{\theta',\theta}^A}. \quad (38)$$

These functional forms are standard in studies that estimate spillover elasticities such as Diamond (2016), 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 in location j .

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

$$\Gamma^P = \max_{\theta} \left\{ \sum_{\theta'} \gamma_{\theta',\theta}^P \right\},$$

$$\Gamma^A = \min_{\theta} \left\{ - \sum_{\theta'} \gamma_{\theta',\theta}^A \right\}.$$

Also, we let $D = \min_j \{d_{H,j}\}$ be the lowest inverse elasticity of housing supply. Under the functional form assumptions (32) to (38) we have the following property.

²⁷I.e., optimization of producers of non-traded goods implies that housing supply is $\ln H_j \propto \frac{1}{d_{H,j}} \ln R_j$.

Proposition 3. *The planning problem is concave if*

$$\Gamma^A > \Gamma^P, \quad (39)$$

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

$$1 - \gamma^A > (1 + \gamma^P) \left(\frac{1 - \alpha_C}{1 + D} + \alpha_C \right). \quad (40)$$

The first condition establishes concavity by imposing that congestion forces, are at least as large as agglomeration forces. Specifically, the congestion from the type that generates the weakest congestion, measured by Γ^A , dominates the agglomeration from the type that generates the strongest agglomeration, measured by Γ^P . Condition (39) may also apply when there is a single type, in which case it simplifies to $-\gamma^A > \gamma^P$. With a single type, further assuming Cobb-Douglas preferences over traded and non-traded goods we obtain the weaker restriction (40). These cases apply to the economic-geography models we reviewed in the introduction, which typically assume a single worker type.²⁸

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, we complement results that characterize uniqueness of the spatial equilibrium under no policy intervention and trade balance (Allen et al., 2014). Second, our result holds in a context with heterogeneous workers and cross-groups spillovers. We note that our uniqueness condition applies under the optimal expenditure distribution. Multiplicity is still possible under sub-optimal policies or no policy intervention, but this poses no limitation for our approach.

Implementation in Changes and Data Requirements We bring the model to the data by assuming that the observed allocation is generated by a decentralized equilibrium consistent with Definition (1) under the functional form assumptions (32) to (38). To compute the optimal allocation, we solve for a planner's problem that optimizes over the changes in the expenditure of each type of worker $\hat{x}_j^\theta$ and in the remaining endogenous variables $\{\hat{P}_i, \hat{p}_i, \hat{Y}_i, \hat{W}_i, \hat{N}_j, \hat{L}_j^\theta, \hat{R}_i, \hat{u}^\theta\}$ relative to this observed equilibrium.²⁹ The planner controls these variables to maximize the welfare change of one arbitrarily chosen group, $\hat{u}^\theta$, relative to the observed allocation, subject to lower bounds for the welfare changes of the remaining groups. We can then trace the utility frontier by solving the problem under different lower bounds for the remaining groups.

The solution to this problem in changes is equivalent to the solution to the problem in levels from Section (3.3), assuming that the same parameters and fundamentals underlie the observed

²⁸The CES restriction (33) on the aggregator of trade flows $Q(\cdot)$ is not needed for any of these results. Therefore, these condition holds regardless of product differentiation across locations. We also note that these conditions sufficient but not necessary for uniqueness.

²⁹As shown in Appendix (A.5), given an arbitrary change in expenditure per capita $\{\hat{x}_j^\theta\}$, an equilibrium in changes relative to the observed allocation consists of $\{\hat{P}_i, \hat{p}_i, \hat{Y}_i, \hat{W}_i, \hat{N}_j, \hat{L}_j^\theta, \hat{R}_i, \hat{u}^\theta\}$ such that the system of equations (A.45) to (A.54) hold.

equilibrium and the planning problem in levels. In the spirit of Dekle et al. (2008), as shown in Section A.5, this strategy allows us identify the data that is sufficient to back out fundamentals and solve for the optimal allocation, as stated in the following proposition.

Proposition 4. *Assume that the observed data is generated by a competitive equilibrium consistent with Definition (1) under the functional forms (32) to (38). Then, relative to the initial equilibrium, the optimal allocation can be fully characterized as function of:*

- i) the distributions of wages, employment and expenditures across labor types and locations;*
- ii) the distribution of bilateral import and export shares across locations;*
- iii) the utility and production function parameters $\{\alpha_C, \sigma, \rho, b_{Y,j}^I, b_{H,j}^I, d_{H,j}\}$; and*
- iv) the spillover elasticities $\{\gamma_{\theta',\theta}^A, \gamma_{\theta',\theta}^P\}$.*

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

Importantly, this implementation does not impose restrictions on the distributional policies across locations in the observed equilibrium. These policies, entering the model through the transfers t_j^θ defined in (15), manifest themselves empirically through the expenditure distribution x_j^θ . As a result, we do not impose that the observed allocation is inefficient: the efficiency of the observed allocation depends on whether the distribution of expenditures lines up with condition (18) in Proposition 1. It could be that the empirical relationship between expenditures, wages and employment is not far from (18), in which case our implementation of the planner's problem would predict small welfare gains from implementing optimal policies.

We also highlight that, by following this approach, we ensure that the model is initially parametrized in a way that exactly matches the observed outcomes enumerated in items i) and ii) of the proposition. In particular, the fundamentals $\{Z_j^\theta, A_j^\theta\}$, TFP shifters $\{z_j^Y, z_j^H\}$, and bilateral trade costs $\{d_{ij}\}$ are chosen such that the equilibrium from Definition (1) generates these outcomes.

5 Data and Calibration

To take the model to the data, we use as an empirical setting the distribution of economic activity across Metropolitan Statistical Areas (MSAs) in the United States in the year 2007. We focus on the spatial allocation of two skill groups, high skill (college) and low skill (non college) workers. In section 5.1 we first describe the data sources used to fulfill the data requirements in points i) and ii) of Proposition 4. Then, in Section 5.2 we describe the strategy to set the parameters required in points iii) and iv) of the proposition. The broad steps are described here while more details are given in Appendix B. We conclude with basic stylized facts that set the stage for the quantitative analysis in the next section.

5.1 Data

As established in point i) of Proposition 4, we need data on income and expenditures by group and MSA. To that end, we rely on the BEA’s Regional Accounts, which reports labor income, capital income and welfare transfers by MSA. A complementary BEA dataset for the years 2000 to 2007 reports total taxes paid by individuals and MSA (Dunbar, 2009). Taken together, these sources give us a dataset at the MSA level. We then split each of these MSA-level totals into two labor groups: high skill, defined as workers who have completed at least four years of college, and low skill, defined as everyone else. To compute these shares, we use data from the March supplement of the Current Population Survey (IPUMS-CPS, Flood et al. (2017)) collected by Census. We compute the shares of labor income, capital income, taxes and transfers that pertains to each group in each MSA and use these shares to apportion the total at the MSA level reported by the BEA into two groups. Since the March CPS supplement is not available for the smallest MSAs, our dataset covers 245 MSAs in the continental US.³⁰

An important concern when measuring these variables is that, in the model, there is no heterogeneity across individuals within each group θ (high skill and low skill workers) whereas in reality these groups are heterogeneous across cities. If we did not control for this heterogeneity, our procedure to implement the model would interpret the observed variation in net individual transfers across MSA’s within a group as place-based transfers, when they reflect, in part, differences in the types of workers that locate in different MSA’s. In principle, this concern can be mitigated by allowing for several groups corresponding to the fine individual characteristics observed in the CPS. 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, and transfers by skill and MSA from compositional effects using a detailed set of socio-demographic controls (age, detailed educational attainment, sectoral composition, racial composition, and labor force participation) measured from Census data (IPUMS). In the quantification we then use measures of income, expenditures, taxes and transfers that are net of variation in composition within groups across MSA’s. We discuss the details of this step in Appendix B.

We use the variables above to construct expenditure per capita, x_i^θ , using its definition (15) 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 (15), implying no variation in capital income across cities for each type. Therefore we construct expenditure as labor income net of taxes and transfers, plus a group-specific measure of capital income constructed from each skill group’s share of national capital income according to the BEA/CPS data.³¹

As implied by ii) of Proposition 4, quantifying the model also requires data on trade flows between MSAs. The Commodity Flow Survey (CFS) reports the flow of manufacturing goods

³⁰These areas correspond to 97% of the population and 98% of Income of all US urban areas. Urban areas in the US in turn cover 83% of the population, and 87% of personal income. In the implementation we limit the analysis to the 218 MSAs studied in Diamond (2016) for which housing supply elasticities are available.

³¹In our data, 59% of non-labor income is owned by high skill workers. This step involves setting a national share of profits in GDP consistent with the general equilibrium of the model. See Appendix B for details.

shipped between CFS zones in the US every five years. The CFS zones correspond to geographic units that are larger than our unit of observation, the MSA. To overcome this data limitation, we adapt the approach in Allen and Arkolakis (2014) and Monte et al. (2015), who use estimates of trade frictions as function of geography to project CFS-level flows to the MSA level. In our context, we use the aggregate gravity equation predicted by the model to find the unique estimates of trade flows between MSAs that are consistent with actual distance between MSAs, estimates of trade frictions with respect to distance taken from Allen and Arkolakis (2014), and observed trade imbalances. Imbalances are computed comparing income in the traded sector to expenditure on traded goods (for both final and intermediate use) in each MSA, and are matched exactly by this procedure.

Finally, to calibrate the labor shares in production in part iii) of Proposition 4, we use CPS data on employment in traded and non-traded sectors by MSA.³² We also adjust this measure to remove variation coming from compositional effects, following a similar approach to the one described above for the income, expenditure, and transfers variables.

5.2 Calibration with Heterogeneous Workers

Our model is consistent with Diamond (2016) and generates the estimating equations used in her analysis. We use the same definition of geographic units (MSA) and skill groups (College and Non College), and we rely on similar data sources for quantification. Therefore, estimates in Diamond (2016) constitute a natural benchmark for our quantification. In what follows, we discuss the benchmark elasticities and caveats to keep in mind for sensitivity analysis.

Utility and Production Function Parameters $\{\alpha_C, \sigma, \rho, b_{Y,j}^I, b_{H,j}^I, d_{H,j}\}$ We use 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 (35)) for each MSA, and of the elasticity of substitution between high and low skill, estimated at 1.6 and implying $\rho = 0.392$.

We calibrate the Cobb-Douglas share of intermediate use in traded good production ($b_{Y,j}^I = 0.468$ for all j in (34)) using the share of material intermediates in all private good industries production in 2007 from the U.S. KLEMS data. The calibrated share of intermediates used in non-traded production in each city ($b_{H,j}$ in (35)) is chosen to match the share of workers in the non-traded sector of each MSA, as detailed in Section B.2. We assume the same elasticity of substitution between traded varieties in (33) of $\sigma = 9$ as Allen and Arkolakis (2014).

Productivity Spillovers $\gamma_{\theta',\theta}^P$ Diamond (2016) reports estimates for the causal effect of MSA population by skill on wages. She estimates an elasticity of wage by skill group with respect to the population of each skill group. As we discuss in Appendix 4.2, these estimates can be mapped to our $\gamma_{\theta,\theta'}^P$ parameters using the wage equation (13) and the parameter ρ reported above. An issue

³²We use the following NAICS sectors: retail, real estate, construction, education, health, entertainment, hotels and restaurants.

raised by this approach is that these elasticities are large compared to existing references estimating city-level agglomeration elasticities such as Ciccone and Hall (1996), Combes et al. (2008), or Kline and Moretti (2014a).³³ As also detailed in Appendix 4.2, the agglomeration coefficient estimated in these studies would correspond, in data generated by a model with our supply-side structure and heterogeneous workers, to $\overline{s^{W,S}}(\gamma_{SS} + \gamma_{US}) + (1 - \overline{s^{W,S}})(\gamma_{UU} + \gamma_{SU})$, where $\overline{s^{W,S}}$ is average share of high skill workers in total wages across cities. Hence, for our benchmark case, we scale uniformly the previous values of $\gamma_{\theta,\theta'}^P$ to match a value of 0.06 from Ciccone and Hall (1996), which represents a common agglomeration elasticity used in the literature. The resulting elasticities are $(\gamma_{UU}^P, \gamma_{US}^P, \gamma_{SU}^P, \gamma_{SS}^P) = (0.003, 0.02, 0.044, 0.053)$. In robustness checks, we also re-scale $\gamma_{\theta,\theta'}^P$ to match the larger value of 0.2 from Kline and Moretti (2014a). In sum, this way of setting the efficiency spillover elasticities allows us to preserve both a city-level elasticity that is consistent with previous estimates from the literature and relative elasticities across skill types that are consistent with Diamond (2016).

Amenity Spillovers $\gamma_{\theta',\theta}^A$ Diamond (2016) estimates elasticities of labor supply by skill group with respect to real wages, as well as an endogenous amenity index of the form $(L_j^S/L_j^U)^{\gamma^a}$ where γ^a captures the elasticity of MSA level amenities with respect to the relative supply of high skilled workers. As we detail in Appendix 4.2, we can directly map her estimates of these elasticities to the amenity spillovers $\gamma_{\theta',\theta}^A$ in our notation using the labor-supply equation implied by our framework. As a result we obtain $(\gamma_{S,S}^A, \gamma_{U,S}^A, \gamma_{U,U}^A, \gamma_{S,U}^A) = (0.77, -1.24, -0.43, 0.18)$ as benchmark for our quantification. We also inspect the implications of scaling down these elasticities or eliminating the cross-spillover elasticities while keeping their overall level constant. It is worth noting that, if our model was specified with idiosyncratic preference draws across MSA's within types discussed in Section 4.1, the welfare-relevant coefficient $\tilde{\gamma}_{\theta,\theta}^A$ would be parametrized to the same level as $\gamma_{\theta,\theta}^A$ is parametrized in a specification without preference draws. Therefore, given the definition of the planner's problem discussed in Section 4.1, whether or not idiosyncratic preference draws are included in our model does not matter for the quantification, once we condition on the moments from the data reported by Diamond (2016).

5.3 Calibration with Homogeneous Workers

To tease out the impact of accounting for heterogeneous workers types, we start the quantitative section by implementing a model that assumes a single type of worker. Proposition 4 establishes the data requirements of that restricted model as well. To implement that version of the model, we use the aggregate MSA-level variables constructed above. To determine the spillover elasticities, we compute one-group elasticities that correspond to the two-groups spillovers calibrated above, in the following sense. We ask: what elasticities (γ^A, γ^P) of the single-group case would be estimated through the lens of the labor supply and demand equations of the single-group model, if one were to

³³These studies find city-level elasticities γ^P typically in the range of (0.02, 0.2). See Rosenthal and Strange (2004) for a review.

use an MSA-level dataset generated by the model with heterogeneous groups and elasticities $\gamma_{\theta',\theta}^P$ and $\gamma_{\theta',\theta}^P$ calibrated above? Given how we scaled the efficiency spillovers $\gamma_{\theta',\theta}^P$ above, this procedure by construction delivers γ^P equal to the value chosen in the normalization of the efficiency spillovers described above (i.e., 0.06 under Ciccone and Hall, 1996, or 0.2 under Kline and Moretti, 2014a). For amenity spillovers, as detailed in the last subsection of Appendix B.2, this step delivers an aggregate elasticity of

$$\gamma^A = \frac{1 + \overline{s^{L,S}} \left(\gamma_{U,S}^A / \gamma_{S,S}^A \right) + \left(1 - \overline{s^{L,S}} \right) \left(\gamma_{S,U}^A / \gamma_{U,U}^A \right)}{\overline{s^{L,S}} / \gamma_{S,S}^A + \left(1 - \overline{s^{L,S}} \right) / \gamma_{U,U}^A}, \quad (41)$$

where $\overline{s^{L,S}}$ is the average share of skill workers in the population across cities. As a result of this step we obtain $\gamma^A = -0.19$.

5.4 Stylized Facts

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

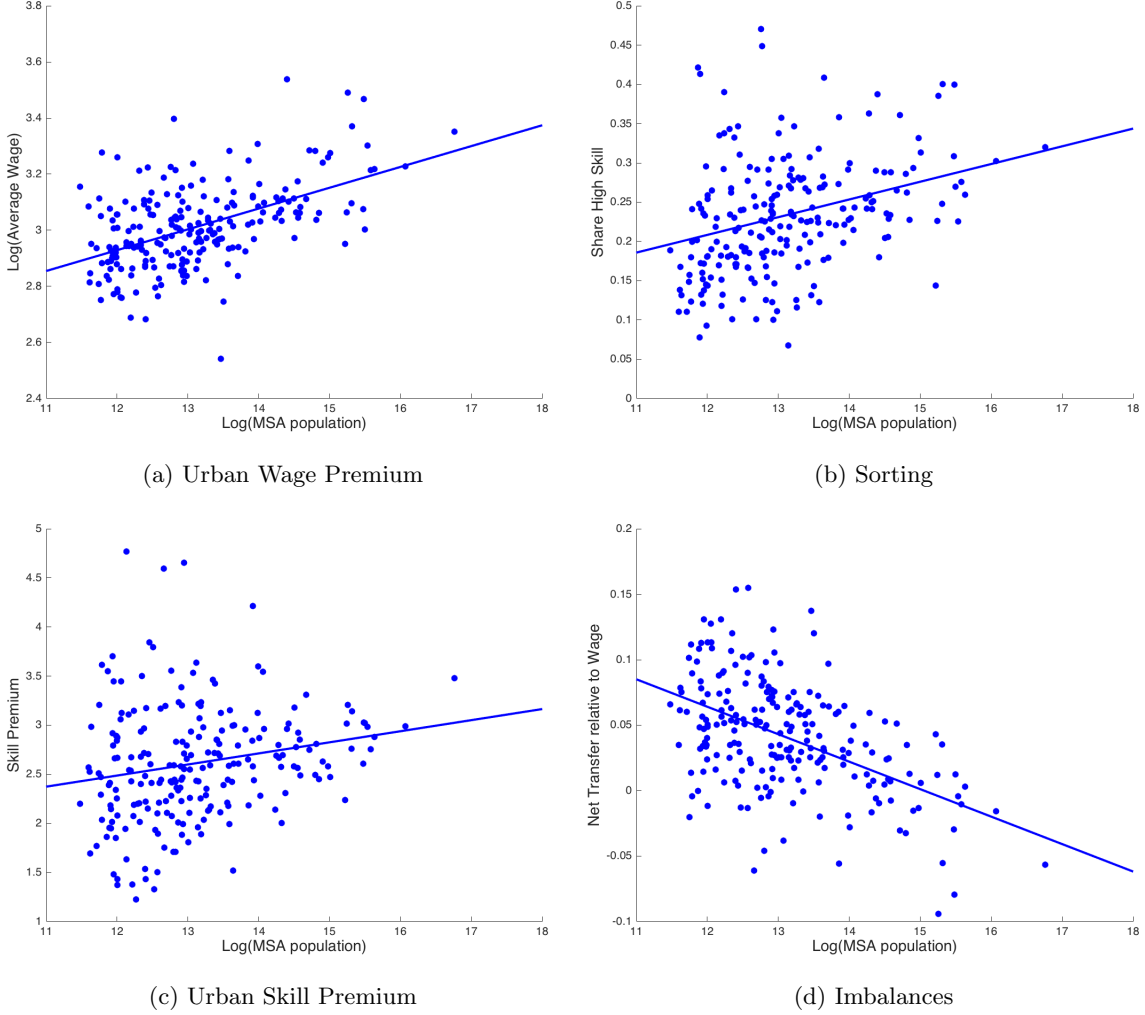
Panels A to C show the standard facts about spatial disparities and sorting. Panel A documents the urban wage premium, defined as the increase in average nominal wages with city size. The elasticity of wages to city size is 7.5%. Panel B shows spatial sorting, in terms of the share of high-skill workers. The semi-elasticity of the share of high skill workers with respect to city size is 2.3%. I.e., doubling population increases the skill share by 2.3 percentage points.³⁴ Panel C shows the urban skill premium, defined as the increase in the ratio of high- to low-skill wage with city size. The slope of 0.11 means that larger cities feature a more unequal nominal wage distribution. The first fact suggests differences in productivity and cost of living across cities, while the last two suggest complementarities between city size and skill.

Panel D shows a somewhat less known fact, the relationship between city size and net imbalances. For each city we construct the absolute value of the net imbalance as the difference between expenditures and total income (from labor and non-labor sources). The graph shows net imbalance relative to labor income at the MSA level across MSA's. We note that, given our construction of the expenditure variable, these differences in imbalances across cities purely result from: i) sorting by skill with differences in capital income by skill; and ii) government policies (taxes and transfers). The first force leads to a positive slope because high-skill workers earn more non-labor income, sort into larger cities. The overall negative slope reflects that government policies redistribute income from larger, high wage, high skill cities to smaller, low wage, low skill cities.³⁵

³⁴This spatial sorting has been increasing over time. Cities with initially higher shares of high skill workers have seen that increase. This trend, often referred to as “Great Divergence”, is at the heart of the analyses of Moretti (2013) and Diamond (2016) on real wage inequality.

³⁵We note that these transfers are net of compositional effects according to detailed demographic characteristics in IPUMS, as mentioned above. Therefore, distributive government distribution policies that vary with these

Figure 1: Urban Premia



Note: each figure shows data across MSA's. All the city level outcomes reported on the vertical axis are adjusted by socio-demographic characteristics of each city, as detailed in Appendix B.1. The slope (SE) in each panel is: Panel (a): 0.074 (0.008); Panel (b): 0.023 (0.005); Panel (c): 0.113 (0.039); Panel (d): -0.02 (0.002).

6 Optimal Spatial Policies in the U.S. Economy

We now explore quantitatively the predictions of our model and discuss the optimal spatial policies that come out of the quantified model. To tease out the role of skill heterogeneity in driving results, we first solve for the optimal allocation in a version of our model with homogeneous workers and we then move to the case with skilled and unskilled workers.

characteristics across individuals are not driving these patterns.

6.1 Homogeneous Workers

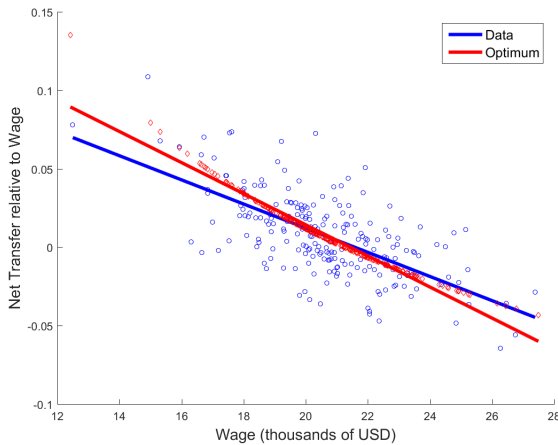
We solve the planner’s problem in relative changes described in Section 4.2 assuming that the economy is populated by one homogeneous group of workers, following the calibration steps described in Section 5.3. In this case the welfare gains from implementing the optimal allocation are negligible, equal to 0.05%. Hence, through the lens of a model with homogeneous workers, the spatial allocation in the U.S. economy seems to be close to efficient in terms of aggregate welfare. The first panel of Figure 2 shows transfers relative wages across MSA’s in the optimal allocation and in the data (corresponding to Panel D from Figure 1). As implied by our discussion of optimal transfers with homogeneous workers in Section 4.1, the optimal net transfers vary one-to-one with the wage.³⁶ Empirically, the relationship between net transfers and wages across cities is quite close to the optimal relationship. However, the optimal transfers induce non-trivial labor reallocations, as about 2% of the workforce changes location relative to the data. The right panel of Figure 2 summarizes this reallocation, by showing changes in city size as function of initial city size.

Why are the gains from optimal spatial policies with homogeneous workers so tiny? The answer may lie in the calibrated spillover elasticities or in the observed spatial distribution. We show next that these small gains are a feature of the particular data that was fed into the quantification, rather than a feature of elasticities.

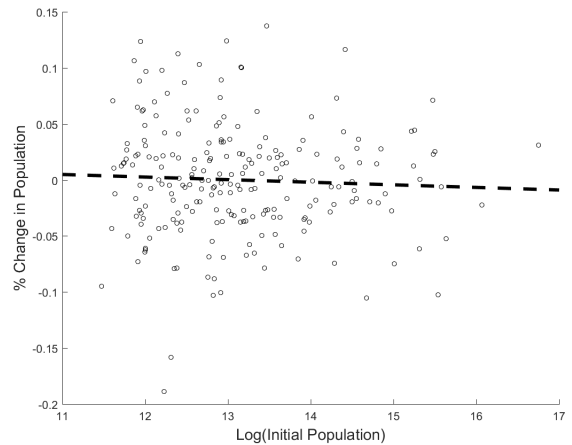
First, the small gains remain under a range of alternative values for the spillover elasticities, given the observed data. When we vary the productivity and amenity spillover elasticities (γ^A, γ^P) over a range from 0 to 4 times their baseline calibrated values, we find that the welfare gains from optimal reallocations are always below 0.2%. Second, the gains increase by up to an order of magnitude if we perturb the transfers per capita across MSA’s entering the quantification. We randomly modify the difference between expenditure and wages across cities (i.e., the ratio

Figure 2: Change in City Size and in Per Capita Transfers in Optimal Allocation

(a) Per Capita Transfers by MSA, Data and Optimal Allocation



(b) Change in Total Population



³⁶I.e., in the light of that previous discussion, the optimal transfer is: $\frac{x_j - w_j}{w_j} = \frac{1 + \gamma^P}{1 - \gamma^A} - \frac{E}{1 - \gamma^A} \frac{1}{w_j}$.

$(x_j - w_j) / w_j$), keeping the mean and variance of these rates at, at most, the level in the data. We then compute the welfare gains from optimal reallocations starting from this distorted equilibrium with random (but small) transfers. In these cases, the welfare gains are always in the ballpark of 0.5% or less.

The previous results suggest that, at the given observed distribution of wages, modifying elasticities or eliminating transfers could lead to somewhat larger, but still small gains. However, when we perturb the data in other dimensions, the inferred gains can be much larger than in the baseline parametrization. We simulate counterfactual market equilibria for random values of equilibrium wages and population, each corresponding to a specific *laissez-faire* equilibrium (without government policies) under counterfactual vectors of exogenous fundamentals (productivity and amenity shifters $\{Z_j, A_j\}$). In particular, we simulate equilibria where the wages and city size distributions exhibit either more or less variance than in the observed U.S. data equilibrium, as well as stronger or weaker covariance patterns. Figure 3 plots the results. The calibrated U.S. economy lies at the point (1,1), where the welfare gains are close to zero. The figure shows that welfare gains from optimal policies, starting from a *laissez-faire* equilibrium without redistribution, could be substantial. In particular, they are large if the equilibrium features either a larger variance of wages across cities relative to the data, or stronger (either positive or negative) correlation between wages and city size.

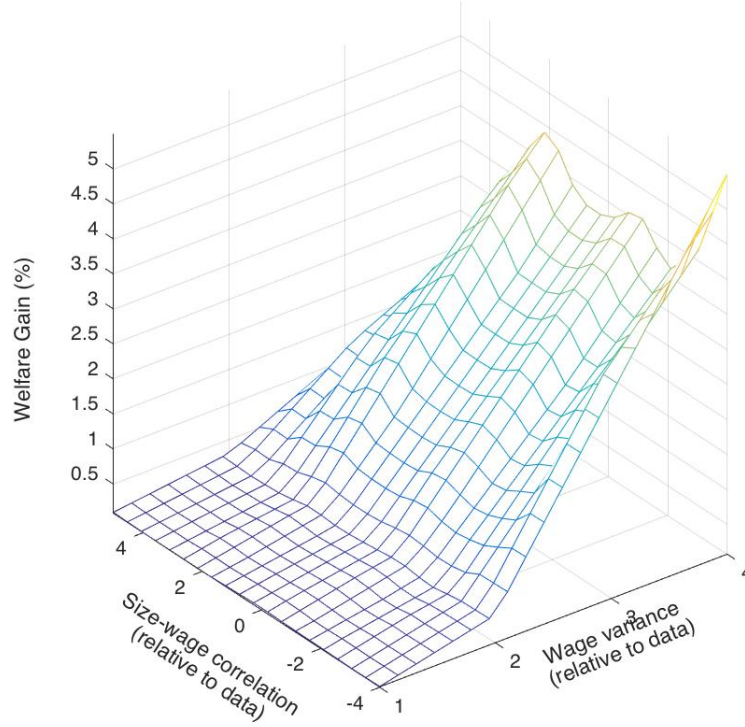
We conclude that, under homogeneous workers, the model does not deliver small gains by design. Instead, through the lens of a model with homogeneous workers, the data suggests that the U.S. transfers are close to efficient.

6.2 Heterogenous Workers

We now move on to the quantitative implications of optimal spatial policies in the context of the full model that features two groups of workers, skilled (S) and unskilled (U). Here again we solve the planner’s problem in changes relative to the observed equilibrium. Specifically, we solve the problem of maximizing the change in utility of skilled workers, $\hat{u}^S$, subject to a lower bound for the change in utility of unskilled worker, $\hat{u}^U$. Varying this lower bound allows to trace out the Pareto frontier.

Figure 4 shows the utility frontier of the U.S. economy, expressed in changes relative to the observed equilibrium. The point (1,1) corresponds to allocations where the welfare of skilled and unskilled workers is unchanged compared to the calibrated equilibrium. In contrast to the one-group case, we now find sizable welfare gains from the optional spatial reallocation. Three points are highlighted on the frontier, corresponding to different social weights for unskilled and skilled workers in the planner’s optimization problem. When the welfare gain of unskilled and skilled workers is restricted to be the same, optimal transfers lead to a 6.1% welfare gain for 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 14.5% for high skill workers and of 10.1% for low skill workers. In what follows, we first describe key differences between the optimal and the data equilibrium.

Figure 3: Welfare Gains from Optimal Reallocation from Different Initial Equilibria



Note: We simulate a range of laissez-faire equilibria with no government transfers in which the joint distribution of wages and city sizes differs from the data in terms of the variance of the wage distribution and the correlation between wages and city sizes. In these equilibria, the distribution of city sizes has the same variance as in the data. Correlation and variances are reported in relative terms compared to the data. For each variance-correlation combination we draw 20 random distributions of wages and city sizes, and report the mean welfare gains across these simulations.

Then we analyze the drivers of the stark difference between the optimal welfare change under the one-group and two-groups calibration.

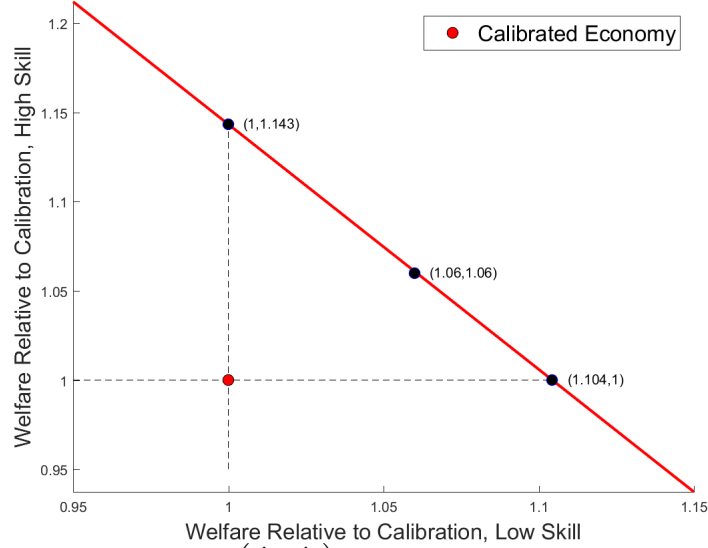
Actual versus Optimal Transfers The optimal allocation is implemented through net transfers by worker type and city. How does the optimal spatial redistribution implied by the model compare to the data? To guide the answer to this question we can return to the optimal policies. By definition, from (15), transfers per capita are the difference between expenditure and income:

$$t_j^\theta = x_j^\theta - (w_j^\theta + b^\theta \Pi). \quad (42)$$

Letting $(\tau^\theta(\cdot), T^\theta)$ be the optimal policies derived from Proposition 2, the optimal transfer received by type θ worker in city j are $t_j^\theta \equiv \tau^\theta(\cdot) w_j^\theta + T^\theta$.

Figure 5 shows a scatter plot of the net transfers per capita relative to wages t_j^θ/w_j^θ by MSA and worker type on the vertical axis against the average wage w_j^θ in both the data (blue circles) and the optimal allocation (red diamonds), for low skill workers (hollow markers) and high skill workers

Figure 4: 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 A.5. Each point corresponds to a maximization of $\hat{u}^H$ subject to a different lower bound on $\hat{u}^L$.

(solid markers).³⁷ If the observed allocation was efficient, the data and the optimum should be the same up to a scale on the vertical axis.

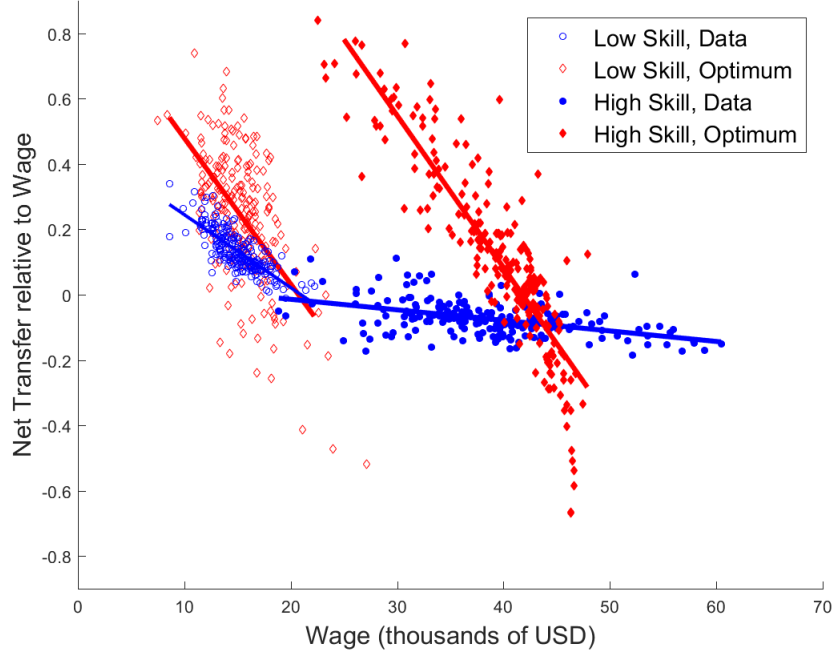
Consider first the transfers in the data. This figure illustrates a clear pattern across groups and city sizes: redistribution flows from high skill workers and large cities towards low skill workers and small cities. Net average transfers are positive for low skill workers in most MSA's, and on average across MSA's, they equal 1.9 thousand dollars or 13% of the average low-skill wage. For high skill workers the net transfers are negative. The corresponding numbers are -3 thousand dollars or -7% of the average wage. Within skill groups, net transfers decrease with the average wage of the MSA, and become negative for skilled workers in MSA's with sufficiently high wages. In cities where high skill workers earn on average more than \$50k per year, net transfers of for skill workers are -6.9 thousand dollars or -12% of wages.

The observations in red show the efficient allocation. By construction, this relationship satisfies the optimality condition from Proposition 1. Across cities, the optimal transfers relative to labor income decrease more steeply with wages than in the data for both labor types, implying a stronger redistribution towards low-wage cities than what is observed empirically. Moreover, the increase in the implicit tax rate is sharper for high-skill workers, as seen from the greater change in the slope of transfers for the high-skilled.

To understand the forces behind these patterns, we note that the slope is partly driven by

³⁷The optimal allocation represented in the figure corresponds to the point on the Pareto frontier where welfare gains are equal for both types of workers. In Section A.5 we show that these patterns are robust to picking efficient allocations in other points of the utility frontier.

Figure 5: Per Capita Transfers by Skill Level and MSA, Data and Optimal Allocation



Note: each point in the figure corresponds to an MSA-skill group combination. The vertical axis show the difference between average expenditure and wage and the horizontal axis shows the average wage. For details of how the data is constructed see Appendix B. The slopes of each linear fit (with SE) are: Low Skill, Data: -0.02 (0.001); Low Skill, Optimum: -0.04 (0.004); High Skill, Data: -0.003 (0.0004); High Skill, Optimum: -0.05 (0.001). The figure corresponds to planner's weights such that both types of workers experience the same welfare gain (i.e., point (1.061,1.061) in Figure 4).

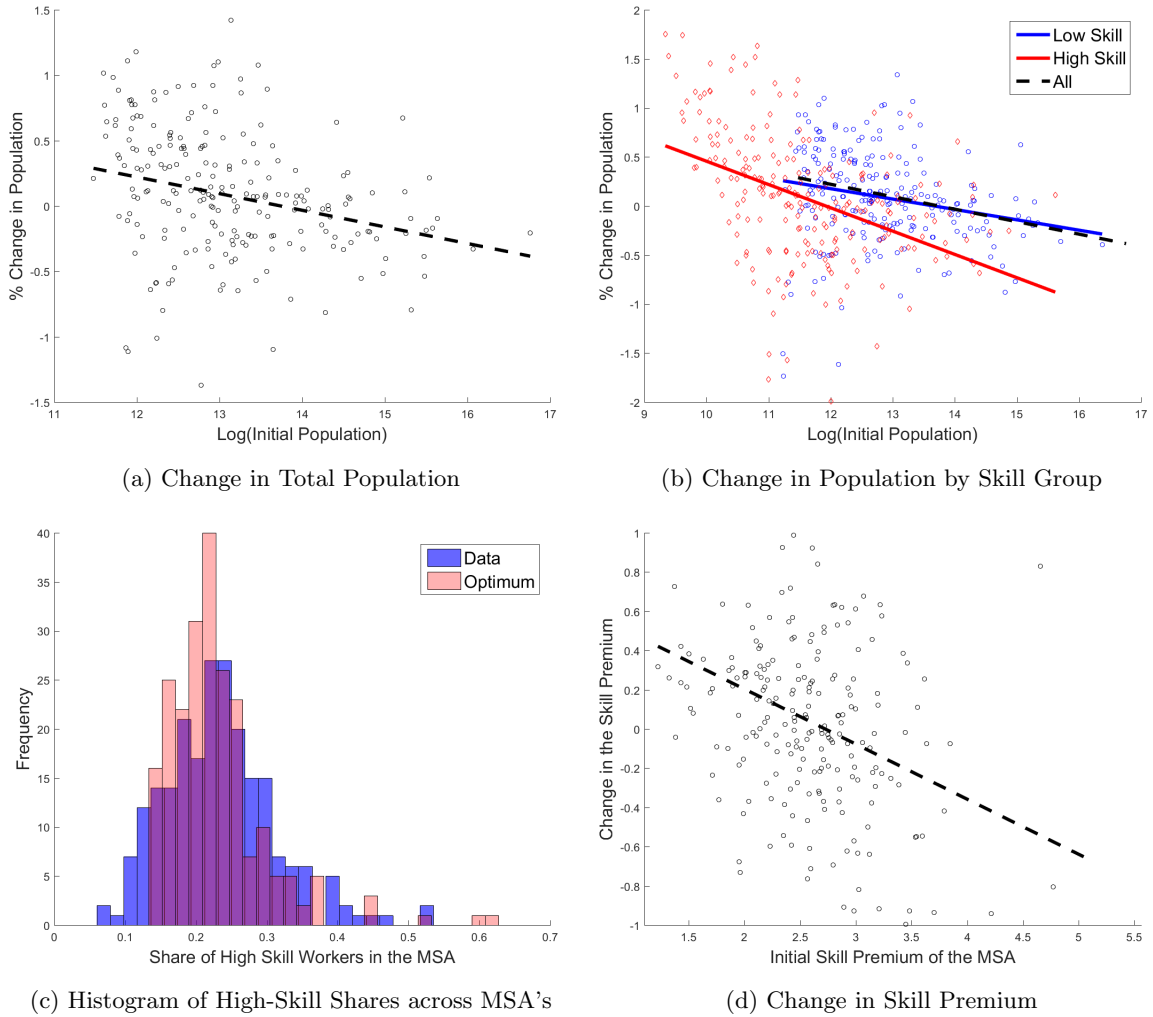
the optimal subsidies $\tau^\theta(\cdot)$ defined in equation (21). The term $\frac{\gamma_{\theta,\theta}^P + \gamma_{\theta,\theta}^A}{1 - \gamma_{\theta,\theta}^A}$ would be the optimal tax if there were no spillovers across types. The dispersion in optimal transfers around a trend is exclusively due to the spillover across types, i.e. the remaining terms in (21). This term also drives part of the slope, depending on the sign of the cross-spillovers. For low-skill workers, the negative own spillovers alone imply a tax, $\frac{\gamma_{U,U}^P + \gamma_{U,U}^A}{1 - \gamma_{U,U}^A} < 0$, although not big enough to generate the net transfers in the optimal allocation. The larger implicit optimal tax is driven by the negative cross spillovers through amenities ($\gamma_{U,S}^A < 0$). Because a larger concentration of group U generates negative spillovers on group S , this channel implies a larger implicit tax on low skill workers in cities where a larger of share of consumption accrues to high skill workers. These places are the large, high-wage cities in both the data and the optimal allocation.

For high skill workers, the role of own- and cross-spillovers is quite different than for low skill workers. High skill workers generate positive own spillovers, $\frac{\gamma_{S,S}^P + \gamma_{S,S}^A}{1 - \gamma_{S,S}^A} > 0$. Absent spillovers on low-skill workers, this force would imply a large proportional labor income *subsidy* for high skill workers. However, this force is more than offset by positive cross spillovers. Because a larger concentration of group S generates positive spillovers on group U , the cross spillovers $\gamma_{S,U}^P$ and $\gamma_{S,U}^A$ lead to a larger optimal subsidy to high skill workers in places with a high share of low skill

workers. These places are the small, low-wage cities in both the data and the optimal allocation. Specifically, high skill workers are more subsidized (or less taxed) where the spending of low skill workers relative to the wage bill of high skill workers, $\frac{x_j^U L_j^U}{w_j^S L_j^S}$, is larger.

Optimal Reallocation and Sorting The optimal transfers change the spatial distribution of economic activity compared to the data. By changing the location incentives of workers, they affect spatial sorting and the city size distribution. These reallocations in turn impact the spatial distributions of labor productivity and wages through agglomeration spillovers, and the distribution

Figure 6: Changes in Population, Skill Shares, and Skill Premium across MSA's



Note: Panel (a) shows the change in population between the optimal allocation and the initially observed equilibrium and the linear fit. Slope (SE): -0.12 (0.03); $R^2=0.06$. Panel (b) displays the same outcomes for high and low skill workers. Slopes (with SE): High Skill: -0.24 (0.04); Low Skill: -0.10 (0.03). Panel (d) displays in the vertical axis the difference in the skill premium between the optimal and initial allocation. Slope (SE): -0.27 (0.06). The figures correspond to planner's weights such that both types of workers experience the same welfare gain (i.e., point (1.061,1.061) in Figure 4).

of urban amenities through amenity spillovers. These effects in turn feed back to location choices, changing the spatial pattern of skill premia and inequality. Our goal here is to describe the spatial equilibrium that results from this process and contrast with the facts on spatial disparities discussed in Section 5.

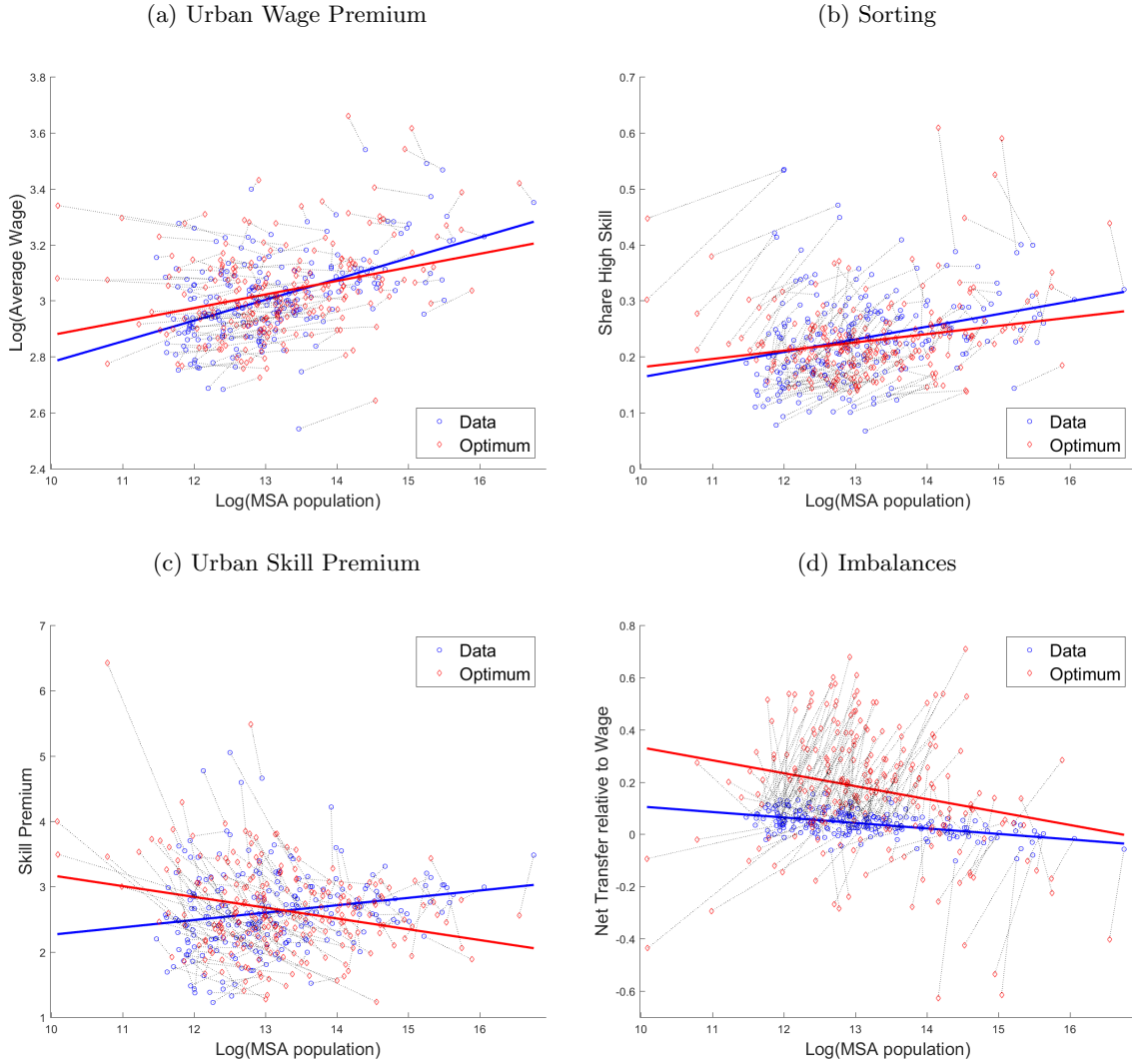
Figure 6 shows the pattern of reallocation. Panel (a) shows the initial total population of each MSA on the horizontal axis and the percentage change in population implied by the optimal allocation on the vertical axis. The stronger redistribution to low-wage locations discussed in the previous section 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 some middle- and small-MSA's that shrink. The R^2 of the linear regression is only 4%, implying that, due to the multiple sources of spatial heterogeneity and interactions, initial city size is a poor predictor of whether a city is too large or too small in the observed allocation. Panel (b) shows the same figure within groups skill, including the linear fit from the left panel for reference. The reallocations towards initially smaller places is sharper for high skill workers than for low skill workers. This is consistent with the initially observed allocation featuring too little mixing of skills due to the spillovers across types.

These differential patterns of reallocation by skill lead to less segregation of skills across cities. Panel (c) shows the histogram of skill shares across MSA's in the initial and optimal allocation, implying a concentration of skill shares across MSA's. An additional implication of these reallocations, shown in panel (d), is that, within-city wage inequality, in terms of the skill premium, tends to decrease in cities that are initially more unequal. Therefore, the optimal spatial transfers reduces inequality within cities that are more unequal in the observed allocation.

The Urban Premia in the Optimal Allocation We now revisit the stylized facts about urban premia mentioned in Section 5 in the optimal allocation vis-à-vis the data. Figure 7 reproduces Figure 1 in the data and the optimal allocation. Each pair of linked observations corresponds to the same MSA. Before discussing each of these figures, it is worth highlighting that in all of them there is a fair amount of heterogeneity in terms of how cities change between the observed and the optimal allocation: conditioning on any initial size, some cities shrink and some expand; in addition, an optimal city expansion is not necessarily associated with an increase in the average wage, skill premium, share of skill workers, or net transfers. These rich responses result from the different dimensions of heterogeneity captured by the model. We now focus on the average response of the various margins, and on how it impacts the patterns of urban premia observed in the observed equilibrium.

We start with Panel D of Figure 7, which summarizes the relationship between city size and imbalances relative to wages. As expected given the previous discussion, redistribution to smaller MSA's is stronger in the optimal allocation than in the data. In the optimal allocation the absolute value of the imbalances is larger, pointing to an inefficiently low overall amount of transfers in the data.

Figure 7: Urban Premia, Data and Optimal Allocation



Note: each panel reports outcomes across MSA's in the data and in the optimal allocation. Each linked pair of observations corresponds to the same MSA. The slope (SE) in each panel is: Panel (a): Data 0.074 (0.008), Opt 0.045 (0.01); Panel (b): Data 0.023 (0.005), Opt 0.015 (0.004) ; Panel (c): Data 0.113 (0.039), Opt -0.17 (0.040); Panel (d): Data -0.034 (0.004), Opt -0.07 (0.02).

We turn to differences between high skill and low skill workers. Panel B shows that spatial sorting is somewhat less pronounced, on average, in the optimal equilibrium compared to the data. That is, the optimal allocation features more mixing of high and low skills in large cities. This dampening of sorting is a result of the larger increase in implicit taxes for high skill workers than for low skill workers in Figure 5 discussed above. Associated with this dampened sorting there is a reduction in the urban skill wage premium, as shown in Panel C. The lack of an urban skill premium reflects the reduction in efficiency spillovers among high skill workers in larger cities. Overall, these results imply that, in the optimal allocation, the sorting patterns end up being detached from the urban skill premium. This suggests that high skill sorting into larger cities ends up being driven

by differential preferences for urban amenities (both endogenous and exogenous).

Finally, as Panel A of Figure 7 indicates, there is still a noticeable wage premium in the large cities, although it is also tempered relative to the data. Given the lack of an urban skill premium, this urban wage premium 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 cities.

6.3 Contrast of the Results under Homogeneous and Heterogeneous Workers

Why are the welfare gains so different with multiple worker types relative to a single worker type? To answer this question, it is worth noting the quantification with heterogeneous workers would deliver the exact same MSA-level outcomes and welfare gains (common to all workers) as the quantification with homogeneous workers if, under heterogeneous workers: (i) the data exhibited no spatial sorting by skill (the share of workers of type θ is constant over space), no urban skill premium (the skill wage premium $\frac{w_j^S}{w_j^U}$ is constant over space), and no relative differences in expenditures (the expenditure ratio $\frac{x_i^S}{x_i^U}$ is the same across space); (ii) high and low skill workers were perfectly substitutable in production ($\rho = 1$); and (iii) spillovers across types were absent, with own spillovers equal the calibration in the one-group case, i.e.: $\gamma_{\theta,\theta'}^j = 0$ if $\theta \neq \theta'$ and $\gamma_{\theta,\theta}^j = \gamma^j$ for $j = A, P$.

This benchmark is a useful reference to inspect why, given the U.S. data, the results are so different under a homogeneous and heterogeneous workers. We can investigate the role played by different forces by releasing each of the restrictions (i)-(iii) above. As we do so, we move from a “one-group calibration” of the model with heterogeneous workers (with no sorting in the data, elasticity of substitution equal to 1, and no cross-spillovers) to the full two-groups case calibration we have previously implemented (where all these forces are included).³⁸

Table 1: Welfare gains under different model specifications

	Data	Elasticity ρ	Cross Spillovers?	Welfare Gain (%)
(1)	No Sorting	1	No	0.05
(2)	No Sorting	1	Yes	1.69
(3)	No Sorting	0.39	Yes	0.79
(4)	Sorting	1	No	0.17
(5)	Sorting	1	Yes	7.57
(6)	Sorting	0.39	Yes	6.10

Table 1 reports the results. Each of the columns allows for a different relaxation of (i), (ii), or (iii). The first row corresponds to the welfare gain under homogeneous workers reported in Section 6.1, and the last row corresponds to the welfare gains from model with heterogeneous workers in Section 6.2, when the gains are common to all workers.

³⁸We note that what we have just called the “one-group calibration” is not the calibration of the model with homogeneous workers, but rather a calibration of a model with heterogeneous workers which, by design, delivers the same welfare effects for all workers and MSA-level reallocation as the model with homogeneous workers.

Starting from a case with perfect substitution across workers, bringing in cross spillovers considerably increases the welfare gains, regardless of whether there is sorting (i.e., comparing rows 4 and 5) or not (comparing 1 and 2). Therefore, the existence of spillovers across workers plays a key role in driving the results. Sorting also plays a relevant role, but only cross-spillovers are present. In the absence of cross spillovers, including sorting in the data (i.e., comparing rows 1 and 4) raises the welfare gain, but still to a small amount. However, the inclusion of sorting when there are cross spillovers (comparing rows 2 and 5, or 3 and 6) leads to considerably larger gains. We conclude that accounting for heterogeneity in skills is important for the implied benefits of implementing optimal spatial policies, and that spillovers across different types of workers play a substantial role in driving this result.

6.4 Robustness

We consider the robustness of the results with heterogeneous workers presented so far to alternative assumptions on the parameters.

Pareto Weights Throughout our discussion of the reallocation patterns in Section 6.2, we compared the data to an optimal allocation corresponding to the same welfare gains to all workers (i.e., the welfare gains $(\hat{u}^U, \hat{u}^S) = (1.061, 1.061)$ in Figure 4). The patterns of optimal redistribution from Figure 18 are similar under optimal allocations corresponding to other points on the utility frontier in Figure 4. Different planner’s weights are implemented through the lump-sum transfers across types, which turn out to have little effect on the optimal spatial transfers.

To highlight this property, Figure A.1 in Appendix C shows the outcomes as in Figure 5 but assuming planner weights that are either 10 times larger for high skill workers (case labeled “low weight on U”) or 10 times larger for low skill workers (case labeled “high weight on U”). The optimal welfare changes relative to the observed allocation associated with those weights are very different from the benchmark case. However, we find a similar pattern of net transfers as in our previous analysis. The lower is the weight for high skill workers, the lower is the intercept of the optimal transfer schedule for that group and the higher is for low skill workers, but the slope remains similar to the benchmark. Therefore the stronger redistribution to smaller and low-wage cities in the optimal allocation is roughly independent from the point on the utility frontier that we target. As a result, as shown in Figure A.2, the qualitative and quantitative aspects of the reallocation patterns in Figure 7 remain unchanged.

Amenity and Efficiency Spillover Elasticities We now revisit the main results under different values of the spillovers. In particular we consider two alternative parametrizations that assume different levels for the spillover elasticities. Noting that, in absolute value, the amenity $\gamma_{\theta, \theta'}^A$ spillovers are larger than the efficiency spillovers $\gamma_{\theta', \theta}^P$, we revisit our main results assuming that $\gamma_{\theta, \theta'}^A$ is 50% below the benchmark parametrization. As a result, the single-group equivalent amenity spillover constructed in 41 falls to $\gamma^A = 0.085$. Second, rather than normalizing the efficiency spillovers $\gamma_{\theta, \theta'}^P$,

to match an average value of 0.06 from Ciccone and Hall (1996), we normalize them to match an average of 0.2 as in Kline and Moretti (2014a).³⁹ This step amounts to multiplying our benchmark spillover elasticities by a factor of 3.3.

Table 2 shows the welfare when the planner’s weight are such that low and high skill workers benefit similarly from the optimal reallocation. The first row corresponds to benchmark in Figure 4. Lowering the amenity spillover by 50% brings the common welfare gain down to 3.3%, while multiplying the efficiency spillovers by 3.3 increases the gain to 7.3%. Importantly, the qualitative patterns of optimal reallocation and spatial premia discussed in the benchmark are also similarly present under these alternative parametrizations. Figure A.3 replicates the optimal patterns of urban premia from Figure 7 under the benchmark and under the alternative parametrizations. Under low amenity spillovers, the basic qualitative and quantitative patterns are very similar to the benchmark. Under high efficiency spillovers, the patterns are also roughly similar, and we observe an even stronger tempering of the sorting pattern by skill, to the point that the optimal allocation features no sorting by skill into larger cities.

Table 2: Welfare gains under different levels of the spillovers

Spillovers	Welfare Gain (%)
Benchmark	6.1
Lower $\gamma_{\theta'\theta}^A$	3.3
Higher $\gamma_{\theta'\theta}^P$	7.3

The table reports the common welfare gains for skilled and unskilled workers under alternative parametrizations. “Lower $\gamma_{\theta'\theta}^A$ ” corresponds to amenity spillovers that are 50% lower than the benchmark, while “Higher $\gamma_{\theta'\theta}^P$ ” corresponds to efficiency spillovers that are 3.3 times larger than the benchmark.

In sum, implementing our analysis under these alternative parametrizations does not alter the main messages from the quantification, in the sense that optimal spatial policies may require stronger redistribution towards low-wage cities than in the data, reduce wage inequality in larger cities, weaken spatial sorting by skill, and lead to significant welfare gains.

7 Conclusion

In this paper we studied optimal spatial policies in quantitative geography frameworks with spillovers and sorting of heterogeneous workers. We rely on a framework that nests recent strands of quantitative spatial research and includes many key determinants of the spatial distribution of economic activity such as geographic frictions and asymmetric spillovers across heterogeneous workers. We generalize these models to allow, in the market allocation, for arbitrary policies that transfer resources across agents and regions.

We first derived a condition on the joint distribution of expenditures and wages across workers and regions that ensures existence of an efficient competitive allocation. Using this condition we showed how to use data on the observed distribution of wages, expenditures and spillover

³⁹The resulting elasticities are $(\gamma_{UU}^P, \gamma_{US}^P, \gamma_{SU}^P, \gamma_{SS}^P) = (0.011, 0.146, 0.065, 0.176)$.

elasticities to assess the efficiency of the observed spatial allocation. We showed that the optimal expenditure distribution can be implemented by labor taxes or subsidies coupled with lump-sum transfers. Under constant-elasticity spillovers and absent spillovers across workers, constant labor income taxes coupled with lump-sum transfers are distortive and sufficient to implement the efficient allocation. When spillovers across workers are present, spatial efficiency may require place-specific labor subsidies and more mixing of heterogeneous workers across cities than in the competitive allocation.

We showed how to use the main framework to quantify efficient outcomes in the data. In order to deal with issues of multiplicity, we first identified conditions on the distributions of spillover and housing supply elasticities to ensure uniqueness of the competitive allocation under the optimal policies. We also showed that the distributions of fundamentals needed to compute the optimal allocation can be backed out from data on the distribution of wages, employment, and expenditures across worker types and regions.

We applied the model to data on distribution of economic activity across MSA's in the U.S. in the year 2007. Under standard existing estimates of the spillover elasticities, our results suggest too little redistribution to low-wage cities. The net transfer per capita falls more sharply with the average wage across cities in the optimal allocation than in the data. As a result, the spatially efficient allocation implies, on average, a reallocation of workers from large to small cities and features smaller dispersion in city sizes compared to the observed allocation. However, there is also considerable dispersion in the distribution of growth rates for any given initial size, including large cities that grow as the economy moves to the optimum.

The results also suggest that spatial optimality requires a greater degree of mixing of different skills. Relative to the data, the optimal allocation exhibits lower skill wage premium in larger cities and weaker spatial sorting of high skill workers into larger cities. Through the lens of a model with homogeneous workers, our approach delivers negligible gains. However, under heterogeneous workers, the gains are considerably larger. These differences are driven by spillovers across heterogeneous workers. Overall, the results imply that accounting for heterogeneity across skills may be crucial for the design and aggregate welfare effects of optimal spatial policies.

References

- Abdel-Rahman, H. and M. Fujita (1990). Product variety, marshallian externalities, and city sizes. *Journal of regional science* 30(2), 165–183.
- Abdel-Rahman, H. M. and A. Anas (2004). Theories of systems of cities. *Handbook of regional and urban economics* 4, 2293–2339.
- Allen, T. and C. Arkolakis (2014). Trade and the topography of the spatial economy. *Quarterly Journal of Economics* 1085, 1139.
- Allen, T., C. Arkolakis, and X. Li (2015). Optimal city structure. *Yale University, mimeograph*.
- Allen, T., C. Arkolakis, and Y. Takahashi (2014). Universal gravity. Technical report, National Bureau of Economic Research.
- Anderson, J. E. and E. van Wincoop (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review* 93(1), 170–192.

- Arkolakis, C., A. Costinot, and A. Rodríguez-Clare (2012). New trade models, same old gains? *The American Economic Review* 102(1), 94–130.
- Baum-Snow, N. and R. Pavan (2013). Inequality and city size. *Review of Economics and Statistics* 95(5), 1535–1548.
- Behrens, K., G. Duranton, and F. Robert-Nicoud (2014). Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy* 122(3), 507–553.
- Behrens, K., G. Mion, Y. Murata, and J. Südekum (2017). Spatial frictions. *Journal of Urban Economics* 97, 40–70.
- Bhagwati, J. and H. G. Johnson (1960). Notes on some controversies in the theory of international trade. *The Economic Journal* 70(277), 74–93.
- Brandt, L., T. Tombe, and X. Zhu (2013). Factor market distortions across time, space and sectors in China. *Review of Economic Dynamics* 16(1), 39–58.
- Caliendo, L., F. Parro, E. Rossi-Hansberg, and P.-D. Sarte (2014). The impact of regional and sectoral productivity changes on the us economy. Technical report, National Bureau of Economic Research.
- Ciccone, A. and R. E. Hall (1996). Productivity and the density of economic activity. *The American Economic Review*, 54–70.
- Combes, P.-P. (2011). The empirics of economic geography: how to draw policy implications? *Review of World Economics* 147(3), 567–592.
- Combes, P.-P., G. Duranton, and L. Gobillon (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics* 63(2), 723–742.
- Davis, D. R. and J. I. Dingel (2012). A spatial knowledge economy. Technical report, National Bureau of Economic Research.
- Dekle, R., J. Eaton, and S. Kortum (2008). Global rebalancing with gravity: Measuring the burden of adjustment. Technical Report 3, International Monetary Fund.
- Desmet, K. and E. Rossi-Hansberg (2013). Urban accounting and welfare. *American Economic Review* 103(6), 2296–2327.
- Diamond, R. (2016). The determinants and welfare implications of us workers’ diverging location choices by skill: 1980–2000. *The American Economic Review* 106(3), 479–524.
- Dixit, A. (1985). Tax policy in open economies. *Handbook of public economics* 1, 313–374.
- Dunbar, A. E. (2009). Metropolitan area disposable personal income: Methodology and results for 2001–2007.
- Eaton, J. and S. Kortum (2002). Technology, geography, and trade. *Econometrica* 70(5), 1741–1779.
- Eeckhout, J. and N. Guner (2015). Optimal spatial taxation: Are big cities too small?
- Eeckhout, J., R. Pinheiro, and K. Schmidheiny (2014). Spatial sorting. *Journal of Political Economy* 122(3), 554–620.
- Fajgelbaum, P. D., E. Morales, J. C. S. Serrato, and O. M. Zidar (2015). State taxes and spatial misallocation. Technical report, National Bureau of Economic Research.
- Fajgelbaum, P. D. and E. Schaal (2017). Optimal transport networks in spatial equilibrium. Technical report, National Bureau of Economic Research.
- Flatters, F., V. Henderson, and P. Mieszkowski (1974). Public goods, efficiency, and regional fiscal equalization. *Journal of Public Economics* 3(2), 99–112.
- Flood, S., M. King, S. Ruggles, and J. R. Warren (2017). Integrated public use microdata series, current population survey: Version 5.0.[dataset]. minneapolis: University of minnesota, 2017.

- Gaubert, C. (2015). Firm sorting and agglomeration. *University of California, Berkeley*.
- Giannone, E. (2017). Skilled-biased technical change and regional convergence. Technical report, University of Chicago Working Paper, available at: http://home.uchicago.edu/~elisagiannone/files/JMP_ElisaG.pdf.
- Glaeser, E. L., J. D. Gottlieb, et al. (2008). The economics of place-making policies. *Brookings Papers on Economic Activity* 39(1 (Spring)), 155–253.
- Helpman, E. (1998). The size of regions: transport and housing as factors in agglomeration. In D. Pines, E. Sadka, and I. Zilcha (Eds.), *Topics in Public Economics*, pp. 33–54. Cambridge University Press Cambridge.
- Henderson, J. V. (1974). The sizes and types of cities. *The American Economic Review*, 640–656.
- Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and manufacturing TFP in China and India. *Quarterly Journal of Economics* 124(4), 1403–1448.
- Hsieh, C.-T. and E. Moretti (2015). Why do cities matter? local growth and aggregate growth. Technical report, National Bureau of Economic Research.
- Khajavirad, A., J. J. Michalek, and N. V. Sahinidis (2014). Relaxations of factorable functions with convex-transformable intermediates. *Mathematical Programming* 144(1-2), 107–140.
- Kline, P. (2010). Place based policies, heterogeneity, and agglomeration. *The American Economic Review* 100(2), 383–387.
- Kline, P. and E. Moretti (2014a). Local economic development, agglomeration economies and the big push: 100 years of evidence from the Tennessee Valley Authority. *Quarterly Journal of Economics*.
- Kline, P. and E. Moretti (2014b). People, places, and public policy: Some simple welfare economics of local economic development programs. *Annual Review of Economics* 6(1), 629–662.
- Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. *American Economic Review*, 950–959.
- Monte, F., S. J. Redding, and E. Rossi-Hansberg (2015). Commuting, migration and local employment elasticities. Technical report, National Bureau of Economic Research.
- Moretti, E. (2012). *The new geography of jobs*. Houghton Mifflin Harcourt.
- Moretti, E. (2013). Real wage inequality. *American Economic Journal: Applied Economics* 5(1), 65–103.
- Neumark, D., H. Simpson, et al. (2015). Place-based policies. *Handbook of Regional and Urban Economics* 5, 1197–1287.
- Ossa, R. (2015). A quantitative analysis of subsidy competition in the us. Technical report, National Bureau of Economic Research.
- Redding, S. J. (2016). Goods trade, factor mobility and welfare. *Journal of International Economics* 101, 148–167.
- Redding, S. J. and E. A. Rossi-Hansberg (2017). Quantitative spatial economics. *Annual Review of Economics* 9(1).
- Redding, S. J. and M. A. Turner (2015). Transportation costs and the spatial organization of economic activity. *Handbook of Regional and Urban Economics* 5, 1339–1398.
- Roback, J. (1982). Wages, rents, and the quality of life. *Journal of Political Economy*, 1257–1278.
- Rosen, S. (1979). Wage-based indexes of urban quality of life. *Current issues in urban economics* 3, 324–345.
- Rosenthal, S. S. and W. C. Strange (2004). Evidence on the nature and sources of agglomeration economies. *Handbook of regional and urban economics* 4, 2119–2171.
- Sandmo, A. (1975). Optimal taxation in the presence of externalities. *The Swedish Journal of Economics*, 86–98.

A Appendix to Section 4

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

The planning problem described as follows.

Definition 2. *The planning problem is*

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

subject to (i) the spatial mobility constraints

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

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

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

(iii) local and notional labor-market clearing,

$$\begin{aligned} N_j^Y + N_j^H &= N \left(z_1^\theta \left(L_j^1, \dots, L_j^\Theta \right) L_j^1, \dots, z_j^\Theta \left(L_j^1, \dots, L_j^\Theta \right) L_j^\Theta \right) \text{ for all } j; \\ \sum_j L_j^\theta &= L^\theta \text{ for all } \theta; \text{ and} \end{aligned}$$

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

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

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

for all j and θ and some constants $\{E^\theta\}$. If the planner's problem is globally concave and (A.1) 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:

$$\begin{aligned}
\mathcal{L} = & u^\theta \\
& - \sum_j \omega_j^\theta L_j^\theta \left(u^\theta - a_j^\theta \left(L_j^1, \dots, L_j^\theta \right) U \left(c_j^\theta, h_j^\theta \right) \right) \\
& - \sum_{\theta' \neq \theta} \sum_j \omega_j^{\theta'} L_j^{\theta'} \left(u^{\theta'} - a_j^{\theta'} \left(L_j^1, \dots, L_j^\theta \right) U \left(c_j^{\theta'}, h_j^{\theta'} \right) \right) \\
& - \sum_j p_j^* \left(\sum_i d_{ji} Q_{ji} - Y_j \left(N_j^Y, I_j^Y \right) \right) \\
& - \sum_j P_j^* \left(\sum_\theta L_j^\theta c_j^\theta + I_j^Y + I_j^H - Q \left(Q_{1j}, \dots, Q_{Jj} \right) \right) - \sum_j R_j^* \left(\sum_\theta L_j^\theta h_j^\theta - H_j \left(N_j^H, I_j^H \right) \right) \\
& - \sum_j W_j^* \left(N_j^I + N_j^H - N \left(z_j^1 \left(L_j^1, \dots, L_j^\theta \right) L_j^1, \dots, z_j^\theta \left(L_j^1, \dots, L_j^\theta \right) L_j^\theta \right) \right) \\
& - \sum_\theta E^\theta \left(\sum_j L_j^\theta - L^\theta \right) + \dots
\end{aligned} \tag{A.2}$$

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

$$[Q_{ji}] \quad P_i^* \frac{\partial Q(Q_{1i}, \dots, Q_{Ji})}{\partial Q_{ji}} \leq p_j^* \tau_{ji}, \tag{A.3}$$

$$[N_j^Y, N_j^H] \quad p_j^* \frac{\partial Y_j}{\partial N_j^Y} \leq W_j^*; R_j^* \frac{\partial H_j}{\partial N_j^H} \leq W_j^*, \tag{A.4}$$

$$[I_j^Y, I_j^H] \quad p_j^* \frac{\partial Y_j}{\partial I_j^Y} \leq P_j^*; R_j^* \frac{\partial H_j}{\partial I_j^H} \leq P_j^*, \tag{A.5}$$

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

$$\begin{aligned}
[c_j^\theta] \quad & \omega_j^\theta a_j^\theta \frac{\partial U(c_j^\theta, h_j^\theta)}{\partial c_j^\theta} c_j^\theta = P_j^* c_j^\theta \\
[h_j^\theta] \quad & \omega_j^\theta a_j^\theta \frac{\partial U(c_j^\theta, h_j^\theta)}{\partial h_j^\theta} h_j^\theta = R_j^* h_j^\theta
\end{aligned}$$

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

$$\omega_j^\theta a_j^\theta U(c_j^\theta, h_j^\theta) = x_j^{\theta*}, \tag{A.6}$$

where

$$x_j^{\theta*} \equiv R_j^* h_j^\theta + P_j^* c_j^\theta. \tag{A.7}$$

Therefore, we can write

$$[c_j^\theta] \quad c_j^\theta = \frac{\alpha_C(c_j^\theta, h_j^\theta)}{P_j^*} x_j^{\theta*} \tag{A.8}$$

$$[h_j^\theta] \quad h_j^\theta = \frac{1 - \alpha_C(c_j^\theta, h_j^\theta)}{R_j^*} x_j^{\theta*} \tag{A.9}$$

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 (A.7) and the slackness condition on the spatial mobility constraint, the first-order condition of the planning

problem with respect to L_j^θ is:

$$\sum_{\theta'} \omega_j^{\theta'} L_j^{\theta'} \frac{\partial a_j^{\theta'}(L_j^1, \dots, L_j^\Theta)}{\partial L_j^\theta} U(c_j^{\theta'}, h_j^{\theta'}) + W_j^* \frac{dN_j}{dL_j^\theta} \leq x_j^{\theta*} + E^\theta, \quad (\text{A.10})$$

with equality if $L_j^\theta > 0$. Further using (A.6), if $L_j^\theta > 0$ then:

$$W_j^* \frac{dN_j}{dL_j^\theta} + \sum_{\theta'} \frac{(x_j^{\theta*})' L_j^{\theta'}}{a_j^{\theta'}} \frac{\partial a_j^{\theta'}}{\partial L_j^\theta} = x_j^{\theta*} + E^\theta. \quad (\text{A.11})$$

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 (A.3)-(A.11) and the constraints enumerated in (i) to (iii) in Definition 2 hold.

It is straightforward to show that (A.3) to (A.9) 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 (A.11). Conversely, if $x_j^\theta = x_j^{\theta*}$ for $x_j^{\theta*}$ defined in (A.1) given the W^θ that solve the planners problem, 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 convex 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 labor subsidies τ_j^θ such that*

$$\tau_j^\theta = \frac{\gamma_{\theta,\theta}^{P,j} + \gamma_{\theta,\theta}^{A,j} + \sum_{\theta' \neq \theta} \frac{\gamma_{\theta,\theta'}^{P,j} w_j^{\theta'} + \gamma_{\theta,\theta'}^{A,j} x_j^{\theta'}}{w_j^\theta} \frac{L_j^{\theta'}}{L_j^\theta}}{1 - \gamma_{\theta,\theta}^A} \quad (\text{A.12})$$

coupled with transfers

$$T_j^\theta = -b^\theta \Pi^* - \frac{E^\theta}{1 - \gamma_{\theta,\theta}^A}.$$

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

Proof. The result follows from combining (1) with the expressions for labor demand (13) and expenditure per capita (15). The labor demand condition (13) implies that the value of the marginal product of labor entering in (18) can be written as function of wages, employment and elasticities:

$$W_j \frac{dN_j}{dL_j^\theta} = w_j^\theta \left(1 + \gamma_{\theta,\theta}^{P,j}\right) + \sum_{\theta' \neq \theta} w_j^{\theta'} \left(\frac{L_j^{\theta'}}{L_j^\theta}\right) \gamma_{\theta,\theta'}^{P,j}. \quad (\text{A.13})$$

Using this expression with and the definition of $\gamma_{\theta,\theta'}^{A,j}$, we can re-write (18) as follows:

$$x_j^\theta = \frac{\left(1 + \gamma_{\theta,\theta}^{P,j}\right) + \sum_{\theta' \neq \theta} \frac{\gamma_{\theta,\theta'}^{P,j} w_j^{\theta'} + \gamma_{\theta,\theta'}^{A,j} x_j^{\theta'}}{w_j^\theta} \frac{L_j^{\theta'}}{L_j^\theta}}{1 - \gamma_{\theta,\theta}^A} w_j^\theta - \frac{E^\theta}{1 - \gamma_{\theta,\theta}^A}. \quad (\text{A.14})$$

Combining this expression with (15) and (16) gives the result. □

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

Proof. We consider the following planning problem defined in section (3.3):

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

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 which are defined in the main text, and which we come back to below. We show here that this problem, noted $\mathcal{P}$, can be recast as a concave problem, under the condition stated in proposition 2. Therefore, a local maximum of $\mathcal{P}$ is necessarily its unique global maximum.

The planning problem $\mathcal{P}$ can be recast as the following equivalent problem $\mathcal{P}'$, after simple algebraic manipulations:

$$\max_{\{v^\theta, U_j^\theta, C_j^\theta, H_j^\theta, \tilde{L}_j^\theta, \tilde{N}_j^k, I_j^k, Q_{ij}, M_j, S_j\}} v^\theta \quad (\text{A.15})$$

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

$$v^\theta - \mathcal{F} \left(\frac{U_j^\theta \prod_{\theta' \neq \theta} (\tilde{L}_j^{\theta'})^{\frac{\gamma_{\theta', \theta}^A}{1 + \alpha}}}{(\tilde{L}_j^\theta)^{\frac{1 - \gamma_{\theta, \theta}^A}{1 + \alpha}}} \right) \leq 0 \quad \text{for all } j; \quad (\text{A.16})$$

$$\underline{v}^{\theta'} - \mathcal{F} \left(\frac{U_j^{\theta'} \prod_{\theta'' \neq \theta'} (\tilde{L}_j^{\theta''})^{\frac{\gamma_{\theta'', \theta'}}{1 + \alpha}}}{(\tilde{L}_j^{\theta'})^{\frac{1 - \gamma_{\theta', \theta'}}{1 + \alpha}}} \right) \leq 0 \quad \text{for all } j \text{ and } \theta' \neq \theta; \quad (\text{A.17})$$

$$U_j^\theta - U(C_j^\theta, H_j^\theta) \leq 0 \quad (\text{A.18})$$

$$\sum_i d_{ji} Q_{ji} - \left(b_Y^N (N_j^Y)^{\beta_Y} + b_Y^I (I_j^Y)^{\beta_Y} \right)^{\frac{1}{\beta_Y}} \leq 0 \quad \text{for all } j, i; \quad (\text{A.19})$$

$$\sum_\theta C_j^\theta + (I_j^Y) + (I_j^H) - Q(Q_{1j}, \dots, Q_{Jj}) \leq 0 \quad \text{for all } j; \quad (\text{A.20})$$

$$\sum_\theta H_j^\theta - \left(b_H^N (N_j^H)^{\beta_H} + b_H^I (I_j^H)^{\beta_H} \right)^{\frac{1}{\beta_H}} \leq 0 \quad (\text{A.21})$$

$$M_j - \left[\sum_\theta \left(Z_j^\theta \prod_{\theta'} (\tilde{L}_j^{\theta'})^{\frac{\gamma_{\theta', \theta}^P}{1 + \alpha}} (\tilde{L}_j^\theta)^{\frac{1}{1 + \alpha}} \right)^\rho \right]^{\frac{1}{\rho}} \leq 0 \quad \text{for all } j; \quad (\text{A.22})$$

$$N_j^Y + N_j^H - M_j \leq 0 \quad (\text{A.23})$$

$$\sum_j \left(\tilde{L}_j^\theta \right)^{\frac{1}{1 + \alpha}} - L^\theta = 0 \quad \text{for all } \theta \quad (\text{A.24})$$

To reach these expressions, we have introduced the auxiliary variables M_j , S_j and U_j^θ and we have used the

following change of variables

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

where $\mathcal{F}(\cdot)$ is a bijection that we define below and where we have written $\alpha = \Gamma^P$. 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:

$$L^\theta - \sum_j \left(\tilde{L}_j^\theta\right)^{\frac{1}{1+\alpha}} \leq 0 \text{ for all } \theta. \quad (\text{A.25})$$

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 constraints (A.16) and (A.17), and specifically the term $\frac{U_j^\theta \prod_{\theta' \neq \theta} (\tilde{L}_j^{\theta'})^{\frac{\gamma_{\theta', \theta}^A}{1+\alpha}}}{(\tilde{L}_j^\theta)^{\frac{1+\sigma_\theta - \gamma_{\theta, \theta}^A}{1+\alpha}}}$. By proposition 11 of Khajavirad et al. (2014), functions of the form $f(y, z) = \frac{\prod_{i=1}^k y_i^{a_i}}{z^b}$ where $a_i > 0$, $b > 0$ and $\sum_{i=1}^k a_i < b$ are G -concave, for functions $G(x)$ that are any concave transform of $-x^{\frac{1}{\sum a_i - b}}$. G -concave means that the function $G(f(y, z))$ is concave in (y, z) . Therefore, we seek a function $\mathcal{F}$ such that (A.16) and (A.17) are convex for all θ whenever $\gamma_{\theta', \theta}^A \geq 0$ for all $\theta' \neq \theta$ and $\frac{1+\sigma_\theta - \gamma_{\theta, \theta}^A}{1+\alpha} > 1 + \frac{\gamma_{\theta', \theta}^A}{1+\alpha}$. This parameter restriction corresponds to the assumption that $\Gamma^A > \Gamma^P$. We use here the common function $\mathcal{F}(x) = -x^b$ for $b = \frac{1+\Gamma^P}{\Gamma^P - \Gamma^A}$ in constraint (A.16) for any j and any θ . Given the definition of Γ^A , $\mathcal{F}(\cdot)$ is a concave transform of $G_\theta(x) = -x^{\frac{1+\Gamma^P}{\Gamma^P - (\sigma_\theta + \sum_{\theta' \neq \theta} \gamma_{\theta', \theta}^A)}}$ for any θ , where G_θ is a transformation that makes the group θ -constraint concave, applying Proposition 11 of Khajavirad et al. (2014). Second, functions of the form $f(x_1, \dots, x_n) = \left[\sum a_i x_i^\beta\right]^\rho$ are concave whenever $\beta \in (0, 1)$ and $\rho\beta \leq 1$. Therefore, constraints (A.19), (A.21) and (A.25) are convex. Constraint (A.18) is convex because $U(\cdot)$ is concave. Constraint (A.20) is convex because the aggregator $Q(\cdot)$ is concave. Next, consider the constraint (A.22). The second term is the negative of a composition of an increasing CES function with exponent $\rho \leq 1$, which is concave, and a series of functions of the form

$$f(x_1, \dots, x_\theta) = \prod_{\theta'} \left(x^{\theta'}\right)^{\frac{\gamma_{\theta', \theta}^P}{1+\alpha}} \left(x^\theta\right)^{\frac{1}{1+\alpha}}.$$

As concave transforms of a geometric mean, these functions are concave, whenever $\frac{1+\sum_{\theta'} \gamma_{\theta', \theta}^P}{1+\alpha} \in (0, 1)$. This restriction holds by definition of α . We finally invoke that the vector composition of a concave function that is increasing in all its elements with a concave function is concave. Therefore, constraint (A.22) is convex. Finally, constraint (A.23) is linear hence convex.

It follows that the relaxed problem $\mathcal{P}''$ is a maximization problem with concave objective and convex inequality constraints. It admits at most one global maximum, and a vector satisfying its first order conditions is necessarily the global maximum. If at this unique optimal point for $\mathcal{P}''$ the relaxed constraint (A.25) binds, so that (A.24) 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 $\mathcal{P}$.⁴⁰

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

⁴⁰We have not proven that (A.25) 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.

in this case to:

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

subject to the constraints (A.19), (A.20), (A.22), (A.23) and (A.25), which are unchanged except that they now hold for only one group. The modified constraint for housing production is:

$$\tilde{H}_j^\theta - \left(b_H^N (\tilde{N}_j^H)^{\beta_H(1+D)} + b_H^I (\tilde{I}_j^H)^{(1+D)\beta_H} \right)^{\frac{1}{\beta_H} \frac{1}{1+D}} \leq 0 \quad (\text{A.26})$$

where we have used the following change of variable $\tilde{H}_j^\theta = (H_j^\theta)^{1+d_{H,j}'}_{H,j}$. The modified housing market constraint (A.26) is convex. The objective of the planner is quasi-concave as the minimum of a ratio of a concave and a convex function, as long as $(1-\alpha_C) \frac{1}{1+d_{H,j}} + \alpha_C \leq \frac{1+\sigma_\theta-\gamma_{\theta,\theta}^A}{1+\alpha}$ 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. $\square$

A.2 Equivalence with Monopolistic Competition

Consider the economic geography environment from Section 4.1. As a reminder, that environment starts from general model from Section 3 and imposes only one labor type, inelastic housing supply ($H_j(N_j^H, I_j^H) = H_j$ is a constant), and only labor used in production of traded goods ($Y_j(N_j^Y, I_j^Y) = N_j^Y = N_j = z_j(L_j)L_j$). Now suppose that, in addition, the production structure in the traded sector is the same as in Krugman (1980): in each location j , M_j homogeneous plants produce differentiated varieties with constant elasticity of substitution κ among them, and setting up a plant in location j requires F_j units of labor. The resulting environment corresponds to the Redding (2016) or to Helpman (1998) in the absence of individual preference shocks ($\sigma = 0$).

We now show that the competitive allocation of such an extended model, as well as their normative implications, are equivalent to the model with homogeneous products analyzed in Section 4.1 under an aggregate production function equal to:

$$\tilde{Y}_j(L_j) = K_j (z_j(L_j)L_j)^{\frac{\kappa}{\kappa-1}}, \quad (\text{A.27})$$

where $K_j \equiv \frac{\kappa-1}{\kappa} (\kappa F_j)^{\frac{1}{1-\kappa}}$ is a constant. Therefore, a monopolistic competition model with no productivity spillovers is equivalent to a homogeneous-product model with perfect competition and spillover elasticity equal to $\gamma^P = \frac{1}{\kappa-1}$. This property relates to the result, dating back to at least Abdel-Rahman and Fujita (1990) and highlighted among others by Allen and Arkolakis (2014), that CES product differentiation with monopolistic competition has the same aggregate implications to constant-elasticity aggregate production function with increasing returns. In our context, we must also demonstrate that the equivalence extends to the welfare implications summarized in Proposition 1.

The key reason why this equivalence holds is that, as it is well-known, 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 4.1 go through. We note that these properties would not go through under monopolistic competition outside of CES. In that case, the entry and bilateral pricing decisions would be inefficient.

Environment We start by describing how the physical environment of this model differs from the environments from section 3. Now, the aggregate bilateral flows are $Q_{ji} = M_j^{\frac{\kappa}{\kappa-1}} q_{ji}$, where M_j is the number of plants in j and q_{ji} is the quantity exported by each of these from j to i . The feasibility constraint for traded goods(4) becomes $z_j(L_j)L_j = M_j (\sum_i \tau_{ji} q_{ji} + F_j)$ to account for the use of labor in setting up plants. Combining these two expressions,

that constraint can be further expressed:

$$M_j^{\frac{1}{\kappa-1}} (z_j (L_j) L_j - F_j M_j) = \sum_i \tau_{ji} Q_{ji}. \quad (\text{A.28})$$

Competitive Equilibrium Now we describe how the market allocation differs from the baseline environments. First, the optimization conditions (11) is replaced by:

$$\max \sum_i (p_{ji} - \tau_{ji} W_j) q_{ji} \quad (\text{A.29})$$

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

$$p_j \equiv M_j^{\frac{1}{1-\kappa}} \frac{\kappa}{\kappa-1} W_j \quad (\text{A.30})$$

As a result, condition (14) still determines the flows in the competitive equilibrium. Combining these pricing rules with (A.29), imposing zero profits and using (4) we obtain the number of producers in a competitive allocation:

$$M_j = \frac{z_j (L_j) L_j}{\kappa F_j}. \quad (\text{A.31})$$

And further combining with A.28, we can write

$$\widetilde{Y}_j (L_j) = \sum_i \tau_{ij} Q_{ij} \quad (\text{A.32})$$

for $\widetilde{Y}_j$ given in A.27.

We conclude that the competitive allocation can be represented as in the model without product differentiation from Definition 1 under the restrictions from Section 4.1 and assuming the aggregate production function $\widetilde{Y}_j (L_j)$. I.e., it is given by quantities $\{c_j, h_j, L_j, Q_{ij}, L_j\}$ and prices P_j, R_j, p_j , such that: (i) consumers optimize (i.e., c_j, h_j are a solution to (8) given expenditures x_i); (ii) trade flows are given by (14); (iii) employment L_j is consistent with the spatial mobility constraint (10); and (iv) all markets clear, i.e. (2), (3) and (A.32) hold.⁴¹

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

$$\begin{aligned} \mathcal{L} = & u - \sum_j \omega_j \left(u - a_j (L_j) U(c_j, h_j) \left(\frac{L_j}{L} \right)^{-\sigma} \right) \\ & - \sum_j p_j^* \left(\sum_i d_{ji} Q_{ji} - M_j^{\frac{1}{\kappa-1}} (z_j (L_j) L_j - F_j M_j) \right) \\ & - \sum_j P_j^* (L_j c_j - Q(Q_{1j}, \dots, Q_{Jj})) - \sum_j R_j^* (L_j h_j - H_j) - W \left(\sum_j L_j - L \right) + \dots \end{aligned} \quad (\text{A.33})$$

Relative to Definition 2, now the planner also chooses the number of firms M_j in each location and faces the constraint (A.28) instead of (4). We readily see that entry is efficient by noting that the first-order condition with respect to M_j implies (A.31). As a result, the market clearing constraint in the second line of (A.33) can be replaced by (A.32). The resulting planning problem is equivalent to Definition (2) applied to the economic geography model in Section 4.1 under the production function $\widetilde{Y}_j (L_j)$.

⁴¹Note that the definition of the competitive allocation can dispense with the wage W_j , which can be determined residually from (A.30).

A.3 Extension to Public Spending

The Lagrangian of planning problem described under the extension to public spending in Section 4 is

$$\begin{aligned}
\mathcal{L} = & u \\
& - \sum_j \omega_j L_j \left(u - a_j \left(G_j^U, L_j \right) U \left(c_j, h_j \right) \right) \\
& - \sum_j p_j^* \left(\sum_i d_{ji} Q_{ji} - Y_j \left(N_j^Y, I_j^Y \right) \right) \\
& - \sum_j P_j^* \left(I_j^G + L_j c_j + I_j^Y + I_j^H - Q \left(Q_{1j}, \dots, Q_{Jj} \right) \right) - \sum_j R_j^* \left(H_j^G + L_j h_j - H_j \left(N_j^H, I_j^H \right) \right) \\
& - \sum_j P_j^{G*} \left(G_j^Y + G_j^U - G_j \left(I_j^G, H_j^G \right) \right) \\
& - \sum_j W_j^* \left(N_j^I + N_j^H - N_j \left(G_j^Y, z_j \left(L_j \right) L_j \right) \right) \\
& - E \left(\sum_j L_j - L \right) + \dots
\end{aligned} \tag{A.34}$$

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

$$x_j^P = \frac{1 + \gamma^P}{1 - \gamma^A} w_j - \frac{E}{1 - \gamma^A}. \tag{A.35}$$

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

$$\begin{aligned}
\left[G_j^U \right] : & \quad L_j x_j^P \gamma^{A,G} = P_j^{G*} G_j^U \\
\left[G_j^Y \right] : & \quad W_j^* N_j \varepsilon_{N,G} = P_j^{G*} G_j^Y
\end{aligned}$$

where $\gamma^{A,G} \equiv \frac{\partial a_j}{\partial G_j} \frac{G_j}{a_j}$ and $\varepsilon_{N,G} = \frac{\partial N}{\partial G_j} \frac{G_j}{N_j}$ are elasticities of amenities and labor efficiency to government spending. Adding up the two previous equations, using the first-order conditions over I_j^G and H_j^G , and applying homogeneity of the the production function of government spending $G_j \left(I_j^G, H_j^G \right)$ we reach condition (30) in the text. Further combining (30) with (A.35) gives (29) in the text. In terms of the efficient implementation, under constant labor subsidies with a lump sum transfer individuals earn $x_j^P = w_j (1 + \tau) + \frac{\Pi}{L} + T$. Hence, the optimal tax must still be (23), as when there is no government spending. Given the value of τ , the lump-sum transfer T is such that

$$\sum_j \left(w_j \tau + x_j^G \right) \frac{L_j}{L} + T = 0$$

so that the national government budget constraint holds. This constraint ensures that each local government can be finance spending per capita x_j^G .

A.4 Extension to Preference Heterogeneity within Groups

The Lagrangian of planning problem described under the extension to preference draws in Section 4 is formally the same as (A.2), except that now the spillover function $a_j^{\theta'} \left(L_j^1, \dots, L_j^\Theta \right)$ is replaced by $a_i^{\theta'} \left(L_i^\theta \right)^{-\sigma_\theta}$. Following the same steps as in the proof of Proposition (1), we find that condition (18) is extended to

$$W_j \frac{dN_j}{dL_j^\theta} + \sum_{\theta'} \frac{x_j^{\theta'} L_j^{\theta'}}{a_j^{\theta'}} \frac{\partial a_j^{\theta'}}{\partial L_j^\theta} = x_j^\theta (1 + \sigma_\theta) + E^\theta \quad \text{if } L_j^\theta > 0. \tag{A.36}$$

Following the same steps as in the proof of Proposition (2), we find that (21) is extended to

$$\tau_j^\theta = \frac{\gamma_{\theta,\theta}^{P,j} + \left(\gamma_{\theta,\theta}^{A,j} - \sigma_\theta\right) + \sum_{\theta' \neq \theta} \frac{\gamma_{\theta,\theta'}^{P,j} w_j^{\theta'*} + \gamma_{\theta,\theta'}^{A,j} x_j^{\theta'*}}{w_j^{\theta'*}} \frac{L_j^{\theta'*}}{L_j^{\theta*}}}{1 - \left(\gamma_{\theta,\theta}^{A,j} - \sigma_\theta\right)}. \quad (\text{A.37})$$

The general-equilibrium structure underlying propositions (3) and (4) under the assumptions of the quantitative model can be expressed exactly as in the proof of Proposition (3) and as in the planning problem in relative changes from Section (A.5) below, the only modification being that the term $\gamma_{\theta,\theta}^A$ is replaced by $\gamma_{\theta,\theta}^A - \sigma_\theta$.

A.5 Planning Problem in Relative Changes and Proof of Proposition 4

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

Preliminaries We adopt the functional forms from Section 4.2. From the optimization problems (11) and (12) and market clearing in the housing market we obtain the following conditions:

$$W_i N_i^Y = \left(1 - b_{Y,i}^I\right) p_i Y_i. \quad (\text{A.38})$$

$$W_i N_i^H = \frac{1 - b_{H,i}^I}{1 + d_{H,i}} (1 - \alpha_C) X_i. \quad (\text{A.39})$$

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

$$\frac{N_i^H}{N_i} = \frac{\frac{1 - b_{H,i}^I}{1 + d_{H,i}} \frac{1 - \alpha_C}{1 - b_{Y,i}^I} \left(\frac{X_i}{p_i Y_i}\right)}{\frac{1 - b_{H,i}^I}{1 + d_{H,i}} \frac{1 - \alpha_C}{1 - b_{Y,i}^I} \left(\frac{X_i}{p_i Y_i}\right) + 1}. \quad (\text{A.40})$$

Using (A.38) and (A.39) along with labor-market clearing (A.4), we can further express gross expenditures over tradeable income as a function the shares of wages in expenditures:

$$\frac{X_i}{p_i Y_i} = \frac{1 - b_{Y,i}^I}{\frac{W_i N_i}{X_i} - \frac{1 - b_{H,i}^I}{1 + d_{H,i}} (1 - \alpha_C)}. \quad (\text{A.41})$$

We now re-formulate some of the equilibrium from Definition 1 conditions to include prices. Consider first the market clearing condition 5. Multiplying both sides by the price of the traded bundle P_j , letting $E_j^Y = P_j Q_j = P_j Q(Q_{1j}, \dots, Q_{Jj})$ be the gross expenditures in tradeable goods in j (used both as intermediate and for final consumption), using equilibrium in the housing market, and using optimality condition for the choice of intermediate inputs in the traded sector, we can re-write that condition as

$$E_j^Y = \left(\alpha_C + (1 - \alpha_C) \frac{b_{H,j}^I}{d_{H,j} + 1}\right) X_j + b_Y^I (p_j Y_j). \quad (\text{A.42})$$

where $X_j = \sum_{\theta'} L_j^{\theta'} x_j^{\theta'}$ are the aggregate expenditures in region j . This condition says that aggregate expenditures in traded goods results from the aggregation of expenditures by consumers and final producers. Second, consider the market condition (4) for traded commodities. Multiplying both sides by the price of traded commodities at j , p_j , this condition is equivalent to

$$\sum_i s_{ji}^X = 1. \quad (\text{A.43})$$

where $s_{ji}^X \equiv \left(\frac{E_i}{p_j Y_j}\right) s_{ji}^M$ is region i 's share of j 's sales of tradeable goods (i.e., the export share of i in j) and

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

$$\sum_j s_{ji}^M \equiv 1. \quad (\text{A.44})$$

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 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:

$$\sum_j s_{ij}^X \left(\frac{\hat{p}_i}{\hat{P}_j} \right)^{1-\sigma} E_j^{\hat{Y}} = \hat{p}_i \hat{Y}_i \text{ for all } i, \quad (\text{A.45})$$

$$\sum_j s_{ji}^M \left(\frac{\hat{p}_j}{\hat{P}_i} \right)^{1-\sigma} = 1 \text{ for all } i, \quad (\text{A.46})$$

$$\left(1 - \frac{N_i^H}{N_i} \right) \hat{p}_i \hat{Y}_i + \frac{N_i^H}{N_i} \hat{X}_i = \hat{W}_i \hat{N}_i \text{ for all } i, \quad (\text{A.47})$$

$$\hat{W}_i^{1-b_{Y,i}^I} \hat{P}_i^{b_{Y,i}^I} = \hat{p}_i \text{ for all } i \quad (\text{A.48})$$

where $\hat{X}_j = \sum_\theta s_j^{X,\theta} \hat{x}_j^\theta \hat{L}_j^\theta$ is the change in and aggregate expenditures by region, where $s_j^{X,\theta}$ is group θ 's share in the consumer expenditures in j in the initial equilibrium. Equations (A.45) and (A.46) follow from expressing (A.43) and (A.44) in relative changes and using the CES functional for $Q(\cdot)$ from (33). In condition (A.45), using (A.42) implies that the change in expenditures in tradeable commodities is:

$$E_j^{\hat{Y}} = \hat{X}_j + \frac{b_Y^I}{\left(\alpha_C + (1 - \alpha_C) \frac{b_H^I}{d_{H,j}+1} \right) \left(\frac{X_j}{p_j Y_j} \right) + b_Y^I} \hat{p}_j \hat{Y}_j. \quad (\text{A.49})$$

Condition (A.47) follows from expressing labor-market clearing (7) in relative changes together with (A.38) and (A.39), where the non-traded labor share $\frac{N_i^H}{N_i}$ is defined in (A.40). Condition (A.48) follows from optimization of producers of tradeable commodities.

The system (A.45) to (A.48) 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 $\hat{X}_i$, and 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 expressions in the market allocation (13) allows us to back out the efficiency of each group:

$$z_i^\theta = \frac{w_i^\theta}{W_i} \left(\frac{L_i^\theta}{N_i} \right)^{1-\rho} \quad (\text{A.50})$$

Expressing the CES functional form for the aggregation of labor types in (36) in relative changes and using (A.50) we obtain:

$$\hat{N}_i = \left(\sum_\theta s_i^{W,\theta} \left(\hat{L}_i^\theta \right)^\rho \prod_{\theta'} \left(\hat{L}_i^{\theta'} \right)^{\gamma_{\theta',\theta}^P} \right)^{\frac{1}{\rho}} \quad (\text{A.51})$$

where $s_j^{W,\theta} = \frac{w_j^\theta L_j^\theta}{\sum_{\theta'} w_j^{\theta'} L_j^{\theta'}}$ is group θ share of wages in city j . This expression 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, (10),

$$\hat{u}^\theta = \prod_{\theta'} \left(\hat{L}_i^{\theta'} \right)^{\gamma_{\theta',\theta}^A} \frac{\hat{x}_i^\theta}{\hat{P}_i^{\alpha_C} \hat{R}_i^{1-\alpha_C}}. \quad (\text{A.52})$$

In this expression, $\hat{R}_i$ is the change in the price of non-traded goods in location i . Solving for the equilibrium in the market from non-traded goods, after some manipulations this relative price can be expressed as solely a function of the price of the own traded good, the price index of traded commodities, of aggregate expenditures in i :

$$\hat{R}_i = \left(\hat{p}_i^{\frac{1-b_{H,i}^I}{1-b_{Y,i}^I}} \hat{P}_i^{b_{H,i}^I - b_{Y,i}^I} \hat{X}_i^{d_{H,i}} \right)^{\frac{1}{1+d_{H,i}}}. \quad (\text{A.53})$$

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

$$\sum_j s_j^{L,\theta} \hat{L}_j^\theta = 1 \text{ for all } \theta. \quad (\text{A.54})$$

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, an equilibrium in changes given a change in expenditure per capita $\{x_j^\theta\}$ consists of $\{\hat{P}_i, \hat{p}_i, \hat{Y}_i, \hat{W}_i, \hat{N}_j, \hat{L}_j^\theta, \hat{R}_i, \hat{u}^\theta\}$ such that equations (A.45) to (A.54) hold. These equations conform a system of $5J + \Theta J + \Theta$ equations in equal number of unknowns, where J is the number of locations and Θ is the number of types.

Planner's Problem in Relative Changes In the implementation, we solve an optimization over $\{x_j^\theta\}$ subject to $\{\hat{P}_i, \hat{p}_i, \hat{Y}_i, \hat{W}_i, \hat{N}_j, \hat{L}_j^\theta, \hat{R}_i, \hat{u}^\theta\}$ consistent with (A.45) to (A.54) in order to maximize the utility of a given group θ , $\hat{u}^\theta$, subject to a lower bound for the change in utility of the other groups ($u^{\theta'} \geq \underline{u}^{\theta'}$ for $\theta' \neq \theta$). This problem (call it $\mathcal{P}_2''$) differs formally from the baseline problem in Definition 2 (call it $\mathcal{P}_2$) for two reasons. First, it features prices, expenditures and incomes rather than being expressed in terms of quantities alone, as in conditions (A.38) to (A.44). We denote $\mathcal{P}_2'$ an intermediary problem expressed in terms of income and expenditure rather than quantities, but still in levels. Second, $\mathcal{P}_2''$ is expressed in changes relative to an initial equilibrium rather than in levels. We show here that the two problems are nevertheless equivalent. Therefore, the problem that we implement has a unique maximizer under the conditions of Proposition 2.

To see that the two problems have the same solutions, we first focus on the first order conditions of problem $\mathcal{P}_2$ and compare them to the problem in levels $\mathcal{P}_2'$ expressed in income and expenditures terms rather than in quantities. Conditions (A.3) and (A.5) 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 $x \rightarrow x_o \hat{x}$ for all variables, where x_o is a constant corresponding to the observed data and $\hat{x}$ the optimization variable in $\mathcal{P}_2'$. The problem in relative changes considered here and the problem $\mathcal{P}_2'$, and in turn problem $\mathcal{P}_2$, have therefore the same solutions, subject to the appropriate change of variables. In particular, a point that satisfies the first order conditions under the conditions of Proposition (3) is the (unique) global maximizer for both problems.

Proof of Proposition 4 Proposition (4) follows from inspecting (A.45) to (A.54) under the planner’s problem in relative changes defined above. Note that, given the elasticities $\{\alpha_C, \rho, b_{Y,j}^I, b_{H,j}^I, d_{H,j}\}$, and as long as $b_Y^I > 0$, computing the change in tradeable expenditures requires information about gross expenditures over tradeable income, $\frac{X_j}{p_j Y_j}$. This information is also needed to compute the non-traded labor share $\frac{N_i^H}{N_i}$ in (A.47). However, as shown in (A.40) and (A.41), $\frac{X_j}{p_j Y_j}$ can be constructed from the elasticities $\{\alpha_C, b_Y^I, b_H^I, d_{H,j}\}$ and the share of wages in gross expenditures, $\frac{W_i N_i}{X_i}$.

B Data Appendix

We detail the construction of the variables used to implement the counterfactuals. We rely on three primary data sources: i) BEA regional economic accounts, CA4 Personal Income and Employment by Major Component (<https://www.bea.gov/regional/downloadzip.cfm>); ii) estimates of disposable income by MSA from Dunbar (2009) based on BEA regional economic accounts (https://www.bea.gov/papers/xls/dpi_msa_working_paper_2001_2007_results.xls); iii) March CPS based on the IPUMS-CPS, ASEC 2007 sample (<https://cps.ipums.org/cps/index.shtml>).

B.1 Appendix to Section 5.1 (Data)

MSA-Level Outcomes We first extract from Dunbar (2009) the following information: *population*, *personal income*, and *personal taxes* paid by MSA, in 2007. To split personal income by source of income, we merge this data with the BEA Regional Economic Accounts. We compute the share of personal income corresponding to each possible source: labor income, capital income, and transfers. Specifically, we measure labor income as BEA’s *earning by place of work*;⁴² capital income as the sum of *proprietor’s income*, and *dividends, interests and rents*; and transfers as *current transfer receipts*.⁴³ Combining these shares with the total personal income and taxes by MSA from Dunbar (2009) provides us with a measure of labor income, capital income, transfers and taxes at the MSA level.

Break-Down By Skill Group We split these totals at the MSA level into two groups, high skill and low skill. To that end, we use March CPS data, part of the Integrated Public Use Microdata Series (Flood et al., 2017). The March CPS reports, at the individual level, income by source (labor, capital and transfers), MSA of residence and level of education. We define as high skill those workers who have completed 4 years of college, or more; and as low skill those who have completed less than 4 years of college, or have not gone to college. We aggregate individual level data from the March CPS at the MSA-group level, to get an MSA-level estimate of capital income, labor income, transfers and taxes by group, as well as the population of both groups.⁴⁴ We do not use this information directly, as the MSA aggregates from individual-level data might be noisy, in particular for smaller MSA’s. Instead, we use this information to construct the shares of the MSA-level outcomes from the BEA described above corresponding to each group of workers. That is, for each MSA i , we compute $s_i^L = \frac{\tilde{X}_i^L}{\tilde{X}_i^L + \tilde{X}_i^H}$ where $\tilde{X}_i^\theta$ denotes total MSA i -group θ level capital income, labor income, transfers, taxes or population in the march CPS data. We use this share s_i^L , together with the MSA-level dataset for income described above, to build our measure $X_i^\theta = X_i s_i^\theta$ of MSA-group

⁴²The BEA’s earning by place of work is comprised of: wages and salaries, supplements to wages and salaries, proprietor’s income, net of contributions for government social insurance, plus adjustment for residence.

⁴³Current transfer receipts is defined as the sum of government social benefits and net current transfer receipts from business (<https://www.bea.gov/glossary/glossary.p.htm>).

⁴⁴Specifically, we aggregate the following categories to measure capital income: income from interest, from dividends, from rents. We aggregate the following categories to measure labor income: wage and salary income, non-farm business income, farm income, income from worker’s compensation, alimony and child support. We aggregate the following categories to measure transfers: welfare income, social security income, income from SSI, income from unemployment benefits, income from veteran’s, survivor’s, disability benefit, income from educational assistance. We aggregate the following categories to measure taxes paid: federal income tax liability, after all credits, and state income tax liability, after all credits.

level population, labor income, capital income, transfers and taxes. We also compute the corresponding per-capita measures for each MSA-group: $x_i^\theta = \frac{X_i^\theta}{L_i^\theta}$.

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

$$x_i^\theta = x_0^\theta + \sum_j \beta_j^\theta DEM_{ij}^\theta + \varepsilon_i^\theta, \quad (\text{A.55})$$

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

$$\hat{x}_i^\theta \equiv x_i^\theta - \sum_j \hat{\beta}_j^\theta \left(DEM_{ij}^\theta - \overline{DEM}_j^\theta \right) \quad (\text{A.56})$$

where $\hat{\beta}_j^\theta$ is the estimate from (A.55) and $\overline{DEM}_j^\theta \equiv \frac{1}{I} \sum_i DEM_{ij}^\theta$.⁴⁵ The adjusted measure $\hat{x}_i^\theta$ can be interpreted as the value that that variable x_i^θ would take if the demographics of group θ in city i were equal to the mean demographics of group θ in the country. Through this procedure, we remove heterogeneity across cities within groups due to observable compositional difference. For the quantification we use $\hat{x}_i^\theta$ to measure each per-capita outcome. The corresponding MSA-level variable is $\hat{X}_i^\theta = \hat{x}_i^\theta L_i^\theta$. The resulting data $(\hat{X}_i^\theta, \hat{x}_i^\theta, L_i^\theta)$ is our MSA-group level dataset, where X stands for labor income, capital income, transfers and taxes.

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

$$x_i^\theta = w_i^\theta (1 + \tau_i^\theta) + T_i^\theta + b^\theta \Pi^H. \quad (\text{A.57})$$

The variables $\{w_i^\theta, \tau_i^\theta, T_i^\theta\}$, respectively labor income per capita, tax paid per capita (where $\tau_i^\theta < 0$ is a tax), and transfer received per capita, are directly taken from the BEA/CPS dataset constructed above. We measure b^θ as the average fraction of national capital income owned by each type θ worker in BEA/CPS dataset. This step gives $b^S L^S = 0.59$ and $b^U L^U = 0.41$. Finally, we set a value for national profits and returns to land Π^H that is consistent with the general equilibrium of the model. Using profit maximization in the housing sector and market clearing in the non-tradeable sector we obtain the following expression for Π^H as function of calibrated elasticities and observable outcomes:

$$\Pi^H = \frac{(1 - \alpha_C) \sum_i \frac{d_{H,i}}{d_{H,i}+1} \sum_\theta L_i^\theta (w_i^\theta (1 + \tau_i^\theta) + T_i^\theta)}{1 - (1 - \alpha_C) \sum_i \frac{d_{H,i}}{d_{H,i}+1} \sum_\theta b^\theta L_i^\theta}. \quad (\text{A.58})$$

Using x_i^θ we then construct X_i (aggregate expenditure by MSA) as $X_i = \sum L_i^\theta x_i^\theta$ and $s_j^{X,\theta}$ (share of expenditures by type within MSA). Following these adjustments, we need to ensure that the sum of transfers paid by the government equal the sum of taxes levied. To that end, we scale all transfers uniformly so that they add up to the sum of taxes. This ensures that the government budget constraint holds.⁴⁶

⁴⁵I.e., we define $\hat{x}_i^\theta \equiv x_0^\theta + \sum_j \hat{\beta}_j^\theta \overline{DEM}_j^\theta + \varepsilon_i^\theta$, where ε_i^θ is the estimated residual from (A.55).

⁴⁶This step implies that transfers are uniformly scaled down by 35%. The fact that total taxes and transfers do not match in our dataset comes in part from having removed heterogeneity that is not place-specific from the data

Traded and Non-Traded Sectors We need data on the relative size of the non-traded sector in each city to calibrate the labor shares by sector. The CPS data also reports the sector of activity of workers. We measure at the MSA level the share of workers who work in the non traded sector by counting all workers in the following NAICS sectors: retail, real estate, construction, education, health, entertainment, hotels and restaurants. This measure is not group-specific. To remove unmodeled heterogeneity in this measure, we compute a series of MSA-level socio-demographic characteristics, as above, and regress the share of workers in the non-traded sector on these demographic characteristics. We compute, as above, the predicted share of workers in the non-traded sector in each city, assuming that demographic characteristics of the city are at the nationwide mean.

Trade Shares We need data on trade shares between MSA's, s_{ij}^M and s_{ij}^X (import and export shares). These flows are observed in the CFS data, but not at the finer geographic level that we consider here (MSA). Therefore, we adapt the procedure in Allen and Arkolakis (2014) and Monte et al. (2015), whereby the import shares from the CFS data are used to parametrize the elasticity of trade with respect to distance. In particular, the model implies the following expression for share of location i 's imports originating from j :

$$s_{ji}^M = \left(\frac{d_{ji}}{P_i} \frac{W_j^{1-b_Y^I} P_j^{b_Y^I}}{z_j} \right)^{1-\sigma} \equiv \left(d_{ji} \delta_j^D \delta_i^O \right)^{1-\sigma} \quad (\text{A.59})$$

where δ_i^O and δ_j^D are origin and destination fixed effects. We assume that trade costs have the form $\ln d_{ji} = \psi \ln \text{dist}_{ji} + e_{ji}$, where dist_{ji} is the great circle distance between MSA's j and i . We use Allen and Arkolakis (2014) estimate for ψ and set trade costs to $d_{ji} = \text{dist}_{ji}^\psi$. We then construct the smoothed import shares s_{ji}^M between MSA's using A.59. To that end we must obtain the values of $\{\delta_j^D, \delta_i^O\}$, which are uniquely pinned down, up to a normalization, by considering the identity that sales equals income:

$$p_j Y_j = \sum_i s_{ji}^M E_i, \quad (\text{A.60})$$

together with equation (A.59) and the definition of the price index, leading to:

$$\left(\delta_i^O \right)^{\sigma-1} = \sum_j \left(d_{ji} \delta_j^D \right)^{1-\sigma}. \quad (\text{A.61})$$

Plugging (A.59) and (A.61) in (A.60), we get a system N equations in N unknowns, which we solve to recover $\{\delta_j^D, \delta_i^O\}$ and in turn s_{ji}^M . The export shares are then constructed using $s_{ji}^X \equiv \left(\frac{E_i}{p_j Y_j} \right) s_{ji}^M$, where spending E_i and traded income $p_j Y_j$.

B.2 Appendix to Section 5.2 (Calibration)

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

$$1 - b_{H,j}^I = \frac{W_j N_j^H}{(1 - \alpha_C) X_j} (1 + d_{H,j}). \quad (\text{A.62})$$

We compute this expression using the observed wage bill of worker sin non-traded sectors $W_j N_j^H$ and total expenditure X_j described in the previous subsection, and our calibrated values for α_C and $d_{H,j}$ described in Section (7). In practice, this step requires observing $\frac{W_j N_j^H}{(1 - \alpha_C) X_j} (1 + d_{H,j}) \in [0, 1]$ in all cities. However, for some cities we find $\frac{W_j N_j^H}{(1 - \alpha_C) X_j} (1 + d_{H,j}) > 1$. In these cases, we recalibrate the housing supply elasticity in the non-traded sector $d_{H,j}$ such

and from our treatment of capital to be consistent with the sources of capital income (profits from housing rents), which scales down its share in income relative to the data.

that $\frac{W_j N_j^H}{(1-\alpha_C)X_j} (1 + d_{H,j}) = 1$, and set $b_{H,j}^I = 0$.⁴⁷ For four MSA's where $\frac{W_j N_j^H}{(1-\alpha_C)X_j} > 1$, we set $W_j N_j^H = (1 - \alpha_C)X_j$. This corresponds in practice to scaling down $W_j N_j^H$ by less than 2 % in these cities.

Efficiency Spillover Elasticities Under the assumptions of the quantitative model, the labor demand condition (13) gives the following expression for the log wage of type- θ worker:

$$\ln w_j^\theta = \left[\rho \left(1 + \gamma_{\theta,\theta}^P \right) - 1 \right] \ln \left(L_j^\theta \right) + \rho \gamma_{\theta',\theta}^P \ln \left(L_j^{\theta'} \right) + \ln W_j - (\rho - 1) \ln N_j + \ln \varepsilon_j^\theta. \quad (\text{A.63})$$

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

$$\Delta \ln w_j^\theta = \left[\rho \left(1 + \gamma_{\theta,\theta}^P \right) - 1 \right] \Delta \ln \left(L_j^\theta \right) + \rho \gamma_{\theta',\theta}^P \Delta \ln \left(L_j^{\theta'} \right) + \Delta \kappa_j + \Delta \ln \varepsilon_j^\theta \quad (\text{A.64})$$

where $\Delta \kappa_j = \Delta \ln W_j - (\rho - 1) \Delta \ln N_j$ is a city effect. We can adopt two approaches to map Diamond (2016) estimates. She estimates equations (27) and (28) in her paper using Bartik shocks as instruments. The only difference between these equations in her paper correspond and (A.64) is the fixed effect $\Delta \kappa_j$ here. Assuming that the inclusion of the fixed effect $\Delta \kappa_j$ would not alter Diamond (2016) estimates, we can directly map her estimates in Column 3 of Table 5, i.e. $\rho \left(1 + \gamma_{S,S}^P \right) - 1 = 0.229$, $\rho \gamma_{U,S}^P = 0.312$, $\rho \left(1 + \gamma_{U,U}^P \right) - 1 = -0.552$, $\rho \gamma_{S,U}^P = 0.697$. Applying the ρ reported from her analysis then gives $(\gamma_{UU}^P, \gamma_{US}^P, \gamma_{SU}^P, \gamma_{SS}^P) = (0.14, 0.8, 1.78, 2.15)$.

We further re-scale these implied spillover elasticities so that they imply a city-level estimate of efficiency spillovers in line with classic references such as Ciccone and Hall (1996) or Combes et al. (2008). The typical estimating equation in these and related studies summarized by Rosenthal and Strange (2004) consists of a regression of average city wages w_j on city population L_j , variously instrumented assuming a single spillover elasticity. In log-changes, such an equation would take the form: $\hat{w}_j = \gamma^P \hat{L}_j + \psi_j$, where ψ_j is a city effect and γ^P is the city-level spillover elasticity. In our environment, city-level wages are $w_j L_j = N_j W_j$. Under the assumptions of the quantitative model, applying (A.51) we find that an exogenous shift in the total population of city j keeping its composition across groups constant would then imply:

$$\hat{w}_j = \left[s_j^{W,S} \left(\gamma_{S,S}^P + \gamma_{U,S}^P \right) + \left(1 - s_j^{W,S} \right) \left(\gamma_{S,U}^P + \gamma_{U,U}^P \right) \right] \hat{L}_j + \hat{W}_j, \quad (\text{A.65})$$

where $s_j^{W,S}$ is the share of skilled workers in wages in city j . Through the lens of our model, the coefficient γ^P estimated at the city level would correspond to $\overline{s^{W,S}} \left(\gamma_{S,S}^P + \gamma_{U,S}^P \right) + \left(1 - \overline{s^{W,S}} \right) \left(\gamma_{S,U}^P + \gamma_{U,U}^P \right)$, where $\overline{s^{W,S}}$ is the average skill share across cities. Therefore, we uniformly re-scale the $\gamma_{\theta,\theta'}^P$ coefficients backed out from Diamond (2016) above such that, under their scaled values, $\overline{s^{W,S}} \left(\gamma_{S,S}^P + \gamma_{U,S}^P \right) + \left(1 - \overline{s^{W,S}} \right) \left(\gamma_{S,U}^P + \gamma_{U,U}^P \right) = \gamma^P$ for a $\gamma^P = 0.06$ from Ciccone and Hall (1996) and $\overline{s^{W,S}} = 0.49$ as observed in our data. The elasticities resulting from this procedure are reported in the text.

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

$$\Delta \ln L_j^\theta = a_0^\theta \Delta \ln \left(\frac{w_j^\theta}{P_j} \right) + a_1^\theta \Delta \ln \left(\frac{R_j}{P_j} \right) + a_2^\theta \Delta \ln \left(a_j^{\theta,D} \right) + \Delta \xi_j^\theta, \quad (\text{A.66})$$

and $a_j^{\theta,D} \equiv \prod_{\theta'} \left(L_j^{\theta'} \right)^{\zeta^{A,\theta'}}$ is the endogenous component of amenities in her analysis and (a_0, a_1, a_2) are estimated coefficients. Columns 3 and 4 of Table 3 of Diamond (2016) reports the following point estimates: $(a_0^U, a_0^S, a_2^U, a_2^S, \zeta^{A,S}, \zeta^{A,U}) = (4.026, 2.116, 0.274, 1.012, 2.6, -2.6)$. We can generate the same expression from our setup and match the coefficients

⁴⁷This step affects the profits generated in the economy, in practice reducing slightly the implied values of Π^H and X_j whose construction we described in the previous subsection of the appendix, which in turn modifies (A.62). We iterate to find the unique fixed point of the procedure.

from our model to these parameters. We do so allowing for idiosyncratic preference draws within each type as in Section 31 (i.e., assuming $\sigma_\theta > 0$), and show that whether or not we allow for these draws has no impact for our quantification, provided we match the point estimates above. The labor-supply equation implied by (31) is

$$\sigma_\theta \ln L_j^\theta = \ln \left(\frac{x_j^\theta}{P_j} \right) - (1 - \alpha_C) \ln \left(\frac{R_j}{P_j} \right) + \ln \left(a_j^\theta \right) + \left(\sigma_\theta \ln L^\theta - \ln u^\theta \right). \quad (\text{A.67})$$

We redefine our amenity index a_j^θ defined in 38 as a function of $a_j^{\theta,D}$: $a_j^\theta = A_j^\theta (L_j^\theta)^{\gamma_{\theta,\theta}^A - \beta^{a,\theta} \zeta^{A,\theta}} a_j^{\theta,D}$, where $\beta^{a,\theta} \equiv \frac{\gamma_{\theta',\theta}^A}{\zeta^{A,\theta'}}$ is by construction constant over θ' . Using this condition in A.67, re-arranging and expressing the equation in changes we obtain

$$\Delta \ln L_j^\theta = \frac{1}{(\sigma_\theta - \gamma_{\theta,\theta}^A) + \beta^{a,\theta} \zeta^{A,\theta}} \Delta \ln \left(\frac{x_j^\theta}{P_j} \right) - \frac{1 - \alpha_C}{(\sigma_\theta - \gamma_{\theta,\theta}^A) + \beta^{a,\theta} \zeta^{A,\theta}} \Delta \ln \left(\frac{R_j}{P_j} \right) + \frac{\beta^{a,\theta}}{(\sigma_\theta - \gamma_{\theta,\theta}^A) + \beta^{a,\theta} \zeta^{A,\theta}} \Delta \ln \left(a_j^D \right) + \Delta \xi_j^\theta \quad (\text{A.68})$$

where $\Delta \xi_j^\theta \equiv \frac{1}{(\sigma_\theta - \gamma_{\theta,\theta}^A) + \beta^{a,\theta} \zeta^{A,\theta}} (\ln A_j^\theta + \sigma_\theta \ln L^\theta - \ln u^\theta)$. Comparing A.66 with A.68 readily allow us to map Diamond (2016) estimates to our parameters as follows:

$$\begin{aligned} \gamma_{\theta,\theta}^A - \sigma_\theta &= \frac{a_2^\theta}{a_0^\theta} \zeta^{A,\theta} - \frac{1}{a_0^\theta} \\ \gamma_{\theta',\theta}^A &= \frac{a_2^\theta}{a_0^\theta} \zeta^{A,\theta'} \end{aligned}$$

for $\theta = U, S$. The resulting numbers are reported in the text. We note that, conditional the estimates of $(a_0^U, a_0^S, a_2^U, a_2^S, \zeta^{A,S}, \zeta^{A,U})$ in A.66, we back out the value of $\gamma_{\theta,\theta}^A - \sigma_\theta$ but are unable to distinguish $\gamma_{\theta,\theta}^A$ from $-\sigma_\theta$. However, as discussed in Section 31, $\gamma_{\theta,\theta}^A - \sigma_\theta$ is the relevant summary statistic to characterize optimal allocations and policies given data under the definition of the planner problem with idiosyncratic preference draws defined in Section 31.

Equivalence of Amenity Spillovers Between Multiple and Single Worker Types In the quantitative section we implement a version of the model with a single type of worker, using a single spillover elasticity γ^A . To find this elasticity, we note that, under a single worker type, the labor-supply equation (A.67) expressed in time differences becomes

$$\Delta \ln L_j = -\frac{1}{\gamma^A - \sigma} \left(\Delta \ln \left(\frac{x_j}{P_j} \right) - (1 - \alpha_C) \Delta \ln \left(\frac{R_j}{P_j} \right) \right) + \Delta \xi_j \quad (\text{A.69})$$

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

$$\Delta \ln L_j^\theta = \sum_\theta s_j^{L,\theta} \frac{1}{\sigma_\theta - \gamma_{\theta,\theta}^A} \left(\Delta \ln \left(\frac{x_j^\theta}{P_j} \right) - (1 - \alpha_C) \Delta \ln \left(\frac{R_j}{P_j} \right) \right) + \sum_\theta s_j^{L,\theta} \sum_{\theta' \neq \theta} \frac{\gamma_{\theta',\theta}^A}{\sigma_\theta - \gamma_{\theta,\theta}^A} \Delta \ln L_j^{\theta'} + \Delta \xi_j^\theta \quad (\text{A.70})$$

where $\Delta \xi_j^\theta$ includes changes in the labor supply type- θ workers and in the exogenous component of amenities, A_j^θ . We can draw an equivalence between the aggregate elasticity that would be estimated assuming homogeneous workers (i.e., using (A.69)) when the true model includes heterogeneous workers, so that the data is generated by (A.70). In the latter, assuming a shock that exogenously changes population and expenditure per capita in the same proportion

for every worker, aggregating the labor supplies by skill we obtain:

$$\hat{L}_j = \left(\frac{\sum_{\theta} \frac{s_j^{L,\theta}}{\sigma_{\theta} - \gamma_{\theta,A}^A}}{1 - \sum_{\theta} \sum_{\theta' \neq \theta} \frac{s_j^{L,\theta} \gamma_{\theta',\theta}^A}{\sigma_{\theta} - \gamma_{\theta',\theta}^A}} \right) \left(\hat{x}_j - \hat{P}_j - (1 - \alpha_C) (\hat{R}_j - \hat{P}_j) \right) + \Delta \bar{\xi}_j. \quad (\text{A.71})$$

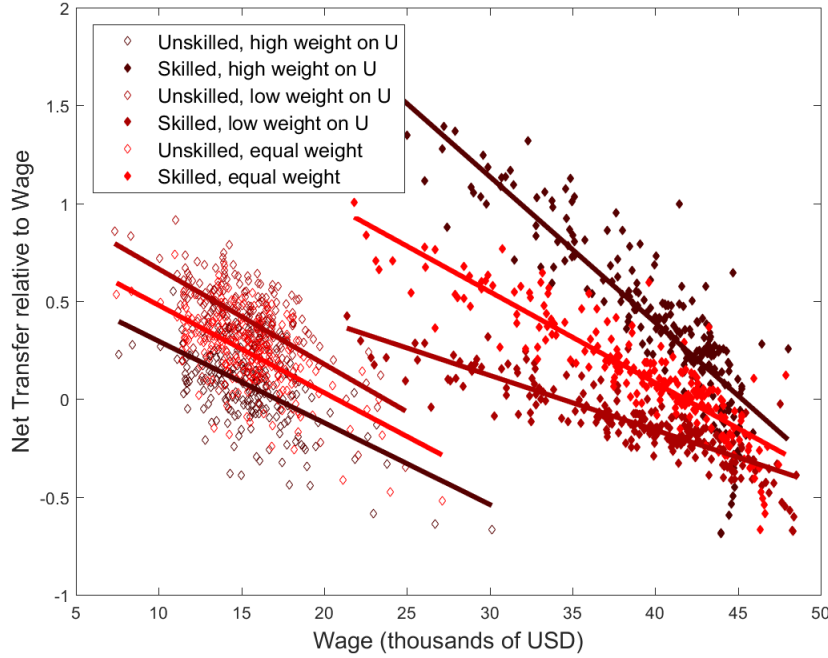
where $s_j^{L,\theta}$ is the share of type θ workers in j and $\Delta \bar{\xi}_j \equiv \sum_{\theta} s_j^{L,\theta} \Delta \xi_j^{\theta}$. Comparing (A.69) with (A.71), we obtain that, at the average share of type- θ workers in the economy, $\bar{s}^{L,\theta} = \frac{1}{J} \sum_j s_j^{L,\theta}$, the coefficient that would be recovered is:

$$\gamma^A - \sigma = \frac{1 + \sum_{\theta} \sum_{\theta' \neq \theta} \frac{s_j^{L,\theta} \gamma_{\theta',\theta}^A}{\gamma_{\theta,\theta}^A - \sigma_{\theta}}}{\sum_{\theta} \frac{s_j^{L,\theta}}{\gamma_{\theta,\theta}^A - \sigma_{\theta}}}. \quad (\text{A.72})$$

When implementing the model with a single worker type we use this expression to determine $\gamma^A - \sigma$. As discussed in Section 31, $\gamma^A - \sigma$ is the relevant summary statistic to characterize optimal allocations and policies given data under the definition of the planner problem with idiosyncratic preference draws defined in Section 31.

C Appendix to Section 6.4 (Robustness)

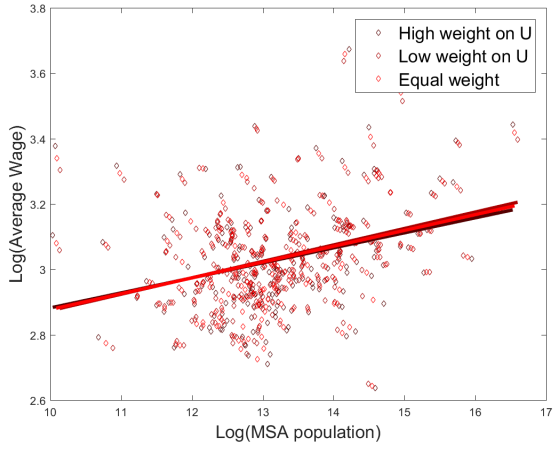
Figure A.1: Transfers relative to Wage by Skill Level and MSA under Different Pareto Weights



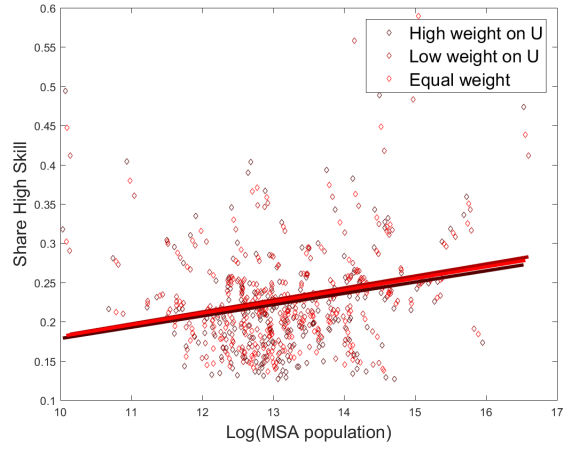
The figure replicates Figure 5 under alternative weights for each skill group in the planning problem. In the case “low weight on U”, the weight is 10 times larger for high skill workers. The welfare changes are $(\hat{u}^U, \hat{u}^S) = (0.89, 1.28)$. In the case “high weight on U”, the weight is 10 times larger for low skill workers. The welfare changes are $(\hat{u}^U, \hat{u}^S) = (1.18, 0.90)$. The case “equal weight” corresponds to the optimal allocation in Figure 5.

Figure A.2: Urban Premia under Different Pareto Weights

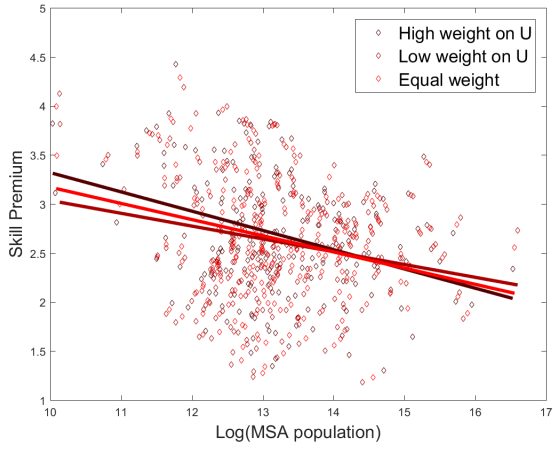
(a) Urban Wage Premium



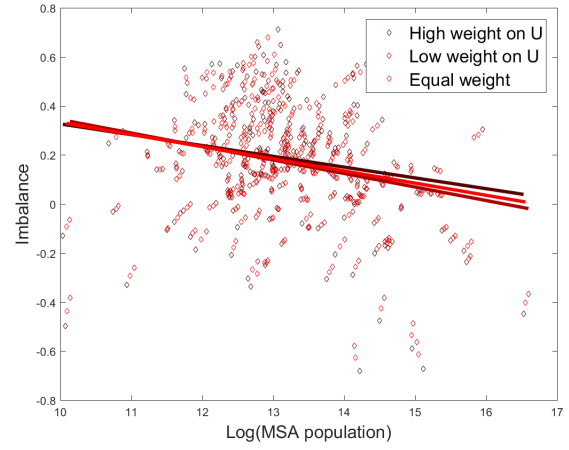
(b) Sorting



(c) Urban Skill Premium

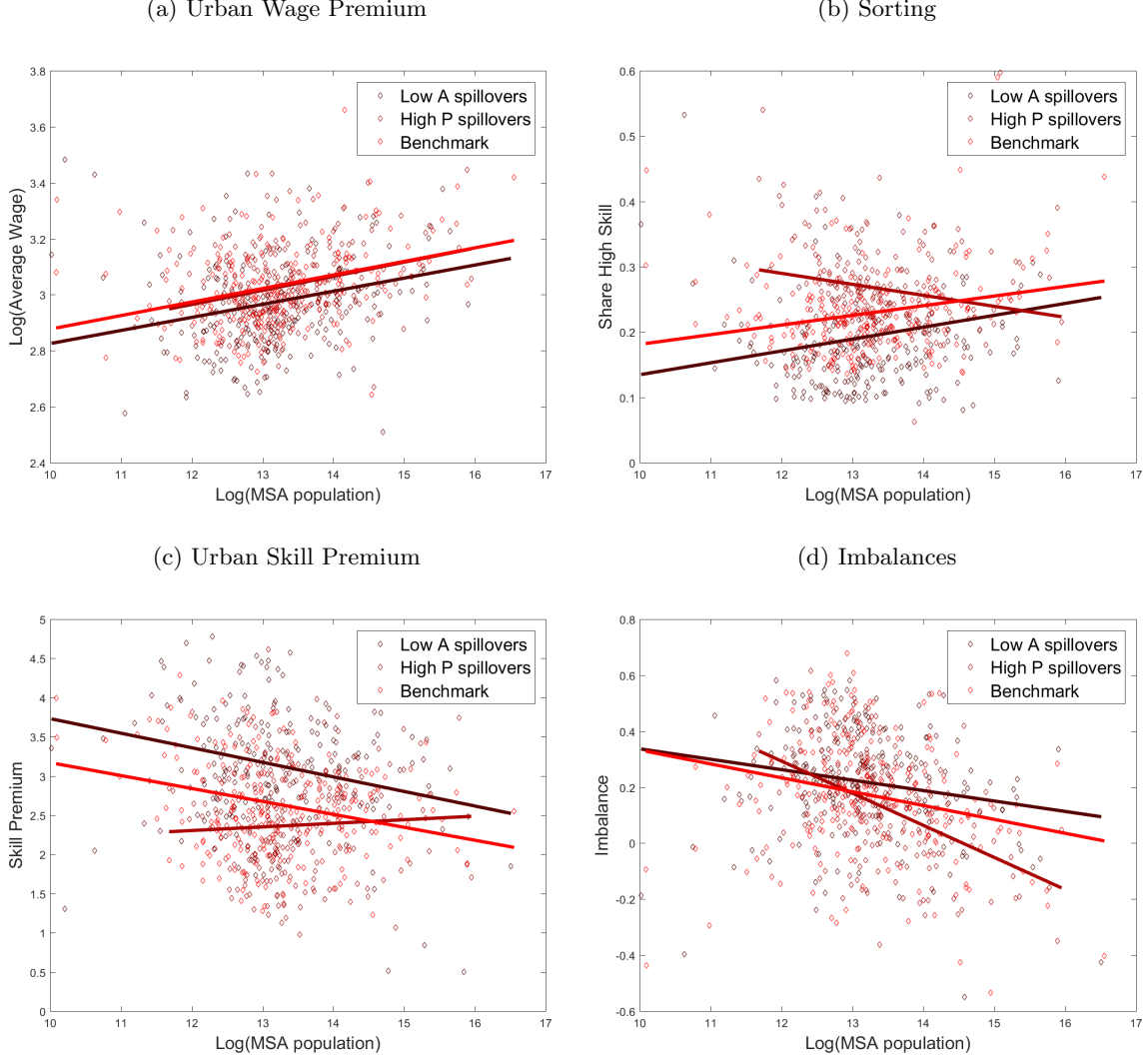


(d) Imbalances



Note: each panel reports outcomes across MSA's in the optimal allocation under alternative weights for each skill group in the planning problem. See the note to the previous graph.

Figure A.3: Urban Premia under Different Spillover Elasticities



Note: each panel reports outcomes across MSA's in the optimal allocation under alternative parametrizations of the spillover elasticities. “Low A spillovers” corresponds to amenity spillovers that are 50% lower than the benchmark. The welfare changes are $\hat{u}^U = \hat{u}^S = 1.033$ in that case. “High P spillovers” corresponds to efficiency spillovers that are 3.3 times larger than the benchmark. The welfare changes are $\hat{u}^U = \hat{u}^S = 1.073$ in that case. The case “benchmark” corresponds to the optimal allocation in Figure 7.