NBER WORKING PAPER SERIES

OPTIMAL SPATIAL POLICIES

Pablo D. Fajgelbaum
Cecile Gaubert

Working Paper 33493
http://www.nber.org/papers/w33493

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2025

We thank Arnaud Costinot for his discussion of the paper and Junseok Lee and Tom Basin for excellent research assistance. Cecile Gaubert acknowledges support from NSF CAREER grant #1941917. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Optimal Spatial Policies
Pablo D. Fajgelbaum and Cecile Gaubert
NBER Working Paper No. 33493
February 2025
JEL No. H21, H23, R12, R13

ABSTRACT

We summarize recent methods to study optimal spatial policies. We center the discussion on policies that implement the optimal distribution of population in the presence of spatial spillovers, spatial transfers to optimally tackle redistribution between rich and poor regions, and optimal transportation investments.

Pablo D. Fajgelbaum
Department of Economics
University of California, Los Angeles
8283 Bunche Hall
Mail Stop: 147703
Los Angeles, CA 90095
and NBER
pfajgelbaum@gmail.com

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

1 Introduction

Governments around the world make widespread use of place-based policies that target resources to specific regions. In the U.S., notable examples include the historical Tennessee Valley Authority as well initiatives like Opportunity Zones, which provides tax breaks to incentivize private investment in economically distressed areas. The European Union Cohesion Policy allocates funds with the goal of reducing regional disparities via a range of initiatives amounting to a third of the EU budget. Another example is China's Western Development Strategy, which aims at developing western regions and reducing the economic gap with more prosperous coastal areas. An abundant empirical literature has examined the effects of many different place-based policies (see for example Neumark et al. (2015) and Ehrlich and Overman (2020) for reviews).

What does economic theory have to say about the rationale for such interventions? In this chapter, we review recent theoretical analyses of policies that reallocate resources across space. We examine how recent developments in spatial equilibrium models help to theoretically characterize and quantitatively solve for optimal spatial policies. We discuss two main motives for government intervention: efficiency and equity. Under the efficiency rationale, spatial policies aim at maximizing national welfare. Under the equity rationale, spatial policies aim at reducing inequality in economic outcomes.

Our general theoretical treatment of optimal policies is focused on implementing first-best taxes and subsidies that vary across space. We also discuss optimal local investments, with a focus on transportation. We use the term "spatial policies" to refer to these particular interventions. In reality, place-based policies take a variety of forms. We exclude from our theoretical discussion many specific policies that target geographic areas such as housing policies, zoning, local education investments, or unemployment insurance. Our treatment also leaves aside a discussion of second-best policies (i.e., to what extent can these policies come close to achieving an efficient allocation).

We first address the efficiency motive. Cities and regions serve as focal points for numerous non-market interactions. The concentration of workers in specific areas generates agglomeration spillovers that increase productivity, while the concentration of residents enhances local amenities, like social networks and entertainment options. These spillovers are external to households, as individuals do not internalize the broader effects of their location choices on others. As a result, market allocations are typically inefficient, creating a justification for spatial policies to address these externalities.

In section 2, we introduce a stylized model based on the Rosen-Roback framework with spillovers to explain the core trade-offs and insights for efficient spatial policies. Section 3 extends the analysis to more complex models, in particular quantitative spatial models, which incorporate additional factors such as input-output linkages and heterogeneous worker types. This section also explains how to take these theoretical frameworks to the data and quantify optimal spatial policies, and illustrates these methods with an application to the distribution of economic activity in the U.S.

Section 4 shifts the focus to the equity motive for spatial policies. Many countries face significant disparities in wages, incomes, and overall economic conditions between regions, with poverty often concentrated in specific areas. This concentration of poverty often motivates policy makers to subsidize laggard regions. We analyze the optimal design of policies aimed at redistributing resources to economically disadvantaged regions. Such policies are often suspected to be inefficient: a commonly held view is that low-income people should be subsidized based on their income, irrespective of where they live. A central question addressed in section 4 is why, in contrast to this view, place-based policies aimed at redistribution may fruitfully be used to complement traditional income-based tools, such as income taxes.

We devote Section 5 to examining optimal transportation infrastructure investments as a particular type of spatial policy. Investments in transportation infrastructure are often central to government policy discussions, and are likewise a major topic in the urban, regional, and trade literatures. We treat investment in transportation infrastructure separately from other policies because of the distinctive mathematical tools and methods required to analyze investments in these networks. The section discusses the role of transportation infrastructure in facilitating economic integration across regions and quantifies the optimal level and distribution of such investments as found in the literature. Section 6 concludes with a summary of areas of ongoing research.

2 Efficient Spatial Allocation

2.1 Introduction

A central topic in spatial economics is the role of *agglomeration spillovers*, which boost productivity in densely populated areas and drive the formation of cities. These spillovers may go hand in hand with negative effects of density, such as traffic congestion and pollution. At its core, the spatial equilibrium is characterized by numerous externalities. As a result, market outcomes may be suboptimal, with some areas that are too large while others remain under-developed. This inefficiency occurs because individuals do not internalize the spatial spillovers they generate when they make location choices. In this context, policy intervention can be justified on efficiency grounds.

What is the optimal distribution of population across locations? How far from the efficient allocation is the observed spatial equilibrium? What policies may achieve efficiency? This section answers these questions through the lens of state-of-the-art frameworks in urban and spatial economics. These canonical themes have been addressed by two classic strands of the literature. The first, following Henderson (1974), assumes that land is infinitely abundant and (typically homogeneous) cities can enter freely, with each city able to choose its optimal size. A second strand, following Flatters et al. (1974), considers a pre-existing set of heterogeneous locations producing local public goods. Our treatment builds upon this second strand, having as a backdrop the recent quantitative spatial models (QSE, see chapters by Allen and Arkolakis and Redding in this handbook). The QSE literature grew out of spatial models with trade frictions such as Krugman (1991) and especially, given modeling details, Helpman (1998). A country is partitioned into many locations, each of which is the locus of various spatial externalities. Workers freely choose where to live and work. The models are typically enriched to fit many margins observed in the data, such as input-output linkages and heterogeneous worker types. In this context, we consider a central planner who may intervene in the market allocation to implement the efficient distribution of population across locations.[1]

The treatment builds largely on Fajgelbaum and Gaubert (2020). Compared to previous presentations of these topics, the main message emerging from this treatment is that the optimal allocation can always be implemented with location-specific Pigouvian subsidies that can be expressed as a function of only spillover elasticities and market outcomes, such as wages and expenditures, at the optimal allocation. The same optimal policy formula holds regardless of many model features, including the microfoundation for the spillover and whether the supply side is complex as in spatial quantitative models or simpler as in earlier urban frameworks. As a result, numerically implementing the optimal allocation and subsidies involves computing a specific model counterfactual in response to a shock –a standard approach in the QSE literature– where the shock implements the spatial subsidies prescribed by that optimal policy formula.

This section lays out the main logic and intuitions in a purposefully stylized framework –a simple version of the Rosen-Roback model augmented with spatial spillovers. We discuss how to characterize optimal spatial policies in this model and discuss their key properties. The next section extends the analysis to cover various assumptions typically made in the QSE literature. These extensions concern the structure of production (heterogeneous goods, input-output linkages, costly trade); the nature and specification of spatial spillovers (spillovers coming from public good provision, heterogeneous spillovers by groups of workers, spillovers that decay over space); and the level of aggregation of the spatial model (regional model of cities, or urban model of locations within cities). The next section also discusses how the optimal allocation and optimal policies can be quantified in a realistic quantitative spatial model. It presents a step-by-step methodology and illustrates some of the results using data on the distribution of economic activity in the U.S.

2.2 Baseline Model

2.2.1 Setup

Preferences and Technology The stylized framework in this section has the minimal ingredients needed to discuss optimal spatial policies. Consider an economy with $j = 1, .., J$ ex-ante heterogeneous locations and L homogeneous workers. Each individual resides and works in the same location. The utility of a resident of location j is

$$u_j\left(c_j, h_j, L_j\right), \tag{1}$$

where c_j is per-capita consumption of a freely traded commodity and h_j is per-capita consumption of housing, a non-tradable commodity. The utility function $u_j\left(\cdot\right)$ can be non-homothetic, accommodating for example the fact that housing consumption has low income elasticity. Utility is

[1] Several excellent reviews of these topics include the handbook chapter by Abdel-Rahman and Anas (2004) on efficiency issues in systems-of-cities models, the handbook chapter by Wildasin (1987) (and its extended version, Wildasin, 1986a) on efficiency in the local public finance literature, and the books by Henderson (1977) and Fujita (2013), which discuss efficiency in both approaches.

indexed by location j to capture exogenous differences in amenities or quality of life in a location. Thus, some locations are nicer to live in, all other variables equal. Finally, one's preferences for a location may also depend on local amenities that respond endogenously to local population. To capture this notion formally, utility may depend directly on local population L_j, conditioning on the previous variables.

On the production side, the aggregate supply of tradable output $Y_j(N_j)$ in location j is a function of labor services N_j. It is indexed by location j as some locations may be more productive for the same amount of labor services. Each worker supplies one unit of labor with productivity $z_j(L_j)$, so a total of $N_j(L_j) = z_j(L_j) L_j$ efficiency units are available in location j. Through the function $z_j(.)$, labor productivity may increase with the local density of economic activity. Finally, housing supply is assumed fixed at an exogenous endowment H_j in location j.

The key driver of inefficiency in this model, which justifies policy intervention, is that the location choice of a worker leads to spillovers on local economic activity. These spillovers are typically not internalized in location decisions. Two types of spillovers are modeled: an amenity spillover that directly impacts other workers' utility through the L_j term in (1), and a productivity spillover that impacts other workers' output through the $z_j(L_j)$ term. These functions are left general at this point; the chapter later discusses how various microfoundations used in the literature map into this reduced-form formulation of the spatial spillovers.

Spatial Mobility Constraint Workers are perfectly mobile across space. Thus, in any feasible allocation, there exists an economy-wide level of utility u such that:

$$u_j \leq u, \tag{2}$$

with equality for all inhabited locations. The free-mobility constraint means that workers arbitrage away spatial differences in welfare. It constitutes a natural benchmark to study efficiency in a spatial equilibrium, especially when thinking about long-run policy objectives. However, this assumption may seem unrealistic especially in the short run. One way forward would be to study spatial policies in a dynamic model with moving frictions. Such models are developing in the literature (Caliendo et al., 2019; Allen and Donaldson, 2020). Studying optimal policies in these contexts a natural avenue for future research. Alternatively, a popular modeling approach in the literature has been to capture mobility frictions while retaining a static model, by breaking away from the assumption of homogeneous workers. These papers introduce heterogenous workers with different idiosyncratic attachments to locations, so that aggregate mobility responses to shocks are limited. Section 4 of this chapter discusses spatial policies in such environments and shows how they map to the discussion here.

The free mobility condition (2) imposes a restriction on policy instruments: allocations where identical agents experience different welfare are unfeasible.[2] This restriction rules out equilibria

[2]Imposing this restriction on feasible allocations rules out the unequal treatment of equals. Incentives for such unequal treatment of equals arise from the perspective of a planner maximizing a social welfare function (Mirrlees, 1972; Wildasin, 1986b), because utility is equalized across locations in the market equilibrium but marginal utilities (of consumption of tradables) generically are not.

with rationing where population size can be capped in a location.

Feasible Allocations A feasible allocation $\{c_j, h_j, L_j, N_j, u\}$ satisfies the following constraints:

$$u_j(c_j, h_j, L_j) = u \text{ if } L_j > 0 \text{ (and } \leq u \text{ otherwise)} \tag{3}$$

$$\sum_j c_j L_j \leq \sum_j Y_j(N_j) \tag{4}$$

$$h_j L_j \leq H_j \tag{5}$$

$$N_j \leq z_j(L_j) L_j \tag{6}$$

$$\sum_i L_j = L \tag{7}$$

In a non-wasteful allocation where all constraints bind, housing consumption per capita h_j as well as labor supply N_j are uniquely determined by local population L_j (equations (5) and (6)). An equilibrium is therefore fully characterized by two margins: consumption of tradables c_j and population L_j. Furthermore, for a given level of welfare in the economy u, the free mobility condition (3) ensures that the level of L_j pins down that of c_j, and vice-versa. Feasible allocations are therefore fully characterized by local consumption of traded goods. This local consumption can be arbitrary, and in particular different from local production, so long as the aggregate market for traded goods clears, i.e. (4) holds. Put differently, allocations are defined by a set of *transfers* of tradable goods across locations. In particular, an optimal allocation will be characterized by a specific set of transfers between places.

2.2.2 Market Allocation

In a decentralized equilibrium, each worker chooses location and consumption to maximize utility, while perfectly competitive producers hire labor to maximize profits. Producers pay W_j per efficiency unit of labor. Being atomistic, they do not internalize that worker's productivity responds to local employment L_j. Letting the tradable commodity be the numeraire, producers in location j maximize profits $\Pi_j^Y = Y_j(N_j) - W_j N_j$, hence the wage per unit of labor is:

$$W_j = Y_j'(N_j). \tag{8}$$

A worker in location j earns this wage per efficiency unit times her number of efficiency units z_i, themselves a function of population

$$w_j = W_j z_j(L_j). \tag{9}$$

Total unearned income in the economy includes the returns to housing and firm profits. The endowment of housing is rented at the market rate R_j, so that the returns to housing in location j are $R_j H_j$. All agents in the economy are assumed to be identical, hence they hold the same endowment of fixed factors. Per capita, the net returns to these fixed factors are:[3]

$$\pi \equiv \frac{1}{L}\sum_j \Pi_j^Y + \frac{1}{L}\sum_j R_j H_j. \tag{10}$$

A central government may put in place taxes and transfers on labor and non-labor income. Such a policy is summarized in each location by a net per-capita subsidy s_j. Conditional on living in j, a worker earns after-tax income

$$x_j \equiv w_j + \pi + s_j. \tag{11}$$

The subsidy s_j may be a function of market outcomes –for instance, it may depend on income. When a policy redistributes only within each location (e.g., a local income tax rebated to local workers), then $s_j = 0$. In contrast, a non-trivial formulation of this policy entails transfers across space. We refer to the set $\{s_j\}$ as a *spatial policy.* It must be revenue neutral, i.e.:

$$\sum_j L_j s_j = 0. \tag{12}$$

Conditional on living in j, a worker obtains the following indirect utility

$$v_j\left(x_j, R_j, L_j\right) \equiv \max_{c_j, h_j} u_j\left(c_j, h_j, L_j\right) \quad s.t. \quad c_j + R_j h_j = x_j. \tag{13}$$

By free mobility, worker resides in a location that yields maximum utility,

$$u = \max_j v_j\left(x_j, R_j, L_j\right), \tag{14}$$

and all populated locations deliver the maximum utility.

A competitive allocation –or spatial equilibrium– given s_j is a set of quantities $\{c_j, h_j, L_j, N_j\}$, utility u and prices $\{w_j, R_j\}$ such that the allocation is feasible according to (3)-(7), firms optimize (so wages are given by (9)) and the consumer solves (14)-(13).

2.2.3 Links to Well-Known Frameworks

In this model, two elasticities determine misallocation and optimal spatial policies: the productivity spillover ε_{z_j,L_j} and the amenity spillover ε_{v_j,L_j}, where we use $\varepsilon_{F,X} \equiv \frac{\partial F}{\partial X}\frac{X}{F}$ to denote the elasticity of F with respect to X. These elasticities capture welfare-relevant impacts of population that are not internalized by market prices. The former captures the increase in labor productivity caused by an increase in local population while the second captures the increase in indirect utility of residents directly caused by an increase in local population. For normative purposes, the utility spillover elasticity only matters up to the curvature of indirect utility, so the term we will often encounter in calculations of optimal policy is $\frac{\varepsilon_{v_j,L_j}}{\varepsilon_{v_j,x_j}}$. The normalization captures that monotone transformations of utility functions are irrelevant for optimal allocations and have no empirical counterpart.

The specific microfoundation of these functions varies across models. The environment set forth

[3] Some papers in the literature tie unearned income to location by assuming that fixed factors are locally owned; as we discuss below, any such deviation from a national, commonly owned portfolio of fixed factors creates a distortion to the allocation of population. We show below how adapt the analysis to that case.

in the baseline model nests several classic frameworks, briefly reviewed here.

Neoclassical Models Neoclassical models as in Roback (1982) assume that both utility $u_j(\cdot)$ and labor productivity $z_j(.)$ are independent from L_j.

Systems-of-Cities Models in the tradition of the systems-of-cities literature following Henderson (1974) are based on the monocentric city model. Cities combine a positive external agglomeration force and a (typically non-external) congestion force, leading to a inverted U-shape relationship between city-level utility and local population. In a common model variant, efficiency units are $N_j = L_j^{\gamma z} h_j(L_j)$, where $L_j^{\gamma z}$ is a positive productivity spillover. At the same time, productive hours per capita $h_j(L_j)$ decrease with city size: as cities grow, they expand geographically to host increasing population, the average commute is longer, and average workers' productive hours decrease. Because land is commonly owned among the population of a city, this reduction in productive hours is not fully internalized and acts as a negative externality.[4] On net, a city's output displays an inverted-U shape with its population L_j, increasing with agglomeration effects at a small scale, but decreasing as congestion ends up dominating at a larger scale (Abdel-Rahman and Anas, 2004). With a fixed number of cities, this setup maps into the one discussed here. However, different from our focus, models of systems-of-cities typically consider an infinite supply of potential city sites, and are concerned with the endogenous number of cities in equilibrium for different governance structures.

Local Public Finance In the local public finance literature, the source of spatial spillovers is public-goods provision, where local public goods are provided by local governments. For instance, Flatters et al. (1974) assume the utility function $\tilde{u}_j(c_j, h_j, g(G_j, L_j))$ where G_j is local public spending (produced 1-1 from tradable goods) and $g(G_j, L_j)$ are the resulting public services provided to each resident. Congestion or non-rivalry of public-service provision enters through the shape of $g(G_j, L_j)$. The discussion of efficiency in this section goes through unchanged in this case by defining $u_j(c_j, h_j, L_j) = \tilde{u}_j(c_j, h_j, g(G_j, L_j))$, where the (possibly socially inefficient) distribution of G_j is taken as given. In this case, the central government finances spending on public goods.[5] A more complete treatment of this case, however, requires taking into account that G_j may be financed by local governments through local taxes rather than by a central planner, a theme discussed in section 3.2.

New Economic Geography In new economic geography models (Krugman, 1991; Helpman, 1998), the standard microfoundation for spillovers is that larger cities host a greater variety of producers of goods that can be only locally consumed or used as intermediate inputs. In the presence of love-of-variety in consumption or production, entry of more producers in larger cities generates spillovers. Formally, differentiated inputs (or final goods) are combined in these models

[4]When the city is circular, this process leads to $h_j(L_j) = 1 - kL_j^{1/2}$ for some constant k.

[5]I.e., (12) is replaced by $\sum_j L_j s_j(\cdot) = \sum G_j$.

using a Constant Elasticity of Substitution function, with elasticity of substitution σ. As noted by Abdel-Rahman and Fujita (1990) this microfoundation generates productivity (respectively, amenity) spillovers that have constant elasticity $\varepsilon_{z_j,L_j} = \frac{1}{\sigma-1}$ in the case of intermediate inputs (respectively, $\frac{\varepsilon_{v_j,L_j}}{\varepsilon_{v_j,x_j}} = \frac{1}{\sigma-1}$ in the case of final goods).

Quantitative Spatial Economics Quantitative spatial models (Redding and Rossi-Hansberg, 2017) typically assume spillovers similar to those in the New Economic Geography literature described earlier, or use reduced-form spillovers without specifying their microfoundations. Spillovers are assumed to have constant elasticity, both as a function of location (they are independent from j) and as a function of population (their elasticity is independent from L_j). On the production side, reduced-form spillovers $z(L_j)$ represent classic Marshallian economies of density such as knowledge spillovers, enhanced access to consumers and suppliers, or labor pooling. On the utility side, workers derive utility from amenities in the place where they live, usually parametrized in multiplicative form: $u_j = a_j(L_j)\,\tilde{u}(c_j, h_j)$. Amenity spillovers may be positive, driven by economies of density in amenities (e.g. diversity in entertainment options and restaurants) or negative, driven by the additional pollution and traffic generated by higher density. Finally, these models assume that preferences are homothetic, so that $\varepsilon_{v_j,x_j} = 1$. Given that these assumptions are both tractable and widespread, we will use them repeatedly to illustrate the analysis, based on the following example:

Example 1. Spillover Elasticities are constant both across space and as a function of population size, $\varepsilon_{z_j,L_j} = \gamma_Z$ and $\frac{\varepsilon_{v_j,L_j}}{\varepsilon_{v_j,x_j}} = \gamma_A$.

2.3 Welfare-increasing Reallocations

This section takes a small perturbation approach to characterize reallocations that increase welfare in the baseline model. We derive a condition that determines when an allocation is efficient. Using that condition we characterize several properties of the equilibrium.

2.3.1 What Places Should Be Subsidized in an Observed Allocation?

Starting at a spatial equilibrium of the economy, we consider a small reallocation of workers between locations $\{dL_j\}$, while keeping total population constant. To a first-order approximation, such a reallocation leads to a change in aggregate welfare of

$$\frac{du}{\chi} = \sum_j (\Gamma_j - x_j + w_j)\, dL_j, \tag{15}$$

where

$$\Gamma_j \equiv \varepsilon_{z_j,L_j} w_j + \frac{\varepsilon_{v_j,L_j}}{\varepsilon_{v_j,x_j}} x_j \tag{16}$$

and where

$$\chi \equiv \left(\sum_j L_j \left(\frac{\partial v_j}{\partial x_j} \right)^{-1} \right)^{-1}. \tag{17}$$

The term Γ_j captures, in money-equivalent terms, the spillover effect of adding a worker to location j. The term χ is the average marginal utility of expenditure in the economy. So, $\frac{du}{d\chi}$ is interpreted as an aggregate equivalent variation, i.e., an increase in income such that, if properly allocated across regions while keeping initial location and prices fixed, leaves the country indifferent with experiencing the labor reallocation.[6]

The formal derivation of (15) is given in the Appendix. Intuitively, the term $\Gamma_j - x_j + w_j$ measures the social welfare impact of an additional worker in location j. The marginal worker in j increases production, both directly through that worker's labor (as measured by w_j) and indirectly through productivity spillovers that benefit incumbent workers (as measured by $\varepsilon_{z_j,L_j} w_j$), so that the social marginal product of labor is $\frac{dY_j}{dL_j} = \left(1 + \varepsilon_{z_j,L_j}\right) w_j$.[7] A marginal worker also generates utility spillovers on incumbent workers equal to $\frac{\varepsilon_{v_j,L_j}}{\varepsilon_{v_j,x_j}} x_j$ per worker when measured in consumption-equivalent terms. At the same time, an additional worker in j takes up local resources (the worker's own consumption x_j). Summing these terms gives the marginal welfare effect of adding a worker to location j, and summing over all cities for a given reallocation $\{dL_j\}$ yields (15).

Noting that $x_j - w_j = s_j + \pi$, it follows that a reallocation of workers from location i to j (keeping population in all other locations constant) would raise aggregate welfare u if and only if:

$$\Gamma_j - s_j > \Gamma_i - s_i. \tag{18}$$

So long as this condition holds in a market allocation, a worker has a larger welfare impact in location j than in i, and location j should grow. This can be achieved by increasing subsidies given to inhabitants of j while reducing i's.

What locations in a given market allocation are too small and should be subsidized? Naturally, everything else the same, locations with more positive (negative) spillovers should be subsidized (taxed). Beyond this observation, (18) provides a first answer to this question as well as an approach to implement welfare-increasing policies, because it depends on variables that are typically observable (wages and expenditures) as well as on elasticities that can a priori be measured. Ranking locations according to their observed $\Gamma_j - s_j$ in (16) helps establish priority to receive further subsidies. For a given (small) amount of subsidies to distribute, welfare will improve if a city higher up in this ranking gets more extra subsidy than one ranked below. A ranking of these cities therefore provides a preliminary measure of "bang-for-the-buck" from implementing subsidies in the observed allocation.

Measuring returns to spatial policies in this way is an inherently "local" exercise concerning small changes compared to the observed equilibrium. An advantage of such an approach is that it relies on elasticities that hold in the current equilibrium and are a priori measurable. It does not require further assumptions on how the equilibrium behaves far away from the observed allocation.

[6] Utility changes measured by du depend on unobservable monotone transformations of u entering through $\frac{\partial v_j}{\partial x_j}$; however, $\frac{du}{d\chi}$ is independent from those transformations. More specifically, $\frac{du}{d\chi} = \sum L_i ev_i$ where ev_i is such that that, around an initial equilibrium, a resident of i is indifferent between receiving an extra ev_i in income and continuing to live in location i at the initial prices, or experiencing the shock du caused by a counterfactual.

[7] The number of workers impacted by the spillovers L_j does not explicitly enter this expression because spillovers are expressed as elasticities (% terms, rather than levels).

However, on the flip side, these subsidies do not necessarily align with subsidies that would yield the globally optimal allocation. Any change in transfers starting at an inefficient allocation potentially changes the ranking of locations. In the next subsection, with further assumptions, we characterize optimal transfers that implement the globally efficient allocation. And in the next section we illustrate both this first-order and exact approaches to quantifying optimal policies.

2.3.2 *Laissez-Faire* Allocation: When is it Efficient?

Condition (18) reveals that the market allocation is generically inefficient except in knife-edge cases. One such case is the *laissez-faire* equilibrium, a standard benchmark. As it features no transfers, meaning $s_j = 0$ everywhere, (18) implies this allocation is efficient if the money-equivalent value of the spillovers Γ_j is the same everywhere. As we would expect, this is the case when there are no spillovers, $\varepsilon_{z_j,L_j} = \varepsilon_{v_j,L_j} = 0$, as in neoclassical frameworks such as Roback (1982). Unsurprisingly, the neoclassical spatial equilibrium without distortions is efficient.

More generally, condition (18) also reveals that the market allocation is efficient when utility and efficiency spillovers are constant *and* exactly compensate each other. To see why, assume there are constant-elasticity spillovers (Example 1). Starting again from $s_j = 0$ (*laissez-faire*), we have $x_j = w_j + \pi$ and condition (18) implies gains from reallocation from i to j as long as:

$$(\gamma_A + \gamma_Z)\,(w_j - w_i) > 0. \tag{19}$$

Wage differences ($w_j \neq w_i$) is the typical outcome due to compensating differentials. Therefore, generically, there are gains from reallocating workers in space when spillovers have constant elasticity (a point we return to in the next subsection). However, in the knife-edge case with $\gamma_A + \gamma_Z = 0$, amenity and efficiency spillovers exactly cancel out, leading to efficiency (a point noted by Allen and Arkolakis (2014)).[8]

2.3.3 Inefficiency of *Laissez-Faire* under Constant-Elasticity Spillovers

We just proved that the *laissez-faire* allocation is generically inefficient. This is the case even when spillovers have constant elasticity. The critical factor for policy is the money equivalent value of the spillover, Γ_j, which varies across locations even when the elasticities are constant. This result stands against a prior in the place-based policy literature, according to which the case with constant-elasticity spillovers is efficient, and therefore policy intervention is justified only if spillover elasticities are heterogeneous across space. Such is the conclusion in Glaeser and Gottlieb (2008) and Kline and Moretti (2014), echoed in Neumark et al. (2015) and Duranton and Venables (2018). In these papers, this conclusion arises because they study the welfare impacts of labor reallocations while assuming zero transfers. However, transfers (which create imbalances between the production

[8]These results may seem counterintuitive in the light of how optimal industrial policies look like under sector-level external effects. As shown for example by Bartelme et al. (2019), when spillovers are common across sectors there are no incentives to subsidize sectors; whereas here a constant spillover in space still leads to incentives for location-specific subsidies.

and consumption of tradeables), are essential for incentivizing spatial reallocations when there is free labor mobility.

More specifically, assuming away transfers is problematic for three reasons. First, without transfers, only one allocation –the market allocation– is consistent with free labor mobility. Any alternative allocation would involve differences in utility across space, conflicting with the concept of free mobility; put in another way, any other zero-transfer allocation would require forced migration. Second, transfers enable Pareto improvements (regardless of whether labor mobility is free or constrained by forced migrations). Thus, if transfers are allowed, the zero-transfers market allocation is inefficient. Third, typical place-based policies –such as wage subsidies implemented by a federal government– involve payment imbalances across regions, with some regions as net contributors and others as net beneficiaries. Therefore, transfers are necessary to think about these policies. In sum, the result that the allocation is inefficient under constant-elasticity spillovers differs from this previous research because our analysis accounts for transfers and for their effects on welfare and mobility.

2.3.4 Inefficiency from Local Ownership

Consider a model without spillovers but with local ownership of fixed factors. In this vein, Caliendo et al. (2018) assume that a fraction ω of the returns to local factors accrue to local residents while Redding (2016) assume full local ownership ($\omega = 1$). Full local ownership is also a standard assumption in systems-of-cities models (Abdel-Rahman and Anas, 2004). Formally, denote π_j the local returns to fixed factors relative to population of j, such that in the laissez-faire allocation:

$$x_j = w_j + \omega \pi_j + (1 - \omega) \pi. \tag{20}$$

Given condition (16), in the absence of spillovers ($\Gamma_j = 0 \; \forall j$), there are gains from reallocation in this economy: condition (18) becomes $-x_j + w_j > -x_i + w_i$, which holds whenever $\pi_i > \pi_j$. This result means it is desirable to reallocate workers towards locations with lower returns to fixed factors. This reallocation occurs through transfers intending to restore the common-ownership assumption, which corresponds to $\omega = 0$. Hence, the assumption on local ownership of fixed factors creates a distortion that gives incorrect signals to individuals who choose a location. Location choices become tied to not just marginal contributions to net social benefits, but also to high rents per se. This process results in high-rents regions being too large.

2.3.5 *Laissez-Faire* Allocation: When are Large Cities too Large?

A natural question in the context of optimal spatial population is whether larger or smaller cities in a free market allocation are undersized. Intuitively, the answer to this question must depend on the signs of spillovers and the distribution of fundamentals. We can answer this question by applying (18) to the free market allocation. Whether larger cities should grow depends on whether they are per-capita rich (in wages or in expenditures) and on the sign and intensity of their spillover elasticities. Said another way, once we know the sign and intensity of spillover elasticities, we need to determine if larger cities are richer to determine if they are too small or too large.

For ease of exposition, imagine a case of our model without housing in consumption. We also assume constant-elasticity spillovers (Example 1) as well as a constant-elasticity production function for tradables $Y_j(N_j)$. Cities with better exogenous components of efficiency (a shifter of Y_j) or amenities (a shifter of v_j) are larger under standard conditions that guarantee a unique interior solution. Alongside the spillover elasticities, a key parameter in this model is $\varepsilon_{Y,N} < 1$, the curvature of the production function with respect to efficiency units. Larger cities pay lower wages whenever $(1+\varepsilon_{z,L})\,\varepsilon_{Y,N} < 1$; in this case, efficiency spillovers $\varepsilon_{z,L}$ are weaker than decreasing returns to scale in production. At the same time, as implied by (19), low-wage cities should grow whenever net spillovers are negative ($\gamma_Z + \gamma_A < 0$).

Table 1 answers whether large cities are too large or too small in the free market allocation in different parameter regions. Large cities are too small only when efficiency spillovers γ_Z are either too large or too small. When the agglomeration spillover γ_Z is small, larger cities pay lower wages and net agglomeration is negative, indicating that these cities should grow. Symmetrically, when γ_Z is large, larger cities pay higher wages and net agglomeration is positive, indicating that these cities should grow. In the right column of the table, large cities are high wage, implying an urban wage premium in the free market allocation. In this case, large cities are too small if and only if spillovers are net agglomerative ($\gamma_Z + \gamma_A > 0$).

Table 1: Are Large Cities Too Large or Too Small in the *Laissez-Faire* Allocation?

	$(1+\gamma_Z)\,\varepsilon_{Y,N} < 1$	$(1+\gamma_Z)\,\varepsilon_{Y,N} > 1$
$\gamma_Z + \gamma_A < 0$	Too small (and low wage)	Too large (and high wage)
$\gamma_Z + \gamma_A > 0$	Too large (and low wage)	Too small (and high wage)

Note: Each entry of the table states whether, in a pairwise comparison of two cities in a free-market allocation, the larger city is too small (hence it should grow at the expense of the smaller city for aggregate welfare to increase) or too large. The table also notes whether the lager city is lower- or higher-wage in the observed allocation given the region of the parameter space.

2.4 Efficient Allocation

This section characterizes an efficient allocation and the spatial policies that may implement it.

2.4.1 Optimal Spatial Policies

If the equilibrium is efficient, then no reallocation of existing population should have a first-order impact on welfare. Therefore, condition (18) must hold with equality for any pair of locations. Using the definition of Γ_j, it follows that the subsidies that implement the optimal allocation must satisfy:[9]

$$s_j^* = \underbrace{\varepsilon_{z_j,L_j} w_j + \frac{\varepsilon_{v_j,L_j}}{\varepsilon_{v_j,x_j}} x_j}_{\Gamma_j} - \Gamma \tag{21}$$

where, using (12), the constant Γ equals the population-weighted average spillovers Γ_j:

$$\Gamma = \sum_j \left(\frac{L_j}{L}\right) \varepsilon_{z_j,L_j} w_j + \sum_j \left(\frac{L_j}{L}\right) \frac{\varepsilon_{v_j,L_j}}{\varepsilon_{v_j,x_j}} x_j.$$

The optimal transfer is a standard Pigouvian correction and it can be interpreted in terms of marginal costs and benefits of allocating workers to a given location. By definition, the subsidy s_j^* is the fiscal cost for the government of adding a worker to location j. The part of the social benefit generated by this worker in j that is not internalized by the market includes the efficiency spillover $\varepsilon_{z_j,L_j} w_j$ and the amenity spillover $\frac{\varepsilon_{v_j,L_j}}{\varepsilon_{v_j,x_j}} x_j$. If the worker does not locate in j, she still generates spillovers on the economy (the average Γ). At the optimal equilibrium, the cost of the subsidy s_j^* equals the net benefit of locating a marginal worker in j, net of the opportunity cost of locating the worker elsewhere, captured by Γ. Locations receive greater subsidies if their spillovers are greater than the average in the economy.

Constant Elasticity Spillovers Simple policy rules obtain in the special case of Example (1). Combining (21) with (11), the optimal subsidies take the form:

$$s_j^* = \frac{\gamma_Z + \gamma_A}{1 - \gamma_A} \left(w_j - \overline{w}\right), \tag{22}$$

where $\overline{w}$ is a population-weighted average of wages. The optimal allocation is implemented by a subsidy that is proportional to labor earnings, with constant proportion $\frac{\gamma_Z+\gamma_A}{1-\gamma_A}$ minus a lump-sum tax $\frac{\gamma_Z+\gamma_A}{1-\gamma_A}\overline{w}$ paid in all locations. To fix ideas, assume negative amenity spillovers ($\gamma_A < 0$), so that higher population creates congestion. So long as the productivity spillover is not too strong ($\gamma_Z < -\gamma_A$), this subsidy scheme implies that richer cities, with above average wage, pay a tax while poorer cities receive a subsidy. The intuition for this result is that the (negative) utility spillover increases with local consumption at a rate γ_A. Because in this example consumption increases one-to-one with wages, net spillovers are proportional to wages, with a negative coefficient. High-wage cities have the most negative spillovers, and should shrink.

Optimal Labor Demand Subsidies The optimal allocation can be implemented both via labor supply subsidies, such as those we characterized ,or labor demand subsidies received by firms. In the context of Example 1, a labor demand subsidy per worker s_j^h means that, for firms, the cost of each worker falls from w_j to $w_j - s_j^h$. Replicating the steps leading up to (22), the optimal hiring subsidy is:

$$s_j^{h*} = \frac{1 - \gamma_A}{1 + \gamma_Z} s_j^*. \tag{23}$$

[9]These conditions are not generically sufficient to guarantee an efficient allocation, as it could be that with such subsidies, multiple equilibria exist.

An optimal hiring subsidy shares the same qualitative features as a direct individual transfer (22), although the overall level of the rate may be different.

Other Distortions to Population Allocation In the model, there are no distortions other than those caused by the spillovers. However, the analysis readily accommodates any distortion that only affects the allocation of population. For instance, if workers face a per-capita tax μ_j in j, perhaps put in place by a local government, the optimal subsidy would obviously be $s_j^* + \mu_j$ for the same s_j^* defined in (21). Another example stems from the particular form of spatial misallocation that arises if one assume local ownership of fixed factors as described above. This type of distortion can be represented by a pseudo-tax $\mu_j = \omega_j (\pi - \pi_j)$ faced by individuals in their location decisions. This pseudo-tax can be again readily added to the current analysis. The optimal policy, in addition to correcting the spillovers, also corrects the distortion caused by local ownership and becomes:

$$s_j^* = \Gamma_j + \omega_j (\pi - \pi_j) - \Gamma. \tag{24}$$

The policy calls for less subsidies (or more taxes), all else equal, in locations where the returns to fixed factors are high. The optimal policy restores the right incentives for location choices, based on real economic returns. Rich locations that are inefficiently large shrink.

2.4.2 Link to Henry George Theorem and Implementation with Land Taxes

A long-standing result in the urban and local public finance literature is the Henry George Theorem (HGT). One way to state this theorem is that, in an economy with spillovers, fully taxing the fixed factor and redistributing the proceeds to local workers exactly allows to reach efficient city size. Put more technically, the value of spillovers at the optimal allocation is equal to the returns to the fixed factor. This result is typically derived for one location in isolation, assuming that a local planner freely chooses the local population to maximize utility per worker of residents, as in the setup of Henderson (1974). Arnott (2004), Abdel-Rahman and Anas (2004), and Wildasin (1986a) show derivations of the theorem in different contexts. The Henry George Theorem can be broadly interpreted as a test for whether an isolated city has optimal population, or as a guide for how to implement this optimal size via taxation and subsidies.

To connect with this result we can ask: can the central planner achieve efficiency by taxing fixed factors and optimally redistributing via gross income subsidies? To answer this question, it proves useful to first consider the aggregate returns to scale in our economy. Imagine we start in an efficient allocation, where (21) holds, and let u^* be associated maximized utility. Then, using (A.5) we obtain:[10]

$$\underbrace{\text{Shadow Value of Labor}}_{\frac{1}{\chi}\frac{du^*}{dL}} = \underbrace{\text{Spillovers}}_{\Gamma} - \underbrace{\text{Return to Land}}_{\pi}. \tag{25}$$

An increase of total population around an efficient equilibrium yields marginal social gains equal

[10]Formally, $\frac{du^*/dL}{\chi}$ is the multiplier of the labor market constraint in the social planner problem, as confirmed in Appendix (25)

to the average spillover Γ but reduces the consumption of the fixed factor for incumbents by π. If our planner could choose L for free (as in the HGT), an interior solution to that problem would correspond to $\frac{du^*}{dL} = 0$, hence to $\Gamma = \pi$ and to spillovers exactly capitalized in fixed factors (HGT).[11]

Does the HGT hold in the framework we presented? Given (21), this means in practice asking whether the returns to fixed factors per capita, π, are at least as big as Γ. In that case, fixed-factor taxes would suffice to implement the optimal gross subsidies Γ_j. Equation (25) shows that $\Gamma = \pi + \frac{1}{\chi}\frac{du}{dL}$, a generalized version of HGT: aggregate spillovers equal returns to fixed factor plus the shadow value of labor (equal to zero in standard HGT with free labor). Fixed factors can finance optimal subsidies if $du^*/dL \leq 0$ (a negative shadow value of labor). Whether this holds depends on parameters. However, cases with positive shadow value can be ruled out by the standard stability requirement that $du_j/dL_j \leq 0$: a marginal local labor increase must not increase local utility per resident. Appendix (2.4.2) shows that assuming stability of the efficient market allocation implies $\pi \geq \Gamma$: taxing all fixed factors would, indeed, generate enough budget implement optimal subsidies (Albouy et al. (2019) show a result in this style).[12]

Furthermore, the same stability argument implies $\Gamma_j \leq \pi_j$, where π_j are local returns to fixed factors, so that *local* taxes on fixed factors at j more than suffice to cover the gross subsidy Γ_j to residents of j that implements efficiency according to (21). Therefore a particular redistribution of fixed factors can implement the spatial allocation. Specifically, from (24), the allocation would be efficient (requiring zero net subsidies, $s_j^* = 0$) given local ownership rates $\omega_j = \frac{\Gamma_j - \Gamma}{\pi_j - \pi}$. Such redistribution from owners of local fixed factors to local residents entails redistribution across space, the landmark of the spatial policies we have discussed, because owners of local factors are located across the economy. This result, however does, not necessarily imply that local governments will choose this policy, a result that would depend on local developers market structure and their objective function.

3 Extensions and Numerical Implementation

Having established the main results and intuitions in a simple model, we now go through a series of extensions corresponding to more realistic assumptions typically made in the QSE literature. Each section considers one deviation from the baseline model. We then turn to implementing the theory quantitatively.

[11] In passing we remind the reader that this result reflects a core insight in urban economics (Henderson, 1974): when developing cities is free (equivalently, when population size can be freely chosen), cities exist in an optimal allocation only if spillovers are net-agglomerative ($\Gamma = \pi > 0$). If the external effects do not exist or are negative ($\Gamma \leq 0$) then the optimal distribution of economic activity is fully dispersed.

[12] Specifically, appendix (2.4.2) finds $sign\left(\frac{du_j}{dL}\right) = sign\left(\Gamma_j - R_j h_j\right)$. That is, local utility grows with local population through local spillovers net of congestion through consumption of fixed factors, with stability ensured when the latter is stronger. If stability holds across locations, adding up across j implies $\Gamma - \pi \leq 0$.

3.1 Trade Frictions and General Production Structure

This section considers more general production structures, where the economy may produce many products, production takes place in several sectors connected by input-output linkages, and trade is frictional. The key result here is that the optimal policy rules we have discussed so far hold in this much more general model, as long as the spillover structure retains the same form. Specifically, amenity and productivity spillovers are local (population in location j only affects productivity or amenities in j), an assumption we relax later. In addition, productivity spillovers must also remain sector-neutral (corresponding to urbanization externalities that are economy-wide, rather than localization externalities that percolate sector-by-sector). As long as these restrictions are satisfied, the exact same optimal policy rule (21) continues to hold. Of course, the spillovers have different general-equilibrium ramifications in a more general model; however, market inefficiencies are still fully tackled through the same policies distorting only location-level labor supply (or demand). This compartmentalization of the inefficiencies reflects a broader "principle of targeting" noted by Bhagwati and Johnson (1960) in trade-policy contexts and by Sandmo (1975) and Dixit (1985) in economies with external effects.

We derive here the optimal policy rule by considering the maximization problem of a social planner aiming to find an allocation that maximizes the common level of welfare in the economy, an alternative to the perturbation approach considered earlier. Workers have concave utility $u_j\left(\left\{c_j^k\right\}, L_j\right)$ over many private goods k with consumption per capita c_j^k. Production of k combines efficiency units of labor N_j^k, unchanged compared to the basic specification, and possibly other industries' l output X_j^{kl} as intermediate. The production function is $Y_j^k\left(N_j^k, \left\{X_j^{kl}\right\}_l\right)$. Finally, trade between locations is costly, with iceberg trade costs: $\left(1+t_{ji}^k\right) Q_{ji}^k$ units of good k from j must be shipped to i for Q_{ji}^k units to arrive. The index k may indicate a sector or a variety produced in a given origin. Hence, this specification nests the typical Armington formulation used in quantitative spatial models at the regional level, where products are differentiated by origin (see Redding and Rossi-Hansberg (2017) and Allen and Arkolakis (in this handbook)). The formulation also nests cases where the same good may be produced in different locations, such as in Sotelo (2020) and Bergquist et al. (2022) analyses of spatially differentiated agricultural markets. Finally, it also allows for gross flows of a homogeneous commodity to be purely in transit in a given location, a case that forms the basis to study optimal transport networks in Section 5.

The optimal allocation maximizes u in the feasible set. The planner's Lagrangian is thus:

$$\begin{aligned}\mathcal{L} =& u - \sum_j \omega_j \left[u - u_j\left(\left\{c_j^k\right\}, L_j\right)\right] - \sum_j W_j \left[N_j - z_j\left(L_j\right) L_j\right] - \Omega\left[\sum_j L_j - L\right] \\ & - \sum_j \sum_k p_j^k \left[c_j^k L_j + \sum_l X_j^{lk} + \sum_i \left(1+t_{ji}^k\right) Q_{ji}^k - Y_j^k\left(N_j^k, \left\{X_j^{kl}\right\}_l\right) - \sum_i Q_{ij}^k\right]. \qquad (26)\end{aligned}$$

The first-order conditions of this problem constitute necessary conditions for a maximum. If, in addition, the objective function is concave and the constraint set forms a convex region, then they are also sufficient for optimality. In this case, finding a set of subsidies that guarantees that these first order condition hold in the market allocation leads unequivocally to efficiency. In the parameterized version of this model described below, Fajgelbaum and Gaubert (2020) derive sufficient conditions for the planner's problem to be concave. Quite intuitively, a sufficient (but strong) condition is that spillovers are negative on net so that congestion dominates.

We wrote the problem such that, if the problem is quasi-concave and first-order conditions are sufficient, the multipliers on inequality constraints are positive. Each multiplier corresponds to a constraint in the feasible set, generalizing (3)-(7). The first constraint is the spatial mobility constraint while the last is the national availability of workers, both equality constraints. The other are inequality constraints: p_j^k is the multiplier on the constraint that consumption of sector-k goods at j is limited by local goods availability, including imports and exports, while W_j captures the local availability of efficiency units for production. In an efficient market allocation, the price of product k in region j equals the multiplier p_j^k and the wage rate per efficiency unit of labor will equal W_j.

The optimal policy is obtained by taking first-order conditions with respect to just consumption per capita in each sector-location, and total employment by location. These two conditions are:

$$\left[\partial c_j^k\right] \qquad \omega_j \frac{\partial u_j}{\partial c_j^k} = p_j^k L_j \tag{27}$$

$$\left[\partial L_j\right] \qquad \omega_j \frac{\partial u_j}{\partial L_j} - \sum_k p_j^k c_j^k + w_j \left(1 + \varepsilon_{z_j, L_j}\right) = \Omega, \tag{28}$$

where the second condition has used that, in a market allocation, wage per worker $w_j = W_j z_j$. Using (27) and (28) along with first-order conditions of the consumer problem in the market allocation, we obtain that, if the market allocation is efficient, the subsidies satisfy:[13]

$$s_j = \varepsilon_{z_j, L_j} w_j + \frac{\varepsilon_{v_j, L_j}}{\varepsilon_{v_j, x_j}} x_j - (\pi + \Omega). \tag{29}$$

The optimal policy is the same as (21). We obtain again in particular, as discussed in the context of the Henry George Theorem, that $\Gamma = \pi + \Omega$ where Ω is the shadow value of labor in the aggregate. Since the derivations only used first-order conditions over c_j^k and L_j, it means that the supply-side assumptions we made determining production, input-output structure, and trade are irrelevant for the optimal policy rule.

3.2 Public Goods

Urban agglomeration of consumer-workers is often driven by the desire for access to public goods that they collectively finance. The model so far treated amenities as a reduced-form object that potentially increases with local population. In reality, public goods such as schools, parks, and transportation systems are amenities that are typically provided by governments and must be

[13]Indirect utility is $v_j\left(x_j, \left\{p_j^k\right\}, L_j\right) \equiv \max_{c_j^k : \sum_k p_j^k c_j^k = x_j} u_j\left(\left\{c_j^k\right\}, L_j\right)$. The first order condition is $\frac{\partial u_j}{\partial c_j^k} = \lambda p_j^k$, where the multiplier λ of the budget constraint equals $\frac{\partial v_j}{\partial x_j}$, so $\frac{\partial u_j}{\partial c_j^k} = \frac{\partial v_j}{\partial x_j} p_j^k$. Combining this condition with (27), and further using (28) and the facts that $\frac{\partial u_j}{\partial L_j} = \frac{\partial v_j}{\partial L_j}$ and $\sum_k p_j^k c_j^k = x_j = w_j + \pi + s_j$ gives (29).

funded. A natural extension to the analysis conducted so far is thus to assume that residents enjoy a pure public good g_j provided by local or central governments, so that utility is $u_j(c_j, h_j, g_j, L_j)$. We focus here on a simple exposition of this case, and refer the reader to the handbook chapter by Wildasin (1987) that discusses many subtleties that arise with the financing of public and semi-public goods. We note that it is also straightforward to extend the reasoning to a productive public good that increases firms' productivity rather than locational amenities.

We assume that the public good is produced one-for-one with the traded good and that it is non-rival and non-excludable, so that its cost is simply g_j. A feasible allocation $\{c_j, h_j, g_j, L_j, N_j, u\}$ now satisfies the following constraints:

$$u_j(c_j, h_j, g_j, L_j) = u \text{ if } L_j > 0 \text{ (and } \leq u \text{ otherwise)}, \tag{30}$$

$$\sum_j c_j L_j + g_j \leq \sum_j Y_j(N_j), \tag{31}$$

with the feasibility constraints (5)-(7) unchanged compared to the baseline, including potentially spillover effects in utility and production. Two margins are potentially distorted: population distribution, like before, and public good provision g_j. A central planner would choose to provide the local public good following the Samuelsonian condition for optimal public expenditure. A local planner maximizing utility per resident without any strategic consideration would also do so.[14] We focus here again on spatial policies that tackle the optimal distribution of population.

Solving the planner's problem as in the baseline model leads to the same optimality condition (18) as before. Assuming for expositional simplicity that there are no amenity or productivity spillovers of the kind discussed earlier ($\Gamma_j = 0$), the optimality condition is

$$w_j - x_j = \Omega, \ \forall j. \tag{32}$$

Whether the allocation without central planner intervention satisfies (32) depends on how the local public goods are financed. Two polar cases are interesting to discuss. First, if local public goods are financed by taxing the local fixed factors, then the budget constraint of local residents is: $x_j = w_j + \pi$, where π is now defined after taxes. At any rate, it satisfies (32). Second, if the local public good is financed by a local per-capita tax on workers, the budget constraint of local residents is: $x_j = w_j - \frac{g_j}{L_j} + \pi$. This allocation is distorted and a central government would subsidize location j to exactly offset this distortion, setting $s_j = \frac{g_j}{L_j} + s_0$ (with s_0 to ensures the central government budget clearing) to offset the distortive local tax. As a result, optimal funding of the public goods is done with a constant lump-sum tax s_0 across space (Helpman and Pines, 1980; Wildasin, 1980). The intuition for these results is that, with a pure public good and a common constant marginal cost of supplying it across space, a marginal worker in a location is not costly. However, variation

[14] In the efficient allocation, the planner chooses the levels of local public goods such that: $\left(\frac{\partial u_j}{\partial c_j}\right)^{-1} \frac{\partial u_j}{\partial g_j} L_j = 1$, the standard Samuelsonian condition for optimal public expenditure. The marginal cost to provide public service g is 1 and equals the sum of marginal rate of substitution across individuals in the location. This condition obtains if a local government maximizes the welfare of its residents taking its number as well as market prices as given; i.e., solving for $g_j = \arg\max_\mu v_j\left(w_j + \pi - \frac{g_j}{L_j}, g_j, L_j\right)$, where g_j is financed by local residents (the tax is $\frac{g_j}{L_j}$). A rich literature studies how local provision of the public good may play out in less stylized environments (Helsley, 2004; Glaeser, 2013).

in per-capita tax rates affects relative appeal across locations, creating a distortion. A tax on local fixed factors, in contrast, does not distort location incentives.

3.3 Heterogeneous Workers

We have so far assumed homogeneous workers. However, different demographic groups make different location choices, resulting in spatial sorting. Moreover, efficiency and amenity spillovers may vary across groups, in terms of both the spillovers each group experiences and generates on the rest of the economy (Diamond, 2016). We introduce these forces and discuss policies that implement efficient spatial sorting.[15]

The population is partitioned into groups $\theta \in \Theta$. Group membership is assumed to be exogenous. These groups could for instance be defined by demographic characteristics like levels of education, age, or ethnicity. On the preference side, different groups may have different utility functions $u_j^\theta\left(c_j^\theta, h_j^\theta, \left\{L_j^{\theta'}\right\}_{\theta'}\right)$. The utility of type θ in location j may depend on the population of all groups in that location, as captured by the vector $\left\{L_j^{\theta'}\right\}_{\theta' \in \Theta}$. On the production side, workers of different types may be imperfect substitutes in production of the traded good, as captured by the aggregator of efficiency units

$$N_j = N\left(z_j^1 L_j^1, .., z_j^\Theta L_j^\Theta\right). \tag{33}$$

The labor productivity of type θ in location j is $z_j^\theta\left(\left\{L_j^{\theta'}\right\}_{\theta' \in \Theta}\right)$, which may again be a function of the population of all groups in the location. Finally, government subsidies s_j^θ as well as the claim to returns to the fixed factors π^θ may be group-specific.

These specifications are general enough to encompass various sources of heterogeneity that drive spatial sorting emphasized in the literature (Diamond and Gaubert, 2022). Sorting may be driven by comparative advantage in production, as some skills may be particularly productive in a given location. The terms z_j^θ capture this force, which may be exogenous - some skills are more productive in some sectors that happen to be in some locations - and endogenous: a given skill group may be subject to agglomeration effects in production. Sorting may also be shaped by different preferences for locations. Again, this could be for reasons that are exogenous to sorting (e.g., retirees like to live in warmer weather) or for reasons that are endogenous to sorting. For instance, local amenities may depend on who lives in the location, and be in turn valued differently by different groups. Finally, sorting may be shaped by cost-of-living differences between locations if they impact more resource-constrained groups more. Non-homothetic preferences u allow for this force.

Although the discussion in this section is not about redistribution, we allow for the planner to have heterogeneous Pareto weights λ_θ on the different groups. As the groups have different utility functions, it is unclear how to compare utils of a given group to utils of another group. Tracing out the whole Pareto frontier with arbitrary weights λ_θ sidesteps this issue. The planner chooses allocations, and in particular population and consumption $\{L_j^\theta, c_j^\theta\}$, to maximize $\sum_{\theta \in \Omega} \lambda^\theta u^\theta$. It is

[15] Our presentation differs from the optimal sorting model of Helsley and Strange (2014) who consider sector-specific agglomeration effects in a model where heterogeneous sectors hire sector-specific workers.

again the case here that population and consumption distribution of a given group are one to one, holding the distribution of other types fixed. Therefore, the planner can implement the efficient population allocation by subsidizing each group so that they reach their optimal distribution of consumption. Solving for the first-order condition of the planner problem yields the following optimal subsidy for group θ in location j:

$$s_j^\theta = \sum_{\theta' \in \Theta} \frac{\varepsilon_{v_j^\theta, L_j^{\theta'}}}{\varepsilon_{v_j^{\theta'}, x_j^{\theta'}}} \frac{x_j^{\theta'} L_j^{\theta'}}{L_j^\theta} + \sum_{\theta' \in \Theta} \varepsilon_{z_j^\theta, L_j^{\theta'}} \frac{w_j^{\theta'} L_j^{\theta'}}{L_j^\theta} - \left(\pi^\theta + \Omega^\theta\right). \tag{34}$$

This optimal subsidy as a similar structure to (21). Optimal subsidies differ across groups because they generate different spillovers. A type θ is subsidized in j when she generates high amenity spillovers on to other inhabitants of the location, the first term in (34), and when she generates high productivity spillovers on to other workers in j, the second term in (34). Finally, the term Ω^θ, which is constant across space for each group, measures how much the planner would be willing to pay for adding a θ-type worker to the economy. It shifts group-specific subsidies in all locations. This group-wide term tackles the pure redistributive concerns of the planner between types. Through this term, subsidies increase monotonically for a group θ with the planner's relative preference for this group, λ^θ. With heterogenous workers, the optimal policy (34) assumes that the planner has access to tax instruments that vary by place and by group.[16]

To illustrate the forces at play more concretely we conclude with a simple example.

Example 2. Consider two skill groups, Low and High. Utility of type $\theta \in \{L, H\}$) is $u_j^\theta = A_j^\theta c_j^\theta$, so that there are no spillovers in amenities. The labor productivity of type θ is $Z_j^\theta \left(L_j^{\theta'}\right)^{\gamma_Z^{\theta'\theta}}$ where $\theta' \neq \theta$. That is, productivity spillovers happen across skill groups, and only across, with an elasticity that is constant across space. Finally, there are constant returns to scale in production so that $\pi = 0$.

In this example, the location-specific component of the optimal subsidy (34) is simply:

$$s_j^\theta = \gamma_Z^{\theta\theta'} \left(\frac{w_j^{\theta'} L_j^{\theta'}}{w_j^\theta L_j^\theta} \right). \tag{35}$$

The subsidy for $\theta-$type workers varies across locations according to the distribution of relative wage bills $\frac{w_j^{\theta'} L_j^{\theta'}}{w_j^\theta L_j^\theta}$. If group θ generates positive productivity spillovers on group θ', i.e. $\gamma_Z^{\theta\theta'} > 0$, then group θ receives higher subsidies in locations where it is relatively scarce. By the same token, if both groups generate positive spillovers on the other group, their subsidies are negatively correlated across cities. The scarcer the group in relative term, the higher the subsidy. Relative to a *laissez-faire* equilibrium, this policy therefore tempers the degree of spatial sorting. More mixing of groups is desirable as it increases productivity through co-agglomeration effects.

[16] More restricted tax instruments will not typically allow to reach the first best allocation, but may still improve welfare.

3.4 Cross-Regions Spillovers

Considering that agglomeration and congestion effects happen within a given location may seem reasonable when considering large regions like states or MSAs. This is therefore a standard assumption, consistent with classic approaches including systems-of-cities models (Henderson, 1974), new economic geography (Krugman, 1991; Helpman, 1998), and quantitative spatial models (Allen and Arkolakis, 2014). However, when studying economic activity at a finer geographic scale, it may be appropriate to consider that the economic conditions in one area spillover to nearby areas. Regional models such as Rossi-Hansberg (2005) and urban models as in Ahlfeldt et al. (2015) assume spatially decaying spillovers. This may be incorporated in the baseline framework by assuming that spillovers in a location depend on the distribution of population in all locations: $u_j\left(c_j, h_j, \{L_k\}_k\right)$ and $z_j\left(\{L_k\}_k\right)$. The specific functional form used in Ahlfeldt et al. (2015) is $\left(\sum_j L_j e^{-\delta t_{jj'}}\right)^{\alpha}$ where $t_{jj'}$ is travel time between blocks j and j' within a city and δ is a decay parameter.

The derivations of the baseline model go through with this more general specification of spillovers, and yield the following optimal subsidy in location j:

$$s_j = \sum_{k=1}^{J} \frac{\varepsilon_{v_k,L_j}}{\varepsilon_{v_k,x_k}} \frac{L_k x_k}{L_j} + \sum_{k=1}^{J} \varepsilon_{z_k,L_j} \frac{w_k L_k}{L_j} - (\pi + \Omega)\,. \tag{36}$$

Non-localized spillovers lead to the intuitive implication that the optimal transfers should be higher in places that generate strong spillovers to more and larger locations, as measured by their total wage bill and total income.

3.5 Commuting and Congestion Tax

A strand of the literature studies economic geography models where individuals may live and work in different locations. Abstracting from that force makes sense when considering large regions, but it is a relevant feature when studying policies targeted at neighborhoods, like the Enterprise Zone programs in the U.K. or the U.S. for instance.

To account for commuting, the structure of the model changes. Households freely choose where to live and where to work. L_{jk} workers chose to commute from residence j to workplace k. More generally, all choice variables can depend on both residence and workplace and are indexed by jk. Worker face a disutility of commuting. Denoting δ_{jk} the time spent commuting between j and k, $\frac{\partial u_{jk}}{\partial \delta_{jk}} < 0$. The economy is potentially subject to three types of spatial spillovers. First, productivity in workplace k is a function of local employment through classic agglomeration effects in production, $z_j\left(\sum_j L_{jk}\right)$. Second, utility depends on local population $L_j = \sum_k L_{jk}$ as it may shape local amenities. Finally, time spent commuting can depend on "traffic" on the route. To build intuition and introduce the topic, we assume that traffic congestion for travelers from j to k depends on the numbers of those travelers only: $\tau_{jk}\left(L_{jk}\right)$, with $\tau'_{jk}\left(L_{jk}\right) \geq 0$.[17] Here, some form of congestion on a route is needed to rationalize the fact that, in the data, residents of a given location

[17]Of course in reality congestion on the route between j and k depends not only on commuters who directly commute from j to but also on all other commuters that use part or all of this route in their commute. This feature can be incorporated using the optimal transport tools described in Section 5.

typically commute to different workplaces that offer different wages. This is consistent with free mobility and utility equalization across all choices if τ_{jk} adjusts endogenously in equilibrium.[18]

The planner now maximizes the common level of utility u, equalized across all residence-workplace pairs that are chosen ($u = u_{jk}\left(c_{jk}, h_{jk}, \sum_k L_{jk}, \delta_{jk}\right)$ whenever $L_{jk} > 0$). The logic remains similar to the baseline case: it can be seen from the set of feasible allocations that each vector of L_{jk} corresponds to only one distribution of consumption x_{jk}. Therefore, the planner chooses subsidies s_{jk} that depend on both residence and workplace to implement an efficient allocation of population. Given the condition for efficient population distribution, the subsidy that implements it is:

$$s_{jk} = \underbrace{\sum_l \frac{\varepsilon_{v_{jl},L_j}}{\varepsilon_{v_{jl},x_{jl}}} \frac{L_{jl}x_{jl}}{L_j}}_{\text{residence}-j\text{ specific}} + \underbrace{w_k \varepsilon_{z_k,L_k}}_{\text{workplace- }k\text{ specific}} + \frac{\varepsilon_{v_{jk},\delta_{jk}}\varepsilon_{\delta_{jk},L_{jk}}}{\varepsilon_{v_{jk},x_{jk}}} x_{jk} - (\Omega + \pi) \tag{37}$$

The optimal subsidy to a jk worker decomposes into three terms. The first two terms are common with amenity and efficiency spillovers that were present in the framework we have developed before. Specifically, the first term tackles the amenity spillover (increasing jk commutes raises the number of residents in j irrespective of their workplace l) and the second term tackles the productivity spillover (subsidizing jk commutes raises employment, hence agglomeration effects, in k). The third term is a congestion tax: adjusting the subsidy to congestion and the origin-destination pair commute pair level. This term is negative and tackles the congestion caused by additional commuters on a route. In popular specifications with constant elasticities, the congestion tax is space-invariant: the overall optimal subsidy simplifies into the sum of a residence-specific and a workplace-specific terms only, as the next example, which augments Example 1 with commuting, shows.

Example 3. Utility is $u_{jk} = \frac{c_{jk}}{\delta_{jk}} (L_j)^{\gamma_A}$, where $\delta_{jk} = \delta_{jk}^0 L_{jk}^{\gamma_T}$. Labor productivity in k is $z_k \left(\sum_j L_{jk}\right)^{\gamma_Z}$. Production has constant returns to scale so that $\pi = 0$.

In this example, the optimal subsidy (37) becomes:

$$s_{jk} = \sum_l \frac{\gamma_A}{1+\gamma_T} \frac{L_{jl}x_{jl}}{L_j} + w_k \left(\frac{\gamma_Z - \gamma_T}{1+\gamma_T}\right) - \frac{\Omega}{1+\gamma_T}. \tag{38}$$

Since the congestion elasticity γ_T does not depend on the route, neither does the congestion-tax component of the overall subsidy, and the subsidy can be recast as the sum of a residence and a workplace-specific terms.

[18] The QSE literature typically assumes no congestion in commuting costs. Instead, in order to generate heterogeneous choices of workplace k for residents of j, it assumes that workers have heterogeneous idiosyncratic preferences for commuting routes (see e.g. Monte et al., 2018). This turns out to lead to isomorphic expressions. However, the interpretation of the normative properties of two models differ. One (presented here) has spillover in congestion, the other does not. One (presented here) has homogeneous workers, the other does not. Therefore, optimal subsidies correct an inefficiency here, but are redistributive in the version with heterogeneous workers and no congestion. Section 4 discusses this theme and the normative implications of models with idiosyncratic preferences more generally.

3.6 Quantitative Implementation

This section moves on to show how to take these results to the data, quantify efficient policies and compute the corresponding allocation. To do it in practice requires further parameterization of the model.

3.6.1 Steps

The optimal allocation in the models we have discussed above corresponds to an equilibrium X^* (for example, $= \left\{c_j^*, h_j^*, N_j^*, L_j^*, w_j^*, u^*\right\}$ in the model of section (2)) that satisfies the definition of a competitive allocation with optimal transfers (i.e. transfers such that equation (21) is satisfied in the example). Implementing the optimal allocation requires knowledge of the model's fundamentals (shifters of production or utility functions, bilateral trade costs) and a parametrization of elasticities (production function and utility parameters and spillover elasticities). We describe here an approach to doing this, in two steps.

The first step ("hat algebra") is standard in quantitative spatial models (as reviewed in other chapters of this handbook). This step derives a system of equations that describes how the allocation changes in response to a policy shock, rather than describing the new equilibrium. The variables of interest are relative changes $\hat{X} = X'/X$ of counterfactual outcomes X' relative to the initial equilibrium X. These changes $\hat{X}$ are expressed as a function of data D (i.e., the initial equilibrium), elasticities Θ, a vector of initial spatial subsidies s, and a vector of shocks Δs to these spatial subsidies: $\hat{X}\left(D, \Theta, s, \Delta s\right)$. This step is convenient because it implicitly calibrates the unobserved fundamentals of the model to exactly match the observed distribution D of economic activity. In this process, the procedure identifies D summary statistics that, together with initial spatial subsidies s, suffice to answer the counterfactual question given the shocks and the elasticities. This data requirement is typically a subset of the variables X of the model.[19]

The second step (optimization) finds the optimal policy shock Δs^*. One way to proceed is to numerically optimize over the distribution of feasible values of Δs to maximize the change in utility per worker compared to an observed allocation:

$$\Delta s^* = \arg\max_{\Delta s} \hat{u} \text{ s.t. } \hat{X} = \hat{X}\left(D, \Theta, s, \Delta s\right) \tag{39}$$

where the $\hat{X}\left(D, \Theta, s, \Delta s\right)$ are the changes in endogenous variables (including $\hat{u}$) given by feasible changes in subsidies Δs. An alternative procedure is to use the closed-form solution that characterizes optimal policies (such as (21) for the model of section (2)). This closed-form solution is an optimal policy expressed as a (subset of) the elasticities Θ and the variables in X *evaluated at the optimum* X^*: $s^*\left(\Theta, X^*\right)$. By construction, the optimal subsidies s^* and optimal reallocations $\hat{X}^*$

[19] This type of derivations was first introduced by Dekle et al. (2008) in the context of international trade models and is often applied in quantitative spatial models, and they apply more generally than to the model described here (see Adao et al. (2017) for a discussion of generality of these models in trade setups). In spatial setups, Dingel and Tintelnot (2020) discuss limitations and alternatives of this approach in high-dimensional models, such as those with commuting we discuss below.

solve:

$$\hat{X}^* = \hat{X}(D, \Theta, s, \Delta s^*), \tag{40}$$

$$\Delta s^* = s^*\left(\Theta, X\hat{X}^*\right) - s. \tag{41}$$

The first line implements a counterfactual using the optimal policy shocks. The second line evaluates the optimal policy rule at those counterfactual values, $X^* \equiv X\hat{X}^*$.[20]

Regardless of how this second step is implemented, the upshot is that the optimal reallocations $\hat{X}^*$, and therefore the optimal allocation and transfers, are ultimately expressed as a function of model elasticities Θ and of the cross-section of economic activity D, which is exactly rationalized by the model (given elasticities). Fajgelbaum and Gaubert (2020) numerically implement the optimization using (39) and then use (41) to verify the solution.

The obvious practical concern is whether the optimization problem in (39) is well-behaved (admitting a unique interior solution) or, equivalently, whether the solution to the system (40)-(41) is interior and unique. Showing concavity of (39) ensures uniqueness of (40)-(41).[21] Fajgelbaum and Gaubert (2020) provide sufficient conditions for concavity, requiring congestion forces (concavity in production of tradable and non-tradables) to be stronger than agglomeration forces entering through $z_j(L_j)$ and $u(\cdot, L_j)$. We apply their result in a special case below.

3.6.2 Data and Measurement Issues

Implementing these steps requires data on both D and s. The content of D depends on model specification (see Allen and Arkolakis (2025) for details in standard spatial frameworks). Measuring the spatial subsidies s is a more specific issue that arises in our context. A challenge in doing so is that spatial policies take many different forms and are managed by different entities targeting different type of recipients (e.g. firms, workers, local governments). A "bottom-up" approach to measuring s would be to compile all existing subsidies. Hanson, Rodrick and Sandhu (2025) compile information on a wide range of place-based programs that could be useful to pursue this approach.

An alternative "top-down" approach, is to start from the condition $s_j = x_j - w_j - \pi$, and measure (x_j, w_j, π) to infer s_j. Fajgelbaum and Gaubert (2020) follow this approach and use census and BLS data data to measure labor income w_j, expenditure x_j and capital income π. This approach has several caveats. First, income and expenditures may be different across space because people are systematically different. Also, different people receive different transfers that might not be place-based. To partially assuage these concerns, Fajgelbaum and Gaubert (2020) use measures of (x_j, w_j) that are net of variation in socio-demographic characteristics across space. A second issue is rooted in the static nature of the model: in a static model, the difference between income and expenditure corresponds to government transfers, but in reality, savings also shape the difference

[20] This approach has been used in quantifications of optimal subsidies and trade policies in international trade frameworks, e.g. see Lashkaripour and Lugovskyy (2023).

[21] A related concern is whether the solution to the hat-algebra system $\hat{X}(d, \Theta, s, \hat{s})$ from the first step is uniquely defined given $\hat{s}$. Allen and Arkolakis (in this handbook) present a detailed discussion of uniqueness in spatial models given shocks to fundamentals.

between income and expenditure. Conditioning on socio-demographic characteristics, differences in savings in space will confound the measurement of transfers.

3.6.3 Implementation in a Simple Case

We illustrate the implementation based on the model of Section 2, which we now parameterize using constant-elasticity functions. Utility is Cobb-Douglas in tradables and housing and multiplicative in the amenity spillover:

$$U_j\left(c_j, h_j, L_j\right) = A_j c^{\alpha_C} h^{1-\alpha_C} L_j^{\gamma_A}. \tag{42}$$

Tradable output is linear in efficiency units:

$$Y_j\left(N_j\right) = z_j^Y N_j, \tag{43}$$

and efficiency units per worker are a power function of workers,

$$z_j\left(L_j\right) = Z_j L_j^{\gamma_Z}. \tag{44}$$

Hat-Algebra Step We first construct the system of equations $\hat{X}\left(D, \Theta, s, \hat{s}\right)$. Conditions (45)-(47) below define the equilibrium changes $\hat{X} = \left\{\hat{L}_j, \hat{x}_j, \hat{u}, \hat{\pi}\right\}$ as function of data $D = \{w_j, L_j, x_j\}$, elasticities $\Theta = \{\alpha_C, \gamma_A, \gamma_Z\}$, initial policies s_j, and policy shocks $\hat{s}_j$. The spatial equilibrium condition (14) yields:

$$\hat{u} = \hat{x}_j^{\alpha_C} \hat{L}_j^{-1-\alpha_C-\gamma_A}, \tag{45}$$

The returns to land $\hat{\pi}$ must be consistent with goods market clearing (4),

$$\sum_j s_j^X \hat{x}_j \hat{L}_j = \sum_j s_j^W \underbrace{\hat{L}_j^{1+\gamma_Z}}_{\hat{w}_j \hat{L}_j}, \tag{46}$$

where s_j^X and s_j^W are shares of j in aggregate expenditures and wages (or tradable output), respectively. Finally, $\hat{u}$ must be consistent with labor market clearing (7),

$$\sum_j s_j^L \hat{L}_j = 1, \tag{47}$$

where s_j^L is region j's labor share. The previous system is written as function of arbitrary shocks to expenditure per capita $\hat{x}_j$. In turn, using (11), expenditures change according to:

$$x_j \hat{x}_j = \underbrace{\hat{L}_j^{\gamma_Z}}_{\hat{w}_j} + \pi\hat{\pi} + s_j \hat{s}_j. \tag{48}$$

We note that the changes in expenditure per capita $\hat{x}_j$ can be treated as an exogenous shock to the system, as there is always a set of $\hat{s}_j$ that can rationalize any desired $\hat{x}_j$. We exploit this property in the optimization step.

Optimization Step One alternative is to solve (39), i.e. numerically solve for $\left\{\hat{u}, \hat{x}_j, \hat{L}_j\right\}$ that maximize $\hat{u}$ subject to (45) to (47). We would maximize over $2J+1$ variables subject to $2J+1$

constraints. The optimal transfers $\hat{s}_j$ can be obtained residually from the definition of expenditures using (48).

Another approach is to exploit the solution for optimal policy (21). Expressed in relative changes, the optimal policy rule is:

$$\hat{x_j} = \frac{1+\gamma_Z}{1-\gamma_A}\frac{s_j^W}{s_j^X}\hat{L}_j^{\gamma_Z} + \frac{Y}{L}\left(\frac{1}{\alpha_C} - \frac{1+\gamma_Z}{1-\gamma_A}\right)\sum_j s_j^W \hat{L}_j^{1+\gamma_Z}, \tag{49}$$

where to obtain this condition we combined (21) with (48) and (46) to solve for $\pi\hat{\pi}$. Furthermore, combining (45) with (47), utility change is a function of the distribution of expenditure changes:

$$\hat{u} = \left(\sum_j s_j^L \hat{x}_j^{\frac{\alpha_C}{1-\alpha_C-\gamma_A}}\right)^{1-\alpha_C-\gamma_A} \tag{50}$$

The optimal allocation is the $\left\{\hat{u}, \hat{L_j}, \hat{x_j}\right\}$ that jointly solves (45), (49) and (50) (a system with $2J+1$ equations with same number of unknowns). Again, optimal changes in subsidies can be recovered using (48). Applying results from Fajgelbaum and Gaubert (2020), the planners' problem is convex if $\gamma_A + (1+\gamma_Z)\,\alpha_C < 1$. This sufficient condition means that agglomeration spillovers must be not too strong, and that higher consumption of housing (a lower α_C) makes concavity more likely. In that case, the system that solves (45), (50), and (49) has a unique solution. As a result, we obtain $\left\{\hat{u}^*, \hat{L_j^*}, \hat{x_j}^*\right\}$ as a function of data $d = \{w_j, L_j, x_j, \pi\}$ and elasticities $\Theta = \{\alpha_C, \gamma_A, \gamma_Z\}$.

3.6.4 Numerical Examples

We now conduct a quantification exercise based on different model specifications. A first specification corresponds to the model we have just solved for in section 3.6.3. This is the baseline model of Section 2 with homogeneous workers, local spillovers, and production of a freely traded good (specialized here to constant returns to scale); we refer to this case as model 1. The second specification augments that model with two groups of workers who generate and benefit differentially from local spillovers as described in section 3.3. The supply of efficiency units in (33) combines the two worker types using a CES production function (model 2). Finally, the third specification augments model 2 with trade frictions and product differentiation as in 3.1; specifically, each region produces a differentiated good that is traded across space subject to trade frictions as in standard quantitative spatial models (model 3). We show that moving from the baseline model to a model with extra layers that make it more realistic impacts the optimal policies in significant ways.

All three examples are based on U.S. MSAs. The models are calibrated following similar similar steps and data sources as Fajgelbaum and Gaubert (2020). In particular, the calibration of spillover elasticities combines estimates from Ciccone and Hall (1996) and Diamond (2016). Model 1 with homogeneous workers employs the constant spillover elasticities $\varepsilon_{z_j,L_j} = 0.06$ and $\frac{\varepsilon_{v_j,L_j}}{\varepsilon_{v_j,x_j}} = -0.19$, which are net congestive. Model 2 and 3 have two types of workers (high- and low-skill corresponding to workers with and without a college degree, respectively) with heterogeneous spillovers. The calibrated matrices of spillovers are: $\left(\gamma_{LL}^P, \gamma_{HL}^P, \gamma_{LH}^P, \gamma_{HH}^P\right) = (.003,\ .044,\ .02,\ .053)$

and $\left(\gamma_{LL}^{A}, \gamma_{HL}^{A}, \gamma_{LH}^{A}, \gamma_{HH}^{A}\right) = (-.43,\ .18,\ -1.24,\ .77)$. These parameters imply strong positive amenity spillovers generated by high skill workers and negative spillovers generated by low skill workers. The single-worker and heterogenous-workers spillovers are consistent, in the sense that a a properly weighted average of the heterogeneous elasticities yield the homogeneous-worker elasticities.[22]

First-Order Gains from Increasing Subsidies We start by implementing the exercise in section 2.3 that ranks cities in the observed equilibrium according to the model-predicted returns to increasing their spatial subsidy. Figure 1 plots Γ_j, s_j, and $\Gamma_j - s_j$ in the cross-section of cities for the homogenous group of model 1 (panel (a))and separately for the low-skill workers (panel (b)) and high-skill workers (panel (c)) in model 2.[23] According to this metric, the cities that would yield higher returns to a marginal increase in spatial subsidies (higher $\Gamma_j - s_j$) tend to be high-wage cities in model 1. This result may seem surprising given that externalities are net congestive, so that in a *laissez-faire* equilibrium high-wage cities generate more congestion. Indeed, in panel (a) of Figure 1, the spillovers Γ_i do decrease with city wage. However, the spatial transfers s_i in the observed equilibrium also decrease with city wage (in the direction the theory suggests is optimal) and they do so faster than the spillovers. Thus, the equilibrium is distorted in the opposite way than what externalities alone would predict: around the observed equilibrium, reducing taxes in high-wage cities increases welfare.

Moving on to the two groups case, we first notice a different scale in the axes: the calibrated spillovers, in particular those coming from amenities, are very large in absolute magnitude for some groups. The spillovers generated by low skill workers are more negative in higher wage cities, as in model 1, but this spillover is much stronger than in the baseline model, with existing subsidies relatively smaller in low-wage areas. Therefore, the next dollar spent on spatially subsidizing low skill workers should go to lower-wage cities. In contrast, high-skill workers generate positive spillovers on both types of workers. These spillovers are higher in higher wage cities, while existing subsidies redistribute away from these cities. Therefore, the next dollar spent on spatially subsidizing high skill workers should go to higher wage cities.

Optimal Subsidies Next, we solve for the optimal system of spatial transfers in these economies. Figure 2 reports s_j^*/w_j (net optimal subsidies relative to current wage) across MSAs, again as a function of the wage in the observed allocation. In model 1, the optimal spatial transfers redistribute towards lower-wage cities. This result comes from externalities being net dispersive in this calibration. However, the actual allocation features stronger redistribution than what these parameters imply, similar to what we observed in the previous local exercise. As a result, larger high-wage cities are being taxed too much from the perspective of efficiency under these parameters. The wel-

[22]See appendixes B.2 and D.1 of Fajgelbaum and Gaubert (2020).

[23]Model 3 is identical to model 2 in terms of their observed population, wages, and expenditure, and therefore in terms of first-order effect of a shock represented here. These models differ however when considering larger shocks.

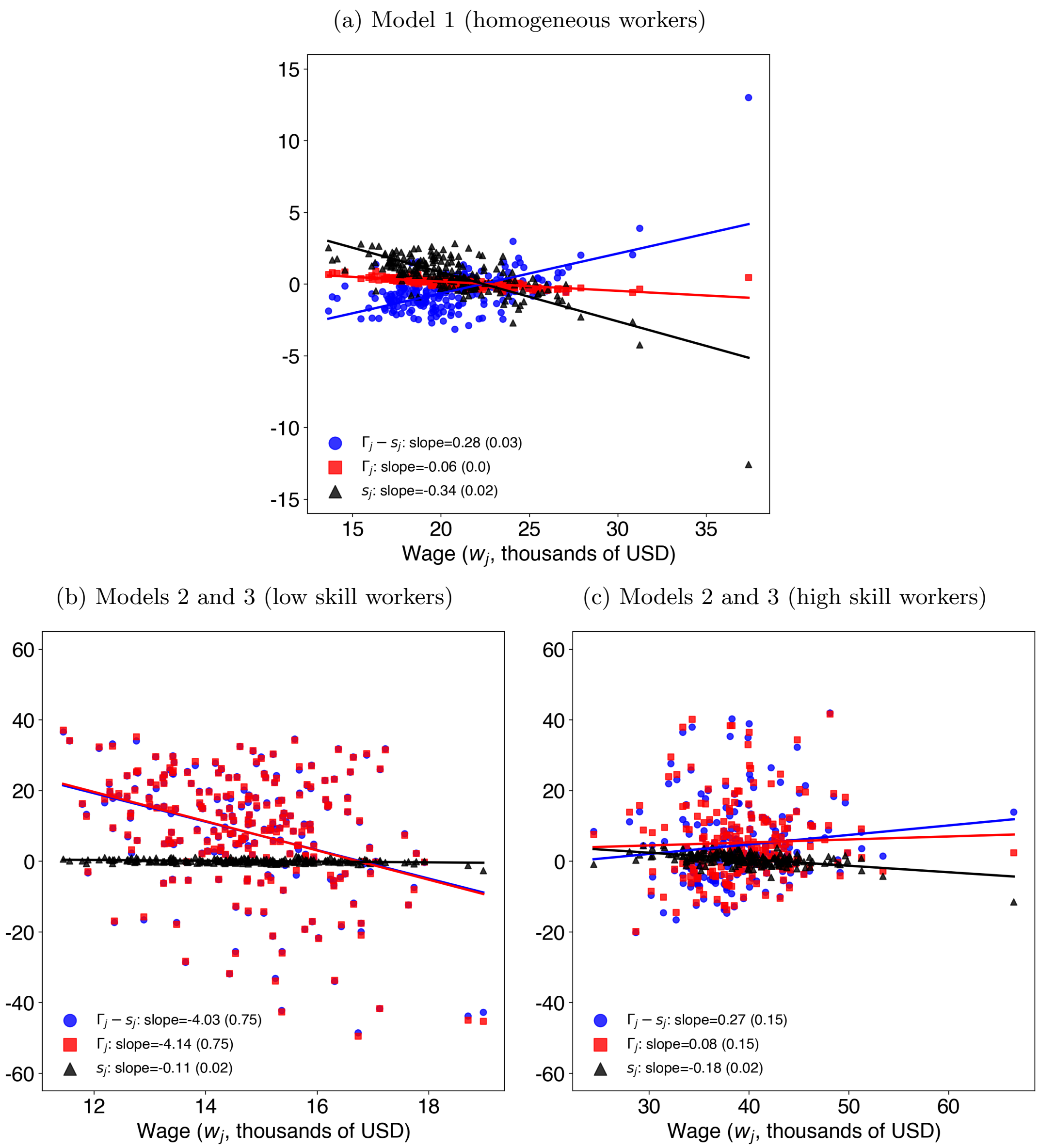

Figure 1: Returns to Marginal Subsidies across Cities

Note: All series are re-centered to their mean.

fare impacts associated with these reallocations are however negligible (around 0.02%, with this order of magnitude not being affected by increasing or decreasing each spillover elasticity by 50%).

Figure 2 reports the optimal subsidy in models 2 and 3 by group of workers. Breaking down the model by skill groups radically changes the qualitative message. Within the low-skill group, optimal transfers redistribute to low-income places much more strongly than in the data, a result of the strong negative congestion spillovers among low-skilled workers in the calibration. In contrast, high-skill workers generate positive productivity and amenity spillovers on both types. The direction of the optimal subsidy is therefore ambiguous: positive spillovers on the own group pushes towards high-skill concentration in high-wage areas, but the positive spillovers on the low-

(a) Model 1 (homogeneous workers)

(b) Model 2 (heterogeneous workers)

(c) Model 3 (het workers and trade frictions)

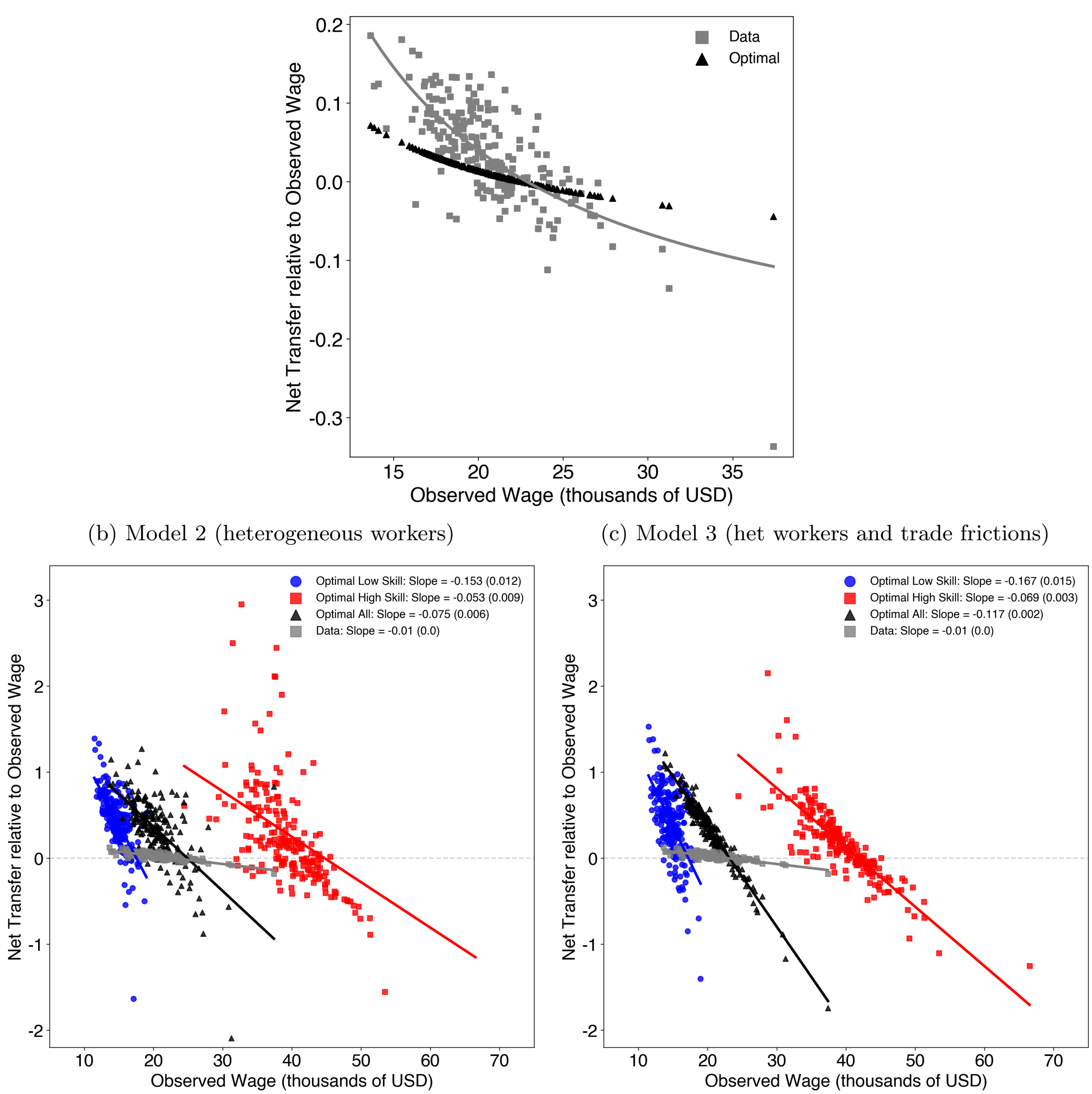

Figure 2: Optimal Transfers

Note: All series are re-centered to their mean.

skill pushes towards subsidizing a higher concentration of high-skill in low-skill intensive places. On net the second force dominates in this calibration, favoring more mixing between low- and high-skill workers in low-wage areas than in the actual allocation. Moreover, this pattern stands in stark contrast with the first-order approach from panel (c) of Figure 1. While the incentives for increasing the concentration of high skill workers dominate locally around the observed allocation, it is more beneficial globally to foster growth in low-income locations through a greater presence of high-skill workers there. The welfare impacts of optimal subsidies are now more significant, with around 1% welfare gains for both worker types.

The patterns appear qualitatively similar between models 2 and 3, implying that incorporating trade frictions does not affect which cities should be subsidized to attract high- and low-skill workers. The quantitative extent of optimal subsidies, however, does differ somewhat across these two cases. The pattern of redistribution towards currently low-wage cities is stronger in the presence of trade frictions.

3.7 Recent Applications

We conclude this section by reviewing research that has quantified the effect of spatial policies, either using the class of models developed above or using different frameworks.

Optimal Population Sizes with Homogeneous Workers Fajgelbaum and Gaubert (2020), Eeckhout and Guner (2017), and Albouy et al. (2019) compute gains from optimal population distribution in the U.S. in a case with homogeneous workers and reduced-form amenity or productivity spillovers. They all impose spillover elasticities in the range typically estimated in the empirical studies. Across papers there are many differences in terms of functional forms, parametrization, forces included in the model, and data. Fajgelbaum and Gaubert (2020) uses a quantitative spatial model with heterogeneity in fundamental productivity, amenities and trade costs to match observed data on wages, expenditures and trade; while Albouy et al. (2019) uses a model with free entry of cities and productivity heterogeneity only. Despite the differences, these exercises give two fairly consistent messages. First, optimal population reallocation under these assumptions leads to small welfare gains (ranging from negligible in Fajgelbaum and Gaubert (2020) to about 1% in the other studies). Second, optimal policies to correct spillovers should on net redistribute from rich to poor locations, but observed redistribution is stronger than what is optimally prescribed on efficiency grounds only, implying that larger cities grow on net in the efficient allocation compared to what is observed in the data. However, both messages are highly sensitive to considering heterogenous rather than homogeneous workers, as we discuss below.

Various papers have studied population misallocation across Chinese cities due to external effects, paying particular attention to frictions caused by the Hukou migration system. Au and Henderson (2006) uses a systems-of-cities model, where utility per worker is inverted-U shape with population. The model features agglomeration effects driven by input differentiation as in New Economic Geography models and reduced-form Marshallian externalities. These forces are dominated at large city scale by internal commuting costs. The calibration of the model yields heterogeneous city-level productivity elasticities, in contrast with the previous papers. They are driven by different input shares in production in different cities, calibrated from the data. Using reforms to the Hukou system as a source of variation, they find evidence consistent with an inverted-U shape in GDP per worker. The paper concludes that Chinese cities are largely undersized, compared to a benchmark where each city would choose its optimal population size by independently maximizing its GDP per-capita. This approach is consistent with cities that can draw workers from rural areas and with workers unable to migrate freely between cities. Wu and You (2023) revisit this conclusion in a framework where Hukou rules are captured by taxes and workers can reallocate freely. They find the largest Chinese cities to be smaller than what a national planner would choose, as in the U.S. papers mentioned above, although they appear too large from a perspective of maximizing city-level utility.

Public Spending and Local Tax Rates The previous papers take a reduced-form approach to model spillovers, without stressing a particular microfoundation. Some papers have quantified optimal policies accounting for the structure of local taxes as a source of distortions in contexts where public spending is the agglomeration source.[24] Fajgelbaum et al. (2018) study state taxes in the U.S. incorporating income, sales, and corporates state taxes in a framework where workers and firms respond to taxes by moving. Rates of mobility and preference for government services are estimated using tax changes. They find that eliminating dispersion in local tax rates would raise national welfare gains by 0.6%. Albouy (2012) and Blouri and Ehrlich (2020) consider optimal intergovernmental grants in Canada and the EU. Albouy (2012) shows evidence suggestive that those transfers do not seem to match criteria of equity or efficiency. Most papers assume that governmental transfers are paid directly to residents or used for a homogeneous government good; Blouri and Ehrlich (2020) instead consider alternative uses and argue that the optimal transfers should be used as wage subsidies in poorer peripheral EU regions and as infrastructure investments, which reduce bilateral travel times, in productive central areas. Henkel et al. (2021) match a quantitative spatial model of Germany to the spatial transfers used by local governments to invest in public goods. Their calibration suggests that optimal transfers would double the welfare gains from spatial policies, with optimal transfers following the same direction as the current system but smaller in scale. Kim (2023) estimates preferences for government services after a fiscal reform in South Korea and argues that jointly accounting for detailed migration and commuting margins matters to compute optimal interregional government transfers.[25]

Optimal Worker Sorting Fajgelbaum and Gaubert (2020) implement optimal spatial policies in the U.S. using a framework with two types of workers, low and high skill. The results contrast with the case with homogeneous workers. They find larger gains and optimal transfers that are very different. Bilateral spillovers across worker types in both amenities and productivities, based on

[24]A large literature dating back to Wilson (1986) and Zodrow and Mieszkowski (1986) has investigated conditions such that taxes and transfers lead to an efficient allocation as a result of inter-governmental competition. Recent quantitative analyses have incorporated endogenous spatial tax rates as a source of misallocation. One of the primary place-based policies used in the United States is business tax incentives aiming to attract firms (Slattery and Zidar, 2020), and Ferrari and Ossa (2023) quantify the losses from a non-cooperative Nash equilibrium in state subsidies. D'Amico (2021) studies a model federal spatial are decided by local voters, whereas Slattery (2024) models the competition between local governments trying to attract firms by offering discretionary subsidies as an auction.

[25]A strand of the literature also studies how existing place-*blind* taxes distort spatial allocations. Specifically, progressive federal income taxes lead to efficiency losses by taxing high-wage locations relatively more. Albouy (2009) argues that the income tax system should be based on real income rather than nominal income to reduce this inefficiency, while Colas and Hutchinson (2021) calibrates a model with heterogeneous households to argues that the efficiency gains of reducing the tax burden on high-income cities needs to be balanced out against the corresponding negative distributional effects.

Diamond (2016), strongly drive the results. Given those elasticities, which imply positive spillovers from high- to low-skill workers, common welfare gains for all workers of around 4% are attained by implementing optimal transfers, with optimal redistribution from high- to low-wage locations being stronger than in the data for both groups. Larger U.S. MSAs are too large on average, and particularly so in terms of high-skill workers, although this result masks considerable heterogeneity across MSAs. The optimal allocation reallocates high-skill workers into small, high-skill-scarce cities to generate more mixing in these locations.

Rossi-Hansberg et al. (2019) also implement optimal transfers with heterogeneous workers. The paper is motivated by the trend towards spatial polarization of occupations in the U.S., with cognitive occupations increasingly concentrated in large cities. The model features industries that differ in how intensively they use cognitive occupations. They estimate productivity spillovers by worker type from the response of model-implied productivity to city size and composition. They find that workers in cognitive occupations benefit from positive spillovers within their group, but face negative spillovers from other groups. On the preference side, the model does not account for amenity spillovers, instead relying on idiosyncratic worker preferences for location as a congestion force. The optimal allocation involves more clustering of cognitive occupations than observed, i.e. higher polarization. However, cities hosting these clusters shrink due to congestion effects.

Optimal Transfers with Local Distortions A literature studies the impact of removing local distortions or "wedges" to correct spatial misallocation, rather than implementing transfer policies that yield efficiency. We mention here only a few papers which are more closely related to our central theme of implementing optimal transfers.[26] In a systems-of-cities framework similar to Au and Henderson (2006) that also includes amenity spillovers, Desmet and Rossi-Hansberg (2013) studies city-specific labor wedges in U.S. and China. They infer these wedges using an approach in the style of Hsieh and Klenow (2009). They find a bigger role for these frictions in driving the size distribution of cities in China than in the U.S., as well as larger welfare impacts of removing them in that country. In this framework, efficiency losses due to agglomeration and congestion spillovers appear very small under homogeneous workers. Hsieh and Moretti (2019) use a neoclassical spatial model to argue that wedges in land markets, in particular heterogeneous land regulations leading to heterogeneous housing supply elasticities, may drive misallocation across U.S. cities. A caveat to their approach is provided by Greaney (2023), who cautions that modeling distortions to land markets as a shock to an elasticity makes the welfare impacts unit-dependent. He proposes a way to adjust the housing elasticities and, when doing so, finds a negligible impacts from heterogeneous housing supply elasticities on welfare.

Optimal Firm Sorting Gaubert (2018) considers a model where heterogeneous firms benefit differently from agglomeration effects. The firms that have the most to gain from agglomeration

[26]Other papers removing frictions in a spatial allocation include for example Behrens et al. (2017) for both trade and spatial frictions in a systems-of-cities model, Tombe and Zhu (2019) and Bryan and Morten (2019) for removing migration frictions in quantitative spatial models, and Franco (2024) for spatial dispersion in markups.

effects are the most productive firms and they sort therefore into larger cities. In this context, the use of place-based policies - which typically subsidize firms in low-productivity areas - turns out to have negative welfare impact, as they prevent firms from sorting to the right place. Furthermore, they hurt areas that have slightly higher productivity, from which firms move out chasing subsidies. These places lose agglomeration effects, hence lose more firms, that move to larger, more productive areas. As a result, large and small cities grow, but middle-sized cities shrink. These policies therefore do not necessarily reduce spatial inequality between cities.

Motivated by the rich and persistent heterogeneity in employment rates across location, Bilal (2023) develop a theory of job search where employers with heterogeneous productivity sort over space. Workers are also free chose where to live. They match subject to search frictions. In equilibrium, high productivity employers have the most to lose in locating in areas where workers are harder to find. Therefore, they sort into slack labor markets. This sorting predicts local labor market flows that are consistent with empirical evidence. Such spatial equilibrium features misallocation because of a labor market pooling externality - low productivity employers are too shielded from competition with high productivity employers on the labor market. Optimal policies incentivizes productive employers to relocate toward high-unemployment locations.[27]

3.8 Future Research

The discussion suggests several areas where more research would be useful. Analyzing optimal spatial policies in dynamic frameworks is one avenue, as a host of real-world considerations cannot be addressed with static models. For example, accounting realistically for migration frictions in designing policies calls for a dynamic approach. Furthermore, place-based policies may be temporary rather than permanent. Central governments may leverage the inter-temporal nature of their budget constraint to subsidize locations, as spillover effects may be dynamic and take time to unfold. The literature has made progress on developing dynamic spatial frameworks with agglomeration effects (Desmet et al., 2018; Allen and Donaldson, 2020; Nagy, 2023); see the chapter on dynamic spatial models in this handbook. These frameworks seem ripe for optimal policy analysis. Donald et al. (2023) and (Wu and You, 2023) study optimal spatial transfers in the dynamic framework with migration frictions of Caliendo et al. (2019).

Another interesting theme that has been unexplored in the quantitative spatial literature concerns optimal policies outside the region of the parameter space where models are "well-behaved" and feature unique equilibria. Whether the market economy operates in this multiplicity region is an empirical question that depends on parameters. However, this question seems important because a fundamental idea behind some place-based policies is the notion of a "big push" to jump-

[27]Recent papers develop models of two-sided sorting of firms and workers across space. Oh (2023) and Hong (2024) characterize the equilibrium of such models and show they predict the colocation of high productivity firms and workers in larger cities. They discuss efficiency in the context of the different spatial externalities that may arise. In Oh (2023), frictional labor markets lead larger cities that are too large compared to the efficient allocation, as they try to find better firms and earn higher wage. In Hong (2024), the optimal spatial policy would encourage more co-location of high productivity firms and workers, while redistributing income to low-earning cities.

start an area into a better, self-sustaining equilibrium. Focusing on the restrictive parameter space where there is no notion of such alternative equilibria, as the current quantitative literature mostly does, seems restrictive against this backdrop. Importantly, the methods reviewed in this section and in the previous one can be used to obtain necessary conditions for optimal transfers even in the multiplicity region of the parameter space; for example, the methods readily apply within the multiplicity region of the classic Krugman (1991)-Helpman (1998) models, and could be equally applied to quantitative versions of these models.

Throughout the discussion we have focused on first-best optimal policies. However, in reality, policymakers may have access to less flexible instruments, such as policies that cannot vary by location but can vary on other dimensions, such as by factor of production or income. Those policies will not in general be able to achieve the first best, but may be able to improve on the market outcome. A fruitful area for future research is to study second-best policies, and in particular to determine how far from the first-best allocations these restricted policies lead to. A closely related question is to investigate to what extent other policies that vary in space, such as zoning or housing policies, help enhance efficiency when first-best instruments are not available. For example, Kline and Moretti (2014) show that, around a free-market allocation without any transfers and with efficiency spillovers, interventions that reallocate capital across locations do not have first-order impacts on welfare via reductions in misallocation, suggesting a limited role for second-best interventions in that context.

Finally, a greater connection between models and empirical estimates is needed. The theory and the numerical examples we discussed show that a clear understanding of both the magnitude and the shape and heterogeneity of spillovers is key to designing optimal spatial policies. Theory may be ahead of empirics in this regard, both concerning agglomeration in productivity and, perhaps even more so, agglomeration (or congestion) effects via amenities. A rich and growing empirical literature estimates impacts of place-based policies in both the short and the long run across many different contexts (e.g., Kline and Moretti (2014) in the U.S., or recently Atalay et al. (2023) in Turkey and Incoronato and Lattanzio (2023) in Italy, among others). Typically these studies are not focused on telling apart arguably efficient market interactions from general equilibrium effects from spillovers, but doing so would be particularly informative for the design of optimal policies. A recent literature focuses on estimating spillovers using rich face-to-face interaction data (e.g. Atkin et al. (2022) or Couture et al., 2024), offering a new promising avenue to identify these effects at highly disaggregated levels. Incorporating these micro measurements and teasing out their aggregate policy implications would also be a promising avenue for research.

4 Spatial Policies as Redistributive Tools

Economic outcomes, including wages and income, vary significantly across regions with some areas enjoying prosperity while others lag behind. These regional inequalities have become a pressing policy concern, leading governments to adopt place-based policies that redistribute resources to economically disadvantaged areas. This section focuses on the redistributive motive for these spatial policies.

Indexing redistribution by place raises an important issue: if the goal is to redistribute from richer to poorer households, governments can use progressive taxes and transfers indexed on income. Is there, in addition, a rationale to use place-based redistribution? The traditional answer to this question is "no" (see e.g. Glaeser (2008)). Place-based transfers are viewed as inefficient because they encourage economic activity in less productive places at the expense of more productive places. This section revisits this issue in the context of recent contributions and discusses specific microfoundations such that policies may optimally condition on place, rather than on income alone.

To analyze this equity rationale separately from the efficiency rationale of the previous sections, the model now deliberately excludes any efficiency-driven reasons for place-based policies such as spillovers or public goods provision. As in these conditions the laissez-faire allocation is efficient, the focus is solely on the redistributive objectives of place-based transfers.

4.1 Model setup

The country is comprised of J locations indexed by j where individuals may choose to live and work. Production is kept as simple as possible. Perfectly competitive firms produce a homogeneous and perfectly tradable good. They produce using labor under constant returns to scale with heterogeneous productivity across locations. Heterogeneous individuals are indexed by $\omega = (\theta; \varepsilon)$, where θ indexes skill and $\varepsilon = (\varepsilon_1, ..., \varepsilon_J)$ is a vector of preferences over locations. These characteristics may be correlated. Workers supply labor inelastically in their location, earning a wage $w_j(\theta)$ for skill θ equal to their output of the tradable good. Individuals consume the freely traded good and make a discrete choice of location to obtain utility:

$$u^\omega = \max_j u_j\left(c_j^\omega\right) + \varepsilon_j^\omega. \tag{51}$$

Both skill and preferences shape the location decision, but only skill influences location-specific income. The utility function is concave in consumption c and is indexed by j, thus allowing locations to differ in terms of amenities.

The central government is a benevolent planner who maximizes a social welfare function with Pareto weights μ^ω on type ω,

$$W = \sum_\omega \mu^\omega u^\omega, \tag{52}$$

subject to budget balance. The planner's taste for redistribution may come from heterogeneous Pareto weights as well as from decreasing marginal utility in consumption. Preference shocks ε_j^ω do not *directly* affect marginal utilities because they enter additively in (51), but they do matter indirectly through location choice. Section 4.4 discusses the challenges posed by non-additive shocks for normative analysis.

Assume first that the planner observes the individuals' types ω and can implement transfers that are ω-specific. In this case, social welfare W is maximized when the marginal utility of consumption of all individuals (weighted by their Pareto weight) is equalized. Individual-specific transfers are chosen such that $\mu^\omega \frac{du^\omega}{dc}$ is equal for all individuals.[28] Such non-distortive transfers would constitute an implementation of the second welfare theorem. In reality, however, the information set of governments is more limited and taxes are typically indexed on observable choices made by individuals. Taxes therefore distort choices, leading to an efficiency cost of redistribution.

We consider here two types of fiscal instruments indexed on choices: place-specific subsidies, which are our main focus, as well as income taxes, which are ubiquitous in practice. Place-specific transfers capture in reduced form the many ways in which a central government may channel resources to specific locations. They summarize in particular how taxes vary across space and how benefits are distributed across locations by the central government.[29]

We now define the planner's problem. Let $L(\theta)$ denote the exogenous measure of agents of skill θ and by $F_\theta(\{\varepsilon\})$ the distribution of preference shocks for skill θ. The planner can choose a vector of place-specific subsidies $\{s_j\}_{j=1,..,N}$ that are constant within a location, i.e. do not depend on income. For realism, we allow these spatial transfers to complement an existing (place-blind) income tax $t(w)$. The consumption of an individual ω located in j is $c_j(\theta^\omega) = w_j(\theta^\omega) - t(w_j(\theta^\omega)) + s_j$, where θ^ω is ω's skill. A measure $L_j(\theta)$ of individuals with skill θ choose location j (equal to the choice probability for j among θ types times $L(\theta)$), while $L_j \equiv \sum_\theta L_j(\theta)$ is the total number of residents in j. As a reminder, location choice probabilities within a type θ are non-degenerate because of the ε shocks within type.

The social welfare function is:

$$W = \sum_\omega \mu^\omega \max_j \left\{ u_j\left(w_j(\theta^\omega) - t(w_j(\theta^\omega)) + s_j\right) + \varepsilon_j^\omega \right\} \tag{53}$$

$$\text{subject to } \sum_\theta \sum_j t(w_j(\theta)) L_j(\theta) - \sum_\theta \sum_j s_j L_j(\theta) \geq 0,$$

where the constraint is the government budget balance (equivalent to national goods market clearing). We will use the notation $u_j(\theta) \equiv u_j(c_j(\theta))$ as a shorthand for the common component of utility in location j for skill θ. Throughout the section, when we talk about changes in social welfare we mean changes in W.

4.2 When is Place-based Redistribution Desirable?

Arguably, place-based redistribution is justified as a redistributive tool only in contexts where the income tax is insufficient to attain that goal. When is this the case? This question is motivated after Atkinson and Stiglitz (1976), who show conditions under which taxing consumption for redistribution purposes is superfluous if an income tax is available. Some of their insights carry

[28] Marginal utility is evaluated at the location chosen in equilibrium, i.e. $\frac{du^\omega}{dc} \equiv \frac{\partial u_{i^*_\omega}}{\partial c_{i^*_\omega}}$ for $i^*_\omega = \arg\max_i u_j\left(c_j^\omega\right) + \varepsilon_j^\omega$.

[29] Taking the U.S. as an example, federal taxes may vary across space explicitly (e.g. the SALT deduction) and implicitly, as analyzed by Albouy (2009). Federal transfers to local communities also vary widely across space (UrbanInstitute, 2024).

through to the context of location choices, but additional forces need to be considered here as, for instance, location decisions impact both consumption and production.

Gaubert et al. (2024) tackle this question. They assume that an income tax system –perhaps optimally chosen– is in place, and ask: can place-based redistribution increase welfare further? For this analysis, the country is partitioned in two regions, i and j. The planner is utilitarian so that $\mu^{\omega} = 1$ for all individuals and the planner's preference for income redistribution stems only from the concavity of $u(\cdot)$. Starting with a place-blind income tax $t(w)$ only, consider introducing small place-based taxes. Specifically, we compute the welfare impact of a small perturbation that introduces a lump-sum place-based transfer $\frac{ds}{L_j}$ to residents of the poorer region financed by a lump-sum tax $\frac{ds}{L_i}$ levied on the rest of the country. This place-based reform of the tax system is *ex-ante* budget neutral. The social welfare change from this reform is:

$$\frac{dW}{\chi} = \bar{\lambda}_j - \bar{\lambda}_i + \sum_{\theta} \left[t\left(w_j\left(\theta\right)\right) - t\left(w_i\left(\theta\right)\right)\right] dL_j\left(\theta\right). \tag{54}$$

Place-based redistribution towards j is desirable when $\frac{dW}{\chi} > 0$. In this expression, $\bar{\lambda}_j$ is the average social marginal welfare weight of location j. Formally, it can be computed as the average marginal utility of income of residents of j:[30]

$$\bar{\lambda}_j \equiv \frac{1}{\chi} \sum_{\theta} \frac{L_j\left(\theta\right)}{L_j} u_j'\left(c_j\left(\theta\right)\right). \tag{55}$$

Intuitively, $\bar{\lambda}_j$ measures the direct welfare effect of a marginal income gain for a resident of j, expressed in terms of the cost of public funds χ (the Lagrange multiplier on the government budget constraint). That is, the planner would be indifferent between giving a dollar to an average inhabitant of j and getting $\bar{\lambda}_j$ dollars of budget.

Assume that j is the poorer region and has a higher social welfare weight $\bar{\lambda}_j$. We come back to why it might be the case next. Then, the term $\bar{\lambda}_j - \bar{\lambda}_i > 0$ in (54) captures the equity gain of the place-based transfer from i to j, while the term $\sum_{\theta} \left[t\left(w_j\left(\theta\right)\right) - t\left(w_i\left(\theta\right)\right)\right] dL_j\left(\theta\right)$ captures an efficiency cost given pre-existing income tax system. Some individuals move to j in response to the subsidy; if, as a consequence, their income tax contributions fall (which they will typically do, with wages being typically lower in j) then government revenues fall, a fiscal externality.[31] As long as this fiscal cost of movers is not too high compared to the planner's valuation of the transfer, introducing some place-based taxation is desirable.

Before moving on to specific examples that illustrate when such a reform may be desirable, we review situations that may yield $\bar{\lambda}_j - \bar{\lambda}_i > 0$. A location may have a lower income for two (possibly complementary) reasons. The first reason is spatial sorting, with low-skill workers rela-

[30]The computation of these welfare weights uses the Williams-Daly-Zachary theorem: in a random utility model such as this one, with agents solving $\max_j u_j + \varepsilon_j$, the derivative of expected utility with respect to the common component of utility of j equals j's choice probability: $\frac{\partial \mathbb{E}\left(\max_j u_j + \varepsilon_j\right)}{\partial u_i} = \Pr\left(i = \arg\max_j u_j + \varepsilon_j\right)$ (McFadden, 1978, 1980).

[31]In this case with linear production, the fiscal externality is the only efficiency cost. Outside the linear production case, place-based taxation generically leads to distortions in production.

tively more likely to locate in j. This spatial sorting may be driven in particular by different skills having different preferences for location as captured by their distribution $F_\theta(\varepsilon)$, or by comparative advantage in production captured by $\{w_j(\theta)\}$. The second reason is location-specific productivity: workers may be less productive in j regardless of skill. These two sources of heterogeneous income across space will have different implications for whether place-based redistribution improves upon a place-blind income tax.The two examples below highlight polar cases where a place-based reform increases welfare in the presence of a place-blind income tax.

Example 4. Preference-driven sorting.

Assume that, conditional on skill, workers have the same wage $w_k(\theta) = \theta$ in the two locations $k = i, j$, hence the same income $\theta - t(\theta)$. The income tax is left general, so long as net income $\theta - t(\theta)$ increases with gross income θ, an assumption that optimal income taxes satisfy (Piketty and Saez, 2013). Spatial sorting is driven exclusively by the different preferences for locations across different skills. An individual of type $(\theta, \{\varepsilon_k\})$ faces utility:

$$u_k = u(c_k(\theta)) + \varepsilon_k \text{ for } k = i, j.$$

We assume $u(.)$ concave, $\varepsilon_i \sim \mathcal{N}(\theta, 1)$, and $\varepsilon_j = 0$. Workers with skill θ get the same level of consumption in both locations, hence choose to locate in i if $\varepsilon_i - \varepsilon_j \geq 0$. This happens with probability $\Phi(\theta)$, increasing in θ, where $\Phi(.)$ is the CDF of the standard normal distribution. Therefore, higher incomes are over-represented in i (in a first-order stochastic dominance sense), implying $\bar{\lambda}_j - \bar{\lambda}_i \geq 0$, an equity motive for redistributing from i to j. When a small place-based subsidy to j is introduced, marginal workers of all skills (those for whom $\varepsilon_i - \varepsilon_j = 0$) move from i to j. As they do, their income θ and hence tax contribution $t(\theta)$ do not change, hence there is no efficiency cost. Following (54), some amount of spatial transfers unambiguously increases welfare, a finding that mirrors classic result in public finance saying that taxing a consumption good may improve upon taxing income alone when preferences are heterogeneous across goods (Saez, 2002). Importantly, this is the case starting from any income tax schedule.

Example 5. Comparative advantage-driven sorting.

Assume now that productivity alone drives sorting. Specifically, below a certain skill level θ^*, workers have the same income θ in both locations; but for skills above θ^*, $w_i = 2\theta$ and $w_j = \theta$. Thus, low-skill workers face no income differences in space, but higher skills earn more in i. Spatial sorting ensues. To fix ideas, one can choose the distribution of net preferences for i, $\varepsilon_i - \varepsilon_j$, to be common across skills and such that all high-skill workers strictly prefer i while low-skill workers (below θ^*) live in both locations. This can be ensured assuming a large enough θ relative to the support of $\varepsilon_i - \varepsilon_j$. In this case, location i has strictly higher incomes and redistributing from i to j yields equity gains. A small subsidy to j induces lower-skill workers to migrate there without experiencing an earning loss. Higher-skill workers, who are all inframarginal, remain in i even after the introduction of the small subsidy. Hence, this place-based transfers yields equity gains at no efficiency cost, implying by (54) that it is desirable.

These polar examples echo trends highlighted in the empirical literature on spatial sorting. College and non-college workers are estimated to have systematically different preferences for large, coastal cities (Diamond, 2016), explaining in part their spatial sorting. In addition, high-skill technical jobs are increasingly concentrated over space (Moretti, 2012; Autor, 2019), disproportionately attracting a high-skill population.

In these examples, place-based redistribution to poorer areas is unambiguously desirable. The key force that potentially dampens the positive welfare impact of a place-based subsidy to the poor region is the earnings response of movers. It arises whenever place has a causal effect on earnings and workers are mobile. In fact, it is possible that a place-based subsidy to the rich, productive region is desirable in some cases to incentivize workers to live in the productive location – despite the increase in income inequality this reform generates.

These examples highlight some general lessons about the desirability of place-based policies. Whenever spatial income differences are driven by spatial sorting and this sorting is driven by preferences, the efficiency losses from spatial redistribution are small compared to their equity gains. Place acts as a "tag", although unlike the immutable tags considered by Akerlof (1978) this one is chosen by households, becoming an attribute that is correlated with the sources of income inequality such as skill. However, whenever sorting is driven by comparative advantage, the desirability of spatial redistribution hinges on whether higher skill workers move into low-income areas in response to the subsidies. In other words, if mobility responses are concentrated in occupations without large earnings difference across space, then spatial redistribution towards lower income places is desirable. Finally, when income differences are largely driven by place-based components of productivity, it is ambiguous whether place-based redistribution is preferred.

4.3 Towards Optimal Spatial Redistribution

The previous section considered a marginal transfer around an initial equilibrium. This section characterizes transfers that implement spatial redistribution optimally. It first considers lump-sum spatial transfers, before moving on to spatial transfers that may be also indexed by income.

4.3.1 Lump-Sum Spatial Subsidies

A planner with social welfare function (53) chooses a vector of spatial transfers $\{s_j\}$ to regions $j = 1, .., J$ optimally. For this analysis, we take the place-blind income tax schedule $t(.)$ as given. The first-order condition of the planner's problem characterizes optimal spatial transfers.

Optimality Test Setting up the Lagrangian and taking first order conditions with respect to s_j, an optimal system of spatial subsidies verifies for all $j \in 1, ..., J$:

$$\left(\bar{\lambda}_j - 1\right) L_j = \sum_{\theta} \sum_{i} \left(s_i - t\left(w_i\left(\theta\right)\right)\right) \frac{\partial L_i\left(\theta\right)}{\partial s_j}, \tag{56}$$

where $\bar{\lambda}_j = \frac{1}{\chi}\frac{\partial W}{\partial s_j}$ again denotes the average marginal social welfare weight of inhabitants of i expressed in terms of cost of public funds χ (the Lagrange multiplier on the government budget constraint).[32]

The left-hand side of expression (56) measures the equity benefits of a marginal dollar of subsidy in j. It corresponds to the gross benefits of the subsidy, valued $\bar{\lambda}_j$ for each resident of j, net of the direct cost to finance it. These net benefits are multiplied by the number of recipients L_j. The right-hand side of the expression measures the (indirect) fiscal cost of the subsidy that is generated by the behavioral response to the subsidy s_j: it triggers migrations, potentially in all locations i, with movers typically changing their income tax payments. At the optimum, (56) says that equity gains of spatial transfers are exactly offset by the corresponding fiscal costs.

Expression (56) provides a test of optimality of a system of spatial transfers, for a given set of Pareto weights. Alternatively, the condition could be used to estimate the social preferences that (best) correspond to an observed set of spatial transfers $\{s_j\}$, an exercise in the spirit of Bourguignon and Spadaro (2012) who estimate Pareto weights revealed by the income tax schedule. [33]

Optimal Formulas under Parametric Restrictions The formula (56) characterizes the system of optimal place-based subsidies. To make further progress and examine the form these subsidies may take, one needs to impose more structure, for which we follow Ales and Sleet (2022), augmented here with heterogeneous skills. Assume next that preference shocks are distributed logit, identical for all skills, i.e. $F_\theta(\varepsilon_i) = \exp\{-(e^{-\varepsilon_i})\}$. Then, the migration responses to a shock in j simplify to:=

$$\frac{\partial L_i(\theta)}{\partial s_j} = L_j(\theta)\left(1_{i=j} - L_i(\theta)\right) u_j'\left(c_j(\theta)\right). \tag{57}$$

In response to a positive subsidy shock in location j, individuals abandon other locations in proportion to their size. Because the ε_i are iiid across space, a shock to one location does not lead to particularly stronger reallocations away from or into any locations –all locations are equally close substitutes. The optimal spatial subsidies can be solved out from (56) and simplify to:

$$s_i - \bar{t}_i = \frac{\bar{\lambda}_i}{\overline{u_i'}} - \frac{1}{\overline{u_i'}},\ \forall i, \tag{58}$$

where bars denote appropriately weighted averages of taxes or marginal utilities in a given location.[34] Equation (58) says that the optimal per-capita subsidy in i, net of fiscal revenues from the income tax, increases with equity gains: it increases with the Pareto weight corresponding to the

[32]With μ^θ that depend on skill θ but not on shocks ε, these weights are $\bar{\lambda}_j = \frac{1}{\chi}\sum_\theta \frac{L_j(\theta)}{L_j}\mu^\theta u_j'\left(c_j(\theta)\right)$.

[33]Expression (56) can be used to show that the laissez-faire equilibrium, without any transfers, maximizes social welfare when weights equal inverse marginal utilities: $\mu^\omega = \frac{\chi}{u'_{j(\omega)}\left(c_{j(\omega)}(\theta)\right)}$, where $j(\omega)$ is the location of individual ω in this equilibrium. Then, $\bar{\lambda}_j = 1$ and both the left- and right-hand side of (56) are zero.

[34]Specifically, $\bar{t}_i = \frac{\sum_\theta L_i(\theta)u_i'(c_i(\theta))}{\sum_\theta L_i(\theta)u_i'(c_i(\theta))}t\left(w_i(\theta)\right)$ is the average of the per capita income tax over skills in i, weighted by their marginal utilities, and $\overline{u_i'} = \frac{\sum_\theta L_i(\theta)}{L_i}u_i'\left(c_i(\theta)\right)$ is the average marginal utility of income in i, so that $\frac{\bar{\lambda}_i}{\overline{u_i'}} = \frac{1}{\chi}\frac{\sum_\theta \frac{L_i(\theta)}{L_i}\mu^\theta u_i'(c_i(\theta))}{\sum_\theta \frac{L_i(\theta)}{L_i}u_i'(c_i(\theta))}$ is a measure of average Pareto weights in i.

location and with marginal utility of consumption.

In the logit case, the optimal subsidy only varies across space because the equity motive does; in contrast the cost of taxation is constant across space. Ales and Sleet (2022) consider alternative specifications that capture more realistic variation in this efficiency cost of taxation. They consider the separable mixed logit. There is no heterogeneity in skill. Individuals are characterized by additive Gumbel-distributed preference shocks and by another parameter that influences their utility, but not their income. One can think of this parameter as modeling a systematic component of preferences over locations across demographic groups. The preference shocks ε can be correlated within demographic groups, as in the BLP model (Berry et al., 1995). This setup accommodates richer substitution patterns between locations than the simple logit: a group might value an amenity (e.g., sunshine) that is particularly present is some locations, treating them as high substitutes, while another might value other dimensions (e.g., presence of an opera) so that these same locations are poor substitutes in the eyes of the second group. With these richer substitution patterns, the response to a subsidy in j is heterogenous across space, as it attracts more workers in locations that are closer substitutes with location j. As a result, not only are optimal subsidies higher in locations that yield high equity gains as in (58), but in addition subsidies are higher, all else equal, in locations that happen to be substitute with locations that yield high equity gains.

4.3.2 Spatial Transfers Indexed on Income

With the lump-sum spatial subsidies considered so far, poor households of wealthy areas are taxed to subsidize rich households of distressed areas. With redistributive goals in mind, it seems intuitive that optimal transfers be indexed instead by both place and income. Gaubert et al. (2024) tackle this question. To do so, they consider a planner that chooses an income tax schedule optimally in each location.[35] They derive the optimal non-linear income tax schedules for two locations. Using numerical simulations, they quantify the corresponding place-based transfers and show how they vary by income for various calibrations.

When sorting is generated entirely by heterogenous preferences for locations of different skills, the optimal spatial transfer to the poorer region is done both through a large per-capita transfer and through lower marginal income tax rates in this region. There are fewer high income workers to tax there, so that high marginal tax rates, which are distortive, bring smaller benefits than in the high income region. The whole non-linear tax schedule is shifted down in the poorer region.

When sorting is driven by comparative advantage in production, the direction of optimal transfers depend on how mobile individuals are. When migration elasticities are low, place-based taxation is similar qualitatively to the case above where sorting is driven by preferences alone. However, when migration elasticities are larger, these patterns change, at least for higher incomes. They now face higher marginal tax rates in the poor and low-productivity region, which incentivizes them to

[35] While we considered labor supply to be exogenous so far, labor supply is now endogenous hence responds to the income tax. This is necessary to face a meaningful trade-off when devising the optimal income tax: it balances equity gains (redistribution between income levels) with efficiency costs (disincentivizing labor supply).

live in the productive location. The direction of optimal transfers changes and the richer region may even be subsidized, on average. A takeaway is that in the presence of comparative advantage the redistributive motive tends to dominate for lower incomes, which pushes towards redistribution to lower incomes in poorer regions. For higher incomes, considerations of the efficiency cost of taxation are more important and push towards lower taxes in the richer region . The mobility of workers determines which forces dominate.

The analysis shows that the causes of spatial sorting as well as mobility responses at various income levels impact the design of optimal spatial transfers. The drivers of spatial sorting and the estimation of migration elasticities have been a point of focus in the literature. More empirical research is needed on the possible heterogeneity by income level - rather than the averages the literature typically focuses on - in the forces that drive optimal place-based redistribution: migration responses, and causal effect of place on earnings. These are key for the efficiency cost of spatial policies.

4.4 Normative Implications of Preference Shocks

Preference shocks for location (the ε_j^ω in equation (51)) lead to heterogeneous location choices of individuals that are observationally identical. Besides their realism, these shocks were introduced in the literature for tractability and to match empirical facts, as they help to rationalize elasticities of labor mobility across space. Tabuchi and Thisse (2002) are an early example of this modeling approach, which is currently ubiquitous in the quantitative spatial economics literature (Redding and Rossi-Hansberg, 2017), including in analyses such as Ahlfeldt et al. (2015) and Diamond (2016). A few methodological points related to these shocks are worth discussing.

4.4.1 Pitfalls of the Utilitarian Planner

As we have shown in this section, including heterogeneity across workers does not change the conclusion from Section 2 that the competitive *laissez-faire* equilibrium is (absent externalities) efficient: no policy intervention is required on efficiency grounds. This is true about preference shocks as much as about other dimensions of heterogeneity often included in spatial research, such as worker skill (e.g., Diamond, 2016) or place of birth (e.g., Bryan and Morten, 2019).

However, in contexts with heterogeneous preference shocks (the ε_j^ω in equation (51)), welfare analysis is often conducted using a utilitarian social welfare function to aggregate across agents with heterogeneous preferences, within a group of observable characteristics (such as skill or place of birth). An issue with this practice is that a redistributive motive for government intervention necessarily creeps in: aggregating via a utilitarian welfare function –or, for that matter, any specific social welfare function– creates a ranking among different pareto-efficient allocations, and therefore a rationale to implement spatial transfers in order to target one (arbitrary) point in the Pareto frontier. As a consequence, in contexts where spatial policies may be justified on efficiency grounds –both to target spillovers, as in Section 2, or to implement physical spatial investments, as in

Section 5– the policy prescriptions from a utilitarian planner conflate efficiency considerations and distributional concerns.[36]

To illustrate this issue, consider a simple example from Fajgelbaum and Gaubert, 2020: a spatial economy with only labor in production, agglomeration effects, and idiosyncratic preferences for location. Preference for location j of agent ω is $u_j^\omega = c_j \varepsilon_j^\omega$, where consumption $c_j = w_j + s_j$ equals the wage plus a place-specific subsidy, and where the wage $w_j = Z_j L_j^{\gamma_Z}$ depends on fundamentals Z_j and population L_j through externalities with elasticity γ_Z. The utilitarian planner solves $\max_{s_j} u \equiv \mathbb{E}\left[\max_j c_j \varepsilon_j^\omega\right]$, where the expectation denotes the integral over the realizations of ε_j^ω. Assuming ε_j is Frechet, $\Pr\left[\varepsilon_j^\omega < x\right] = \exp\left(-x^{-1/\sigma}\right)$, this problem is equivalent to $\max_{s_j} u$ subject to $c_j L_j^{-\sigma} = u$ for all j. This structure is mathematically identical to Example 1 but conceptually different: the parameter $-\sigma$ stands in place of the amenity spillover elasticity γ_A to capture that the aggregate (utilitarian) utility of residents in a location decreases as marginal agents select into a location. From 22, the outcome of the utilitarian planner's problem is a wage subsidy rate equal to $\frac{\gamma_Z - \sigma}{1+\sigma}$. This policy combines efficiency considerations (entering through γ_Z) and redistributional considerations (through σ). When $\gamma_Z = 0$, a utilitarian planner redistributes to low-productivity locations, where marginal utility of consumption is higher.

In such context where agents have idiosyncratic preferences for location, how to define optimal policies driven exclusively by efficiency without redistribution also influencing the policy prescriptions is an open methodological question. One approach is simply to show robustness of policies with respect to social welfare functions –it would be interesting to learn the extent to which qualitative and quantitative features of optimal policies driven by efficiency vary with (arbitrarily chosen) preferences for redistribution. Another option would be to carry a multi-dimensional object (the distribution of welfare changes across groups) for policy evaluation. In this case, the outcome of a planner problem would be a set of optimal policies, all of which bring the economy to the Pareto frontier. Such a practice is standard when the groups are discrete and represent, for example, different skills (such as in Fajgelbaum and Gaubert, 2020 across skills) but has not been implemented in the case of a continuum of preference shocks. Yet another approach could be to devise a way to "purge" the welfare implications of counterfactuals from redistribution using decomposition such as those in Donald et al. (2023).

4.4.2 Normative Implications of Non-Additive Taste Shocks

Except in special cases (see next subsection), non-additive taste shocks cannot be generically identified in the data but affect the normative implications. We can see this in a simple example.

[36] The utilitarian welfare function also corresponds to the expected utility of individuals who are *ex-ante* identical (before their taste shocks are realized), but value resources differently *ex-post* depending on their spatial choice (driven by the shock realization). The optimal spatial transfers from this *ex-ante* problem are identical to those we have derived in this section, but have a different interpretation: the *laissez-faire* allocation is interpreted as an incomplete market allocations, and these spatial transfers as "insurance" again states of the world where one chooses to live in a low income location (Mongey and Waugh, 2024). Formula (56) can in this case be seen as a $J-$choice extension of the Bailey-Chetty formula that characterizes an optimal insurance policy with two states (employment and unemployment) (Chetty, 2006), as pointed out by Davis and Gregory (2021) and Donald et al. (2023).

Consider the simple utility function $u(c+\varepsilon_j)$ with $u(.)$ increasing and concave, where $\{\varepsilon_j\}_j$ are taste shocks. Individuals choose location j over location i whenever

$$\varepsilon_j - \varepsilon_i > w_i - w_j. \tag{59}$$

The values of ε_i and ε_j separately play no role in observable choices. Therefore, these two shocks cannot separately identified. However, separately identifying them is necessary for normative statements.

We show this in an example. Consider two alternative models: (a) $\varepsilon_j = 0$ while ε_i is an idiosyncratic shock with some positive variance; and (b) $\varepsilon_i = 0$ while ε_j is an idiosyncratic shock with some positive variance. We next show these two models make opposite normative predictions because they imply different rankings of social welfare weights $\left(\overline{\lambda}_i, \overline{\lambda}_j\right)$. In model (a), residents of i have shocks $\varepsilon_i \geq w_j - w_i$. Therefore, they have social welfare weights (under a utilitarian planner):

$$\overline{\lambda}_i = \frac{u'(w_i+\varepsilon_i)}{\chi} \leq \frac{u'(w_j)}{\chi} = \overline{\lambda}_j,$$

where the inequality stems from the concavity of u. In contrast, in model (b), residents of j have shocks $\varepsilon_j > w_i - w_j$. Therefore, they have welfare weights:

$$\overline{\lambda}_j = \frac{u'(w_j+\varepsilon_j)}{\chi} < \frac{u'(w_i)}{\chi} = \overline{\lambda}_i.$$

In model (a), the planner wants to redistribute to j; while in model (b), she wants to redistribute to i. That two models are observationally equivalent yet make opposite welfare prescriptions is undesirable, a point made by Davis and Gregory (2021) and Gaubert et al. (2024). We chose a particular non-additive example to make this point, but the concern is more general than in this example. With additive taste shocks, however, this issue does not arise: conditional on rationalizing a given set of location choices, the normative implications of the model are independent from the distribution of the additive shock.

4.4.3 Multiplicative Frechet Shocks

A popular specification of utility, especially in quantitative spatial models (Redding and Rossi-Hansberg, 2017) is a multiplicative specification: $v^\omega = \max_j v_j \varepsilon_j^\omega$ where there is a single skill, v_j is typically a Cobb Douglas utility function, and $\left\{\varepsilon_j^\omega\right\}$ are Frechet distributed taste shocks with shape parameter κ ($F(x) = \exp\left[-x^{-\kappa}\right], \quad x > 0$). Using properties of the Frechet distribution, migration is governed by:

$$\frac{1}{L_i}\frac{\partial L_j}{\partial c_i} = \kappa\left(1_{i=j} - L_j\right)\frac{v_i'(c)}{v_i(c)}.$$

One can then directly verify that the optimal spatial transfer formula (56) leads to $s_i = \frac{1}{\kappa}\left(\frac{1}{\chi} - \frac{v_i(c)}{v_i'(c)}\right)$, or, denoting $u_i = \log v_i$,

$$s_i = \frac{1}{\kappa}\left(\frac{1}{\chi} - \frac{1}{u_i'(c)}\right).$$

Therefore, when the shape parameter is 1, this formula is identical to that of the logit case (58). This result is not surprising. The two models are isomorphic, as can be seen through a log transformation of utility.[37] For this reason, the Frechet case is a knife-edge case such that, even if shocks are multiplicative, the normative implications of the model are only dependent on observed location choices like in the additive case.

5 Optimal Transport Networks

We turn to studying investments in transport infrastructure. We devote a separate section to this spatial policy given the central role played by transport costs in spatial economics and because they are studied using a distinct set of methods. Transport infrastructure is naturally one important determinant of trade and mobility costs, with infrastructure impacting monetary, time, and utility costs of goods transport and travel (Atkin and Donaldson, 2015; Bryan and Morten, 2019). A large body of research examines the impact of actual or counterfactual improvements to these transportation networks on economic outcomes (see review by Redding and Turner (2015) and Donaldson (2024) in this handbook). The problem of characterizing optimal transport networks is a priori challenging due to its high dimensionality, with improvements in one part of the transport network affecting the entire system. We review recent papers in this literature that leverage mathematical methods designed to manage this complexity, as well as some of the key findings. We conclude with some suggestions for future research.

5.1 Background: Least-Cost Routes in Spatial Models

Real-world economies have complex transport sectors whose efficiency depends on infrastructure and market structure. In quantitative spatial models, this complexity is usually not modeled, and instead transport costs for goods or people are summarized by matrices of "iceberg" costs τ_{ij}. For one unit of a commodity from location i to arrive location j, $\tau_{ij} > 1$ units of that commodity must be shipped, with the magnitude τ_{ij} usually parametrized as a function of time, distance, or other frictions (Head and Mayer, 2014). In commuting models, where networks transport people, iceberg commuting costs are either direct utility costs (Ahlfeldt et al., 2015) or correspond to productive hours lost (Dingel and Tintelnot, 2020). From now on, for readability, we use the terminology of trade and shipping goods when laying out the model (rather than transporting passengers), but these methods can be adapted to the latter case.

Before thinking about optimal infrastructure investment, a preliminary step is to conceptualize how the spatial distribution of infrastructure affects the transport costs τ_{ij}. In the models that are typically used, goods are shipped on a network defined by a graph. The graph consists of a set of nodes that represent locations where production and consumption take place or where goods may be just in transit; as well as a set of links (or edges) connecting the nodes. For each node j there is

[37] Denote $u^\omega = \log v^\omega = \max_i u_i + \log \varepsilon_i^\omega$. Given that the logarithm of a Frechet distribution is a Gumbel distribution, this model is the logit model described above.

a set of "neighboring" nodes $\mathcal{N}(j)$ that are connected to j via a link. The connected nodes need not be actual neighbors in a geographical sense, and their definition depends on the application. Transporting goods from j to $k \in \mathcal{N}(j)$ entails a cost t_{jk}, so that $1 + t_{jk}$ units must be shipped from j for one to arrive in k. Goods can only be transported from j to $i \notin \mathcal{N}(j)$ through a route r of connected nodes. The corresponding cost from j to i is the product of the individual iceberg costs corresponding to all the links on the route r. The chosen route from j to i is selected from the set $\mathcal{R}_{ij}$ of all possible routes connecting j and i to minimize the "long-distance" iceberg cost:[38]

$$\tau_{ji} = \min_{r \in \mathcal{R}_{ji}} \prod_{(kl) \in r} (1 + t_{kl}) . \tag{60}$$

Expression (60) captures how improvements to a transport network –both small such as improving a railroad link or large such as a new highway– affect the matrix of bilateral trade costs τ_{ji}. When a set of links is improved, τ_{ji} falls for origin-destination pairs that use the improved links, including both origin-destination pairs that already used those links before the improvement and pairs that switch to using them after.

Mathematically, finding τ_{ji} and the corresponding route on a given graph is called the *least-cost route* problem. It is a special case in a broader class of "optimal transport problems" for which solution methods have been developed (Galichon (2016) provides a textbook introduction).[39] For economists using these tools, the key practical takeaway is the duality approach to numerically solving these problems. When applicable (which depends on the convexity of the planner's problem), duality reduces the problem's dimensionality by shifting the problem from optimizing over possible routes –usually a huge space– to optimizing over the shadow value of commodities at each node, a much smaller space. These values characterize the direction and intensity of trades: commodities flow from locations where they have lower shadow value to locations with higher shadow value.[40] Powerful off-the-shelf optimization algorithms are available and used by standard numerical solvers to quickly solve the least-cost route problem (60) and more complex optimal transport problems (Bertsekas (1998) reviews a range of methods).[41]

Least-cost route problems have been introduced in quantitative spatial models to study the effects of specific transportation improvements in a network. Among the best known and earliest studies in this spirit is Donaldson (2018), who studies the impact of 19th century railroad growth in India. He uses a quantitative trade model where iceberg costs are parametrized as function of lowest-cost routes chosen by shippers with access to alternative transport modes in each link, with

[38]This formulation assumes link-level iceberg costs, with the link-level t_{jk} often parametrized as a function of time in transit and other frictions from j to k. An alternative formulation that is also often used defines t_{ij} as an additive cost, with the optimal route minimizing total cost ($r_{ij} = \arg\min_{r \in \mathcal{R}_{ij}} \sum_{(kl) \in r} t_{jk}$) and the "long-distance" iceberg cost τ_{ij} defined as some function of the minimized total cost.

[39]Despite their name, these methods tackle problems that are not limited to physical transportation. They more generally tackle how different entities or agents (like workers and jobs) should be matched in a way that maximizes the net payoffs.

[40]Intuitively, even if no actual trade occurs from the perspective of an agent in the economic model solving the least-cost route problem, the solution to this problem can be interpreted as that of a sequence of competitive agents buying goods in each node and reselling at the connecting nodes that pay the highest value net of the shipping cost.

[41]For example, the "distances" command in Matlab quickly computes least cost routes among sets of origins and destinations on a graph.

mode-specific costs chosen to match observed price gaps. Allen and Arkolakis (2014) simulate the economic impact of the Interstate Highway System in the U.S. by modeling transport costs by car as a function of the shortest distance by road between locations.

5.2 Optimal Transport in a Spatial Equilibrium Model

Armed with the tractability of the least-cost route problem, one could hope to use it to tackle the *optimal network* problem. Broadly speaking, the optimal network problem consists of finding the optimal distribution of link-level transport costs t_{kl} from a feasible set (typically determined by a government budget constraint). However, optimizing directly over t_{kl} is not necessarily tractable because of the dimensionality of the problem (corresponding to the number of links) and the potential non-convexity of a planner's problem over t_{kl}. Fajgelbaum and Schaal (2020) show a way to tackle this optimal network problem. We follow their approach here.

We consider the same spatial equilibrium with multiple goods k and trade frictions as in Section 3.1. Assuming iceberg trade costs at the link level, when $\left(1+t_{ji}^k\right) Q_{ji}^k$ units of product k are shipped from j to a connected neighbor $i \in \mathcal{N}(j)$ then Q_{ji}^k units arrive in i. The only difference with respect to the framework in Section 3.1 is that now the trade cost t_{ij}^k on each link is endogenous. It is decreasing in the corresponding trade flow Q_{ji}^k, capturing congestion (more flows on a link increase travel time given infrastructure), and it is increasing in the amount of infrastructure invested on the link, I_{ji} (more lanes of road decrease travel times given trade flows).

The objective of the planner is to maximize welfare by choosing how to invest in transport infrastructure, I_{ij}. It is instructive to break down the optimization problem in two parts and first treat infrastructure I_{ji} as given. Given infrastructure, the planner chooses the allocation to maximize utility, as summarized in the Lagrangian:

$$\begin{aligned}\mathcal{L} =& u-\sum_j \omega_j\left[u-u_j\left(\left\{c_j^k\right\}, L_j\right)\right]-\sum_j W_j\left[\sum_k N_j-z_j\left(L_j\right) L_j\right]-\Omega\left[\sum_j L_j-L\right] \\ & -\sum_j \sum_k p_j^k\left[c_j^k L_j+\sum_l X_j^{lk}+\sum_{i \in \mathcal{N}(j)}\left(1+t_{ji}^k\left(Q_{ji}^k, I_{ji}\right)\right) Q_{ji}^k-Y_j^k\left(N_j^k,\left\{X_j^{kl}\right\}_l\right)-\sum_{i \in \mathcal{N}(j)} Q_{ij}^k\right].\end{aligned} \tag{61}$$

This problem, without congestion, is the same as in Section 3.1. In that section, we characterized the optimal population allocation given the flows. Here, we characterize in addition the optimal flows entering in the second line. This constraint captures goods availability per node –the "conservation of flows" constraint. A quantity $Y_j^k+\sum_{i \in \mathcal{N}(j)} Q_{ij}^k$ of type k goods (production plus exports to neighbors) is available at j, $c_j^k L_j+\sum_l X_j^{lk}$ is domestically absorbed, and $\sum_{i \in \mathcal{N}(j)} t_{ji} Q_{ji}^k$ is reshipped. The multiplier of this constraint, p_j^k, is the shadow value of good k at j: in an efficient market allocation, this is the price of good k at j.

The sub-problem of finding the optimal Q_{ji}^k conditional on c_j^k, L_j, Y_j^k while treating t_{ji}^k as exogenous is an optimal transport problem, similar to the least-cost route problem discussed above.

Here, however, elementary trade costs t_{ji}^k are endogenous. With congestion, production and consumption feed back into trade costs and impact the choice of a least-cost route.[42] The first-order conditions of (61) with respect to Q_{ji}^k yield:

$$\frac{p_i^k}{p_j^k} \leq 1 + t_{ji}^k + \frac{\partial t_{ji}^k}{\partial Q_{ji}^k} Q_{ji}^k, \tag{62}$$

with equality if $Q_{ji}^k > 0$. This is a no-arbitrage condition: the price differential between a location and its connected neighbors must be less than or equal to the marginal transport cost. From the planner's perspective, this marginal cost takes into account the average cost $1 + t_{ji}^k$ and diminishing returns from congestion. The market allocation may not internalize this congestion, in which case optimal shipping is implemented with congestion taxes. We highlight below some cases in the literature where the market allocation does not satisfy optimal congestion, as well as recent papers focused on implementing optimal congestion taxes. Inverting this conditions yields trade flows Q_{ji}^k as a function of price differentials $\frac{p_i^k}{p_j^k}$.

Solving the problem (61) yields $u^* (\{I_{ji}\})$, the maximal level of utility that can be attained given the infrastructure network I_{ji}. The next step is to solve for the optimal distribution of infrastructure investment I_{ji} given a resource constraint. A natural modeling question at this stage concerns the technology to produce infrastructure. There is no formal obstacle to modeling I_{ji} as the output of a sector that uses fixed factors and tradeables.[43] However, implementation and exposition are simplified when I_{ji} is a homogeneous input that can be only be used for infrastructure, with fixed total endowment K. The cost of building I_{ji} units of infrastructure (e.g., lanes of road) on a link ji is $\delta_{ji}^I I_{ji}$ units of the homogenous input. The parameter δ_{ji}^I captures that building infrastructure is more costly in some locations. The final network optimization problem is then:

$$\max_{\{I_{ji}\}} u^* (\{I_{ji}\}) \text{ subject to} \sum_j \sum_{i \in \mathcal{N}(j)} \delta_{ji}^I I_{ji} \leq K. \tag{63}$$

Using envelope conditions from (61), the first-order condition over I_{ji} yields:

$$\Gamma_{ji}^I \equiv \underbrace{\left(\sum_k p_j^k Q_{ji}^k \right) \left(- \frac{\partial t_{ji}^k}{\partial I_{ji}} \right)}_{\text{Marginal Gain from Infrastructure}} \leq \underbrace{\delta_{ji}^I P_K}_{\text{Marginal Building Cost}}, \tag{64}$$

where P_K is the shadow cost of the budget constraint in (63), corresponding to the price of K in an

[42]The Armington model with product differentiation by origin within sectors, commonly used in quantitative spatial models, is a special case when each commodity k is produced in only one location. In that case, the solution to the least-cost route is implemented independently from the solution to the economic allocation because there is no feedback between optimal transport and economic activity. As a result, the vast majority of quantitative spatial models, even those applied to infrastructure, do not explicitly define the conservation of flows constraint in the second line of 61. Instead, these models are cast with τ_{ij} (the "long-distance" iceberg cost) as a primitive, with this cost already representing the solution to the least-cost route, as we have mentioned. Outside of Armington, when identical goods are produced in multiple locations, the optimal transport problem must solve for the assignment between producers and consumers within a commodity. Recent studies applied to agricultural markets with homogeneous goods, such as in Sotelo (2020) and Bergquist et al. (2022), have this feature. Castro-Vincenzi (2022) solves an optimal transport problem to characterize the assignment of products to export platforms within a firm.

[43]For example, the tradeoff between using land for infrastructure or for consumption is at the heart of early thinking about optimal transport investments; see the discussion in Henderson (1977).

efficient market allocation where production of infrastructure is decentralized and K is commonly owned across individuals. Equation (64) holds with equality on links with actual investment ($I_{ji} > 0$) at the optimum. The marginal gain from infrastructure, Γ^{I}_{ji} in the left-hand side, is a familiar expression in trade models, dating back to at least Fogel (1964): the marginal benefit from a reduction in trade cost equals the value of trade directly exposed to the trade cost reduction, times the trade cost reduction.[44] Similar to Hulten (1978), a technology shock (in this case, to the trading technology) affects aggregate welfare only through the gross value of commodities affected by a shock, and regardless of the input-output structure of the economy.

Substituting the solution for Q^{k}_{ji} as a function of the price differentials p^{n}_{k}/p^{n}_{j} from (62) into (64) implies that the optimal infrastructure I_{ji} between locations j and i is, like the trade flows, only a function of prices these two locations. Expressing both I_{ji} and Q^{k}_{ji} as function of prices allows to compute the solution of the problem using efficient numerical solutions relying on duality, as in as the optimal transport literature. Formally, letting $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$ denote the Lagrangian of the full problem (combining (61) and (63)), where $\mathbf{x}$ denotes the control variables (consumption, population, trade flows, infrastructure) and where $\boldsymbol{\lambda}$ denotes the Lagrange multipliers (i.e., shadow prices) of the combined problem, the dual solves: $\inf_{\boldsymbol{\lambda}\geq \mathbf{0}} \sup_{\mathbf{x}} \mathscr{L}(\mathbf{x}, \boldsymbol{\lambda})$, a minimization over node-level prices subject to non-negativity constraints.

An important caveat, however, is that for this problem to be numerically tractable and guaranteed to yield a global optimum using standard solvers, the full planner's problem must be a convex optimization problem. In addition to the concavity conditions on the spillovers discussed in Sections 2 and 3, convex optimization here also requires the link-level shipping costs $t_{jk}(Q, I)$ to be a convex function of both Q and I. If shipping costs are convex in Q only, the (61) subproblem can be solved using the dual but non-convex optimization methods must be used for (63).

To get an example of a solution to the whole optimal network problem, assume a power function for trade costs:

$$t^{k}_{ji}(Q, I) = \delta^{\tau}_{jik} \frac{Q^{\beta}}{I^{\gamma}}. \tag{65}$$

In this case, $t_{jk}(Q, I)$ is a convex function of both Q and I when $\beta > \gamma$, i.e., when congestion due to transport β offsets the increasing returns inherent to optimizing over the network. The optimal transport problem embedded in (61) is convex as long as there is congestion, $\beta > 0$. To get intuition on the solution, we assume further that only one commodity flows from j to k. Then, (omitting superscripts that denote commodities) the optimal infrastructure investment on ij is

$$I^{*}_{ji} = \kappa_{I} \left[\frac{1}{\delta^{I}_{ji} \left(\delta^{\tau}_{ji}\right)^{\frac{1}{\beta}}} \left(p_{j} \left(\frac{p_{i}}{p_{j}} - 1 \right)^{\frac{1+\beta}{\beta}} \right) \right]^{\frac{\beta}{\beta-\gamma}}, \tag{66}$$

where the κ_{I} is a constant. Condition (66) reveals determinants of optimal investments. Assuming decreasing returns to the transport sector ($\beta > \gamma$), (66) dictates lower investments in locations

[44] This condition hinges on the iceberg trade cost assumption; more generally, the benefits would equal the value of resources used to ship the trade flow affected by the cost reduction.

with greater natural barriers to trade or invest (greater δ^I_{ji} or δ^τ_{ji}). Conditional on these frictions, investments should be larger in links with higher price differentials, as those differentials call for more equilibrium trade, and therefore more benefits to investing. Conditional on price differentials, investments are larger in links departing from more expensive origins, as these are origins where goods are socially more valuable. The condition also implies that the dispersion of infrastructure across space is magnified when $\beta - \gamma$ is low, as in this case the transport sector has weak decreasing returns scale.

5.3 Improving Transport Networks: Quantification Approaches

This section discusses recent research that has quantified the benefits of transportation improvement in complex transport networks. The discussion is broadly structured around methods, and along the way we highlight the more substantive issues addressed in the applications.

5.3.1 Marginal Benefits by Link

A interesting object for policy is the quantification of the marginal benefits of improving different links in the transport network. In the framework above, it is summarized by Γ^I_{ji} in equation (64). A key challenge in measuring Γ^I_{ji} is that trade flows at the level of individual links, $p^k_j Q^k_{ji}$, are not generally observed in standard trade datasets. A possible solution is to use data on vehicle flows as a proxy for trade, as in Allen and Arkolakis (2022) for highways or Heiland et al. (2019) and Do et al. (2024) for sea transport. Another strategy to measure $p^k_j Q^k_{ji}$ is to obtain them from the solution of the transport problem (61) given infrastructure. Fajgelbaum and Schaal (2020) parametrize that problem to match a cross-sectional distribution of economic activity, generating $p^k_j Q^k_{ji}$ as a by-product.

Allen and Arkolakis (2022) calculate the marginal benefit of improving each link of the U.S. Interstate Highway Network using an Armington trade model with labor mobility and external agglomeration and congestion effects, as in Allen and Arkolakis (2014). Exports X_{kl} from a production origin k to a final destination l are done by a continuum of traders that evenly split the value of trade, each one receiving idiosyncratic extreme-value preference shocks over each possible route. As a result every possible route is used, with smaller probability on longer routes. This assumption yields a closed-form solution for the fraction of the trade flow value X_{kl} from k to l that crosses link ji:

$$\frac{p^k_j Q^k_{ji}}{X_{kl}} = \left(\frac{\tau_{kl}}{\tau_{kj} t_{ji} \tau_{il}} \right)^\theta, \tag{67}$$

where the τ_{kl} also is a closed-form function of the full matrix of link-level costs t_{kl}. In addition, congestion at the level of each link is modeled for tractability as function of the value rather the volume of trade through a link. In their implementation, Allen and Arkolakis (2022) represent the U.S. trade network as a graph where the nodes are the major highways intersections, aggregated to broader statistical areas. They use data on traffic between these areas and assume that traffic flows are linear in trade value to get a proxy for link-level trade. They parameterize the trade costs function t^k_{ji} as a function of number of lanes of road and traffic flows, and the costs δ^I_{ji} as a dollar-equivalent cost of observed lane-miles. Given the structure of the model, they can use these measures to compute counterfactual outcomes following specific link-by-link investments. Consistent with the first-order intuition from Γ^I_{ji} in (64), they find the highest-traffic segments to be typically those with higher marginal benefits to invest. They find large variance in returns across link but also, on average, a positive and large rate of return of investing in the U.S. highway system, with benefit-cost ratios well above 1, a result that stands in contrast with calculations from Duranton et al. (2020) using a more aggregate model and different data sources.

Brancaccio et al. (2024) study the welfare impacts of investments in U.S. ports by computing the gains from a small expansion of capacity in each port. The model is different from those in the quantitative spatial literature: it explicitly models the demand and supply of transport systems rather than assuming iceberg trade costs; however, the model only accounts for the partial equilibrium effect of these improvements. Brancaccio et al. (2024) estimate port-level efficiency in supplying port services (measured as vessel waiting time) using a queuing model where waiting times depend on labor and capital inputs at the port. When a port capacity expands, traders reallocate their shipments across ports and the whole distribution of waitings times changes. They find large impacts of port expansion on trade and welfare, with 20% of those benefits stemming from indirect congestion reductions across the system, and positive net returns from expanding capacity at the margin in about one-third of U.S. ports.

5.3.2 Globally Optimal (Road) Networks

Computing the impacts of relatively small investments "link-by-link", as described above, gives a sense of what are the returns to the next dollar injected to the system. However, this approach does not give the solution to the global investment problem (63). It could be that improvements that should be prioritized according to the link-by-link criterion are different from the most relevant ones when considering larger investments, or from investments that impact several segments of the system simultaneously.

Fajgelbaum and Schaal (2020) quantify the global solution to (61) and (63) and apply to the European road network. The nodes are population centroids of one-degree cells and the links connect neighboring cells. The fundamentals of (61) (i.e., productivities and amenities) are parameterized such that the solution to (61) matches population and value added across these cells, given observed I_{ji} measured by the dollar value of existing roads. The link-level trade costs t^k_{ji} match internal trade within a country and the building cost δ^I_{ji} is taken from World Bank road infrastructure projects (Collier et al., 2016). The crucial values of β and γ in (65) are taken from Couture et al. (2018), who estimate the relationship between speed, number of vehicles, and the number of lanes on a road. Assuming that trade costs τ^k_{ji} and traffic are respectively proportional to shipping time and shipments, their estimates imply $\beta > \gamma$, the convex region of the parameter space. Across 24 European countries, Fajgelbaum and Schaal (2020) compute the optimal I_{ij} (a free reallocation of the network) and optimal expansions. The former provides an efficient benchmark against which to gauge the existing network. They find welfare losses of 1.8% on average due to misallocation of roads, a roughly similar magnitude to the gross benefit from optimally expanding current road networks by 50%. Within countries, the optimal network investments tend to reallocate activity in poorer regions, reflecting that aggregate utility increases by reducing differences in marginal utility of consumption across locations.

Gorton and Ianchovichina (2022) and Graff (2024) apply similar methods to quantify optimal road networks across the majority of Latin American countries respectively, with average losses of 1.3% and 1.6% percent. The numbers are not directly comparable across these studies due to differences in implementation details, but the results seem to point to a similar order of magnitude for misallocation across levels of developments, on average. However, the most extremely misallocated countries appear to be in Africa (such as Somalia and Sudan, with losses around 6-7%). African countries also present high returns to an optimal expansion of their road networks (on average, 0.8% welfare gains from a 10% optimal expansion of their road networks). These papers compute optimal networks within countries, which biases towards finding misallocation near borders when routes facilitate international trade. Krantz (2024) explores how optimal road network expansions across the 47 largest cities and ports in Africa depends on the size of the expansion and key model elasticities such as product differentiation and congestion.

5.3.3 Non-Convex Network Optimization

Two features are necessary to make the network investment problem (63) a convex optimization problem. First, it requires strong enough congestion in transport, the condition $\beta > \gamma$ that ensures convexity in specification (65). This condition fails without congestion, i.e. $\beta = 0$, a common assumption in trade models. Second, the problem may also be non-convex if spatial spillovers, of the kind discussed in Section 2, are not internalized by the planner. When the optimization problem is non-convex, standard global search algorithms find local optima but are not guaranteed to identify a global solution. To explore the model properties outside of the convexity region, Fajgelbaum and Schaal (2020) repeat their exercises on European countries assuming that $\beta < \gamma$. Their numerical global search algorithms are still aided by the use of the dual to solve for (61). In this region of the parameter space, the gains from optimal investments and the losses from misallocation are larger (although with a similar ranking across countries), and within-country patterns of optimal investments are more concentrated in fewer links.

Various papers, such as Blouri and Ehrlich (2020), Santamaria (2022), and Bordeu (2023), study such non-convex problems by incorporating market-clearing conditions as additional restrictions when solving (63) rather than solving (61) from a planner problem. An equivalent approach in Fajgelbaum and Schaal (2020) is to solve for (61) using a planner that take agglomeration effects as given. Santamaria (2022) in particular solves for the optimal road network in a spatial equilibrium model with monopolistic competition and trade frictions (Krugman, 1980), which in the context of (61) and (63) corresponds to an Armington trade model with efficiency spillovers and no congestion ($\beta = 0$). She calibrates the initial equilibrium to West Germany in 1950, after the division following the Second World War, and computes the optimal highways for West Germany. She uses optimization algorithms that search locally starting from the initial network available to West German planners right after the division and argues that, despite the global problem being non-convex, starting the algorithm around the observed network in 1950 will likely identify a close local optimum that was reasonably targeted by planners. This approach echoes the idea, in Ahlfeldt et al. (2015), of searching for a new equilibrium after a shock around the initial equilibrium when there are potentially multiple market equilibria. This approach is intuitive but no formal statement rationalizes it, and it implicitly assumes that the dynamic process of network optimization in the real world mimics the iterative steps of the numerical solver.

A third key feature making (63) a convex optimization problem is that the I_{ji} are treated as a continuous variable denoting the intensity of infrastructure investment, rather than as binary or discrete. However, it may be natural to think about whether a particular set of locations or nodes is connected at all. In the parameterization (65), because of Inada conditions, a non-zero level of infrastructure in every link is optimal. Allowing for zero optimal infrastructure on a link necessitates either an alternative functional form than (65), or to model I_{ji} as binary rather than continuous. A binary I_{ji} gives rise to a discrete optimization problem over a huge space (with dimension equal to 2 to a power equal to the number of links) and precludes using standard calculus and continuous numerical solvers. Existing algorithms that help solving discrete choice problems following Tarski (1955) (such as Jia (2008) or Arkolakis and Eckert (2017)) do not readily apply to this case, as they rely on properties of investments across links that do not hold here.

Fajgelbaum et al. (2023), studying California's High Speed Rail, computes the optimal placement of rail stations along railway lines in a framework where travelers may use different modes of transport. Specifically, they optimize over the location of the 24 stations of California's High-Speed Rail. Rather than asking whether each node of the railway line should host one of the 24 stations (a high dimensional combinatorial problem), they parametrize the station placement problem as a continuous choice over geographic coordinates for each station. Their optimization problem remains non-convex, but can be solved using numerical methods with continuous choice. They rely on the fact that transport engineers predetermined the feasible routes linking Northern and Southern California, sidestepping the route design problem in their optimization. In principle, optimization over the geographic coordinates of connected nodes could be used to fully characterize a transport network in more general cases, as long as it is combined with a criterion for the placement of links. Standard cost-minimization criteria (such as least-cost-spanning trees) are good candidates for the latter, as they have been found to rationalize observed transport investments well (see Faber (2014) for China).

Alder (2016) and Kreindler et al. (2023) consider the optimization of transport networks with binary choices on each link. In the context of a gravity trade model, Alder (2016) proposes a heuristic algorithm and applies it to study the efficiency of Indian roads. His algorithm starts from a fully connected network and, at each iteration, removes the least beneficial link or adds the most beneficial one. He shows that an optimal network constrained to connecting middle-size cities, which resembles features of the Chinese road networks, would have outperformed the large road expansions observed in India. Kreindler et al. (2023) solve the problem of designing optimal bus routes and their frequency in a partial equilibrium model with inattentive consumers and applies it to Jakarta's transit system. The planner receives an idiosyncratic preference shock ε_N for each possible network N and chooses $N^*(\varepsilon) = \max_N W(N) + \varepsilon_N$. Standard methods allow them to draw a random sample from the distribution of $N^*(\varepsilon)$. This distribution degenerates to the global optimum that solves $\max_N W(N)$ as the variance of ε_N vanishes (at the cost of increasing computation times). This approach allows them to quickly compute how an expected feature of the optimal network (across realizations of ε) changes to a first order when parameters change. Their method seems particularly fruitful for applications where the goal is to understand this type of comparative statics, which are more costly to compute in the non-probabilistic approach.

5.4 Determinants of Inefficient Investments

The papers we have reviewed here invariantly conclude that observed networks are inefficient, meaning that the observed networks could be reconfigured in welfare-enhancing ways. The literature has put forward various potential explanations for apparent misallocation: path dependency, political preferences, mismeasured investments costs, and decentralized network design.

A natural explanation for the apparent misallocation of transport networks is that economic fundamentals or planners' preferences change at higher frequency than transport networks, which are path-dependent investments. Therefore, networks are designed in a given time period may look inefficient later on. The frameworks we have reviewed can be used to assess this type of path-dependency. Santamaria (2022) calculates that the optimal West German highway network in 1974 would have been quite different from an optimal expansion from 1950 to 1974 (constrained by the pre-division network), resulting in 20% larger optimal gains. Graff (2024) finds that areas in Africa that a century ago received more railroad investments tend to have too many roads today. Santamaria (2022) finds high correlation between actual expansions of German highways after division and model-predicted optimal investments, and Graff (2024) find that the modern road network in Germany has small levels of misallocation. These finding could suggest that German planners are efficient and that the four-fold increase of the road network since reunification has erased the historical legacy. Alternatively, it could be that historical roads were not inefficiently located from the perspective of the reunified country.

A second explanation for the apparent inefficiency of transport investments is that planners' objective functions may not correspond to the national planners maximizing aggregate welfare that is typically assumed. The empirical evidence suggests there is high political discretion behind transportation investments, heightened in non-democratic contexts (Burgess et al., 2015; Voigtländer and Voth, 2014). In democratic environments, transport investments may be geared at maximizing political approval rather than aggregate income. Fajgelbaum et al. (2023) use the case study of California's High Speed Rail to estimate the political preferences of a hypothetical social planner who chose the placement of rail stations. They take advantage of the fact that the project was submitted for approval through a popular vote. Using voting data, they estimate a significant elasticity of local votes to the potential economic impacts of the project, as predicted by a quantitative model of commuting and long-distance travel. Then, they compare the actual design of California's High Speed rail with counterfactual station placements that were not adopted to obtain bounds on the planner's preferences over different demographics and over political approval. Their estimates imply strong planner's preferences for gaining votes. Solving for the counterfactual "apolitical" network, they find that political preferences drove the network design away from large urban centers, where it would have been hard to gain additional votes given the strong political preferences of local voters for the train already. Across Chinese counties, Alder and Kondo (2020) show a correlation between road over-investment calculated using the Alder (2016) framework and political influence proxied by the birthplace of government officials.

A third general reason is misspecification of infrastructure investment costs, which are often modeled in simple ways to enhance tractability. One can always choose costs that rationalize observed networks as being efficient, an approach that Fajgelbaum and Schaal (2020) propose to simulate optimal expansions. The evidence on misallocation usually persists under the range of cost measures used in the literature, including information on actual road projects Fajgelbaum and Schaal (2020), public agency guidelines (Allen and Arkolakis, 2022), or cost function estimation (Brancaccio et al., 2024). Sorin (2024) demonstrates the importance of institutions governing public land acquisitions for infrastructure, such as eminent domain, in driving optimal road investments in a developing country context.

A final possible explanation is that infrastructure misallocation may result from the choice of local rather than central planners. Mohring and Harwitz (1962) show that developers who charge tolls equal to the optimal congestion taxes implement the optimal infrastructure investment, a result that Fajgelbaum and Schaal (2020) generalize to the context of network investments in (63) if developers take prices as given. Outside this market structure, infrastructure investments of competing governments or of developers who can manipulate market outcomes are generically inefficient. Felbermayr and Tarasov (2022) study investments of competing planners in a trade model on a linear geography. Bordeu (2023) studies road investments of neighboring municipalities in Chile, highlighting inefficiencies that arise from the lack of coordination between municipalities, and Chen (2024) embeds an oligopoly model into the optimal transport model to study the investments made by 19th century US railroad companies. Inefficiencies arise as their objective functions may not coincide with welfare.

5.5 Optimal Frequency and Congestion Policies

Rather than studying the optimal spatial distribution of local transport investments, another strand of research studies the allocation of transport systems at more aggregate levels, as well network-wide pricing or tax policies to deal with congestion externalities. These papers often abstract from general equilibrium interactions or optimal routing which are a focus of quantitative spatial models. However, in order to design optimal policies, these papers quantify in greater detail the transport choices made by commuters and the supply of transportation services made by private and public agents. Moreover, they rely on detailed micro datasets to estimate heterogeneity in traveler preferences or in the production function of the transport sector, using techniques from industrial organization.

The value of waiting times and travel times takes center stage in this research. The estimation of the welfare value of time is a key input to compute optimal policies, recognizing that these times are endogenous to the decentralized decisions of both travelers and service providers. Hence, these papers bring to the forefront some classic themes in transport economics, such as travel demand estimation (McFadden, 1974) and scale economies of transport providers (Mohring, 1972), in order to compute optimal investments (for a textbook treatment see Small and Verhoef, 2007).

Almagro et al. (2024) and Barwick et al. (2021) estimate models where households choose their mode of travel in the presence of queuing and congestion. In Barwick et al. (2021), households choose residence and commuting mode. Their location choice affects house prices and commuting speed through congestion. The paper analyzes congestion pricing, driving restrictions and transport expansion in Beijing. In Almagro et al. (2024), household location is fixed and households choose their travel mode subject to congestion externalities. They characterize the optimal frequency and pricing of public transit options in Chicago. Conwell (2023) estimates optimal subsidies to suppliers of minibuses who do not internalize their impacts on waiting times. In these frameworks, congestion externalities are a function of city aggregates or numbers of commuters by residence-workplace pair, hence optimal congestion taxes can be computed using the approach we laid out in Section 3.5. Hierons (2024) uses optimal transport tools to incorporate the network structure of the transport network, so that congestion is defined on a link of a transport network and depends on gross travelers through that link. Brancaccio et al. (2023) studies optimal subsidy schemes in decentralized transport markets with search externalities, and Kreindler (2024) studies effectiveness of peak-hour congestion pricing.

These papers complement the quantitative spatial frameworks we have described in both method and focus. There could be exciting research avenues from further combining the tools and questions. For instance, network interactions can impact congestion policies; these policies may be more effective when targeted to specific segments of the network. Kreindler (2024) combines elements from both strands by solving the problem of optimal routes and service frequency in a context where passengers care differentially about travel and waiting times. Likewise, spatial models could benefit from the richer consumer choice features of this strand research. Fajgelbaum et al. (2023) go in this direction by introducing modal choice and estimating the heterogeneous preferences of different demographics groups over modes of travel. The spatial distribution of these demographic groups ends up being an important determinant of optimal network design.

5.6 Future Research

This section reviewed a recent body of work that models and quantifies optimal transport investments. These models can be useful to provide efficient benchmarks to measure potential misallocation of actual transport networks and their expansion. The analysis and quantification of optimal transport models can also provide insights on pressing policy tradeoffs. Using statistics on transportation spending and usage in the U.S., Duranton et al. (2020) argue that large-scale expansions of the country's transport infrastructure may not be fully warranted, and that reallocating investments across different modes of transport could offer higher returns. The methods review here can help inform this type of debate.

A number of important themes are so far absent from the literature. Transport investments are highly persistent, yet the current methods ignore dynamic tradeoffs and uncertainty. Sunk investments that will unfold slowly over time do not necessarily determine optimal networks that look in the steady state like the static optimal networks. In a growing economy, optimal investments should foresee the importance of certain regions in the future. Many current policy debates concern resilience of supply chains to unexpected shocks and ways to protect against increasing climate hazards. Arguably, investment in transport networks should trade-off the reasons to invest that maximize static welfare against ways to minimize impacts of future risk. These topics have so far not been explored.

The literature also does not pay much attention to the idea of optimal transport investments geared towards development. A classic idea from new economic geography models is that regional development is highly uneven due to agglomeration, with trade costs largely determining whether the economy operates in a multiplicity region. The local returns to transport investments in this context may be highly convex. A related issue concerns the complementarity between the optimal investments we have studied in this section and the optimal transfers we studied in previous sections. Some of the approaches we have reviewed simultaneously consider optimal transport investments and optimal spatial transfers, while others assume that the former are implemented but not the latter. A better understanding is lacking of the different prescriptions of these two approaches, and of when they are more appropriate to guide optimal investments.

In more practical terms, the applications in the literature focus overwhelmingly on road investments or urban policies concerning road transport, with few cases studying railroads or ports. Analysis of air transport and, more generally, analyses of optimal investments that account for inter-modality are also lacking.

6 Conclusion

In this chapter we discussed optimal spatial policies in spatial equilibrium models. Our emphasis has been on general equilibrium: determining whether a policy that affects a specific location is optimal cannot be done in isolation by focusing only on the returns to that location. Instead, an equilibrium model is required to assess the shadow cost of resources reallocated to a given region, i.e., to measure the next-best use of these resources in the rest of the economy. Empirical studies that demonstrate positive net returns to a policy do not fully determine whether the policy is desirable unless they account for this shadow cost. We showed how to conduct this measurement in standard models. Our analysis focused on three topics. First, we reviewed place-specific policies such as taxes and subsidies aimed at reallocating workers to address spillovers from urban agglomeration. This was done across a range of models, including canonical urban frameworks and more complex quantitative spatial models. Second, we examined spatial transfers in contexts without spillovers, where the market allocation is Pareto efficient, but where social preferences for redistribution justify spatial redistribution even in the presence of personal income taxes. Third, we discussed optimal transport investments using methods that account for the optimal routing and network structure of the economy.

Our analysis was limited to topics which have received relatively more attention in recent years. Current work is deploying related quantitative spatial models, often in combination with elements from other fields in economics such as industrial organization and development economics, to quantify policies across a range of pressing issues that feature spatial tradeoffs. Recent topics include the allocation of medical facilities in contexts where accessing these facilities is costly (Agte, 2024; Dingel et al., 2023; Hsiao, 2022); the location of schools and educational investments in contexts where the access is costly and the returns to these investments depend on student composition (Agostinelli et al., 2024; Eckert and Kleineberg, 2024); policies that affect exposure to environmental risk, such as zoning laws (Ospital, 2022), building regulations (Ostriker and Russo, 2022), local taxes (Aronoff and Rafey, 2023), or transport investments (Balboni, 2019); local carbon taxes (Conte et al., 2022); investments in renewable technologies (Arkolakis and Walsh, 2023; Colas and Salumier, 2023); or lockdown policies (Fajgelbaum et al., 2021; Giannone et al., 2022). Many studies limit the analysis to the impacts of specific policies that have been implemented or that are in the process of being implemented. The tools described in this chapter could be used to further advance the analysis of these pressing issues from a normative perspective.

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.

Adao, R., A. Costinot, and D. Donaldson (2017). Nonparametric counterfactual predictions in neoclassical models of international trade. *American Economic Review 107*(3), 633–89.

Agostinelli, F., M. Luflade, and P. Martellini (2024). On the spatial determinants of educational access. Technical report, National Bureau of Economic Research.

Agte, P. (2024). Fighting silent killers: How india public healthcare staffing expansion saves lives by improving access and market quality. Technical report, Technical report.

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

Akerlof, G. A. (1978). The economics of "tagging" as applied to the optimal income tax, welfare programs, and manpower planning. *The American economic review 68*(1), 8–19.

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

Albouy, D. (2012). Evaluating the efficiency and equity of federal fiscal equalization. *Journal of Public Economics 96*(9-10), 824–839.

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

Alder, S. (2016). Chinese roads in india: The effect of transport infrastructure on economic development. *Available at SSRN 2856050*.

Alder, S. and I. Kondo (2020). Political distortions and infrastructure networks in china: A quantitative spatial equilibrium analysis.

Ales, L. and C. Sleet (2022). Optimal taxation of income-generating choice. *Econometrica 90*(5), 2397–2436.

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

Allen, T. and C. Arkolakis (2022). The welfare effects of transportation infrastructure improvements. *The Review of Economic Studies 89*(6), 2911–2957.

Allen, T. and D. Donaldson (2020). Persistence and path dependence in the spatial economy. Technical report, National Bureau of Economic Research.

Almagro, M., F. Barbieri, J. C. Castillo, N. G. Hickok, and T. Salz (2024). Optimal urban transportation policy: Evidence from chicago. Technical report, National Bureau of Economic Research.

Arkolakis, C. and F. Eckert (2017). Combinatorial discrete choice. *Available at SSRN 3455353*.

Arkolakis, C. and C. Walsh (2023). Clean growth. Technical report, National Bureau of Economic Research.

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

Aronoff, D. and W. Rafey (2023). Conservation priorities and environmental offsets: Markets for florida wetlands. Technical report, National Bureau of Economic Research.

Atalay, E., A. Hortaçsu, M. Runyun, C. Syverson, and M. F. Ulu (2023). Micro-and macroeconomic impacts of a place-based industrial policy. Technical report, National Bureau of Economic Research.

Atkin, D., M. K. Chen, and A. Popov (2022). The returns to face-to-face interactions: Knowledge spillovers in silicon valley. Technical report, National Bureau of Economic Research.

Atkin, D. and D. Donaldson (2015). *Who's Getting Globalized?: The Size and Implications of Intra-national Trade Costs*. Number w21439. National Bureau of Economic Research Cambridge, MA.

Atkinson, A. B. and J. E. Stiglitz (1976). The design of tax structure: direct versus indirect taxation. *Journal of public Economics 6*(1-2), 55–75.

Au, C.-C. and J. V. Henderson (2006). Are chinese cities too small? *The Review of Economic Studies 73*(3), 549–576.

Autor, D. H. (2019). Work of the past, work of the future. In *AEA Papers and Proceedings*, Volume 109, pp. 1–32. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Balboni, C. A. (2019). *In harm s way? infrastructure investments and the persistence of coastal cities*. Ph. D. thesis, London School of Economics and Political Science.

Bartelme, D. G., A. Costinot, D. Donaldson, and A. Rodriguez-Clare (2019). The textbook case for industrial policy: Theory meets data. Technical report, National Bureau of Economic Research.

Barwick, P. J., S. Li, A. R. Waxman, J. Wu, and T. Xia (2021). Efficiency and equity impacts of urban transportation policies with equilibrium sorting. Technical report, National Bureau of Economic Research.

Behrens, K., G. Mion, Y. Murata, and J. Suedekum (2017). Spatial frictions. *Journal of Urban Economics 97*, 40–70.

Bergquist, L. F., B. Faber, T. Fally, M. Hoelzlein, E. Miguel, and A. Rodríguez-Clare (2022). Scaling agricultural policy interventions. Technical report, National Bureau of Economic Research.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica*.

Bertsekas, D. P. (1998). *Network optimization: continuous and discrete models*. Citeseer.

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

Bilal, A. (2023). The geography of unemployment. *The Quarterly Journal of Economics 138*(3), 1507–1576.

Blouri, Y. and M. v. Ehrlich (2020). On the optimal design of place-based policies: A structural evaluation of eu regional transfers. *Journal of international economics 125*, 103319.

Bordeu, O. (2023). Commuting infrastructure in fragmented cities. *Job Market Paper, University of Chicago Booth School of Business*.

Bourguignon, F. and A. Spadaro (2012). Tax–benefit revealed social preferences. *The Journal of Economic Inequality 10*, 75–108.

Brancaccio, G., M. Kalouptsidi, and T. Papageorgiou (2024). Investment in infrastructure and trade: The case of ports. Technical report, National Bureau of Economic Research.

Brancaccio, G., M. Kalouptsidi, T. Papageorgiou, and N. Rosaia (2023). Search frictions and efficiency in decentralized transport markets. *The Quarterly Journal of Economics 138*(4), 2451–2503.

Bryan, G. and M. Morten (2019). The aggregate productivity effects of internal migration: Evidence from indonesia. *Journal of Political Economy 127*(5), 2229–2268.

Burgess, R., R. Jedwab, E. Miguel, A. Morjaria, and G. Padró i Miquel (2015). The value of democracy: evidence from road building in kenya. *American Economic Review 105*(6), 1817–51.

Caliendo, L., M. Dvorkin, and F. Parro (2019). Trade and labor market dynamics: General equilibrium analysis of the china trade shock. *Econometrica 87*(3), 741–835.

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

Castro-Vincenzi, J. (2022). Climate hazards and resilience in the global car industry. *Princeton University manuscript*.

Chen, Y. (2024). Network structure and the efficiency gains from mergers: Evidence from us freight railroads. *Available at SSRN 4742990*.

Chetty, R. (2006, December). A new method of estimating risk aversion. *American Economic Review 96*(5), 1821–1834.

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

Colas, M. and K. Hutchinson (2021). Heterogeneous workers and federal income taxes in a spatial equilibrium. *American Economic Journal: Economic Policy 13*(2), 100–134.

Colas, M. and E. Salumier (2023). Optimal subsidies for residential solar. Technical report, working paper.

Collier, P., M. Kirchberger, and M. Söderbom (2016). The cost of road infrastructure in low-and middle-income countries. *The World Bank Economic Review 30*(3), 522–548.

Conte, B., K. Desmet, and E. Rossi-Hansberg (2022). On the geographic implications of carbon taxes. Technical report, National Bureau of Economic Research.

Conwell, L. (2023). Subways or minibuses? privatized provision of public transit. Technical report.

Couture, V., J. Dingel, A. Green, and J. Handbury (2024). Demographic preferences and income segregation. *University of Pennsylvania, mimeograph*.

Couture, V., G. Duranton, and M. A. Turner (2018). Speed. *Review of Economics and Statistics 100*(4), 725–739.

D'Amico, L. (2021). Place based policies with local voting: Lessons from the eu cohesion policy. *Available at SSRN 3866098*.

Davis, M. and J. M. Gregory (2021). Place-based redistribution in location choice models. NBER Working Paper 29045, 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., D. K. Nagy, and E. Rossi-Hansberg (2018). The geography of development. *Journal of Political Economy 126*(3), 903–983.

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.

Diamond, R. and C. Gaubert (2022). Spatial sorting and inequality. *Annual Review of Economics 14*(1), 795–819.

Dingel, J. I., J. D. Gottlieb, M. Lozinski, and P. Mourot (2023). Market size and trade in medical services. Technical report, National Bureau of Economic Research.

Dingel, J. I. and F. Tintelnot (2020). Spatial economics for granular settings. Technical report, National Bureau of Economic Research.

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

Do, A., S. Ganapati, W. F. Wong, and O. Ziv (2024). Transshipment hubs, trade, and supply chains.

Donald, E., M. Fukui, and Y. Miyauchi (2023). Optimal dynamic spatial policy. Technical report, Mimeo, Boston University.

Donaldson, D. (2018). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review 108*(4-5), 899–934.

Duranton, G., G. Nagpal, and M. Turner (2020). Transportation infrastructure in the us. Technical report, National Bureau of Economic Research.

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

Eckert, F. and T. Kleineberg (2024). The geography of opportunity: Education, work, and intergenerational mobility across us counties. Technical report, Working Paper.

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

Ehrlich, M. v. and H. G. Overman (2020, August). Place-based policies and spatial disparities across european cities. *Journal of Economic Perspectives 34*(3), 128–49.

Faber, B. (2014). Trade integration, market size, and industrialization: evidence from china's national trunk highway system. *Review of Economic Studies 81*(3), 1046–1070.

Fajgelbaum, P. D. and C. Gaubert (2020). Optimal spatial policies, geography, and sorting. *The Quarterly Journal of Economics 135*(2), 959–1036.

Fajgelbaum, P. D., C. Gaubert, N. Gorton, E. Morales, and E. Schaal (2023). Political preferences and the spatial distribution of infrastructure: Evidence from california's high-speed rail. Technical report, National Bureau of Economic Research.

Fajgelbaum, P. D., A. Khandelwal, W. Kim, C. Mantovani, and E. Schaal (2021). Optimal lockdown in a commuting network. *American Economic Review: Insights 3*(4), 503–522.

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

Fajgelbaum, P. D. and E. Schaal (2020). Optimal transport networks in spatial equilibrium. *Econometrica 88*(4), 1411–1452.

Felbermayr, G. J. and A. Tarasov (2022). Trade and the spatial distribution of transport infrastructure. *Journal of Urban Economics 130*, 103473.

Ferrari, A. and R. Ossa (2023). A quantitative analysis of subsidy competition in the us. *Journal of Public Economics 224*, 104919.

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

Fogel, R. (1964). Railroads and american economic growth.

Franco, S. (2024). *Output market power and spatial misallocation.* The University of Chicago.

Fujita, M. (2013). Urban land use theory. In *Regional and Urban Economics Parts 1 & 2*, pp. 111–187. Routledge.

Galichon, A. (2016). *Optimal Transport Methods in Economics.* Princeton University Press.

Gaubert, C. (2018). Firm sorting and agglomeration. *American Economic Review 108*(11), 3117–3153.

Gaubert, C., P. M. Kline, D. Vergara, and D. Yagan (2024). Place-based redistribution. Technical report, National Bureau of Economic Research.

Giannone, E., N. Paixão, and X. Pang (2022). Jue insight: The geography of pandemic containment. *Journal of Urban Economics 127*, 103373.

Glaeser, E. L. (2008). *Cities, agglomeration, and spatial equilibrium.* Oxford University Press.

Glaeser, E. L. (2013). Chapter 4 - urban public finance. In A. J. Auerbach, R. Chetty, M. Feldstein, and E. Saez (Eds.), *handbook of public economics, vol. 5*, Volume 5 of *Handbook of Public Economics*, pp. 195–256. Elsevier.

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

Gorton, N. and E. Ianchovichina (2022). Trade networks in latin america: Spatial inefficiencies and optimal expansions. *Available at SSRN 4041108*.

Graff, T. (2024). Spatial inefficiencies in africa trade network. *Journal of Development Economics*.

Greaney, B. (2023). Housing constraints and spatial misallocation: Comment. *University of Washington, mimeograph*.

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

Heiland, I., A. Moxnes, K. H. Ulltveit-Moe, and Y. Zi (2019). Trade from space: Shipping networks and the global implications of local shocks.

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

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

Helsley, R. W. (2004). Chapter 54 - urban political economics. In J. V. Henderson and J.-F. Thisse (Eds.), *Cities and Geography*, Volume 4 of *Handbook of Regional and Urban Economics*, pp. 2381–2421. Elsevier.

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

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

Henderson, J. V. (1977). *Economic theory and the cities*. Academic Press.

Henkel, M., T. Seidel, and J. Suedekum (2021, November). Fiscal transfers in the spatial economy. *American Economic Journal: Economic Policy 13*(4), 433–68.

Hierons, T. (2024). Spreading the jam: Optimal congestion pricing in general equilibrium. *Job Market Paper, University of Chicago*.

Hong, G. (2024). Two-sided sorting of workers and firms: Implications for spatial inequality and welfare. Technical report, Canadian Labour Economics Forum (CLEF), University of Waterloo.

Hsiao, A. (2022). Democratization and infrastructure investment: Evidence from healthcare in indonesia.

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

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

Hulten, C. R. (1978). Growth accounting with intermediate inputs. *The Review of Economic Studies 45*(3), 511–518.

Incoronato, L. and S. Lattanzio (2023). Place-based industrial policies and local agglomeration in the long run.

Jia, P. (2008). What happens when wal-mart comes to town: An empirical analysis of the discount retailing industry. *Econometrica 76*(6), 1263–1316.

Kim, W. (2023). Migration, commuting, and the spatial distribution of public spending. Technical report.

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

Krantz, S. (2024). Optimal investments in africa's road network. Technical report, Kiel Working Paper.

Kreindler, G. (2024). Peak-hour road congestion pricing: Experimental evidence and equilibrium implications. *Econometrica 92*(4), 1233–1268.

Kreindler, G., A. Gaduh, T. Graff, R. Hanna, and B. A. Olken (2023). Optimal public transportation networks: Evidence from the world's largest bus rapid transit system in jakarta. Technical report, National Bureau of Economic Research.

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

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy 99*(3), 483–499.

Lashkaripour, A. and V. Lugovskyy (2023). Profits, scale economies, and the gains from trade and industrial policy. *American Economic Review 113*(10), 2759–2808.

McFadden, D. (1974). The measurement of urban travel demand. *Journal of public economics 3*(4), 303–328.

McFadden, D. (1978). Modeling the choice of residential location. In A. Karlqvist, L. Lundqvist, L. Snickars, and J. Weibull (Eds.), *Spatial Intearction Theory and Planning Models*, pp. 531–551. North Holland.

McFadden, D. (1980). Econometric models for probabilistic choice among products. *Journal of Business*, S13–S29.

Mirrlees, J. A. (1972). The optimum town. *The Swedish Journal of Economics*, 114–135.

Mohring, H. (1972). Optimization and scale economies in urban bus transportation. *The American Economic Review 62*(4), 591–604.

Mohring, H. and M. Harwitz (1962). Highway benefits: An analytical framework. Technical report, Northwestern University.

Mongey, S. and M. E. Waugh (2024). Discrete choice, complete markets, and equilibrium. NBER Working Paper 32135, National Bureau of Economic Research.

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

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

Nagy, D. K. (2023). Hinterlands, city formation and growth: Evidence from the us westward expansion. *Review of Economic Studies 90*(6), 3238–3281.

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

Oh, R. (2023). Spatial sorting of workers and firms. Technical report, Technical report.

Ospital, A. (2022). Urban policy and spatial exposure to environemntal risk. *Unpublished Manuscript*.

Ostriker, A. and A. Russo (2022). The effects of floodplain regulation on housing markets. *URL https://ostriker. github. io/papers/Ostriker-Russo floodplain-regulations. pdf*.

Piketty, T. and E. Saez (2013). Optimal labor income taxation. In *Handbook of public economics*, Volume 5, pp. 391–474. Elsevier.

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.

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

Rossi-Hansberg, E., P.-D. Sarte, and F. Schwartzman (2019). Cognitive hubs and spatial redistribution. Technical report, National Bureau of Economic Research.

Saez, E. (2002). The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes. *Journal of Public Economics 83*(2), 217–230.

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

Santamaria, M. (2022). Reshaping infrastructure: Evidence from the division of germany.

Slattery, C. (2024). Bidding for firms: Subsidy competition in the us. *Available at SSRN 3250356*.

Slattery, C. and O. Zidar (2020). Evaluating state and local business incentives. *Journal of Economic Perspectives 34*(2), 90–118.

Small, K. A. and E. T. Verhoef (2007). *The economics of urban transportation.* Routledge.

Sorin, J. (2024). Public roads on private lands: Land costs and optimal road improvements in urban uganda. *Job Market Paper, University of Chicago*.

Sotelo, S. (2020). Domestic trade frictions and agriculture. *Journal of Political Economy 128*(7), 2690–2738.

Tabuchi, T. and J.-F. Thisse (2002). Taste heterogeneity, labor mobility and economic geography. *Journal of Development Economics 69*(1), 155–177.

Tarski, A. (1955). A lattice-theoretical fixpoint theorem and its applications.

Tombe, T. and X. Zhu (2019). Trade, migration, and productivity: A quantitative analysis of china. *American Economic Review 109*(5), 1843–1872.

UrbanInstitute (2024). State and local finance data.

Voigtländer, N. and H.-J. Voth (2014). Highway to hitler. *NBER Working Paper* (w20150).

Wildasin, D. (1986a). *Urban public finance*. Routledge.

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

Wildasin, D. E. (1986b). Spatial variation of the marginal utility of income and unequal treatment of equals. *Journal of Urban Economics 19*(1), 125–129.

Wildasin, D. E. (1987). Theoretical analysis of local public economics. In *Handbook of regional and urban economics*, Volume 2, pp. 1131–1178. Elsevier.

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

Wu, W. and W. You (2023). Should governments promote or restrain urbanization? a quantitative analysis of the internal migration restrictions in china. *A Quantitative Analysis of the Internal Migration Restrictions in China*.

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

A Appendix to Section 2 (Efficient Spatial Allocation)

A.1 Allocations are characterized by $\{c_j\}$

From (3), population in a location is an implicit function of consumption in that location and utility, such that $u_j\left(c_j, \frac{H_j}{L_j}, L_j\right) = u$. Call it $\tilde{L}_j(c_j, u)$. Replacing $\tilde{L}_j(c_j, u)$ in (4), and further using $h_j = \frac{H_j}{L_j}$ and $N_j = z_j(L_j)L_j$ from (5) and (6) gives an implicit solution for utility as function of the distribution of consumption that we denote $\tilde{u}(\{c_j\})$. Attainable levels of utility in this economy are then given by any combination of $\{c_j\}$ consistent with aggregate labor market clearing (7), $\sum_i \tilde{L}_j(c_j, \tilde{u}(\{c_j\})) = L$.

A.2 Appendix to (2.3) (Welfare-Increasing Reallocations)

Consider a perturbation (dc_j, dL_j) around a market allocation. Using the feasibility condition (4) with equality yields:

$$\sum_j dc_j L_j + \sum_j c_j dL_j = \sum_j Y_j'(z_j(L_j)L_j)\left[z_j'(L_j)L_j + z_j(L_j)\right]dL_j. \tag{A.1}$$

Using the market equilibrium conditions (8) and (9),

$$\sum_j dc_j L_j + \sum_j c_j dL_j = \sum_j \left(\varepsilon_{z_j,L_j} + 1\right) w_j dL_j. \tag{A.2}$$

Using the spatial mobility condition (1) with equality, the consumer's problem (13) implies that indirect utility is equalized in space across populated locations:

$$v_j(x_j, R_j, L_j) = u. \tag{A.3}$$

Totally differentiating this condition and using Roy's identity and the individual budget constraint yields:

$$dc_j = \frac{x_j}{\varepsilon_{v_j,x_j}}\frac{du}{u} - \frac{\varepsilon_{v_j,L_j}}{\varepsilon_{v_j,x_j}}\frac{dL_j}{L_j}x_j + (x_j - c_j)\frac{dL_j}{L_j} \tag{A.4}$$

Combining (A.2) and (A.4) yields:

$$\frac{du}{u} = \frac{\sum_j \left(\varepsilon_{z_j,L_j} w_j + \frac{\varepsilon_{v_j,L_j}}{\varepsilon_{v_j,x_j}} x_j - x_j + w_j\right) dL_j}{\sum_j \frac{x_j L_j}{\varepsilon_{v_j,x_j}}}.$$

Defining $\Gamma_j \equiv \varepsilon_{z_j,L_j} w_j + \frac{\varepsilon_{v_j,L_j}}{\varepsilon_{v_j,x_j}} x_j$ and $\chi \equiv \left[\sum_j \frac{L_j}{\partial v_j/\partial x_j}\right]^{-1}$ yields:

$$\frac{du}{\chi} = \sum_j (\Gamma_j - x_j + w_j)\, dL_j. \tag{A.5}$$

Suppose now that we evaluate (A.5) at $dL_j = -dL_i = dL_0$ and $dL_n = 0$ for $n \neq i, j$. Then $du > 0$ whenever (18) holds.

A.3 Appendix to 2.4.2 (Link to Henry George Theorem)

The equilibrium is stable if $\frac{\partial u_j}{\partial L_j} \leq 0$. In this derivation, we do not assume that $u_j = u$ is maintained following a local population shock. Instead, we assume that, in this perturbation, net transfers between locations are unchanged. However, production, the housing market and the corresponding consumption adjusts internally in j . A change in local population leads to:

$$du_j = \frac{\partial u_j}{\partial c_j} dc_j + \frac{\partial u_j}{\partial h_j} dh_j + \frac{\partial u_j}{\partial L_j} dL_j$$

where, assuming Y_j has constant returns in N_j, $\frac{\partial c_j}{\partial L_j} = \frac{\partial w_j}{\partial L_j} = \frac{w_j}{L_j} \varepsilon_{z_j,L_j}$ (assuming that local changes in production are consumed locally) while $\frac{\partial h_j}{\partial L_j} = -\frac{h_j}{L_j}$ (by housing market clearing in j). Using the first order conditions of the consumer problem and rearranging yields:

$$\begin{aligned} \frac{x_j}{\varepsilon_{v_j,x_j}} \frac{du_j}{u_j} &= \left(w_j \varepsilon_{z_j,L_j} - R_j h_j + \frac{\varepsilon_{v_j,L_j}}{\varepsilon_{v_j,x_j}} x_j \right) \frac{dL_j}{L_j} \\ &= \left(\Gamma_j - R_j h_j \right) \frac{dL_j}{L_j} \end{aligned}$$

Stability of the equilibrium yields

$$\Gamma_j - R_j h_j \leq 0, \tag{A.6}$$

and when the equilibrium is efficient, $\sum_j \frac{L_j}{L} \Gamma_j = \Gamma$ so that summing over (A.6) for all j yields:

$$\Gamma - \pi \leq 0 \tag{A.7}$$

in the stable efficient equilibrium.

A.4 Planner's Problem in Baseline model

The efficient allocation has $\{c_j, h_j, L_j, N_j, u\}$ that deliver maximal u in the feasible set (i.e. satisfying (3)-(7)). The planner's Lagrangian is thus:

$$\begin{aligned} \mathcal{L} =& u - \sum_j \omega_j \left[u - u_j \left(c_j, h_j, L_j \right) \right] - P^* \left[\sum_j c_j L_j - \sum_j Y_j \left(N_j \right) \right] \\ & - \sum_j R_j^* \left[h_j L_j - H_j \right] - \sum_j W_j^* \left[N_j - z_j \left(L_j \right) L_j \right] - \Omega \left[\sum_j L_j - L \right] \end{aligned} \tag{A.8}$$

We wrote the Lagrangian such that, if the problem is quasi-concave and first-order conditions are sufficient, all the multipliers on inequality constraints are positive. In an efficient allocation, the multipliers P, W_j, and R_j of (A.8) are the equilibrium prices for tradable goods, nontradables in j, and efficiency units in j, respectively. Without loss of generality, one can normalize $P^* = 1$ in the planner's problem by choice of units of u. The multiplier Ω is the opportunity cost of a worker in any location. It is also the impact of an adding a worker to the overall economy on the average utility of incumbents ($\frac{du}{dL}$) at the efficient allocation.

A necessary condition for the allocation to be efficient(sufficient when the planner's problem is concave, which we assume here) is that the first-order conditions with respect to consumption and employment by location hold:

$$[\partial c_j] \qquad \omega_j \frac{\partial u_j}{\partial c_j} = P^* L_j \tag{A.9}$$

$$[\partial h_j] \qquad \omega_j \frac{\partial u_j}{\partial h_j} = R_j^* L_j \tag{A.10}$$

$$[\partial L_j] \qquad \omega_j \frac{\partial u_j}{\partial L_j} - P^* c_j - R_j^* h_j + W_j^* z_j \left(1 + \varepsilon_{z_j, L_j}\right) - \Omega = 0. \tag{A.11}$$

Assume that the allocation can be implemented in a market setting with appropriate labor subsidies - a conjecture that can be verified ex post. In a market allocation, consumers maximize utility subject to their budget constraint (problem (13)). The first-order conditions of that problem hold, and in particular $\frac{\partial v_j}{\partial x_j}$ equals the multiplier of this budget constraint, so that:

$$\frac{\partial u_j}{\partial c_j} = \frac{\partial v_j}{\partial x_j} \tag{A.12}$$

$$\frac{\partial u_j}{\partial h_j} = \frac{\partial v_j}{\partial x_j} R_j \tag{A.13}$$

If the market allocation with subsidies is efficient, then (A.9) and (A.12) hold so that:

$$\omega_j = \frac{L_j}{\left(\frac{\partial v_j}{\partial x_j}\right)}, \tag{A.14}$$

In addition, (A.10) and (A.13) hold leading to $R_j = R_j^*$. Finally, in the market allocation $\frac{\partial u_j}{\partial L_j} = \frac{\partial v_j}{\partial L_j}$ and wage is given by (9) . Combining these conditions with (A.11) yields:

$$x_j = \frac{\varepsilon_{v_j, L_j}}{\varepsilon_{v_j, x_j}} x_j + w_j \left(1 + \varepsilon_{z_j, L_j}\right) - \Omega \tag{A.15}$$

Combined with (11),

$$s_j = \varepsilon_{z_j, L_j} w_j + \frac{\varepsilon_{v_j, L_j}}{\varepsilon_{v_j, x_j}} x_j - (\pi + \Omega). \tag{A.16}$$

The optimal subsidy is thus (21) with $\Gamma = \pi + \Omega$.

B Appendix to Section 3 (Extensions and Implementation)

B.1 Planner Problem with Public Good Provision

The planner now chooses g_j in addition to $\{c_j, h_j, L_j, N_j, u\}$. Her maximization problem is:

$$\begin{aligned}\mathcal{L} =& u - \sum_j \omega_j \left[u - u_j\left(c_j, h_j, g_j, L_j\right)\right] - P^* \left[\sum_j c_j L_j + G\left(g_j, L_j\right) - \sum_j Y_j\left(N_j\right)\right] \\ & - \sum_j R_j^* \left[h_j L_j - H_j\right] - \sum_j W_j^* \left[N_j - z_j\left(L_j\right) L_j\right] - \Omega \left[\sum_j L_j - L\right]\end{aligned} \tag{A.17}$$

The first-order conditions with respect to local consumptions are still (A.9) and (A.10). The first-order condition with respect to public good consumption and local population change to:

$$[\partial g_j] \qquad \omega_j \frac{\partial u_j}{\partial g_j} = P^* \frac{\partial G}{\partial g_j} \tag{A.18}$$

$$[\partial L_j] \qquad \omega_j \frac{\partial u_j}{\partial L_j} - P^* c_j - R_j^* h_j - P^* \frac{\partial G}{\partial L_j} + W_j^* z_j \left(1 + \varepsilon_{z_j, L_j}\right) - \Omega = 0 \tag{A.19}$$

Solving as in the baseline case using (A.9) and (A.10), we get the optimality condition for population distribution:

$$x_j = \frac{\varepsilon_{v_j, L_j}}{\varepsilon_{v_j, x_j}} x_j + w_j + w_j \varepsilon_{z_j, L_j} - \frac{\partial G}{\partial L_j} - \Omega \tag{A.20}$$

Comparing to the decentralized allocation with a local tax $\mu_j = \frac{G_j}{L_j}$,such that $x_j = w_j + \pi - \mu_j + s_j$, yields the optimal subsidy:

$$s_j = \frac{\varepsilon_{v_j, L_j}}{\varepsilon_{v_j, x_j}} x_j + \varepsilon_{z_j, L_j} w_j - \frac{\partial G}{\partial L_j} + \frac{G_j}{L_j} + s_0 \tag{A.21}$$

where s_0 ensures government budget balance, i.e. $s_0 = -\Gamma + \frac{\sum L_j \frac{\partial G}{\partial L_j}}{\sum L_j} - \sum G_j$.

B.2 Heterogeneous Workers

The planner is utilitarian within type and has Pareto weights λ^θ across types. She maximizes $\sum_\theta \lambda^\theta u^\theta$ under feasibility constraint by choosing $\left\{c_j^\theta, L_j^\theta, N_j, u^\theta\right\}\}$. The Lagrangian is:

$$\mathcal{L} = \sum_\theta \lambda^\theta u^\theta - \sum_\theta \sum_j \omega_j^\theta \left[u^\theta - u_j^\theta \left(c_j^\theta, h_j^\theta, \left\{L_j^{\theta'}\right\}_{\theta'}\right)\right] - P^* \left[\sum_\theta \sum_j c_j^\theta L_j^\theta - \sum_j Y_j\left(N_j\right)\right]$$
$$- \sum_j R_j^* \left[\sum_\theta h_j^\theta L_j^\theta - H_j\right] - \sum_j W_j^* \left[N_j - N\left(z_j^1 L_j^1, .., z_j^\Theta L_j^\Theta\right)\right] - \sum_\theta \Omega^\theta \left[\sum_j L_j^\theta - L^\theta\right] \tag{A.22}$$

The first order condition with respect to population of each type in each location yields:

$$\left[\partial L_j^\theta\right] \qquad \sum_{\theta' \in \Theta} \omega_j^{\theta'} \frac{\partial u_j^{\theta'}}{\partial L_j^\theta} - P^* c_j^\theta - R_j^* h_j^\theta + W_j^* \frac{dN_j}{dL_j^\theta} - \Omega^\theta = 0 \tag{A.23}$$

The next steps are similar to those described in A.4, where the first order conditions for consumption are applied type by type. The optimal allocation satisfies:

$$x_j^\theta = w_j^\theta + \sum_{\theta' \in \Theta} x_j^{\theta'} \frac{\varepsilon_{v_j^{\theta'}, L_j^\theta}}{\varepsilon_{v_j^{\theta'}, x_j^{\theta'}}} \frac{L_j^{\theta'}}{L_j^\theta} + \sum_{\theta' \in \Theta} w_j^{\theta'} \varepsilon_{z_j^\theta, L_j^{\theta'}} \frac{L_j^{\theta'}}{L_j^\theta} - \Omega^\theta,$$

for some Ω^θ that measures how much the planner would value an additional θ-type worker in the economy. The relative valuation of different types depends in particular directly on their relative welfare weights λ^θ. The optimal subsidy (36) follows from the consumer's budget constraint in the market allocation, $x_j^\theta = w_j^\theta + \pi^\theta + s_j^\theta$.

B.3 Cross-Region Spillovers

With cross-region spillovers, expression (A.11) in the baseline case becomes

$$[\partial L_j] \qquad \sum_k \omega_k \frac{\partial u_k}{\partial L_j} - P^* c_j - R_j^* h_j + W_j^* z_j + \sum_k W_k^* \frac{L_k}{L_j} z_k \varepsilon_{z_k, L_j} - \Omega = 0.$$

Using the same steps, and in particular plugging in (A.9), yields the following characterization of an efficient allocation:

$$x_j = \sum_k \frac{\varepsilon_{v_k, L_j}}{\varepsilon_{v_k, x_k}} \frac{L_k}{L_j} x_k + w_j \left(1 + \sum_k \frac{w_k L_k}{L_j} \varepsilon_{z_k, L_j}\right) - \Omega.$$

B.4 Commuting

The planner chooses a feasible allocation $\{c_{jk}, L_{jk}, u_{jk}\}$ to maximize welfare. Her constrained maximization problem corresponds to maximizing the Lagrangian:

$$\begin{aligned}\mathcal{L} =& u - \sum_j \omega_{jk} \left[u - u_{jk}\left(c_{jk}, h_{jk}, \sum_l L_{jl}, \delta_{jk}\right)\right] - P^* \left[\sum_{jk} c_{jk} L_{jk} - \sum_j Y_j\left(N_j\right)\right] \\ &- \sum_j R_j^* \left[\sum_k L_{jk} h_{jk} - H_j\right] - \sum_j W_j^* \left[N_j - z_j\left(\sum_i L_{ij}\right) \sum_i L_{ij}\right] - \Omega \left[\sum_{jk} L_{jk} - L\right]\end{aligned} \tag{A.24}$$

The first-order condition on population allocation, writing $\sum_l L_{jl} = L_j$, is:

$$[\partial L_{jk}] \qquad \sum_l \omega_{jl} \frac{\partial u_{jl}}{\partial L_j} + \omega_{jk} \frac{\partial u_{jk}}{\partial \tau_{jk}} \frac{\partial \delta_{jk}}{\partial L_{jk}} - R_{jk} h_{jk} - c_{jk} + W_k^* z_k \left(1 + \varepsilon_{z_k, L_k}\right) - \Omega = 0$$

Therefore:

$$s_{jk} = \underbrace{\sum_l \frac{\varepsilon_{v_{jl}, L_j}}{\varepsilon_{v_{jl}, x_{jl}}} \frac{L_{jl} x_{jl}}{L_j}}_{\text{residence}-j \text{ specific}} + \frac{\varepsilon_{v_{jk}, \delta_{jk}} \varepsilon_{\delta_{jk}, L_{jk}}}{\varepsilon_{v_{jk}, x_{jk}}} x_{jk} + \underbrace{w_k \varepsilon_{z_k, L_k}}_{\text{workplace-}k\text{ specific}} - (\Omega + \pi)$$

In the constant elasticity example 3, this simplifies to:

$$s_{jk} = \sum_l \gamma_A \frac{L_{jl} x_{jl}}{L_j} - \gamma_T x_{jk} + w_k \gamma_z - \Omega,$$

hence plugging in $x_{jk} = w_k + s_{jk}$ yields (37).

C Appendix to Section 4 (Redistribution)

The social welfare function is:

$$W = \sum_\theta \mu^\omega \underbrace{\max_j \left\{u_j\left(c_j\left(\theta\right)\right) + \varepsilon_j^\omega\right\}}_{\equiv W(\theta)}$$

with $c_j(\theta) = w_j(\theta) - t(w_j(\theta)) + s_j$. The planner's problem is summarized by the Lagrangian:

$$\mathcal{L} = W - \chi\left(\sum_\theta \sum_j s_j L_j(\theta) - \sum_\theta \sum_j t(w_j(\theta)) L_j(\theta)\right)$$

The introduction of a place-based subsidy to j, ds_j, starting from a pure place-blind redistribution system $t(.)$, yields a welfare change of:

$$\frac{dW}{ds_j} = \frac{\partial W}{\partial s_j} - \chi\left(\sum_\theta \sum_i t(w_i(\theta)) \frac{\partial L_i(\theta)}{\partial c_j(\theta)}\right).$$

If Pareto weights depend only on skill θ but not on preference shocks ε ($\mu^\omega = \mu^\theta$), then:

$$\frac{\partial W}{\partial s_j} = \sum_\theta \mu^\theta \frac{\partial u_j(\theta)}{\partial c_j(\theta)} L_j(\theta).$$

This last equality uses that $\frac{\partial W}{\partial s_j} = \sum_\theta \mu^\theta \frac{\partial W(\theta)}{\partial u_i(\theta)} \frac{\partial u_j(\theta)}{\partial s_j(\theta)}$ and the Williams-Daly-Zachary theorem (McFadden, 1978, 1980), which yields that for such a discrete choice model, $\frac{\partial W(\theta)}{\partial u_i(\theta)} = L_i(\theta)$. Defining

$$\bar{\lambda}_j \equiv \frac{1}{\chi}\frac{\partial W}{\partial s_j} \tag{A.25}$$

and manipulating the equation further yields:

$$\frac{1}{L_j}\frac{dW}{ds_j} = \chi\left(\bar{\lambda}_j - 1 - \frac{\sum_\theta \sum_i t(w_i(\theta)) \frac{\partial L_i(\theta)}{\partial c_j(\theta)}}{L_j}\right)$$

Consider a country with two locations $\{i, j\}$, with $L_i + L_j = L$, hence $dL_i + dL_j = 0$. Implementing a small place-based subsidy $ds_j = \frac{ds}{L_j}$ in j financed by a lump-sum tax $ds_i = \frac{ds}{L_i}$ elsewhere leads to the welfare change

$$dW = \frac{dW}{ds_j} ds_j + \frac{dW}{ds_i} ds_i,$$

which after some algebra yields (54).

To characterize the optimal $\{s_j\}$ given $t(.)$, take the first order condition of the planner problem (53):

$$\frac{\partial \mathcal{L}}{\partial s_j} = \frac{\partial W}{\partial s_j} - \chi L_j - \chi \sum_\theta \sum_i (s_i - t(w_i(\theta))) \frac{\partial L_i(\theta)}{\partial c_j(\theta)} = 0.$$

Rearranging and using (A.25) yields expression (56).