NBER WORKING PAPER SERIES

OPTIMAL TRANSPORT NETWORKS IN SPATIAL EQUILIBRIUM

Pablo D. Fajgelbaum
Edouard Schaal

Working Paper 23200
http://www.nber.org/papers/w23200

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2017, Revised March 2017

We thank Thomas Chaney, Arnaud Costinot, Don Davis, Jonathan Eaton, Cecile Gaubert, Gueorgui Kambourov, Ezra Oberfield, Stephen Redding, Esteban Rossi-Hansberg and Jonathan Vogel for helpful comments. We thank seminar participants at Chicago Booth, CEMFI, Chinese University of Hong Kong, Columbia, Dartmouth, Humboldt, MIT, National University of Singapore, Penn State, Princeton, Ridge, Science Po, SED, SMU, Toronto, Universidad de Barcelona, USC, and West Coast Trade Workshop. Edouard Schaal acknowledges financial support from the Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2015-0563) and the Barcelona GSE (grant SG-2017-07). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Optimal Transport Networks in Spatial Equilibrium
Pablo D. Fajgelbaum and Edouard Schaal
NBER Working Paper No. 23200
February 2017, Revised March 2017
JEL No. F11,O18,R13

ABSTRACT

We study optimal transport networks in spatial equilibrium. We develop a framework consisting of a neoclassical trade model with labor mobility in which locations are arranged on a graph. Goods must be shipped through linked locations, and transport costs depend on congestion and on the infrastructure in each link, giving rise to an optimal transport problem in general equilibrium. The optimal transport network is the solution to a social planner's problem of building infrastructure in each link. We provide conditions such that this problem is globally convex, guaranteeing its numerical tractability. We also study cases with increasing returns to transport technologies in which global convexity fails. We apply the framework to assess optimal investments and inefficiencies in observed road networks in 25 European countries. The counterfactuals suggest larger gains from road network expansion and larger losses from misallocation of current roads in lower-income countries.

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

Edouard Schaal
Centre de Recerca en Economia Internacional
Universitat Pompeu Fabra
Ramon Trias Fargas, 23-25
08005 Barcelona
Spain
Spain
edouard.schaal@gmail.com

1 Introduction

Trade costs are a ubiquitous force in international trade and economic geography, as they rationalize spatial distributions of prices, real incomes, and trade flows. In reality, trade costs result from a diverse set of policies and frictions in the economic environment. Transport infrastructure, in particular, stands out as an important force (Limao and Venables, 2001; Atkin and Donaldson, 2015).[1] Following the pioneering work of Eaton and Kortum (2002), a standard approach to study the gains from market integration, and of infrastructure improvements in particular, is to fit a quantitative trade model to data on the geographic distribution of economic activity, and then ask what would happen if trade costs between specific locations were to change by some predetermined amount.[2]

A range of questions related to transport infrastructure calls for a somewhat different approach. Consider, for example, a problem confronted by countries at the time of allocating resources: how should infrastructure investments be allocated across regions, and how do the aggregate gains depend on the magnitude of the total investment? Relatedly, infrastructure investments may be sensitive to frictions, local interests or corruption, potentially leading to suboptimal transport networks that may hinder trade and development.[3] How important are these inefficiencies? To answer these questions, it is necessary to pinpoint the best set of infrastructure investments, to then ask what would happen if trade costs were to change in the way implied by the efficient transport network.

In this paper, we develop and apply a framework to study optimal transport networks in general equilibrium spatial models. We solve a global optimization over the space of networks, given any primitive fundamentals, in a general neoclassical framework. In contrast to the standard approach, here trade costs are an outcome rather than a primitive, endogenously responding to fundamentals such as resource endowments and geographic frictions through optimal investments in the transport network. We apply the framework to European road networks, where we assess the aggregate and regional impacts of optimal infrastructure growth, the inefficiencies of observed networks, and the optimal placement of roads as a function of observable regional characteristics.

The point of departure for the framework is a neoclassical economy with multiple goods, factors, and locations, nesting standard trade models (such as the Ricardian, Armington, and factor-endowment models) and allowing for either a fixed spatial distribution of the primary factors (as in

[1]For a review of various determinants of trade costs see Anderson and Van Wincoop (2004).

[2]Costinot and Rodríguez-Clare (2013) review the quantitative gravity literature on changes in trade costs focused on measuring gains from international trade. Redding and Rossi-Hansberg (2016) review a body of research using similar frameworks to study counterfactuals involving changes in infrastructure within countries. See Donaldson (2015) and Redding and Turner (2015) for reviews of empirical analyses of actual changes in transport infrastructure, as well as the literature review below for additional references.

[3]See WorldBank (2011) and IADB (2013) for assessments of transport costs and infrastructure in Africa and Latin-America, respectively. Collier et al. (2016) provide evidence that the costs of building road networks in low- and middle-income countries are related to political conflict. Such inefficiencies are not only apparent in developing countries; e.g., Castells and Solé-Ollé (2005) consider the role of political factors in driving the allocation of infrastructure investment across Spanish departments.

international trade models) or for labor to be mobile (as in economic geography models).[4] The key methodological innovation is that locations are arranged on a graph and goods can only be shipped through connected locations subject to transport costs that depend both on how much is shipped (e.g., because of congestion or decreasing returns to shipping technologies) and on how much is invested in infrastructure (e.g., the number of lanes or the quality of the road). We tackle the planner's problem of simultaneously choosing the transport network (i.e., the set of infrastructure investments), the allocation of production and consumption, and the gross trade flows across the graph.

Solving this problem may be challenging because of dimensionality—the space of all networks is large—and interactions—an investment in one link asymmetrically impacts the returns to investments across the network. It is also complicated by the potential presence of increasing returns due to the complementarity between infrastructure investments and shipping. We exploit the fact that the planner's subproblem of choosing gross trade flows is an optimal flow problem on a network, a well understood problem in the operations research and optimal transport literatures. A key insight from these literatures is that the optimal flows derive from a "potential field"—prices in our context—that can be efficiently solved numerically using duality techniques.[5] We make assumptions such that the full planner's problem, involving the general equilibrium allocation and the network investments alongside the optimal transport, inherits the tractability of optimal flow problems. Our assumptions, including a continuous mapping from infrastructure investments to trade costs and curvature in the technology to transport goods, ensure that the full planner's problem is convex, and that the set of optimal infrastructure investments can be expressed as a function of equilibrium prices. As a result, we solve the full planner's problem while avoiding a direct search in the space of networks. Instead, we search in the space of equilibrium prices applying the numerical methods typically used for optimal transport problems.

While strong enough congestion in transport guarantees convexity of the planner's problem, our framework can also be used when congestion is weak or absent—a case that implies increasing returns in the overall transport technology. We numerically approximate the global solution in non-convex cases by combining the duality approach to obtain the optimal flows and infrastructure as a function of prices with global-search numerical methods that build upon standard simulated annealing techniques. Even though in non-convex cases we only find local optima, the ensuing networks display the qualitative features that we expect in the presence of economies of scale. In particular, in a simple case with a single commodity, we demonstrate that the optimal network is a "tree" under increasing returns: every pair of locations is necessarily connected by only one route when convexity fails, but generically connected by multiple routes if it holds. In tune with this feature, our numerical solutions in more complex environments with multiple commodities show that the network is sparser in the region of the parameter space where convexity fails: the distribution of infrastructure investments is more concentrated in fewer links and includes a larger

[4] We limit the analysis to transport of goods. In the case with labor mobility, labor is perfectly mobile.
[5] See the references in the literature review.

amount of zeros.

The framework has enough flexibility to be matched to real-world data and then used to undertake counterfactuals involving the optimal transport network. The quantification relies on two steps. First, the model's fundamentals can be calibrated such that the solution to the planner's optimal allocation of consumption, production, and gross flows matches spatially disaggregated data on economic activity given an observed transport network. This step is enabled by the fact that, given the transport network, the welfare theorems hold. Second, assuming a specific technology to build infrastructure makes it possible to undertake counterfactuals involving the optimal network.

We apply these steps in the context of European road networks. For the quantification, we allow locations in the model to be heterogeneous in productivity and in the supply of non-traded goods. We discipline these fundamentals such that, given the observed road networks, the model reproduces the observed population and value added at a 0.5 x 0.5 degree spatial resolution (approximately 50km x 50km cells) across the 25 European countries in our data. For this step we construct a measure of the road infrastructure (number of lanes and type of roads) linking any two contiguous cells in the data, and we entertain different assumptions on labor mobility and on the returns to infrastructure, encompassing both convex and non-convex cases. Then, we impose alternative assumptions on road building costs. We either assume that the observed road network is the outcome of the full planning problem—allowing us to back out these costs from the first-order conditions of the planner's problem—or use existing estimates from the literature for how building costs vary with observable geographic features.

Our counterfactuals in the benchmark parametrization with convex costs imply that, across countries, the average welfare gain from an optimal 50% expansion in the resources used to build the observed road networks and the average welfare loss from road misallocation are between 3% and 6%, depending on the assumptions on building costs and labor mobility. Regardless of these assumptions, we find larger returns to optimal road expansion and larger losses from road misallocation in poorer economies. Within countries, the optimal expansion or reallocation of roads reduces regional inequalities in real consumption, reflecting the fact that the goal of optimal infrastructure investments is to reduce dispersion in the marginal utility of consumption of traded commodities. However, different assumptions on building costs and returns to scale imply different ways of achieving this goal of reducing spatial inequalities by changing the optimal placement of infrastructure. We illustrate the alternative road investment plans implied by the different assumptions and counterfactuals by considering two of the largest economies in our data, France and Spain.

The rest of the paper proceeds as follows. Section 2 discusses the connection to the literature. Section 3 develops the framework, establishes its key properties, and discusses the numerical implementation. Section 4 presents simple illustrative examples. Section 5 applies the model to road networks in Europe. Section 6 concludes. We relegate proofs, additional derivations, details of the quantitative exercise, tables, and figures to the appendix.

2 Relation to the Literature

Our paper is related to a recent quantitative literature in international trade and spatial economics that studies the role of trade costs in rich geographic settings. Eaton and Kortum (2002) and Anderson and Van Wincoop (2003) developed quantitative versions of the Ricardian and Armington trade models, respectively, allowing counterfactuals with respect to trade costs in a multi-country competitive equilibrium.

Some recent studies introduce traders who choose the least cost route to ship their goods within a given transport network. These studies undertake counterfactuals with respect to the cost of shipping across specific links, but do not optimize in the space of networks like we do.[6] In this vein, Allen and Arkolakis (2014) measure the aggregate effect of the U.S. highway system, Donaldson and Hornbeck (2016) calculate the historical impact of railroads on the U.S. economy, and Redding (2016) compares the impact of infrastructure changes in models with varying degrees of increasing returns. Alder (2016) simulates counterfactual transport networks in India, Nagy (2016) studies how the development of U.S. railways affected city formation, and Sotelo (2016) simulates the impact of highway investments on agricultural productivity in Peru. Other recent studies allowing for factor mobility and trade frictions within countries include Bartelme (2015), Caliendo et al. (2014) and Ramondo et al. (2012).[7]

To the best of our knowledge, only a few papers feature some form of search or optimization over transport networks: Alder (2016) applies a heuristic algorithm that progressively eliminates links according to their impact on market access and Felbermayr and Tarasov (2015) study optimal infrastructure investments by competing planners in an Armington model where locations are arranged on a line.[8] Allen and Arkolakis (2016) compute the first-order welfare impact of reductions to the cost of shipping across specific links in an Armington model, but do not optimize over the space of networks.[9] In contrast, we solve a global optimization over the space of networks in a neoclassical framework with or without labor mobility.[10] The fundamentals can be chosen to match data on economic activity and actual transport networks at high spatial resolution, as in our application to Europe.[11]

Both our model and the studies cited above include an optimal transport problem, defined as

[6] Chaney (2014a) studies endogenous networks of traders in contexts with imperfect information. For a review of recent literature on the role of various types of networks in international trade see Chaney (2014b).

[7] Redding et al. (2016) study innovations to urban transport systems and apply their analysis to Berlin.

[8] Some recent studies allow for endogenous transport costs in different historical contexts: Swisher IV (2015) model U.S. transport investments as the result of a Nash Equilibrium across competing companies in the context and Trew (2016) endogeneizes trade costs in the spatial-development framework of Desmet and Rossi-Hansberg (2014) by making them depend on the amount of activity in a location, and studies the role of transport infrastructure in shaping structural change in England and Wales.

[9] Allen et al. (2014) apply related envelope conditions to compute the maximal welfare gradient with respect to local changes in trade costs in gravity models without optimizing over the transport network.

[10] The model also nests the spatial equilibrium model of Rosen-Roback (Roback, 1982).

[11] The studies computing the gains from reducing trade costs or improving infrastructure rarely account for the costs of doing so, while we consider both sides of the trade-off by including a cost of building infrastructure in each link. In our counterfactuals, the parametrization of these costs has important implications for where optimal infrastructure is placed.

the trader's problem of choosing least-cost routes across pairs of locations.[12] However, in the studies cited above, the optimal transport problem does not include congestion and can therefore be solved independently from any general-equilibrium outcome. In addition, in these previous studies, each location sources each good from only one origin, as in the Armington model where each commodity is produced in only one location.[13] In contrast, the solution to the optimal transport problem in this paper depends on the solution to the general equilibrium of the neoclassical allocation problem, and markets may source the same good from different locations. The least-cost route optimization present in the applications of the gravity trade models discussed before corresponds to the solution of our optimal transport problem in the special case in which there is no congestion.

As mentioned in the introduction, the planner's subproblem of choosing how to ship goods given demand, supply and infrastructure formally defines a type of optimal transport problem. Optimal transport problems were studied early on by Monge (1781) and Kantorovich (1942).[14] More specifically, because we analyze the optimal route problem instead of just the direct assignment of sources to destinations, our approach is more closely related to optimal flow problems on a network as studied in Chapter 8 of Galichon (2016) and Chapter 4 of Santambrogio (2015).[15] However, our problem differs from this literature in two important aspects. First, in our model, consumption and production in every location are endogenous because they respond to standard general-equilibrium forces. Instead, the aforementioned optimal flows problems are concerned with mapping sources with fixed supply to sinks with fixed demand.[16] Second, our ultimate focus is on the optimization over the transport network itself and in the application of the model to optimal network investments in the presence of general-equilibrium forces, whereas this literature usually takes the transport costs between links as a primitive. In that regard, the problem that we study is akin to the optimal transport network problems in non-economic environments analyzed in Bernot et al. (2009).

Despite these differences, our model inherits key appealing properties of optimal transport problems. While the optimal transport literature shows that strong duality holds under weak conditions in a wide variety of environments, it holds under some conditions in our model as a special case of convex duality. Hence, our way of embedding an optimal transport problem into a general neoclassical equilibrium model extended with a network design problem does not preclude the validity of key earlier insights from the optimal transport literature. The main benefit of duality, in our context, is a reduction of the search space and substantial gains in computation times.[17]

[12] Note that "optimal transport" refers to the optimal shipping of goods throughout the network. This is one of the subproblems embedded in our framework, alongside the optimal network design problem.

[13] An exception is Sotelo (2016), who models a factor-endowment economy where different locations may produce the same agricultural good.

[14] See Villani (2003) for a textbook treatment of the subject.

[15] See also Bertsekas (1998) for a survey of algorithms and numerical methods for optimal flow and transport problems on a network.

[16] See Beckmann (1952) for an early continuous-space example of such an optimal transport problem in economics. See also Carlier (2010) and Ekeland (2010) for introductory lecture notes to the mathematical theory of optimal transport and its connection to economics.

[17] Our paper also relates to the network-design and planning literature in operations research, which studies related network-design problems in telecommunications and transport industries without embedding them in general-

A large body of empirical research estimates how actual changes in transport costs impact economic activity. For instance, Fernald (1999) estimates the impact of road expansion on productivity across U.S. industries; Chandra and Thompson (2000), Baum-Snow (2007) and Duranton et al. (2014) estimate the impact of the U.S. highways on various regional economic outcomes; Donaldson (2010) estimates the impact of access to railways in India; and Faber (2014) estimates the impact of connecting regions to the expressway system in China.[18] Our application measures the aggregate country-level welfare gains from optimally expanding current road networks. In the counterfactuals, we inspect the relationship between infrastructure investment and growth across regions. Feyrer (2009) and Pascali (2014) assess how the arrival of new transport technologies impacted countries or cities whose geographic position made them differentially likely to use the new transport mode. In Section 4 we illustrate how our model could be used to determine the impact of new transport technologies operating through the optimal investments reshaping the network.

Finally, we also apply the model to measure the potential losses from misallocation of current roads. In that sense, this paper is broadly related to the literature on the aggregate effects of misallocation such as Restuccia and Rogerson (2008) and Hsieh and Klenow (2009). Recent papers such as Desmet and Rossi-Hansberg (2013), Brandt et al. (2013), and, more recently, Hsieh and Moretti (2015) and Fajgelbaum et al. (2015), specifically focus on misallocation across geographic units. Asturias et al. (2016) study how transport infrastructure impacts misallocation in a model where misallocation is endogenous through variable markups. In our case, the counterfactuals study the inefficient placement of roads in space from the perspective of a welfare-maximizing central planner.

3 Model

3.1 Environment

Preferences The economy consists of a discrete set of locations $\mathcal{J} = \{1, .., J\}$. We let L_j be the number of workers located in $j \in \mathcal{J}$, and L be the total number of workers. We will entertain cases with labor mobility, where L_j is determined endogenously, and cases without mobility, where L_j is given. Workers consume a bundle of traded goods and a non-traded good in fixed supply, such as land or housing. Utility of an individual worker who consumes c units of the traded goods bundle and h units of the non-traded good is

$$u = U(c, h), \tag{1}$$

where the utility function U is homothetic and concave in both of its arguments.[19]

equilibrium spatial models. See Ahuja et al. (1989) for a handbook treatment of the subject.

[18] See also Coşar and Demir (2016) and Martincus et al. (2017) for empirical studies of how infrastructure investments impact international shipments.

[19] Except when noted explicitly, we do not impose the Inada condition. The utility function could also vary by location to encompass cases where they vary in how attractive they are, e.g., because of amenities.

In location j, per-capita consumption of traded goods is

$$c_j = \frac{C_j}{L_j},$$

where C_j is the aggregate supply of the traded goods bundle in location j. There is a discrete set of tradable sectors $n = 1, .., N$, combined into C_j through a homogeneous of degree 1 and concave aggregator,

$$C_j = C_j^T \left(C_j^1, \ldots, C_j^N\right) \tag{2}$$

where C_j^n is the total quantity of sector n's output consumed in location j. The typically assumed CES aggregator is a special case of this technology.[20]

Production The supply-side of the economy corresponds to a general neoclassical economy. In addition to labor, there is a fixed supply $\mathbf{V}_j = \left(V_j^1, \ldots, V_j^M\right)'$ of primary factors $m = 1, .., M$ in location j. These factors are immobile across regions but mobile across sectors. The production process may also use goods from other sectors as intermediate inputs. Output of sector n in location j is:

$$Y_j^n = F_j^n \left(L_j^n, \mathbf{V}_j^n, \mathbf{X}_j^n\right), \tag{3}$$

where L_j^n is the number of workers, $\mathbf{V}_j^n = \left(V_j^{1n}, \ldots, V_j^{Mn}\right)'$ is the quantity of other primary factors, and $\mathbf{X}_j^n = \left(X_j^{1n}, \ldots, X_j^{Nn}\right)$ is the quantity of each sector's output allocated to the production of sector n in location j. The production function F_j^n is either neoclassical (constant returns to scale, increasing and concave in all its arguments) or a constant (endowment economy). Therefore, the production structure encompasses the neoclassical trade models. The Armington model (Anderson and Van Wincoop, 2003) corresponds to $N = J$ (as many sectors as regions) and $F_j^n = 0$ for $n \neq j$, so that Y_j^j is region j's output in the differentiated commodity that (only) region j provides. The Ricardian model corresponds to labor as the only factor of production and linear technologies, $Y_j^n = z_j^n L_j^n$. The specific-factors and Hecksher-Ohlin models are also special cases of this production structure.

Underlying Graph The locations $\mathcal{J}$ are arranged on an undirected graph $(\mathcal{J}, \mathcal{E})$, where $\mathcal{E}$ denotes the set of edges (i.e., unordered pairs of $\mathcal{J}$). For each location j there is a set $\mathcal{N}(j)$ of connected locations, or neighbors. Goods can be shipped only through connected locations; i.e., goods shipped from j can be sent to any $k \in \mathcal{N}(j)$, but to reach any $k' \notin \mathcal{N}(j)$ they must transit through a sequence of connected locations. The transport network design problem will consist of determining the level of infrastructure linking each pair of connected locations.

A natural case encompassed by this setup corresponds to j being a geographic unit such as county, $\mathcal{N}(j)$ being its bordering counties, and shipments being done by land. More generally,

[20] Under the CES assumption with elasticity of substitution σ, since we only require C_j^T to be concave, our formulation allows the traded sectors to be either complements ($\sigma < 1$) or substitutes ($\sigma > 1$).

neighbors in our theory do not need to be geographically contiguous, since it could be possible to ship directly between geographically distant locations by land, air or sea. The fully connected case in which every location may ship directly to every other location, $\mathcal{N}(j) = \mathcal{J}$ for all j, is one special case.

Transport Technology In the model, goods will typically transit through several locations before reaching a point where they are consumed or used as intermediate input. We let Q^n_{jk} be the quantity of goods in sector n shipped from j to $k \in \mathcal{N}(j)$, regardless of where the good was produced.[21] Transporting Q^n_{jk} from j to k requires τ^n_{jk} units of the good n itself, where τ^n_{jk} denotes the per-unit cost of transporting good n from j to k. The term $1 + \tau^n_{jk}$ corresponds to the iceberg cost typically considered in the literature. Here, the per-unit cost τ^n_{jk} may depend on the quantity shipped, Q^n_{jk}, and on the level of infrastructure I_{jk} along link jk through the following transport technology:

$$\tau^n_{jk} = \tau_{jk}\left(Q^n_{jk}, I_{jk}\right), \tag{4}$$

where

$$\frac{\partial \tau_{jk}}{\partial Q^n_{jk}} \geq 0. \tag{5}$$

This assumption allows for decreasing returns in the shipping sector. We refer to these decreasing returns as congestion, with the understanding that this concept encapsulates several real-world forces whereby an increase in shipping activity leads to higher marginal transport costs. These forces include increased road use, as well as the fact that the transport sector may operate subject to decreasing returns to scale due to land-intensive fixed factors such as warehousing or specialized physical and human capital.[22] In short, the more is shipped, the higher the per-unit shipping cost. While the transport technology (4) assumes that the per-unit cost for commodity n depends on the quantity shipped of commodity n only, the framework can accommodate congestion externalities across goods, as we show in Section 3.6.

We interpret I_{jk} as capturing features that lead to reductions in the cost of transporting goods. For example, when shipping over land, I_{jk} may correspond to whether a road linking j and k is paved, its number of lanes or the availability of roadside services. Hence, we assume:

$$\frac{\partial \tau_{jk}}{\partial I_{jk}} \leqslant 0.$$

We adopt the conventions that, in the absence of infrastructure, transport along jk is prohibitively costly, $\tau_{jk}(Q_{jk}, 0) = \infty$, and that only when infrastructure goes to infinity is there free transport,

[21] We adopt the convention that $\mathcal{N}(j)$ does not include j, i.e., j is not defined as a neighbor of itself.

[22] The Handbook on Estimation of External Costs in the Transport Sector commissioned by the European Commission (Maibach et al., 2013) lists higher travel times, higher accident rate, and road damage as reasons why increased road use may impact transport costs. Other social costs include environmental damage and noise.

$\tau_{jk}(Q_{jk}, \infty) = 0$.[23]

The transport technology $\tau_{jk}(\cdot)$ is allowed to vary by jk, denoting that shipping along some links may be more costly than along others for the same quantity shipped and infrastructure. This variation may reflect geographic characteristics such as distance or ruggedness. The per-unit cost function $\tau_{jk}(Q, I)$ may also depend on the direction of the flow; e.g., if elevation is higher in j than k and it is cheaper drive downhill then $\tau_{jk}(Q, I) > \tau_{kj}(Q, I)$.

Flow Constraint In every location there may be tradable commodities being produced, as well as coming in or out. The balance of these flows requires that, for all locations $j = 1, .., J$ and commodities $n = 1, .., N$:

$$\underbrace{C_j^n + \sum_{n'} X_j^{nn'} + \sum_{k \in \mathcal{N}(j)} \left(1 + \tau_{jk}^n\right) Q_{jk}^n}_{\text{Consumption +Intermediate Use+ Exports}} \quad \leq \quad \underbrace{Y_j^n + \sum_{i \in \mathcal{N}(j)} Q_{ij}^n}_{\text{Production + Imports}} \quad . \tag{6}$$

The left-hand side of this inequality is location j's consumption C_j^n of good n, intermediate-input use $X_j^{nn'}$ by each sector n', and exports to neighbors Q_{jk}^n. These flows are bounded by the local production Y_j^n and imports from neighbors Q_{ij}^n.[24]

We let P_j^n be the multiplier of this constraint. This multiplier reflects society's valuation of a marginal unit of good n in location j. In the decentralized allocation, this multiplier will equal the price of good n in location j; therefore, we simply refer to P_j^n as the price of good n in location j.

Network Building Technology We define the transport network as the distribution of infrastructure $\{I_{jk}\}_{j \in \mathcal{J}, k \in \mathcal{N}(j)}$. The network-design problem will determine this distribution. For simplicity, we assume that building infrastructure requires a mobile resource such as "concrete" or "asphalt", in fixed aggregate supply K, which cannot be used for other purposes. This assumption represents a situation where society has sunk an amount of resources into network-building, but must still decide how to allocate these resources across different places. At the time of characterizing the planner's problem, it will lead to the intuitive property that the opportunity cost of building infrastructure in any location is simply foregoing infrastructure elsewhere. Section 3.6 discusses how to endogenize the supply of infrastructure.

Importantly, the cost of setting up infrastructure may vary across links jk. Specifically, building a level of infrastructure I_{jk} on the link jk requires an investment of $\delta_{jk}^I I_{jk}$ units of K. The network-building constraint therefore is:

$$\sum_j \sum_{k \in \mathcal{N}(j)} \delta_{jk}^I I_{jk} = K. \tag{7}$$

While both the transport technology $\tau_{jk}(Q, I)$ in (4) and the infrastructure building cost δ_{jk}^I

[23] The locations $k \notin \mathcal{N}(j)$ unconnected to j can be equivalently modeled as connected locations for which $\tau_{jk}(Q, I) = \infty$ for all Q and I.

[24] In standard minimum-cost flow problems this restriction is referred to as "conservation of flows constraint". E.g., see Bertsekas (1998) and Chapter 8 of Galichon (2016).

in (7) vary across links according to similar geographic features, each type of variation reflects conceptually different forces that will manifest themselves differently in the data at the time of the quantitative application. Variation in the transport technology $\tau_{jk}(Q, I)$ by jk given Q and I captures how features of the terrain impact per-unit shipping costs given quantity shipped and infrastructure, whereas δ^I_{jk} captures the trade-off, in terms of real resources, between setting up a given level of infrastructure in one link versus another. Importantly, in the planner's problem below, δ^I_{jk} will not impact the allocation other than through infrastructure I_{jk}.

We allow the network-design problem to take place when some infrastructure $\underline{I}_{jk}$ is already in place, and we also allow (but do not require) an upper bound $\overline{I}_{jk}$ to how much can be built in each link, possibly representing geographic constraints on the capacity to build on a specific link. Assuming that existing infrastructure cannot be reallocated implies the constraints:

$$0 \leq \underline{I}_{jk} \leq I_{jk} \leq \overline{I}_{jk} \leq \infty.$$

In our application to European countries, we will compute the optimal road network expansions starting from an observed road network $\underline{I}_{jk}$.

While the graph $(\mathcal{J}, \mathcal{E})$ is undirected, there is no need to impose symmetry in investments or costs between connected locations, i.e., we can accommodate $I_{jk} \neq I_{kj}$. We note that the actual direction of the flows Q^n_{jk} is endogenous and that the marginal transport cost $\tau_{jk}\left(Q^n_{jk}, I_{jk}\right)$ varies depending on the direction, due to geographic features, quantities shipped and the level of infrastructure.

3.2 Planner's Problem

We solve the problem of a utilitarian social planner who maximizes welfare under two extreme scenarios: either labor is immobile or freely mobile. The first scenario corresponds to the standard assumption in international trade models, while the second corresponds to standard urban economics model in the tradition of Rosen-Roback (Roback, 1982). In the former case, we let ω_j be the planner's weight attached to each worker located in region j. We define each problem in turn.

Definition 1. *The planner's problem with immobile labor is*

$$W = \max_{\substack{c_j, h_j, C_j, \{I_{jk}\}_{k \in \mathcal{N}(j)}, \\ \left\{C^n_j, L^n_j, \mathbf{V}^n_j, \mathbf{X}^n_j, \{Q^n_{jk}\}_{k \in \mathcal{N}(j)}\right\}_n}} \sum_j \omega_j L_j U\left(c_j, h_j\right)$$

subject to:

(i) availability of traded commodities,

$$c_j L_j \leq C^T_j\left(C^1_j, \ldots, C^N_j\right) \text{ for all } j;$$

and availability of non-traded commodities,

$$h_j L_j \leq H_j \text{ for all } j;$$

(ii) the balanced-flows constraint,

$$C_j^n + \sum_{n'} X_j^{nn'} + \sum_{k\in\mathcal{N}(j)} \left(1+\tau_{jk}\left(Q_{jk}^n, I_{jk}\right)\right) Q_{jk}^n \leq F_j^n\left(L_j^n, \mathbf{V}_j^n, \mathbf{X}_j^n\right) + \sum_{i\in\mathcal{N}(j)} Q_{ij}^n \text{ for all } j,\ n;$$

(iii) the network-building constraint,

$$\sum_j \sum_{k\in\mathcal{N}(j)} \delta_{jk}^I I_{jk} \leq K,$$

subject to a pre-existing network,

$$0 \leq \underline{I}_{jk} \leq I_{jk} \leq \overline{I}_{jk} \leq \infty \text{ for all } j, k \in \mathcal{N}(j);$$

(iv) local labor-market clearing,

$$\sum_n L_j^n \leq L_j \text{ for all } j;$$

and local factor market clearing for the remaining factors,

$$\sum_n V_j^{mn} \leq V_j^m \text{ for all } j \text{ and } m; \text{ and}$$

(v) non-negativity constraints on consumption, flows, and factor use,

$$\begin{aligned} C_j^n, c_j, h_j &\geq 0 \text{ for all } j \in \mathcal{N}(j), n \\ Q_{jk}^n &\geq 0 \text{ for all } j, k \in \mathcal{N}(j), n \\ L_j^n, V_j^{mn} &\geq 0 \text{ for all } j, m, n. \end{aligned}$$

If labor is freely mobile then the problem is defined as follows.

Definition 2. *The planner's problem with labor mobility is*

$$W = \max_{\substack{u, c_j, h_j, C_j, \{I_{jk}\}_{k\in\mathcal{N}(j)}, L_j, \\ \left\{C_j^n, L_j^n, \mathbf{V}_j^n, \mathbf{X}_j^n, \{Q_{jk}^n\}_{k\in\mathcal{N}(j)}\right\}_n}} u$$

subject to restrictions (i)-(v) above; as well as:

(vi) free labor mobility,

$$L_j u \leq L_j U\left(c_j, h_j\right) \text{ for all } j; \text{ and}$$

(vii) aggregate labor-market clearing,

$$\sum_j L_j = L.$$

This formulation restricts the planner's problem to allocations satisfying utility equalization across locations, a condition that must hold in the competitive allocation. Since U is strictly increasing, restriction (vi) implies that the planner will allocate $u = U\left(c_j, h_j\right)$ across all populated locations, and $c_j = 0$ otherwise.[25]

[25] Note that both planner's problems are defined assuming weak inequality constraints except for the the aggregate

The planner's problem from Definition 1 can be expressed as nesting three problems:

$$W = \max_{I_{jk}} \max_{Q_{jk}^n} \max_{\left\{C_j^n, L_j^n, \mathbf{V}_j^n, \mathbf{X}_j^n\right\}} \sum_j \omega_j L_j U\left(c_j, h_j\right)$$

subject to the constraints. A similar nesting can be expressed in the case with labor mobility from Definition 2. We now discuss some intuitive features of the planner's solution to each subproblem.

Optimal Allocation The innermost maximization problem over $\left(C_j^n, L_j^n, \mathbf{V}_j^n, \mathbf{X}_j^n\right)$ is a rather standard allocation problem of choosing consumption and factor use subject to the production possibility frontier and the availability of goods in each location. In what follows we refer to it as the "optimal allocation" subproblem.

Optimal Flows The optimal flow problem that determines the gross flows Q_{jk}^n through the network combines an optimal transport problem—how to map production sources to destinations—and a least-cost route problem under congestion. Such a problem, under the assumption that consumption C_j^n and production Y_j^n are taken as given, is well known in the optimal transport literature (see, for instance, Chapter 8 of Galichon (2016) or Chapter 4 of Santambrogio (2015)) and in operations research (Bertsekas, 1998). A general lesson from these two literatures is that these problems are well behaved and admit strong duality. In other words, while the least-cost route problem and the optimal coupling of sources to destinations may appear to be high-dimensional combinatorial problems, the solution boils down to finding a "potential field", meaning one Lagrange multiplier (or price) for each location/good, and expressing the flows as a function of the difference between the multipliers of two locations.

The optimal flow problem in our model inherits these properties as a special case of convex duality. To understand the solution, remember that P_j^n is the multiplier of the flows constraint (ii), equal to the price of good n in location j in the market allocation according to Proposition 4 below. The first-order condition from the planner's problem gives the following equilibrium price differential for commodity n between j and $k \in \mathcal{N}(j)$:[26]

$$\frac{P_k^n}{P_j^n} \leq 1 + \tau_{jk}^n + \frac{\partial \tau_{jk}^n}{\partial Q_{jk}^n} Q_{jk}^n, = \text{if } Q_{jk}^n > 0. \tag{8}$$

Condition (8) is a no-arbitrage condition: the price differential between a location and any of its neighbors must be less than or equal to the marginal transport cost. From the planner's perspective, this marginal cost takes into account the diminishing returns due to congestion. In the absence of congestion, $\partial \tau_{jk}/\partial Q_{jk}^n = 0$, the price differential would be bounded by the iceberg cost, $1 + \tau_{jk}^n$.

This expression carries a number of intuitive properties that we exploit throughout the analysis.

labor-market clearing condition (vii), which must hold with equality. The weak inequalities allow for some locations to be unpopulated ($L_j^n = h_j = C_j^n = c_j = 0$), as well as for some factors to be used in only some sectors ($V_j^n = 0$).

[26] Appendices A.1 and A.2 present the first-order conditions from the planner's problem from which the expressions discussed through the paper are derived.

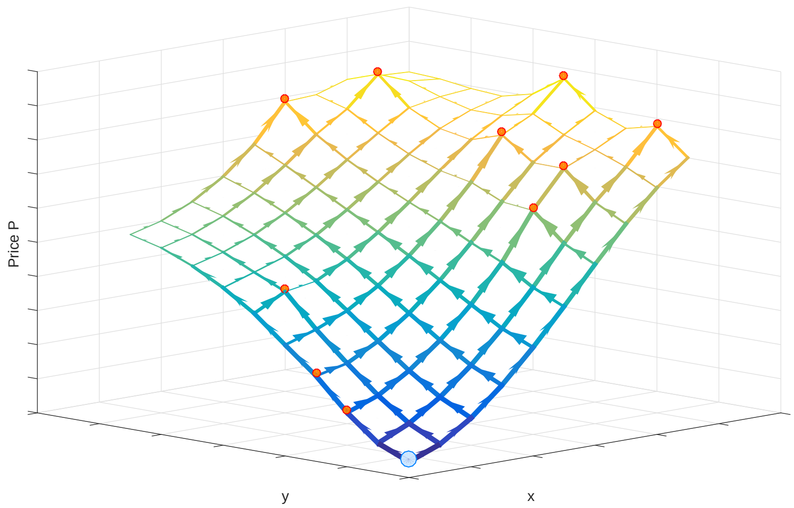


Figure 1: Example of Optimal Flows as a Function of the Price Field

Notes: The picture shows an example of optimal flows in a 15×15 square network with uniform infrastructure across links and one good produced at the origin (blue circle) and consumed in 10 other locations (orange circles). The price in each location is indicated by the z-axis coordinate, and corresponds to a solution of the optimal flow problem given production, consumption and population. The density of flows is represented by the thickness of links and their direction is indicated by the arrows.

Given the network investment, it identifies the trade flow Q^n_{jk} as a function of the price differential as long as the right-hand side can be inverted. This inversion is possible under the condition that the total transport cost, $Q^n_{jk}\tau^n_{jk}$, is convex in the quantity shipped. Under that condition, the gross trade flow Q^n_{jk} is increasing in the price differential P^n_k/P^n_j: the larger the difference in marginal valuations, the higher the flow to the location where the product is more scarce. Condition (8) also implies that goods in each sector flow in only one direction; i.e. $Q^n_{jk} > 0 \Rightarrow Q^n_{kj} = 0$. However, along a given link there may be flows in opposite directions corresponding to different sectors.

To help visualize the geometric properties of the problem, Figure 1 illustrates how a price field can implement the optimal flows given consumption and production. In the example, a good is produced in the location at the origin (blue circle) and demanded in ten locations (orange circles). The prices, represented on the z-axis, attain their lowest value at the point of production, and gradually increase with the distance from that point. The optimal flows follow the price gradient according to equation (8) under equality. The locations where consumption takes place are local peaks of the price field, as long as these locations do not re-ship the good.[27]

The least-cost route optimization present in the applications of gravity trade models discussed in the literature review corresponds to the solution to this optimal transport problem assuming

[27]In this example there are some shipments in every link, although they become negligible in regions faraway from the points of production and consumption. As shown below, links with zero flows may arise depending on the shape of the transport technology. To construct the example in Figure 1, we have used the convenient property that, with congestion, the right-hand side of equation (8) can be inverted to express the flows as a function of prices. The case without congestion lacks such an inversion but is a linear programming problem that can be tackled with the simplex algorithm (Bertsekas, 1998).

no congestion. In that case, the optimal transport problem can be solved independently from the rest of the model. In our case, determining the least-cost routes requires information about the flows, the supply, and the demand for each good, which are endogenously solved as part of the allocation. Therefore, the optimal transport problem must be solved jointly with the optimal allocation problem.

Optimal Network Consider now the outer problem of choosing the transport network I_{jk} for all $j \in \mathcal{J}$ and $k \in \mathcal{N}(j)$ given the optimal transport and the neoclassical allocation. Letting μ be the multiplier of the network-building constraint (iii), the planner's choice for I_{jk} implies

$$\underbrace{\mu\delta^I_{jk}}_{\text{Marginal Building Cost}} \geq \underbrace{\sum_n P^n_j Q^n_{jk}\left(-\frac{\partial \tau^n_{jk}}{\partial I_{jk}}\right)}_{\text{Marginal Gain from Infrastructure}}, \tag{9}$$

with equality if there is actual investment, $I_{jk} > \underline{I}_{jk}$. This condition compares the marginal cost and benefits from investing on the link jk. The left-hand side is the opportunity cost of building an extra unit of infrastructure along jk, equal to the marginal value of the scarce resource K in the economy (the multiplier μ of the the network building constraint (7)) times the rate δ^I_{jk} at which that resource translates to infrastructure. In turn, the gain from the additional infrastructure, on the right hand side of (9), is the reduction in per-unit shipping costs, $-\partial\tau^n_{jk}/\partial I_{jk}$, applied to the total value of the goods used as input in the transport technology, the trade flows $\sum_n P^n_j Q^n_{jk}$.[28]

Importantly, the network investment problem inherits the properties that make the optimal transport problem tractable. Substituting the solution for Q^n_{jk} as function of the price differentials P^n_k/P^n_j into (9) implies that the optimal infrastructure I_{jk} between locations j and k is only a function of prices in each location. Hence, rather than searching in the very large space of all networks, this condition allows us to solve for the optimal investment link by link given the considerably smaller set of all prices.

3.3 Properties

Convexity We establish conditions for the convexity of the planner's problem, which guarantee its numerical tractability.

Proposition 1. *(Convexity of the Planner's Problem) (i) Given the network $\{I_{jk}\}$, the joint optimal transport and allocation problem in the fixed (resp. mobile) labor case is a convex (resp. quasiconvex) optimization problem if $Q\tau_{jk}(Q, I_{jk})$ is convex in Q for all j and $k \in \mathcal{N}(j)$; and (ii) if in addition $Q\tau_{jk}(Q, I)$ is convex in both Q and I for all j and $k \in \mathcal{N}(j)$, then the full planner's*

[28] Recent papers measure the first-order impact of changes in bilateral trade costs on world welfare (Atkeson and Burstein, 2010; Burstein and Cravino, 2015; Lai et al. 2015; Allen et al., 2014) or in trade costs in specific links of a transport network on country-level welfare (Allen and Arkolakis, 2016) around an observed equilibrium. The right-hand side of (9) could be used for a similar purpose, given a specific set of changes in trade costs. In our context, this expression is one part of the full characterization of the global optimum, alongside with the optimal allocation and optimal flows.

problem including the network design problem from Definition 1 (resp. Definition 2) is a convex (resp. quasiconvex) optimization problem. In either the joint transport and allocation problem, or the full planner's problem, strong duality holds when labor is fixed.

The first result establishes that the joint optimal allocation and optimal transport subproblems, taking the infrastructure network $\{I_{jk}\}$ as given, define a convex problem for which strong duality holds under the mild requirement that the transport technology $Q\tau_{jk}(Q, I_{jk})$ is (weakly) convex in Q. This property ensures that our specific way of introducing an optimal-transport problem into a general neoclassical economy is tractable. Specifically, it guarantees the existence of Lagrange multipliers that implement the optimal allocation and transport subproblems and ensures the sufficiency of the Karush-Kuhn-Tucker (KKT) conditions, in turn allowing us to apply a duality approach to solve the model numerically—an approach which, as discussed in Section 3.5, substantially reduces computation times. Even if the full problem, including the network design, is not convex due to increasing returns to the network building technology (i.e., if part (ii) of the proposition fails but part (i) holds), a large subset of the full problem can be solved using these efficient numerical methods.[29]

The second result establishes the convexity of the full planner's problem, including the network design, under the stronger requirement that the transport cost function $Q\tau_{jk}(Q, I_{jk})$ is jointly convex in Q and I. This condition restricts how congestion in shipping and the returns to infrastructure enter in the transport technology in each link through $\tau_{jk}(Q, I)$. In the absence of congestion (i.e., if $\partial\tau_{jk}/\partial Q = 0$), convexity fails unless τ_{jk} is a constant.

Example: Log-Linear Parametrization of Transport Costs A convenient parametrization of (4) is the constant-elasticity transport technology,

$$\tau_{jk}(Q, I) = \delta^{\tau}_{jk}\frac{Q^{\beta}}{I^{\gamma}} \text{ with } \beta \geq 0, \gamma \geq 0. \tag{10}$$

If $\beta > 0$, this formulation implies congestion in shipping: the more is shipped, the higher the per-unit shipping cost; when $\beta = 0$, the marginal cost of shipping is invariant to the quantity shipped, as in the standard iceberg formulation. In turn, γ captures the elasticity of the per-unit cost to infrastructure. The scalar δ^{τ}_{jk} captures the geographic trade frictions that affect per-unit transport costs given the quantity shipped Q and the infrastructure I, such as distance, ruggedness, or difference in elevation.

When the transport technology is given by (10), many of the preceding results admit intuitive closed-form formulations. First, the restriction that $Q\tau_{jk}(Q, I)$ is convex in both arguments from Proposition 1 holds if and only if $\beta \geq \gamma$. This inequality captures a form of diminishing returns

[29] The proof of Proposition is 1 is immediate and can be summarized here. Given the neoclassical assumptions, the objective function is concave and the constraints are convex, except possibly for the balanced-flows constraint. Convexity of the transport cost $Q\tau_{jk}(Q, I_{jk})$ ensures convexity of that constraint as well. In the case with labor mobility, the planner's problem can only be recast as a quasiconvex optimization problem, but the Arrow-Enthoven theorem for the sufficiency of the Karush-Kuhn-Tucker conditions under quasiconvexity, requiring that the gradient of the objective function is different from zero at the optimal point, is satisfied (Arrow and Enthoven, 1961).

to the overall transport technology: the elasticity of per-unit transport costs to investment in infrastructure is smaller than its elasticity with respect to shipments. Second, from the no-arbitrage condition (8), we obtain the following solution for total flows from j to k as function of prices:

$$Q_{jk}^n = \left[\frac{1}{1+\beta} \frac{I_{jk}^{\gamma}}{\delta_{jk}^{\tau}} \max \left\{ \frac{P_k^n}{P_j^n} - 1, 0 \right\} \right]^{\frac{1}{\beta}}. \tag{11}$$

This solution naturally implies that better infrastructure is associated with higher flows given prices and geographic trade frictions. Third, using the log-linear transport technology (10), whenever the planner chooses to build on top of existing infrastructure ($I_{jk} > \underline{I}_{jk}$), the optimal infrastructure (9) arising from the optimal-network problem is

$$I_{jk}^* = \left[\frac{\gamma}{\mu} \frac{\delta_{jk}^{\tau}}{\delta_{jk}^{I}} \left(\sum_n P_j^n \left(Q_{jk}^n \right)^{1+\beta} \right) \right]^{\frac{1}{1+\gamma}}. \tag{12}$$

Given the prices at origin, the optimal infrastructure increases with the gross flows Q_{jk}^n. Given these flows, infrastructure also increases with prices at origin: because shipping requires the good being shipped as an input, a higher sourcing price implies a higher marginal saving from investing. Conditioning on these outcomes, infrastructure increases with δ_{jk}^{τ}, reflecting that optimal infrastructure investments offset geographic trade frictions, and decreases with δ_{jk}^{I}, reflecting that the investment is smaller where it is more costly to build.

Expression (12) determines the level of infrastructure when there actually is investment ($I_{jk} > \underline{I}_{jk}$). Because it satisfies the Inada condition, the log-linear specification (10) implies that the solution to the planner's problem features a positive investment whenever the price of any good varies between neighboring locations, $P_j^n \neq P_k^n$ for any n. Specifically, the optimal level of infrastructure is

$$I_{jk} = \max \left\{ I_{jk}^*, \underline{I}_{jk} \right\}. \tag{13}$$

where, combining (11) with (9), we reach an explicit characterization of the optimal infrastructure in each link as function of equilibrium prices alone:

$$I_{jk}^* = \left[\frac{\kappa}{\mu \delta_{jk}^{I} \left(\delta_{jk}^{\tau} \right)^{\frac{1}{\beta}}} \left(\sum_{n: P_k^n > P_j^n} P_j^n \left(\frac{P_k^n}{P_j^n} - 1 \right)^{\frac{1+\beta}{\beta}} \right) \right]^{\frac{\beta}{\beta-\gamma}}. \tag{14}$$

where $\kappa \equiv \gamma (1+\beta)^{-\frac{1+\beta}{\beta}}$ is a constant and the multiplier μ is such that the network-building constraint (7) is satisfied.

Proposition 2. *(Optimal Network in Log-Linear Case) When the transport technology is given by* (10)*, the full planner's problem is a convex (resp. quasiconvex) optimization problem if* $\beta \geq \gamma$*. The optimal infrastructure is given by* (13) *implying that, in the absence of a pre-existing network*

($\underline{I}_{jk} = 0$), then $I_{jk} = 0 \Leftrightarrow P_k^n = P_j^n$ for all n.

Under a general formulation of the transport technology $\tau_{jk}(Q, I)$ before imposing the log-linear form (10), and in the absence of a pre-existing network ($\underline{I}_{jk} = 0$), then: i) the solution to the full planner's problem may feature no infrastructure and no trade in some links even if the prices vary between the pairs of nodes connected by those links; and ii) in the presence of a pre-existing network ($\underline{I}_{jk} \geq 0$), the optimal transport subproblem may feature zero flows along links with positive infrastructure even if prices are different (i.e., there may be unused roads). However, when the transport technology takes the loglinear form (10), Proposition 2 implies that these possibilities arise if and only if there are no incentives to trade ($P_j^n = P_k^n$ for all n) due to i) the Inada condition on I_{jk} in the transport technology (10), and ii) the property that the marginal shipping costs are zero when no shipping is done as long as $\beta \geq 0$, respectively.

Non-Convexity: the Case of Increasing Returns to Transport When the condition guaranteeing global convexity in Proposition 1 fails, the constraint set in the planner's problem is not convex, and the sufficiency of the first-order conditions is not guaranteed. We may nonetheless implement these cases numerically, as we discuss in Section 3.5, and characterize certain properties of the optimal network theoretically, as we do now. Focusing on the log-linear specification (10) introduced above, such nonconvexities arise when the transport technology features economies of scale, $\gamma > \beta$.

We show in a simple special case how the qualitative properties of the optimal network are affected by such economies of scale. In particular, increasing returns to investment in infrastructure create an incentive for the planner to concentrate flows on few links. As a result, the optimal network may take the form of a *tree,* a property already highlighted for various non-economic environments in the optimal transport literature.[30]

Proposition 3. *In the absence of a pre-existing network (i.e., $\underline{I}_{jk} = 0$), if the transport technology is given by* (10) *and satisfies $\gamma > \beta$, and if there is a unique commodity produced in a single location, the optimal transport network is a tree.*

A tree is a connected graph with no loops (see Figure 2). Intuitively, under the conditions of the proposition, loops cannot be optimal, because they waste resources. On the margin, it is always better to remove alternative paths linking pairs of nodes and concentrate infrastructure investments and flows in fewer links. As a result, in the optimal network a single path connects any two locations, a defining characteristic of a tree.

Note that this property only holds when there is only one source for one commodity. When goods are produced in multiple regions, or when there are multiple goods, it may still be optimal to maintain loops depending on the underlying graph and comparative advantages. However, the incentives to concentrate flows on fewer but larger routes remain. In Section 4 we present several

[30]E.g., these applications range from the formation of blood vessels to irrigation or electric power supply systems (Banavar et al., 2000; Bernot et al., 2009).

Figure 2: Examples of tree and non-tree networks

(a) Non-tree (b) Tree

examples with multiple goods and multiple productive locations where, if $\gamma > \beta$, the topology of the optimal network is sparser and concentrated on fewer links relative to cases with $\gamma \leq \beta$. Similar patterns arise in non-convex cases in the application from Section 5.

3.4 Decentralized Allocation Given the Network

We establish that the planner's optimal allocation ($\max_{C_j^n, L_j^n, \mathbf{V}_j^n, \mathbf{X}_j^n}$) and optimal transport ($\max_{Q_{jk}^n}$) subproblems given the network $\{I_{jk}\}$ correspond to a decentralized competitive equilibrium. For the decentralization of these subproblems, we do not need to take a particular stand on whether the network is the result of a planner's optimization.

Given the network, the decentralized economy corresponds to the perfectly competitive equilibrium of a standard neoclassical economy where consumers maximize utility given their budget, producers maximize profits subject to their production possibilities, and goods and factor markets clear. The only less standard feature is the existence of a transport sector with congestion. We assume free entry of atomistic traders into the business of purchasing goods in any sector at origin o and delivering at destination d for all $(o, d) \in \mathcal{J}^2$. The traders are price-takers and use a constant-returns to scale shipping technology. Each trader has a cost equal to $\tau_{jk}^n q_{jk}^n$ of delivering q_{jk}^n units of good n from j to $k \in \mathcal{N}(j)$ and takes the iceberg trade cost τ_{jk}^n as given, although this trade cost is determined endogenously through (10) as function of the aggregate quantity shipped.

As long as there is congestion in shipping, the traders will engage in an inefficient amount of shipping. We assume that the market allocation features policies that correct this externality. While there are multiple ways to achieve efficiency, we allow here for Pigouvian sales taxes t_{jk}^n on companies shipping good n on leg $j \to k$.

Consider a trader purchasing good n at location o and delivering it to location d. This company maximizes profits by optimizing over the route $r = (j_0, \ldots, j_\rho) \in \mathcal{R}_{od}$, where $j_0, \ldots, j_r$ is a sequence of nodes leading from o to d and $\mathcal{R}_{od}$ is the set of all such routes. Since transport technologies are linear, the optimal route r^n_{od} must maximize the per-unit profits:

$$\pi^n_{od} = \max_{r=(j_0,\dots,j_\rho)\in\mathcal{R}_{od}} p^n_d - \underbrace{p^n_o T^n_{r,0}}_{\text{Sourcing Costs}} - \underbrace{\sum_{k=0}^{\rho-1} p_{j_{k+1}} t^n_{j_k j_{k+1}} T^n_{r,k+1}}_{\text{Taxes}}, \tag{15}$$

where p^n_j is the price of good n in location j in the market allocation, and $T^n_{r,k}$ is the accumulated iceberg cost from location j_k to the final destination d along path r.[31] For each unit delivered at d, a shipper from d obtains the price p^n_d and must ship $T^n_{r,0}$ out of o, purchased there at the price p^n_o. In addition, shippers must pay the "toll" $p_{j_{k+1}} t^n_{j_k j_{k+1}}$ on each of the $T^n_{r,k+1}$ units that cross from j_k to j_{k+1}. In the absence of congestion taxes ($t^n_{j_k j_{k+1}} = 0$), shippers just choose the route that minimizes the iceberg cost from o to d, $T^n_\pi(o,d)$. That solution would correspond to the least-cost route optimization present in the applications of gravity trade models discussed in the literature review. Otherwise, shippers also take into account the taxes on the gross flows on each link to decide the optimal path.

To define the competitive equilibrium, we must also allocate the returns to factors other than labor. Under no labor mobility we assume that, in addition to the wage, each worker in location j receives a transfer t_j such that $\sum_{j=1}^J t_j L_j = \Pi$, where Π is an aggregate portfolio including all the sources of income except for labor.[32] Hence, workers are rebated all tax revenues and own all the primary factors and non-traded goods in the economy. This formulation allows for trade imbalances, which are needed to implement the planner's allocation under arbitrary weights.

Since it is standard, we relegate the Definition 3 of the competitive allocation with and without labor mobility to the appendix. Using that definition, we establish that the welfare theorems given the transport network hold.

Proposition 4. *(First and Second Welfare Theorems) If the sales tax on shipments of product n from j to k is*

$$1 - t^n_{jk} = \frac{1+\tau^n_{jk}}{1+\left(\varepsilon^n_{Q,jk}+1\right)\tau^n_{jk}},$$

where $\varepsilon^n_{Q,jk} = \partial \log \tau^n_{jk} / \partial \log Q^n_{jk}$, then:

(i) if labor is immobile, the competitive allocation coincides with the planner's problem under specific planner's weights ω_j. Conversely, the planner's allocation can be implemented by a market allocation with specific transfers t_j; and

(ii) if labor is mobile, the competitive allocation coincides with the planner's problem if and only if all workers own an equal share of fixed factors and tax revenue regardless of their location, i.e., $t_j = \frac{\Pi}{L}$.

[31] See condition (1)(c) of Definition 3 of the general equilibrium in Appendix A.3 for the definition of $T^n_{r,k}$.

[32] For simplicity we refer to t_j as a transfer, although it encompasses both ownership of fixed factors and government transfers. This formulation encompasses the case where returns to the fixed factors and the tax revenue in each location are owned by residents of that location. In that particular case, there would be no trade imbalances in the market allocation.

In either case, the price of good n in location j, p_j^n, equals the multiplier on the balanced-flows constraint in the planner's allocation, P_j^n.

These results are useful for bringing our model to the data in the application. Under the assumption that the observed allocation corresponds to the decentralized equilibrium, the first welfare theorem will enable us to calibrate the model using the planner's solution to the optimal allocation and optimal transport subproblems given the network. As we discuss in Section 3.6, it would also be possible to calibrate the model assuming that the observed market allocation does not feature policies correcting the externality and is therefore inefficient.

3.5 Numerical Implementation

In this section we broadly discuss our numerical implementation and relegate details to Appendix A.4.

Convex Cases Under the conditions of Proposition 1, the full planner's problem is a convex optimization problem and the KKT conditions are both necessary and sufficient. The system of first-order conditions is, however, a large system of non-linear equations with many unknowns. Fortunately, gradient-descent based algorithms make large-scale convex optimization problems like ours numerically tractable, meaning that these algorithms are guaranteed to converge to the unique global optimum (Boyd and Vandenberghe, 2004).[33]

Our problem can be tackled numerically using two equally valid approaches. The first one is to feed the numerical solver with the *primal* problem, in other words the full planner's problem exactly as written in Definition 1. Specifically, letting $\mathscr{L}$ be the Lagrangian of the planner's problem as a function of the variables controlled by the planner, $\mathbf{x} = \left(C_j^n, L_j^n, \mathbf{V}_j^n, Q_{jk}^n, \ldots\right)$, and the multipliers $\boldsymbol{\lambda} = \left(P_j^n, \ldots\right)$ on the various constraints,[34] the primal problem consists of solving the saddle-point problem

$$\sup_{\mathbf{x}} \inf_{\boldsymbol{\lambda} \geq \mathbf{0}} \mathscr{L}\left(\mathbf{x}, \boldsymbol{\lambda}\right).$$

The second approach, usually preferred in the optimal transport literature, is to solve instead the *dual* problem obtained by inverting the order of optimization, i.e.,

$$\inf_{\boldsymbol{\lambda} \geq \mathbf{0}} \sup_{\mathbf{x}} \mathscr{L}\left(\mathbf{x}, \boldsymbol{\lambda}\right).$$

In our context, the convexity of the full planner's problem without labor mobility ensures that the dual problem coincides with the primal under weak conditions (Proposition 1), i.e., strong duality holds. The advantage of the dual is that we can use the first-order conditions from the optimal

[33] We use the open-source large-scale optimization package IPOPT (https://projects.coin-or.org/Ipopt) which is based on an interior point method and is able to handle thousands of variables as long the problem is sufficiently sparse. The software converges in polynomial time, in the sense that the resolution time is $O\left(n^a m^b\right)$, where n is the number of variables, m is the number of constraints and a, b some real numbers (Nesterov and Nemirovskii, 1994).

[34] These expressions are defined explicitly in Appendix A.1.

transport and the optimal investment problems, (8) and (12), as well as those from the neoclassical allocation problem, to express the control variables as functions of the multipliers, $\mathbf{x}(\lambda)$. The remaining minimization problem, $\inf_{\lambda \geq 0} \mathscr{L}(\mathbf{x}(\lambda), \lambda)$, is a convex minimization problem over fewer variables, subject to only non-negativity constraints.

Non-Convex Cases When the condition stated in Proposition 1 fails, the full planner's problem is no longer globally convex, and the method described above is not guaranteed to find the global optimum. To solve for such non-convex cases, we exploit the property, stated at the beginning of Proposition 1, that the joint neoclassical allocation and optimal transport problem nested within the planner's problem is convex as long as $Q\tau_{jk}(Q, I_{jk})$ is convex in Q. This condition is weaker and holds under the log-linear specification as long as $\beta \geq 0$, including the standard case without congestion ($\beta = 0$). We combine the primal and dual approaches to solve for the joint neoclassical allocation and optimal transport problems with an iterative procedure over the infrastructure investments. Specifically, starting from a guess on the network investment I_{jk}, we solve for the optimum over C_j^n, L_j^n, $\mathbf{V}_j^n$ and Q_{jk}^n, and then use the optimal network investment condition (9) to obtain a new guess over I_{jk}, and then repeat until convergence. We then refine the solution using a simulated annealing method that perturbs the local optimum and gradually reaches better solutions. See Appendix A.4 for additional details.

3.6 Extensions

In this section, we briefly consider various extensions and discuss how to preserve the convexity property in each case.

Congestion Across Goods We have assumed that congestion only applies within good types. A natural extension is to allow for congestion across goods. A simple way to model this feature while preserving the convexity of the problem is to assume that the per-unit cost τ_{jk}^n is denominated in units of the bundle of traded goods aggregated through C_j^T rather that in units of the good itself. Specifically, we can assume that transporting each unit of good n from j to $k \in \mathcal{N}(n)$ requires

$$\tau_{jk}^n = m^n \tau_{jk}(Q_{jk}, I_{jk}) \tag{16}$$

units of the traded goods bundle, where the parameters $(m_0, \ldots, m_N)$ capture the unit weight or volume of goods in each sector, and where $Q_{jk} = \sum_{n=1}^N m_n Q_{jk}^n$ is the total weight or volume transported from j to k. The total units of the traded goods bundle used to transport goods from j is $\sum_k Q_{jk}\tau_{jk}(Q_{jk}, I_{jk})$. After properly adjusting the resource constraints in the definition of the planner's problem,[35] the convexity of the full planner's problem is preserved under the exact same conditions stated in Proposition 1. In this case, under the log-linear specification (10), heavy enough

[35]This correction requires adding up $\sum_k Q_{jk}\tau_{jk}(Q_{jk}, I_{jk})$ to the left-hand side of (i) and eliminating $\tau_{jk}\left(Q_{jk}^n, I_{jk}\right)$ from (ii).

goods are not shipped even in the presence of price differentials between connected locations, due to the marginal congestion cost that they exert on other goods.[36]

Endogenous Supply of Resources in Infrastructure We have assumed that the network building technology uses a single input, asphalt, in aggregate fixed supply. The framework can also accommodate a network building technology using ordinary factors of production, endogenously supplied at the local level. For instance, we can assume that building I_{jk} requires local factors of production at j and k by letting $I_{jk} = \frac{1}{\delta^I_{jk}} F^I \left(L^I_j + L^I_k, H^I_j + H^I_k \right)$ where F^I is a neoclassical production function, and where $L^I_j + L^I_k$ and $H^I_j + H^I_k$ are, respectively, the amounts of labor and non-traded goods from j and k used to build infrastructure on the link between j and k. This formulation would encompass the feature that building up the network takes up resources, such as land, that could be used to produce goods or consumed in the form of housing. After properly modifying the factor resource constraints, the full planner's problem is convex under the same conditions as before as in Proposition 1 as long as F^I is concave in all its arguments.

Externalities and Inefficiencies in the Market Allocation In Section 3.4, we assumed that the decentralized allocation is efficient. However, in some cases it may be desirable to consider an inefficient market allocation. For example, a standard formulation with agglomeration spillovers is to assume that the production technology is $Y^n_j = F^n_j \left(L^n_j, \mathbf{V}^n_j, \mathbf{X}^n_j; L_j \right)$, where the spillover from the total number of workers L_j on output Y^n_j is not internalized in the market allocation. Similarly, without the Pigouvian taxes t^n_{jk} correcting the congestion externality in shipping, the market allocation is inefficient. In these cases, it is still possible to calibrate the model and to undertake counterfactuals using a "fictitious" planner who ignores the dependence of Y^n_j on L_j or of τ^n_{jk} on Q^n_{jk}. For example, in the case of size spillovers, the fictitious planner problem is defined exactly as in Definition 2 under the assumption that the vector of aggregate population levels $\overline{\mathbf{L}} = \left\{ \overline{L}_j \right\}$ in $Y^n_j = F^n_j \left(L^n_j, \mathbf{V}^n_j, \mathbf{X}^n_j; \overline{L}_j \right)$ is taken as given.[37] As long as $F^n_j(\cdot)$ is neoclassical given $\overline{L}_j$, the statement of Proposition 1 remains the same. However, this approach requires solving an additional loop imposing that the vector of population $\mathbf{L} = \{L_j\}$ that solves the fictitious planner problem coincides with the perceived aggregate distribution $\overline{\mathbf{L}}$. It is straightforward to show that, if it exists, every distribution of population $\overline{\mathbf{L}}$ satisfying this fixed point problem corresponds to an inefficient market allocation and vice-versa.[38]

---

[36] The no-arbitrage condition (11) in this case implies all goods for which which $m^n \geq \frac{1}{\tau_{jk}(\beta+1)} \frac{P^n_k - P^n_j}{P^C_j}$, are not traded from j to k.

[37] Similarly, under congestion externalities, the fictitious planner problem is defined exactly as in Definition 2 given the shipments $\bar{\mathbf{Q}} = \left\{ \overline{Q^n_{jk}} \right\}_{j,k,n}$ in $\tau_{jk} \left(\overline{Q^n_{jk}}, I_{jk} \right)$.

[38] Whether such a fixed point exists depends on the specifics of the environment. It is beyond the scope of this paper to determine the conditions under which that is the case, but we note that, given the network $\{I_{kl}\}$, our environment can accommodate the specific parametric assumptions that guarantee existence or uniqueness of an inefficient decentralized allocation found in the previous literature. E.g., see Allen and Arkolakis (2014) for conditions that lead to existence and uniqueness in an Armington model with labor mobility and size spillovers.

4 Illustrative Examples

In this section we implement examples that illustrate the basic economic forces captured by the framework and its potential uses. We start with an endowment economy without labor mobility and only one traded and one non-traded good in a symmetric graph. Then, we progressively move to more complex cases with multiple locations in asymmetric spaces, multiple sectors, labor mobility, and heterogeneous building costs due to geographic features. Throughout the examples, we illustrate the contrast between the globally optimal networks in convex cases, where the congestion forces dominate the returns to network building, and the approximate optimal networks in cases where global convexity of the planner's problem fails. In all the examples, preferences are CRRA over a Cobb-Douglas bundle of traded and non-traded goods, $U = \left(c^{\alpha} h^{1-\alpha}\right)^{1-\rho} / (1-\rho)$ with $\alpha = \frac{1}{2}$ and $\rho = 2$. There is a single factor of production, labor, and all technologies are linear. We adopt the constant-elasticity functional forms (10) for the transport and network-building technologies.

4.1 One Good on a Regular Geometry

Comparative Statics over K in a Symmetric Network To start we impose $\beta = \gamma = 1$, which lies at the boundary of the parameter space guaranteeing global convexity. We assume a single good, no labor mobility and no geographic frictions, $\delta^{\tau}_{jk} = \delta^{I}_{jk} = 1$.

Figure A.1 presents a network with 9×9 locations uniformly distributed in a square, each connected to 8 neighbors. All fundamentals except for productivity are symmetric: $(L_j, H_j) = (1, 1)$. Labor productivity is $z_j = 1$ at the center and 10 times smaller elsewhere.

Figure A.2 shows the globally optimal network when $K = 1$ (panel (a)) and when $K = 100$ (panel (b)). The upper-left figure in each panel displays the optimal infrastructure network I_{jk} corresponding to (12). The optimal network investments radiate from the center, and so do shipments. The bottom figures in each panel display the multipliers of the flows constraint (6)—the prices in the market allocation—and consumption. Because tradable goods are scarcer in the outskirts, marginal utility is higher and so are prices. As the aggregate investment grows from $K = 1$ to $K = 100$, the network grows into the outskirts and the differences in the marginal utility shrink. Panel (a) of Figure A.3 displays the spatial distribution of prices (upper panels) and consumption (bottom). The left panels display outcomes across locations ordered by Euclidean distance to the center. As the network grows, relative prices and consumption converge to the center, and spatial inequalities are reduced.

Panel (b) of Figure A.3 illustrates the difference between the welfare gains from uniform and optimal network expansion. For K close to zero, the levels of infrastructure I_{jk} are small everywhere and every location is close to autarky. We simulate an increase in K in two cases: a proportional increase in infrastructure across all links (a "rescaled" network) and the optimal one. The figure reports the welfare increase associated with each network. Broadly speaking, the uniform network expansion corresponds to the standard counterfactual implemented in international trade, in which trade costs are reduced uniformly from autarky to trade. As K grows, the economy converges to the level of welfare under free internal trade regardless of whether the network is optimal. Moving from close to autarky to close to free trade across locations increases aggregate welfare by 5%. However, investing optimally leads to faster convergence to the free-trade welfare level. In the example, the welfare level attained in the uniform network when $K = 10^6$ is attained in the optimal network when $K = 10^3$.

Randomly Located Cities and Non-Convex Cases We now explore more complex networks and non-convex cases. Figure A.4 shows 20 "cities" randomly located in a space where each location has six neighbors. Population is $L_j = 1$ in each city and 0 otherwise. Productivity is again ten times larger at the center. The top panel shows the infrastructure and goods' flows in the optimal network. The optimal network radiates from the center to reach all destinations. Due to congestion, some destinations are reached through multiple routes. However, to reach some faraway locations such as the one in the northwest, only one route is built.

The middle panel inspects the same spatial configuration but assumes $\gamma = 2$. Now, the sufficient condition for global convexity from Proposition 1 fails. We see a qualitative change in the shape of the network. Due to increasing returns to network building, fewer roads are built but each has higher capacity. In particular, there is now only one route linking any two destinations, consistent with the no-loops result in Proposition 3.[39]

Because in the non-convex network we can only guarantee convergence to a local optimum, we refine the solution by applying the numerical approach discussed in Appendix A.4 involving simulated annealing. The bottom panel compares the non-convex network before and after the annealing refinement. The refined network economizes on the number of links, leading to a welfare increase but preserving the no-loops property.

4.2 Many Sectors, Labor Mobility, and Non-Convexity

We now further introduce multiple traded goods and labor mobility. We allow for 11 traded commodities, one "agricultural" good (good 1) that may be produced everywhere outside of "cities" ($z_j^1 = 1$ in all "countryside" locations) and ten "industrial" goods, each produced in one random city only ($z_j^n = 1$ in only one city j and $z_j^n = 0$ otherwise). These goods are combined via a constant elasticity of substitution aggregator with elasticity of substitution $\sigma = 2$. Labor continues to be the sole factor of production, but is now mobile. The supply of the non-traded good is uniform, $H_j = 1$ for all j.

Figure A.5 shows the convex case ($\beta = \gamma = 1$). The first panel shows the optimal network. In the figure, each circle's size denotes the population share. The remaining figures show the shipments of each good, with the circle sizes representing the shares in total production for the corresponding good. Figure A.6 shows the optimal network with annealing in the nonconvex case when $\gamma = 2$.

In these examples, we observe complex shipping patterns. There are bilateral flows over each

[39] While we derive the no-loop result when there is only one producer, in this example every populated location produces the good.

link, now involving several commodities. Overall, the optimal network in the first panel reflects the spatial distribution of comparative advantages. Since industrial goods are relatively scarce, wages and population are higher in the cities that produce them. Due to the need to ship industrial goods to the entire economy and to bring agricultural goods to the more populated cities, the transport network has better infrastructure around the producers of industrial products. As Panel (a) of each figure illustrates, the optimal network links the industrial cities through wider routes branching out into the countryside. The agricultural good, being produced in many locations, travels short distances and each industrial city is surrounded by its agricultural hinterland.

The comparison between Figures A.5 and A.6 confirms the intuition that, in the presence of economies of scale in transportation, the optimal network becomes more skewed towards fewer but wider "highways". Note, however, that the tree property from Proposition 3 no longer holds because there are multiple goods.

4.3 Geographic Features and New Transport Technologies

We now show how the framework can accommodate geographic accidents. To highlight the role of these frictions we revert to a case with a single good and no factor mobility. Panel (a) of Figure A.7 shows 20 cities randomly allocated in a space where each location is connected to 8 other locations. Population equals 1 in all cities and productivity is the same everywhere (equal to 0.1) except in the central city, displayed in red, where it is 10 times larger. Each city's size in the figure varies in proportion to consumption.

As implied by condition (12), the optimal infrastructure in a given link depends on the link-specific building cost δ^I_{jk}. In panel (a) we show the optimal network under the assumption that the cost of building infrastructure is proportional to the Euclidean distance:

$$\delta^I_{jk} = \delta_0 \text{Distance}^{\delta_1}_{jk}. \tag{17}$$

As in our first set of examples, the optimal network radiates from the highest-productivity city to alleviate differences in marginal utility.

In panel (b), we add a "mountain" by adding an elevation dimension to each link and reconfiguring the building cost as

$$\delta^I_{jk} = \delta_0 \text{Distance}^{\delta_1}_{jk} \left(1 + |\Delta \text{Elevation}|_{jk}\right)^{\delta_2}. \tag{18}$$

Because it is more costly to build through the mountain, the optimal network circles around it to reach the cities in the northeast. Because more resources are invested in that region, the network shrinks elsewhere.

In the subsequent figures, we either increase or decrease the cost of building the network in specific links. Specifically, we allow for the more general specification:

$$\delta^I_{jk} = \delta_0 \text{Distance}^{\delta_1}_{jk} \left(1 + |\Delta \text{Elevation}|_{jk}\right)^{\delta_2} \delta_3^{\text{CrossingRiver}_{jk}} \delta_4^{\text{AlongRiver}_{jk}}. \tag{19}$$

In panel (c) we include a river and assume that $\delta_3 = \delta_4 = \infty$, so that investing in infrastructure either across or along the river is prohibitively costly. The optimal network linking cities across the river can only be built through the one patch of dry land. In that natural crossing there is a "bottleneck", and a large amount of infrastructure is optimally built.

In panel (d) we assume instead that no dry patch exists and that building bridges is feasible, $1 < \delta_3 < \infty$. Now, the planner builds two bridges, directly connecting the pairs of cities across the river. Panel (e) further allows for water transport by allowing to building transport capacity along the river ($\delta_4 < \infty$). The planner retains the bridges, but now faraway locations in the southeast are reached by water instead of ground transport.

Finally, panel (f) moves to the non-convex case, $\gamma = 2 > \beta$, implemented through the combination of first-order conditions and simulated annealing approach described in Section 3.5. Now, a unique route links any two cities, water transport is not used, and a single bridge is built.

We conclude by showing how the optimal reconfiguration of the transport network triggered by the arrival of a new transport technology can lead to a drastic reconfiguration of city sizes. Both panels of Figure A.8 correspond to an economy with random cities, all with same population, where productivity is 10 times larger in the city represented in red. The circle sizes again represent consumption per capita. Panel (a) shows an economy with strong dependence on water transport, with low δ_3 in (19). The optimal network implies high consumption in the city near the river. In panel (b) we assume that ground transport becomes cheap (e.g., due to the arrival of railways), represented by a lower δ_1 in (19). As a result, water transport is abandoned and the spatial distribution of consumption per worker is reconfigured. The city near the river shrinks and other cities that become more central to the new network, as well as those in their hinterland, grow.

5 Road Network Expansion and Misallocation in Europe

We apply the framework for quantitative analysis of road networks in Europe. We start by describing the data sources and the steps we used to represent data on economic activity and road networks in terms of the graph of our model. Then we choose the fundamentals to match the observed distribution of economic activity within each country. We conclude by implementing counterfactuals involving the optimal transport network. The counterfactuals tackle two related questions: how large would the gains from optimal expansions of current road networks be, and how large are the losses from misallocation of current roads? The first question is motivated by the fact that a large fraction of public investment is directed to expansion of roads, yet no quantitative general-equilibrium analysis exists of the optimal placement of these investments and their impact across and within countries. The second question is motivated by the fact that the allocation of regional investments in transportation is often sensitive to frictions and political interests, potentially leading to inefficiencies in the observed transport networks.

5.1 Data and Discretization

Sources We combine geocoded data on the shape of road networks, population, and income across 25 European countries. The road network data is from EuroRegionalMap by EuroGeographics. The dataset combines shapefiles on the current road network from each European country's mapping and cadastral agencies. For example, the French road network is represented by 38699 segments connecting 159519 distinct geographic points.

An appealing feature of this dataset is that each segment of a road network has information about objective measures of road quality including road use (national, primary, secondary, or local) and number of lanes, as well as other features such as whether it is paved or includes a median. National roads encompass each country's highway system, and, as shown in Table A.1 in Appendix C, they are always paved, more likely to include a median, and feature twice as many average lanes relative to other types of roads.[40] Since the roads labeled as primary, secondary and tertiary have similar characteristics, we bundle them into a single "non-national roads" category. In our analysis, we use two features of each segment: whether it belongs to a national road, and its number of lanes.

We use population data from NASA-SEDAC's Gridded Population of the World (GPW) v.4, and value added from Yale's G-Econ 4.0. The GPW population data is reported for 30 arc-second cells (approximately 1 kilometer), and the G-Econ value-added data is reported for 1 arc-degree cells (approximately 100 km). We undertake our analysis using 0.5 arc-degree cells (approximately 50 km). The resulting number of cells within each country is in most cases in between the number of level-3 NUTS subdivisions (provinces or counties) and the number of LAU subdivisions (municipalities or communes). We allocate population to each 0.5-degree cell by aggregating the smaller cells in GPW, and we allocate income by apportioning the G-Econ cells according to the GPW-based population measure.

We denote by L_j^{obs} and GDP_j^{obs} the population and value added observed in each cell j of each country. Using these data we also construct empirical counterparts to the underlying geography $(\mathcal{J}, \mathcal{E})$ corresponding to the locations and links in the graph of our model, as well as an observed measure of infrastructure I_{jk}^{obs} for each link.

We perform all the analysis separately for each of the 25 countries included in EuroRegionalMap for which data on number of lanes is available. This set includes rich and poor countries, as well as geographically large and small. Table A.2 in Appendix C reports the list of countries with summary statistics about the size and average features of their road networks, the number of cells, and features of their discretized road networks.[41]

[40]E.g., roads labeled as national in the data include the Autobahn highway system in Germany, *autovias* and *autopistas* in Spain, and the *autoroute* system in France.

[41]The 25 countries included in our data are Austria, Belgium, Cyprus, Czech Republic, Denmark, Finland, France, Georgia, Germany, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Macedonia, Moldova, Netherlands, Northern Ireland, Portugal, Serbia, Slovakia, Slovenia, Spain and Switzerland. We exclude the following 10 countries for which road lane data is not available in EuroRegionalMap: Bulgaria, Croatia, Great Britain, Greece, Estonia, Iceland, Norway, Poland, Romania, and Sweden. For Luxembourg we use 0.25 arc-degree cells to allow for a significant number of cells.

Underlying Graph To define the set of nodes $\mathcal{J}$ in each country, we use the high-resolution GPW population data to locate the population centroid of each cell. The population centroids are usually very close to a node on the road network. We relocate each population centroid to the closest point on a national road crossing through the cell, or on other types of roads if no national roads cross through the cell.[42] We define the observed population and income of each node $j \in \mathcal{J}$ to be equal to the total income GDP_j^{obs} and the population L_j^{obs} of the cell that contains it.

In turn, we define the set of edges $\mathcal{E}$ as the links between nodes in contiguous cells. This step defines a set of up to eight neighbors $\mathcal{N}(j)$ for each node $j \in \mathcal{J}$: the 4 nodes in horizontal or vertical neighbors and the 4 nodes along the diagonals.

Discretized Road Network To construct a measure of infrastructure corresponding to I_{jk} in our model, we first aggregate the observed attributes of the road network over the actual roads linking each $j \in \mathcal{J}$ and $k \in \mathcal{N}(j)$. We use information on whether each segment s on the actual road network belongs to a national road and its number of lanes. We define the average number of lanes and average road type for the link between j and k as follows:

$$\begin{aligned} lanes_{jk} &= \sum_{s \in S} \omega_{jk}(s)\, lanes(s), \\ nat_{jk} &= \sum_{s \in S} \omega_{jk}(s)\, nat(s), \end{aligned}$$

where $lanes(s)$ is the number of lanes on each segment s on the actual road network S, $nat(s)$ indicates whether segment s belongs to a national road, and $\omega_{jk}(s)$ is the weight attached to the infrastructure of each segment when computing the level of infrastructure from j to k. The weights $\omega_{jk}(s)$ should be larger on segments of the road network that are more likely to be used when shipping from j to k, and equal to zero for all $s \in S$ if no direct route exists linking j and k. We define $\omega_{jk}(s)$ based on the fraction of the cheapest path $\mathcal{P}(j,k)$ from j to k corresponding to that segment:

$$\omega_{jk}(s) = \begin{cases} \frac{length(s)}{\sum_{s' \in \mathcal{P}(j,k)} length(s')} & s \in \mathcal{P}(j,k) \\ 0 & s \notin \mathcal{P}(j,k) \end{cases}$$

where $length(s)$ is the length of segment s and $\mathcal{P}(j,k)$ is the cheapest path from j to k on the actual road network.[43] We follow these steps as long as the cheapest path does not stray from the cells containing j and k.[44] When that happens, we assume that no direct path from j to k exists

[42] This leads to a very small adjustment: on average across countries, the average relocation across all cells within a country is 6.2 km.

[43] This step does not involve solving the model. In this step, for each pair of nodes $j \in \mathcal{J}$ and $k \in \mathcal{N}(j)$ we ask: what are the average characteristics (number of lanes and type of road) of the actual route connecting these two locations in the real world? For this we must pick some route in the real world, and the cheapest-route criterion is a selection device. This cheapest path is constructed weighting each segment s by its road user cost based on data from Combes and Lafourcade (2005) and other sources. See Appendix C for details on these weights.

[44] We classify a path from j to k as straying from the cells containing j and k if more than 50% of the path steps over cells that do not contain j or k.

in the actual road network, $\mathcal{P}(j,k) = \varnothing$, in which case $\omega_{jk}(s) = 0$ for all segments $s \in S$.

Observed Measured of Infrastructure After implementing the previous steps, we obtain the measures $lanes_{jk}$ and nat_{jk} capturing the average number of lanes and the likelihood of using national roads between j and k on the real network. However, our model includes a single index of infrastructure, I_{jk}. Hence, we define the observed measure of infrastructure for each $j \in \mathcal{J}$ and $k \in \mathcal{N}(j)$ by aggregating the observed attributes $lanes_{jk}$ and nat_{jk} into a single index:

$$I_{jk}^{obs} = lanes_{jk} \times \chi_{nat}^{1-nat_{jk}}. \tag{20}$$

To compute the index we must assign a value to χ_{nat}. We note that, in the model, the resource cost of building a level of infrastructure I_{jk}^{obs} is $\delta_{jk}^{I} I_{jk}^{obs}$, with features of the terrain entering through δ_{jk}^{I}. Therefore, the coefficient $1/\chi_{nat} > 1$ in (20) captures the extent by which the features associated with national roads raise construction and maintenance costs relative to a non-national road. We set $1/\chi_{nat} = 5$, which corresponds to expenditures in road construction and maintenance per kilometer of federal motorways relative to the cost per kilometer of other trunk roads in Germany in 2007, as reported by Doll et al. (2008).

In sum, we construct the observed infrastructure I_{jk}^{obs} as the average number of national road lanes over the path from j to k on the actual road network, if a direct path exists.[45]

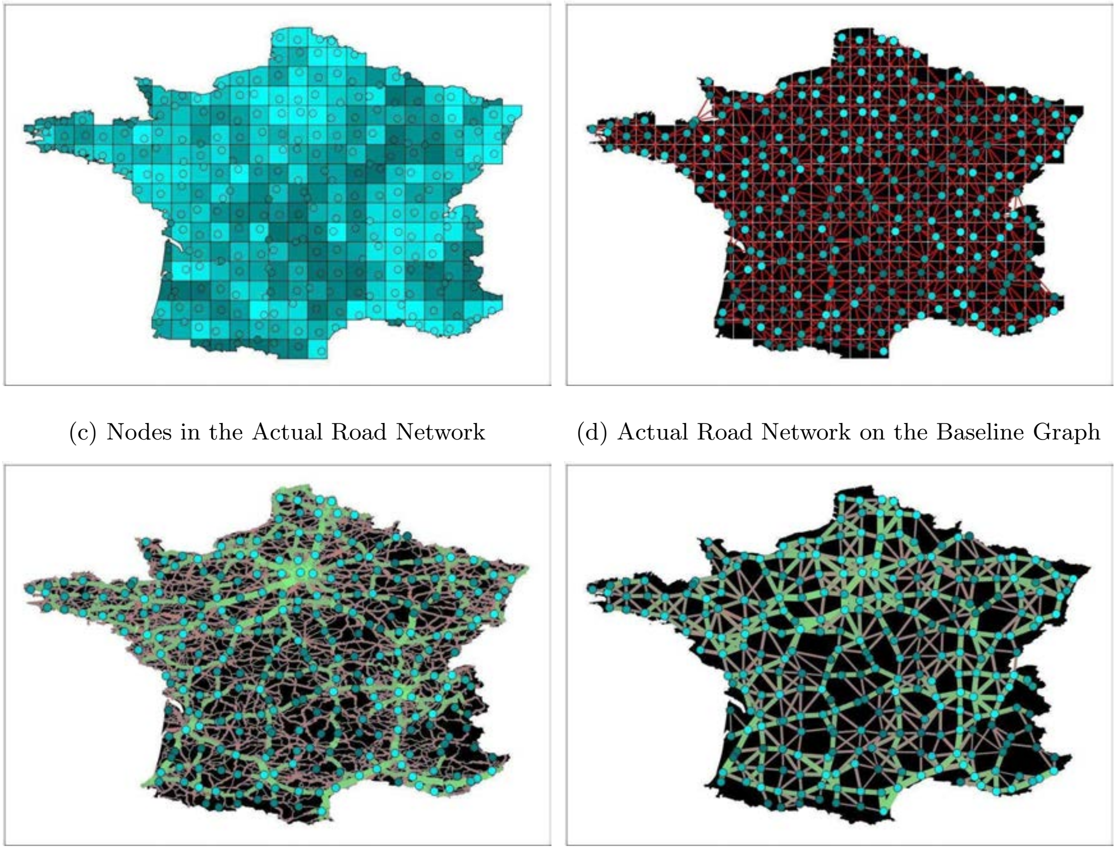
We verify that I_{jk}^{obs} correlates with external measures of road quality: first, across countries, the average of this infrastructure measure has a correlation of 0.45 with the road-quality index from the Global Competitiveness Report (WorldEconomicForum, 2016);[46] second, across all connected nodes in all countries in the discretized network, there is a correlation of 0.67 between I_{jk}^{obs} and the speed on the quickest path according to GoogleMaps. This relationship between speed and infrastructure is depicted in Figure A.10.

Examples: France and Spain Figures 3 and 4 represent each of the steps described above for two large countries in our data, France and Spain. Panel (a) of each panel shows the discretized map and associated population. Brighter cells are more populated, corresponding to higher deciles of the population distribution across cells. The (b) panels display the cells, the centroids (light blue circles) and the edges (red segments) of the underlying graph. The (c) panels show the centroids and the full road network. Green segments correspond to national roads and red segments correspond to other roads, and the width of each road is proportional to its number of lanes.

[45] To understand the units, we note that $nat_{jk} = 1$ implies $I_{jk}^{obs} = lanes_{jk}$. Therefore, this measure can be interpreted as saying that the resource cost on the path linking j to k is equivalent to the resource cost, for that same link, of a national road with I_{jk}^{obs} lanes.

[46] The average infrastructure of each country is constructed as $\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{N}(j)} \omega_{jk} I_{jk}^{obs}$ where $\omega_{jk} = \frac{dist_{jk}}{\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{N}(j)} disst_{jk}}$ is the fraction of total distance in the discretized network corresponding to the link from j to k. Column (6) of Table A.2 in Appendix C reports this value for each country. Note that the average infrastructure index captures both the number of lanes and the prevalence of national roads, and it is therefore not directly comparable to the the average lane per kilometer reported in Column (3).

Figure 3: Discretization of the French Road Network



(a) Population on the Discretized Map

(b) Nodes and Edges in the Baseline Graph

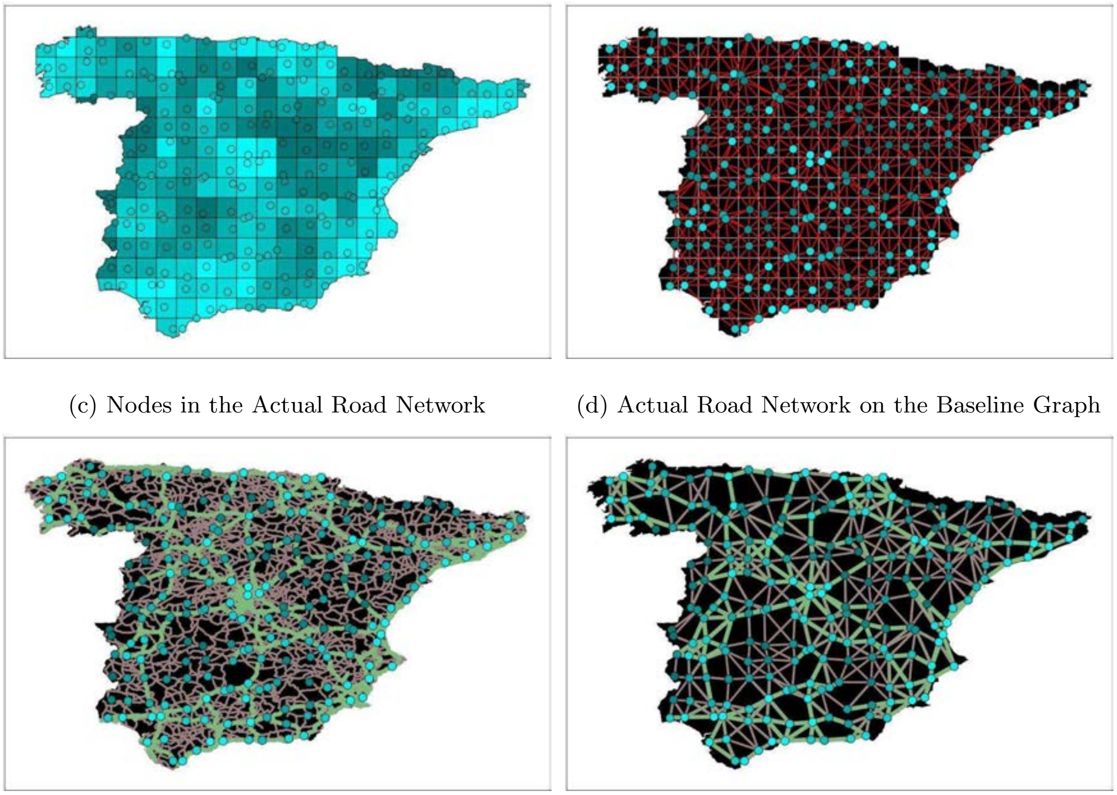
(c) Nodes in the Actual Road Network

(d) Actual Road Network on the Baseline Graph

Notes: Panel (a) shows total population from GPW aggregated into 50 km cells. Panel (b) shows the nodes $\mathcal{J}$ corresponding to the population centroids of each cell in Panel (a), reallocated to their closest point on the actual road network, and the edges $\mathcal{E}$ corresponding to all the vertical and diagonal links between cells. Panel (c) shows the centroids and the actual road network. Green segments correspond to national roads, red segments are all other roads, and the width of each segment is proportional to the number of lanes. Panel (d) shows the same centroids and the edges as the baseline graph in Panel (b), where each edge is weighted proportionally to the average number lanes on the cheapest path between each pair of nodes on the road network. The color shade ranges from red to green according to the fraction of the shortest path traveled on a national road.

Finally, the (d) panels show the infrastructure in the discretized road network. Each of the edges from the (b) panels is now assigned a width depending on the average number of lanes, $lanes_{jk}$, and a color ranging from red to green depending on the likelihood of using a national road, nat_{jk}. The width and color scale are the same as in panel (c). When no direct link from j to k is identified by our procedure, no edge is shown. The resulting discretized networks on the baseline grids clearly mirror the actual road networks for both countries, but they are now expressed in terms of the nodes and edges of our model and therefore allow us to quantify it.

Figure 4: Discretization of the Spanish Road Network



(a) Population on the Discretized Map (b) Nodes and Edges in the Baseline Graph

(c) Nodes in the Actual Road Network (d) Actual Road Network on the Baseline Graph

Notes: Panel (a) shows total population from GPW aggregated into 50 km cells. Panel (b) shows the nodes $\mathcal{J}$ corresponding to the population centroids of each cell in Panel (a), reallocated to their closest point on the actual road network, and the edges $\mathcal{E}$ corresponding to all the vertical and diagonal links between cells. Panel (c) shows the centroids and the actual road network. Green segments correspond to national roads, red segments are all other roads, and the width of each segment is proportional to the number of lanes. Panel (d) shows the same centroids and the edges as the baseline graph in Panel (b), where each edge is weighted proportionally to the average number lanes on the cheapest path between each pair of nodes on the road network. The color shade ranges from red to green according to the fraction of the shortest path traveled on a national road.

5.2 Parametrization

We discuss the specific parametric assumptions to implement the general model described in Section 3.

Preferences and Technologies The individual utility over traded and non-traded goods defined in (1) is assumed to be Cobb-Douglas,

$$U = c^{\alpha} h^{1-\alpha},$$

while the aggregator of traded goods (2) is CES:

$$C_j = \left(\sum_{n=1}^{N} \left(C_j^n \right)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tag{21}$$

where $\sigma > 0$ is the elasticity of substitution. Labor is the only factor of production and the production technologies (3) are assumed to be linear:

$$Y_j^n = z_j^n L_j^n.$$

We need to impose values to the preference parameters (α, σ). We assume $\alpha = 0.4$ to match a standard share of non-traded goods in consumption and $\sigma = 5$ which corresponds to a central value of the demand elasticities reported by Head and Mayer (2014) across estimates from the international trade literature. As we discuss below, the calibrated model gives a reasonable prediction for the distance elasticity of trade, which is typically closely linked to σ in existing studies.

Labor Mobility We undertake the entire analysis for the case in which labor is fixed and for the case in which it is perfectly mobile.

Transport Technology We adopt the log-linear transport technology (10). Under this assumption we must parametrize the congestion parameter β, the parameter γ capturing the return to infrastructure investments, and the frictions δ_{jk}^{τ}.

As discussed in Section 3.1, the congestion parameter β admits several interpretations. Here, we associate congestion in the model with the impact of traffic on speed on actual roads, under the assumptions that shipments translate linearly into traffic, and that lower speed translates linearly into higher per-unit shipping cost. Wang et al. (2011) review and estimate a standard class of traffic density-speed relationship from traffic flow theory and transportation engineering. As detailed in Appendix C, we choose β such that the relationship between flows and inverse-shipping costs in our model matches the empirical relationship between traffic density and speed reported in their paper. As a result of this step, we obtain $\beta = 1.245$.[47]

Having set β, we perform the entire analysis, including the calibration with fixed and mobile labor, for values of γ that span convex and non-convex cases: $\gamma = \{0.5 \times \beta, \beta, 1.5 \times \beta\}$.

Geographic Trade Frictions We also need to calibrate the matrix of trade frictions δ_{jk}^{τ} applying to the transport technology (10). As far as we know, data on trade flows within countries is not readily available at a reasonable level of spatial disaggregation for almost any of the countries in our data. Therefore, trade flows are not observed across the cells in our discretization; if that data were available, δ_{jk}^{τ} could be backed out for each pair of links as part of our calibration to rationalize the observed trade flows as an equilibrium outcome.

[47]To preserve the global convexity of the optimal flows problem, we adopt the log-linear specification of the transport technology rather than the logistic relationship assumed in Wang et al. (2011).

To sidestep this shortcoming we follow the standard approach of assuming that δ^τ_{jk} is a function of distance,

$$\delta^\tau_{jk} = \delta^\tau_0 dist_{jk}^{\delta^\tau_1}. \tag{22}$$

We set δ^τ_0 such that the model matches the share of total intra-regional trade in total intra-national trade of 39% reported by Llano et al. (2010) using average flows from 1995 to 2005 across Spanish regions. Because this ratio is only available for Spain, we undertake the calibration of δ_0 using our model predictions for Spain, and then, using this estimate, we construct δ^τ_{jk} across all cells in each country. In Table A.3 in Appendix C, we report the mean and standard deviation of the intra-regional trade share across the countries in our data, as well as the calibrated δ^τ_0, for each value of γ and assumption on labor mobility. The average intra-regional trade share in total domestic trade is often close to 50%.

This approach to calibrating geographic frictions within countries is in the spirit of Ramondo et al. (2012), who study a model featuring within-country trade without access to within-country trade data except for one country (in their case, the U.S.).[48] They jointly set $(\delta^\tau_0, \delta^\tau_1)$ to target the elasticity of trade with respect to distance from a standard gravity equation, as well as the share of intra-regional trade in domestic trade within the U.S., and then apply these coefficients to all other regions in their data. As we discuss in Appendix C, the coefficient δ^τ_1 has approximately no impact on the elasticity of shipping costs with respect to distance in our model, and therefore it has close to no impact on the trade-distance elasticity recovered from a standard gravity regression run on data generated by our model. Therefore, we normalize $\delta^\tau_1 = 1$.[49]

While the trade distance elasticity is rather insensitive to δ_1, it is sensitive to both β and σ. We have calibrated these parameters to match external sources, but we note that the calibrated model makes reasonable predictions for the trade-distance elasticity. As reported in Table A.3 in Appendix C, across countries the within-country trade-distance elasticity is centered around 1.1. A trade-distance elasticity around one corresponds to the typical value of existing estimates on both intra-national and inter-national trade data as summarized by Ramondo et al. (2012).

Productivities and Endowments We must impose values for the productivities z^n_j and the endowment of non-traded services H_j. In the case with perfect labor mobility we interpret $\left(L^{obs}_j, GDP^{obs}_j\right)$ as outcomes of the planner's solution for the optimal allocation and optimal flows problems discussed in Section 3.2 taking the observed network I^{obs}_{jk} as given, and use this information to back out the fundamentals $\left(z^n_j, H^n_j\right)$. In the case with fixed labor, we interpret GDP^{obs}_j as the outcome of the planner solution and use this information to back out the productivities z^n_j, normalizing $H_j = 1$ and setting the planner's weights $\omega_j = 1$ everywhere.

[48] In their formulation, Ramondo et al. (2012) also include an international border effect, which is not present here because we implement the analysis separately for each country.

[49] We experimented with a range of alternative values for δ^τ_1, and we always found that this parameter has no effect on the calibration and counterfactuals other than a re-scaling of the calibrated value of δ^τ_0, consistent with our discussion in Appendix C.

Since our data only includes aggregate measures of economic activity for each cell, we assume that each location produces only one tradable good. We allow for N different sectors: $N-1$ "industrial" goods, and one "agricultural" good. We assume that each of the $N-1$ industrial goods is produced in each of the $N-1$ cells with the largest observed population, and that the agricultural good is produced by all the remaining cells.[50] As a benchmark, we assume 10 different sectors ($N = 10$) and explore the robustness of this assumption to alternative values of N at the end of this section ($N = 5$ or $N = 15$). In geographically small countries where $N > J$ we set the number of goods equal to the number of locations, $N = J$.

This approach leaves us with J productivity parameters z_j, each corresponding to the productivity of a different location. Given the observed infrastructure I_{jk}^{obs} and the previous parameter choices, we choose each location's productivity and supply of non-traded goods such that, taking the observed network I_{jk}^{obs} as given, the planner's solution to the optimal allocation and optimal flows problems from Definition 2 reproduces the observed value-added and population as an outcome.[51] The model solution readily yields the level of population in each location, L_j. As for the model's prediction for GDP, we invoke the second welfare theorem from Proposition 4 to recover the prices in the observed allocation as the multipliers of the various constraints in the planner's problem.[52]

The various panels in Figure A.11 in Appendix C show the results of the calibration for the case of $\gamma = \beta$ (similar relationships hold for alternative values of γ). Panels (a) and (b) contrast the model-implied population share and income share of each location against the data, over all locations in the 25 countries. Except for very few locations, both population and income shares are matched with high precision. Panels (c) and (d) show the calibrated fundamentals (productivity and endowment of non-traded services per capita) in the vertical axes against income and population shares in the data, respectively, for the case with labor mobility. The calibration implies higher productivity and slightly lower supply of non-traded goods per capita in more populated places. Panel (e) shows a similar positive relationship between productivity and income share in the calibration of the model with fixed labor.

[50] Because data on industry-level value added or trade flows within countries is not always available, the approach in many of the related studies cited in the literature review has been to assume a pattern of specialization where each location produces a different product, i.e. the Armington assumption. Here, when $N = J$ the production structure corresponds to that assumption. While assuming Armington is sensible at somewhat higher levels of aggregation, it is arguably less appealing for the high geographic resolution (50 km x 50 km cells) that we consider.

[51] Based on the discussion of the preceding section, we first implement this step for Spain, where we jointly calibrate $\{z_j, H_j\}$ and δ_0^τ. For the remaining countries, we apply that δ_0^τ, but still back out $\{z_j, H_j\}$ using each country's income and value-added distribution.

[52] More specifically, in the solution of the planner's problem each location's value added is $P_j^{n(j)} z_j L_j + P_j^H H_j + \sum_n \sum_{k\in\mathcal{N}(j)} \left[P_k^n - P_j^n \left(1+\tau_{jk}\left(Q_{jk}^n, I_{jk}^{obs}\right)\right)\right] Q_{jk}^n$, where $n(j)$ denotes the good produced by location j, P_j^n is the price of good n in location j (i.e., multiplier of the flows constraint for good n in j in the planner's problem), and P_j^H is the price of non-traded services in sector j (i.e., the multiplier of the availability of non-traded goods constraint in the planner's problem). This step assumes that value added in the transport sector is empirically accounted to the exporting node.

Cost of Building Infrastructure To implement the optimal transport network in counterfactual scenarios we must parametrize the cost of infrastructure along each edge, δ^I_{jk}. We follow two approaches. In the first approach, we interpret the observed infrastructure I^{obs}_{jk} as the result of the full planner's problem. We do so under the assumption that $\delta^I_{jk} = \delta^I_{kj}$, so that it is equally costly to build in either direction, and that $I^{obs}_{jk} = I^{obs}_{kj}$, implying that infrastructure applies equally in either direction. In this case the observed network, I^{obs}_{jk}, is consistent with the planner's first-order condition for I_{jk} in (12) under the assumption that $\underline{I}_{jk} = 0$. Imposing symmetry on that first-order condition we then recover the cost of infrastructure as function of outcomes from the calibrated model (see Appendix A.2). We refer to this measure as the "FOC-based" measure of building costs, $\delta^{I,FOC}_{jk}$.

Our second approach is agnostic about whether the observed network results from any sort of optimization by a central authority, but takes a stand about how the building costs depend on geographic features. Specifically, we rely on data from Collier et al. (2016), who estimate highway building costs from more than three thousand World-Bank investment projects across the world, and then relate these costs to a host of geographic and non-geographic frictions.[53] We assume that δ^I_{jk} is a function of two geographic features included in their study, distance and ruggedness of the terrain. We refer to this building-cost measure as the "geographic" measure, $\delta^{I,GEO}_{jk}$. In our notation, their estimates imply:

$$\ln\left(\frac{\delta^{I,GEO}_{jk}}{dist_{jk}}\right) = \ln\left(\delta^I_0\right) - 0.11 * (dist_{jk} > 50km) + 0.12 * \ln\left(rugged_{jk}\right), \tag{23}$$

where $dist_{jk}$ is the distance between j and k and $rugged_{jk}$ is the average over the ruggedness in locations j and k.[54] This expression implies that it is more costly to build on rugged terrain, but less costly per kilometer to build on longer links. We assume that the elasticity of building costs with respect to features of the terrain is the same across all countries, but that the constant δ^I_0 may be country-specific.

These steps give two alternative measurements of δ^I_{jk} up to scale in each country. In the case of $\delta^{I,FOC}_{jk}$ the scale corresponds to the multiplier μ of the planner's resource constraint (see (A.3) in Appendix A.2), and in the case of $\delta^{I,GEO}_{jk}$ it corresponds to δ^I_0 in (23). In either case, the network-building constraint (7) must be satisfied. Hence, we set $K = 1$ in every country and re-scale each of the two measures of δ^I_{jk} to satisfy the network-building constraint with equality in each country.

[53] The investment projects in their data are concentrated in low- and middle-income countries, of which three (Lithuania, Georgia, and Macedonia) are in our data. The coefficients from their study introduced in our equation (23) correspond to the average of the coefficients over the distance dummy and the ruggedness index across the 6 specifications in Tables 4 and 5 of their paper.

[54] We use elevation data from the ETOPO1 Global Relief Model. The ETOPO1 dataset corresponds to a 1 arc-minute degree grid. We construct ruggedness for each cell as the average ruggedness across the 900 arc-minute cells from the ETOPO1 dataset contained in each 0.5 arc-degree cell in our discretized maps. We use the standard ruggedness index by Riley et al. (1999). Letting $\mathcal{J}^{etopo}(j)$ be the set of cells in ETOPO1 contained in each cell $j \in \mathcal{J}$ of our discretization and $\mathcal{N}^{etopo}(i)$ be the 8 neighboring cells to each cell in ETOPO1, this index is defined as: $rugged_j = \left(\sum_{i\in\mathcal{J}^{etopo}(j)} \sum_{k\in\mathcal{N}^{etopo}(i)} (elev_i - elev_k)^2\right)^{1/2}$; i.e., it is the standard deviation of the difference in elevation across neighboring cells. Then, we define $rugged_{jk}$ in (23) as $rugged_{jk} = \frac{1}{2}\left(rugged_j + rugged_k\right)$.

5.3 Optimal Expansion and Reallocation

We simulate two types of counterfactuals. First, we measure the aggregate gains from the optimal expansion of the observed road network within each country. For that, we assume that the total resources K are increased by 50% relative to the observed network, constraining the planner to build on top of the existing network, I_{jk}^{obs}. I.e., in the notation of restriction (iii) in definitions 1 and 2, $\underline{I}_{jk} = I_{jk}^{obs}$. Second, we measure the potential losses due to misallocation of current roads within each country. For that, we assume that the total resources K are the same as in the observed network, without constraining the planner to build on top of the existing network, I_{jk}^{obs}. I.e., in the notation of restriction (iii) in Definitions 1 and 2, $\underline{I}_{jk} = 0$.[55]

In short, the first "optimal expansion" counterfactual amounts to optimally expanding the network on top of what is already observed. In turn, the second "optimal reallocation" counterfactual amounts to optimally reallocating the existing roads or, equivalently, to building the globally optimal network employing the same amount of resources as those used to build the observed network. The first counterfactual is clearly more policy-relevant, as it prescribes where new roads should be built and yields the aggregate gains of those investments. The second counterfactual is unfeasible in reality, but it gives a sense of the losses from misallocation of existing roads.

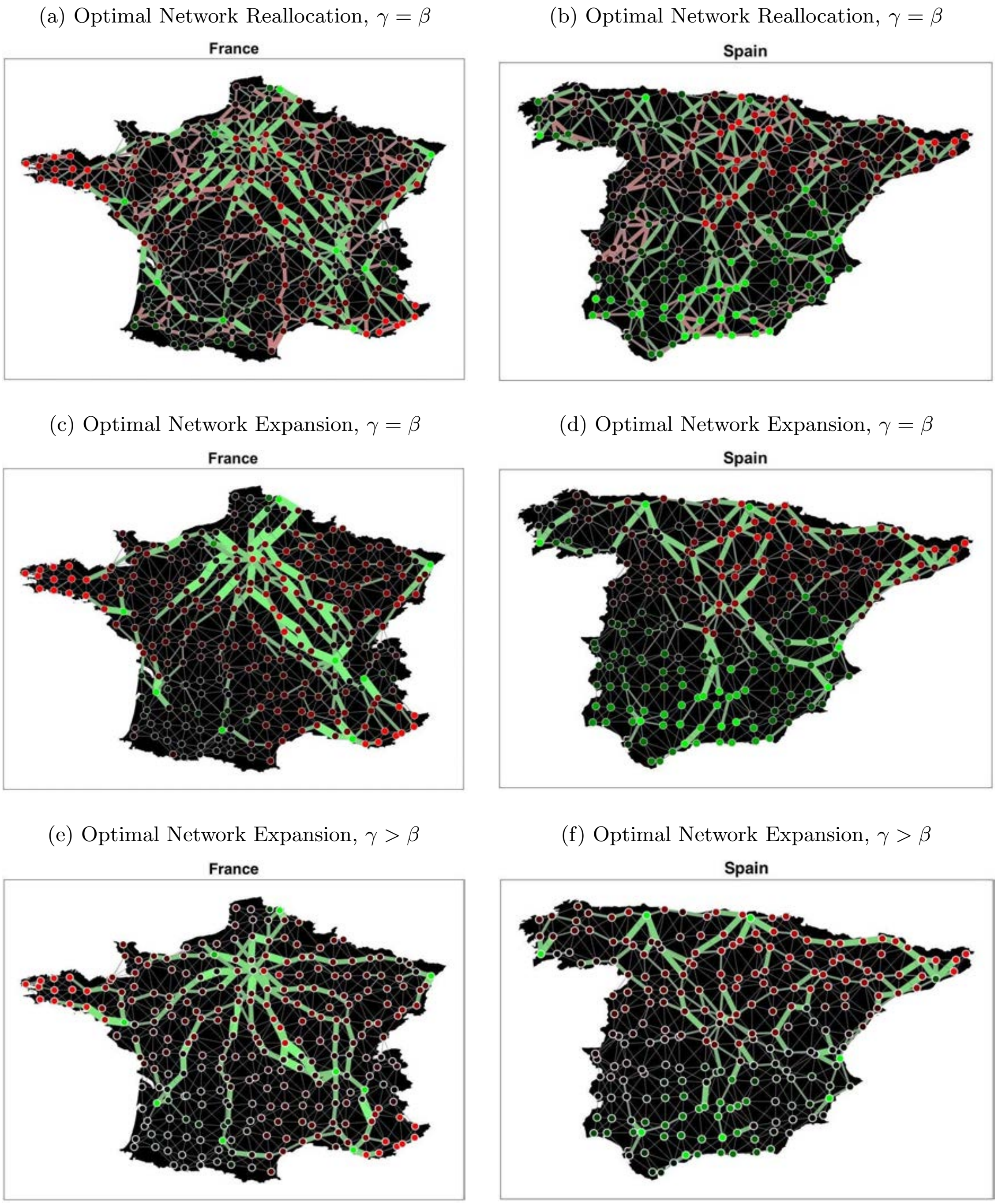
We implement the optimal expansion under the two measures of building costs, the FOC-based measure $\delta_{jk}^{I,FOC}$ and the geographic measure $\delta_{jk}^{I,GEO}$. The optimal reallocation is only meaningful under the geographic measure, since, by construction, the observed network is optimal and cannot be improved under $\delta_{jk}^{I,FOC}$. We implement each of these three counterfactuals for each of the three values $\gamma = \{0.5\beta, \beta, 1.5\beta\}$, assuming both fixed and mobile labor, separately for each of the 25 countries. We re-calibrate the model for each value of γ, assumption on labor mobility, and country following the steps from the previous section.

Regional Impact within Countries We inspect first the within-country regional implications for two of the largest countries in our data, Spain and France. Figure 5 depicts the pattern of investment and population change for counterfactuals under the geographic measure of building costs, $\delta^{I,GEO}$. Panels (a) and (b) show the optimal reallocation and panels (c) and (d) show the optimal expansion when $\gamma = \beta$ (convex case). Panels (e) and (f) reproduce the optimal expansion assuming $\gamma > \beta$ (non-convex case). All the figures correspond to assuming mobile labor. The thickness of each link increases with the absolute value of the investment, defined as the difference between the counterfactual and the observed infrastructure, $I_{jk}^{*} - I_{jk}^{obs}$. In the reallocation counterfactual, links with negative investment, $I_{jk}^{*} - I_{jk}^{obs} < 0$, are shown in red, while all other links are shown in green. In turn, green nodes denote positive population change, and red nodes denote negative population change. Brighter nodes represent a larger absolute value of population change.

In the optimal reallocation counterfactual, we observe positive investments radiating away from

[55] In every case, we set the upper bound on infrastructure, $\overline{I}_{jk}$, to be 50% above the largest level of infrastructure observed in each country.

Figure 5: Optimal Network Reallocation and Expansion, $\delta^I = \delta^{I,GEO}$



(a) Optimal Network Reallocation, $\gamma = \beta$

(b) Optimal Network Reallocation, $\gamma = \beta$

(c) Optimal Network Expansion, $\gamma = \beta$

(d) Optimal Network Expansion, $\gamma = \beta$

(e) Optimal Network Expansion, $\gamma > \beta$

(f) Optimal Network Expansion, $\gamma > \beta$

Notes: The width and brightness of each link is proportional to the difference between the optimal counterfactual network and the observed network, $I^*_{jk} - I^{obs}_{jk}$, for each link $jk \in \mathcal{E}$ shown in panel (b) of Figures 3 and 4. The color scale is the same as in Figure 3. In the misallocation counterfactuals, red links represent negative investment. Brighter green (red) nodes represent larger population increase (decrease).

some areas with higher economic activity in the case of France, but a more dispersed investment pattern in Spain. As we compare panels (a) and (b) with panels (c) and (d), we observe similar investment patterns in the optimal reallocation and expansion counterfactuals within each country: the links identified as having too much infrastructure, shown in red in panels (a) and (b), typically feature no expansion in panels (c) and (d). The comparison between panels (c)-(d) and panels (e)-(f) reveals that under increasing returns to infrastructure the optimal road expansion follows a similar pattern as in the convex case, but is more sparse and concentrated on fewer roads, in tune with our results in Proposition 3.

Despite the different investment patterns, population is reallocated to the same set of regions within each country across the counterfactuals. Due to the labor mobility constraint in the planner's problem, changes in labor are perfectly correlated with changes in consumption of traded commodities per worker, c_j.[56] For the cases without labor mobility, there is a similar consistency across the counterfactuals in the changes in consumption of traded commodities per capita c_j across locations.

What observable characteristics make specific regions more likely to receive infrastructure or to grow? Is growth correlated with receiving infrastructure? To answer these questions, we inspect, across the 25 countries, how a few typically observable regional characteristics map to infrastructure investment and population growth using data from the counterfactuals. Table 1 reports results from regressions of infrastructure and population growth on each location's initial population, income per capita, consumption of traded goods per capita and level of infrastructure, and on whether the location produces differentiated products. We report here results corresponding to the case where $\gamma = \beta$, but note that the results are very similar under the alternative values of γ or assuming fixed labor (and using the change in consumption per capita instead of population as dependent variable).

The odd columns imply that, regardless of the measure of building costs, optimal road investments are directed to locations with initially lower levels of infrastructure, capturing decreasing marginal aggregate welfare gains from infrastructure in specific links. Under the geographic measure of building costs in columns (1) and (3), optimal road investments are also more intensely directed to locations with initially higher levels of population and consumption per worker, as well as to producers of differentiated goods. However, these patterns are not present under the FOC-based measure of building costs. Importantly, in every case, the few variables in the regression have reasonable explanatory power (R^2 in the order of 30-40%). Therefore, even though the model implies a complex mapping from the fundamentals to the investments, observable features of each location guide a considerable fraction of the optimal investment decisions.

Looking at population changes, the even columns imply that the handful of variables in the regression explain between about 60% and 90% percent of the population changes. But, perhaps surprisingly, infrastructure growth in a location does not show up as an important determinant. Therefore, the model suggests that, in the context of a centrally and optimally planned infrastructure expansion, it is not necessarily true that the locations receiving more investments are also

[56] The labor mobility constraint (vi) from Definition 2 implies $\alpha \Delta \ln c_j = (1-\alpha)\, \Delta \ln L_j + \Delta \ln u$, where $\Delta \ln x$ denotes the difference in the log of variable x between the counterfactual and calibrated allocation.

Table 1: Optimal Infrastructure Investment, Population Growth and Local Characteristics

	Reallocation ($\delta = \delta^{I,GEO}$)		Expansion ($\delta = \delta^{I,GEO}$)		Expansion ($\delta = \delta^{I,FOC}$)	
Dependent variable:	Investment	Pop. Growth	Investment	Pop. Growth	Investment	Pop. Growth
	(1)	(2)	(3)	(4)	(5)	(6)
Population	0.308***	0.002	0.104***	0.002**	0.004	0.002**
Income per Capita	0.127	0.003	0.007	-0.002	-0.020	0.031**
Consumption per Capita	0.290**	-0.143***	0.179***	-0.134***	0.130	-0.179***
Infrastructure	-0.362***	-0.003	-0.195***	-0.001	-0.067**	0.000
Differentiated Producer	0.271***	0.017***	0.133***	0.028***	-0.099***	0.031***
R^2	0.38	0.56	0.32	0.65	0.38	0.90

Each column corresponds to a different regression pooling all locations across the 25 countries assuming $\gamma = \beta$, mobile labor, $\delta = \delta^{I,GEO}$, and N=10. All regressions include country fixed effects. Standard errors are clustered at the country level. ***=1% significance, **=5%, *=10%. Dependent variables: Investment is defined as $\Delta \ln \overline{I_j}$, where $\overline{I_j} = \frac{1}{\#\mathcal{N}(j)} \sum_{k \in \mathcal{N}(j)} I_{jk}$ is the average level of infrastructure across all the links of location j, and population growth is defined as $\Delta \ln L_j$, where $\Delta \ln x$ denotes the difference between the log of variable x in the counterfactual and in the calibrated allocation. Independent variables: all correspond to the log of the level of each variable in the calibrated model. Population and income per capita are the two outcomes matched by the calibration. Consumption per capita corresponds to traded goods, c_j in the model. Infrastructure is the average infrastructure of each location, $\overline{I_j}$. Differentiated producer is a dummy for whether the location is a producer of differentiated goods in the calibration.

those more likely to grow.[57]

Instead, two other variables have a significant relationship with regional growth: consumption of traded goods per capita and whether the location is a producer of differentiated goods. Consumption per capita is a strong determinant, with a negative elasticity of growth with respect to initial consumption in the order of 13%-18%. If consumption per capita was excluded, then the coefficient on income per capita would become negative and significant, with a negative elasticity of growth with respect to income per capita of 10% across the three counterfactuals. Hence, the impact of initial income on growth in the optimal investment plan operates through the level of consumption.

This reallocation pattern reflects that the goal of the optimal investments is to reduce variation in the marginal utility of consumption of traded commodities across locations. Since changes in population and consumption per capita between the counterfactual and initial allocation are perfectly correlated, the optimal investment plan leads to an increase in consumption of traded commodities in locations where consumption per capita is initially low. We conclude that the optimal investment in infrastructure reduces spatial inequalities, although different assumptions on building costs imply different ways of achieving this goal by changing the optimal placement of infrastructure, as implied by our previous discussion.

[57] The empirical literature on the impacts of trade costs on regional outcomes referenced in Section 2, and summarized by Donaldson (2015), does not always find that improvements in a location's market access leads to an increase in that location's economic activity. For example, Faber (2014) finds lower growth in peripheral counties that gained market access relative to other counties that did not improved access in the context of China's expansion of its National Trunk System.

Table 2: Average Welfare Gains Across Countries

Returns to Scale:	$\gamma = 0.5\beta$		$\gamma = \beta$		$\gamma = 1.5\beta$	
Labor:	Fixed	Mobile	Fixed	Mobile	Fixed	Mobile
Optimal Reallocation						
$\delta = \delta^{I,GEO}$	3.2%	3.0%	4.6%	4.8%	5.6%	6.6%
Optimal Expansion						
$\delta = \delta^{I,GEO}$	3.9%	3.5%	5.8%	5.7%	7.2%	7.9%
$\delta = \delta^{I,FOC}$	1.2%	1.0%	3.4%	5.8%	11.3%	12.4%

Each element of the table shows the average welfare gain in the corresponding counterfactual across the 25 countries.

Aggregate Impact Across Countries We conclude with the aggregate welfare effects. Table 2 shows the average welfare gain for each counterfactual across all 25 countries in our dataset. Tables A.4 and A.5 in Appendix C show the results for each country with fixed and mobile labor, respectively.

Starting from the case $\gamma = \beta$, we find average welfare gains across countries of between 3% and 6%, depending on the type of counterfactual, whether labor is allowed to be mobile, and whether building costs are measured according to the geographic or FOC-based measure. The average gains are increasing in the returns to scale, γ, particularly so under the FOC-based measure of building costs.

These effects vary considerably across countries. Each panel of Figure 6 shows the welfare gain across countries in each counterfactual for the case of $\gamma = \beta$, under both fixed and mobile labor, against each country's real income per capita. We see considerable variation in the gains across countries, ranging from around 2% to 15%. We also find negative relationships between income per capita and the welfare gains from either optimally expanding or reallocating current roads, suggesting larger returns to infrastructure and larger misallocation of existing roads in poorer economies.

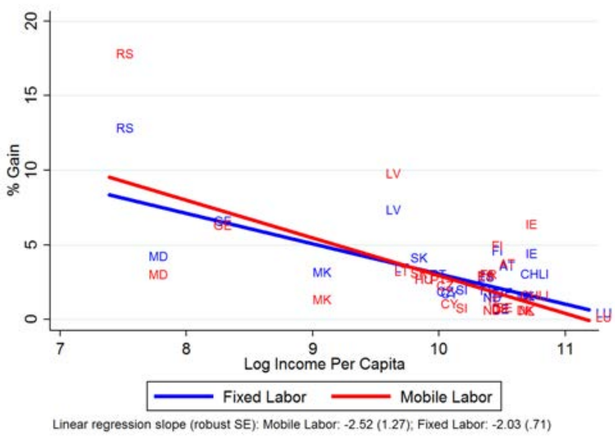
This distribution of welfare gains across countries is very stable regardless of the parametrization of γ, the assumption on labor mobility, the parametrization of the building costs δ^I, or the type of counterfactual. For example, across all the parametrizations of γ and labor mobility, the correlation between the gains from optimally expanding the network under the two measures of building costs, $\delta^{I,GEO}$ and $\delta^{I,FOC}$, is between 0.88 and 0.98.[58] Therefore, the answers to the questions of which countries would gain more from optimally expanding their current road networks and which countries suffer larger losses from misallocation of current roads is robust across all these cases.

Robustness to the Number of Sectors The analysis was implemented assuming $N = 10$ sectors. We also implement the calibration and counterfactuals, under both mobile and fixed labor,

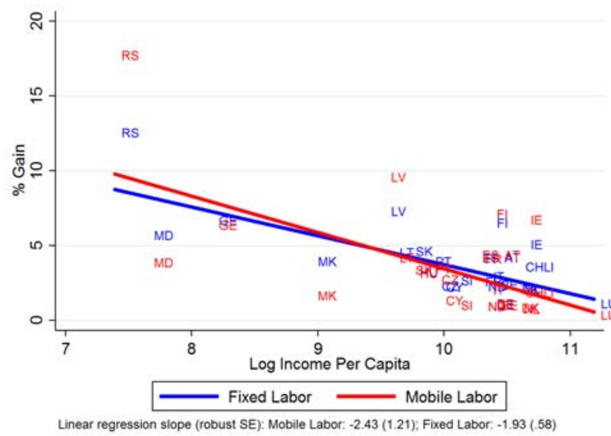
[58]In the benchmark case of $N = 10$, for each country we run 18 counterfactuals spanning the assumptions on γ, δ^I, type of counterfactual (expansion or reallocation), and labor mobility. Across all pairwise comparisons of these 18 cases, the lowest correlation in welfare gains across countries is 0.79.

Figure 6: Gains from Optimal Reallocation and Expansion and Income Per Capita

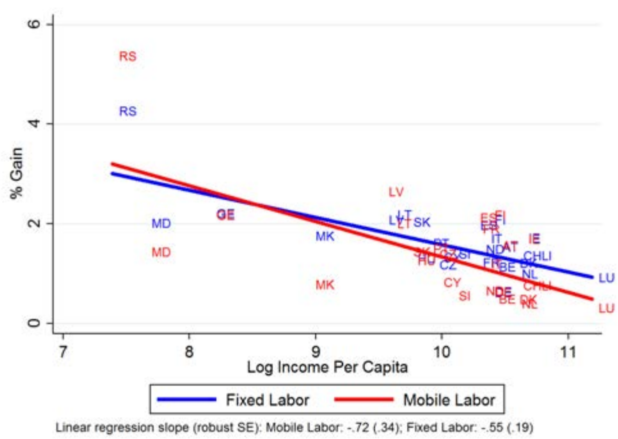
(a) Optimal Reallocation with $\delta^{I,GEO}$



(b) Optimal Expansion with $\delta^{I,GEO}$



(c) Optimal Expansion with $\delta^{I,FOC}$



Notes: Each figure displays the % welfare gains across countries in each counterfactual against each country's log-income per capita, for the case $\gamma = \beta$. The same patterns are present for $\gamma = 0.5\beta$ and $\gamma = 1.5\beta$.

assuming either that $N = 5$ or $N = 15$, for the case $\gamma = \beta$. We find that both the regional impact within countries and the aggregate impact across countries are very similar to the benchmark with $N = 10$. Table A.6 in Appendix C reports the coefficients from columns (3) and (4) of Table 1, corresponding to the optimal expansion under calibrations that assume $N = 5$ or $N = 15$. In both cases, the patterns described above remain unchanged, and the magnitude of most of the coefficients does not exhibit large variation. Similarly, Table A.7 reproduces Table 2 for different values of N. The aggregate gains change little with the number of sectors. The correlation between the aggregate welfare effects across countries under $N = 5$ or $N = 15$ and under $N = 10$ is between 0.8 and 0.9 depending on the type of counterfactual and the assumption on labor mobility.

6 Conclusion

In this paper, we develop a framework to study optimal transport networks in spatial equilibrium models. The framework combines a neoclassical environment where each location is a node in a graph, an optimal transport problem subject to congestion in shipping, and an optimal network design. It nests commonly used neoclassical trade models and it allows for either fixed or mobile factors across space. We provide conditions such that the full planner's problem, involving the optimal flow of goods as well as the general-equilibrium and network-design problems, is globally convex and numerically tractable using standard numerical methods typically applied to tackle optimal transport problems.

In the application, we match the model to data on road networks and economic activity at a 0.5 x 0.5 degree spatial resolution across 25 European countries. Given the observed road network, the model reproduces the population and value added observed across the cells in the data. Using the calibrated model, we find larger gains from road expansion and larger losses from misallocation of current roads in lower-income countries. We also find that the optimal expansion of current road networks reduces regional inequalities within countries. These results hold consistently across different parametrizations.

The framework could serve as basis for other applications. For instance, it could be used to study political-economy issues associated with infrastructure, such as spatial competition among planning authorities. Our application was limited to European countries, but low-income economies are likely to benefit more from infrastructure investment. It is also well understood that systems of cities and transport networks are highly persistent;[59] the model could be used to study inefficient network lock-in due to existing investments corresponding to dated economic fundamentals. The empirical literature mentioned in Section 2 and summarized by Donaldson (2015) relies on exogenous sources of variation for the placement of infrastructure investments; the framework may be used to construct instruments for investments in transport infrastructure as function of observable regional characteristics. Finally, a number of forces such as commuting or dynamic adjustment were left out of our analysis. We believe these are all interesting avenues to pursue in future research.

References

Ahuja, R. K., T. L. Magnanti, and J. B. Orlin (1989). Chapter iv network flows. *Handbooks in operations research and management science 1*, 211–369.

Alder, S. (2016). Chinese roads in india: The effect of transport infrastructure on economic development. *Manuscript, Univ. North Carolina, Chapel Hill*.

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 (2016). The welfare effects of transportation infrastructure improvements. *Manuscript, Dartmouth and Yale*.

Allen, T., C. Arkolakis, and Y. Takahashi (2014). Universal gravity. *NBER Working Paper* (w20787).

[59] See Davis and Weinstein (2002) and Michaels and Rauch (2013).

Anderson, J. E. and E. Van Wincoop (2003). Gravity with gravitas: a solution to the border puzzle. *the american economic review 93*(1), 170–192.

Anderson, J. E. and E. Van Wincoop (2004). Trade costs. *Journal of Economic literature 42*(3), 691–751.

Arrow, K. J. and A. C. Enthoven (1961). Quasi-concave programming. *Econometrica: Journal of the Econometric Society*, 779–800.

Asturias, J., M. García-Santana, and R. Ramos Magdaleno (2016). Competition and the welfare gains from transportation infrastructure: Evidence from the golden quadrilateral of india.

Atkeson, A. and A. T. Burstein (2010). Innovation, firm dynamics, and international trade. *Journal of political economy 118*(3), 433–484.

Atkin, D. and D. Donaldson (2015). Who's getting globalized? the size and implications of intra-national trade costs. Technical report, National Bureau of Economic Research.

Banavar, J. R., F. Colaiori, A. Flammini, A. Maritan, and A. Rinaldo (2000). Topology of the fittest transportation network. *Physical Review Letters 84*(20), 4745.

Bartelme, D. (2015). Trade costs and economic geography: Evidence from the U.S. Technical report, University of Michigan.

Baum-Snow, N. (2007). Did highways cause suburbanization? *The Quarterly Journal of Economics*, 775–805.

Beckmann, M. (1952). A continuous model of transportation. *Econometrica: Journal of the Econometric Society*, 643–660.

Bernot, M., V. Caselles, and J.-M. Morel (2009). *Optimal transportation networks: models and theory*, Volume 1955. Springer Science & Business Media.

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

Boyd, S. and L. Vandenberghe (2004). *Convex optimization*. Cambridge university press.

Brandt, L., T. Tombe, and X. Zhu (2013). Factor market distortions across time, space and sectors in China. *Review of Economic Dynamics 16*(1), 39–58.

Burstein, A. and J. Cravino (2015). Measured aggregate gains from international trade. *American Economic Journal: Macroeconomics 7*(2), 181–218.

Caliendo, L., F. Parro, E. Rossi-Hansberg, and P.-D. Sarte (2014). The impact of regional and sectoral productivity changes on the U.S. economy. Technical Report 20168, National Bureau of Economic Research.

Carlier, G. (2010). Technical report, National Bureau of Economic Research.

Castells, A. and A. Solé-Ollé (2005). The regional allocation of infrastructure investment: The role of equity, efficiency and political factors. *European Economic Review 49*(5), 1165–1205.

Chandra, A. and E. Thompson (2000). Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system. *Regional Science and Urban Economics 30*(4), 457–490.

Chaney, T. (2014a). The network structure of international trade. *The American Economic Review 104*(11), 3600–3634.

Chaney, T. (2014b). Networks in international trade. In *The Oxford Handbook of the Economics of Networks*.

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.

Combes, P.-P. and M. Lafourcade (2005). Transport costs: measures, determinants, and regional policy implications for france. *Journal of Economic Geography 5*(3), 319–349.

Coşar, A. K. and B. Demir (2016). Domestic road infrastructure and international trade: Evidence from turkey. *Journal of Development Economics 118*, 232–244.

Costinot, A. and A. Rodríguez-Clare (2013). Trade theory with numbers: Quantifying the consequences of globalization, vol. 4 of handbook of international economics.

Davis, D. R. and D. E. Weinstein (2002). Bones, bombs, and break points: the geography of economic activity. *The American Economic Review 92*(5), 1269–1289.

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

Desmet, K. and E. Rossi-Hansberg (2014). Spatial development. *The American Economic Review 104*(4), 1211–1243.

Doll, C., H. van Essen, et al. (2008). Road infrastructure cost and revenue in europe. *Produced within the study Internalisation Measures and Policies for all external cost of Transport (IMPACT) P Deliverable 2.*

Donaldson, D. (2010). Railroads of the raj: Estimating the impact of transportation infrastructure. Technical report, National Bureau of Economic Research.

Donaldson, D. (2015). The gains from market integration. *Annual Review of Economics 7*(1), 619–647.

Donaldson, D. and R. Hornbeck (2016). Railroads and american economic growth: A market access approach. *The Quarterly Journal of Economics*, qjw002.

Duranton, G., P. M. Morrow, and M. A. Turner (2014). Roads and trade: Evidence from the us. *The Review of Economic Studies 81*(2), 681–724.

Eaton, J. and S. Kortum (2002). Technology, geography, and trade. *Econometrica 70*(5), 1741–1779.

Ekeland, I. (2010). Notes on optimal transportation. *Economic Theory 42*(2), 437–459.

Faber, B. (2014). Trade integration, market size, and industrialization: evidence from china's national trunk highway system. *The Review of Economic Studies*, rdu010.

Fajgelbaum, P. D., E. Morales, J. C. S. Serrato, and O. M. Zidar (2015). State taxes and spatial misallocation. *NBER Working Paper* (w21760).

Felbermayr, G. J. and A. Tarasov (2015). Trade and the spatial distribution of transport infrastructure.

Fernald, J. G. (1999). Roads to prosperity? assessing the link between public capital and productivity. *American Economic Review*, 619–638.

Feyrer, J. (2009). Trade and income–exploiting time series in geography. Technical report, National Bureau of Economic Research.

Figueroa, C., B. Fotsch, S. M. Hubbard, and J. Haddock (2013). Assessment procedures for paved and gravel roads. Technical report, Purdue University School of Civil Engineering.

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

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

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

Hsieh, C.-T. and E. Moretti (2015). Why do cities matter? local growth and aggregate growth. Technical Report 21154, National Bureau of Economic Research.

IADB (2013). Too far to export: Domestic transport costs and regional export disparities in latin america and the caribbean.

Kantorovich, L. V. (1942). On the translocation of masses. In *Dokl. Akad. Nauk SSSR*, Volume 37, pp. 199–201.

Lai, E. L.-C., H. Fan, and H. S. Qi (2015). Global gains from reduction of trade costs.

Limao, N. and A. J. Venables (2001). Infrastructure, geographical disadvantage, transport costs, and trade. *The World Bank Economic Review 15*(3), 451–479.

Llano, C., A. Esteban, J. Pérez, and A. Pulido (2010). Opening the interregional trade black box: The c-intereg database for the spanish economy (1995–2005). *International Regional Science Review*.

Maibach, M., C. Schreyer, D. Sutter, H. Van Essen, B. Boon, R. Smokers, A. Schroten, C. Doll, B. Pawlowska, and M. Bak (2013). Handbook on estimation of external costs in the transport sector.

Martincus, C. V., J. Carballo, and A. Cusolito (2017). Roads, exports and employment: Evidence from a developing country. *Journal of Development Economics 125*, 21–39.

Michaels, G. and F. Rauch (2013). Resetting the urban network: 117-2012.

Monge, G. (1781). *Mémoire sur la théorie des déblais et des remblais*. De l'Imprimerie Royale.

Nagy, D. (2016). City location and economic development. *Manuscript, CREI*.

Nesterov, Y. and A. Nemirovskii (1994). *Interior-point polynomial algorithms in convex programming*. SIAM.

Pascali, L. (2014). The wind of change: Maritime technology, trade and economic development.

Ramondo, N., A. Rodríguez-Clare, and M. Saborío-Rodríguez (2012). Trade, domestic frictions, and scale effects. Technical report, National Bureau of Economic Research.

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

Redding, S. J., G. Ahlfeldt, and D. Sturm (2016). A quantitative framework for evaluating the impact of urban transport improvements. *Manuscript, Princeton University*.

Redding, S. J. and E. Rossi-Hansberg (2016). Quantitative spatial economics. Technical report, National Bureau of Economic Research.

Redding, S. J. and M. A. Turner (2015). Transportation costs and the spatial organization of economic activity, vol. 5 of handbook of regional and urban economics.

Restuccia, D. and R. Rogerson (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics 11*(4), 707–720.

Riley, S., S. DeGloria, and R. Elliot (1999). A terrain ruggedness index that quantifies topographic heterogeneity. *Intermountain Journal of Sciences 5*(1–4), 23–27.

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

Santambrogio, F. (2015). Optimal transport for applied mathematicians. *Birkäuser, NY*.

Sotelo, S. (2016). Domestic trade frictions and agriculture. *Manuscript, University of Michigan*.

Swisher IV, S. (2015). Reassessing railroads and growth: Accounting for transport network endogeneity. Technical report, mimeo.

Takayama, A. (1985). *Mathematical economics*. Cambridge University Press.

Tay, R. and A. Churchill (2007). Effect of different median barriers on traffic speed. *Canadian Journal of Transportation 1*(1).

Trew, A. W. (2016). Endogenous infrastructure development and spatial takeoff.

Villani, C. (2003). *Topics in optimal transportation*. Number 58. American Mathematical Soc.

Wang, H., J. Li, Q.-Y. Chen, and D. Ni (2011). Logistic modeling of the equilibrium speed–density relationship. *Transportation research part A: policy and practice 45*(6), 554–566.

WorldBank (2011). *Africa's Transport Infrastructure: Mainstreaming Maintenance and Management*. World Bank Publications.

WorldEconomicForum (2016). *The Global Competitive Report, 2016-2017*. World Economic Forum Publications.

A Appendix to Section 3 (Model)

A.1 Planner's Problem

In this section we present the first-order conditions to the planner's problem. We refer to these first-order conditions in some of the characterizations in the text and in the proofs below.

Immobile Labor

The Lagrangian of the problem in Definition 1 is

$$\begin{aligned}
\mathscr{L} = & \sum_j \omega_j L_j U\left(c_j, h_j\right) - \sum_j P_j^C \left[c_j L_j - C_j^T \left(C_j^1, .., C_j^N\right)\right] - \sum_j P_j^H \left(h_j L_j - H_j\right) \\
& - \sum_j \sum_n P_j^n \left[C_j^n + \sum_{k \in \mathcal{N}(j)} \left(Q_{jk}^n + \tau_{jk}\left(Q_{jk}^n, I_{jk}\right) Q_{jk}^n\right) - F_j^n \left(L_j^n, \mathbf{V}_j^n, \mathbf{X}_j^n\right) - \sum_{i \in \mathcal{N}(j)} Q_{ij}^n\right] \\
& - \sum_j W_j \left[\sum_n L_j^n - L_j\right] - \sum_j \sum_l R_j^m \left[\sum_n V_j^{mn} - V_j^m\right] \\
& - \mu \left(\sum_j \sum_{k \in \mathcal{N}(j)} \delta_{jk}^I I_{jk} - K\right) + \sum_{j,k} \zeta_{jk}^{\underline{I}} \left(I_{jk} - \underline{I}_{jk}\right) + \sum_{j,k} \zeta_{jk}^{\overline{I}} \left(\overline{I}_{jk} - I_{jk}\right) \\
& + \sum_{j,k,n} \zeta_{jkn}^Q Q_{jk}^n + \sum_{j,n} \zeta_{jn}^L L_j^n + \sum_{j,n,l} \zeta_{jnl}^V V_j^{ln} + \sum_{j,n,l} \zeta_{jnl}^X X_j^{ln} + \sum_{j,n} \zeta_{jn}^C C_j^n + \sum_j \zeta_j^c c_j + \sum_j \zeta_j^h h_j
\end{aligned}$$

where P_j^C, P_j^H, P_j^N, W_j, R_j^l, μ, ζ_{jk}^I, ζ_{jkn}^Q, ζ_{jn}^L, ζ_{jnl}^V, ζ_{jn}^C, ζ_j^c, ζ_j^h are the multipliers of all constraints implied by (i)-(v) in Definition 1. The first-order conditions with respect to consumption and production are:

$$\begin{aligned}
&[c_j] && \omega_j L_j U_C\left(c_j, h_j\right) + \zeta_j^c = P_j^C L_j \\
&[h_j] && \omega_j L_j U_H\left(c_j, h_j\right) + \zeta_j^h = P_j^H L_j \\
&\left[C_j^n\right] && P_j^C \frac{\partial C_j^T}{\partial C_j^n} + \zeta_{jn}^C = P_j^n \\
&\left[L_j^n\right] && P_j^n \frac{\partial Y_j^n}{\partial L_j^n} + \zeta_{jn}^L = W_j \\
&\left[V_j^n\right] && P_j^n \frac{\partial Y_j^n}{\partial V_j^{ln}} + \zeta_{jnl}^V = R_j^l \\
&\left[X_j^n\right] && P_j^n \frac{\partial Y_j^n}{\partial X_j^{ln}} + \zeta_{jnl}^X = P_j^l
\end{aligned}$$

The first-order conditions with respect to flows is:

$$\left[Q_{jk}^n\right] \qquad - P_j^n \left(1 + \tau_{jk}^n + \frac{\partial \tau_{jk}^n}{\partial Q_{jk}^n} Q_{jk}^n\right) + P_k^n + \zeta_{jkn}^Q = 0$$

which, along with the complementary slackness condition for Q_{jk}^n, implies (8) in the main text.

Finally, the first order condition with respect to the network investment is

$$[I_{jk}] \qquad \sum_n P_j^n Q_{jk}^n \left(-\frac{\partial \tau_{jk}^n}{\partial I_{jk}}\right) + \left(\zeta_{jk}^{\underline{I}} - \zeta_{jk}^{\overline{I}}\right) = \mu \delta_{jk}^I$$

which, along with the complementary slackness condition for I_{jk}^n, implies (9) in the text.

Mobile Labor

The Lagrangian of the problem in Definition 2 is

$$
\begin{aligned}
\mathscr{L} = & \, u-\sum_j \tilde{\omega}_j L_j\left(u-U\left(c_j, h_j\right)\right)-W^L\left(\sum_j L_j-L\right) \\
& -\sum_j P_j^C\left[c_j L_j-C_j^T\left(C_j^1, .., C_j^N\right)\right]-\sum_j P_j^H\left(h_j L_j-H_j\right) \\
& -\sum_j \sum_n P_j^n\left[C_j^n+\sum_{k \in \mathcal{N}(j)}\left(Q_{jk}^n+\tau_{jk}\left(Q_{jk}^n, I_{jk}\right) Q_{jk}^n\right)-F_j^n\left(L_j^n, \mathbf{V}_j^n, \mathbf{X}_j^n\right)-\sum_{i \in N(j)} Q_{ij}^n\right] \\
& -\sum_j W_j\left[\sum_n L_j^n-L_j\right]-\sum_j \sum_l R_j^l\left[\sum_n V_j^n-V_j\right] \\
& -\mu\left(\sum_j \sum_{k \in \mathcal{N}(j)} \delta_{jk}^I I_{jk}-K\right)+\sum_{j,k} \zeta_{jk}^{\underline{I}}\left(I_{jk}-\underline{I}_{jk}\right)+\sum_{j,k} \zeta_{jk}^{\overline{I}}\left(\overline{I}_{jk}-I_{jk}\right) \\
& +\sum_{j,k,n} \zeta_{jkn}^Q Q_{jk}^n+\sum_{j,n} \zeta_{jn}^L L_j^n+\sum_{j,n,l} \zeta_{jnl}^V V_j^{ln}+\sum_{j,n,l} \zeta_{jnl}^X X_j^{ln}+\sum_{j,n} \zeta_{jn}^C C_j^n+\sum_j \zeta_j^c c_j+\sum_j \zeta_j^h h_j
\end{aligned}
$$

where, in addition to the previous notation for the multipliers, in the first line we have defined $\tilde{\omega}_j$ and W^L as the multipliers of constraints (vi) and (vii) in Definition 2.

The first-order conditions with respect to consumption of traded services $\left[C_j^n\right]$, factor allocation within locations $\left[L_j^n\right]$, $\left[V_j^n\right]$ and $\left[X_j^n\right]$, optimal transport $\left[Q_{jk}^n\right]$, and optimal investment $[I_{jk}]$ are the same as in the problem without labor mobility. The first-order conditions with respect to u and L_j are:

$$
\begin{aligned}
& [u] \qquad 1=\sum_j L_j \tilde{\omega}_j \\
& [L_j] \qquad P_j^C c_j+P_j^H h_j-\tilde{\omega}_j\left[U\left(c_j, h_j\right)-u\right]=W_j-W^L
\end{aligned}
$$

where from monotonicity of $U\left(c_j, h_j\right)$ it follows that

$$
U\left(c_j, h_j\right)= \begin{cases}u & \text{if } L_j>0, \\ 0 & \text{if } L_j=0.\end{cases}
$$

In addition, the first-order conditions with respect to consumption of traded and non-traded services, $[c_j]$ and $[h_j]$, are the same as in the problem without labor mobility replacing the planner's weights ω_j with the multipliers of the mobility constraint $\tilde{\omega}_j$. Combining $[L_j]$ with $[c_j]$ and $[h_j]$ gives the multiplier on the labor-mobility constraint. For populated locations:

$$
\tilde{\omega}_j=\frac{W_j-W^L}{U_C\left(c_j, h_j\right) c_j+U_H\left(c_j, h_j\right) h_j}.
$$

A.2 Symmetry in Infrastructure Investments

For the applications in Section 5 we impose symmetry in infrastructure levels, i.e., $I_{jk}=I_{kj}$. This section provides the first-order condition for I_{jk} in that case. The first-order condition with respect to I_{jk} is

$$
[I_{jk}] \qquad \sum_n P_j^n Q_{jk}^n\left(-\frac{\partial \tau_{jk}^n}{\partial I_{jk}}\right)+\sum_n P_k^n Q_{kj}^n\left(-\frac{\partial \tau_{jk}^n}{\partial I_{jk}}\right)+\left(\zeta_{jk}^{\underline{I}}-\zeta_{jk}^{\overline{I}}\right)=\mu\left(\delta_{jk}^I+\delta_{jk}^I\right). \tag{A.1}
$$

Assuming symmetry leaves all the remaining first-order conditions presented in Section A.1 unchanged. Under the log-linear specification (10) of the transport technology, the optimal infrastructure investment, conditional on $I_{jk} \in\left(\zeta_{jk}^{\underline{I}}, \zeta_{jk}^{\overline{I}}\right)$, is

$$I_{jk}^{*}=\left[\frac{\gamma}{\mu\left(\delta_{jk}^{I}+\delta_{kj}^{I}\right)}\left(\sum\nolimits_{n}\delta_{jk}^{\tau}P_{j}^{n}\left(Q_{jk}^{n}\right)^{1+\beta}+\sum_{n}\delta_{kj}^{\tau}P_{k}^{n}\left(Q_{kj}^{n}\right)^{1+\beta}\right)\right]^{\frac{1}{1+\gamma}}, \tag{A.2}$$

which, substituting for the optimal flows, yields:

$$I_{jk}^{*}=\left(\frac{\kappa}{\mu\left(\delta_{jk}^{I}+\delta_{kj}^{I}\right)}\left(\sum_{n:P_{k}^{n}>P_{j}^{n}}\left(\delta_{jk}^{\tau}\right)^{1/\beta}P_{j}^{n}\left(\frac{P_{k}^{n}}{P_{j}^{n}}-1\right)^{\frac{1+\beta}{\beta}}+\sum_{n:P_{j}^{n}>P_{k}^{n}}\left(\delta_{kj}^{\tau}\right)^{1/\beta}P_{k}^{n}\left(\frac{P_{j}^{n}}{P_{k}^{n}}-1\right)^{\frac{1+\beta}{\beta}}\right)\right)^{\frac{\beta}{\beta-\gamma}}. \tag{A.3}$$

Expressions (12) and (14) in Section 3 are analogous to these conditions when symmetry is not imposed. As discussed in Section 5.2, to build $\delta_{jk}^{I,GEO}$ we use (A.2) under the symmetry assumption $\delta_{jk}^{I}=\delta_{kj}^{I}$. Setting $I_{jk}^{*}=I_{jk}^{obs}$,$\delta_{jk}^{I,GEO}$ can be backed out as function of calibrated parameters, the observed network I_{jk}^{OBS}, and the equilibrium prices generated by the calibrated model. Note that, to generate these prices, we use the model calibrated given the network I_{jk}^{OBS}, as discussed in Section 5.2.

A.3 Proofs of the Propositions

Proposition 1. *(Convexity of the Planner's Problem) (i) Given the network $\{I_{jk}\}$, the joint optimal transport and allocation problem in the fixed (resp. mobile) labor case is a convex (resp.quasiconvex) optimization problem if $Q\tau_{jk}(Q,I_{jk})$ is convex in Q for all j and $k\in\mathcal{N}(j)$; and (ii) if in addition $Q\tau_{jk}(Q,I)$ is convex in both Q and I for all j and $k\in\mathcal{N}(j)$, then the full planner's problem including the network design problem from Definition (1) (resp. Definition (2)) is a convex (resp. quasiconvex) optimization problem. In either the joint transport and allocation problem, or the full planner's problem, strong duality holds when labor is fixed.*

Proof. Consider the planner's problem from Definition 1. We can write it as

$$\max_{\left\{C_j,\left\{C_j^n,\left\{Q_{jk}^n,I_{jk}\right\}_{k\in\mathcal{N}(j)}\right\}\right\}_{\forall j}} f=\sum_j \omega_j L_j U\left(\frac{C_j}{L_j},\frac{H_j}{L_j}\right)$$

subject to: (i) availability of traded commodities,

$$g_j^1=C_j-C_j^T\left(C_j^1,..,C_j^N\right)\leqslant 0 \text{ for all } j;$$

(ii) the balanced-flows constraint,

$$g_{jn}^2\equiv C_j^n+\sum_{k\in\mathcal{N}(j)}Q_{jk}^n\left[1+\tau_{jk}\left(Q_{jk}^n,I_{jk}\right)\right]-F_j^n\left(L_j^n,\mathbf{V}_j^n,\mathbf{X}_j^n\right)-\sum_{i\in\mathcal{N}(j)}Q_{ij}^n\le 0 \text{ for all } j,\ n;$$

(iii) the network-building constraint,

$$\sum_j\sum_{k\in\mathcal{N}(j)}\delta_{jk}^I I_{jk}\le K;$$

and conditions (iv)-(v) in the text. Since constraints (iii)-(v) are linear, we need f to be concave and g_j^1 and g_{jn}^2 to be convex. Since U is jointly concave in both its arguments, f is concave. $C_j\left(\left\{C_j^n\right\}\right)$ is concave, hence g_j^1 is convex. If $Q\tau_{jk}(Q,I)$ is convex in then g_{jn}^2 is the sum of linear and convex functions, hence it is convex. To show that this problem admits strong duality, a constraint qualification is required. Note first that constraints g_j^1 and g_{jn}^2 must hold with equality at an optimum and therefore can be substituted into the objective function. The remaining constraints (iii)-(v) are all linear and thus satisfy the Arrow-Hurwicz-Uzawa qualification constraint (Takayama (1985), Theorem 1.D.4). Hence, the global optimum must satisfy the KKT conditions and the duality gap is 0.[60]

[60] Despite having substituted constraints g_j^1 and g_{jn}^2 into the objective function, the multipliers for these constraints, P_j^C and P_j^n, can be recovered from the above KKT conditions such that $\omega_j U_C(c_j,h_j)=P_j^C$ and $P_j^C\partial C_j^T/\partial C_j^n=P_j^n$.

Consider now the planner's problem with labor mobility from Definition 2. Because U is homothetic, we can express it as $U = G(U_0(c,h))$, where G is an increasing continuous function and U_0 is homogeneous of degree 1. Therefore, imposing the change of variables $\tilde{u} = G^{-1}(u)$, the planner's problem can be restated as

$$\max \tilde{u}$$

subject to the convex constraints (i)-(v) and $L_j \tilde{u} \leqslant U_0(C_j, H_j)$. To make the latter constraint convex, let us denote $U_j = L_j \tilde{u}$ and replace $\tilde{u}$ in the objective function by $\min_{j|L_j>0}\left\{\frac{U_j}{L_j}\right\}$,[61] so that the problem becomes

$$\max_{C_j,\left\{C_j^n, L_j^n, \mathbf{V}_j^n, \left\{Q_{jk}^n, I_{jk}^n\right\}_{k\in\mathcal{N}(j)}\right\}, U_j, L_j} \min_{j|L_j>0}\left\{\frac{U_j}{L_j}\right\}$$

subject to the convex restrictions (i)-(v) above as well as

$$U_j \le U_0(C_j, H_j) \text{ for all } j.$$

The objective function is quasiconcave because U_j/L_j is quasiconcave and the minimum of quasiconcave functions is quasiconcave. In addition, all the restrictions are convex. Arrow and Enthoven (1961) then implies that the Karush-Kuhn-Tucker conditions are sufficient if the gradient of the objective function is different from zero at the candidate for an optimum, and here the gradient never vanishes.

□

Proposition 2. *(Optimal Network in Log-Linear Case) When the transport technology is given by* (10)*, the full planner's problem is a convex (resp. quasiconvex) optimization problem if $\beta \geqslant \gamma$. The optimal infrastructure is given by* (13) *implying that, in the absence of a pre-existing network (i.e., if $I^0_{jk} = 0$), then $I_{jk} = 0 \leftrightarrow P_k^n = P_j^n$ for all n.*

Proof. First, note that if $\beta \ge \gamma$ then $Q\tau(Q,I) \propto Q^{1+\beta} I^{-\gamma}$ is convex in $Q \in \mathbb{R}_+$ and $I \in \mathbb{R}_+$. To see that, note that the determinant of the Hessian of $Q^{1+\beta}I^{-\gamma}$ is $(1+\beta)\gamma(\beta-\gamma)Q^{2\beta}I^{-2(\gamma+1)}$, which is positive for $Q \in \mathbb{R}_+$ and $I \in \mathbb{R}_+$ if $\beta \ge \gamma \ge 0$. Next, from the first-order condition for optimal infrastructure (9), if the solution to the planning problem implies $I_{jk} = \underline{I}_{jk}$, so that there is no investment, then:

$$\begin{aligned}
\mu &\ge -\frac{1}{\delta^I_{jk}} \sum_n P_j^n Q_{jk}^n \left.\frac{\partial \tau_{jk}^n}{\partial I_{jk}}\right|_{I_{jk}=\underline{I}_{jk}} \\
&\ge \gamma \frac{\delta^\tau_{jk}}{\delta^I_{jk}} \frac{\sum_n P_j^n \left(Q_{jk}^n\right)^{1+\beta}}{\underline{I}_{jk}^{\gamma+1}} \\
&\ge \frac{\gamma(1+\beta)^{-\frac{1+\beta}{\beta}}}{\delta^I_{jk}\left(\delta^\tau_{jk}\right)^{\frac{1}{\beta}}} \frac{\sum_{n:P_k^n>P_j^n} P_j^n \left(\frac{P_k^n}{P_j^n}-1\right)^{\frac{1+\beta}{\beta}}}{\underline{I}_{jk}^{\frac{\beta-\gamma}{\beta}}},
\end{aligned}$$

where the second line follows from (10) and the third line follows from (11). Note that the last inequality is equivalent to $\underline{I}_{jk} \ge I^*_{jk}$ for I^*_{jk}defined in (14). Therefore, if $\underline{I}_{jk} < I^*_{jk}$ then $I_{jk} > \underline{I}_{jk}$ and $I_{jk} = I^*_{jk}$. Moreover, if there is any n such that $P_k^n \ne P_j^n$ then $I^*_{jk} > 0$.

□

Proposition 3. *(Tree Property) Assume that $\lim_{c\to 0^+} U_C(c,h) = \infty$. In the absence of a pre-existing network (i.e., $\underline{I}_{jk} = 0$), if the transport technology is given by* (10) *and satisfies $\gamma > \beta$, and if there is a unique commodity produced in a single location, then the optimal transport network is a tree.*

[61]Since the objective function is strictly increasing in $\tilde{u}$ and because $\tilde{u}$ only shows up in the constraints $L_j\tilde{u} \leqslant U_0(C_j, H_j)$ for all j, it is necessarily the case that $\tilde{u} = \min_{j|L_j>0} U_j/L_j$.

Proof. To establish the result, we focus on the case with fixed labor.[62] We assume WLOG that production Y_j is exogenous (endowment economy) since there is only one commodity and factors are immobile. To fix ideas, let us assume that $\mathcal{I} = \{0, 1, \ldots, J-1\}$ and $Y_0 > 0$ but $Y_j = 0$ for $j \geq 1$. We write down the Lagrangian of the problem

$$\mathscr{L} = \sum_j \omega_j L_j U(c_j, h_j) - \sum_j P_j \left[L_j c_j + \sum_{k \in N(j)} \left(1 + \delta^\tau_{jk} \frac{Q^\beta_{jk}}{I^\gamma_{jk}} \right) Q_{jk} - Y_j - \sum_{k \in N(j)} Q_{kj} \right]$$
$$- \mu \left(\sum_j \sum_{k \in N(j)} \delta^I_{jk} I_{jk} - K \right) + \sum_{j,k} \zeta^Q_{jk} Q_{jk} + \sum_{j,k} \zeta^I_{jk} I_{jk}, \quad \zeta^Q_{jk} \geqslant 0, \zeta^I_{jk} \geqslant 0.$$

Despite being a nonconvex optimization problem, there must exist a vector of Lagrange multipliers such that the KKT conditions hold.[63] As a preliminary step, we eliminate the infrastructure investment I_{jk} using (12), so that $I_{jk} = \left(\frac{\gamma}{\mu} \frac{\delta^\tau_{jk}}{\delta^I_{jk}} P_j Q^{1+\beta}_{jk} \right)^{\frac{1}{\gamma+1}}$ whenever $Q_{jk} > 0$, otherwise $I_{jk} = 0$. Solving for the multiplier μ such that (7) is satisfied, we reformulate the problem with allocation and flows only as follows

$$\mathscr{L} = \sum_j \omega_j L_j U(c_j, h_j) - \sum_j P_j \left[L_j c_j + \sum_{k \in N(j)} Q_{jk} - Y_j - \sum_{k \in \mathcal{N}(j)} Q_{kj} \right]$$
$$- K^{-\gamma} \left[\sum_{j,k} \left(\delta^I_{jk} / \delta^\tau_{jk} \right)^{\frac{\gamma}{\gamma+1}} \left(P_j Q^{1+\beta}_{jk} \right)^{\frac{1}{\gamma+1}} \right]^{\gamma+1} + \sum_{j,k} \zeta^Q_{jk} Q_{jk}, \quad \zeta^Q_{jk} \geqslant 0. \tag{A.4}$$

□

The source of nonconvexities is the term $\left[\sum_{j,k} \left(\delta^I_{jk} / \delta^\tau_{jk} \right)^{\frac{\gamma}{\gamma+1}} \left(P_j Q^{1+\beta}_{jk} \right)^{\frac{1}{\gamma+1}} \right]^{\gamma+1}$, which is convex when $\beta \geqslant \gamma$, but neither convex nor concave when $\gamma > \beta$. Let us now assume that $(\mathbf{c}^*, \mathbf{Q}^*)$ with $\mathbf{c}^* = (c^*_0, \ldots c^*_{J-1})'$ and $\mathbf{Q}^* = \left(Q^*_{jk}\right)_{j,k \in \mathcal{N}(j)}$ is a local optimum, i.e., it satisfies the FOCs and SOCs of the Lagrangian (A.4). We are going to show that the graph associated with $\mathbf{Q}^*$ is a tree. Define the (undirected) graph associated to $\mathbf{Q}^*$ as the tuple $(\mathcal{I}, \mathcal{E}^*)$ such that $\mathcal{E}^* \subset \mathcal{E}$ is a subset of the edges of the underlying geography such that

$$\mathcal{E}^* = \left\{ \{j, k\} \in \mathcal{E} \mid Q^*_{jk} > 0 \right\}.$$

Note that since I_{jk} is non-zero whenever $Q_{jk} > 0$ or $Q_{kj} > 0$, the support of graph $(\mathcal{I}, \mathcal{E}^*)$ coincides with that of the transport network $\{I_{jk}\}$. After this preparatory work, we now refer the reader to Proposition 5 in Appendix D which establishes that $\mathcal{E}^*$ is a tree.

Definition 3. *The decentralized equilibrium without labor mobility consists of quantities* $c_j, h_j, C_j, C^n_j, L^n_j, \mathbf{V}^n_j, \mathbf{X}^n_j, \{Q^n_{jk}\}_{k \in \mathcal{N}(j)}$, *goods prices* $\{p^n_j\}_n, p^C_j, p^H_j$ *and factor prices* $w_j, \{r^m_j\}_m$ *in each location jsuch that:*

(i)(a) consumers optimize:

$$\{c_j, h_j\} = \arg\max_{\hat{c}_j, \hat{h}_j} U\left(\hat{c}_j, \hat{h}_j\right)$$
$$p^C_j \hat{c}_j + p^H_j \hat{h}_j = e_j \equiv w_j + t_j,$$

where e_j *are expenditures per worker in* j *and where* p^C_j *is the price index associated with* $C_j\left(c^1_j, .., c^N_j\right)$ *at prices*

[62] In the labor mobility case, an identical argument can be made by taking the optimal allocation of L_j as given and replacing the Pareto weights ω_j with the Lagrange multipliers of the constraints $L_j u \leq L_j U(c_j, h_j)$.

[63] The resource constraint can be substituted in the objective function to yield $P_j = \omega_j U_C(c_j, h_j)$. The Arrow-Hurwicz-Uzawa theorem (Takayama (1985), Theorem 1.D.4) implies that, the remaining constraints being affine, there must exist a vector of Lagrange multipliers such that the KKT conditions hold.

$\left\{p_j^n\right\}_n$ *and* t_j *is a transfer per worker located in* j*. The set of transfers satisfy*

$$\sum_j t_j L_j = \Pi$$

where Π *adds up the aggregate returns to the portfolio of fixed factors and the government tax revenue,*

$$\Pi = \sum_j p_j^H H_j + \sum_j \sum_m r_j^m V_j^m + \sum_j \sum_{k \in \mathcal{N}(j)} \sum_n t_{jk}^n p_k^n Q_{jk}^n;$$

(i)(b) firms optimize:

$$\left\{L_j^n, \mathbf{V}_j^n, \mathbf{X}_j^n\right\} = \underset{\hat{L}_j^n, \hat{\mathbf{V}}_j^n, \hat{\mathbf{X}}_j^n}{argmax}\ p_j^n F_j^n \left(\hat{L}_j^n, \hat{\mathbf{V}}_j^n, \hat{\mathbf{X}}_j^n\right) - w_j \hat{L}_j^n - \sum_m r_j^m \hat{V}_j^{mn};$$

(i)(c) the transport companies optimize,

$$\pi_{od}^n = \max_{r=(j_0,\ldots,j_\rho)\in\mathcal{R}_{od}} p_d^n - p_o^n T_{r,0}^n - \sum_{k=0}^{\rho-1} p_{j_{k+1}} t_{j_k j_{k+1}}^n T_{r,k+1}^n,$$

for all $(o,d) \in \mathcal{J}^2$*, where* $\mathcal{R}_{od} = \left\{(j_0,\ldots,j_\rho) \in \mathcal{J}^{\rho+1}, \rho \in \mathbb{N} \mid j_0 = o,\ j_\rho = d,\ j_{k+1} \in \mathcal{N}(j_k) \text{ for all } 0 \le k < \rho\right\}$ *is the set of routes from* o *to* d*, and* $T_{r,k}^n$ *is the accumulated iceberg cost from location* j_k *to* d *along route* r*,*

$$T_{r,k}^n = \begin{cases} \prod_{m=k}^{\rho-1}\left(1+\tau_{j_m j_{m+1}}^n\right) & \text{for } 0 \le k \le \rho - 1 \\ 1 & \text{for } k = p; \end{cases}$$

and there is free entry to delivering products from every source to every destination: $\pi_{od}^n \le 0$ *for all* $(o,d) \in \mathcal{J}^2$*,* $=$ *if good* n *is shipped from* o *to* d*.*

(i)(d) producers of final commodities optimize:

$$\left\{C_j^n\right\} = \underset{C_j^n}{argmax}\ C_j\left(\left\{\hat{C}_j^n\right\}\right) - \sum_j p_j^n \hat{C}_j^n;$$

as well as the market-clearing and non-negativity constraints (i), (ii), (iv), and (v) from Definition 1.

If, in addition,if labor is mobile, then the decentralized equilibrium also consists of utility u *and employment* $\{L_j\}$ *such that*

$$u = U_j(c_j, h_j)$$

whenever $L_j > 0$*, and the labor market clearing condition (vii) from Definition 2 holds.*

Proposition 4. *(First and Second Welfare Theorems) If the sales tax on shipments of product* n *from* j *to* k *is* $1 - t_{jk}^n = \frac{1+\tau_{jk}^n}{1+\left(\varepsilon_{Q,jk}^n+1\right)\tau_{jk}^n}$ *where* $\varepsilon_{Q,jk}^n = \partial \log \tau_{jk}^n / \partial \log Q_{jk}^n$*, then: (i) if labor is immobile, the competitive allocation coincides with the planner's problem under specific planner's weights* ω_j*. Conversely, the planner's allocation can be implemented by a market allocation with specific transfers* t_j*; and (ii) if labor is mobile, the competitive allocation coincides with the planner's problem if and only if all workers own an equal share of fixed factors and tax revenue, i.e.,* $t_j = \frac{\Pi}{L}$*. In either case, the price of good* n *in location* j*,* p_j^n*, equals the multiplier on the balanced-flows constraint in the planner's allocation,* P_j^n*.*

Proof. **Equivalence of the First-order Conditions.** Condition (i)(c) from the definition of the market allocation implies that the free entry condition of shippers holds for every pair of neighbors; i.e., for every $j \in \mathcal{J}$ and $k \in \mathcal{N}(j)$,

$$p_k^n\left(1 - t_{jk}^n\right) \le p_j^n\left(1+\tau_{jk}^n\right),\ = \text{if } Q_{jk}^n > 0. \tag{A.5}$$

This condition is consistent with the first-order condition (8) from the planner's problem if and only if the tax scheme is defined as in the proposition. We must further show that, under this tax scheme, a route is the solution to (i)(c) if and only if it is used in the solution to the planner's problem, which we establish at the end of this proof.

Without labor mobility, the rest of the allocation corresponds to a standard neoclassical economy with convex technologies and preferences where the welfare theorems hold. Specifically, the first-order conditions from the consumer and firm optimization problems (i)(a) and (i)(b) yield:

$$
\begin{aligned}
&[\hat{c}_j] && \left(\frac{1}{\lambda_j}\right) U_C\left(c_j, h_j\right) = p_j^C \\
&\left[\hat{h}_j\right] && \left(\frac{1}{\lambda_j}\right) U_H\left(c_j, h_j\right) = p_j^H \\
&\left[\hat{C}_j^n\right] && p_j^C \frac{\partial C_j^T}{\partial C_j^n} = p_j^n \\
&\left[\hat{L}_j^n\right] && \frac{\partial Y_j^n}{\partial L_j^n} P_j^n \leq w_j, = \text{if } L_{jn} > 0 \\
&\left[\hat{V}_j^{mn}\right] && \frac{\partial Y_j^n}{\partial V_j^{mn}} P_j^n \leq r_j^m, = \text{if } V_j^{mn} > 0.
\end{aligned}
$$

Since the market clearing constraints are the same in the market's and the planner's allocation, the planner's allocation coincides with the market if the planner's weights are such that the planner's FOC for C_j coincide with the market. This is the case if the weight ω_j from the planner's problem equals the inverse of the multiplier on the budget constraint from the consumer's optimization problem (i)(a) in the market allocation. To find that weight, using that U is homothetic we can write $U = G\left(U_0\left(c, h\right)\right)$, where U_0 is homogeneous of degree 1. Then, the planner's allocation coincide with the market's under weights

$$
\omega_j = \frac{e_j}{G'\left(U_0\left(c_j, h_j\right)\right) U_0\left(c_j, h_j\right)},
$$

where e_j is the expenditure per worker and c_j, h_j are the consumption per worker of the traded and non-traded good in the market allocation. If U is homogeneous of degree one, then $\omega_j = P_j^U$, where P_j^U is the price index associated with $U\left(c_j, h_j\right)$ at the market equilibrium prices p_j^C, p_j^H. In the opposite direction, given arbitrary weights ω_j, the market allocation implements the planner's under the transfers $t_j = P_j^C c_j + P_j^H h_j - W_j$ constructed using the quantities $\{c_j, h_j\}$ from the planner's allocation and the multipliers $\left\{P_j^C, P_j^H\right\}$ and W_j corresponding to the constraints (i) and (iv) of the planner's problem, respectively.

For the case with labor mobility, note that, for populated locations, the planner's first-order condition with respect to L_j implies:

$$
P_j^C c_j + P_j^C h_j = W_j - W^L.
$$

Therefore, the market allocation and the planner's solution coincide if and only if in the market allocation expenditure per worker in location j takes the form $e_j = w_j + \textit{Constant}$ for all j. The only transfer scheme delivering the same transfer per capita is $t_j = \frac{\Pi}{L}$.

Equivalence of Least Cost Routes. We want to establish that a route is a solution to (i)(c) in Definition 3 under the proposed tax scheme if and only if it is used in the planner's problem. Fix good n. We introduce the following notation: for all $r \in \mathcal{R}_{od}$, we denote t_r^n the matrix:

$$
\left(t_r^n\right)_{j \in \mathcal{I}, k \in \mathcal{N}(j)} = \begin{cases} T_r^n\left(j_{l+1}, d\right) & \text{for } 0 \leqslant l \leqslant p-1 \text{ such that } j = j_l, k = j_{l+1} \\ 0 & \text{otherwise,} \end{cases}
$$

where $T_r^n\left(j_k, d\right)$ is the accumulated iceberg cost from j_k to d on route r as introduced in Definition 3. Matrix t_r^n captures the fact that shipping ε additional units of good n (at destination) from o to d through route r requires modifying the trade flows to $\mathbf{Q} + \varepsilon t_r^n$. Indeed, $Q_{j_{p-1}d}^n$ must increase by $\varepsilon \times T_r\left(j_p, d\right) = \varepsilon$, $Q_{j_{p-2}j_{p-1}}^n$ by $\varepsilon \times T_r\left(j_{p-1,d}\right) = \varepsilon\left(1 + \tau_{j_{p-1}j_p}^n\right)$, and so on.

Consider an optimal route from o to d, $r^* = \left(j_0^*, \ldots, j_{p^*}^*\right) \in \mathcal{R}_{od}$, i.e., such that $Q_{j_k^* j_{k+1}^*}^n > 0$ at the optimum of the planner's problem. We now consider redirecting a marginal amount of goods $\varepsilon > 0$ from π^* to some other route $r = (j_0, \ldots, j_p) \in \mathcal{R}_{od}$. In other words, we consider the flows $\mathbf{Q} + \varepsilon t_r^n - \varepsilon t_{r^*}^n$. The first order impact on the Lagrangian can be decomposed into its various contributions to the resource constraint in locations along the two paths, r and r^*:

1. *Destination d*: the resource constraint in the final destination, d, is unaffected: the flows from route r^* are reduced by ε, but an additional amount ε of goods arrive from route r.

2. *Locations along the path*: for any location j_k^* along route r^*, other than that the origin and the destination, the amount of goods that arrives from j_{k-1}^* decreases by $\varepsilon T_{r^*}^n(j_k^*, d)$ and the amount of goods re-expedited to j_{k+1}^* decrease by $\varepsilon T_{r^*}^n(j_{k+1}^* d)$, implying a reduction in gross shipments of $\left(1 + \tau_{j_k j_{k+1}}^n\right) \times \varepsilon T_{r^*}^n(j_{k+1}^*, d) = \varepsilon T_{r^*}^n(j_k^*, d)$. Both reductions of flows thus fully offset each other. The sole remaining effect on the resource constraint at j_k^* is a reduction in total trade costs through lower congestion. The first-order contribution of these savings in trade costs corresponds to a gain of

$$P_{j_k^*}^n \frac{\partial \tau_{j_k^* j_{k+1}^*}^n}{\partial Q_{j_k^* j_{k+1}^*}^n} Q_{j_k^* j_{k+1}^*}^n \times \varepsilon T_{r^*}^n(j_{k+1}^*, d).$$

   Symmetrically, the impact on a location j_k along route r amounts to a first-order impact on the Lagrangian of

$$\left[-P_{j_k}^n \frac{\partial \tau_{j_k j_{k+1}}^n}{\partial Q_{j_k j_{k+1}}^n} Q_{j_k j_{k+1}}^n + \zeta_{j_k j_{k+1},n}^Q\right] \times \varepsilon T_r^n(j_{k+1}, d),$$

   where we have also included the multiplier $\zeta_{j_k j_{k+1},n}^Q$ which can be strictly positive along initially unused links, as opposed to links along the optimal path.

3. *Origin o*: shipments through path r^* are reduced by $\varepsilon T_{r^*}^n(j_1^*, d)$, implying a saving of $\left(1 + \tau_{oj_1^*}^n\right) \times \varepsilon T_{r^*}^n(j_1^*, d) = \varepsilon T_{r^*}^n(o, d)$ goods; while shipments through r are increased by $\varepsilon T_r^n(j_1, d)$, implying an additional demand of resources of $\left(1 + \tau_{oj_1}^n\right) \times \varepsilon T_r^n(j_1, d) = \varepsilon T_r^n(o, d)$. In addition, total trade costs are affected by changes in congestion along both paths. Lower congestion along path r^* amounts to a saving in trade costs of

$$P_o^n \frac{\partial \tau_{oj_1^*}^n}{\partial Q_{oj_1^*}^n} Q_{oj_1^*}^n \times \varepsilon T_{r^*}^n(j_1^*, d),$$

   but increased congestion along r, combined with the non-negativity constraint on $Q_{oj_1}^n$, results in a Lagrangian first-order impact of

$$\left[-P_o^n \frac{\partial \tau_{oj_1}^n}{\partial Q_{oj_1}^n} Q_{oj_1}^n + \zeta_{oj_1,n}^Q\right] \times \varepsilon T_r^n(j_1, d).$$

By definition, r^* being the solution to the planner's saddle point problem of the Lagrangian, the redirection of flows from r^* to r should have a negative first-order effect on the Lagrangian. Summing up, the overall impact on the Lagrangian, this implies

$$P_o^n T_{r^*}^n(o, d) + \sum_{k=0}^{p^*-1} \left[P_{j_k^*}^n \frac{\partial \tau_{j_k^* j_{k+1}^*}^n}{\partial Q_{j_k^* j_{k+1}^*}^n} Q_{j_k^* j_{k+1}^*}^n\right] T_{r^*}^n(j_{k+1}^*, d)$$
$$\leqslant P_o^n T_r^n(o, d) + \sum_{k=0}^{p-1} \left(P_{j_k}^n \frac{\partial \tau_{j_k j_{k+1}}^n}{\partial Q_{j_k j_{k+1}}^n} Q_{j_k j_{k+1}}^n - \zeta_{j_k j_{k+1},n}^Q\right) T_r^n(j_{k+1}, d).$$

To simplify the expression, we use the first-order condition of the Lagrangian with respect to every Q_{jk}^n,

$$P_j^n \frac{\partial \tau_{jk}^n}{\partial Q_{jk}^n} Q_{jk}^n - \zeta_{jk,n}^Q = P_k^n - P_j^n\left(1 + \tau_{jk}^n\right),$$

and obtain that the optimal route r^* must be the solution to the least-cost route problem

$$\min_{r\in\mathcal{R}_{od}} P_o^n T_r^n(o,d) + \sum_{k=0}^{p-1}\left[P_{j_{k+1}}^n - \left(1+\tau_{j_k j_{k+1}}^n\right)P_{j_k}^n\right]T_r^n(j_{k+1},d).$$

Under the proposed Pigouvian tax scheme, the above least-cost route problem can be equivalently restated as a minimization over total shipping costs and total tax liabilities. More specifically, the free-entry condition on every link (j,k) tells us that

$$P_k^n - \left(1+\tau_{jk}^n\right)P_j^n \leqslant P_k^n t_{jk}^n,$$

with equality if the link is used at the optimum. Hence, for all routes $r\in\mathcal{R}_{od}$, we have

$$P_o^n T_r^n(o,d) + \sum_{k=0}^{p-1}\left[P_{j_{k+1}}^n - \left(1+\tau_{j_k j_{k+1}}^n\right)P_{j_k}^n\right]T_r^n(j_{k+1},d) \leqslant P_o^n T_r^n(o,d) + \sum_{k=0}^{p-1} P_{j_{k+1}}^n t_{j_k j_{k+1}}^n T_r^n(j_{k+1},d),$$

with equality for the optimal route r^*. Hence, the optimal route r^* is also solution to the least-cost route problem

$$\min_{r\in\mathcal{R}_{od}} P_o^n T_r^n(o,d) + \sum_{k=0}^{p-1} P_{j_{k+1}}^n t_{j_k j_{k+1}}^n T_r^n(j_{k+1},d),$$

where we recognize the equivalence with condition (i)(c) of Definition 3.

□

A.4 Appendix to Section 3.5 (Numerical Implementation)

In this section, we provide a more detailed explanation of the numerical algorithms we use to solve the model.

Duality Approach

As explained in section 3.5, our preferred approach to solve the model relies on solving the dual Lagrangian problem of the planner. We provide, here, a simple example of how to solve the joint optimal allocation and transport problem taking the infrastructure network $\{I_{jk}\}$ as given. This example can easily be generalized to the full problem, including the network design problem, in the convex case, but is also part of our resolution method for the nonconvex case. We focus on the case studied in the quantitative part of the paper, in which: i) we use the log-linear specification of transport costs, $\tau_{jk}^n = \delta_{jk}^\tau \left(Q_{jk}^n\right)^\beta I_{jk}^{-\gamma}$; ii) labor is the sole production factor, $F_j^n\left(L_j^n\right) = z_j^n\left(L_j^n\right)^a$; and iii) C^T is a CES aggregator with elasticity of substitution σ. We consider the case with immobile labor.[64]

We write the Lagrangian of the problem

$$\begin{aligned}\mathscr{L} = &\sum_j \omega_j L_j U(c_j,h_j) - \sum_j P_j^C\left[c_j L_j - \left(\sum_n \left(C_j^n\right)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}\right]\\ &-\sum_j\sum_n P_j^n\left[C_j^n + \sum_{k\in\mathcal{N}(j)}\left(Q_{jk}^n + \delta_{jk}^\tau\left(Q_{jk}^n\right)^{1+\beta} I_{jk}^{-\gamma}\right) - z_j^n\left(L_j^n\right)^a - \sum_{i\in\mathcal{N}(j)} Q_{ij}^n\right]\\ &-\sum_j W_j\left[\sum_n L_j^n - L_j\right] + \sum_{j,k,n}\zeta_{jkn}^Q Q_{jk}^n + \sum_{j,n}\zeta_{jn}^L L_j^n + \sum_{j,n}\zeta_{jn}^C C_j^n + \sum_j \zeta_j^c c_j.\end{aligned}$$

Recall that the dual problem consists of solving

$$\inf_{\lambda\geq\mathbf{0}}\sup_{\mathbf{x}} \mathscr{L}(\mathbf{x},\boldsymbol{\lambda}).$$

[64] In the mobile labor case, we can only show that the planner's problem is a quasiconvex optimization problem. Hence, a duality gap may exist. We therefore adopt a (slower) primal approach in that case.

We start by expressing our control variables $\mathbf{x} = \left(c_j, C_j^n, Q_{jk}^n, L_j^n\right)$ as functions of the Lagrange multipliers $\boldsymbol{\lambda} = \left(P_j^C, P_j^n, W_j, \zeta_{jkn}^Q, \zeta_{jn}^L, \zeta_{jn}^C, \zeta_j^c\right)$. Using the optimality conditions, one obtains the following expressions:

$$c_j = U_c^{<-1>}\left(\left(\sum_{n'}\left(P_j^{n'}\right)^{1-\sigma}\right)^{\frac{1}{1-\sigma}}/\omega_j, h_j\right)$$

$$C_j^n = \left[\frac{P_j^n}{\left(\sum_{n'}\left(P_j^{n'}\right)^{1-\sigma}\right)^{\frac{1}{1-\sigma}}}\right]^{-\sigma} L_j c_j$$

$$Q_{jk}^n = \left[\frac{1}{1+\beta}\frac{I_{jk}^{\gamma}}{\delta_{jk}^{\tau}}\max\left(\frac{P_k^n}{P_j^n} - 1, 0\right)\right]^{\frac{1}{\beta}}$$

$$L_j^n = \frac{\left(P_j^n z_j^n\right)^{\frac{1}{1-a}}}{\sum_{n'}\left(P_j^{n'} z_j^{n'}\right)^{\frac{1}{1-a}}} L_j.$$

As these expressions illustrate, we have been able to eliminate a large number of the multipliers directly, so that the only remaining Lagrange multipliers are $\boldsymbol{\lambda} = \left(P_j^n\right)_{j,n}$. We may now compute the inner part of the saddle-point problem:[65]

$$\begin{aligned}\mathscr{L}\left(\mathbf{x}\left(\boldsymbol{\lambda}\right), \boldsymbol{\lambda}\right) = &\sum_j \omega_j L_j U\left(c_j\left(\boldsymbol{\lambda}\right), h_j\right) \\ &- \sum_j\sum_n P_j^n\left[C_j^n\left(\boldsymbol{\lambda}\right) + \sum_{k\in\mathcal{N}(j)}\left(Q_{jk}^n\left(\boldsymbol{\lambda}\right) + \delta_{jk}^{\tau}\left(Q_{jk}^n\left(\boldsymbol{\lambda}\right)\right)^{1+\beta} I_{jk}^{-\gamma}\right) - z_j^n\left(L_j^n\left(\boldsymbol{\lambda}\right)\right)^a - \sum_{i\in\mathcal{N}(j)} Q_{ij}^n\left(\boldsymbol{\lambda}\right)\right].\end{aligned}$$

The dual problem then consists of the simple unconstrained, convex[66] minimization problem in $J \times N$ unknowns:

$$\min_{\boldsymbol{\lambda}\geqslant 0} \mathscr{L}\left(\mathbf{x}\left(\boldsymbol{\lambda}\right), \boldsymbol{\lambda}\right).$$

This problem can be readily fed into a numerical optimization software. Faster convergence can be achieved by providing the software with an analytical gradient and hessian. Note that, as a direct implication of the envelope theorem, the gradient of the dual problem is simply the vector of constraints:

$$\nabla\mathscr{L}\left(\mathbf{x}\left(\boldsymbol{\lambda}\right), \boldsymbol{\lambda}\right) = -\left(\begin{array}{c}\vdots \\ C_j^n\left(\boldsymbol{\lambda}\right) + \sum_{k\in\mathcal{N}(j)}\left(Q_{jk}^n\left(\boldsymbol{\lambda}\right) + \delta_{jk}^{\tau}\left(Q_{jk}^n\left(\boldsymbol{\lambda}\right)\right)^{1+\beta} I_{jk}^{-\gamma}\right) - z_j^n\left(L_j^n\left(\boldsymbol{\lambda}\right)\right)^a - \sum_{i\in\mathcal{N}(j)} Q_{ij}^n\left(\boldsymbol{\lambda}\right) \\ \vdots\end{array}\right).$$

Nonconvex cases

When the conditions for convexity fail to obtain, the full planner's problem is not a convex optimization problem. It is, however, easy to find local optima by using the following iterative procedure. We then search for a global maximum using a simulated annealing method that we describe below.

Finding Local Optima Despite the failure of global convexity for the full planner's problem, the joint optimal allocation and transport problems, taking the network as given, is always convex as long as $\beta \geqslant 0$. We thus use our duality approach to solve for $\left(c_j, C_j^n, Q_{jk}^n, L_j^n\right)$ for a given level of infrastructure I_{jk}, and then iterate on the (necessary)

[65] Note that, due to complementary slackness, we can drop the constraints that correspond to all the Lagrange multipliers that we were able to solve by hand. As a result, only the balanced flows constraints remain.

[66] Dual problems are always convex, by construction, even when the primal problem is not.

first order conditions that characterize the optimal network. The procedure can be summarized in pseudo-code as follows.

1. Let $l := 1$. Guess some initial level of infrastructure $\left\{I_{jk}^{(1)}\right\}$ that satisfies the network building constraint.

2. Given the network $\left\{I_{jk}^{(l)}\right\}$, solve for $\left(c_j, C_j^n, Q_{jk}^n, L_j^n\right)$ using a duality approach.

3. Given the flows Q_{jk}^n and the prices P_j^n, get a new guess $I_{jk}^{(l+1)} = \left[\frac{\gamma}{\mu}\frac{\delta_{jk}^\tau}{\delta_{jk}^I}\left(\sum_n P_j^n \left(Q_{jk}^n\right)^{1+\beta}\right)\right]^{\frac{1}{1+\gamma}}$ and set μ such that $\sum \delta_{jk}^I I_{jk}^{(l+1)} = K$.

4. If $\sum_{j,k}\left|I_{jk}^{(l+1)} - I_{jk}^{(l)}\right| \leqslant \varepsilon$, then we have converged to a potential candidate for a local optimum. If not, set $l := l+1$ and go back to (2).

Simulated Annealing In the absence of global convexity results, the above iterative procedure is likely to end up in a local extremum. Unfortunately, there exists to our knowledge few global optimization methods that would guarantee convergence to a global maximum in a reasonable amount of time.[67] We opt for the simple but widely used heuristic method of simulated annealing, which is a very popular probabilistic method to search for the global optimum of high dimensional problems such as, for instance, the traveling salesman problem. Simulated annealing can be described as follows:

1. Let $l := 1$. Set the initial network $\left\{I_{jk}^{(1)}\right\}$ to a local optimum from the previous section and compute its welfare $v^{(1)}$. Set the initial "temperature" T of the system to some number.

2. Draw a new candidate network $\left\{\hat{I}_{jk}\right\}$ by perturbing $\left\{I_{jk}^{(l)}\right\}$ (see below). [Optional: deepen the network.] Compute the corresponding optimal allocation and transport $\left\{c_j, C_j^n, Q_{jk}^n, L_j^n\right\}$. Compute associated welfare $\hat{v}$.

3. Accept the new network, i.e., set $I_{jk}^{(l+1)} = \hat{v}$ and $v^{(l+1)} = \hat{v}$ with probability $\min\left[\exp\left(\left(\hat{v} - v^{(l)}\right)/T\right), 1\right]$, if not keep the same network, $\left\{I_{jk}^{(l+1)}\right\} = \left\{I_{jk}^{(l)}\right\}$ and $v^{(l+1)} = v^{(l)}$.

4. Stop when $T < T_{min}$. Otherwise set $l := l+1$ and $T := \rho_T T$ and return to (2),

where $\rho_T < 1$ controls the speed of convergence. Note that we improve the algorithm by allowing to "deepen" the network in step (2), meaning that we additionally apply the iterative procedure from the previous section for a pre-specified number of iterations so that the candidate network $\left\{\hat{I}_{jk}\right\}$ is more likely to be a local optimum itself.

Drawing Candidate Networks The performance of the simulated annealing depends on how new candidate networks are drawn. Because of the complex network structure, purely random perturbations are likely to be rejected and the algorithm may easily fail to improve the initial network. We design a simple algorithm that exploits the structure of the problem to make educated guesses for the candidate networks. The algorithm builds on the idea that, under increasing returns, a welfare improvement can be achieved by directly connecting locations to more central locations. Since a lower price level indicates that a location has higher relative availability of goods produced anywhere in the economy, we use the price level as a proxy for centrality. We thus construct candidate networks where random locations are better connected to their lowest-price neighbors. The algorithm can be described as follows:

1. Given an initial network $I_{jk}^{(l)}$, draw a random set of locations $I \subset \mathcal{J}$.

[67] Techniques such as the branch-and-bound method are guaranteed to converge to the global optimum, but remain heavy to implement and computationally intensive.

2. For each $j \in I$, identify $m(j)$ as the neighbor with the lowest price index for the bundle of tradable goods, $m(j) = \underset{k \in \mathcal{N}(j)}{\operatorname{argmin}} P_k^C$, and $n(j)$ as the "parent" with the highest level of infrastructure, $n(j) = \underset{k \in \mathcal{N}(j) | P_k^n \leqslant P_j^n}{\operatorname{argmax}} I_{kj}$.

3. For each $j \in I$, define the candidate network $I_{jk}^{(l+1)}$ by switching the infrastructure levels of $m(j)$ and $n(j)$:

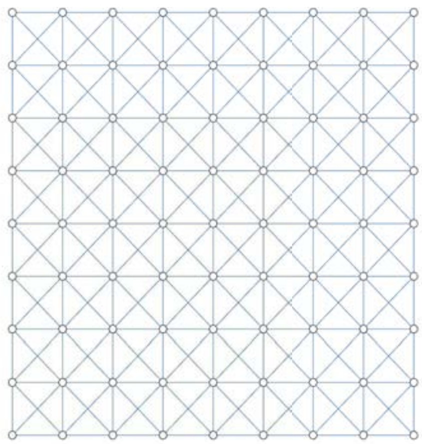
$$I_{kj}^{(l+1)} = \begin{cases} I_{m(j)j}^{(l)} & \text{if } k = n(j), j \in I \\ I_{n(j)j}^{(l)} & \text{if } k = m(j), j \in I \\ I_{kj}^{(l)} & \text{if } j \notin I \text{ or } (j \in I \text{ and } k \notin \{m(j), n(j)\}), \end{cases}$$

which, by construction, satisfies the network building constraint.

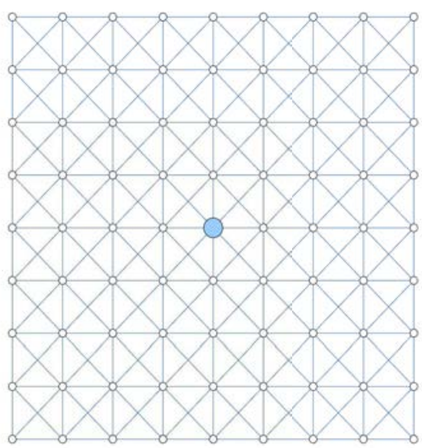
B Appendix to Section 4 (Illustrative Examples)

Figure A.1: A Simple Underlying Geography

(a) Population



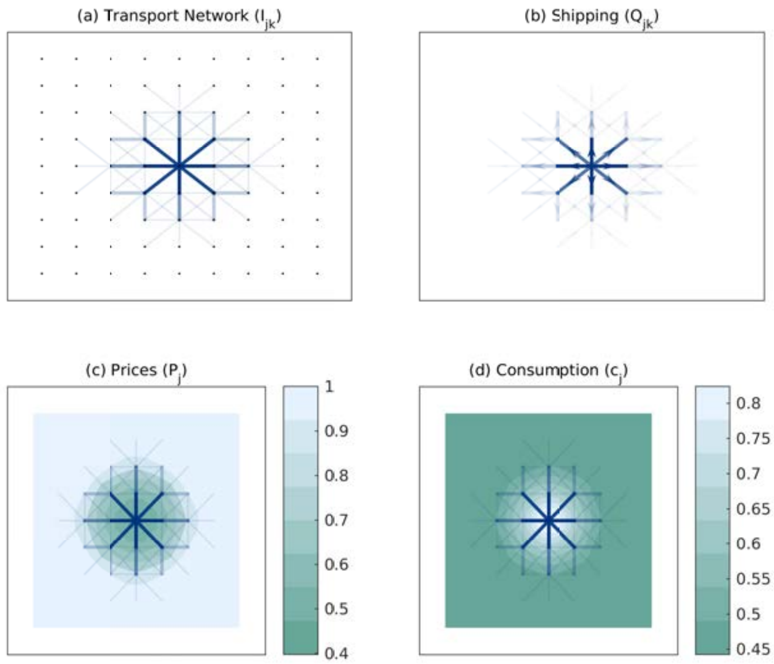
(b) Productivity



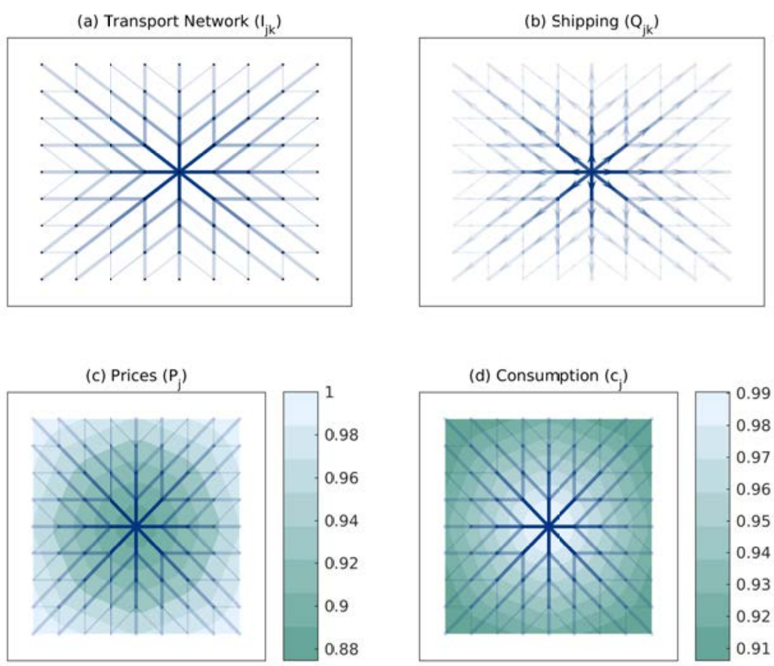
Notes: On panel (a), each circle represents a location. The links represent the underlying network, i.e., links upon which the transport network may be built. Population and housing are uniform across space, normalized to 1. On panel (b), the size of the circles represent the productivity of each location.

Figure A.2: The Optimal Network for $K = 1$ and $K = 100$

(a) K=1



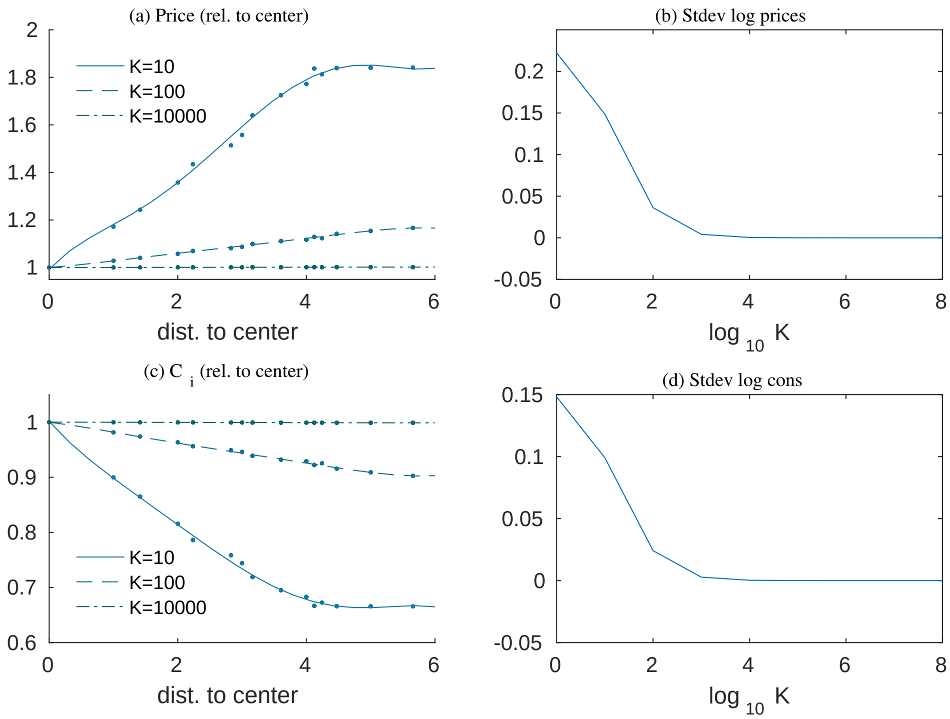
(b) K=100



Notes: On each panel, the thickness and color of the segments reflects the level of infrastructure built on a given link. Thicker and darker colors represent more infrastructure. On the bottom panels, the heat map represents the level of prices and consumption, normalized to 1 at the center. Lighter color represents higher values for prices and consumption. Prices and consumption levels are linearly interpolated across space to obtain smooth contour plots.

Figure A.3: Optimal Network Growth

(a) Spatial Inequalities



(b) Uniform versus Optimal Network

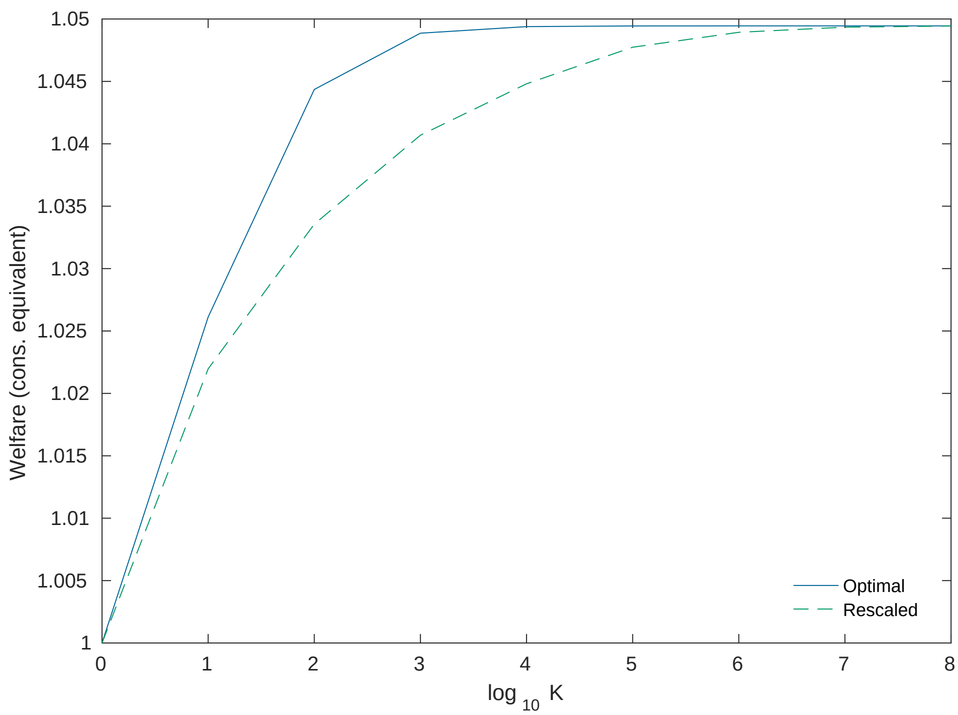
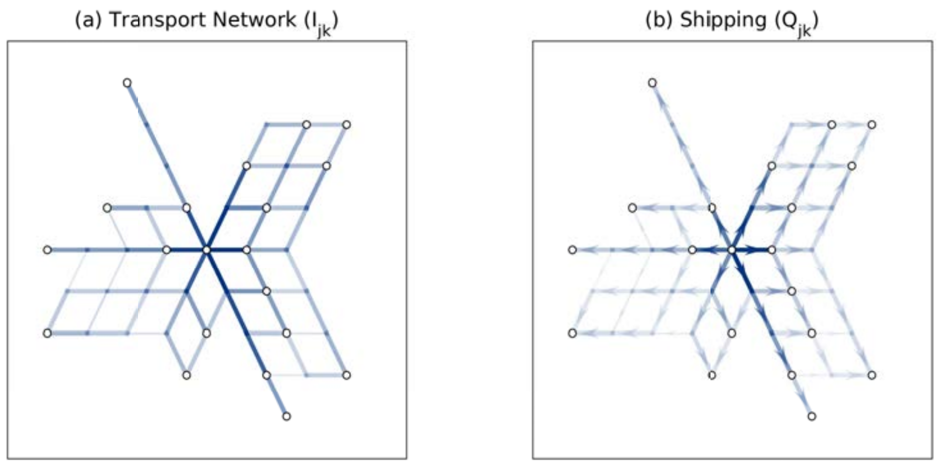
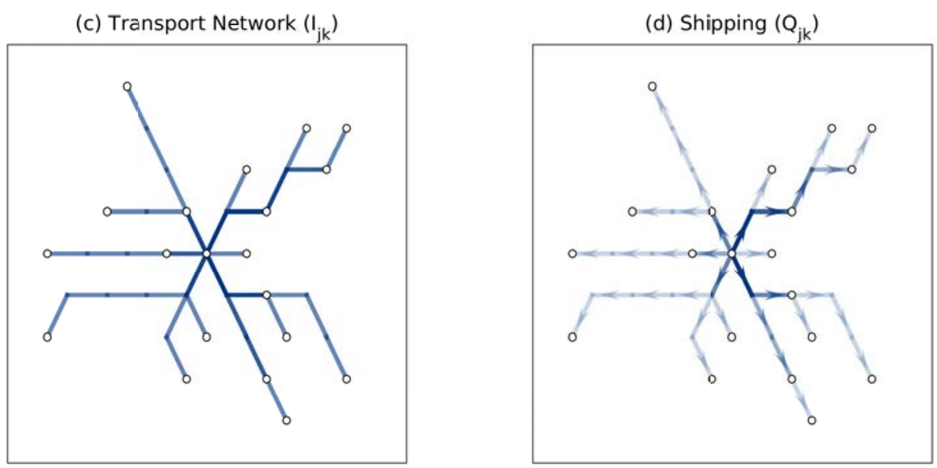


Figure A.4: Optimal Network with Randomly Located Cities

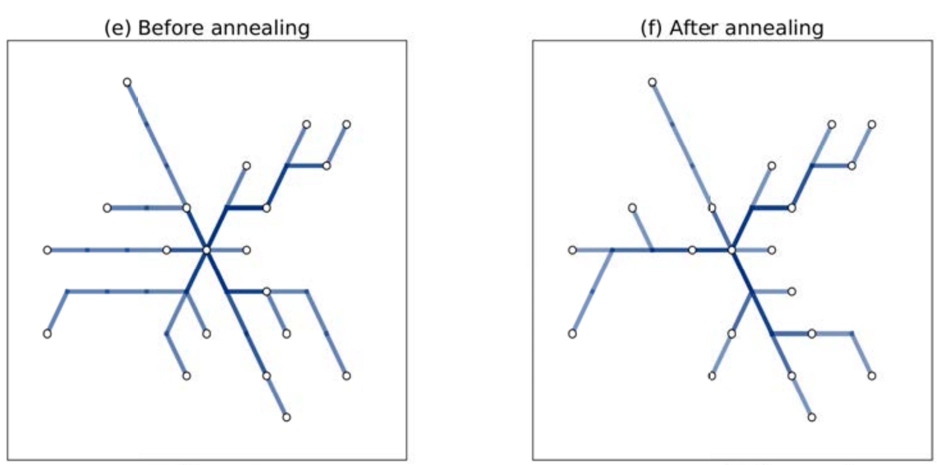
(a) Convex Case: $\gamma = \beta = 1$



(b) Non-Convex Case: $\gamma = 2 > \beta = 1$

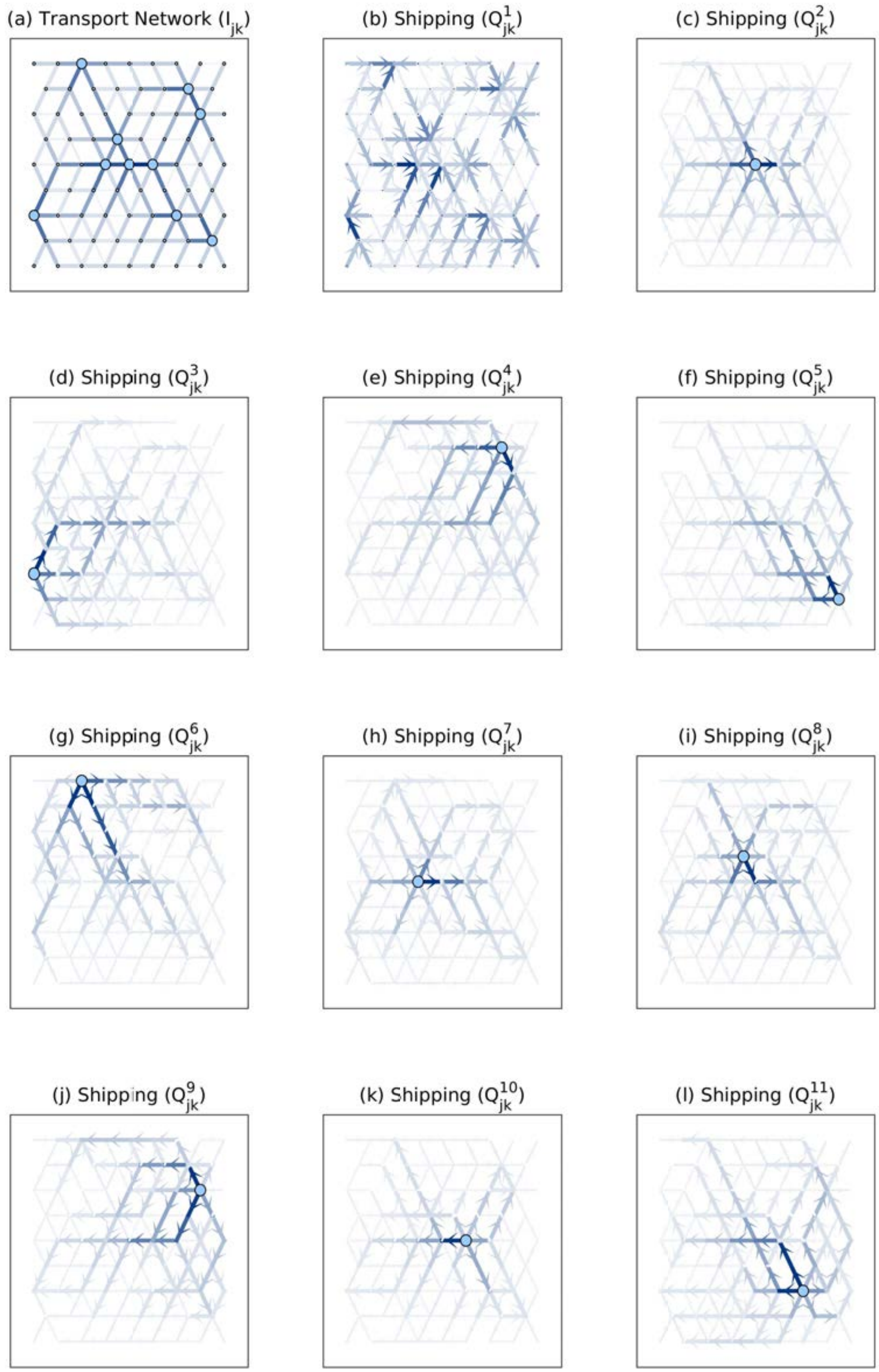


(c) Optimal Network Before and After Annealing Refinement in Non-Convex Case



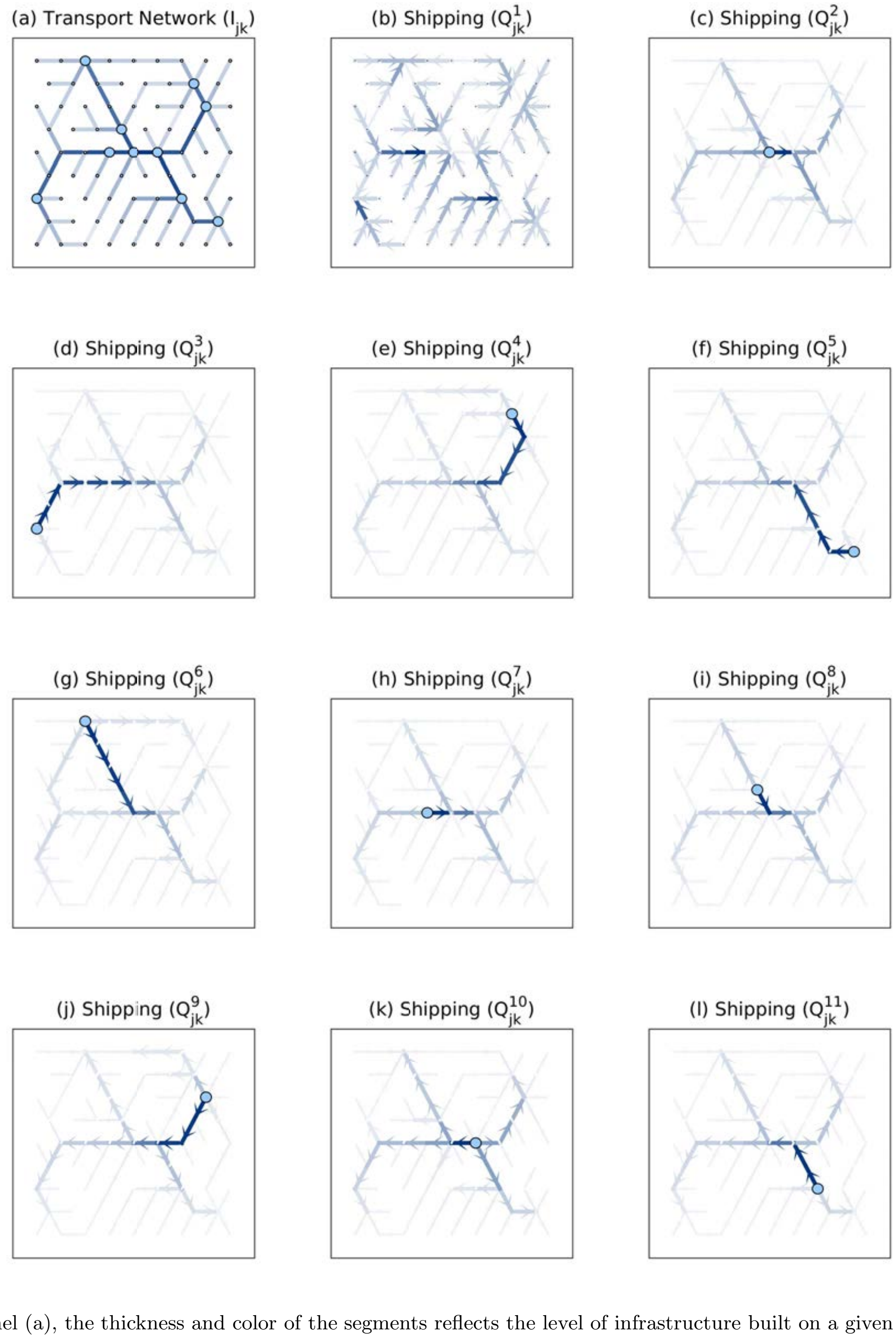
Notes: On each panel, the thickness and color of the segments reflects the level of infrastructure built or the shipment sent on a given link. Thicker and darker colors represent higher infrastructure or quantity.

Figure A.5: Optimal Network with 10+1 Goods, Convex Case ($\beta = \gamma = 1$), Labor Mobility



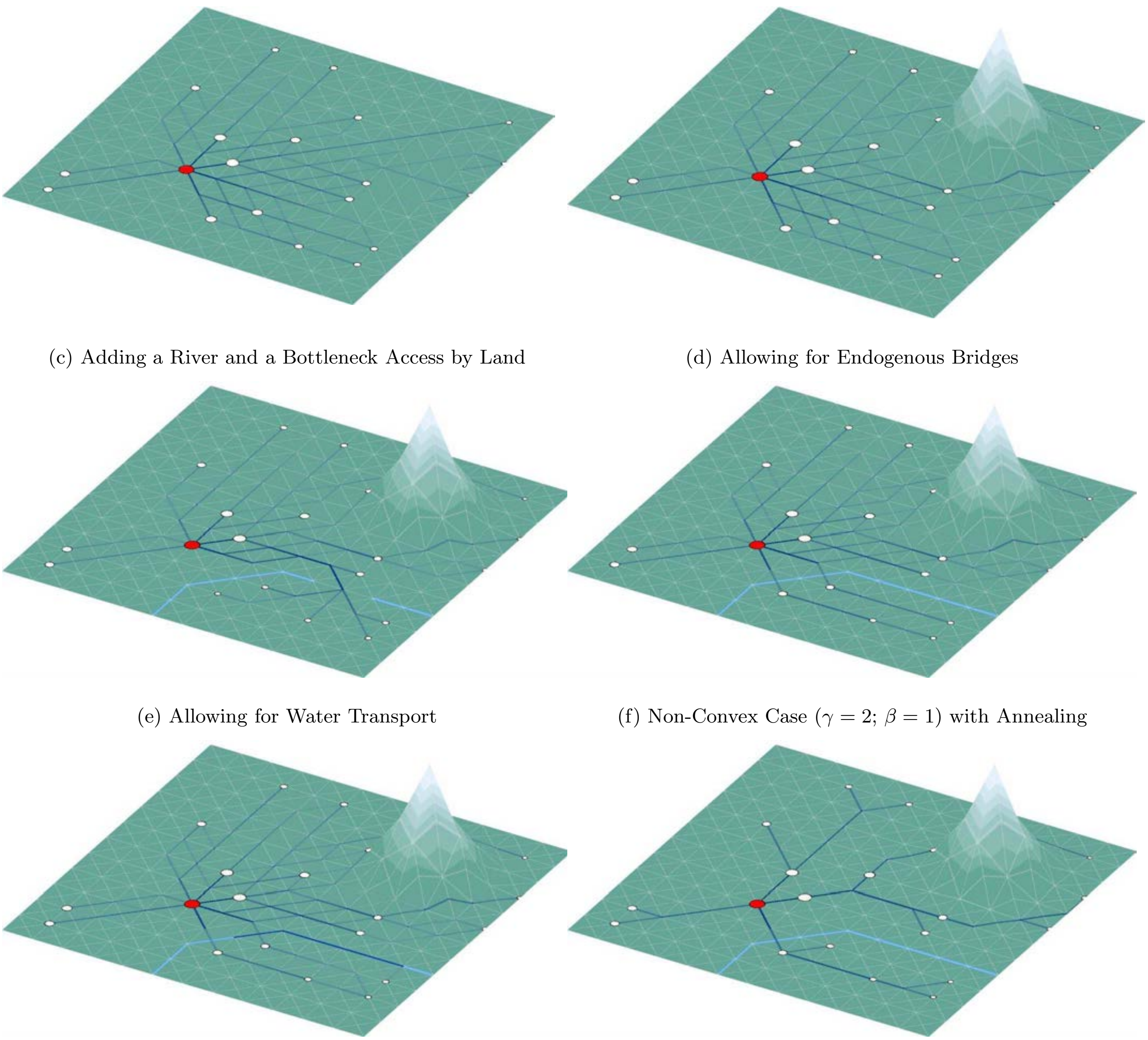
Notes: On panel (a), the thickness and color of the segments reflects the level of infrastructure built on a given link, and the size of each circle is the population share. On the other panels, the segments represent the quantity shipped through each link and the circles represent the location of producers.

Figure A.6: Optimal Network with 10+1 Goods, Nonconvex Case ($\beta = 1$,$\gamma = 2$), Labor Mobility



Notes: On panel (a), the thickness and color of the segments reflects the level of infrastructure built on a given link, and the size of each circle is the population share. On the other panels, the segments represent the quantity shipped through each link and the circles represent the location of producers.

Figure A.7: The Optimal Transport Network under Alternative Building Costs



(a) Baseline Geography

(b) Adding a Mountain

(c) Adding a River and a Bottleneck Access by Land

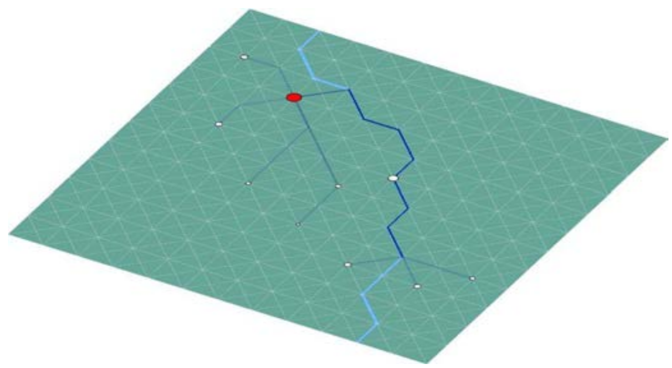
(d) Allowing for Endogenous Bridges

(e) Allowing for Water Transport

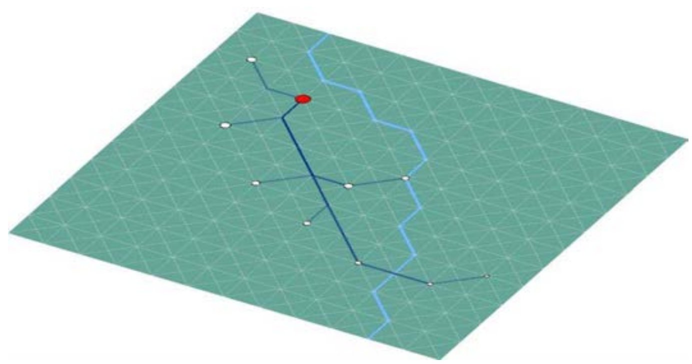
(f) Non-Convex Case ($\gamma = 2$; $\beta = 1$) with Annealing

Notes: The thickness and color of the segments reflects the level of infrastructure built on a given link. Thicker and darker colors represent more infrastructure and quantities. The circles represent the 20 cities randomly allocated across spaces. The larger red circle represents the city with the highest productivity. The different panels vary in the parametrization of the cost of building infrastructure. In panel (a), it is only a function of Euclidean distance. In panel (b), we add a mountain and assume that the cost also depend on difference in elevation. In panel (c), we add a river with a natural land crossing and assume that the cost of building along or across the river is infinite. In panel (d) there is no natural land crossing but allow for construction of bridges. In panel (e) we additionally allow for investment in water transport. Panel (d) makes the assumptions as Panel (e) but assumes increasing returns to network building.

Figure A.8: Arrival of a New Transport Technology and Network Reoptimization



(a) Initial Geography Dependence on Water Transport



(b) Allowing for Cheap Land Transport

Notes: The bright blue curve represents a river. The thickness and color of the other segments reflects the level of infrastructure built on a given link. Thicker and darker colors represent more infrastructure/quantities. The circles represent the 10 cities randomly allocated across spaces. The larger red circle represents the city with the highest productivity.

C Appendix to Section 4 (Calibration and Counterfactuals)

Construction of $\mathcal{P}(j,k)$ The definition of the weights $\omega_{jk}(s)$ assigned to the construction of I_{jk}^{obs} involves the cheapest path $\mathcal{P}(j,k)$ for all $j \in \mathcal{J}$ and $k \in \mathcal{N}(j)$ in every country. To find $\mathcal{P}(j,k)$, we first convert the shapefile with all the road segments from EuroGeographics into a weighted graph, where each edge corresponds to a segment s on the road network. We define $\mathcal{P}(j,k)$ as the shortest path between j and k under the segment-specific weights $length_s * lanes_s^{-\chi_{lane}} * \chi_{use}^{1-nat_s} * \chi_{paved}^{1-paved_s} * \chi_{median}^{1-median_s}$, where $length_s$ is the length of the segment, $lanes_s$ is the number of lanes, nat_s equals 1 if the segment belongs to a national road, $paved_s$ equals 1 if the segment is paved, and $median_s$ equals 1 if the segment has a median. We parametrize χ_{lane}, χ_{use}, χ_{paved}, and χ_{median} based on the extent by which adding a lane, using a national road, using paved road, or using a road with a median reduces road user costs. Specifically, Table 4 of Combes and Lafourcade (2005) reports that, in France, the reference cost per km. in a national road with at least 4 lanes is 25% higher than in other national roads. In our road network data for France, the average number of lanes in national roads with at least 4 lanes is 4.43, and the average number of lanes in national roads with less than 4 lanes is 1.9. From this, we infer that adding 2.5 on top of 2 lanes, a 125% increase in the number of lanes, reduces costs by 25%, implying an elasticity of costs with respect to number of lanes of $\chi_{lane} = \frac{25\%}{125\%} = 0.2$ in absolute value. In addition, Table 4 in Combes and Lafourcade (2005) reports that the total reference cost is about 7% higher on "secondary roads" relative to "other national roads", from which we infer $\chi_{use} = 1.07$. According to Figueroa et al. (2013), road user costs are 35% higher on gravel relative to paved roads, implying $\chi_{paved} = 1.35$, and according to Tay and Churchill (2007), adding a median increases speed by 5%, implying $\chi_{median} = 1.05$.

Calibration of β Under the assumption that the transport cost per unit of transported good, $\tau = \delta^{\tau} \frac{Q^{\beta}}{I^{\gamma}}$, is proportional to the travel time and that the flow of goods, Q, is proportional to the flow of vehicles on a highway, we calibrate the elasticity β to empirical observations relating speed of vehicles on highways to observed car density. We use estimates from Wang et al. (2011) who assembled data from various segments of the GA500 route in Georgia, USA. The data was collected at 5min frequency over the span of year 2003 with speed and density computed over 20s windows. The authors estimate the five-parameter logistic relationship

$$v(k,\theta) = v_b + \frac{v_f - v_b}{\left(1 + \exp\left(\frac{k-k_t}{\theta_1}\right)\right)^{\theta_2}},$$

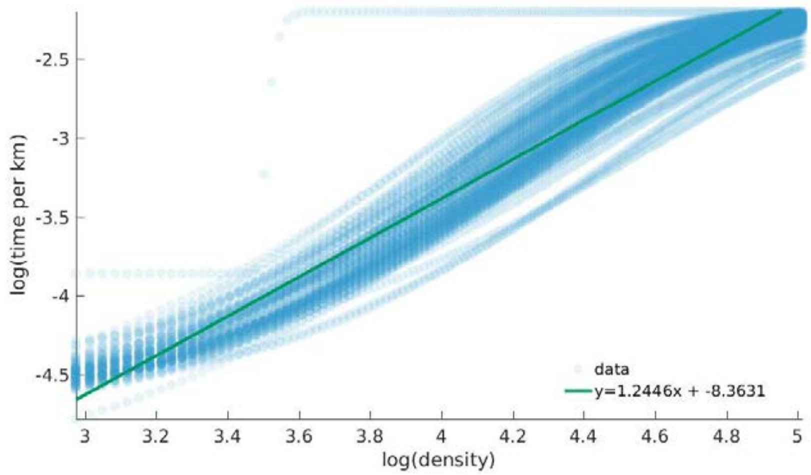
where speed v (km/h) is a function of car density k (cars/km). Parameter v_f is the free flow speed, v_b is the average travel speed at stop-and-go conditions, k_t is the threshold parameter at which traffic transitions from free flow to congested flow, and (θ_1, θ_2) are specific scale and shape parameters. Wang et al. (2011) report estimates of these parameters for 63 sections of the route. We use these estimates to produce artificial observations of speed and density ranging from 18.96 (average threshold k_t at which congestion starts) to 150 cars per km (maximum reported by the authors) for all sections. We then compute the average time per km ($1/v$) and regress its log on the log density to obtain an estimate of the elasticity $\beta = 1.2446$. Figure A.9 below presents the fit of our log-linear model to the data generated by their empirical logistic model.

Calibration of δ_0^{τ} As mentioned in the text, we calibrate the coefficients δ_0^{τ} entering in 22 to match the share of total intra-regional trade to total intra-national trade (sum of intra-regional trade and exports from Spanish regions to other Spanish regions) of 39% reported by Llano et al. (2010).[68] In our model, this summary statistic is:

$$\frac{\sum_j \sum_n P_j^n Y_j^n}{\sum_j \left(\sum_n P_j^n Y_j^n + \sum_{k \in \mathcal{N}(j)} X_{jk}\right)}, \tag{A.6}$$

[68] This is the ratio of the value in the last row of Column 1 of Table 1 in their paper to the sum of that value and the value reported in the last row of Column 2 of that table.

Figure A.9: Fit of linear model on log(time) to log(density)



Notes: The blue scatter plot displays all the pooled artificial observations across the 63 stations on GA500. The green curve is the fitted relationship.

where X_{jk} are total exports from j to $k \in \mathcal{N}(j)$:

$$X_{jk} = \sum_n P_j^n Q_{jk}^n. \tag{A.7}$$

The numerator of this expression is the sum value added in the tradable sector across all regions. Because in the model there are no international flows, this term corresponds to the sum of total intra-regional trade.[69] The denominator equals total intra-national trade, defined as the sum total of intra-regional trade and total exports to other regions. When all regions gross exports equal $x\%$ of their value added, this ratio equal $\frac{1}{1+x}$.[70]

Impact of δ_1^τ on Equilibrium Outcomes We show here that δ_1^τ does not impact the trade-distance elasticity because it does not impact the elasticity of per-unit shipping costs with respect to distance. Consider first a fully symmetric configuration of the model. In logs, the number of units of product n that must be shipped from j_0 for one unit to arrive in j_n through the intervening locations $j_1, .., j_{n-1}$ is:

$$\begin{aligned} log\left(\prod_{i=0}^{N-1}\left(1+\delta_{j_i,j_{i+1}}^\tau \frac{\left(Q_{j_i,j_{i+1}}^n\right)^\beta}{I_{j_i,j_{i+1}}^\gamma}\right)\right) &\simeq \sum_{i=0}^{N-1} \delta_{j_i,j_{i+1}}^\tau \frac{\left(Q_{j_i,j_{i+1}}^n\right)^\beta}{I_{j_i,j_{i+1}}^\gamma} \\ &= DIST_{j_0,j_N} \Delta_0 \frac{(Q^n)^\beta}{I^\gamma}, \end{aligned} \tag{A.8}$$

where $\Delta_0 \equiv \delta_0^\tau dist^{\delta_1^\tau - 1}$ is a constant, $DIST_{j_0,j_N} = N * dist$ is the total distance between locations j_0 and j_n, and where $dist$ is the distance between any two connected locations. The approximation in the first line follows from

[69] Specifically, if we let $D_j = \sum_n P_j^n Y_j^n + M_j - X_j$ be the domestic absorption of region j, where M_j are region j imports and X_j are region j exports, then intra-regional trade at the country level is $\sum_j D_j = \sum_j \sum_n P_j^n Y_j^n$ because, each country being a closed economy, $\sum_j M_j = \sum_j X_j$

[70] I.e., this ratio can be defined as $\frac{\sum D_j}{\sum D_j + \sum_j X_j}$, where $D_j = \sum_n P_j^n Y_j^n + M_j - X_j$ is the domestic absorption of region j, $X_j = \sum_n \sum_{k \in N(j)} P_j^n Q_{jk}^n$ are gross exports, and $M_j = \sum_n \sum_{i \in N(j)} P_j^n Q_{ij}^n$ are gross imports. If each region gross exports is fraction x of its value added, then $D_j = X\left(\frac{1}{x} - 1\right) + M_j$, hence the ratio of intra-regional to intra-national trade becomes $\frac{1}{1+x}$. If each region openness coefficient is $\frac{X_j + M_j}{\sum_n P_j^n Y_j^n} = x$, then then $D_j = \frac{X_j + M_j}{x} + M_j - X_j$, hence the ratio of intra-regional to intra-national trade becomes $\frac{1}{1+x/2}$.

assuming that per-unit shipping costs between connected locations in our model are not large, and the second line follows from assuming symmetry, $Q^n_{j_i,j_{i+1}} = Q^n$, $I_{j_i,j_{i+1}} = I$, and $dist_{j_i,j_{i+1}} = dist$. The expression above implies that, in the model, the elasticity of per-unit transport costs between any two faraway locations j_0 and j_n to the distance $DIST_{j_0,j_n}$ is equal to 1. This means that δ_1^τ impacts the level of per-unit costs, but not the elasticity of per-unit costs with respect to distance in the cross-section. It also implies that δ_1^τ and δ_0^τ impact the overall level trade costs through the constant Δ_0. Our calibration strategy chooses δ_0^τ to match the intra-regional trade share. Furthermore, we can note that δ_1^τ and δ_0^τ only impact the economy through Δ_0, so that, once we have matched the intra-regional trade share the value of δ_1^τ is not relevant for any equilibrium outcome.

Our assumption that $dist_{j_i,j_{i+1}}$ is constant is not a bad approximation to our actual implementation since all cells are equally-sized. However, the equilibria that we study through the paper are clearly asymmetric. In that case, (A.8) becomes:

$$DIST_{j_0,j_N}\Delta_0 \left(\frac{1}{N} \sum_{i=0}^{N-1} \frac{\left(Q^n_{j_i,j_{i+1}}\right)^\beta}{I^\gamma_{j_i,j_{i+1}}} \right).$$

Hence, as long as the average per-unit cost over links (the term between parenthesis in the last expression) does not vary systematically with the total distance of shipments, the model preserves the property we just described.

Type of Road	Local	Secondary	Primary	National
Fraction of Km. of Road Network	10%	49%	31%	9%
Average Number of Lanes	1.84	2.18	1.92	4.12
Standard Deviation of Number of Lanes	0.24	0.18	0.49	0.49
% with a median	0.4%	0.5%	4.4%	92.7%
% paved	88%	98%	100%	100%

Table A.1: Average Features of the Road Network across Countries, by Type of Road

Note: The table reports summary statistics from the EuroRegionalMap by EuroGeographics. The table reports the average of each summary statistic across the 25 countries included in our data.

Country	Code	Actual Road Network			Discretization		
		Length (Km.)	Number of Segments	Average Lanes per Km.	Number of Cells	Length (Km.)	Average Infrastructure Index
		(1)	(2)	(3)	(4)	(5)	(6)
Austria	AT	17230	6161	2.36	46	9968	1.54
Belgium	BE	19702	10496	2.49	20	3400	1.80
Cyprus	CY	2818	947	2.29	30	3653	0.81
Czech Republic	CZ	28665	10194	2.20	48	10935	0.99
Denmark	DK	11443	4296	2.18	21	4102	1.11
Finland	FI	70394	9221	2.04	73	34262	0.29
France	FR	128822	38699	2.05	276	78405	1.45
Georgia	GE	28682	9009	1.95	32	7895	0.38
Germany	DE	115177	66428	2.42	196	56410	1.80
Hungary	HU	32740	9244	2.10	50	12017	0.90
Ireland	IE	24952	4144	2.10	47	11299	0.63
Italy	IT	77608	44159	2.32	126	32640	1.57
Latvia	LV	11495	2103	2.03	47	10599	0.34
Lithuania	LT	10682	1586	2.39	43	9736	0.58
Luxembourg	LU	1779	866	2.31	8	674	0.78
Macedonia	MK	5578	908	2.15	14	2282	0.26
Moldova	MD	8540	1462	2.21	20	4611	0.40
Netherlands	NL	14333	8387	2.68	20	4263	2.62
Northern Ireland	ND	7087	1888	2.18	12	1714	0.81
Portugal	PT	15034	4933	2.10	43	11000	1.43
Serbia	RS	18992	3656	2.14	45	12261	0.68
Slovakia	SK	11420	2610	2.18	30	6130	0.87
Slovenia	SI	7801	2441	2.19	12	1853	1.87
Spain	ES	101990	18048	2.39	227	68162	1.30
Switzerland	CHLI	14526	11102	2.26	25	5127	1.37

Table A.2: Summary Statistics of Actual and Discretized Road Network by Country

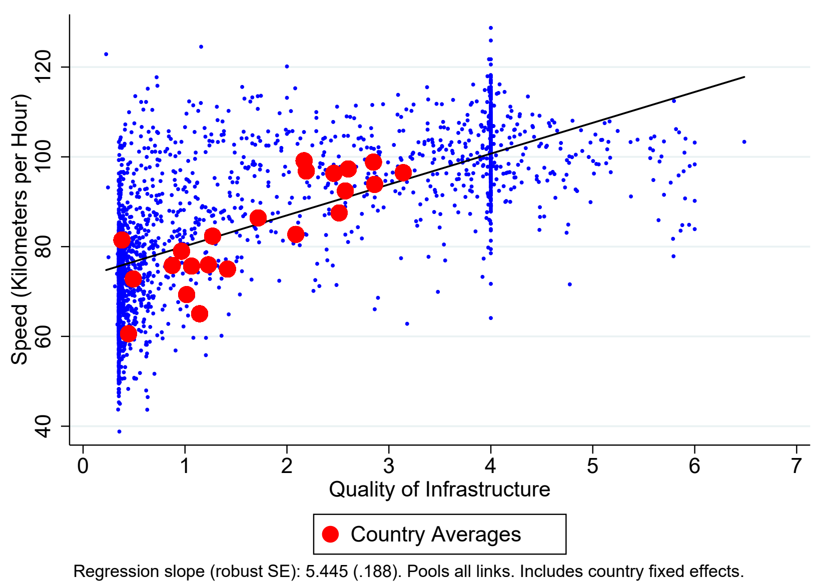
Note: Columns (1) to (3) report statistics from EuroRegionalMap, and Columns (4) to (6) report statistics from the discretization of road networks described in Section 5.1.

Value of γ:	0.5β		β		1.5β	
Labor:	Fixed	Mobile	Fixed	Mobile	Fixed	Mobile
Mean (SD) Trade-Dist. Elasticity	-1.08 (0.17)	-1.11 (0.21)	-1.12 (0.17)	-1.19 (0.26)	-1.16 (0.22)	-1.17 (0.27)
Mean (SD) Intra-regional Share	0.54 (0.13)	0.48 (0.09)	0.57 (0.14)	0.52 (0.11)	0.60 (0.17)	0.54 (0.14)
Calibrated δ_0^τ	1.20	2.74	2.19	6.39	4.33	9.90

Table A.3: Trade-Distance Elasticity, Intra-regional Trade Share, and Calibrated δ_0

Note: The table reports the mean and standard deviation of the trade-distance and elasticity and intra-regional trade share across the 25 countries in our data in the calibrated model under each value of γ and each assumption for labor mobility, as well as the calibrated value of δ_0^τ in each case. To compute the trade-distance elasticity we run, in the calibrated model: $\ln(X_{jk}) = a_0 \ln(GDP_j) + a_1 \ln(GDP_k) + b \ln(dist_{jk}) + \varepsilon_{jk}$, where X_{jk} are exports from j to k, GDP_j is the GDP of region j defined in Footnote 52 , and $dist_{jk}$ is the geographic distance between locations j and k. The intra-regional trade share is computed using the definition in (A.6) below.

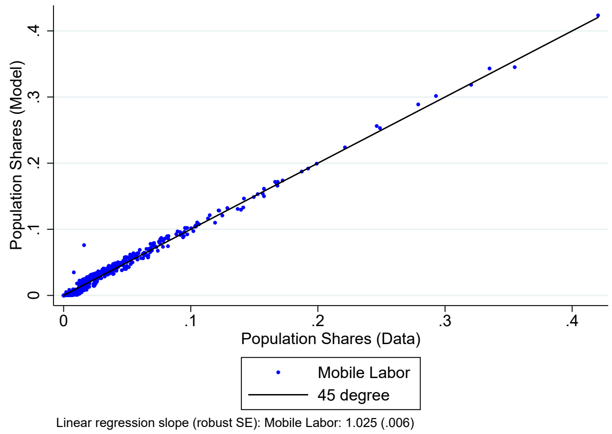
Figure A.10: Quality of Infrastructure Measure and Speed



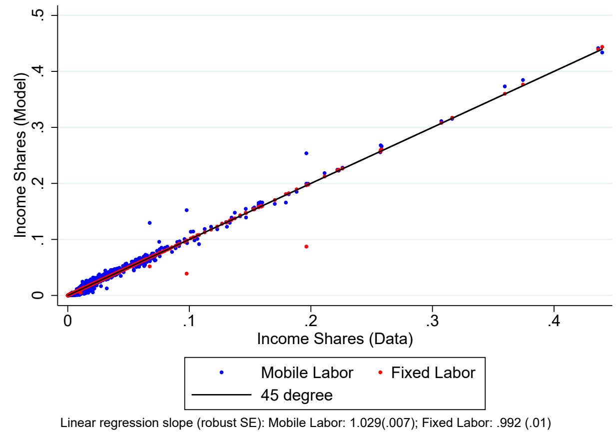
Notes: The figure shows the speed, according to GoogleMaps, on the fastest route linking connected pairs of nodes in the discretized network, against our measure of infrastructure I_{jk}^{obs}. The figure pools all links across all countries. The red circles correspond to the average infrastructure and speed across all links within each country.

Figure A.11: Calibration of Population and Income Shares, All Locations and Countries

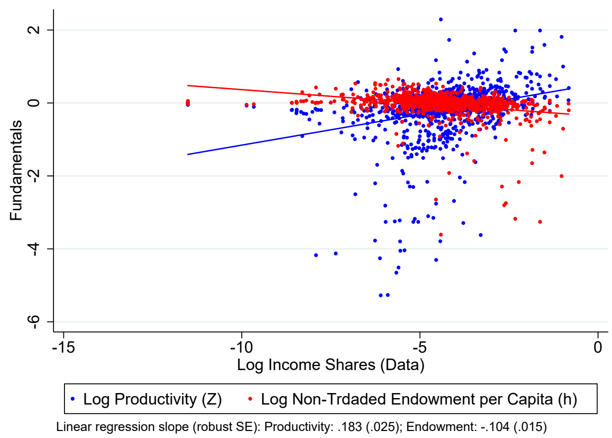
(a) Population Shares in Model and Data



(b) Income Shares in Model and Data

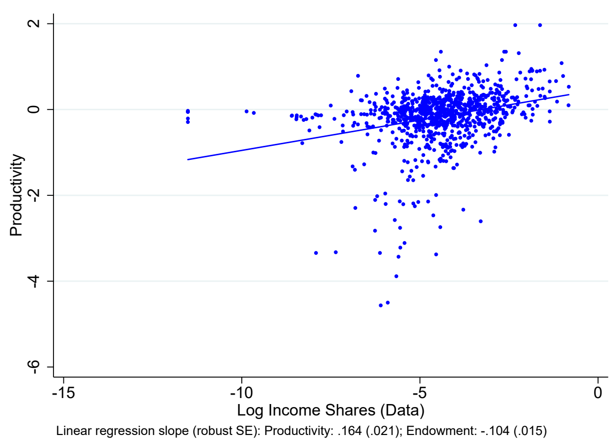


(c) Fundamentals and Income Shares, Mobile Labor



(d) Fundamentals and Population Shares, Mobile Labor

(e) Fundamentals and Income Shares, Fixed Labor



Notes: All the figures pool the 1511 cells from the 25 countries when $\gamma = \beta$. Similar relationships hold for the alternative values of γ assumed in the calibration. In the Panels (c) to (e), log-productivity and log-endowment of the non-traded good per capita are demeaned within each country.

	Low Gamma			Middle Gamma			High Gamma		
Counterfactual	Misallocation	Expansion	Expansion	Misallocation	Expansion	Expansion	Misallocation	Expansion	Expansion
Building Costs	GEO	GEO	FOC	GEO	GEO	FOC	GEO	GEO	FOC
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Austria	4.0%	4.8%	1.3%	5.6%	6.9%	4.6%	6.6%	8.9%	11.6%
Belgium	2.3%	3.0%	1.0%	3.2%	4.6%	2.2%	4.2%	5.6%	6.5%
Cyprus	2.7%	3.6%	1.5%	3.7%	5.0%	3.7%	4.5%	6.7%	9.5%
Czech Republic	2.2%	2.8%	1.0%	3.3%	4.4%	2.5%	4.5%	6.2%	10.4%
Denmark	2.2%	2.9%	1.1%	2.7%	3.8%	3.0%	2.8%	4.5%	7.0%
Finland	3.3%	4.3%	1.2%	4.7%	6.3%	5.6%	5.1%	7.2%	7.8%
France	2.4%	3.2%	1.2%	3.4%	4.7%	3.2%	4.2%	6.1%	7.4%
Georgia	3.9%	4.4%	1.3%	6.7%	6.9%	4.3%	8.4%	8.5%	23.1%
Germany	1.2%	1.8%	0.9%	1.6%	2.4%	1.7%	1.8%	2.7%	3.3%
Hungary	4.1%	4.9%	1.2%	5.2%	6.7%	3.3%	6.6%	8.6%	10.1%
Ireland	3.9%	4.4%	1.1%	6.4%	7.2%	4.1%	7.9%	9.6%	14.9%
Italy	2.6%	3.9%	1.6%	2.9%	5.3%	3.8%	3.1%	5.8%	6.4%
Latvia	5.0%	5.4%	1.2%	8.2%	8.3%	4.1%	10.2%	10.6%	21.1%
Lithuania	3.1%	3.9%	1.4%	4.6%	6.2%	4.7%	6.2%	8.6%	19.1%
Luxembourg	0.4%	0.9%	0.7%	0.8%	1.9%	1.5%	1.3%	2.9%	5.0%
Macedonia	3.1%	3.5%	1.0%	3.3%	4.0%	2.8%	2.8%	4.0%	3.2%
Moldova	3.1%	4.0%	1.1%	3.7%	5.3%	2.9%	3.7%	5.6%	5.4%
Netherlands	2.1%	2.7%	1.0%	3.4%	4.3%	2.2%	4.0%	5.2%	6.6%
Northern Ireland	1.5%	2.1%	1.0%	2.2%	3.4%	2.2%	3.4%	5.0%	8.8%
Portugal	2.7%	3.4%	1.0%	3.6%	5.1%	2.5%	4.0%	6.2%	6.4%
Serbia	11.4%	11.3%	2.6%	18.1%	18.0%	7.7%	24.1%	25.0%	49.4%
Slovakia	4.7%	5.2%	1.4%	6.2%	7.0%	3.9%	5.7%	8.2%	16.4%
Slovenia	2.6%	3.2%	1.1%	3.2%	4.4%	2.3%	3.6%	4.9%	6.2%
Spain	3.0%	4.1%	1.5%	3.8%	6.1%	4.1%	4.2%	7.4%	8.5%
Switzerland	3.5%	4.1%	1.1%	4.9%	6.0%	3.2%	6.2%	7.4%	9.1%
Average	3.2%	3.9%	1.2%	4.6%	5.8%	3.4%	5.6%	7.2%	11.3%

Table A.4: Welfare Gains From Optimal Reallocation or Expansion of Current Networks, Fixed Labor

	Low Gamma			Middle Gamma			High Gamma		
Counterfactual	Misallocation	Expansion	Expansion	Misallocation	Expansion	Expansion	Misallocation	Expansion	Expansion
Building Costs	GEO	GEO	FOC	GEO	GEO	FOC	GEO	GEO	FOC
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Austria	3.9%	4.6%	1.2%	5.6%	6.8%	7.7%	6.7%	8.5%	10.7%
Belgium	1.3%	1.6%	0.6%	1.7%	2.2%	1.6%	2.0%	2.4%	2.6%
Cyprus	1.7%	2.1%	0.9%	2.2%	2.9%	3.2%	2.7%	3.8%	5.4%
Czech Republic	2.0%	2.4%	0.9%	3.6%	4.4%	4.0%	4.9%	6.1%	9.6%
Denmark	0.9%	1.2%	0.5%	1.4%	1.8%	2.1%	1.5%	2.0%	2.9%
Finland	3.9%	5.0%	1.3%	4.8%	6.8%	5.9%	5.8%	9.3%	8.5%
France	3.3%	4.3%	1.5%	4.7%	6.6%	8.2%	5.7%	8.5%	11.6%
Georgia	3.7%	4.1%	1.2%	7.1%	7.2%	10.5%	10.9%	10.7%	22.4%
Germany	0.9%	1.4%	0.7%	1.5%	2.2%	2.2%	1.4%	2.0%	2.3%
Hungary	4.8%	5.7%	1.5%	5.0%	6.3%	5.1%	7.4%	9.7%	14.7%
Ireland	3.9%	4.4%	1.1%	6.6%	7.3%	8.1%	9.0%	9.7%	14.9%
Italy	2.0%	2.9%	1.2%	2.4%	4.1%	3.6%	2.3%	4.1%	4.2%
Latvia	6.0%	6.3%	1.4%	12.0%	11.8%	13.3%	19.4%	17.9%	42.2%
Lithuania	3.0%	3.7%	1.3%	4.6%	6.2%	9.5%	8.7%	11.1%	21.6%
Luxembourg	0.3%	0.6%	0.4%	0.5%	1.2%	1.0%	1.0%	2.4%	3.9%
Macedonia	1.2%	1.3%	0.4%	2.7%	3.3%	2.3%	3.6%	5.0%	4.2%
Moldova	2.4%	2.9%	0.8%	4.2%	5.8%	4.8%	5.3%	7.8%	7.3%
Netherlands	0.7%	0.9%	0.4%	1.1%	1.4%	1.1%	0.9%	1.1%	1.3%
Northern Ireland	0.8%	1.1%	0.6%	1.3%	1.8%	1.7%	2.2%	2.9%	4.6%
Portugal	2.6%	3.3%	1.0%	3.5%	4.8%	3.9%	4.2%	5.9%	5.9%
Serbia	17.6%	17.5%	3.3%	29.8%	28.5%	27.1%	43.7%	43.8%	78.7%
Slovakia	3.2%	3.5%	1.0%	5.0%	5.6%	5.6%	5.2%	7.2%	12.7%
Slovenia	1.0%	1.2%	0.5%	1.3%	1.8%	1.4%	1.4%	1.9%	2.3%
Spain	3.0%	4.3%	1.6%	3.9%	6.4%	6.4%	4.3%	7.8%	9.3%
Switzerland	2.1%	2.4%	0.7%	3.5%	4.1%	4.0%	4.2%	5.0%	5.4%
Average	3.0%	3.5%	1.0%	4.8%	5.7%	5.8%	6.6%	7.9%	12.4%

Table A.5: Welfare Gains From Optimal Reallocation or Expansion of Current Networks, Mobile Labor

Table A.6: Optimal Infrastructure Investment, Population Growth and Local Characteristics for Different Number Sectors

Number Sectors	N=5		N=10		N=15	
Dependent variable:	Investment	Pop. Growth	Investment	Pop. Growth	Investment	Pop. Growth
	(1)	(2)	(3)	(4)	(5)	(6)
Population	0.082***	0.000	0.104***	0.002**	0.107***	0.002***
Income per Capita	-0.074	0.009	0.007	-0.002	-0.035	0.001
Consumption per Capita	0.307***	-0.158***	0.179***	-0.134***	0.267***	-0.143***
Infrastructure	-0.179***	0.002	-0.195***	-0.001	-0.206***	-0.000
Differentiated Producer	0.254***	0.029***	0.133***	0.028***	0.101***	0.025***
R^2	0.28	0.67	0.32	0.65	0.34	0.63

Each column corresponds to a different regression pooling all locations in the optimal expansion counterfactual across the 25 countries assuming $\gamma = \beta$, mobile labor, and $\delta = \delta^{I,GEO}$. All regressions include country fixed effects. Standard errors are clustered at the country level. ***=1% significance, **=5%, *=10%. Dependent variables: Investment is defined as $\Delta \ln \overline{I_j}$, where $\overline{I_j} = \frac{1}{\#\mathcal{N}(j)} \sum_{k \in \mathcal{N}(j)} I_{jk}$ is the average level of infrastructure across all the links of location j, and population growth is defined as $\Delta \ln L_j$, where $\Delta \ln x$ denotes the difference between the log of variable x in the counterfactual and in the calibrated allocation. Independent variables: all correspond to the log of the level of each variable in the calibrated model. Population and income per capita are the two outcomes matched by the calibration. Consumption per capita corresponds to traded goods, c_j in the model. Infrastructure is the average infrastructure of each location, $\overline{I_j}$. Differentiated producer is a dummy for whether the location is a producer of differentiated goods in the calibration.

Table A.7: Average Welfare Gains Across Countries for Different Number of Sectors

Number of Sectors	N=5		N=10		N=15	
Labor:	Fixed	Mobile	Fixed	Mobile	Fixed	Mobile
Optimal Reallocation						
$\delta = \delta^{I,GEO}$	3.5%	3.7%	4.6%	4.8%	4.6%	4.7%
Optimal Expansion						
$\delta = \delta^{I,GEO}$	4.4%	4.3%	5.8%	5.7%	5.8%	5.6%
$\delta = \delta^{I,FOC}$	5.7%	2.6%	3.4%	5.8%	3.7%	5.9%

Each element of the table shows the average welfare gain in the corresponding counterfactual across the 25 countries for the case $\gamma = \beta$.

D Online Appendix I: Auxiliary Material to Proposition 3

D.1 Definitions

Let $G = (\mathcal{I}, \mathcal{E})$ be an undirected graph. We say that a **path** of length $n \in \mathbb{N}^*$ from a node $a \in \mathcal{I}$ to $b \in \mathcal{I}$ is a finite sequence of nodes $(i_1, \ldots, i_n)$ such that $i_k \in \mathcal{I}$ for $1 \leq k \leq n$, $i_1 = a$ and $i_n = b$ and $\{i_k, i_{k+1}\} \in \mathcal{E}$. A **simple path** is a path that contains no repeated node, i.e., $i_k \neq i_l$ for all $1 \leq k, l \leq n$ and $k \neq l$. A **cycle** of length n is a path $p = (i_1, \ldots, i_n)$ such that $i_1 = i_n$. A **simple cycle** of length n is a cycle that contains no repeated node other than the starting and ending nodes, i.e., $i_k \neq i_l$ for $1 \leqslant k, l \leqslant n-1$ and $k \neq l$. A **tree** is a connected graph such that has no simple cycle. Equivalently, in a tree, there is a unique simple path connecting any two nodes.

D.2 Propositions and Lemmas

Proposition 5. *$\mathcal{E}^*$ is a tree.*

Proof. Because node 0 is the unique productive center and there is an Inada condition in consumption, there must exist a path connecting each node to 0. Hence, $\mathcal{E}^*$ must be connected. It remains to show that $\mathcal{E}^*$ cannot have simple cycles. We proceed by contradiction. Assume there exists a simple cycle $p = (i_1, \ldots, i_n)$.Figure A.12 illustrates the different types of cycles that can arise. Case (i) is a cycle with circular flows that run in only one direction. Lemma 1 tells us that such cycle cannot arise if $(\mathbf{c}^*, \mathbf{Q}^*)$ is locally optimal, as they inefficiently waste goods in transportation. Cases (ii) and (iii) correspond to cycles along which flows run into different directions. Lemma 3 establishes that whenever there is a cycle of type (iii), then there must exists a cycle of type (ii). We conclude with Lemma 4 by showing that cycles of type (ii) cannot arise if $(\mathbf{c}^*, \mathbf{Q}^*)$ is locally optimal. The reason is that one is better off redirecting flows into one of the two branches because of economies of scale in the transport technology when $\gamma > \beta$. Hence, simple cycles may not exist and $\mathcal{E}^*$ is a tree.

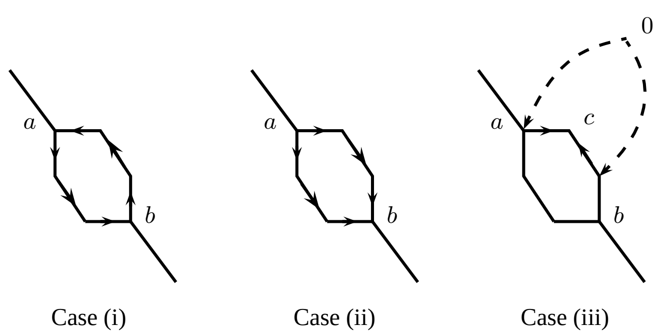


Figure A.12: Different types of simple cycles

□

Lemma 1. *If $(\mathbf{c}^*, \mathbf{Q}^*)$ is a local optimum with $(\mathcal{I}, \mathcal{E}^*)$ its associated graph, then there exists no simple cycle $p = (i_1, \ldots, i_n)$ such that $Q^*_{i_k, i_{k+1}} > 0$ for all $1 \leqslant k \leqslant n-1$.*

Proof. Case (i) in Figure A.12 presents the type of cycle with circular flows that cannot exist in a local optimum. By contradiction, assume that there exists such a cycle $p = (i_1, \ldots, i_n)$. Then, for $\varepsilon > 0$ small, consider the allocation of flows

$$Q^{\varepsilon}_{jk} = \begin{cases} Q_{jk} - \varepsilon & \text{if } \exists l, 1 \leqslant l \leqslant n-1, \text{ such that } j = i_l \text{ and } k = i_{l+1} \\ Q^*_{jk} & \text{elsewhere} \end{cases}.$$

If $\varepsilon \leqslant \min_{1 \leqslant k \leqslant n-1} Q_{i_k,i_{k+1}}$, then $(\{c_j\}, \{Q^{\varepsilon}_{jk}\})$ is a feasible allocation that is strictly preferable to $(\{c_j\}, Q^*)$ since it yields the same utility at a lower transport cost. Hence, the gradient of the Lagrangian with respect to ε is strictly greater than 0 (recall that $P_j = u_c(c_j, h_j) > 0$), contradicting the assumption that $(\mathbf{c}^*, \mathbf{Q}^*)$ is a local optimum.

□

Lemma 2. *For every node $a \in \mathcal{I}$ distinct from the productive center $0 \in \mathcal{I}$ and such that $L_a > 0$, there exists a simple path $p = (i_1, \ldots, i_n)$ that connects 0 to a and such that $Q_{i_k,i_{k+1}} > 0$ for $1 \leqslant k \leqslant n-1$.*

Proof. The proof is constructive. We build a simple path $p = (i_1, \ldots, i_n)$ with $i_1 = a$, $i_k \neq i_l$ for all $1 \leqslant k, l \leqslant n$ and $k \neq l$ and such that $Q_{i_k,i_{k-1}} > 0$. We proceed by recursion on the length of path p, which we denote by $|p| = n$. We start the recursion by setting $i_1 = a$. Because of the Inada conditions in the utility function and $L_a > 0$, we know that $c_a L_a > 0$. The balanced flow constraint in a,

$$c_a L_a = \sum_{k \in \mathcal{N}(a)} Q_{ka} - \sum_{k \in \mathcal{N}(a)} Q_{ak} \left[1 + \delta^{\tau}_{ak} \frac{Q^{\beta}_{ak}}{I^{\gamma}_{ak}}\right] > 0,$$

tells us that location a must be a net recipient of goods from its neighbors. Hence, there exists $k \in \mathcal{N}(a)$ such that $Q_{ka} > 0$. Let $i_2 = k$. If $i_2 = 0$, then we have found a simple path connecting a to 0 with positive flows from 0 to a. If not, we now have a path $p_2 = (i_1, i_2)$ of length 2 such that $i_1 = a$, $i_1 \neq i_2 \neq 0$ and $Q_{i_2 i_1} > 0$. Assume now that $n \geqslant 2$ and, by recursion hypothesis, that we have a path $p_n = (i_1, \ldots i_n)$ with $i_1 = a$, $i_k \neq i_l \neq 0$ for all $1 \leqslant k, l \leqslant n$ and $k \neq l$ and such that $Q_{i_k,i_{k-1}} > 0$. Consider location i_n. The balanced flow constraint at i_n tells us that

$$c_{i_n} L_{i_n} = \sum_{k \in \mathcal{N}(i_n)} Q_{k,i_n} - \sum_{k \in \mathcal{N}(i_n)} Q_{i_n,k} \left[1 + \delta^{\tau}_{i_n,k} \frac{Q^{\beta}_{i_n,k}}{I^{\gamma}_{i_n,k}}\right] \geqslant 0.$$

Since we know by recursion hypothesis that $Q_{i_n,i_{n-1}} > 0$, then there exists a $k \in \mathcal{N}(i_n)$ such that $Q_{k,i_n} > 0$. We know that $k \neq i_l$ for all $1 \leqslant l \leqslant n$ because otherwise there would exists a cycle with circular flows, which is ruled out by Lemma 1. If $k = 0$, then we have found a path $p_{n+1} = (i_1, \ldots, i_n, 0)$ that connects a to 0 with only positive flows from 0 to a. If not, then set $i_{n+1} = k$. We then have a path $p_{n+1} = (i_1, \ldots i_{n+1})$ with $i_1 = a$, $i_k \neq i_l \neq 0$ for all $1 \leqslant k, l \leqslant n+1$ and $k \neq l$ and such that $Q_{i_k,i_{k-1}} > 0$.

We conclude as follows. Since $\mathcal{I}$ is finite, the above recursion must finish in a finite number of iterations. Since the recursion only stops after finding a path that ends in 0, then there must exists a simple path p of size $n < |\mathcal{I}|$ with $p = (i_1, \ldots, i_n)$ such that $i_1 = a$, $i_n = 0$ and $Q_{i_k,i_{k+1}} > 0$ for $1 \leqslant k \leqslant n-1$. By construction, the path $\tilde{p} = (i_n, i_{n-1}, \ldots, i_1)$ proves the statement.

□

Lemma 3. *Assume there exists a simple cycle $p = (i_1, \ldots, i_n)$. Then, there exists $(a, b) \in \mathcal{I}^2$, $a \neq b$, such that there exists two distinct simple paths from a to b, $p_1 = (i^1_k)_{1 \leq k \leq n_1}$ and $p_2 = (i^2_k)_{1 \leq k \leq n_2}$ with $i^1_1 = i^2_1 = a$ and $i^1_{n_1} = i^2_{n_2} = b$, such that the flows are strictly positive from a to b along both paths, i.e., $Q^*_{i^l_k, i^l_{k+1}} > 0$ for $l \in \{1, 2\}$ and $1 \leq k \leq n_l - 1$.*

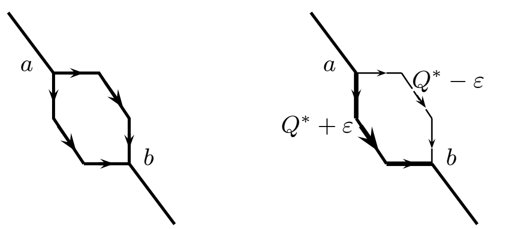
Proof. The objective of this lemma is to establish that if there exists a simple cycle, then there must exist a cycle of type (ii) as illustrated on Figure A.12.

Consider the simple cycle $p = (i_1, \ldots, i_n)$. For convenience of notation, denote $i_0 = i_{n-1}$ and $i_{n+1} = i_2$. Denote $\tilde{Q}_{i_k,i_{k+1}} = Q^*_{i_k,i_{k+1}} - Q^*_{i_{k+1},i_k}$ the net flow from i_k to i_{k+1} for $0 \leqslant k \leqslant n$, which can be either strictly positive or strictly negative. We know from Lemma 1 that the net flows $\tilde{Q}_{i_k,i_{k+1}}$ cannot have the same sign, otherwise we would have a cycle with circular flow, violating the local optimality condition of (c^*, Q^*). Hence, there must exist $1 \leqslant k \leqslant n$ such that $\tilde{Q}_{i_{k-1},i_k} > 0$ and $\tilde{Q}_{i_k,i_{k+1}} < 0$. Node k is a location that receives goods from its two neighbors on the cycle, as illustrated by node c in case (iii) of Figure A.12. Set $a = 0$ and $b = i_k$. We know from Lemma 2 that there exists a path $p_1 = (j^1_1, \ldots, j^1_{n_1})$ such that $j^1_1 = a = 0$, $j^1_{n_1} = i_{k-1}$ and $\tilde{Q}_{j^1_l, j^1_{l+1}} > 0$ for all $1 \leqslant l \leqslant n_1$. Similarly, there exists a path $p_2 = (j^2_1, \ldots, j^2_{n_2})$ such that $j^2_1 = a = 0$, $j^2_{n_2} = i_{k+1}$ and $\tilde{Q}_{j^2_l, j^2_{l+1}} > 0$ for all $1 \leqslant l \leqslant n_2$. We now argue that the paths $\tilde{p}_1 = \left(j_1^1, \ldots, j_{n_1}^1, b\right)$ and $\tilde{p}_2 = \left(j_1^2, \ldots, j_{n_2}^2, b\right)$ are two distinct simple paths from a to b with strictly positive flows. By construction, we know that $\tilde{Q}_{j_{n_1}^1, i_k} > 0$ and $\tilde{Q}_{j_{n_2}^2, i_k} > 0$ so that the flows are strictly positive along both paths. We must only check that they are simple paths, i.e., that the nodes are not repeated. Let us treat the case of $\tilde{p}_1$. The other one follows symmetrically. We must show in particular that there is no l with $1 \leqslant l \leqslant n_1$ such that $j_l^1 = b$. If this was the case, then $\left(b, , j_{l+1}^1 \ldots, j_{n_1}^1, b\right)$ would be a cycle with circular flows running in the same direction, which Lemma 1 rules out. Hence, $\tilde{p}_1$ is a simple path.

□

Lemma 4. *For all $(a,b) \in \mathcal{I}^2$, $a \neq b$, if there are two simple paths p_1 and p_2 connecting a to b, i.e., $p_1 = \left(i_k^1\right)_{1 \leq k \leq n_1}$ and $p_2 = \left(i_k^2\right)_{1 \leq k \leq n_2}$ with $i_1^1 = i_1^2 = a$ and $i_{n_1}^1 = i_{n_2}^2 = b$, such that $Q_{i_k^l, i_{k+1}^l} > 0$ for $l \in \{1,2\}$ and $1 \leq k \leq n_l$, then $p_1 = p_2$.*

Proof. The objective of this lemma is to show that cycles of the type (ii) in Figure A.12 cannot exist. Assume by contradiction that such a cycle exists and that $p_1 \neq p_2$. Note that we can assume WLOG that $i_k^1 \neq i_l^2$ for all $1 < k < n_1$ and $1 < l < n_2$. To see this, let $\underline{k} = \min\left\{k | i_{k+1}^1 \neq i_{k+1}^2\right\}$ and $\overline{k}_1 = \min\left\{k > \underline{k} | \exists l > \underline{k}, i_k^1 = i_l^2\right\}$ and $\overline{k}_2$ be such that $i_{\overline{k}_1}^1 = i_{\overline{k}_2}^2$. By construction, the path $p_1' = \left(i_{\underline{k}}^1, \ldots, i_{\overline{k}_1}^1\right)$ and $p_2' = \left(i_{\underline{k}}^2, \ldots, i_{\overline{k}_2}^2\right)$ are two paths such that $i_{\underline{k}}^1 = i_{\underline{k}}^2$, $i_{\overline{k}_1}^1 = i_{\overline{k}_2}^2$, and $i_k^1 \neq i_l^2$ for all $\underline{k} < k < \overline{k}_1$ and $\underline{k} < l < \overline{k}_2$.



(a) Initial (b) After redirection

Figure A.13: Redirecting the flows to one branch

We are now going to show that $p_1 \neq p_2$ leads to a contradiction. The idea behind the proof is illustrated in Figure A.13 below. We are going to show that if there exists two distinct simple paths with positive flows going from a to b, then it would be strictly preferable to redirect the flows from one branch to the other due to the non-concavity of the Lagrangian, violating the local optimality of $(\mathbf{c}^*, \mathbf{Q}^*)$. Consider the allocation $Q^\varepsilon = \left\{Q_{jk}^\varepsilon\right\}$ for $\varepsilon \in \mathbb{R}$ such that

$$Q_{jk}^\varepsilon = \begin{cases} Q_{jk}^* + \varepsilon & \text{if } \exists l \text{ such that } j = i_l^1 \text{ and } k = i_{l+1}^1 \\ Q_{jk}^* - \varepsilon & \text{if } \exists l \text{ such that } j = i_l^2 \text{ and } k = i_{l+1}^2 \\ Q_{jk}^* & \text{elsewhere.} \end{cases}$$

In other words, Q_{jk}^ε corresponds to the pattern of flows Q_{jk}^* where a volume ε of the flows going through path 2 are redirected through path 1. By construction, Q_{jk}^ε is feasible (we are redirecting a fraction of flows that were running through locations on path 2 but not serving any of these locations). In particular, it leaves the value of the Lagrangian (A.4) unchanged except through the term $\left[\sum_{j,k} \hat{\delta}_{jk}^{\frac{\gamma}{\gamma+1}} \left(P_j Q_{jk}^{1+\beta}\right)^{\frac{1}{\gamma+1}}\right]^{\gamma+1}$ where $\hat{\delta}_{jk} = \delta_{jk}^I / \delta_{jk}^\tau$.

Consider the derivative of the Lagrangian with respect to ε:

$$\frac{\partial \mathscr{L}}{\partial \varepsilon} = -(1+\beta)\left[\sum_{j,k} \hat{\delta}_{jk}^{\frac{\gamma}{\gamma+1}} \left(P_j Q_{jk}^{1+\beta}\right)^{\frac{1}{\gamma+1}}\right]^{\gamma} \left[\sum_{1 \leqslant k \leqslant n_1 - 1} \hat{\delta}_{i_k^1 i_{k+1}^1}^{\frac{\gamma}{\gamma+1}} P_{i_k^1}^{\frac{1}{\gamma+1}} Q_{i_k^1 i_{k+1}^1}^{\frac{1+\beta}{1+\gamma}-1} - \sum_{1 \leqslant k \leqslant n_2 - 1} \hat{\delta}_{i_k^2 i_{k+1}^2}^{\frac{\gamma}{\gamma+1}} P_{i_k^2}^{\frac{1}{\gamma+1}} Q_{i_k^2 i_{k+1}^2}^{\frac{1+\beta}{1+\gamma}-1}\right]$$

which satisfies $\frac{\partial \mathscr{L}}{\partial \varepsilon} = 0$ by assumption (local optimum). Let us examine the second order condition:

$$\begin{aligned}\frac{\partial^2 \mathscr{L}}{\partial \varepsilon^2} = &-(1+\beta)\left(\frac{1+\beta}{1+\gamma}-1\right)\left[\sum_{j,k}\hat{\delta}_{jk}^{\frac{\gamma}{\gamma+1}}\left(P_j Q_{jk}^{1+\beta}\right)^{\frac{1}{\gamma+1}}\right]^{\gamma} \times \\ &\left[\sum_{1\leqslant k\leqslant n_1-1}\hat{\delta}_{i_k^1 i_{k+1}^1}^{\frac{\gamma}{\gamma+1}} P_{i_k^1}^{\frac{1}{\gamma+1}} Q_{i_k^1 i_{k+1}^1}^{\frac{1+\beta}{1+\gamma}-2} + \sum_{1\leqslant k\leqslant n_2-1}\hat{\delta}_{i_k^2 i_{k+1}^2}^{\frac{\gamma}{\gamma+1}} P_{i_k^2}^{\frac{1}{\gamma+1}} Q_{i_k^2 i_{k+1}^2}^{\frac{1+\beta}{1+\gamma}-2}\right] \\ &-(1+\beta)^2\frac{\gamma}{\gamma+1}\left[\sum_{j,k}\hat{\delta}_{jk}^{\frac{\gamma}{\gamma+1}}\left(P_j Q_{jk}^{1+\beta}\right)^{\frac{1}{\gamma+1}}\right]^{\gamma-1} \times \\ &\left[\underbrace{\sum_{1\leqslant k\leqslant n_1-1}\hat{\delta}_{i_k^1 i_{k+1}^1}^{\frac{\gamma}{\gamma+1}} P_{i_k^1}^{\frac{1}{\gamma+1}} Q_{i_k^1 i_{k+1}^1}^{\frac{1+\beta}{1+\gamma}-1} - \sum_{1\leqslant k\leqslant n_2-1}\hat{\delta}_{i_k^2 i_{k+1}^2}^{\frac{\gamma}{\gamma+1}} P_{i_k^2}^{\frac{1}{\gamma+1}} Q_{i_k^2 i_{k+1}^2}^{\frac{1+\beta}{1+\gamma}-1}}_{=0}\right]^2 .\end{aligned}$$

Hence, we see that $\frac{\partial^2 \mathscr{L}}{\partial \varepsilon^2} \geqslant 0$ when $\gamma > \beta$. Therefore, the point under consideration cannot be a local maximum. A tiny deviation in either direction for ε would increase welfare, thereby yielding a contradiction.

□