Optimal Spatial Policies, Geography and Sorting

Pablo D. Fajgelbaum
UCLA and NBER

Cecile Gaubert
UC Berkeley, CEPR and NBER

First Draft: December 2017

[Preliminary and Incomplete]

Abstract

We study optimal spatial policies in quantitative geography frameworks with spillovers and sorting of heterogeneous workers. Due to the spillovers, the spatial allocation is inefficient except under a particular set of net transfers across regions and workers. We first characterize spatially efficient transfers, as well as the labor taxes and subsidies that would implement them. Then, we quantify their aggregate and distributional effects in the U.S. economy. Relative to the current equilibrium, spatial efficiency requires stronger redistribution towards low-wage cities and leads to smaller spatial income disparities, including a smaller urban skill premium. By reaching the utility frontier, the welfare of both college and non-college workers increases by 3.4%.

1 Introduction

A classic idea in economics is that the concentration of economic activity leads to spillovers. Agglomeration economies drive productive efficiency and coexist with congestion forces such as traffic or pollution. Arguably, different groups of workers vary in how much they contribute to these spillovers and in how much they are impacted by them. These forces shape, in part, the uneven distribution of economic activity over space. In particular, they contribute to the strong spatial correlations observed across cities between employment density, average wages, share of skilled workers and skill wage premium. Being external in nature, it is likely that these forces shaping spatial disparities also lead to inefficient outcomes. Is the observed spatial distribution of economic activity inefficient? If so, what policies would guarantee efficiency? Would an optimal spatial distribution feature stronger, or weaker, spatial disparities and sorting than what is observed?

We develop and implement a new approach to provide quantitative answers to these questions. We rely on a framework that nests two recent strands of quantitative spatial research with spillovers: the economic geography models of Allen and Arkolakis (2014) and Redding (2016), and the location choice model with sorting of heterogeneous workers of Diamond (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 for the United States to quantify the impact of optimal net transfers across MSA's on the spatial allocation, wage inequality and welfare across low and high skill workers. relative to the current 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, labor and non-labor income may be arbitrarily distributed across agents and regions through redistributive policies and ownership patterns. 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 existence of an efficient competitive allocation. The condition generalizes the 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 of workers are not separable, these opportunity costs are measured in terms of local consumption expenditures. Moreover, they vary across locations due to the compensating differentials in the cost of living created by geographic forces (congestion, 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 distribution based on data on wages, expenditures, and spillover elasticities.

Our results imply that any particular policy implementing the spatially efficient allocation must deliver a specific set of optimal expenditures and transfers. We show that the optimal expenditure distribution can be implemented by labor taxes or subsidies coupled with lump-sum transfers. The first-best taxes or subsidies on labor income can be defined using just amenity and productivity spillover elasticities, and independently from micro features such as dispersion in fundamentals, trade frictions, housing supply elasticities, and product differentiation across locations (although these forces must be taken into account to quantify the impact of implementing the policy). Under standard constant-elasticity spillovers, spatial variation in optimal policies arises only if there are spillovers across types; for example, if the productivity of low skill workers depends on the presence of high skill workers in the same location. In this case, spatial efficiency may require more mixing of heterogeneous workers across cities than in the competitive allocation, and the decentralized pattern of sorting by skill can be too strong. In the absence of spillovers across types, the first best policies are invariant over space and take the form of constant labor income taxes or subsidies coupled with lump-sum transfers.

Before quantifying the optimal spatial allocation, we take a detour to apply these 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 $\eta\left(\overline{w}-w\right)$, where $\overline{w}$ 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 different implications for whether the optimal policy should redistribute income from high- to low-wage regions or vice-versa.

We show how to use the main framework to quantify efficient outcomes in 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 particular, we show 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 take the distribution of economic activity across MSA's in the U.S. in the year 2007 as our empirical setting. We allow for 2 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. To parametrize the spillover elasticities we rely on estimates from Diamond (2016), who estimates equations that are consistent with our model in a similar empirical setting.

We compute spatially efficient allocations under different Pareto weights for high and low skill workers. Our results suggest that, regardless of these weights, the current allocation features too little redistribution to low-wage cities. The net transfer, defined as the difference between disposable income and wage, 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 current allocation. However, there is also considerable dispersion in the distribution of growth rates for any given initial size.

In the optimal allocation, the ranking of cities by skill intensity is roughly preserved. However, due to positive spillovers from high to low skill workers, optimality requires a greater degree of mixing of different skills. As a result, the distribution of high-skill shares across cities is strongly compressed, in particular because larger cities feature a more balanced mix of high skill and low skill workers. High skill workers remain disproportionately attracted to larger cities, as they continue to value endogenous urban amenities relatively more, but the urban skill premium vanishes. The preliminary estimates also suggest that an optimal reallocation of economic activity increases the real consumption of both low and high skill workers by 3.4%.

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

[1]These 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

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. Optimal spatial policies have not been studied in this context. Instead, the standard applications tend to ignore place specific policies and impose restrictions on net transfers through assumptions on how the income generated by fixed factors is distributed across locations. For instance, Allen and Arkolakis (2014) and Redding (2016) assume away imbalances, while Helpman (1998) assume equal ownership of fixed factors, giving raise to trade surpluses in more productive locations. We show that these assumptions have strong welfare implications, and we characterize the distribution of imbalances that ensures spatial efficiency.

Our framework also includes the model of Diamond (2016), who estimates asymmetric amenity and efficiency spillovers across workers with different skills in a Rosen (1979)-Roback (1982) location model. 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).[2] Giannone (2017) argues that productivity spillovers across workers are relevant to understand the spatial effects of skill biased technological change. Here, we move from the positive to 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 efficiency.[3] Our results suggest that spatially efficient allocations would typically feature less spatial wage disparities and lower wage dispersion within 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), Gaubert (2015) and Ossa (2015) that study specific spatial policies such as taxes or subsidies. These studies compare the actual allocation to counterfactual ones with different distributions of wedges or policies, but do not consider the globally optimal spatial allocation or the spatial policies that implement it.[4] In addition, these models usually proceed under the assumption of a single worker type, while a key component of our analysis is the allocation of heterogeneous workers with asymmetric spillovers. Relative to the typical impact of removing dispersion in spatial wedges, our

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.

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

[3]Recent 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.

[4]In gravity frameworks, Allen et al. (2015) characterize to a first approximation the optimal zoning restrictions within a city and Ossa (2015) shows that no subsidies are superior to the dispersion in subsidies caused by local competition. Eeckhout and Guner (2015) numerically find a proportional income tax that maximizes welfare in a model without trade frictions, homogeneous labor and efficiency spillovers only.

results suggest considerable welfare effects from moving to an optimal allocation.

Starting with Flatters et al. (1974), a large theoretical literature in urban economics studies optimal city sizes.[5] Relative to that literature, we 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, there is no systematic prediction as to whether larger cities are too large or too small in a decentralized allocation. Our results suggest that reallocations towards currently smaller cities would be optimal on average, but there is considerable heterogeneity in optimal city growth at any given level of initial size.

Finally, our paper also relates to analyses of place-based policies. A large empirical literature evaluates the local impact of specific policies, such as government investments in public goods and regional subsidies. This body of research, summarized by Kline and Moretti (2014) and Neumark et al. (2015), typically measures elasticities of local outcomes to a policy. Glaeser et al. (2008) and Kline (2010) draw some of the general-equilibrium implications of these policies in specific environments. We study the normative general-equilibrium implications of some of these estimates in a quantitative spatial framework accounting for multiple indirect effects across economy.[6] While we focus on place-based taxes and subsidies, our approach can be easily adapted to accommodate optimal spatial investments in public goods.

3 Environment

3.1 Economic Geography Model with Sorting

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

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

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

[5] For a review of this extensive theoretical literature see sections 2 and 8 of Abdel-Rahman and Anas (2004). A common approach in this literature since Henderson (1974) is to assume that spillovers are inverted-U shaped within a city. Our main efficiency result in Proposition 1 is consistent with any functional form for the spillovers in either amenities or productive efficiency. Our quantification assumes log-linear spillovers, allowing us to draw from the existing empirical estimates that typically adopt this functional form.

[6] In his review of the policy implications of empirical economic-geography studies, Combes (2011) notes the lack of a general-equilibrium analysis of the optimal allocation of employment in a model of regional trade allowing for geographic inter-dependencies.

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). There is no heterogeneity across workers within each type. The utility function $U(c,h)$ is homogeneous of degree 1.[7]

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

$$\sum_j L_j^\theta = L^\theta, \tag{2}$$

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

We move on to the supply side. Every location produces traded and non-traded goods. Tradeable output uses the aggregate technology $Y_j\left(N_j^Y, I_j^Y\right)$ requiring services of labor N_j^Y and intermediates I_j^Y. Similarly, production in the non-traded sector is $H_j\left(N_j^H, I_j^H\right)$. The functions Y_j and H_j may be city-specific and feature constant or decreasing returns to scale. 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.[8] This framework allows in particular for different cities to feature different housing supply elasticities, shaped in the model by city-level decreasing returns to scale to the production of H.[9]

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

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

Goods in the traded sector can be shipped domestically or to other locations. The country's geography is captured by iceberg trade frictions $\tau_{ji} \geq 1$. Because of these frictions, $\tau_{ji}Q_{ji}$ units must be shipped from location j 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:

$$Y_j\left(N_j^Y, I_j^Y\right) = \sum_i \tau_{ji} Q_{ji}. \tag{4}$$

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 $Q(Q_{1i}, .., Q_{Ji})$. This bundle of traded commodities

[7] The analysis can be easily generalized to homothetic utility functions which are not homogeneous of degree 1 at the cost of more burdensome notation.

[8] To keep notations light, we do not feature land explicitly as a variable input. For simplicity we assume that the fixed factors in production cannot be reallocated between the traded and non-traded sectors. This assumption could be easily relaxed.

[9] If $H_j\left(N_j^H, I_j^H\right)$ is homothetic, so that it can be expressed as $H_j\left(N_j^H, I_j^H\right) = H_j^1\left(H_j^0\left(N_j^H, I_j^H\right)\right)$ for some $H_j^0\left(N_j^H, I_j^H\right)$ that is homogeneous of degree 1, then in the competitive allocation the elasticity of H_j^1 is the housing supply elasticity.

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:

$$Q\left(Q_{1i},..,Q_{Ji}\right)=\sum_{\theta}L_{i}^{\theta}c_{i}^{\theta}+I_{i}^{Y}+I_{i}^{H}. \tag{5}$$

The only feature left to specify is the mapping from the distribution of employment across types, $\left\{L_{j}^{\theta}\right\}$, 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

$$z_{j}^{\theta}=z_{j}^{\theta}\left(L_{j}^{1},..,L_{j}^{\Theta}\right) \tag{6}$$

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 $N\left(z_{j}^{1}L_{j}^{1},..,z_{j}^{\Theta}L_{j}^{\Theta}\right)$. This aggregator also captures imperfect substitution across workers. Feasibility in the use of labor services dictates that

$$N_{j}^{Y}+N_{j}^{H}=N\left(z_{j}^{1}L_{j}^{1},..,z_{j}^{\Theta}L_{j}^{\Theta}\right). \tag{7}$$

This completes the description of the environment. We highlight two key features of our model relative to an otherwise standard neoclassical framework. First, the location choice 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 $N\left(\cdot\right)$ may feature increasing returns to the number of workers in a particular group through $\left\{z_{j}^{\theta}L_{j}^{\theta}\right\}$. Both sources of non-convexities operate though the same margin, L_{j}^{θ}. These property will drive some key features of the welfare analysis and the quantitative implementation.

3.2 Competitive Equilibrium

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 $a_{j}^{\theta}\left(L_{j}^{1},..,L_{j}^{\Theta}\right)$ and $z_{j}^{\theta}\left(L_{j}^{1},..,L_{j}^{\Theta}\right)$.

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

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

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

$$u_j^\theta = a_j^\theta\left(L_j^1, .., L_j^\Theta\right) \frac{x_j^\theta}{\psi\left(P_j, R_j\right)}, \tag{9}$$

where $\psi\left(P, R\right)$ is the price index associated with U.

Since workers are homogeneous within each type, their utility is equalized. Worker's location choice implies that

$$u_j^\theta \leq u^\theta, \ = \text{if } L_j^\theta > 0. \tag{10}$$

Firms Producers of traded and non-traded commodities solve:

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

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

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 $\left\{w_j^\theta\right\}$, the wage per efficiency unit is the solution to a standard cost-minimization problem,

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

Importantly, this labor hiring decision takes the efficiency $z_j^\theta\left(L_j^1, .., L_j^\theta\right)$ 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 $\left\{z_j^\theta\right\}$:

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

In every location there is free entry to the shipping sector. A no-arbitrage condition implies the

price in location i of the traded good from j equals $\tau_{ji}p_j$. Given these prices, the trade flows are[10]

$$P_i \frac{\partial Q\left(Q_{1i},..,Q_{Jj}\right)}{\partial Q_{ji}} = \tau_{ji}p_j. \tag{14}$$

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

Expenditure Function The only component of the competitive allocation left to define is the per capita expenditure x_j^θ. By definition, x_j^θ equals to the labor and non-labor income of a type-θ worker who lives in j net of government taxes and transfers. Rather than imposing a specific set of tax instruments and transfers, we formulate a general expenditure function $\chi_j^\theta\left(w_j^\theta,...\right)$ that maps the wage of type-θ workers in location j and potentially other outcomes across the economy to disposable income:

$$x_j^\theta = \chi_j^\theta\left(w_j^\theta,..\right). \tag{15}$$

We treat the expenditure function $\chi_j^\theta\left(\cdot\right)$ as a primitive. It must be positive and satisfy the restriction that aggregate consumer expenditures equal value added:

$$\sum_j \sum_\theta L_j^\theta \chi_j^\theta\left(w_j^\theta,..\right) = \sum_j \left[\left(\sum_\theta L_j^\theta w_j^\theta\right) + \Pi_j^Y + \Pi_j^H\right]. \tag{16}$$

For example, a *laissez-fare* economy where each worker θ owns a fraction b^θ of the national portfolio of profits and returns to land on top of the wage implies $\chi_j^\theta\left(w_j^\theta,..\right) = w_j^\theta + b^\theta\Pi$, where $\Pi = \sum_j \Pi_j^Y + \Pi_j^H$. By Walras law the government budget constraint holds in any competitive equilibrium.

We have all the ingredients to define the competitive allocation.

Definition 1. *A competitive equilibrium 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)*; 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 $\left\{L_j^\theta\right\}$, consumption of traded and non-traded commodities $\left\{c_j^\theta, h_j^\theta\right\}$, trade flows $\{Q_{ij}\}$ and allocation of efficiency units and

[10] Because the aggregator of traded commodities $Q\left(\cdot\right)$ has constant returns to scale, it is equivalent to assume that consumers and firms choose the optimal mix of traded commodities or that a competitive sector purchases traded commodities and combines them into a final bundle that is sold to local consumers and firms.

intermediate inputs, $\left\{N_j^Y, I_j^Y, N_j^H, I_j^H\right\}$. The planner implement policies that treat all individuals within a group in the same way and is therefore 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 $\left\{u^\theta\right\}$. The optimal planning problem consists in reaching the frontier of that set,

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

The set of solutions to this this problem given all feasible values of $\underline{u}^{\theta'} \in \mathcal{U}$ defines the utility frontier. Competitive equilibria according to Definition 1 may not correspond to a solution to the planning problem due to *spatial inefficiencies*: workers do not internalize the impact of their location choice on amenities through a_j^θ and firms do not internalize the impact of their hiring decisions on efficiency through z_j^θ. We turn in the next section to the solution of the planning problem.

4 Optimal Spatial Allocation and 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 the expenditure function $\chi_j^\theta(\cdot)$. Finally, we identify tax and transfers policies that deliver the desired optimal expenditure function in a market equilibrium. Formal derivations are detailed in Appendix A.1.

4.1 Efficiency Condition and Optimal Expenditure Function

The full planning problem is detailed in Definition (2) of Appendix (A.1). Using the planner's optimality conditions and comparing them to the decentralized allocation we obtain the following result.

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

$$W_j \frac{dN_j}{dL_j^\theta} + \sum_{\theta'} \frac{x_j^{\theta'} L_j^{\theta'}}{a_j^{\theta'}} \frac{\partial a_j^{\theta'}}{\partial L_j^\theta} = x_j^\theta + E^\theta, \quad \text{if } L_j^\theta > 0, \tag{17}$$

for all j and θ and some constants $\left\{E^\theta\right\}$. Furthermore, if the planner's problem is globally concave and (17) *holds for some specific $\left\{E^\theta\right\}$, then the competitive equilibrium is efficient.*

Condition (17) defines a relationship between expenditure per capita and employment that must hold in any efficient allocation. It reflects an optimal balance between the marginal benefit and cost of a marginal type-θ worker in j.

On the left-hand side, the first term is labor's marginal product, capturing both the direct output effect as well as the productivity spillovers.[11] The second term captures the marginal benefits (or costs if negative) on local amenities, equal to the marginal amenity spillover generated on each group of workers living in j, measured in consumption equivalent terms. In populated locations, these marginal benefits equal the marginal costs in the right hand side. An extra type-θ worker in j requires x_j^θ units of expenditures and has an opportunity cost of being employed elsewhere equal to E^θ, the multiplier of the national market clearing constraint (2).

We can draw several useful implications from this result. First, it reflects that the set of equations defining the competitive allocation coincides with the set defining the the planner's allocation, except potentially for the expenditure distribution.[12] The expenditure distribution is given by (17) in the planning problem and by (15) in the competitive allocation, respectively. Therefore, asking whether the spatial allocation is efficient is equivalent to asking whether the x_j^θ's in the market allocation satisfy (17). As shown in Corollary 1, this property motivates an efficiency test of the observed allocation based on wage and expenditure data. A further implication is that market inefficiencies can be tackled through policies acting directly on x_j^θ. We construct such policies in the next section.[13]

Second, Proposition 1 extends a familiar efficiency condition to spatial environments. In this spatial economy, "space" enters through trade costs, non-traded goods, congestion and amenities. In the absence of these forces, there would be no compensating differentials across locations (i.e., no spatial variation in expenditure per capita x_j^θ). In that case, (17) would collapse to the familiar efficiency condition, central in analyses of misallocation in the spirit of Hsieh and Klenow (2009), that the marginal value-product of labor be equalized across production units.[14] Under constant-elasticity production functions, that condition motivates the common practice of inferring allocative inefficiencies from differences in average income. As we just mentioned, Proposition 1 shows that, in quantitative geography environments, the distribution of expenditure per capita includes additional information that can be used to assess efficiency.

Finally, even if (17) holds, there could be inefficient market equilibria consistent with (17).

---

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

[12] The market allocation of efficiency units $\left\{N_j^H, N_j^Y\right\}$, intermediates $\left\{I_j^H, I_j^Y\right\}$, 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, respectively. In the planner's allocation, x_j^θ coincides with shadow value of the local goods assigned to each θ worker in j. See the proof of Proposition 1.

[13] This compartmentalization of the inefficiencies reflects a broader "principle of targeting" noted by Bhagwati and Johnson (1960) in trade-policy contexts and by Sandmo (1975) and Dixit (1985) in economies with external effects.

[14] Assuming away trade frictions ($\tau_{ij} = 1$, implying $P_j = P$), non-traded goods ($U\left(c_j^\theta, h_j^\theta\right) = c_j^\theta$), congestion ($\frac{\partial a_j^{\theta'}}{\partial L_j^\theta} = 0$) and preferences for location ($a_j^\theta = a_j$) then the model is equivalent to one describing the allocation of workers across firms selling differentiated products as in Hsieh and Klenow (2009). In that case, the spatial mobility constraint implies equalization of nominal income across all locations, $\frac{x_j^\theta}{P^\theta} = u^\theta$. As a result, (A.1) collapses to $W_j \frac{dN_j}{dL_j^\theta} = u^\theta P^\theta + E^\theta$ for all j. I.e., for each worker type the value of the marginal product is the same across all productive units j.

However, these cases can be ruled out when the social planner's problem is globally concave. In this 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 under the appropriate expenditure function, in that case only one competitive equilibrium exists and it must be efficient.[15] In Section 4.4 we introduce conditions for global concavity of the planner's problem.

In the definition of the competitive equilibrium we allowed for a general expenditure function $\chi^{\theta}(\cdot)$, defined in (15) as a primitive of the model. Proposition 1 implies that efficient allocations can be implemented by appropriately choosing this function. Technically, this is the case if the solution to the planner's welfare maximization problem is interior, which we assume from now on.[16] We now identify the optimal expenditure function as a stepping stone to identifying optimal policies in the next section. Combining Proposition 1 with the labor-demand condition (13) we reach the following corollary.[17]

Corollary 1. *Let* $\gamma_{\theta,\theta'}^{P,j} \equiv \frac{L_j^{\theta'}}{z_j^{\theta'}} \frac{\partial z_j^{\theta'}}{\partial L_j^{\theta}}$ *and* $\gamma_{\theta,\theta'}^{A,j} \equiv \frac{L_j^{\theta'}}{a_j^{\theta'}} \frac{\partial a_j^{\theta'}}{\partial L_j^{\theta}}$ *be the productivity and amenity spillover elasticities. Also let* $\mathbf{w}_j = \left\{w_j^1, ..., w_j^{\Theta}\right\}$ *be the wage distribution and* $\mathbf{\Omega}_j = \left\{w_j^1 L_j^1, ..., w_j^{\Theta} L_j^{\Theta}\right\}$ *be the wage bill distribution in location* j *in a given competitive allocation. Then, the expenditure function* $\chi_j^{\theta}(\mathbf{w}, \mathbf{\Omega})$ *implicitly defined as*

$$\left(1 - \gamma_{\theta,\theta}^{A,j}\right) \chi_j^{\theta}(\mathbf{w}, \mathbf{\Omega}) \equiv \left[1 + \gamma_{\theta,\theta}^{P,j} + \sum_{\theta' \neq \theta} \left(\gamma_{\theta,\theta'}^{P,j} + \frac{\gamma_{\theta,\theta'}^{A,j}}{w^{\theta'}} \chi_j^{\theta'}(\mathbf{w}, \mathbf{\Omega})\right) \frac{w^{\theta'} L^{\theta'}}{w^{\theta} L^{\theta}}\right] w^{\theta} - E^{\theta} \tag{18}$$

for some specific $\left\{E^{\theta}\right\}$ *implements an efficient competitive equilibrium.*

If $\chi_j^{\theta}(\mathbf{w}, \mathbf{\Omega})$ is defined according to (18) then, in equilibrium, the expenditure distribution is $x_j^{\theta} = \chi_j^{\theta}(\mathbf{w}_j, \mathbf{\Omega}_j)$ and condition (17) is guaranteed to hold. The optimal expenditure function $\chi_j^{\theta}(\mathbf{w}, \mathbf{\Omega})$ varies across types and locations through the *spillover elasticities* $\left\{\gamma_{\theta,\theta'}^{A,j}, \gamma_{\theta,\theta'}^{P,j}\right\}$. These elasticities capture the effect of type-θ workers on the efficiency and experienced amenities of type-θ' workers, respectively.

4.2 First-Best Subsidies

Using the optimal expenditure function we can find first-best policies. For that we make our first functional-form assumptions. Specifically, from now on we assume that the spillover functions

[15] In that case there could still be multiple market equilibria for alternative expenditure distributions that do not satisfy (17).

[16] The constant-elasticity spillover functions along with the conditions for concavity of the planner's problem imposed in section 4.4 will ensure that this is the case.

[17] To reach 18 we use that, from (13), the competitive wage w_j^{θ} equals the marginal product of a worker without taking into account efficiency spillovers. As a result, the marginal product of labor, $\frac{dN_j}{dL_j^{\theta}}$, can be expressed as function of wages and the spillover elasticities.

$z_j^\theta(\cdot)$ and $a_j^\theta(\cdot)$ take the constant-elasticity forms:

$$z_j^\theta\left(L_j^1,..,L_j^\Theta\right) = Z_j^\theta \prod_{\theta'} \left(L_j^{\theta'}\right)^{\gamma^P_{\theta',\theta}}, \tag{19}$$

$$a_j^\theta\left(L_j^1,..,L_j^\Theta\right) \equiv A_j^\theta \prod_{\theta'} \left(L_j^{\theta'}\right)^{\gamma^A_{\theta',\theta}}. \tag{20}$$

These functional forms are standard in studies that estimate spillover elasticities, and we adopt them here with the quantitative application in mind.[18]

We refer to $\left\{Z_j^\theta, A_j^\theta\right\}$ as the fundamental components of productivity and amenities in location j. The Z_j^θ capture exogenous comparative advantages in production across types. For future reference, whenever we assume that there is a single worker type ($\Theta = 1$) we omit the subscripts and simply denote these elasticities by γ^P and γ^A.

To construct policies that implement the optimal expenditure distribution in (18) we apply the standard Pigovian principle of subsidizing (taxing) the generators of net positive (negative) externalities so that they face the right prices. We posit a labor income subsidy (or tax if negative) $\tau_j^\theta = \tau^\theta\left(\mathbf{w}_j, \mathbf{\Omega}_j\right)$ to tackle spatial inefficiencies, coupled with type-specific lump-sum transfer T^θ to address distributional concerns without introducing further distortions.[19] Assuming that each type-θ worker owns a fraction b^θ of the national returns to fixed factors $\Pi = \Pi^Y + \Pi^H$, the expenditure function introduced in (15) becomes

$$\chi_j^\theta\left(w_j^\theta, ..\right) = \left(1+\tau_j^\theta\right) w_j^\theta + \left(b^\theta \Pi + T^\theta\right). \tag{21}$$

To attain efficiency we must then choose $\left\{\tau_j^\theta, T^\theta\right\}$ such that (18) holds. As a result the optimal labor subsidy is

$$\tau^\theta\left(\mathbf{w}, \mathbf{\Omega}\right) \equiv \frac{\gamma^P_{\theta,\theta} + \gamma^A_{\theta,\theta} + \sum_{\theta' \neq \theta}\left(\gamma^P_{\theta,\theta'} + \frac{\gamma^A_{\theta,\theta'}}{w^{\theta'}}\chi^{\theta'}\left(\mathbf{w}, \mathbf{\Omega}\right)\right)\frac{w^{\theta'}L^{\theta'}}{w^\theta L^\theta}}{1-\gamma^A_{\theta,\theta}}, \tag{22}$$

where $\chi^{\theta'}\left(\mathbf{w}, \mathbf{\Omega}\right)$ is defined in (18).[20]

Many details and micro features of the environment do not matter to define the first-best policy. In particular, neither the spatial distribution of fundamental productivity and amenities $\left\{Z_j^\theta, A_j^\theta\right\}$, the bilateral trade costs $\{\tau_{ij}\}$, the utility function $U(\cdot)$, the production functions $N(\cdot)$, $Y(\cdot)$ and $H(\cdot)$, nor the aggregator of trade flows $Q(\cdot)$ matter for the shape of the policy.

[18] We assume that the spillover elasticities do not vary across locations. This restriction would be straightforward to relax in both the theory and the application to the U.S. economy. However, to the best of our knowledge there is no empirical evidence about heterogeneous spillover elasticities across U.S. cities that we could use to discipline heterogeneity in this dimension, and therefore we refrain from that level of generality.

[19] This policy shifts labor supply. Equivalently, we could apply the policy to labor demand through a hiring subsidy of $\frac{\tau^\theta(\mathbf{w},\mathbf{\Omega})}{1+\tau^\theta(\mathbf{w},\mathbf{\Omega})}$ in (13).

[20] The optimal lump-sum transfers are $T^\theta = -\frac{E^\theta}{1-\gamma^A_{\theta,\theta}} - b^\theta\Pi$.

Instead, the optimal subsidy $\tau^{\theta}(\mathbf{w}, \mathbf{\Omega})$ inherits the properties of the optimal expenditure function defined in (18). As a result, it can be expressed as simply a function of wages, wage bills, and spillover elasticities.[21] However, the multiple dimensions of spatial heterogeneity and details of the environment will matter to determine the equilibrium wage distributions, and therefore the gains of implementing this policy.

We can use (22) to gain insight about the shape of optimal subsidies in special cases. Under the assumptions in (19) and (20), so that spillover elasticities are constant across locations, spatial variation in policies arises due to spillovers across types. In the absence of spillovers across types ($\gamma^{P}_{\theta,\theta'} = \gamma^{A}_{\theta,\theta'} = 0$ for all $\theta \neq \theta'$) the optimal labor subsidy is:

$$\tau^{\theta} = \frac{\gamma^{P}_{\theta,\theta} + \gamma^{A}_{\theta,\theta}}{1 - \gamma^{A}_{\theta,\theta}}. \tag{23}$$

Under negative congestion spillovers ($\gamma^{A}_{\theta,\theta} < 0$), if the agglomeration spillover is not too strong ($\gamma^{P}_{\theta,\theta} < -\gamma^{A}_{\theta,\theta}$), then type-$\theta$ workers should pay the same tax $\tau^{\theta} < 0$ equal to the net marginal congestion everywhere. Despite being constant across space, in this case the optimal labor income subsidy τ^{θ} still distorts the allocation because it affects the relative expenditures across locations for a given type.

To gain intuition about the role of spillovers across types, consider a polar case without amenity spillovers ($\gamma^{A}_{\theta,\theta'} = 0$ for all θ, θ') and no efficiency spillovers on the own type ($\gamma^{P}_{\theta,\theta} = 0$). Then, (22) becomes

$$\tau^{\theta}(\mathbf{\Omega}) = \gamma^{P}_{\theta,\theta'} \left(\frac{w^{\theta'} L^{\theta'}}{w^{\theta} L^{\theta}} \right). \tag{24}$$

In this special case the subsidy varies across locations according to the distribution of wage bills. A positive efficiency spillover of type θ on type θ' ($\gamma^{P}_{\theta,\theta'} > 0$) implies a higher marginal gain from attracting type-θ workers to locations with larger prevalence of type θ', as captured by their relative wage bill, and therefore a higher subsidy. This observation serves to assess the efficiency of observed spatial sorting patterns. Positive spillovers across types call for a degree of mixing across types that is not internalized by the market. Therefore, the *laissez-faire* equilibrium may feature too much dispersion in labor shares across cities. For example, in a case with only unskilled and skilled workers denoted by $\theta = U, S$ where only skilled workers generate spillovers on the other type, the optimal subsidies are $\tau^{S} = \gamma^{P}_{\theta,\theta'} \left(\frac{w^{U} L^{U}}{w^{S} L^{S}} \right)$ and $\tau^{U} = 0$.[22] Relative to a *laissez-faire* equilibrium and under standard assumptions, this policy introduces a subsidy to skilled workers where they are more scarce.[23] Therefore, it is likely to temper the dispersion in skill shares across locations.

[21] It also depends on the constants $\{E^{\theta}\}$ if and only if $\gamma^{A}_{\theta,\theta'} \neq 0$ for $\theta' \neq \theta$, given the implicit definition of $\chi^{\theta}(\mathbf{w}, \mathbf{\Omega})$ in (18).

[22] More generally, in this case $\frac{d\tau^{S}(\mathbf{\Omega})}{d\tau^{U}(\mathbf{\Omega})} < 0 \longleftrightarrow \gamma^{P}_{S,U}\gamma^{P}_{U,S} > 0$, so that subsidies are negatively correlated if both types generate positive efficiency spillovers on the other type.

[23] For example, under a CES aggregator of skills, $N = \left(\left(z^{U} L^{U} \right)^{\rho} + \left(z^{S} L^{S} \right)^{\rho} \right)^{\frac{1}{\rho}}$ we obtain $\frac{w^{S}_{j} L^{S}_{j}}{w^{U}_{j} L^{U}_{j}} = \left(\frac{z^{S}_{j}}{z^{U}_{j}} \frac{\left(L^{S}_{j} \right)^{1-\gamma_{S,U}}}{L^{U}_{j}} \right)^{\rho}$. Therefore, the relative wage bill of skilled workers increases with L^{S}_{j} if $\gamma_{S,U} < 1$ and decreases

In Section 6 we will return to these properties. They will helps us rationalize our 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.

4.3 Optimal Transfers in Trade and Economic Geography Models

The environment laid out in Section 3 includes standard economic geography models, such as Helpman (1998), Allen and Arkolakis (2014) and Redding (2016).[24] These models are the backbone of a growing body of quantitative research studying the spatial implications of regional shocks such as infrastructure investments and regional productivity changes, summarized by Redding and Rossi-Hansberg (2017).[25] We now apply the previous results to shed light on the optimal transfers and policies in these models. Our key finding is that a simple linear relationship between trade imbalances and wages implements the efficient allocation regardless of micro-level heterogeneity in amenities, productivity, and trade costs across regions.

To specialize the framework to these models we assume a single worker type and Cobb-Douglas preferences with weight α_C on the traded goods:

$$U(c,h) = c^{\alpha_C} h^{1-\alpha_C}. \tag{25}$$

Therefore, the common component of utility 1 becomes $u_j = A_j L_j^{1+\gamma^A} c_j^{\alpha_C} h_j^{1-\alpha_C}$. Production only uses labor, 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. As before, trade frictions are captured by $\{\tau_{ij}\}$. In a competitive allocation without subsidies or transfers, workers in j receive a wage equal to tradeable output per worker, $w_j = Z_j L_j^{\gamma^P}$. The existing applications also make different assumptions about how, before taxes and transfers, the returns to the fixed factor H_j are allocated across workers, ranging from local redistribution as in Redding (2016) to equal ownership across workers in Helpman (1998). To nest these cases we 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.

After 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 $\omega = 1$.[26]

---

with L_j^U.

[24] Our presentation assumed that each location j sells a homogeneous tradeable product under perfect competition. This presentation is consistent with Allen and Arkolakis (2014). We show in Appendix A.3 that the normative implications of this model are equivalent to those of a differentiated-products model with free entry of monopolistic producers but without productivity spillovers. In 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 a homogeneous-product model with spillover elasticity equal to $\gamma^P = \frac{1}{\kappa - 1}$. Therefore, our analysis also encompasses the models in Helpman (1998) and Redding (2016).

[25] Some 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).

[26] We remain more general than these models in that the trade-flows aggregator $Q(Q_{1i}, .., Q_{Ji})$ is not restricted

Applying Corollary 1 we obtain the optimal expenditure distribution. Using that preferences are Cobb-Douglas, we find that the net transfer received by workers in j in the optimal allocation is:

$$P_j c_j - w_j = \eta \left(\bar{w} - w_j \right), \tag{26}$$

where $\bar{w} \equiv \sum w_i \frac{L_i}{L}$ is the average wage and $\eta \equiv 1 - \frac{\alpha_C \left(1+\gamma^P\right)}{1-\gamma^A}$ is a constant that combines the key elasticities.[27] When $\eta > 0$, agglomeration forces dominate congestion effects. In contrast, when $\eta < 0$, congestion dominates, and the model fits the sufficient concavity condition from Proposition 2 below. In that case, there is a unique competitive allocation consistent with (26).

Condition (26) implies that, except for knife-edge cases on the parameters ($\eta = 0$) or the fundamentals (such that $w_j = w$), the efficient allocation features trade imbalances. For example, when $\eta < 0$ the efficient trade imbalances imply a pattern of redistribution from high- to low-income regions: optimal transfers decrease with wages and are positive in locations with below-average wage, $w_j < w$. These predictions hold for any underlying distribution of amenities, productivities, and trade costs $\{A_j, Z_j, \tau_{ij}\}$. Therefore, regardless of the distribution of fundamentals, the efficiency of an observed allocation in data generated by these models can be determined based on whether the observed trade imbalances per capita align with (26).

What policy implements the optimal trade imbalances (26)? Proceeding as in Section 4.2, the optimal policy can be expressed as a labor subsidy coupled with a lump-sump payment, so that a worker located in j receives a net transfer of $\tau w_j + T$. The optimal subsidy is common across locations and equal to

$$\tau = \frac{\gamma^P + \gamma^A}{1 - \gamma^A} - \frac{1 + \gamma^P}{1 - \gamma^A} \left(1 - \alpha_C\right) \omega, \tag{27}$$

with lump-sum transfer of $T = -\tau \bar{w}$.[28] Except for knife-edge values of the elasticities, the optimal policy may entail a labor subsidy with lump-sum tax (if $\tau > 0$), or a constant labor tax with a lump-sum payment ($\tau < 0$). The net payoff of the policy to each worker in region j is $\tau \left(w_j - \bar{w}\right)$, so that optimal policies entail redistribution away from regions with low wages if $\tau > 0$, or into these regions if $\tau < 0$.

We are now able to identify the optimal policy in different models. We consider the case when $\eta < 0$. In Helpman (1998), who assumes even ownership ($\omega = 0$), the default transfers to low-income regions in the competitive equilibrium are too large form the perspective of spatial efficiency; therefore, to attain (26) 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, to attain (26) the optimal policy redistributes income to low-wage regions

to the constant-elasticity (CES) form. Therefore, we do not assume a constant trade elasticity. The trade elasticity turns out to have no bearing on the optimal spatial transfers.

[27] Net 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.

[28] The only difference between this optimal subsidy and (23) is that here the optimal policy is derived under the possibility of local redistribution ($\omega \geq 0$).

$(\tau < 0)$.

4.4 Quantitative Implementation

We come back now to the main model and its quantitative implementation. We first impose further functional-form assumptions, under which 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 expenditure distribution (18). Then, we define the set of data necessary to identify the model and implement the optimal policies given elasticities.

Functional Forms So far, the only functional form assumptions that we imposed are the constant-elasticity forms for the spillovers in (20) and (19). For the quantitative implication of the model, we impose, in addition, that preferences for traded and non-traded goods are Cobb-Douglas as in (25). The remaining functional forms are assumed to have constant elasticities of substitution. In particular, the production functions of traded and non-traded goods are Cobb-Douglas in labor and intermediate inputs:

$$Y_j\left(N_j^Y, I_j^Y\right) = z_j^Y \left(N_j^Y\right)^{1-b_{Y,j}^I} \left(I_j^Y\right)^{b_{Y,j}^I}, \tag{28}$$

$$H_j\left(N_j^H, I_j^H\right) = z_j^H \left(\left(N_j^H\right)^{1-b_{H,j}^I} \left(I_j^H\right)^{b_{H,j}^I}\right)^{\frac{1}{1+d_{H,j}}}, \tag{29}$$

where $d_{H,j} \geq 0$. Traded goods are produced under constant returns to scale, but we allow for decreasing returns in the housing sector. The coefficient $d_{H,j}$ is the inverse housing supply elasticity of location j, allowed to vary across regions.[29] The aggregator of labor types is CES,

$$N_j\left(L_j^1.., L_j^\Theta\right) = \left[\sum_\theta \left(z_j^\theta \left(L_j^1, .., L_j^\Theta\right) L_j^\theta\right)^\rho\right]^{\frac{1}{\rho}}, \tag{30}$$

and so is the aggregator of traded commodities:

$$Q\left(Q_{1i}, .., Q_{Ji}\right) = \left(\sum_i Q_{ji}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}. \tag{31}$$

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

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

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

[29] I.e., optimization of producers of non-traded goods implies $\ln R_j \propto d_{H,j} \ln H_j$.

Also, we let $D = \min_j \{d_{H,j}\}$ summarize the decreasing returns. Under the functional form assumptions made so far we have the following property.

Proposition 2. *The planning problem is concave if*

$$1 + \Gamma^A > \frac{1 + \Gamma^P}{1 + D}, \tag{32}$$

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

$$1 + \gamma^A > \left(1 + \gamma^P\right) \left[(1 - \alpha_C) \frac{1}{1 + d_{H,j}} + \alpha_C \frac{1}{1 + d_Y}\right] \tag{33}$$

for all j.

The first condition establishes concavity by imposing that congestion forces, driven by amenity spillovers and curvature in production, are at least as large as agglomeration forces. In particular, $\Gamma^A > \Gamma^P$ is sufficient. In that case, 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.[30]

Condition (32) may also apply when there is a single type, in which case it simplifies to $1 - \gamma^A > \frac{1+\gamma^P}{1+D}$. However, with a single type, further assuming Cobb-Douglas preferences over traded and non-traded goods we obtain the restriction (33) involving the fraction of expenditures in the non-traded sector $1 - \alpha_C$. These cases apply to the economic-geography models we reviewed in the introduction, which typically assume a single worker type.[31]

Finally, note that proposition (2) establishes conditions under which the market allocation is unique under optimal transfers. It extends existing results in the literature that focus on characterizing uniqueness of spatial equilibrium under the assumption of trade balance. It also extends uniqueness results to the context of several worker groups with cross-groups spillovers.

Implementation in Changes and Data Requirements To quantify the efficient allocation we cast the planning problem relative to an initial equilibrium consistent with Definition (1) given an arbitrary expenditure function. We solve for a planner's problem that optimizes over the changes in the expenditure of each type of worker and in the remaining endogenous variables subject to the market clearing constraints. The solution to this problem in relative changes is equivalent to the solution of the planning problem in levels from Section (3.3), assuming that the same parameters and fundamentals underlie the equilibrium of Definition (1) and the planning problem in levels. However, in the spirit of Dekle et al. (2008), by expressing the solution in changes we can then identify the data that is sufficient to back out fundamentals and characterize the model.

[30] When $\Gamma^A > \Gamma^P$, concavity of the problem is ensured for any concave $H_j\left(N_j^H, I_j^H\right)$ and $Y_j\left(N_j^Y, I_j^Y\right)$; in that case, assuming the specific functional forms (29) and (28) is not necessary.

[31] Regardless of the number of types, the CES restriction (31) 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.

We define the planning problem in changes in Appendix (A.2). Our discussion in that section implies the following.

Proposition 3. *Assume that observed data is generated by a competitive equilibrium consistent with Definition* (1) *under an arbitrary expenditure function* χ_j^θ *in* (15)*, and under the functional forms* (19)*,* (20)*,* (25) *and* (28) *to* (31)*. Then, relative to the initial equilibrium, the optimal quantities and prices can be characterized as function of:*

i) distributions of wages, employment and expenditures across labor types and locations;

ii) distribution of bilateral import and export shares across locations;

iii) utility and production function parameters $\left\{\alpha_C, \sigma, \rho, b_{Y,j}^I, b_{H,j}^I, d_{H,j}\right\}$*; and*

iv) spillover elasticities $\left\{\gamma_{\theta',\theta}^A, \gamma_{\theta',\theta}^P\right\}$.

The proposition establishes that we can characterize the optimal allocation given the observed data in i) and ii) and the elasticities in iii) and iv). We do not impose restrictions on the distributional policies in the initial equilibrium. The efficiency of the observed allocation depends on whether the distribution of expenditures, which accounts for existing policies that generate transfers between cities, lines up with (18). We also highlight that, by following this approach, we ensure that the model is initially parametrized in a way that exactly matches the outcomes enumerated in i) and ii) above. In particular, the productivities and preferences across workers and locations $\left\{Z_j^\theta, A_j^\theta\right\}$, TFP shifters $\left\{z_j^Y, z_j^H\right\}$, and bilateral trade costs $\{\tau_{ij}\}$ are implicitly 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 to fulfill the data requirements in points i) and ii) of Proposition 3. Then, in Section 5.2 we describe the calibration strategy for 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 Main Data Sources and Stylized Facts

To measure income and expenditure at the MSA level, we start by combining two datasets. First, the BEA's Regional Accounts report every year, at the MSA level, individual income split between labor income, capital income and welfare transfers. Second, a supplemental dataset compiled by the BEA for the years 2000 to 2007 reports total taxes paid by individuals at the MSA level (Dunbar, 2009). We use income net of taxes and transfers, corresponding to the BEA's definition of disposable

income, to match the notion of expenditure in the model.[32] We apportion total personal taxes paid at the MSA level between labor and capital income proportionally to the respective income level.

Second, we use the March supplement of the Current Population Survey, part of the Integrated Public Use Microdata Series (Flood et al., 2017), to split each of these MSA-level totals into two groups: high skill and low skill. Following Diamond (2016) we define high skill workers as workers who have completed at least four years of college, while the remaining workers are classified as low skill. For each individual the CPS reports the skill level, the MSA of residence, income by source (labor, capital and transfers), and taxes paid. We aggregate this information to the MSA level using weights from the CPS to estimate the share of population, income (by source) and tax liabilities that accrue to high and low skill workers in each MSA.[33]

Third, we compute a series of MSA-group level measures of detailed socio-demographic characteristics. Our goal here is to control for heterogeneity within groups that is not part of the model, but which we can observe in the data. We calibrate our model using measures that take out this heterogeneity in order to avoid attributing to a place (MSA) differences in the local composition of the high skill and low skill groups. To that end, we use Census data (IPUMS) on age, detailed educational attainment, sectoral composition, racial composition, and labor force participation in each MSA-group. We project our measures of per-capita labor income, capital income, transfers and taxes paid at the MSA-group level on these characteristics. We compute predicted values for income, taxes and transfers computed at the mean across MSAs of these socio-demographic characteristics. We use these values, stripped off of within group-MSA heterogeneity, for per-capita labor income, transfers and taxes to calibrate the model. Since there is no source of variation in capital income across cities within the model, we use a common level of capital income for each group to calibrate capital income, as well as taxes paid on this income, at the MSA-group level. The share of total capital income that accrues to each group comes from the procedure above, while the level is computed to match the calibrated value of profits in the economy as described below.

Fourth, we also use information reported in the CPS on the sector of activity of workers to measure, at the MSA level, the share of workers in the non traded sector.[34] This measure is not group-specific. We again remove unmodelled heterogeneity in this measure, following a similar approach to the one described above.

Finally, quantifying the model requires data on trade flows between MSAs. The Commodity Flow Survey (CFS) reports the flow of manufacturing goods shipped between CFS zones in the US every five years. The CFS zones correspond to geographic units that are larger than our unit of observation, the MSA. To overcome this data limitation, we adapt the approach taken by Allen and Arkolakis (2014), who use estimates of trade frictions as function of geography to project CFS-

[32] In the BEA publications, disposable income is defined as the "income available for spending", which matches well to our definition of expenditure.

[33] Since the March CPS supplement is not available for the smallest MSAs, our dataset covers 245 MSAs in the continental US that correspond to 97% of the population and 98% of Income of all US urban areas. Urban areas in the US cover 83% of the US population, and 87% of US personal income.

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

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 (1) actual distance between MSAs, (2) estimates of trade frictions with respect to distance taken from Allen and Arkolakis (2014), and (3) the regional trade imbalances in the data.[35]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.

5.2 Calibration

With this MSA-group level dataset at hand, we use equilibrium relationships in the model and calibrated elasticities to construct the variables needed for the model quantification. Whenever possible we rely on Diamond (2016) estimates. Our model generalizes the model in that paper and generates the estimating equations used in her analysis. We also the U.S. as an empirical context under the same definition of geographic units (MSA) and skill groups (College-Non College), and relying on the same data sources for the variables in our model that are common with her analysis. Therefore, her estimates constitute a natural benchmark for our quantification. In what follows, we discuss these benchmark elasticities and caveats to keep in mind for sensitivity analysis. We also discuss the calibration of parameters not included in Diamond (2016).

Utility and Production Function Parameters $\left\{\alpha_C, \sigma, \rho, b^I_{Y,j}, b^I_{H,j}, d_{H,j}\right\}$ 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 (29)) 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 ($b^I_{Y,j} = 0.468$ for all j in (28)) 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 (29)) is chosen to exactly match the share of workers in the non-traded sector of each MSA, as detailed in Section The housing sector generates profits ; we split these profits between high skill and low skill workers according to the share of total capital income accruing to high skill vs low skill workers in the data. Capital income per capita, by group, is common across all MSAs. Labor income and transfers are taken from the dataset built above. Expenditures by group are the sum of labor income, transfers and capital income. We assume the same elasticity of substitution between traded varieties in (31) of $\sigma = 9$ as Allen and Arkolakis (2014).

[35]Allen and Arkolakis (2014) rely on the aggregate gravity equation from a model that corresponds to our general environment under (i) a single skill group, (ii) no trade imbalances, (iii) no intermediate inputs, and (iv) no non traded good. In our case, the aggregate gravity equation only relies on aggregate expenditures and income regardless of the composition of expenditures across skill groups. In addition, for the current quantification we assume away the non-traded sector intermediate inputs, as in Allen and Arkolakis (2014). Therefore, we only need to adapt their approach to our model to account for trade imbalances.

Productivity Spillovers $\left\{\gamma^P_{\theta',\theta}\right\}$ Diamond (2016) reports estimates of the productivity spillovers within and across groups, denoted in our model $(\gamma^P_{UU}, \gamma^P_{US}, \gamma^P_{SU}, \gamma^P_{SS})$, where we use the index $\theta = S$ for high skill, and $\theta = U$ for low skill. Specifically, she reports estimates for the causal effect of skill group size on wages, estimated using a regression of log wages on log population of each group, instrumented using Bartik shocks. Her estimates can be mapped to our $\gamma^P_{\theta,\theta'}$ parameters by log-linearizing the CES production model, and transforming her estimates into the proper units using the CES parameter estimate ρ reported above. Doing so delivers the following values for the productivity spillovers, given her elasticity estimates: $(\gamma^P_{UU}, \gamma^P_{US}, \gamma^P_{SU}, \gamma^P_{SS}) = (0.7, 0.07, 0.12, 1.47)$.

This approach raises two issues. First, the estimates in Diamond (2016) are noisy; for example it cannot be rejected that the productivity spillovers received by high skilled workers are zero. Second, the point estimates for these elasticities are large compared to classic references estimating city-level agglomeration elasticities such as Ciccone and Hall (1996) or Combes et al. (2008), who find city-level elasticities γ^P in the range of $(0.02, 0.15)$.[36] In the notation of our model, the agglomeration coefficient estimated in that literature corresponds to the outcome of a causal regression of $\overline{w}_j = s^S_j w^S_j + s^U_j w^U_j$, the average wage in a city, on $L^S_j + L^U_j$, the total employment in the city. As a result, in the light of our model that elasticity corresponds to $\gamma^P = \overline{s^{W,S}}(\gamma_{SS} + \gamma_{US}) + \overline{s^{W,U}}(\gamma_{UU} + \gamma_{SU})$, where $\overline{s^{W,\theta}}$ is average share of the wage bill of type-θ workers across cities.

Our approach to tackle these issues is scale down uniformly the value of $(\gamma^P_{UU}, \gamma^P_{US}, \gamma^P_{SU}, \gamma^P_{SS})$ inferred from the estimation of Diamond (2016), so as to match a value of 0.06 from Ciccone and Hall (1996). In that way, we preserve a city-level agglomeration elasticity that is consistent with the standard estimates from the literature, alongside relative elasticities across skill types that is consistent with Diamond (2016). We explore the sensitivity of our results to alternative values of this scaling parameter.

Amenity Spillovers $\left\{\gamma^A_{\theta',\theta}\right\}$ To interpret estimates from Diamond (2016) in the context of our model, we re-write expression (20) with the structure of local amenities as:

$$a^\theta_j = A^\theta_j \left(L^\theta_j\right)^{\beta^{own}_{\theta\theta}} \left[\prod_{\theta'} \left(L^{\theta'}_j\right)^{\xi^A_{\theta'}}\right]^{\beta^a_\theta}. \tag{34}$$

Condition (34) maps to our previous equation (20) assuming $\gamma^A_{\theta\theta} = \beta^{own}_{\theta\theta} + \beta^a_\theta \xi^A_\theta$ and $\gamma^A_{\theta'\theta} = \beta^a_\theta \xi^A_{\theta'}$ for $\theta' \neq \theta$. This equivalent formulation means that local amenities enjoyed by group θ result from from (i) exogenous differences between places A^θ_j; (ii) an amenity component that depends on the size of own group in city j, L^θ_j, mitigated by the elasticity $\beta^{own}_{\theta\theta}$; and (iii) an amenity spillover $a^{D\theta}_j = \prod_{\theta'} \left(L^{\theta'}_j\right)^{\xi^A_{\theta'}}$ generated by both groups with elasticity ξ^A_θ and $\xi^A_{\theta'}$, but differentially enjoyed by group θ and θ' as captured by elasticities β^a_θ and $\beta^a_{\theta'}$. The estimate in Diamond (2016) uses the free mobility condition (9) under the assumption that disposable income equals the wage. In logs,

[36] See Rosenthal and Strange (2004) for a review.

that relationship implies

$$\log L_j^\theta = cst_\theta - \frac{\beta_\theta^a}{\beta_{\theta\theta}^{own}} \log a_j^{D\theta} - \frac{1}{\beta_{\theta\theta}^{own}} \left(\log w_j^\theta - \alpha_C \log R_j - (1-\alpha_C) \log P_j \right) + \epsilon_j^\theta \text{ for all } j \text{ and } \theta = U, S, \tag{35}$$

corresponding to the labor supply equation estimated by

Diamond (2016). The estimates from Column 4 of Table 3 of her paper imply: $(\beta_{UU}^{own}, \beta_{SS}^{own}) = -(\frac{1}{4.026}, \frac{1}{2.116})$ and $\left(\frac{\beta_a^U}{-\beta_{UU}^{own}}, \frac{\beta_a^S}{-\beta_{SS}^{own}}\right) = (0.27, 1.01)$.In addition, Column 3 of Table 5 reports an elasticity of the amenity index to the ratio between high and low skill workers of 2.6, which in our notation implies $a_j^{D\theta} = \left(\frac{L_j^H}{L_j^L}\right)^{2.6}$. Taken together, these estimates deliver: $(\gamma_{LL}^A, \gamma_{LH}^A, \gamma_{HL}^A, \gamma_{HH}^A) = (-0.80, -0.65, 0.33, 0.40)$. According to these estimates, amenity spillovers correspond to a positive agglomeration effect for high-skill workers and to a congestion effect for low skill workers.

There is little further guidance in the literature on estimates for these amenity spillovers. In the absence of alternative estimates, we explore a range of values for these elasticities in our sensitivity analysis.

5.3 Stylized Facts

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

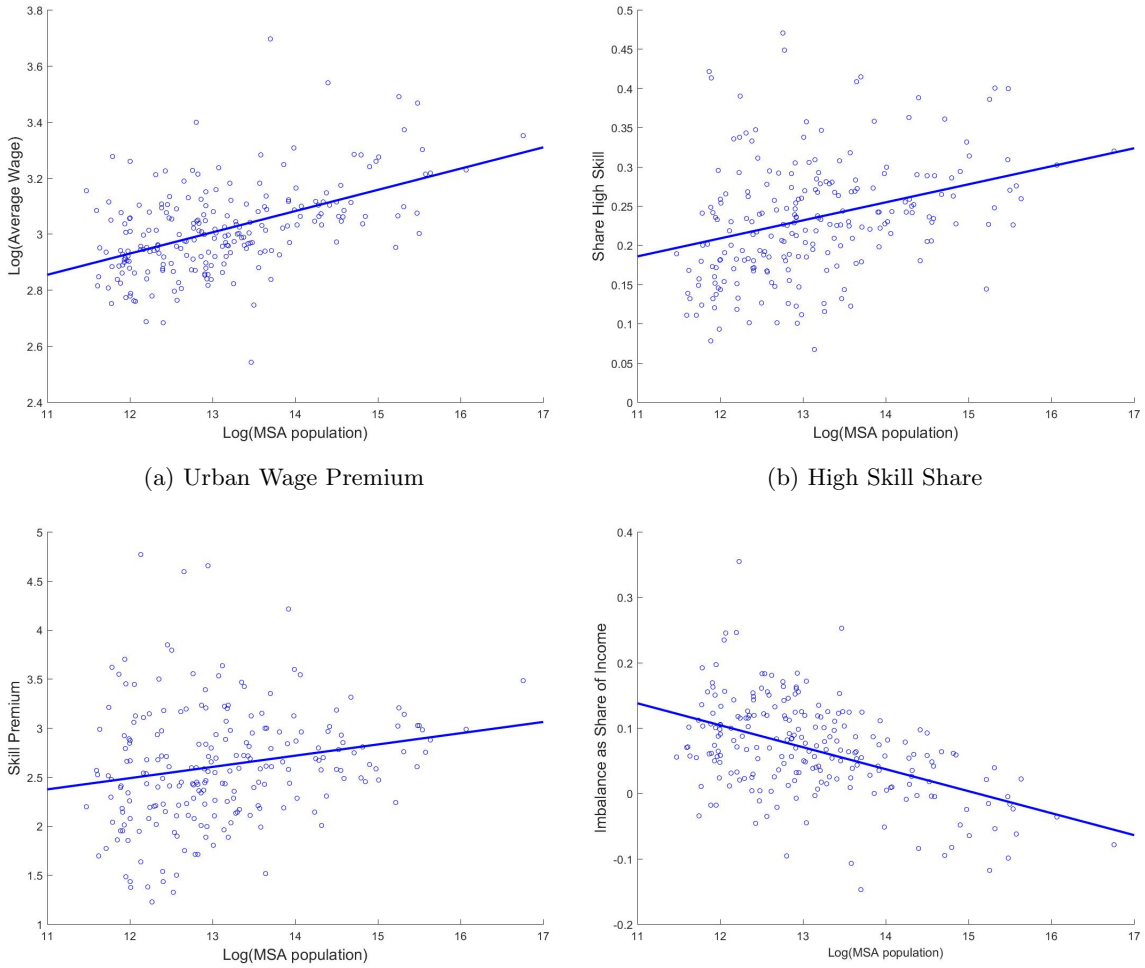
Panels A to C show the standard facts about spatial disparities and sorting. Panel A documents that nominal wages increase with city size. The elasticity of wages to city size is 7.5%. This fact is suggestive of exogenous or endogenous productivity advantages as well as higher prices in larger cities. Panel B shows spatial sorting. 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.[37] Panel C shows that the skill premium, defined as the ratio of high-skill to low-skill wage, increases with city size with a slope of 11.4%. Hence, larger cities feature a more unequal nominal wage distribution. The last two facts suggest complementarities between city size and skill.

Panel D reports the relationship between city size and transfers. The vertical axis shows the net transfer per capita (i.e., governments transfers minus taxes) as a share of income by skill group and city. In the notation of our model, this is $\frac{x_j^\theta - w_j^\theta}{w_j^\theta}$. A higher share low-skill worker's income disposable income is due to redistribution, implying net redistribution to low skill workers. This share decreases with city size with a semi elasticity of 2%, implying net redistribution to low skill workers in smaller cities. On average, net transfers represent about 20% of income in the smaller MSAs. The share of high skill workers' income that is taxed away and redistributed increases with city size, with a semi-elasticity of 2%. This figure illustrates a clear pattern across groups and city

[37] This 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.

sizes: redistribution flows from high skill workers and large cities towards low skill workers and small cities.

Figure 1: Urban Premia



(a) Urban Wage Premium

(b) High Skill Share

(c) Urban Skill Premium

(d) Imbalance as Share of Income

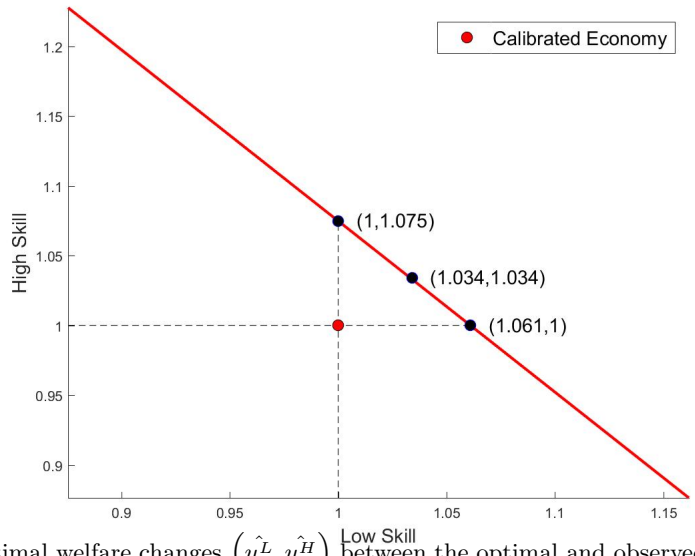
Note: each panel reports outcomes across MSA's. All the city level outcomes reported on the vertical axis are adjusted by socio-demographic characteristics of each city, as detailed in Appendix B.1. The slope (SE) in each panel is: Panel (a): 0.076 (0.009); Panel (b): 0.023 (0.004); Panel (c): 0.115 (0.039); Panel (d): -0.034 (0.004).

6 Optimal Spatial Policies in the U.S. Economy

We now explore quantitatively the predictions of our model and discuss the optimal spatial policies. We solve the planner's problem in relative changes described in Section 4.4. Specifically, given that in our current application there are 2 groups of workers (skilled and unskilled), we solve the problem of maximizing the change in utility of skilled workers, $\hat{u^S}$, subject to a minimum bound for the change in utility of unskilled worker, $\hat{u^U}$.

Utility Frontier Figure 2 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 utility of skilled and unskilled workers is unchanged compared to the calibrated equilibrium. We find sizable welfare gains through the spatial reallocations associated with optimal spatial transfers. 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 3.4% 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 7.5% for high skill workers and of 6.1% for low skill workers.

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



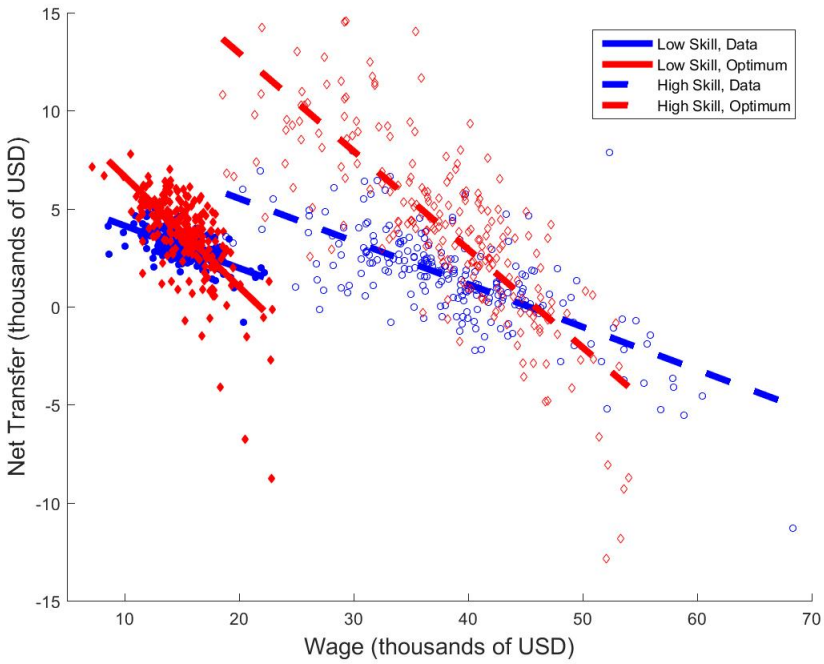
The figure shows the optimal welfare changes $\left(\hat{u^L}, \hat{u^H}\right)$ between the optimal and observed allocation, corresponding to the solution of the planner's problem in relative changes described in Appendix A.2. Each point corresponds to a maximization of $\hat{u^H}$ subject to a different lower bound on $\hat{u^L}$.

Actual versus Optimal Transfers The optimal allocation is implemented through a particular set of 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 note that condition (18) defines the difference between expenditure and wages in an optimal allocation. These transfers are given by:

$$x_j^\theta - w_j^\theta = \left[\frac{\gamma_{\theta,\theta}^P + \gamma_{\theta,\theta}^A}{1-\gamma_{\theta,\theta}^A} + \frac{1}{1-\gamma_{\theta,\theta}^A}\left(\gamma_{\theta,\theta'}^P \frac{w_j^{\theta'} L_j^{\theta'}}{w_j^\theta L_j^\theta} + \gamma_{\theta,\theta'}^A \frac{x_j^{\theta'} L_j^{\theta'}}{w_j^\theta L_j^\theta}\right)\right] w_j^\theta - E^\theta. \tag{36}$$

Figure 3 shows a scatter plot of the net transfers $x_j^\theta - w_j^\theta$ by city and worker type on the vertical axis against the wage w_j^θ of each city and worker type on the horizontal axis, in both the data (blue circles) and the optimal allocation (red circles). The optimal allocation represented in the figure

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



Note: each point in the figure corresponds to an MSA-skill group combination. The vertical axis 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: -0.22 (0.02); Low Skill, Optimum: -0.56 (0.04); High Skill, Data: -0.22 (0.01); High Skill, Optimum: -0.50 (0.03). The figure corresponds to planner's weights such that both types of workers experience the same welfare gain (i.e., point (1.034,1.034) in Figure 2).

corresponds to the point $(1.034, 1.034)$ in the utility frontier, where the welfare of both worker types increases in the same proportion (below we discuss the implications of choosing alternative points on the frontier).The best linear fit is shown in solid lines for low skill workers and in dashed lines for high skill workers.

Consider first the transfers in the data. By construction, these transfers are matched by the model in the initial equilibrium. Average net transfers for low skill workers are positive in almost every city. Taking a simple average across cities, the disposable income of low skill workers is 3 thousand dollars above their average wage and represent 20% of the wage. For high skill workers the corresponding numbers are one thousand dollars and 5% of the wage. On average, the net transfers are negative for high skill workers in larger cities, implying net redistribution from high-skill workers in larger cities to the rest of the economy. The slope of net transfers with respect to wage income across cities is -0.22 for both skilled and unskilled workers, which is consistent with an implicit labor income tax of 22%. We note that, given the steps involved in constructing these data, the intercept of each group depends on determinants of expenditures that independent from the location of an individual worker, including redistribution between groups and capital ownership.

The red circles show the efficient allocation. By construction, this relationship satisfies the

optimality condition 36.Across cities, the optimal transfers 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. Net transfers decrease with wages, with a slope of 0.56 for low skill workers and 0.50 for high skill workers.

To understand the forces behind these patterns, we note that the slope is partly driven by the first term within the square brackets in the right hand side of 36, $\frac{\gamma^P_{\theta,\theta}+\gamma^A_{\theta,\theta}}{1-\gamma^A_{\theta,\theta}}$. This term would be the optimal tax shown in (23) if there were no spillovers across types. The dispersion in optimal transfers around a linear fit is exclusively due to the spillover across types, i.e. the second term within the square brackets in 36. This term also drives part of the slope, depending on the sign of the cross-spillovers.

We can now inspect the role of the different spillovers in driving the optimal spatial transfers. For low-skill workers, the negative own spillovers alone would imply $\frac{\gamma^P_{U,U}+\gamma^A_{U,U}}{1-\gamma^A_{U,U}} = -0.43$, i.e., a labor tax of 43%. The larger implicit optimal tax is driven by the negative cross spillovers through amenities ($\gamma^A_{U,S} < 0$). 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. High skill workers generate on net positive own spillovers of $\frac{\gamma^P_{S,S}+\gamma^A_{S,S}}{1-\gamma^A_{S,S}} = 0.8$. Absent spillovers on low-skill workers across workers, this force would imply an 80% labor *subsidy*. However, this force is more than offset by positive cross spillovers. Because a larger concentration of group S generates positive spillovers on group U, this force leads to a larger implicit subsidy to high skill workers in places with a high share of low skill workers, corresponding. These places are the small, low-wage cities in both the data and the optimal allocation.[38] To highlight this key role of spillovers from high to low skill workers, Figure A.1 in Appendix C replicates the linear relationships of Figure 3 alongside the best linear fit when these spillovers are assumed away ($\gamma^A_{S,U} = \gamma^P_{S,U} = 0$). In that case, the optimal net transfers for high skill workers imply a wage subsidy and look radically different from the data.

Throughout this discussion we compared the data to an optimal allocation corresponding to the same welfare gains to all workers (i.e., the welfare gains $\left(\hat{u}^U, \hat{u}^S\right) = (1.034, 1.034)$ in Figure 2). The patterns of optimal redistribution discussed so far are robust to choosing allocations corresponding to other points on the frontier. Different planner's weights are largely 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.2 in Appendix C shows the same outcomes as in Figure 3 assuming planner weights that are either 10 times larger for high skill workers (in the left panel) or 10 times larger for low skill workers (right panel). The optimal welfare changes relative to the observed allocation associated with those weights are considerably different from our benchmark case: $\left(\hat{u}^U, \hat{u}^S\right) = (-13\%, 23\%)$ under higher weight for high skill workers, and $\left(\hat{u}^U, \hat{u}^S\right) = (16\%, -13\%)$

[38] 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, $\frac{x^U_j L^U_j}{w^S_j L^S_j}$, is larger.

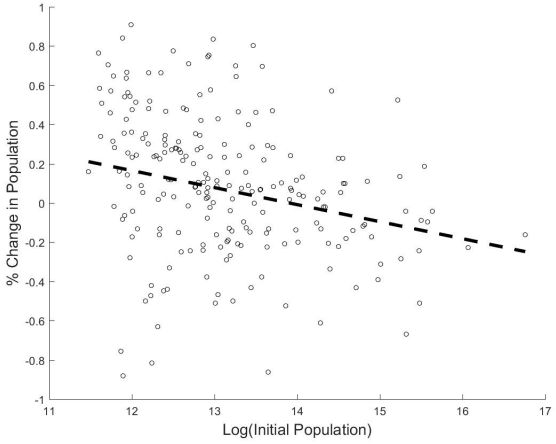
under higher weight for low skill. However, we observe a very similar spatial pattern of net transfers as in our previous analysis, suggesting that the stronger redistribution to smaller and low-wage cities in the optimal allocation is independent from the point on the utility frontier that we target.

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

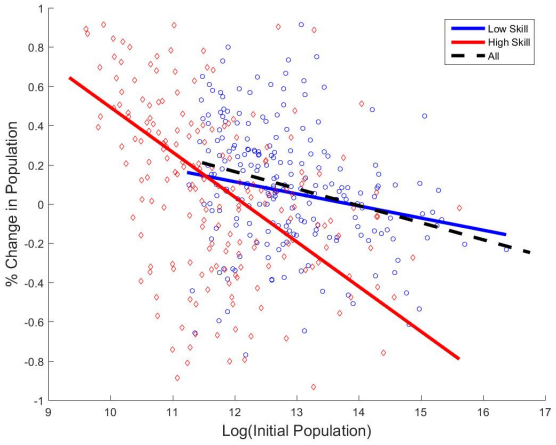
Figure 4 shows the pattern of reallocation. Panel (a) shows the initial total population of each MSA on the horizontal axis and the percentage change in population implied by the optimal allocatino on the vertical axis. The stronger redistirbution to low-wage locations discussed in the previous section implies that, on average, there is reallocation from large to small cities. However, there is 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. 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 (slope of -0.23) than for low skill workers (-0.06). This is consistent with the initially observed allocation featuring too little mixing of skills due to the spillovers across types, as discussed in the previous section.

These differential patterns of reallocation by skill lead to less seggregation 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. A second implication of these reallocations is that, on average, within-city wage inequality, in terms of the skill premium, tends to decrease in cities that initially more unequal.

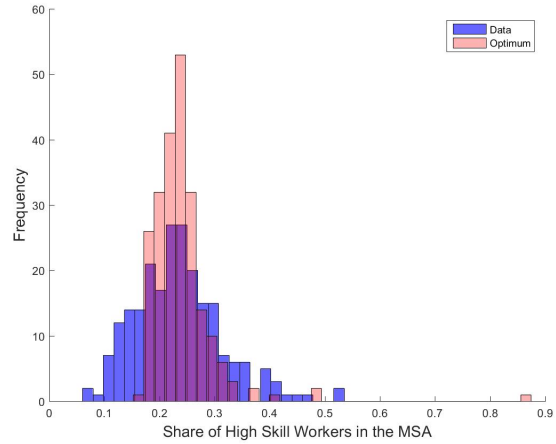
Figure 4: Changes in Population, Skill Shares, and Skill Premium across MSA's



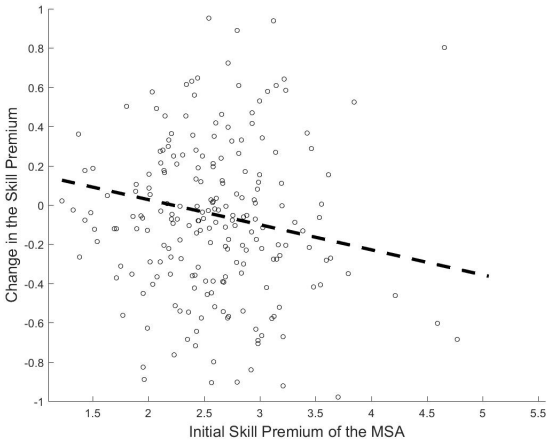
(a) Change in Total Population



(b) Change in Population by Skill Group



(c) Histogram of High-Skill Shares across MSA's



(d) Change in Skill Premium

Note: Panel (a) shows the change in population between the optimal allocation and the initially observed equilibrium and the linear fit. Slope (SE): -0.09 (0.03); R^2=0.05. Panel (b) displays the same outcomes for high and low skill workers. Slopes (with SE): High Skill: -0.23(0.03); Low Skill: -0.06(0.02). Panel (d) displays in the verical axis the difference in the skill premium between the optimal and initial allocation. Slope (SE): -0.12 (0.06). The figures correspond to planner's weights such that both types of workers experience the same welfare gain (i.e., point (1.034,1.034) in Figure 2).

The Urban Premia in the Optimal Allocation We now revisit the stylized facts about urban premia mentioned in Section 5. Figure 5 reproduces these facts in both the data and the optimal allocation. We start with Panel D of Figure 5, which summarizes the relationship between city size and imbalances (as a share of income). As expected given the previous discussion, there are more transfers from large to small cities in the optimal allocation than in the data.

We then turn to differences between high skill and low skill workers. Panel B of 5 shows that spatial sorting is slightly less pronounced, on average, in the optimal equilibrium compared to the data. The slope of the best linear fit of the high skill share with respect to city size decreases from 2.3% to 1.6%. That is, the optimal allocation results in more mixing of high and low skills

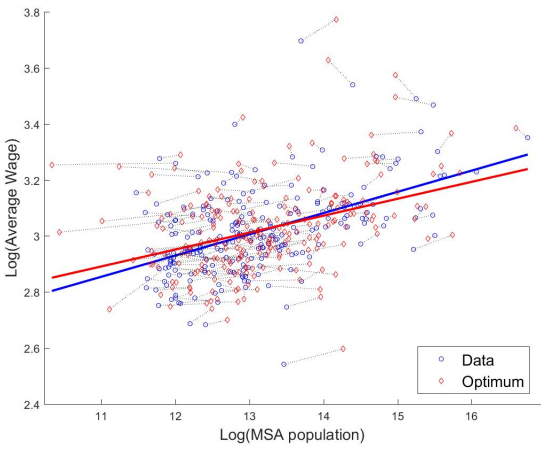
in large cities. This dampened sorting is in part driven by a lower urban skill premium in the optimal allocation compared to the data. The relative wage benefit of locating in a large city is not as large, for high skill workers, in the optimal allocation as it is in the data, so they sort less into larger cities. As Panel C illustrates, the skill premium increases with city size with a slope (semi-elasticity) of 11.4% in the data whereas, in the optimal allocation, this urban skill premium vanishes. In particular, the slope becomes negative, as the skill premium decreases with city size with a semi-elasticity of -7.4%. This reversal of the urban skill premium comes mainly from the change in city size distribution described above. Large cities shrink and small cities grow as a result of the implementation of the optimal transfers.

At the same time, larger cities in the data are revealed to be relatively more productive for high skill workers than for low skill workers, above and beyond endogenous forces. As a consequence, as relative city sizes change, the city-size skill premium shrinks. That is, cities in which high skill workers have a comparative advantage become smaller, and cities in which low skill have a comparative advantage becomes larger. The skill premium is shaped by these difference, so that in the optimal allocation, it tends to decrease, on average, with city size. Finally, as Panel A of Figure 5 indicates, there is still a wage premium in the large cities, but given the lack of an urban skill premium it is driven by an average productivity advantage (i.e., across both skill groups) of larger cities, instead of a comparative advantage in high-skill tasks.

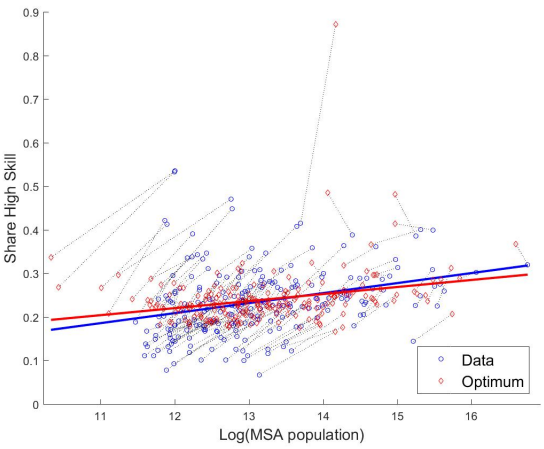
An interesting aspect of the analysis highlighted by these results is that, in the optimal allocation, the sorting patterns end up being detached from the urban skill premium. Instead, sorting is driven by differential preference for urban amenities (including the ones created by spillovers) and spatial transfers.

Figure 5: Urban Premia, Data and Optimal Allocation

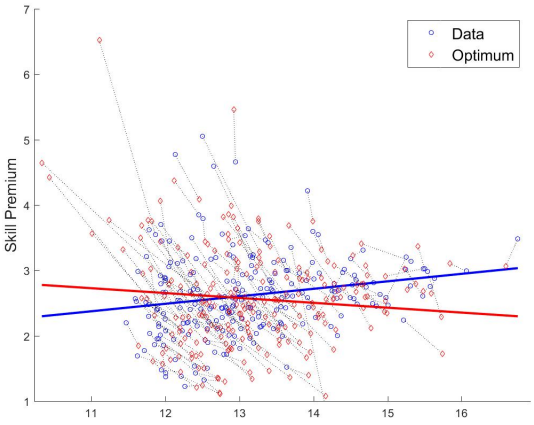
(a) Urban Wage Premium



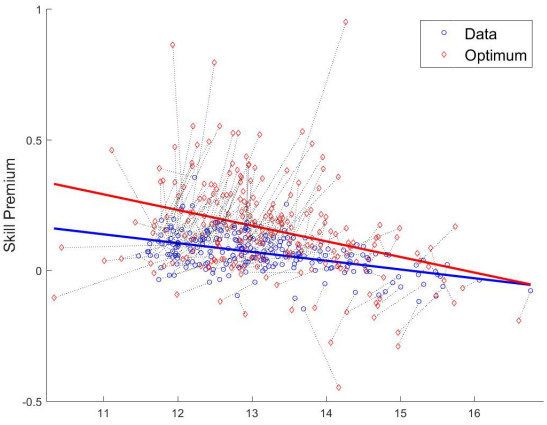
(b) High Skill Share



(c) Urban Skill Premium



(d) Imbalance as Share of Income



Note: each panel reports outcomes across MSA's in the data and in the optimal allocation. Each pair of linked circles is an MSA. The slope (SE) in each panel is: Panel (a): Data 0.076 (0.009), Opt 0.060 (0.009); Panel (b): Data 0.023 (0.004), Opt 0.016 (0.004) ; Panel (c): Data 0.115 (0.039), Opt -0.075 (0.050); Panel (d): Data -0.034 (0.004), Opt -0.059 (0.011).

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. Crucially, 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 elasticities to assess the efficiency of the observed spatial allocation. We show 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 with lump sun transfers are distortive and sufficient to implement the efficient allocation. When spillovers across workers are present, spatial efficiency may require more mixing of heterogeneous workers across cities than in the competitive allocation, and the decentralized pattern of sorting by skill can be too strong.

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 estimates of the spillover elasticities from the literature, our results suggest too little redistribution to low-wage cities. The net transfer, defined as the difference between disposable income and wage, 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 current allocation. However, there is also considerable dispersion in the distribution of growth rates for any given initial size. The results also suggest that, due to positive spillovers from high to low skill workers, optimality requires a greater degree of mixing of different skills. The preliminary estimates also suggest that an optimal reallocation of economic activity increases the real consumption of both low and high skill workers by 3.4%.

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*.

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.

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 (2014). 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 and 2

The planning problem described as follows.

Definition 2. *The planning problem is*

$$\max L^{\theta} u^{\theta}$$

subject to (i) the spatial mobility constraints

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

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

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

(iii) local and notional labor-market clearing,

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

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

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

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

for all j and θ and some constants $\left\{E^{\theta}\right\}$. Furthermore, if the planner's problem is globally concave and (A.1) *holds for some specific $\left\{E^{\theta}\right\}$, then the competitive equilibrium is efficient.*

Proof. First we present the system of necessary first order conditions in the planner's problem. Then we contrast

it with the market allocation. The Lagrangian of the planning problem is:

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

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

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

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

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

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:

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

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

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

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

where

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

Therefore, we can write

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

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

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:

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

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

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

An optimal allocation is given by quantities $\left\{Q_{ji}, N_j^Y, N_j^H, I_j^Y, I_j^H, c_j^\theta, h_j^\theta, L_j^\theta, u^\theta\right\}$ and multipliers $\left\{P_j^*, p_j^*, R_j^*, W_j^*, \omega_j^\theta\right\}$ such that the first-order conditions (A.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 $\left\{P_j^*, p_j^*, R_j^*, W_j^*\right\}$ and decentralized expenditure x_j^θ equal to $x_j^{\theta *}$. In addition, the restrictions (i) to (iii) from definition 2 of the planning problem are the same as restriction (iii) from the competitive equilibrium. Therefore, the system characterizing the competitive solution for $\left\{Q_{ji}, N_j^Y, N_j^H, I_j^Y, I_j^H, c_j^\theta, h_j^\theta, L_j^\theta\right\}$ given the prices $\{P_j, p_j, R_j, W_j\}$ and the expenditure x_j^θ is the same as the system characterizing the planner allocation for those same quantities given the multipliers $\left\{P_j^*, p_j^*, R_j^*, W_j^*\right\}$ and $x_j^{\theta *}$. As a result, if the competitive allocation is efficient, then $x_j^\theta = x_j^{\theta *}$ where $x_j^{\theta *}$ is given by (A.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 $\{P_j, p_j, R_j, W_j\} = \left\{P_j^*, p_j^*, R_j^*, W_j^*\right\}$. If the planning problem is convex then there is a unique solution to the system characterizing the planner's allocation, in which case $\{P_j, p_j, R_j, W_j\} = \left\{P_j^*, p_j^*, R_j^*, W_j^*\right\}$ is the only competitive equilibrium.

□

Proposition 2. *The planning problem is concave if* $1+\Gamma^A > \frac{1+\Gamma^P}{1+D}$, $\Gamma^A \geq 0$ *and* $\gamma^A_{\theta,\theta'} > 0$ *for* $\theta \neq \theta'$. *Under a single worker type (*$\Theta = 1$*), the planning problem is quasi-concave if* $1 + \gamma^A > \left(1+\gamma^P\right)\left[(1-\alpha_C)\frac{1}{1+d_{H,j}} + \alpha_C \frac{1}{1+d_Y}\right]$ *for all* j.

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

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

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

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

$$\max_{\left\{v^\theta, U_j^\theta, C_j^\theta, H_j^\theta, \tilde{L}_j^\theta, \tilde{N}_j^k, I_j^k, Q_{ij}, M_j, S_j\right\}} v^\theta \tag{A.12}$$

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

$$v^{\theta}-\mathcal{F}\left(\frac{U_{j}^{\theta}\prod_{\theta'\neq\theta}\left(\widetilde{L}_{j}^{\theta'}\right)^{\frac{\gamma_{\theta',\theta}^{A}}{1+\alpha}}}{\left(\widetilde{L}_{j}^{\theta}\right)^{\frac{1-\gamma_{\theta,\theta}^{A}}{1+\alpha}}}\right)\leq 0 \text{ for all } j; \tag{A.13}$$

$$\underline{v}^{\theta'}-\mathcal{F}\left(\frac{U_{j}^{\theta'}\prod_{\theta''\neq\theta'}\left(\widetilde{L}_{j}^{\theta''}\right)^{\frac{\gamma_{\theta'',\theta'}^{A}}{1+\alpha}}}{\left(\widetilde{L}_{j}^{\theta'}\right)^{\frac{1-\gamma_{\theta',\theta'}^{A}}{1+\alpha}}}\right)\leq 0 \text{ for all } j \text{ and } \theta'\neq\theta; \tag{A.14}$$

$$U_{j}^{\theta}-U\left(C_{j}^{\theta},H_{j}^{\theta}\right)\leq 0 \tag{A.15}$$

$$\sum_{i}\tau_{ji}Q_{ji}-\left(b_{Y}^{N}\left(\widetilde{N}_{j}^{Y}\right)^{\beta_{Y}(1+D)}+b_{Y}^{I}\left(\widetilde{I_{j}^{Y}}\right)^{(1+D)\beta_{Y}}\right)^{\frac{1}{\beta_{Y}}\frac{1}{1+D}}\leq 0 \text{ for all } j,i; \tag{A.16}$$

$$\sum_{\theta}C_{j}^{\theta}+\left(\widetilde{I_{j}^{Y}}\right)^{1+D}+\left(\widetilde{I_{j}^{H}}\right)^{1+D}-Q\left(Q_{1j},..,Q_{Jj}\right)\leq 0 \text{ for all } j; \tag{A.17}$$

$$\sum_{\theta}H_{j}^{\theta}-\left(b_{H}^{N}\left(\widetilde{N_{j}^{H}}\right)^{\beta_{H}(1+D)}+b_{H}^{I}\left(\widetilde{I_{j}^{H}}\right)^{(1+D)\beta_{H}}\right)^{\frac{1}{\beta_{H}}\frac{1}{1+D}}\leq 0 \tag{A.18}$$

$$M_{j}^{1+D}-\left[\sum_{\theta}\left(Z_{j}^{\theta}\prod_{\theta'}\left(\widetilde{L}_{j}^{\theta'}\right)^{\frac{\gamma_{\theta',\theta}^{P}}{(1+D)(1+\alpha)}}\left(\widetilde{L}_{j}^{\theta}\right)^{\frac{1}{(1+D)(1+\alpha)}}.\right)^{\rho}\right]^{\frac{1}{\rho}}\leq 0 \text{ for all } j; \tag{A.19}$$

$$\left(\widetilde{N_{j}^{Y}}\right)^{1+D}M_{j}^{-D}+\left(\widetilde{N_{j}^{H}}\right)^{1+D}M_{j}^{-D}-M_{j}\leq 0 \tag{A.20}$$

$$\sum_{j}\left(\widetilde{L_{j}^{|\theta}}\right)^{\frac{1}{1+\alpha}}-L^{\theta}=0 \text{ for all } \theta \tag{A.21}$$

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

$$\begin{aligned}
v^{\theta} &= \mathcal{F}\left(u^{\theta}\right)\\
H_{j}^{\theta} &= L_{j}^{\theta}h_{j}^{\theta}\\
C_{j}^{\theta} &= L_{j}^{\theta}c_{j}^{\theta}\\
\widetilde{N_{j}^{k}} &= \left(N_{j}^{k}\right)^{\frac{1}{1+D}} \text{ for all } j\in 1,...,N \text{ and for } k\in\{Y,H\};\\
\widetilde{I_{j}^{k}} &= \left(I_{j}^{k}\right)^{\frac{1}{1+D}} \text{ for all } j\in 1,...,N \text{ and for } k\in\{Y,H\};\\
\widetilde{L}_{j}^{\theta} &= \left(L_{j}^{\theta}\right)^{1+\alpha} \text{ for all } j\in 1,...,N \text{ and } \theta,
\end{aligned}$$

where $\mathcal{F}(.)$ is a bijection that we define below and where we have defined

$$\alpha=\frac{1+\Gamma^{P}}{1+D}-1. \tag{A.22}$$

Problems $\mathcal{P}$ and $\mathcal{P}'$ are equivalent: any solution to $\mathcal{P}'$ is a solution to $\mathcal{P}$ and vice-versa. We then consider the relaxed problem $\mathcal{P}''$ that is identical to $\mathcal{P}'$ except that the last constraint of $\mathcal{P}'$ is relaxed into an inequality constraint:

$$L^{\theta}-\sum_{j}\left(\widetilde{L_{j}^{|\theta}}\right)^{\frac{1}{1+\alpha}}\leq 0 \text{ for all } \theta. \tag{A.23}$$

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.13) and (A.14), and specifically the term $\frac{U_j^\theta \prod_{\theta' \neq \theta} \left(\widetilde{L}_j^{\theta'}\right)^{\frac{\gamma^A_{\theta',\theta}}{1+\alpha}}}{\left(\widetilde{L}_j^\theta\right)^{\frac{1+\sigma_\theta-\gamma^A_{\theta,\theta}}{1+\alpha}}}$. 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.13) and (A.14) are convex for all θ whenever $\gamma^A_{\theta',\theta} \geq 0$ for all $\theta' \neq \theta$ and $\frac{1+\sigma_\theta-\gamma^A_{\theta,\theta}}{1+\alpha} > 1 + \frac{\gamma^A_{\theta',\theta}}{1+\alpha}$. This parameter restriction corresponds to assumption (i): $1 + \Gamma^A > \frac{1+\Gamma^P}{1+D}$. We use here the common function $\mathcal{F}(x) = -x^b$ for $b = \frac{1+\Gamma^P}{1+\Gamma^P-\left(1+\Gamma^A\right)(1+D)}$ in constraint (A.13) for any j and any θ. Given the definition of Γ^A, $\mathcal{F}(.)$ is a concave transform of $G_\theta(x) = -x^{\frac{1+\Gamma^P}{\left(1+\Gamma^P\right)-\left(1+\sigma_\theta+\sum_{\theta'}\gamma^A_{\theta',\theta}\right)(1+D)}}$ 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 $f(x_1,...,x_n) = \left[\sum a_i x_i^\beta\right]^\rho$ are concave whenever $\beta \in (0,1)$ and $\rho\beta \leq 1$. Therefore, constraints (A.16), (A.18) and (A.23) are convex. Constraint (A.15) is convex because $U(.)$ is concave. Constraint (A.17) is convex because $D \geq 0$ so that $x \to x^{1+D}$ is convex, and the aggregator $Q(.)$ is concave. Next, consider the constraint (A.19). The second term is the negative of a composition of an increasing CES function with exponent $\rho \leq 1$, which is concave, and a series of functions of the form

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

As concave transforms of a geometric mean, these functions are concave, a whenever $\frac{1+\sum_{\theta'}\gamma^P_{\theta',\theta}}{(1+D)(1+\alpha)} \in (0,1)$. This restriction holds by definition of α in (A.22). 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.19) is convex. Turning to constraint (A.20), note that $x \to x^{1+D}$ is convex whenever $D \geq 0$. Therefore, $f(x,t) = t\left(\frac{x}{t}\right)^{1+D} = x^{1+D}t^{-D}$ is also convex, as it is the perspective of the first function. In turn, constraint (A.20) is 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.23) binds, so that (A.21) 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 $\mathcal{P}$.[39]

We now specialize to the case of a single type of workers ($\Theta = 1$). For simplicity, we make the assumption that decreasing returns to scale are stronger in the non traded sector and we allows for city-level decreasing returns to scale in the non traded sector. We define:

$$1 + d'_{H,j} = \frac{1 + d_{H,j}}{1 + d_Y} \geq 1$$

The proof below is easily transposed to the opposite case using symmetric arguments. The relaxed planner's problem $\mathcal{P}''$ can be further simplified in this case to:

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

subject to the constraints (A.16), (A.17), (A.19), (A.20) and (A.23), which are unchanged except that they now hold

[39] We have not proven that (A.23) 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.

for only one group. The modified constraint for housing production is:

$$\widetilde{H}_j^\theta - \left(b_H^N \left(\widetilde{N_j^H} \right)^{\beta_H (1+D)} + b_H^I \left(\widetilde{I_j^H} \right)^{(1+D)\beta_H} \right)^{\frac{1}{\beta_H} \frac{1}{1+D}} \leq 0 \tag{A.24}$$

where we have used the following change of variable $\widetilde{H}_j^\theta = \left(H_j^\theta \right)^{1+d'_{H,j}}$. The modified housing market constraint (A.24) 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 $(1-\alpha_C) \frac{1}{1+d_{H,j}} + \alpha_C \leq \frac{1+\sigma_\theta - \gamma^A_{\theta,\theta}}{1+\alpha}$ 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.

□

A.2 Planning Problem in Relative Changes

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 4.4. From the optimization problems (11) and (12) and market clearing in the housing market we obtain the following conditions:

$$W_i N_i^Y = \left(1 - b_{Y,i}^I \right) p_i Y_i. \tag{A.25}$$

$$W_i N_i^H = \frac{1 - b_{H,i}^I}{1 + d_{H,i}} \left(1 - \alpha_C \right) X_i. \tag{A.26}$$

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

$$\frac{N_i^H}{N_i} = \frac{\frac{1-b_{H,i}^I}{1+d_{H,i}} \frac{1-\alpha_C}{1-b_{Y,i}^I} \left(\frac{X_i}{p_i Y_i} \right)}{\frac{1-b_H^I}{1+d_{H,i}} \frac{1-\alpha_C}{1-b_{Y,i}^I} \left(\frac{X_i}{p_i Y_i} \right) + 1.} \tag{A.27}$$

Using (A.25) and (A.26) 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:

$$\frac{X_i}{p_i Y_i} = \frac{1 - b_{Y,i}^I}{\frac{W_i N_i}{X_i} - \frac{1-b_{H,i}^I}{1+d_{H,i}} \left(1 - \alpha_C \right)}. \tag{A.28}$$

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 $E_j^Y = P_j Q_j = P_j Q \left(Q_{1j}, .., Q_{Jj} \right)$ 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

$$E_j^Y = \left(\alpha_C + \left(1 - \alpha_C \right) \frac{b_{H,j}^I}{d_{H,j} + 1} \right) X_j + b_Y^I \left(p_j Y_j \right). \tag{A.29}$$

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

$$\sum_i s_{ji}^X = 1. \tag{A.30}$$

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'sshare of i's purchases of tradeable goods (i.e., the import share of region j in i). Finally, aggregating the budget constraints of individual consumers gives

$$\sum_j s_{ji}^M \equiv 1. \tag{A.31}$$

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 expenditure function $\chi_j^\theta(\cdot)$ defined in (15). We now show that the outcomes in the new equilibrium relative to the initial equilibrium are given by a set of changes in prices $\left\{\hat{P}_i, \hat{p}_i, \hat{R}_i\right\}$, wages $\left\{\hat{W}_i\right\}$, employment by group $\left\{\hat{L_i^\theta}\right\}$, supply of efficiency $\left\{\hat{N}_i\right\}$, production of tradeable goods $\left\{\hat{Y}_i\right\}$, and utility levels $\left\{\hat{u^\theta}\right\}$ that satisfy a set of conditions given the equilibrium change in expenditure per capita by group and location $\left\{\hat{x_i^\theta}\right\}$. The planner problem in relative changes will then choose the optimal $\left\{\hat{x_i^\theta}\right\}$.

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

$$\sum_j s_{ij}^X \left(\frac{\hat{p_i}}{\hat{P_j}}\right)^{1-\sigma} \hat{E_j^Y} = \hat{p_i}\hat{Y_i} \text{ for all } i, \tag{A.32}$$

$$\sum_j s_{ji}^M \left(\frac{\hat{p_j}}{\hat{P_i}}\right)^{1-\sigma} = 1 \text{ for all } i, \tag{A.33}$$

$$\left(1 - \frac{N_i^H}{N_i}\right) \hat{p_{ii}}\hat{Y_i} + \frac{N_i^H}{N_i}\hat{X_i} = \hat{W_i}\hat{N_i} \text{ for all } i, \tag{A.34}$$

$$\hat{W_i}^{1-b_{Y,i}^I} \hat{P_i}^{b_{Y,i}^I} = \hat{p_i} \text{ for all } i \tag{A.35}$$

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.32) and (A.33) follow from expressing (A.30) and (A.31) in relative changes and using the CES functional for $Q(\cdot)$ form (31). In condition (A.32), using (A.29) implies that the change in expenditures in tradeable commodities is:

$$\hat{E_j^Y} = \hat{X_j} + \frac{b_Y^I}{\left(\alpha_C + (1-\alpha_C)\frac{b_H^I}{d_{H,j}+1}\right)\left(\frac{X_j}{p_j Y_j}\right) + b_Y^I} \hat{p_j}\hat{Y_j}. \tag{A.36}$$

Condition (A.34) follows from expressing labor-market clearing (7) in relative changes together with (A.25) and (A.26), where the non-traded labor share $\frac{N_i^H}{N_i}$ is defined in (A.27). Condition (A.35) follows from optimization of producers of tradeable commodities.

The system (A.32) to (A.35) defines a solution for $\left\{\hat{P_j}, \hat{p_j}, \hat{Y_j}, \hat{W_j}\right\}$ given the change in the number of efficiency units $\hat{N_i}$. This system holds given an arbitrary change in total expenditures in each region $\hat{X_i}$ and independently from heterogeneity across groups or spillovers. Heterogeneous groups and spillovers enter through $\hat{N_i}$. Expressing the CES functional form for the aggregation of labor types in (30) in relative changes and relying on the labor demand expressions in the market allocation (13) to back out the efficiency of each group[40] we obtain:

$$\hat{N_i} = \left(\sum_\theta s_i^{W,\theta} \left(\hat{L_i^\theta}\right)^\rho \prod_{\theta'} \left(\hat{L_i^{\theta'}}\right)^{\gamma_{\theta',\theta}^P}\right)^{\frac{1}{\rho}} \tag{A.37}$$

[40]This step gives productivity by type and city equal to $z_i^\theta = \frac{w_i^\theta}{W_i}\left(\frac{L_i^\theta}{N_i}\right)^{1-\rho}$. This expression is then introduced in $\hat{N_i}$ to reach (A.37).

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. The change in the number of workers of each type in every location that is initially populated must be consistent with the spatial mobility constraint (10),

$$\hat{u^\theta} = \prod_{\theta'} \left(\hat{L_i^{\theta'}} \right)^{\gamma^A_{\theta',\theta}} \frac{\hat{x_i^\theta}}{\hat{P_i}^{\alpha_C} \hat{R_i}^{1-\alpha_C}} \tag{A.38}$$

where, after solving for the equilibrium price of non-traded goods and expressing it relative changes, the change in the cost of housing is

$$\hat{R_i} = \left(\hat{p_{ii}}^{\frac{1-b^I_{H,i}}{1-b^I_{Y,i}}} \hat{P_i}^{b^I_{H,i} - b^I_{Y,i} \frac{1-b^I_{H,i}}{1-b^I_{Y,i}}} \hat{X_i}^{d_{H,i}} \right)^{\frac{1}{1+d_{H,i}}}. \tag{A.39}$$

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

$$\sum_j s_j^{L,\theta} \hat{L_j^\theta} = 1 \text{ for all } \theta. \tag{A.40}$$

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

In sum, an equilibrium in changes given a change in expenditure per capita $\left\{ \hat{x_j^\theta} \right\}$ consists of $\left\{ \hat{P_i}, \hat{p_i}, \hat{Y_i}, \hat{W_i}, \hat{N_j}, \hat{L_j^\theta}, \hat{R_i}, \hat{u^\theta} \right\}$ such that equations (A.32) to (A.40) 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.

Given the elasticities $\left\{ \alpha_C, \rho, b^I_{Y,j}, b^I_{H,j}, d_{H,j} \right\}$, and as long as $b^I_Y > 0$, computing the change in tradeable expenditures requires information about gross expenditures over tradeable income, $\frac{X_j}{p_j Y_j}$. This information is also needed to compute the non-traded labor share $\frac{N_i^H}{N_i}$ in (A.34). As shown in (A.27) and (A.28), $\frac{X_j}{p_j Y_j}$ can be constructed using the elasticities $\left\{ \alpha_C, b^I_Y, b^I_H, d_{H,j} \right\}$ and the share of wages in gross expenditures, $\frac{W_i N_i}{X_i}$.

Planner's Problem in Relative Changes In the implementation, we solve an optimization over $\left\{ \hat{x_j^\theta} \right\}$ subject to $\left\{ \hat{P_i}, \hat{p_i}, \hat{Y_i}, \hat{W_i}, \hat{N_j}, \hat{L_j^\theta}, \hat{R_i}, \hat{u^\theta} \right\}$ consistent with (A.32) to (A.40) in order to maximize the utility of a given group θ, $\hat{u^\theta}$, subject to a lower bound for the change in the level of utility of the other groups ($\hat{u^{\theta'}} \geq \underline{\hat{u^{\theta'}}}$ for $\theta' \neq \theta$). This problem is equivalent to the problem in levels in Definition 2 and therefore concave under the conditions of Proposition 2. The steps to show that the two problems have the same solutions are summarized here: (1) The planner's allocation problem over quantities from 2 under the conditions of Proposition 2. We denote this problem by $\mathcal{P}_2$. (2) Any solution to this problem also maximizes utility subject to feasibility constraints, good, and labor market clearing conditions, expressed using prices and expenditures rather than quantities. We denote this problem $\mathcal{P}_2$'. This is because the constraints that prices must satisfy in $\mathcal{P}_2$' correspond to a subset of the first order conditions of problem $\mathcal{P}_2$ (namely, the ones that define the Lagrange multipliers corresponding to good and factor prices). Therefore, a vector satisfies the first order conditions for $\mathcal{P}_2$ if and only if it satisfies the first order conditions for $\mathcal{P}_2$' ; (3) The problem in changes stated here is simply problem $\mathcal{P}_2'$ modified with the changes of variable $x \to x_o \widehat{x}$ for all variables, where x_o is a constant and $\widehat{x}$ a variable considered in $\mathcal{P}_2'$. The two problems have therefore the same solutions, subject to the appropriate change of variables. In particular a point that satisfies the first order conditions is the (unique) global maximizer for the problem.

A.3 Equivalence with Monopolistic Competition

Consider the economic geography environment from Section 4.3. As a reminder, that environment starts from general model from Section 3 and imposes only one labor type, inelastic housing supply ($H_j \left(N_j^H, I_j^H \right) = H_j$ is a constant), and only labor used in production of traded goods ($Y_j \left(N_j^Y, I_j^Y \right) = N_j^Y = N_j = z_j \left(L_j \right) L_j$). Now suppose that, in addition, the production structure in the traded sector is the same as in Krugman (1980): in each location j, M_j homogeneous plants produce differentiated varieties with constant elasticity of substitution κ 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 4.3 under an aggregate production function equal to:

$$\widetilde{Y_j}\left(L_j\right) = K_j \left(z_j\left(L_j\right) L_j\right)^{\frac{\kappa}{\kappa-1}}, \tag{A.41}$$

where $K_j \equiv \frac{\kappa-1}{\kappa}\left(\kappa F_j\right)^{\frac{1}{1-\kappa}}$ is a constant. Therefore, a monopolistic competition model with no productivity spillovers is equivalent to a homogeneous-product model with perfect competition and spillover elasticity equal to $\gamma^P = \frac{1}{\kappa-1}$. This property relates to the result, dating back to at least Abdel-Rahman and Fujita (1990) and 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 4.3 go through.[41]

Environment We start by describing how the physical environment of this model differs from the environments from sections 3 and 4.3. 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 $z_j\left(L_j\right) L_j = M_j\left(\sum_i \tau_{ji} q_{ji} + F_j\right)$ to account for the use of labor in setting up plants. Combining these two expressions, that constraint can be further expressed:

$$M_j^{\frac{1}{\kappa-1}}\left(z_j\left(L_j\right) L_j - F_j M_j\right) = \sum_i \tau_{ji} Q_{ji}. \tag{A.42}$$

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

$$\max \sum_i \left(\tilde{p_{ji}} - \tau_{ji} W_j\right) q_{ji} \tag{A.43}$$

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

$$p_j \equiv M_j^{\frac{1}{1-\kappa}} \frac{\kappa}{\kappa-1} W_j \tag{A.44}$$

As a result, condition (14) still determines the flows in the competitive equilibrium. Combining these pricing rules with (A.43), imposing zero profits and using (4) we obtain the number of producers in a competitive allocation:

$$M_j = \frac{z_j\left(L_j\right) L_j}{\kappa F_j}. \tag{A.45}$$

And further combining with A.42, we can write

$$\widetilde{Y_j}\left(L_j\right) = \sum_j \tau_{ij} Q_{ij} \tag{A.46}$$

for $\widetilde{Y_j}$ given in A.41.

[41] These properties would not go through under monopolistic competition outside of CES. In that case, the entry and bilateral pricing decisions would be inefficient.

We conclude that the competitive allocation can be represented as in the model without product differentiation from Definition 1 under the restrictions from Section 4.3 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) for expenditures x_i satisfying (15)); (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.46) hold.[42]

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

$$\begin{aligned}\mathscr{L} = u &- \sum_j \omega_j \left(u - a_j(L_j) U(c_j, h_j) \left(\frac{L_j}{L} \right)^{-\sigma} \right) \\ &- \sum_j p_j^* \left(\sum_i \tau_{ji} Q_{ji} - M_j^{\frac{1}{\kappa-1}} \left(z_j(L_j) L_j - F_j M_j \right) \right) \\ &- \sum_j P_j^* \left(L_j c_j - Q(Q_{1j}, .., Q_{Jj}) \right) - \sum_j R_j^* \left(L_j h_j - H_j \right) - W \left(\sum_j L_j - L \right) + ... \end{aligned} \tag{A.47}$$

Relative to Definition 2, now the planner also chooses the number of firms M_j in each location and faces the constraint (A.42) instead of (4). We readily see that entry is efficient by noting that the first-order condition with respect to M_j implies (A.45). As a result, the market clearing constraint in the second line of (A.47) can be replaced by (A.46). The resulting planning problem is equivalent to Definition (2) applied to the economic geography model in Section 4.3 under the production function $\widetilde{Y_j}(L_j)$.

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).

We build a dataset at the MSA-skill group level. We first use the BEA data to construct income data at the MSA level, split by source of income (labor, capital and transfers). We supplement it with data from Dunbar (2009) to get an MSA-level measure of taxes paid. This provides a picture of income and disposable income in each MSA. Second, we use March CPS data to allocate income and disposable income in each MSA to the two skill groups we retain in the analysis (high skill and low skill). We also use the rich socio-demographic information contained in the March CPS to control for within-group heterogeneity when building our variables. Finally, we use this dataset to calibrate the model. We detail this procedure here.

B.1 Appendix to Section 5.1 (Data Sources)

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. The Regional Economic Accounts report, at the MSA level, total income by source of income. 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*;[43] capital income as the sum of *proprietor's income*, and *dividends, interests and rents*; and transfers as *current transfer*

[42] Note that the definition of the competitive allocation can dispense with the wage W_j, which can be determined residually from (A.44).

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

receipts.[44] 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. In the absence of detailed information on taxes paid, we further apportion taxes paid between taxes on labor income and capital income, proportionally to the corresponding incomes.

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 which reports, at the individual level, information on various sources of income, as well as the MSA of residence and level of education of each worker. We group workers into two groups following Diamond (2016): we define high skill are those who have completed 4 years of college, or more; and low skill as 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.[45] We do not use this information directly, as the MSA aggregates from individual-level data might be noisy, especially in smaller MSA's. However, we use this information to construct the shares of the MSA-level outcomesfrom the BEA data corresponding to each group of workers. That is, for each MSA i, we compute $s_i^L = \frac{\widetilde{X}_i^L}{\widetilde{X}_i^L + \widetilde{X}_i^H}$ where $\widetilde{X}_i^\theta$ denotes total MSA i-group θ level capital income, labor income, transfers, taxes or population in the 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 level population, labor income, capital income, transfers and taxes. We also compute the corresponding per-capita measures for each MSA-group: $x_i^\theta = \frac{X_i^\theta}{L_i^\theta}$.

Finally, we compute a series of MSA-group level measures of detailed socio-economic-demographic characteristics. We do this in order to control, in the measures we use to quantify the model, for heterogeneity within group that is not modeled, but which we can observe in the data. We build measures that take out this heterogeneity in order to avoid attributing to a place (MSA) differences in composition of our groups. To that end, we use Census data (IPUMS), and aggregate from individual data to the MSA-group level to get the share of individuals with following characteristics for each MSA-group: age by bins :<20, 20-40, 40-60, >60; fine level of educational attainment: less that 8th grade, grade 9-12, some college (those are relevant for the low skill group) and bachelor, masters or professional degree (for the high skill group); share black; share male ; share unemployed, share out of the labor force ; share working in manufacturing, services, agriculture.

To control for heterogeneity within group in our quantification, we proceed as follows. Denote 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:

$$x_i^\theta = x_0^\theta + \sum_j \beta^{\theta j} DEM_i^{\theta,j} + \varepsilon_i^\theta,$$

where $DEM_i^{\theta,j}$ for $j \in 1, ..., J$ are all of the demographic measures explained above, in MSA i and for group θ. We then compute the predicted value of x_i^θ at the mean demographic characteristics for group θ across all MSA's, i.e.:

$$\hat{x}_i^\theta = x_0^\theta + \sum_j \beta^{\theta j} \overline{DEM^{\theta,j}} + \varepsilon_i^\theta$$

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

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

and we use the value $\hat{x}_i^\theta$ to measure a given per-capita outcome, stripped out of within group heterogeneity, in our final dataset. The corresponding stock variable, stripped of within group heterogeneity, are $\hat{X}_i^\theta = \hat{x}_i^\theta L_i^\theta$. The 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.[46]

In addition, we need data on the relative size of the non-traded sector in each city. 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, except it is not group-specific, and we 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. We supplement our dataset with this measure.

Income and expenditure The measures of population (L_i^θ in the model), labor income, transfers and taxes paid on labor income are directly taken from the dataset. Transfers are uniformly scaled so that they sum, in the aggregate, to the aggregate sum of taxes.[47] This ensures that the government budget constraint holds. Per-capita measures such as wages (w_i^θ=labor income/population), labor taxes (τ_i^θ =taxes/population) and transfers per capita (t_i^θ=transfers/population) immediately follow. We also use the measures of capital income in the data to compute b^θ, the per-capita share of national capital income owned by each individual in group θ. Since the model does not feature heterogeneity in capital income across cities, we shut down heterogeneity in capital holding by measuring the average share of the national portfolio owned in each group, and imputing it uniformly (per capita) in each city. In turn, we construct expenditure by group and by MSA, i.e. the variable x_j^θ in the model, as disposable income by group. We construct this variable using (21):

$$x_i^\theta = w_i^\theta(1-\tau_i^\theta) + t_i^\theta + b^\theta \Pi^H. \tag{A.48}$$

The variables $\left\{w_i^\theta, t_i^\theta, b^\theta\right\}$ are directly observed (see above). We must ensure a value for national profits and returns to land, Π^H, that is consistent with the general equilibrium of the model. For that we use profit maximization in the housing sector and market clearing in the non-tradeable sector to obtain the following expression for Π^H as function of calibrated elasticities or observable outcomes:

$$\Pi^H = \frac{(1-\alpha_C)\sum_i \frac{d_{H,i}}{d_{H,i}+1}\left(\sum_\theta w_i^\theta L_i^\theta - T_i\right)}{1-(1-\alpha_C)\sum_i \frac{d_{H,i}}{d_{H,i}+1}\sum_\theta b^\theta L_i^\theta}, \tag{A.49}$$

where T_i is total taxes minus transfers by MSA, $\sum_\theta w_i^\theta L_i^\theta$ is total labor income by MSA, and the elasticities α_C and $d_{H,i}$ of the model are calibrated as explained below. 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).

Trade Shares The last piece of information we need are 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) Monte et al. (2015), whereby we use the import shares from the CFS data in order to parametrize the elasticity of trade with respect to distance. In

[46] We only use taxes paid on non capital income in what follows. Income and taxes corresponding to capital are imputed in the procedure below to be consistent with the model, as explained below.

[47] Transfers are uniformly scaled down by 35% . The fact that total taxes and transfers do not match in the dataset comes in part from the fact that we have removed heterogeneity that is not place-specific from the data, and more importantly from our treatment of capital income and the corresponding taxes - they are computed to be consistent with the sources of capital income (profits from housing rents) in the model, so that their scale is lower than in the data.

particular, we start from the model expression for import shares,

$$s_{ji}^M = \left(\frac{\tau_{ji}}{P_i} \frac{W_j^{1-b_Y^I} P_j^{b_Y^I}}{z_j} \right)^{1-\sigma} = \left(\tau_{ji} \delta_j^D \delta_i^O \right)^{1-\sigma} \tag{A.50}$$

where δ_j^O and δ_j^D are origin and destination fixed effects. We assume that trade costs have the form $\ln \tau_{ji} = \psi \ln d_{ni} + e_{ni}$. We use Allen and Arkolakis (2014) estimate for ψ and set $\tau_{ji} = d_{ni}^{\psi}$. We then construct the smoothed import shares s_{ji}^M between MSA's using (A.50). For that, we need to obtain the values of $\left\{\delta_j^D, \delta_i^O\right\}$. Following and we know that they are uniquely pinned down, up to a normalization, by considering the identity that sales equals income

$$p_j Y_j = \sum_i s_{ji}^M E_i \tag{A.51}$$

together with equation (A.50) and the definition of the price index[48].which leads to

$$\left(\delta_i^O \right)^{\sigma-1} = \sum_j \left(\tau_{ji} \delta_j^D \right)^{1-\sigma} \tag{A.52}$$

. After plugging (A.50), and (A.52) in (A.51) we get a system N equations in N unknowns, which we solve to recover s_{ji}^M. The export shares are then constructed from the definition $s_{ji}^X \equiv \left(\frac{E_i}{p_j Y_j} \right) s_{ji}^M$.

B.2 Appendix to Section 5.2 (Calibration)

We provide details about the calibration of the intermediate input share in non-traded goods. We use the following equilibrium relationship that comes from the market clearing condition in the non traded sector in city i in the model:

$$1 - b_{H,j}^i = \frac{W_j N_j^{NT}}{(1-\alpha_C) X_j} \left(1 + d_{H,i}\right). \tag{A.53}$$

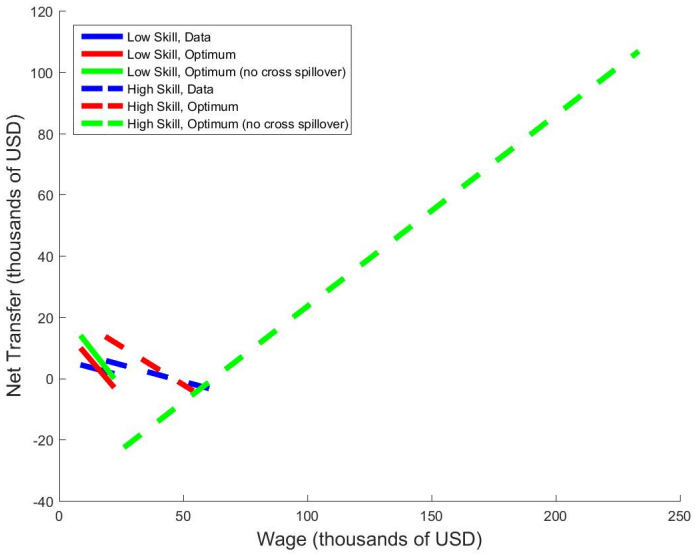
We compute this expression using our estimates of $W_j N_j^{NT}$ (coming from the share of workers in the non traded sector in MSA i in our dataset), X_i (total expenditure in city i) and our calibrated values for α_C and $d_{H,i}$. In practice, this step requires $\frac{W_j N_j^{NT}}{(1-\alpha_C) X_j} \left(1 + d_{H,i}\right) \in [0, 1)$ in all cities. However, for some cities we find $\frac{W_j N_j^{NT}}{(1-\alpha_C) X_j} \left(1 + d_{H,i}\right) > 1$. In these cases, we recalibrate the housing supply elasticity in the non-traded sector $d_{H,i}$ such that $\frac{W_j N_j^{NT}}{(1-\alpha_C) X_j} \left(1 + d_{H,i}\right) = 1$, and set $b_{H,j}^i = 0$.[49] For four MSA's where $\frac{W_j N_j^{NT}}{(1-\alpha_C) X_j} > 1$, this procedure requires an additional adjustment. In this case, we set $W_j N_j^{NT} = (1-\alpha_C) X_j$, which corresponds in practice to scaling down $W_j N_j^{NT}$ by less than 2 % in these cities.

[48] Formally, $P_i^{1-\sigma} = \sum_j \left(\tau_{ji} \frac{W_j^{1-b_Y^I} P_j^{b_Y^I}}{z_j} \right)^{1-\sigma}$.

[49] This procedure in turn reduces slightly the profits generated in the economy, in turn reducing the implied values of Π^H and X_i whose construction we described in the previous subsection of the appendix, and this in turn modifies (A.53). We iterate to find the unique fixed point of the procedure.

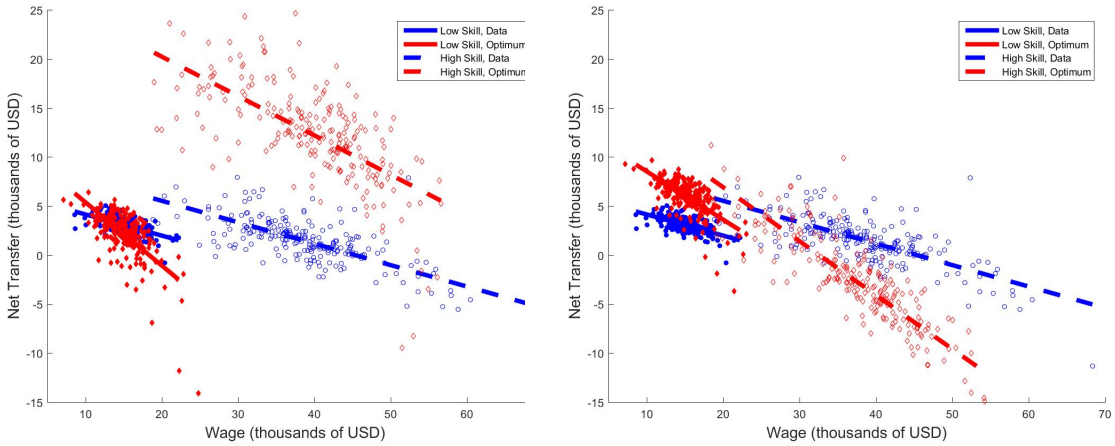
C Appendix to Section 6

Figure A.1: Per Capita Transfers by Skill Level and MSA, Comparison Against Case without Cross-Spillovers for High Skill Workers



Note: this figure replicates the slopes of Figure 3 (blue and red lines). In addition, it shows in green the best linear fit for the optimal transfers of low skill workers (in solid line) and high skill workers (dashed) assuming no spillovers from high skill workers to low skill workers ($\gamma^A_{SU} = \gamma^P_{SU} = 0$). Each linear schedule is displayed on the support of its corresponding variable.

Figure A.2: Per Capita Transfers by Skill Level and MSA under high weight for high skill workers (left panel) and high weight for low skill workers (right panel).



(a) High Weight for High Skill Workers

(b) High Weight for Low Skill Workers

Note: each panel replicates Figure 3 under alternative weights for each skill group in the planning problem. In the left panel, the weight is 10 times larger for high skill workers. In the right panel, the weight is 10 times larger for low skill workers. For details of how the data is constructed see Appendix B. The slopes of the linear fit of the optimum transfers (with SE) for the left panel are: Low Skill: -0.64 (0.05); High Skill: -0.40 (0.04). The slopes of the linear fit of the optimum transfers are: Low Skill: -0.49 (0.04); High Skill: -0.55 (0.02). In the data the slopes are: Low Skill: -0.22 (0.02); High Skill: -0.22 (0.01).