Optimal Transport Networks in Spatial Equilibrium*

Pablo D. Fajgelbaum\textsuperscript{1} and Edouard Schaal\textsuperscript{2}

\textsuperscript{1}UCLA and NBER
\textsuperscript{2}CREI, Universitat Pompeu Fabra, Barcelona GSE and CEPR

July 2019

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 European countries. The counterfactuals suggest considerable losses from suboptimal placement of roads. We also find a significant role for international trade in driving the optimal location of roads within countries.

JEL Classification: F11, O18, R13

\*E-mail: pfajgelbaum@econ.ucla.edu, eschaal@crei.cat. We thank Thomas Chaney, Arnaud Costinot, Don Davis, Jonathan Eaton, Cecile Gaubert, Geoguui Kambourov, Ezra Oberfield, Stephen Redding, Esteban Rossi-Hansberg and Jonathan Vogel for helpful comments. 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), the Barcelona GSE (grant SG-2017-07) and the European Research Council Starting Grant 804095.
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-Óllé (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). 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. 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 (2019) 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 (2019) 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) endogenizes 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. 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. 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). 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). 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. 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.

---

12Note 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.

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


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

16See 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.

17Our 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.\(^\text{18}\) 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 \( J = \{1, \ldots, J\} \). We let \( L_j \) be the number of workers located in \( j \in J \), 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

\[
\begin{align*}
    u &= U(c, h),
\end{align*}
\]

where the utility function \( U \) is homothetic and concave in both of its arguments.

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

\[
\begin{align*}
    c_j &= \frac{C_j}{L_j},
\end{align*}
\]

---

\(^{18}\)See also Coşar and Demir (2016) and Martincus et al. (2017) for empirical studies of how infrastructure investments impact international shipments.
where $C_j$ is the aggregate demand of the traded goods bundle in location $j$.

There is a discrete set of tradable sectors $n = 1, \ldots, N$, combined into $C_j$ through a homogeneous of degree 1 and concave aggregator,

$$C_j = D_j (D_1^j, \ldots, D_N^j),$$

where $D_n^j$ is the total quantity of sector $n$’s output used in location $j$. The typically assumed CES aggregator is a special case of this technology.

**Production** The supply-side of the economy corresponds to a standard neoclassical economy. In addition to labor, there is a fixed supply $V_j = (V_1^j, \ldots, V_M^j)$ of primary factors $m = 1, \ldots, M$ 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:

$$Y_n^j = F_n^j (L_n^j, V_n^j, X_n^j),$$

where $L_n^j$ is the number of workers, $V_n^j = (V_1^{jn}, \ldots, V_M^{jn})$ is the quantity of other primary factors, and $X_n^j = (X_1^{jn}, \ldots, X_N^{jn})$ is the quantity of each sector’s output allocated to the production of sector $n$ in location $j$. The production function $F_n^j$ is either neoclassical (constant returns to scale, increasing and concave in all its arguments) or a constant (endowment economy).

**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, 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 transit through several locations before reaching a point where they are consumed or used as intermediate input. We let $Q_{jk}^n$ be the quantity of goods in sector $n$ shipped from $j$ to $k \in \mathcal{N}(j)$, regardless of where the good was produced. Transporting each unit of good $n$ from $j$ to $k$ requires $\tau_{jk}^n$ units of the good $n$ itself, so that $1 + \tau_{jk}^n$ corresponds to

---

19 We adopt the convention that $\mathcal{N}(j)$ does not include $j$. 

7
the iceberg cost typically considered in the literature. This per-unit cost is specified as a function of the total quantity $Q_{jk}^n$ of good $n$ shipped along the link $jk$ and of the level of infrastructure $I_{jk}$:

$$
\tau_{jk}^n = \tau_{jk} (Q_{jk}^n, I_{jk}).
$$

(4)

For simplicity of exposition, this formulation assumes own-good congestion only. We discuss congestion across goods, which is the benchmark specification in the quantification, in section 3.7. The per-unit cost of shipping is increasing in the quantity of commodities shipped:

$$
\frac{\partial \tau_{jk} (Q, I)}{\partial Q} \geq 0.
$$

(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 may include higher travel times or road damage, as decreasing returns to scale in transportation activities due to land-intensive fixed factors such as warehousing or specialized physical and human capital. In short, the more is shipped, the higher the per-unit shipping cost.\footnote{For a review of the early literature on production-function estimates of returns to scale in the transport sector see Winston (1985). Newbery (1988) theoretically studies road damage externalities, whereby the the road damage caused by one vehicle increases the operating costs of subsequent vehicles. 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.}

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:

$$
\frac{\partial \tau_{jk} (Q, I)}{\partial I} \leq 0.
$$

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

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 $\tau_{jk} (Q, I) < \tau_{kj} (Q, I)$.\footnote{For a review of the early literature on production-function estimates of returns to scale in the transport sector see Winston (1985). Newbery (1988) theoretically studies road damage externalities, whereby the the road damage caused by one vehicle increases the operating costs of subsequent vehicles. 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.}

**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 $j = 1, \ldots, J$ and
commodities \( n = 1, \ldots, N \):

\[
D^n_j + \sum_{n'} X^n_{jn'} + \sum_{k \in \mathcal{N}(j)} (1 + \tau^n_{jk}) Q^n_{jk} \leq Y^n_j + \sum_{i \in \mathcal{N}(j)} Q^n_{ij}.
\] (6)

The left-hand side of this inequality is location \( j \)'s consumption \( D^n_j \) of good \( n \), intermediate-input use \( X^n_{jn'} \) by each sector \( n' \), and exports to neighbors \( Q^n_{jk} \). These flows are bounded by the local production \( Y^n_j \) and imports from neighbors \( Q^n_{ij} \). In standard minimum-cost flow problems this restriction is known as “conservation of flows constraint”.

We let \( P^n_j \) 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 \( P^n_j \) as the price of good \( n \) in location \( j \).

**Network Building Technology** We define the transport network as the distribution of infrastructure \( \{I_{jk}\}_{j \in J, k \in \mathcal{N}(j)} \). 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 3.7 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 \( \delta^I_{jk} I_{jk} \) units of \( K \). The network-building constraint therefore is:

\[
\sum_j \sum_{k \in \mathcal{N}(j)} \delta^I_{jk} I_{jk} \leq K.
\] (7)

While both the transport technology \( \tau_{jk}(Q, I) \) in (4) and the infrastructure building cost \( \delta^I_{jk} \) 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 \( \delta^I_{jk} \) 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, \( \delta^I_{jk} \) will not impact the allocation other than through infrastructure \( I_{jk} \).

We allow the network-design problem to take place when some infrastructure \( L_{jk} \) is already in place, and we also allow (but do not require) an upper bound \( T_{jk} \) 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:

\[ 0 \leq I_{jk} \leq I_{jk} \leq I_{jk} \leq \infty. \]

In our application to European countries, we will compute the optimal road network expansions starting from an observed road network \( I_{jk} \).

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 \( Q_{jk}^n \) is endogenous and that the marginal transport cost \( \tau_{jk}^n \) 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. In the former case, we let \( \omega_j \) 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

\[
W = \max_{c_j, h_j, \{I_{jk}\}_{k \in \mathcal{N}(j)}, \{D^n_j, \mathbf{V}^n_j, \mathbf{X}^n_j, \{Q^n_{jk}\}_{k \in \mathcal{N}(j)}\}} \sum_j \omega_j L_j U(c_j, h_j)
\]

subject to:

(i) availability of traded commodities,

\[ c_j L_j \leq D^n_j (D^n_1, \ldots, D^n_N) \text{ for all } j; \]

and availability of non-traded commodities,

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

(ii) the balanced-flows constraint,

\[ D^n_j + \sum_{n'} X_{nj}^{n'n'} + \sum_{k \in \mathcal{N}(j)} (1 + \tau_{jk} (Q^n_{jk}, I_{jk})) Q^n_{jk} \leq F^n_j (L^n_j, \mathbf{V}^n_j, \mathbf{X}^n_j) + \sum_{i \in \mathcal{N}(j)} Q^n_{ij} \text{ for all } j, n; \]

(iii) the network-building constraint,

\[ \sum_j \sum_{k \in \mathcal{N}(j)} \delta_{jk} I_{jk} \leq K, \]

subject to a pre-existing network,

\[ 0 \leq L_{jk} \leq I_{jk} \leq T_{jk} \leq \infty \text{ for all } j, k \in \mathcal{N}(j); \]
(iv) local labor-market clearing,
\[ \sum_n L^n_j \leq L_j \text{ for all } j; \]
and local factor market clearing for the remaining factors,
\[ \sum_n V^{mn}_j \leq V^m_j \text{ for all } j \text{ and } m; \]
and (v) non-negativity constraints on consumption, flows, and factor use,
\[ C^n_j, c_j, h_j \geq 0 \text{ for all } j \in N(j), n \]
\[ Q^n_{jk} \geq 0 \text{ for all } j, k \in N(j), n \]
\[ L^n_j, V^{mn}_j \geq 0 \text{ for all } j, m, n. \]

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

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

\[ W = \max_{u, c_j, h_j, (L_j)_{j \in N(j), n}} u \]
\[ \{D^n_j, L^n_j, V^n_j, X^n_j, (Q^n_{jk})_{k \in N(j)}\}_n \]

subject to restrictions (i)-(v) above; as well as:
(vi) free labor mobility,
\[ L_j u \leq L_j U(c_j, h_j) \text{ for all } j; \]
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 \( u = U(c_j, h_j) \) across all populated locations, and \( c_j = 0 \) otherwise.

We stop for a moment to discuss the generality achieved in the previous definitions. The case without labor mobility corresponds to international trade models. The production structure encompasses neoclassical trade models. When labor mobility is allowed, the model nests urban economics model in the tradition of Roback (1982). Since we have assumed neoclassical production functions, this formulation does not encompass new economic geography models such as Krugman (1991) and Helpman (1998) nor most recent quantitative extensions (Allen and Arkolakis, 2014; Redding, 2016), as these models are equivalent to frameworks with economies of scale. In Section

22The Armington model (Anderson and Van Wincoop, 2003) corresponds to \( N = J \) (as many sectors as regions) and \( F^n_j = 0 \) for \( n \neq j \), so that \( Y^n_j \) 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, \( Y^n_j = z^n_j L^n_j \). The specific-factors and Hecksher-Ohlin models are also special cases. As discussed below, the solution to the planner problem will generally be decentralized through transfers between countries except for a specific set of weights \( \omega_j \).

23The Rosen-Roback model assumes a single homogeneous tradable good.
(3.7) we discuss how to implement cases with increasing returns and provide some examples. The planner’s problem from Definition 1 can be expressed as nesting three problems:

\[
W = \max_{I_{jk}} \max_{Q^n_{jk}} \max_{\{c_j, h_j, D^n_j, L^n_j, V^n_j, X^n_j\}} \sum_j \omega_j L_j U(c_j, h_j)
\]

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 solution to each subproblem.

**Optimal Allocation** The innermost maximization problem over \((c_j, h_j, D^n_j, L^n_j, V^n_j, X^n_j)\) is a 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 \(Q^n_{jk}\) through the network combines an optimal transport problem—how to map production sources to destinations—and a least-cost route problem under congestion. This problem, under the assumption that consumption \(D^n_j\) and production \(Y^n_j\) 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 literatures is that these problems are well behaved and admit strong duality. In other words, while the least-cost route and the optimal coupling of sources to destinations may appear to be high-dimensional problems, the solution boils down to finding a “potential field”, meaning one Lagrange multiplier (or price) for each location-good pair, and then expressing the flows as a function of the difference between the multipliers across locations.

The optimal flow problem in our model inherits these properties as a special case of convex duality. To understand the solution, remember that \(P^n_j\) 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 (A.1) from the planner’s problem in Appendix (A.1) gives the following equilibrium price differential for commodity \(n\) between \(j\) and \(k\) ∈ \(N\) (j):

\[
\frac{P^n_k}{P^n_j} \leq 1 + \tau^n_{jk} + \frac{\partial \tau^n_{jk}}{\partial Q^n_{jk}} Q^n_{jk}, = \text{if } Q^n_{jk} > 0.
\]

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, \(\partial \tau^n_{jk}/\partial Q_{jk} = 0\), the price differential would be bounded by the trade cost.

This expression has a number of intuitive properties. 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. The inversion is possible, for instance, if the total transport cost \(Q^n_{jk}\tau^n_{jk}\) is convex in
the quantity shipped. In that case, the gross trade flow $Q_{jk}^n$ is increasing in the price differential. Condition (8) also implies that goods in each sector flow in only one direction, although a given link may have flows in opposite directions corresponding to different sectors. In addition, not all goods need to be shipped, and some links may be unused even if they have positive infrastructure. This may occur if the price gap is not large enough at zero trade to justify shipping.

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 an example with a single traded commodity, $N = 1$. In the example, the good is produced in the location at the origin (blue circle) and demanded in ten locations (orange circles). This example uses the functional form for $\tau_{jk}$ in (10), which implies that there are some shipments in every link although they become negligible in regions far away from the points of production and consumption. 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.

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 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 $j \in J$ and $k \in N(j)$ given the optimal transport and the neoclassical allocation. Letting $P_K$ be the multiplier of the network-building constraint (iii) (in other words, the shadow price of asphalt), the planner’s choice for $I_{jk}$ implies

$$P_K \delta_{jk} \geq \sum_{n} P_n^m Q_{njk} \left(-\frac{\partial \tau_{njk}}{\partial I_{jk}}\right),$$

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 $P_K$ of the network building constraint (7)) times the rate $\delta_{jk}$ 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, $-\partial \tau_{jk}/\partial I_{jk}$, applied to the value of the goods used as input in the transport technology.\(^{24}\)

Importantly, the network investment problem inherits the properties that make the optimal transport problem tractable. Substituting the solution for $Q_{njk}$ as function of the price differentials $P_n^m/P_j^n$ 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. Similar properties can be attained in a case with cross-goods congestion, as described in Appendix Section (3.6).

### 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 $k \in N(j)$; and (ii) if in addition $Q_{\tau_{jk}}(Q, I)$ is convex in both $Q$ and $I$ for all $j$ and $k \in N(j)$, then the full planner’s problem including the network design problem from Definition 1 (resp. Definition 2) is a convex

\(^{24}\)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.
(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 3.6, 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.

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 \( I \).

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

\[
\tau_{jk}(Q,I) = \delta_{jk}^\gamma \frac{Q^\beta}{I^\gamma} \text{ with } \beta \geq 0, \gamma \geq 0.
\]

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, \( \gamma \) captures the elasticity of the per-unit cost to infrastructure. The scalar \( \delta_{jk}^\gamma \) 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 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:

$$Q_{jk}^n = \left[ \frac{1}{1 + \beta \delta_{jk}} I_{jk}^\gamma \max \left\{ \frac{P_{jn}^n}{P_{jn}^n} - 1, 0 \right\} \right]^{\frac{1}{\beta}}. \quad (11)$$

This solution naturally implies that better infrastructure is associated with higher flows given prices and geographic trade frictions. It also shows that the total flows fall with congestion $\beta$ and increase with the average price differentials. Third, using the log-linear transport technology (10), whenever the planner chooses to build on top of existing infrastructure ($I_{jk} > L_{jk}$), the optimal infrastructure (9) arising from the optimal-network problem is

$$I_{jk}^* = \left[ \frac{\gamma \delta_{jk}^r}{P_K \delta_{jk}^r} \left( \sum_n P_{jn}^n (Q_{jk}^n)^{1+\beta} \right) \right]^{\frac{1}{1+\gamma}}. \quad (12)$$

Given the prices at origin, the optimal infrastructure increases with the gross flows. Given these flows, infrastructure also increases with prices at origin, as a higher sourcing cost implies a higher marginal saving from investing. Conditioning on these outcomes, infrastructure increases with $\delta_{jk}^r$, reflecting that optimal infrastructure investments offset geographic trade frictions, and decreases with $\delta_{jk}^I$, 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 ($I_{jk} > L_{jk}$). 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, $P_{jn}^n \neq P_{kn}^n$ for any $n$. Specifically, the optimal level of infrastructure is

$$I_{jk} = \max \{ I_{jk}^*, L_{jk} \}. \quad (13)$$

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

$$I_{jk}^* = \left[ \frac{\gamma}{P_K \delta_{jk}^I} \left( \frac{1}{1 + \beta \delta_{jk}^I} \sum_{n, P_{jn}^n > P_{jn}^n} P_{jn}^n \left( \frac{P_{jn}^n}{P_{jn}^n} - 1 \right)^{\frac{1+\beta}{\beta}} \right) \right]^{\frac{1}{1+\gamma}}. \quad (14)$$
where the multiplier $P_K$ 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 ($L_{jk} = 0$), then $I_{jk} = 0 \Leftrightarrow P^n_k = P^n_j$ 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 ($L_{jk} = 0$), 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 ($L_{jk} \geq 0$), 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 log-linear form (10), Proposition 2 implies that these possibilities arise if and only if there are no incentives to trade ($P^n_j = P^n_k$ 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.

**Other Convex Transport Technologies** We provide two additional families of transport technologies and the conditions that satisfy their convexity.

1. **Exponential:** $Q\tau_{jk}(Q, I) = \delta_{jk}^\tau max (\exp (\beta Q - \gamma I) - 1, 0)$, convex for all $\beta \geq 0, \gamma \geq 0$;

2. **CES:** $Q\tau_{jk}(Q, I) = \delta_{jk}^\tau max (Q^{\beta} - \zeta I^\gamma, 0)^{\frac{1}{\gamma}}$, $\zeta \geq 0$, convex for $0 \leq \gamma \leq 1$ and $\beta \geq \gamma$.

**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 3.6, 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 in other applications of optimal transport such as formation of blood vessels, irrigation or electric power supply systems (Banavar et al., 2000; Bernot et al., 2009).

**Proposition 3.** In the absence of a pre-existing network (i.e., $L_{jk} = 0$), 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. 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 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 \((\max_{c_j, h_j, D^n_j, L^n_j, V^n_j, X^n_j})\) and optimal transport \((\max_{q^n_{j,k}})\) 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 \(\tau^n_{j,k} q^n_{j,k}\) of delivering \(q^n_{j,k}\) units of good \(n\) from \(j\) to \(k \in \mathcal{N}(j)\) and takes the iceberg trade cost \(\tau^n_{j,k}\) 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 ad-valorem taxes \(t^n_{j,k}\) on companies shipping good \(n\) on leg \(j \rightarrow k\), denominated in units of the good at origin.

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 \(j_0, \ldots, j_\rho\) 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:

\[
\pi^n_{od} = \max_{r=(j_0, \ldots, j_\rho) \in \mathcal{R}_{od}} p^n_d - \sum_{k=0}^{\rho-1} P^n_{j_k, j_{k+1}} - \sum_{k=0}^{\rho-1} P^n_{j_k, j_{k+1}},
\]

where 

\[
\pi^n_{od} = \max_{r=(j_0, \ldots, j_\rho) \in \mathcal{R}_{od}} p^n_d - \sum_{k=0}^{\rho-1} P^n_{j_k, j_{k+1}} - \sum_{k=0}^{\rho-1} P^n_{j_k, j_{k+1}},
\]

18
where $p^n_j$ is the price of good $n$ in location $j$ in the market allocation. A shipper from $o$ to $d$ purchases each unit at price $p^o_0$ and obtains the price $p^n_d$. In addition, shippers must pay the transport costs $p^n_{jk} \tau^n_{jk+1}$ as well as the “toll” $p^n_{jk} t^n_{jk+1}$ on each of the route segments $j_k$ to $j_{k+1}$. In the absence of tolls, the shipping cost from $o$ to $d$ would equal the total iceberg cost, and the solution would correspond to a standard least-cost route optimization.

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 $\sum_{j=1}^J t_j = \Pi$, where $\Pi$ is an aggregate portfolio including ownership of fixed factors and government transfers. 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.

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

\[ t^n_{jk} = \varepsilon_{Q,jkn}^n \tau^n_{jk}, \]

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

(i) if labor is immobile, the competitive allocation coincides with the planner’s problem under specific planner’s weights $\omega_j$. 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., $t_j = \Pi/L$.

In either case, the price of good $n$ in location $j$, $p^n_j$, equals the multiplier on the balanced-flows constraint in the planner’s allocation, $P^n_j$.

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 enables us to calibrate the model using the planner’s solution to the optimal allocation and optimal transport subproblems given the network. In Section 3.7, we discuss how to calibrate the model assuming that the observed market allocation does not feature policies correcting the externality and is therefore inefficient.

3.5 Decentralization of Network Investments

Infrastructure investments can be decentralized using a very simple rule in some specific cases. We provide an example in the log-linear specification of the transport technology (10).

Consider a set of local monopolies constituted by an operating company for each link of the transport network in charge investing/maintaining infrastructure and charging a good-specific toll.
\( \theta^n_{jk} \) regulated by the government. We assume that each company can purchase “asphalt” from the government or social planner at the price \( p_K \), set equal to the Lagrange multiplier on the network-building constraint \( P_K \). The operating company takes prices and the toll amount as given and only chooses how much infrastructure \( I_{jk} \) to invest. The result can be stated as follows:

**Proposition 5.** If the transport technology is convex \((\beta \geq \gamma)\) and if the toll is set equal to the optimal Pigovian tax, \( \theta^n_{jk} = P^n_j \varepsilon_{Q,jkn}^\tau \gamma^n_{jk} \), then the decentralized infrastructure choice implements the optimal network investment along with specific transfers \( t_j \).

This result echoes the self-financing theorem of Mohring and Harwitz (1962) who showed that the revenue from the optimal congestion tax are sufficient to cover the capital costs of the roads when the transport cost function can be expressed a a function of the ratio \( Q/I \). Consistent with their finding, we show that in the specific case of the log-linear transport technology, setting the tolls equal to the level of the Pigovian tax suffice to align private and social incentives. Our result is, however, not restricted to the case with constant returns to scale \((\gamma = \beta)\). When \( \gamma < \beta \), tolls are more than enough to cover infrastructure investment costs and operating companies make positive profits, which must be rebated to consumer as part of the general transfer scheme \( t_j \) that implements the optimal allocation.

### 3.6 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).\(^\text{26}\)

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 \( \mathcal{Z} \) be the Lagrangian of the planner’s problem as a function of the variables controlled by the planner, \( x = (c_j, h_j, D^n_{j}, L^n_{j}, V^n_{j}, Q^n_{jk}, \ldots) \), and the multipliers \( \lambda = (P^n_j, \ldots) \) on the various constraints, the primal problem consists of solving the saddle-point problem

\[
\sup_x \inf_{\lambda \geq 0} \mathcal{Z} (x, \lambda).
\]

\(^\text{26}\)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 (Nesterov and Nemirovskii, 1994).
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.,

\[
\inf_{\lambda \geq 0} \sup_x \mathcal{L}(x, \lambda).
\]

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 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, \(x(\lambda)\). The remaining minimization problem, \(\inf_{\lambda \geq 0} \mathcal{L}(x(\lambda), \lambda)\), 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 \(c_j, h_j, D^n_j, V^n_j, L^n_{jk}, Q^n_{jk}\), 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.

### 3.7 Alternative Assumptions

**Congestion Across Goods** We have assumed that congestion only applies within good types. A natural assumption is that congestion takes place across goods. A simple way to model this feature while preserving the convexity of the problem is to assume that the per-unit cost \(\tau^n_{jk}\) is denominated in units of the bundle of traded goods aggregated through \(D_j(\cdot)\) rather than in units of the good itself. We assume that transporting each unit of good \(n\) from \(j\) to \(k \in \mathcal{N}(n)\) requires

\[
\tau^n_{jk} = m^n \tau_{jk}(Q_{jk}, I_{jk})
\]

units of the traded goods bundle, where where \(m^n\) measures product-specific characteristics, such as weight or volume, and \(Q_{jk} = \sum_{n=1}^{N} m^n q^n_{jk}\) is the total weight or volume transported from \(j\) to \(k\). Then, the total units of the traded goods bundle \(D_j\) used to transport goods from \(j\) to its
neighbors is

\[ T_j = \sum_{k \in N(j)} \tau_{jk} (Q_{jk}, l_{jk}) Q_{jk} \]

After properly adjusting the resource constraints in the definition of the planner’s problem, the convexity of the full planner’s problem is preserved under the same conditions stated in Proposition 1. Given the price differentials, heavy enough goods may not be shipped due to the marginal congestion cost that they exert on other goods. This is true even under the log-linear specification (10).

Appendices A.1 and A.2 present the definition and first-order conditions of the planner problem in a general case encompassing iceberg costs with own-good congestion, as well as this alternative formulation with congestion across goods. We adopt the case with congestion across goods as the benchmark in our application.

**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

\[ Y^j_n = F^n_j \left( L^j_n, V^j_n, X^j_n; L_j \right) \]

where the spillover from the total number of workers \( L_j \) on output \( Y^j_n \) is not internalized in the market allocation. Another common formulation is to assume externalities in the consumption of amenities as, for instance, in \( U (c_j, h_j; L_j) \), where \( L_j \) is taken as given. 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 undertake counterfactuals using a “fictitious” planner who ignores the dependence of \( Y^j_n \) on \( L_j \) or of \( \tau^n_{jk} \) on \( Q^n_{jk} \). For example, in the case of production spillovers, the fictitious planner problem is defined exactly as in Definition 2 under the assumption that the vector of aggregate population levels \( \mathbf{L} = \{ L_j \} \) in \( Y^j_n = F^n_j \left( L^j_n, V^j_n, X^j_n; L_j \right) \) is taken as given.\(^{27}\) As long as \( F^n_j (\cdot) \) is neoclassical given \( \mathbf{L} \), 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 \( \mathbf{L} \). It is straightforward to show that, if it exists, every distribution of population \( \mathbf{L} \) satisfying this fixed point problem corresponds to an inefficient market allocation and vice-versa.\(^{28}\) An important caveat, however, is that we can no longer establish the general convexity of the problem in the presence of externalities or efficiencies. Section (E) of the Supplementary Material nonetheless provides detailed explanation on how to implement such cases along with a

\(^{27}\)Similarly, the fictitious planner problem is defined exactly as in Definition 2 with \( U (c_j, h_j; \mathbf{L}_j) \) in the case of consumption externalities, taking \( \mathbf{L}_j \) as given, or with \( \tau_{jk} \left( Q^n_{jk}, l_{jk} \right) \) in the case of congestion externalities, taking the shipments \( \mathbf{Q} = \{ Q^n_{jk} \} \) as given.

\(^{28}\)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.

22
method to derive the infrastructure gradient and conduct local optimization.

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, \( U = \left( c^{h(1-\rho)} \right)^{1-\rho} / (1-\rho) \) 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, \( \delta_{jk}^I = \delta_{jk}^L = \text{Distance}_{jk} \).

Figure A.1 presents a network with 9 \( \times \) 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. Panels (c) and (d) display the spatial distribution of prices and consumption. As the network grows, relative prices and consumption converge, and spatial inequalities are reduced.

Randomly Located Cities and Non-Convex Cases We now explore more complex networks and non-convex cases. Figure A.3 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 commodity 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.

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” ($z^1_j = 1$ in all “countryside” locations) and ten “industrial” goods, each produced in one random city only ($z^n_j = 1$ in only one city $j$ and $z^n_j = 0$ 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.4 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.5 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 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.4 and A.5 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 Geographical Features

We now show how the framework can accommodate geographical features like mountains and rivers. 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.6 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 $\delta_{jk}$. In panel (a) we show the optimal network under the assumption that the cost of building infrastructure is proportional to the Euclidean distance:

$$\delta_{jk}^I = \delta_0 \text{Distance}_{jk}^\delta_1.$$  \hfill (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

$$\delta_{jk}^I = \delta_0 \text{Distance}_{jk}^\delta_1 \left(1 + |\Delta \text{Elevation}|_{jk}\right)^\delta_2.$$  \hfill (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:

$$\delta_{jk}^I = \delta_0 \text{Distance}_{jk}^\delta_1 \left(1 + |\Delta \text{Elevation}|_{jk}\right)^\delta_2 \delta_3^{\text{CrossingRiver}_{jk}} \delta_4^{\text{AlongRiver}_{jk}}.$$  \hfill (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 3.6. Now, a unique route links any two cities, water transport is not used, and a single bridge is built.
5 Road Network Expansion and Misallocation in Europe

We apply the framework for quantitative analysis of road networks in European countries. We start by describing the steps 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. We implement the calibrations and counterfactuals country by country. In a final exercise, we also implement the analysis simultaneously for a connected set of countries in continental Europe.

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

Sources We combine geocoded data on road networks, population, and income across European countries. The road network data is from EuroGlobalMap (EGM) by EuroGeographics.29 The dataset combines shapefiles on the road network from each European country’s mapping and cadastral agencies, and it includes all major highways and roads connecting populated areas.30 For example, the French road network is represented by 38668 segments of active roads connecting 159258 geographic points with a total length of about 130000 km.

We perform the calibration and counterfactual analysis separately for each of the 24 countries included in EGM 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.1 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.

An appealing feature of this dataset is that each segment of a road network has information about objective measures of road quality including type of road use (national, primary, secondary, or local), number of lanes, and whether it is paved or includes a median. National roads encompass each country’s highway system.31 On average across the countries in our data, national roads

29This product includes Intellectual Property from European National Mapping and Cadastral Authorities and is licensed on behalf of these by EuroGeographics. Original product is available for free at www.eurogeographics.org Terms of the licence available https://eurogeographics.org/services/open-data/topographic-data/
30We use GlobalMap data v8, corresponding to the year 2016, which is the earliest year we have access to. We only use road segments that are reported as “operational”.
31E.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.
represent 10% of the road network, feature twice as many lanes per kilometer as other types of roads, are always paved (while 94% of the non-national networks are), and are more likely to include a median (87% relative to 4% of non-national networks). Since the roads labeled as primary, secondary and tertiary have very similar characteristics along these dimensions, we bundle them into a single “non-national roads” category.

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. Both datasets correspond to the year 2005. 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). To implement the model we must take a stand on the geographic units corresponding to each node. To strike a balance between the high spatial resolution of the EGM and GPW datasets and the coarser resolution of the G-Econ dataset, in most countries we use 0.5 arc-degree cells (approximately 50 km by 50 km) as benchmark. We adopt squares as geographic units so that the boundaries of the units in the G-Econ dataset coincide with ours. We allocate population to each 0.5-degree cell by aggregating the smaller cells in GPW, and apportion income from the G-Econ cells according to the GPW-based population measure. In a few countries we use smaller or larger cells in order to allow for either a significant number of cells or avoid having a very large number.\footnote{We use the default 0.5 arc-degree cells in 17 countries (Austria, Belgium, Czech Republic, Denmark, Georgia, Hungary, Ireland, Latvia, Lithuania, Macedonia, Moldova, Netherlands, Northern Ireland, Portugal, Slovakia, Slovenia and Switzerland). Whenever assuming 0.5 arc-degree cells would lead to more than 200 cells, we use 1 arc-degree cells (Finland, France, Germany, Italy and Spain); and whenever doing so would lead to less than 20 cells, we use 0.25 arc-degree cells (Luxembourg and Cyprus).}

We denote the population and value added observed in each cell \(j\) of each country by \(I_{jk}^{\text{obs}}\) and \(GDP_{jk}^{\text{obs}}\). The resulting number of cells in each country is reported in Table A.1. In the 19 countries where the NUTS subdivision of geographic units exists, the number of cells is larger than the number of level-2 NUTS regions. In most countries it is also larger than the number of level-3 NUTS regions.\footnote{The NUTS (Nomenclature of Territorial Units for Statistics) is a standard developed by the European Union to divide the territory. The NUTS 2 correspond to “basic regions for the application of regional policies”, and NUTS 3 correspond to “small regions for specific diagnoses” (https://ec.europa.eu/eurostat/web/nuts/background). Excluding overseas territories Spain has 15 level-2 NUTS (autonomous communities) and 47 level-3 NUTS (provinces), whereas our partition has 61 cells. In France there are 21 level-2 NUTS (regions) and 94 level-3 NUTS (departments), whereas our partition has 74 cells.}

Underlying Graph Using these data we construct empirical counterparts to the underlying geography \( (J, E) \) corresponding to the locations and links in the graph of our model, as well as an observed measure of infrastructure \( I_{jk}^{\text{obs}} \) for each link.

To define the set of nodes \( J \) in each country, we use the GPW 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.\footnote{This leads to a very small adjustment: on average across countries, the average relocation across all cells within a country is 5.3 km.} We define the observed population and income of each node \( j \in J \) to be equal to the total income \( GDP_{jk}^{\text{obs}} \) and the
population $L_{j}^{\text{obs}}$ of the cell that contains it.

In turn, we define the set of edges $\mathcal{E}$ as all the links between nodes in contiguous cells. This step defines a set of up to 8 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, commonly referred to as the Moore’s neighborhood of location $j$.

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:

\[
\text{lanes}_{jk} = \sum_{s \in S} \omega_{jk}(s) \text{lanes}(s),
\]

\[
\text{nat}_{jk} = \sum_{s \in S} \omega_{jk}(s) \text{nat}(s),
\]

where $\text{lanes}(s)$ is the number of lanes on each segment $s$ on the actual road network $S$, $\text{nat}(s)$ 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:

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

where $\text{length}(s)$ is the length of segment $s$ and $\mathcal{P}(j,k)$ is the cheapest path from $j$ to $k$ on the actual road network. We follow these steps as long as the cheapest path does not stray from the cells containing $j$ and $k$. When that happens, we assume that no direct path from $j$ to $k$ exists in the actual road network, $\mathcal{P}(j,k) = \emptyset$, 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 $\text{lanes}_{jk}$ and $\text{nat}_{jk}$ 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 $j \in J$ and $k \in \mathcal{N}(j)$ by aggregating the observed attributes $\text{lanes}_{jk}$ and $\text{nat}_{jk}$ into a single index:

$$I_{jk}^{\text{obs}} = \text{lanes}_{jk} \times \kappa_{\text{non-nat}_{jk}}^{1-\text{nat}_{jk}}.$$  \hfill (20)

We assume $I_{jk}^{\text{obs}} = I_{kj}^{\text{obs}}$, implying that infrastructure applies equally in either direction. To compute the index we must assign a value to the cost $\kappa_{\text{non-nat}_{jk}}$. We note that, in the model, the resource cost of building a level of infrastructure $I_{jk}^{\text{obs}}$ is $\delta_{jk}^{I_{jk}^{\text{obs}}}$, with features of the terrain entering through $\delta_{jk}^{I_{jk}^{\text{obs}}}$. Therefore, the coefficient $\kappa_{\text{non-nat}_{jk}} < 1$ in (20) captures the costs of building infrastructure on non-national roads relative to national roads. We set $\kappa_{\text{nat}} = 1/5$, which corresponds to the cost of road construction and maintenance per kilometer on trunk roads relative to federal motorways in Germany in 2007, as reported by Doll et al. (2008).

In sum, we construct the observed infrastructure $I_{jk}^{\text{obs}}$ 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. To understand the units, we note that $\text{nat}_{jk} = 1$ implies $I_{jk}^{\text{obs}} = \text{lanes}_{jk}$. 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 $I_{jk}^{\text{obs}}$ lanes. Across connected nodes the discretized network, there is a correlation of 0.67 between $I_{jk}^{\text{obs}}$ and the speed on the quickest path according to GoogleMaps.

**Examples: France and Spain** Figures 2 and 3 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.

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, $\text{lanes}_{jk}$, and a color ranging from red to green depending on the likelihood of using a national road, $\text{nat}_{jk}$. 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.

### 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,
Figure 2: 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 0.5 arc-degree (approximately 50 km) cells. Panel (b) shows the nodes $J$ corresponding to the population centroids of each cell in Panel (a), reallocated to their closest point on the actual road network, and the edges $E$ 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.

\[ U = c^\alpha h^{1-\alpha}, \]
while the aggregator of traded goods (2) is CES:

\[ C^*_j = \left( \sum_{n=1}^{N} (C^n_j)^{\sigma-1} \right)^{\frac{1}{\sigma-1}}, \]

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:

\[ Y^*_j = z^*_j L^*_j. \]

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 closely linked to $\sigma$.

**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) with cross-good congestion that we describe in section 3.7. We must parametrize the congestion parameter $\beta$ and the returns to infrastructure parameter $\gamma$. Ideally, we would like to set these parameters to elasticities of total trade costs with respect to trade flows and infrastructure that account for as many of the forces discussed in Section 3.1. While such elasticities are not readily available, we can proceed by focusing on the relationship between shipping time and trade costs.\textsuperscript{37} Since the majority of inland

\textsuperscript{37} In different contests, several studies point to a relevant value of time in international goods shipping. E.g., Anderson and Van Wincoop (2004) calculate 9-percent tax equivalent of the average ocean shipping time cost in the US over the second half of the 20th century. Hummels and Schaur (2013) quantify that one additional day in transit is equivalent to 0.6 to 2.1 percent tariff based on information on modal choice between air and ocean. Djankov et al. (2010) estimate a modified gravity equation and conclude that each additional day of delay is equivalent to a country
shipments in the EU are done via road, it is likely that road shipments within European countries indeed include goods that are relatively time-sensitive.\textsuperscript{38} We assume that: i) trade cost is a linear function of shipping time; ii) the number of vehicles is a linear function of shipping flows; and iii) speed is a log-linear function of the number of vehicles and road lane kilometers. As shown in Appendix C, under these assumptions we can calibrate $\beta$ and $\gamma$ to match the empirical between speed, roads, and vehicles estimated by Couture et al. (2018) in U.S. data. Their estimates imply $\beta = 0.13$ and $\gamma = 0.10$, suggesting decreasing returns to scale.\textsuperscript{39} We use these parameters as benchmark, and we also implement the analysis in a case with increasing returns. Specifically, we consider a higher value of $\gamma$ such that the ratio between $\gamma$ and $\beta$ is the mirroring case, $\gamma/\beta = 0.13/0.10$ for $\beta = 0.13$.\textsuperscript{40}

Productivities, Endowments, and Geographic Frictions 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 $(I_{j}^{obs}, GDP_{j}^{obs})$ as outcomes of the planner’s solution for the optimal allocation and optimal flows problems discussed in Section 3.2 taking the observed network $I_{jk}^{obs}$ as given, and use this information to back out the fundamentals $(z^n_j, H_j)$. In the case with fixed labor, we interpret $GDP_{j}^{obs}$ as the outcome of the planner solution and use this information to back out the productivities $z^n_j$, normalizing $H_j/I_{j}^{obs} = 1$ and setting the planner’s weights $\omega_j = 1$ everywhere.

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 $N + 1$ different sectors: $N$ differentiated goods, and one homogeneous good. As a benchmark, we assume that each of the differentiated goods is produced in each of the $N$ cells with the largest observed population, and that the agricultural good is produced by all the remaining cells. We assume 10 different sectors and explore the robustness of this assumption to alternative values of $N$. As a robustness check we also implement an alternative calibration where the differentiated products are allocated to the $N$ largest level-2 NUTS regions within each country. 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, when allowing for labor mobility) such

\textsuperscript{38}According to Eurostat, 75% of the tonne-kilometer shipped via inland transport modes (rail, waterways or road) within the EU-28 are done by road. The share is 50% when considering all transport modes. See https://ec.europa.eu/eurostat/statistics-explained/index.php/Freight_transport_statistics_modal_split%20.

\textsuperscript{39}Couture et al. (2018) find decreasing returns to scale ($\alpha < \theta$ in their notation, corresponding to $\gamma < \beta$ in ours) across all their specification (Tables 5 and 6 of their paper) using a variety of OLS and IV approaches, although the difference is typically small. They refer to this result as suggesting “modest decreasing returns to scale”. Their preferred estimate, used for our calibration, column 6 of table 5 of their paper.

\textsuperscript{40}An alternative approach to calibrate $\alpha$ and $\beta$ could have been to draw from production function estimates of the transport industry. Existing evidence on increasing returns is mixed. In an early review of the transport economics literature, Winston (1985) noted that production-function estimates of returns to scale (defined, in our notation, as the elasticity of $\tau$ to a simultaneous increase in $Q$ and $I$) show mixed results, with different studies showing evidence of decreasing, constant, and increasing returns to scale depending on the mode of transport and details of the estimation approach and functional form.
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, when allowing for labor mobility). 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.\footnote{More specifically, in the solution of the planner’s problem each location’s value added is $P_n^j \cdot z_j L_j + P_H^j H_j$, where $n(j)$ denotes the good produced by location $j$, $P_n^j$ 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 $P_H^j$ 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).}

We must also determine the values of the the geographic trade frictions $\delta_{jk}^\tau$ entering in the transport technology (10). We assume that all goods have the same weight ($m_n = 1$) and that frictions depend on distance, $\delta_{jk}^\tau = \delta_0^\tau \text{dist}_{jk}$. To calibrate $\delta_0^\tau$, we target the level of intra-regional trade in Spain, one of the few countries in our dataset for which regional-level trade data is available. Specifically, we estimate $\delta_0^\tau$ jointly with the other fundamentals so that the model matches the 44% share of intra-regional trade in intra-national trade among tradeables sectors across the 15 continental level-2 NUTS regions of Spain from 2001-2005, according to C-Interreg Dataset (Llano et al., 2010). To generate model-based regional trade flows, we aggregate the bilateral node-to-node trade flows in differentiated goods to the bilateral region-to-region level.\footnote{The model does not make a prediction for bilateral region level flows of homogeneous goods. Since this good is produced in every region, most of its production is not traded across region boundaries. In the calibration, the intra-regional trade share of this sector is 93%.} We then use this value of $\delta_0^\tau$ in the remaining countries when calibrating fundamentals.

Figure A.7 in Appendix C shows the results of the calibration for the convex case of the parameters ($\beta = 0.13, \gamma = 0.10$). Similar relationships hold for the non-convex case ($\beta = 0.13, \gamma = 0.169$). Panels (a) and (b) show the model-implied population and income shares of each location against the data, over all locations in the 24 countries. Except for a few locations, both population and income shares are matched with high precision. The internal trade share for Spain is also precisely matched.\footnote{Across the 24 countries in our data the implied average internal trade share is 35%, with a standard deviation of 12%. The intra-regional trade flow in total flows is defined as $\Sigma_i \Sigma_j \Sigma_{i,j} X_{i,j}$, where $X_{i,j}$ is the total flow of differentiated products from region $i$ to $j$. We calibrate $\delta_0^\tau = 2.05 \times 10^{-3}$ and $\delta_0^\tau = 2.00 \times 10^{-3}$ in the benchmark convex case with fixed and mobile labor, respectively. In the non-convex case we find $\delta_0^\tau = 2.07 \times 10^{-3}$ and $\delta_0^\tau = 2.15 \times 10^{-3}$.}

Panels (c) and (d) show the calibrated fundamentals (productivity and endowment of non-traded good) in the vertical axes against income and population shares in the data, respectively, for the case with labor mobility. Both fundamentals are strongly correlated with observables. A similar positive relationship between productivity and income share holds in the calibration of the model with fixed labor.

**Cost of Building Infrastructure** To implement the optimal transport network in counterfactual scenarios, we must parametrize the cost of infrastructure along each edge, $I_{jk}^\delta$. We follow two approaches. In the first approach, we interpret the observed infrastructure $I_{jk}^{obs}$ as the outcome of...
the planner’s problem under the assumption that it is equally costly to build in either direction and that \( \delta_{jk}^I = \delta_{kj}^I \). In this case the observed network, \( I_{jk}^{obs} \), is consistent with the planner’s first-order condition for \( I_{jk} \) in (12) under the assumption that \( I_{jk} = 0 \). 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, \( \delta_{jk}^{I,FOC} \).

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 thousands of World-Bank infrastructure investment projects across the world, and then relate these costs to a host of geographic and non-geographic frictions.\(^{44}\) We interpret an improvement to the connection between any pair of nodes in our counterfactuals as a different infrastructure project in that dataset. We assume that \( \delta_{jk}^I \) 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, \( \delta_{jk}^{I,GEO} \). In our notation, their estimates imply:

\[
\ln \left( \frac{\delta_{jk}^{I,GEO}}{\delta_{jk}^I} \right) = \ln (\delta_0^I) - 0.11 \times (\text{dist}_{jk} > 50\text{ km}) + 0.12 \times \ln (\text{rugged}_{jk}),
\]

where \( \text{dist}_{jk} \) is the distance between \( j \) and \( k \) and \( \text{rugged}_{jk} \) is the average over the ruggedness in locations \( j \) and \( k \).\(^{45}\) 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 \( \delta_0^I \) may be country-specific.

These steps give two alternative measurements of \( \delta_{jk}^I \) up to scale in each country. Hence, we set \( K = 1 \) in every country and re-scale each of the two measures of \( \delta_{jk}^I \) to satisfy the network-building constraint with equality in each country.

### 5.3 Model-Implied Trade Flows

Our calibration matches intra-national trade share in total trade. We now check the model predictions for bilateral trade flows against the data. We use bilateral trade data across Spanish

---

\(^{44}\)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 (22) 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.

\(^{45}\)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 \( J_{\text{etopo}}(j) \) be the set of cells in ETOPO1 contained in each cell \( j \in J \) of our discretization and \( N_{\text{etopo}}(i) \) be the 8 neighboring cells to each cell in ETOPO1, this index is defined as: \( \text{rugged}_j = \left( \sum_{i \in J_{\text{etopo}}(j)} \sum_{k \in N_{\text{etopo}}(i)} (\text{elev}_i - \text{elev}_k)^2 \right)^{1/2} \); i.e., it is the standard deviation of the difference in elevation across neighboring cells. Then, we define \( \text{rugged}_{jk} \) in (22) as \( \text{rugged}_{jk} = \frac{1}{2} (\text{rugged}_j + \text{rugged}_k) \).
regions defined at the level-2 NUTS subdivision from Spain’s C-Interreg Dataset. Excluding islands, this gives 15 Spanish regions. Panel (a) of Figure 4 shows observed and and model-based trade flows in differentiated products for the calibration where each differentiated good is allocated to a different region. Own-region trade flows are shown as red crosses in the figure. The model-based predictions for bilateral flows are broadly comparable to the data, with a correlation of 0.77.

Another way to assess the implied trade flows of the model is to look at the gravity implications. The standard gravity model posits a log-linear relationship between bilateral trade shares and trade costs. The gravity model typically gives a good fit of the international data (Head and Mayer, 2014), and it is also often applied within countries (Allen and Arkolakis, 2014). A common approach to bring the gravity model to the data is to parametrize bilateral trade costs as a function of geographic frictions. In our parametrization, bilateral trade costs among locations depend on the distance through the network, but also on the equilibrium levels of congestion, the observed levels of infrastructure, and the calibrated values of $\beta$ and $\gamma$.\footnote{\textsuperscript{46}E.g., Ramondo et al. (2016) impose that the bilateral trade cost between two locations $j$ and $k$ take the form $T_{jk} = \delta_0 \text{dist}_{jk}$. In our model, the bilateral cost $T_{jk}$ is endogenous, as it depends on the optimal path and the equilibrium levels of congestion and infrastructure. Our parametrization implies that, on any optimal path given an optimal path $\{j_0, ... , j_p\}$ from $j_0 = j$ to $j_p = k$, $T_{jk} = 1 + \delta_0 \sum_{l=0}^{p-1} \frac{P_j^D \text{dist}_{j_l,j_{l+1}}}{P_j^D \left(\text{obs}_{j_l,j_{l+1}}\right)^\gamma} Q_{j_l,j_{l+1}}^{\beta}$.}

A linear regression yields an elasticity of $-0.91$ in the model and $-1.37$ in the data.

Overall the figures suggest that the model makes reasonable predictions for the distribution of trade flows and for the relationship between bilateral trade flows and distance, despite the fact that the elasticity of trade with respect to distance is not targeted as part of the calibration.\footnote{\textsuperscript{47}I.e., we run, in both model-generated and observed data, the regression $\ln \left(\lambda_{jk}^{\text{NUTS}}\right) = \delta \ln \left(\text{dist}_{jk}\right) + \psi_j + \varepsilon_{jk}$, where $\lambda_{jk}$ is the import share and $\text{dist}_{jk}$ is the bilateral distance between the level-2 NUTS divisions. The figure shows both the import share and distance as residuals from exporter fixed effects. Therefore, the slope of the linear fit reported in the figure is the estimate of $\delta$ in this regression. Distance is computed between geographic centroids of regions. We exclude zero flows as well as flows to the own level-2 NUTS.} We finally examine the level of congestion taxes that the model’s trade flows imply. Due to the log-linear specification of our transport technology, the taxes are always a fraction $\beta = 13\%$ of the transport costs on a given link. Numerically, we obtain that the mean ad valorem tax is about $\lambda_{jk} = \delta_0 \text{dist}_{jk}$. In our model, the bilateral cost $T_{jk}$ is endogenous, as it depends on the optimal path and the equilibrium levels of congestion and infrastructure. Our parametrization implies that, on any optimal path given an optimal path $\{j_0, ... , j_p\}$ from $j_0 = j$ to $j_p = k$, $T_{jk} = 1 + \delta_0 \sum_{l=0}^{p-1} \frac{P_j^D \text{dist}_{j_l,j_{l+1}}}{P_j^D \left(\text{obs}_{j_l,j_{l+1}}\right)^\gamma} Q_{j_l,j_{l+1}}^{\beta}$, where $P_j^D$ is the observed measure of infrastructure and $P_j^D$ is the price at $j$ of the bundle of traded goods used as an input in the transport sector in terms of the shipped good itself. We let the gravity properties of our model be dictated by the parameters $\beta$ and $\gamma$ calibrated to match external sources.\footnote{\textsuperscript{48}The results are very similar in the calibrations that assign a different number of differentiated products to the largest locations in the country when comparing the model-based non-zero bilateral trade flows and the data. Figure A.8 shows that in that case the correlations are between 0.75 and 0.85 depending on how many goods are included in the market allocation. They are also similar in calibrations that assign one good to each region but assumes away the homogeneous product, so that every location within a region produces the differentiated good of that region. Their elasticities between trade and distance for France as it does for Spain, suggesting that the key gravity properties are dictated by the common elasticities rather than by the distribution of fundamentals.}
0.5% in the case with labor mobility (0.59% without) and that the total taxes paid represent 0.31% of GDP with labor mobility (0.33% without).

5.4 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, $L_{jk} = I_{jk}^{obs}$. Second, we measure the 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, $L_{jk} = 0$.\(^{49}\)

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. The first counterfactual is 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 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 $\delta_{jk}^{l,FOC}$ and the geographic measure $\delta_{jk}^{l,GEO}$. The optimal reallocation is only meaningful

\(^{49}\)In every case, we set the upper bound on infrastructure, $T_{jk}$, to be 50% above the largest level of infrastructure observed in each country.
Figure 5: Optimal Network Reallocation and Expansion: Spain and France

(a) Optimal Network Expansion with Labor Mobility, France
(b) Optimal Network Expansion with Labor Mobility, Spain
(c) Optimal Network Reallocation with Labor Mobility, France
(d) Optimal Network Reallocation with Labor Mobility, Spain
(e) Optimal Network Reallocation with Fixed Labor France
(f) Optimal Network Reallocation with Fixed Labor, Spain

Notes: All counterfactuals use the geographic measure of building costs, $\delta^{I,GEO}$ and the benchmark parametrization of $\beta$ and $\gamma$. The width and brightness of each link is proportional to the difference between the optimal counterfactual network and the observed network, $I^*_jk - I^*_{jk}$, for each link $jk \in \mathcal{E}$ shown in panel (b) of Figures 2 and 3. The color scale is the same as in Figure 2. Red links represent negative investment. With labor mobility, brighter green (red) nodes represent larger population increase (decrease).

under the geographic measure, since, by construction, the observed network is optimal and cannot be improved under $\delta^{I,FOC}_{jk}$. We implement each of these three counterfactuals for each of the two values of $\gamma$, assuming both fixed and mobile labor, separately for each of the 24 countries. We re-calibrate the model for each value of $\gamma$, 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 large countries in our data, Spain and France. Figure 5 depicts the pattern of investment and population change under the geographic measure of building costs $\delta^{1,\text{GEO}}$. Panels (a) and (b) show the optimal expansion and panels (c) and (d) show the optimal reallocation when under labor mobility. Panels (e) and (f) reproduce the optimal reallocation assuming that labor is not mobile. All the figures correspond to assuming the benchmark calibration of $\beta$ and $\gamma$. The thickness of each link increases with the absolute value of the investment, defined as the difference between the counterfactual and the observed infrastructure, \( I_{jk}^* - I_{jk}^{\text{obs}} \). In the reallocation counterfactual, links with negative investment, \( I_{jk}^* - I_{jk}^{\text{obs}} < 0 \), are shown in red, while all other links are shown in green. In turn, with labor mobility, 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 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 (c) and (d), typically feature no expansion in panels (a) and (b). The comparison between panels (c)-(d) and panels (e)-(f) reveals that allowing labor mobility does not fundamentally affect the optimal infrastructure investments, as there is a high correlation between the patterns of optimal investment in both cases.

Despite the different investment patterns, in the cases of optimal expansion and optimal reallocation population is reallocated to the same set of regions within each country. 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, $c_j$.\textsuperscript{50} 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.

We inspect, across the 24 countries, how a few typically observable regional outcomes map to infrastructure investment. Panel (a) of Table 1 reports results from regressions of infrastructure growth on each location’s initial population and tradeable income per capita. We report here results corresponding to the case with labor mobility and the benchmark parametrization of $\gamma$ and $\beta$, but the qualitative features we discuss are similar under the alternative value of $\gamma$, or assuming fixed labor and using the change in consumption per capita instead of population as dependent variable. Optimal road investments are directed to locations with initially lower levels of infrastructure, reflecting decreasing returns to infrastructure at the link level. Under the geographic measure of building costs in columns (1) and (2), the investments are also more intensely directed to locations with initially higher levels of population and income per worker. Since the model implies a complex mapping from the fundamentals to the investments, these observable outcomes guide only a fraction of the optimal investment decisions ($R^2$ in the order of 20-30%).

\textsuperscript{50}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

(a) Dependent Variable: Infrastructure Growth

<table>
<thead>
<tr>
<th>Counterfactual</th>
<th>Reallocation ($\delta = \delta^{L,GEO}$)</th>
<th>Expansion ($\delta = \delta^{L,GEO}$)</th>
<th>Expansion ($\delta = \delta^{L,FOC}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Population</td>
<td>0.358***</td>
<td>0.136***</td>
<td>0.003</td>
</tr>
<tr>
<td>Tradeable Income per Capita</td>
<td>0.335***</td>
<td>0.164***</td>
<td>-0.003</td>
</tr>
<tr>
<td>Infrastructure</td>
<td>-0.439***</td>
<td>-0.242***</td>
<td>-0.082***</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.32</td>
<td>0.27</td>
<td>0.21</td>
</tr>
</tbody>
</table>

(b) Dependent Variable: Population Growth

<table>
<thead>
<tr>
<th>Counterfactual</th>
<th>Reallocation ($\delta = \delta^{L,GEO}$)</th>
<th>Expansion ($\delta = \delta^{L,GEO}$)</th>
<th>Expansion ($\delta = \delta^{L,FOC}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Population</td>
<td>0.000</td>
<td>-0.001**</td>
<td>-0.000</td>
</tr>
<tr>
<td>Tradeable Income per Capita</td>
<td>0.026***</td>
<td>0.020***</td>
<td>0.004***</td>
</tr>
<tr>
<td>Consumption per Capita</td>
<td>-0.074***</td>
<td>-0.062***</td>
<td>-0.011***</td>
</tr>
<tr>
<td>Infrastructure Growth</td>
<td>-0.001**</td>
<td>-0.001**</td>
<td>-0.000</td>
</tr>
<tr>
<td>Differentiated Producer</td>
<td>0.008**</td>
<td>0.001***</td>
<td>0.001***</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.47</td>
<td>0.50</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Each column corresponds to a different regression pooling all locations across the 24 countries in the benchmark parametrization of $\gamma$ and $\beta$, assuming mobile labor and $N=10$. All regressions include country fixed effects. Standard errors are clustered at the country level. ***=1% significance, **=5%, *=10%. Dependent variables: population growth is defined as $\Delta \ln L_j$, where $\Delta x = x' - x$ denotes the difference between variable $x$ in the counterfactual ($x'$) and in the calibrated allocation ($x$). Investment growth is defined as the difference over the average, $\Delta I_j / (\frac{1}{2} (I_j + I'_j))$, where total infrastructure at the node level defined as $I_j = \sum_{k \in N(j)} I_{jk}$. Independent variables 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 calibrated model. Differentiated producer is a dummy for whether the location is a producer of differentiated goods in the calibration.

In Panel (b), the dependent variable is population growth. To understand the patterns of optimal reallocation of population, in addition to the variables from the previous regression we also include infrastructure growth, consumption per capita, and a dummy for whether a location is a differentiated producer. When included in the regressions of Panel (a), consumption per capita and the index for being a differentiated producer are not significant and have a very small effect on the $R^2$. Because consumption and income are highly correlated, in that case income per capita is also not significant.

51

39
other locations in potentially complex ways. If we randomly improve individual links, population growth does appear correlated with infrastructure growth.\footnote{Using the benchmark calibration for Spain, we ran 100 counterfactuals where we improved infrastructure in one randomly chosen link by a uniform draw between 0\% and 10\%. In 73\% of cases, the nodes connected by the improved link grew. A regression of population growth on infrastructure growth across these 100 observations gives an elasticity of 0.2\% with standard error of 0.04\%.} Consumption of traded goods per capita is a strong determinant, with a negative elasticity of population growth with respect to initial consumption in the order of 1\%-7\%. 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 1\% across the three counterfactuals. Hence, the impact of initial income on population 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.

**Aggregate Impact Across Countries** We now show the aggregate welfare effects. Table 2 shows the average welfare gain for each counterfactual across the 24 countries in our data. Tables A.2 and A.3 in Appendix C show the results for each country with fixed and mobile labor, respectively. In the benchmark parametrization of $\gamma$ and $\beta$, using the geographic measure of building costs we find average welfare gains across countries of around 1.5\% – 1.7\%. The effects are much smaller under the FOC-based measure, because in that case the optimal expansion does not address a suboptimal placement of existing roads. The average gains are increasing in the returns to scale $\gamma$, with the average welfare gains increasing to between 1.9\% and 2.8\% under geographic measure of building costs. These effects vary considerably across countries, ranging from around 0.5\% to 5\%. There is no clear pattern between misallocation and country size or income.

This distribution of welfare gains across countries is in general stable regardless of the parametrization of $\gamma$, the assumption on labor mobility, the parametrization of the building costs $\delta^I$, or the type of counterfactual. For example, across the parametrizations of $\gamma$ 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 0.76 and 0.88. 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 these cases.
Table 2: Average Welfare Gains Across Countries

<table>
<thead>
<tr>
<th>Returns to Scale:</th>
<th>Benchmark</th>
<th>Non-Convex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labor:</td>
<td>Fixed</td>
<td>Mobile</td>
</tr>
<tr>
<td>Optimal Reallocation</td>
<td>$\delta = \delta^{I,GEO}$</td>
<td>1.6%</td>
</tr>
<tr>
<td>Optimal Expansion</td>
<td>$\delta = \delta^{I,GEO}$</td>
<td>1.7%</td>
</tr>
<tr>
<td></td>
<td>$\delta = \delta^{I,FOC}$</td>
<td>0.3%</td>
</tr>
</tbody>
</table>

Each element of the table shows the average welfare gain in the corresponding counterfactual across the 24 countries.

Alternative Assumptions The analysis was implemented assuming $N = 10$ sectors. We also implement the calibration and counterfactuals, under both mobile and fixed labor, assuming either that $N = 5$ or $N = 15$, for the benchmark calibration of $\beta$ and $\gamma$. 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.4 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 these alternative cases, the patterns described above remain unchanged, and the magnitude of most of the coefficients does not exhibit large variation.

Similarly, Table A.5 reproduces Table 2 for the benchmark and for $N = 15$. The aggregate gains change little with the number of sectors. The correlation between the aggregate welfare effects across countries under $N = 15$ and under $N = 10$ is above 0.9 for each possible type of counterfactual and assumptions on labor mobility and value of $\gamma$. The table also reports average welfare effects under an alternative calibration where each of the largest $N$ regions in each country, defined as level-2 NUTS political subdivisions, is assigned a differentiated product. In the misallocation and expansion counterfactuals under the geographic measure of costs, the correlation in welfare gains between the benchmark case and this alternative allocation is between 0.88 and 0.97 across assumptions of labor mobility, convexity of the parameter space, and number of goods; and between 0.65 and 0.79 in the expansion counterfactual under the FOC measure of costs.

5.5 Application to Multiple Countries within Europe

Our previous applications considered each country in isolation. In doing so, we have ignored a role for international trade. We now implement the analysis for a region of western Europe including 11 countries in our data. Figure A.10 in the appendix shows the baseline map and the discretized

\footnote{Within each region, the differentiated product is allocated to the largest location by population in the calibration. We implement this version for the 15 countries in our data where the NUTS classification is available and there are at least 4 level-2 NUTS regions.}

\footnote{For computational reason, we restrict ourselves to a subsample of connected countries in western continental Europe which includes: Austria, Belgium, Denmark, France, Germany, Italy, Luxembourg, Netherlands, Portugal, Spain and Switzerland. See Figure (A.10) in the Appendix for the map. We use 1 degree by 1 degree cells, resulting in 261 cells.}
network for this connected set of countries. We assume that each country produces a country-specific differentiated product, in addition to a homogeneous good. We use the same parameters as in the benchmark parametrization, and re-calibrate the fundamentals assuming that the 5 largest locations in terms of observed population within each country produce the differentiated product, while the remaining locations produce the homogeneous product. In addition to the cases with perfect labor mobility and with immobile labor, now we also implement a case with partial mobility where labor is mobile within countries but not across countries, recalibrating the fundamentals each time.

Figure 6 shows the optimal network expansion under different assumptions of labor mobility. These counterfactuals highlight the set of areas where European investments would be more profitable. The investments are concentrated in Benelux countries, France, Germany, and Northern Italy. Within Spain, the optimal expansion looks quite different from what we found in panel (b) of Figure 5. Specifically, we now observe differences in the pattern of the optimal network borne out of the planner’s incentives to deal with international trade. The planner chooses to prioritize two corridors connecting Spain to the North of Portugal and the center of Spain to France, whereas before it chose a higher density of investment in the South of the country. Within France, the pattern of investments radiating from Paris is similar to the case in (a) of Figure 5, but now we see optimal investments in the connection with Spain as well as investments in the Northeast in order to connect with neighboring countries.

We also find that the optimal network expansion is very similar in the three cases, suggesting that assumptions on labor mobility do not impact the best investment plan. The welfare gains are very close to 2% in the three cases. When labor is mobile across Europe, the optimal network investment reallocates workers to southern Spain and Portugal, much like in the country-by-country analysis we found reallocation to areas with relatively lower income per worker. As in the previous cases, these changes in population do not correlate very strongly with the investments.

To conclude, we can ask whether these patterns are approximately comparable with the Trans-European Transport Network (TEN-T), a European Commission policy that supports the development of Europe-wide transport networks for different transport modes through funding facilities and a common regulatory framework. The network is multimodal, including roads, air and inland waterways. The TEN-T defines a core network of “strategic importance” for future investments and upgrades based on different criteria such as eliminating bottlenecks or following suggestions from member states. Figure A.11 in Appendix C shows these corridors for the area of Europe covered by our counterfactual. Broadly speaking, our planning problem identifies some priorities for investment which appear to be similar to what real-world planners have decided, such as the high density of investment in Benelux countries and Germany; the international corridor from Paris to the West of France, North of Spain and Portugal; and the connection between the Germany and Denmark. However, we also see some differences, as the solution to our planning problem does not

---

identify the need to invest in roads connecting the South of France, Spain and Portugal.
Figure 6: Optimal Network Expansion: Europe

(a) Full Mobility

(b) Labor Mobility within Countries

(c) No Mobility

Notes: All counterfactuals use the geographic measure of building costs, $\delta^{I,GEO}$ and the benchmark parametrization of $\beta$ and $\gamma$. The width and brightness of each link is proportional to the difference between the optimal counterfactual network and the observed network, $I^*_j k - I^{obs}_{jk}$, for each link $jk \in \mathcal{E}$ shown in panel (b) of Figure A.10. The color scale is the same as in Figure 2. Red links represent negative investment. With labor mobility, brighter green (red) nodes represent larger population increase (decrease).
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 across European countries. Using the calibrated model, we compute the gains from road expansion and losses from misallocation. Across countries, we find real consumption losses in the order of 2% associated with misallocation of roads.

We expect the framework to serve as basis for future work. It could be used to study political-economy issues associated with infrastructure, such as spatial competition among planning authorities non-cooperatively deciding investments. We have refrained from explaining the sources of misallocation, but it would be interesting to know how regional characteristics such as institutional quality relate to misallocation. Our application was limited to European countries, but low-income economies are likely to benefit more from infrastructure investment and are perhaps more prone to inefficient investments due to their institutional environments.

The persistence of transport networks also raises interesting issues. The model could be extended to study inefficient network lock-in due to existing investments corresponding to dated economic fundamentals. Relatedly, we have abstracted from decision-making under uncertainty, but it would be interesting to study a planner who decides in anticipation of changing conditions about technology or fundamentals. In the spirit of Barjamovic et al. (2017), who use the prediction of gravity models to infer the location of cities, one could also envision applications where the model is used to infer the location of roads in historical data.

Finally, 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


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


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. We present the problem adopting a formulation of the transport technology that nests the benchmark approach with own-good congestion in which the transport cost is denominated in units of the good being shipped, as well as an alternative approach with congestion across goods in which the transport cost is denominated in units of the bundle of traded goods. In the formulations below, the parameter $\chi \in \{0, 1\}$ takes a value of 0 in the former benchmark case with own-good congestion, and the value of 1 in the latter with cross-goods congestion. The case $\chi = 0$ corresponds to the equations presented in the body of the paper.

Immobile Labor

The Lagrangian of the problem in Definition 1 is

$\mathcal{L} = \sum_j \omega_j L_j U \left( c_j, h_j \right) - \sum_j P_j^D \left[ c_j L_j + \chi \sum_{k \in N(j)} \tau_{jk} (Q_{jk}, I_{jk}) Q_{jk} - D_j \left( D_j^1, \ldots, D_j^N \right) \right] - \sum_j P_j^H (h_j L_j - H_j)$

$- \sum_j \sum_n P_j^n \left[ D_j^n + \sum_{k \in N(j)} \left( Q_{jk}^n + (1 - \chi) \tau_{jk} (Q_{jk}^n, I_{jk}) Q_{jk}^n \right) - F_j^n \left( L_j^n, V_j^n, X_j^n \right) \right] - \sum_i \sum_{i \in N(j)} Q_{ij}^n$

$- \sum_j W_j \left[ \sum_n L_j^n - L_j \right] - \sum_j \sum_l R_{jn}^m \left[ \sum_n V_{jn}^m - V_{jn}^m \right]$

$- P_K \left( \sum_j \sum_{k \in N(j)} \delta_{jk} I_{jk} - K \right) + \sum_{j,k} \zeta_{jk}^L (I_{jk} - L_{jk}) + \sum_{j,k} \zeta_{jk}^T (T_{jk} - I_{jk})$

$+ \sum_{j,k,n} \zeta_{jkn}^Q Q_{jkn} + \sum_{j,n} \zeta_{jn}^L L_j^n + \sum_{j,n,l} \zeta_{jn}^V V_j^n + \sum_{j,n} \zeta_{jn}^X X_j^n + \sum_{j,n} \zeta_{jn}^D D_j^n + \sum_j \zeta_j^c c_j + \sum_j \zeta_j^h h_j$

where $Q_{jk} = \sum_{n=1}^N m^n Q_{jk}^n$ and where $P_j^D$, $P_j^H$, $P_j^N$, $W_j$, $R_j^1$, $P_K$, $\zeta_{jk}^L$, $\zeta_{jk}^Q$, $\zeta_{jn}^L$, $\zeta_{jn}^X$, $\zeta_{jn}^V$, $\zeta_{jn}^D$, $\zeta_j^c$, $\zeta_j^h$ are the multipliers of all constraints implied by (i)-(v) in Definition 1. The first-order conditions with respect to consumption and production are:

$[c_j] \quad \omega_j L_j U \left( c_j, h_j \right) + \zeta_j^c = P_j^D L_j$

$[h_j] \quad \omega_j L_j U \left( c_j, h_j \right) + \zeta_j^h = P_j^H L_j$

$[D_j^n] \quad P_j^D \frac{\partial D_j^n}{\partial D_j^n} + \zeta_j^n = P_j^n$

$[L_j^n] \quad P_j^n \frac{\partial Y_j^n}{\partial L_j^n} + \zeta_{jn}^L = W_j$

$[V_j^n] \quad P_j^n \frac{\partial Y_j^n}{\partial V_j^n} + \zeta_{jn}^V = R_j^1$

$[X_j^n] \quad P_j^n \frac{\partial Y_j^n}{\partial X_j^n} + \zeta_{jn}^X = P_j^D$
The first-order condition with respect to flows is:

\[
[Q^n_{jk}] = -\chi P_j^D \left( \tau^n_{jk} (Q^n_{jk}, I_{jk}) + \frac{\partial \tau^n_{jk} (Q^n_{jk}, I_{jk})}{\partial Q_{jk}} Q_{jk} \right) - (1 - \chi) P^n Q^n_{jk} \left( \tau^n_{jk} (Q^n_{jk}, I_{jk}) + \frac{\partial \tau^n_{jk} (Q^n_{jk}, I_{jk})}{\partial Q_{jk}} Q_{jk} \right) + P^n_k - P^n_j + \zeta^n_{jk} = 0 \quad (A.1)
\]

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

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

\[
[I_{jk}] = \chi \sum_{n=1}^N P_j^D Q^n_{jk} \left( -\frac{\partial \tau^n_{jk} (Q^n_{jk}, I_{jk})}{\partial I_{jk}} \right) + (1 - \chi) \sum_{n=1}^N P^n Q^n_{jk} \left( -\frac{\partial \tau^n_{jk} (Q^n_{jk}, I_{jk})}{\partial I_{jk}} \right) + \left( \zeta^n_{jk} - \zeta^n_{jk} \right) = P_k \delta^n_{jk} \quad (A.2)
\]

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

**Mobile Labor**

The Lagrangian of the problem in Definition 2 is

\[
\mathcal{L} = u - \sum_{j} \tilde{\omega}_j L_j (u - U (c_j, h_j)) - W^L \left( \sum_j L_j - L \right) - \sum_j P_j^D \left[ c_j L_j + \chi \sum_{n=1}^N \sum_{k \in \mathcal{N}(j)} \tau^n_{jk} \left( \sum_{n=1}^N m^n Q^n_{jk}, I_{jk} \right) Q^n_{jk} - D_j \left( D_j, \ldots, D_j^N \right) \right] - \sum_j P^n_j (h_j L_j - H_j)
\]

\[
- \sum_j \sum_{n} P^n_{j} D^n_{j} - \sum_{j,k \in \mathcal{N}(j)} \tau^n_{jk} (Q^n_{jk}, I_{jk}) Q^n_{jk} - F^n_{j} \left( L^n_{j}, V^n_{j}, X^n_{j} \right) - \sum_{i \in \mathcal{N}(j)} Q^n_{ij}
\]

\[
- \sum_j W_j \left[ \sum_{n} L^n_j - L_j \right] - \sum_j \sum_{j} R^n_j \left[ \sum_{n} V^n_j - V_j \right] - \sum_j \zeta^n_{jk} (I_{jk} - L_{jk}) + \sum_{j,k} \zeta^n_{jk} (I_{jk} - I_{jk}) + \sum_{j,k,n} \zeta^n_{jk} Q^n_{jk} + \sum_{j,n} \zeta^n_{jnl} L^n_j + \sum_{j,n} \zeta^n_{jnl} V^n_j + \sum_{j,n} \zeta^n_{jnl} X^n_j + \sum_{j,n} \zeta^n_{jnl} D^n_j + \sum_j \zeta^n_{jnl} c_j + \sum_j \zeta^n_{jnl} h_j
\]

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 \( [C^n_{j}] \), factor allocation within locations \( [L^n_{j}] \), \( [V^n_{j}] \), and \( [X^n_{j}] \), optimal transport \( [Q^n_{jk}] \) 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:

\[
[u] = \sum_j L_j \tilde{\omega}_j
\]

\[
[L_j] = P_j^D c_j + P^n_j h_j - \tilde{\omega}_j [U (c_j, h_j) - u] = W_j - W^L
\]

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

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

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 \( \omega_j \) with the multipliers of the
mobility constraint $\tilde{\omega}_j$. Combining $[L_i]$ with $[c_j]$ and $[h_j]$ gives the multiplier on the labor-mobility constraint. For populated locations:

$$\tilde{\omega}_j = \frac{W_i - W^L}{U^C (c_j, h_j) c_j + U^H (c_j, h_j) h_j}.$$  

### A.2 Symmetry in Infrastructure Investments

For the applications in Section 5 we impose symmetry in infrastructure levels as an additional restriction in the planner’s problem, 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

$$[I_{jk}] - \chi \sum_{n=1}^{N} \left( P^n_j Q^n_j \frac{\partial \tau^n_{jk} (Q^n_{jk}, I_{jk})}{\partial I_{jk}} + P^n_k Q^n_k \frac{\partial \tau^n_{kj} (Q^n_{kj}, I_{jk})}{\partial I_{jk}} \right)$$

$$- (1 - \chi) \sum_{n=1}^{N} \left( P^n_j Q^n_j \frac{\partial \tau^n_{jk} (Q^n_{jk}, I_{jk})}{\partial I_{jk}} + P^n_k Q^n_k \frac{\partial \tau^n_{kj} (Q^n_{kj}, I_{jk})}{\partial I_{jk}} \right) + \left( \zeta^j_{jk} - \zeta^j_{kj} \right) = P_K \left( \delta^j_{jk} + \delta^j_{kj} \right). \tag{A.3}$$

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 $I_{jk} \in ([H^j, \zeta^j_{jk}])$, is

$$I^*_j = \left[ \frac{\gamma}{P_K \left( \delta^j_{jk} + \delta^j_{kj} \right)} \left( \chi \left( \delta^j_{jk} P_j^D Q^n_{jk}^{1+\beta} + \delta^j_{kj} P_k^D Q^n_{kj}^{1+\beta} \right) + \sum_{n=1}^{N} (1 - \chi) \left( \delta^j_{jk} P_j^n Q^n_{jk}^{1+\beta} + \delta^j_{kj} P_k^n Q^n_{kj}^{1+\beta} \right) \right) \right]^{\frac{1}{\gamma - 1}}. \tag{A.4}$$

As discussed in Section 5.2, to build $\delta^{j,FOC}_{jk}$ we use (A.4) under the symmetry assumption $\delta^j_{jk} = \delta^j_{kj}$. Setting $I^*_j = I^{BS}_{jk}$, $\delta^{j,FOC}_{jk}$ can be backed out as function of calibrated parameters, the observed network $I^{BS}_{jk}$, and the equilibrium prices generated by the calibrated model. Note that, to generate these prices, we use the model calibrated given the network $I^{BS}_{jk}$, 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 $k \in N (j)$; and (ii) if in addition $Q \tau_{jk} (Q, I)$ is convex in both $Q$ and $I$ for all $j$ and $k \in N (j)$, 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

$$\max \left\{ c_j \left( D^n_j Q^n_j \right)_{k \in N(\cdot)} \right\}_{q_j}$$

subject to: (i) availability of traded commodities,

$$g^1_j = c_j + \chi \sum_{k \in N(\cdot)} \tau_{jk} (Q_{jk}, I_{jk}) Q_{jk} - D_j \left( D^1_j, ..., D^N_j \right) \leq 0 \text{ for all } j;$$

(ii) the balanced-flows constraint,

$$g^2_{jn} \equiv D^n_j + \sum_{k \in N(\cdot)} Q^n_{jk} \left[ 1 + (1 - \chi) \tau_{jk} (Q^n_{jk}, I_{jk}) \right] - F_j^n (L^n_j, V^n_j, X^n_j) - \sum_{i \in N(\cdot)} Q^n_{ij} \leq 0 \text{ for all } j, n;$$

$$52$$
(iii) the network-building constraint,

\[ \sum_{j} \sum_{k \in \mathcal{N}(j)} \delta_{jk} l_{jk} \leq K; \]

and conditions (iv)-(v) in the text. Since constraints (iii)-(v) are linear, we need \( f \) to be concave and \( g^1_j \) and \( g^2_{jn} \) to be convex. Since \( U \) is jointly concave in both its arguments, \( f \) is concave. \( D_j \{ D^n_j \} \) is concave, hence \( g^1_j \) is convex. If \( Q_{jk} (Q, I) \) is convex in then \( g^2_{jn} \) 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 \( g^1_j \) and \( g^2_{jn} \) 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 1D.4). Hence, the global optimum must satisfy the KKT conditions and the duality gap is 0.\(^{56}\)

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 \( L_j \tilde{u} \leq U_0(C_j, H_j) \). To make the latter constraint convex, let us denote \( U_j = L_j \tilde{u} \) and replace \( \tilde{u} \) in the objective function by \( \min_{j, L_j > 0} \left\{ \frac{U_j}{L_j} \right\} \), so that the problem becomes

\[
\max_{c_j, \left\{ D^n_j, L^n_j \right\}, \left\{ Q^n_{jk}, l^n_{jk} \right\}_{k \in \mathcal{N}(j)}; j L_j > 0} \min_{j, L_j > 0} \left\{ \frac{U_j}{L_j} \right\}
\]

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

\[
U_j \leq U_0(C_j, H_j) \quad \text{for all } j.
\]

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.

\[ \square \]

**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 (i.e., if \( L_{jk} = 0 \)), then \( l_{jk} = 0 \leftrightarrow P^n_j = P^n_j \) for all \( n \).

**Proof.** First, note that if \( \beta \geq \gamma \) 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 \( \beta \geq \gamma \geq 0 \). Next, from the first-order condition for optimal infrastructure (9), if the solution to the planning problem

---

\(^{56}\) Despite having substituted constraints \( g^1_j \) and \( g^2_{jn} \) into the objective function, the multipliers for these constraints, \( P^D_j \) and \( P^n_j \), can be recovered from the above KKT conditions such that \( \omega_j U_C(c_j, h_j) = P^D_j \) and \( P^n_j \partial C_j^n / \partial C_j^n = P^n_j \).

\(^{57}\) Since the objective function is strictly increasing in \( \tilde{u} \) and because \( \tilde{u} \) only shows up in the constraints \( L_j \tilde{u} \leq U_0(C_j, H_j) \) for all \( j \), it is necessarily the case that \( \tilde{u} = \min_{j, L_j > 0} U_j / L_j \).
implies \( I_{jk} = L_{jk} \), so that there is no investment, then:
\[
P_K \geq -\frac{1}{\delta_{jk}} \sum_n (\frac{P_n}{P_j}) \frac{\partial^*_n}{\partial I_{jk}} \bigg|_{I_{jk} = L_{jk}} \geq \frac{\delta_{jk}^*}{\delta_{jk}} \sum_n (\frac{P_n}{P_j}) \frac{(Q^n_{jk})^{1+\beta}}{L_{jk}^{\beta+1}} \geq \gamma (1 + \beta) \frac{\sum_n (\frac{P_n}{P_j}) (\frac{\delta_{jk}^*}{\delta_{jk}})^{\beta+1}}{L_{jk}^{\beta+1}},
\]
where the second line follows from (10) and the third line follows from (11). Note that the last inequality is equivalent to \( L_{jk} \geq I_{jk}^* \) for \( I_{jk}^* \) defined in (14). Therefore, if \( L_{jk} < I_{jk}^* \) then \( I_{jk}^* > L_{jk} \) and \( I_{jk} = I_{jk}^* \). Moreover, if there is any \( n \) such that \( P_n^* \neq P_j^* \) then \( I_{jk}^* > 0 \).

\[\Box\]

**Proposition 3.** (Tree Property) Assume that \( \lim_{c,h \to 0^+} U_C(c,h) = \infty \). In the absence of a pre-existing network (i.e., \( L_{jk} = 0 \)), 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.

**Proof.** To establish the result, we focus on the case with fixed labor.\(^{58}\) Note further that in the case with a unique tradable commodity the own-good and cross-good congestion cases are identical. 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 \( I = \{0, 1, \ldots, J - 1\} \) and \( Y_0 > 0 \) but \( Y_j = 0 \) for \( j \geq 1 \). We write down the Lagrangian of the problem
\[
\mathcal{L} = \sum_j \omega_j L_j U(c_j, h_j) - \sum_j P_j \left[ L_j c_j + \sum_{k \in N(j)} \left( 1 + \frac{Q^j_{jk}}{P_j} \right) Q_{jk} - Y_j - \sum_{k \in N(j)} Q_{kj} \right] - P_K \left( \sum_j \sum_{k \in N(j)} \delta_{jk}^I I_{jk} - K \right) + \sum_{j,k} \zeta^Q_{jk} Q_{jk} + \sum_{j,k} \zeta^I_{jk} I_{jk}, \quad \zeta^Q_{jk} \geq 0, \zeta^I_{jk} \geq 0.
\]

Despite being a nonconvex optimization problem, there must exist a vector of Lagrange multipliers such that the KKT conditions hold.\(^{59}\) As a preliminary step, we eliminate the infrastructure investment \( I_{jk} \) using (12), so that \( I_{jk} = \left( \frac{\omega_j}{P_j} \frac{\delta_{jk}^I}{\delta_{jk}} P_j Q^{1+\beta}_{jk} \right)^{-\gamma} \) whenever \( Q_{jk} > 0 \), otherwise \( I_{jk} = 0 \). Solving for the multiplier \( P_K \) such that (7) is satisfied, we reformulate the problem with allocation and flows only as follows
\[
\mathcal{L} = \sum_j \omega_j L_j U(c_j, h_j) - \sum_j P_j \left[ L_j c_j + \sum_{k \in N(j)} Q_{jk} - Y_j - \sum_{k \in N(j)} Q_{kj} \right] - K^{-\gamma} \left[ \sum_{j,k} \left( \frac{\delta_{jk}^I}{\delta_{jk}^Q} \right)^{\gamma+1} \left( P_j Q^{1+\beta}_{jk} \right)^{\gamma+1} \right] + \sum_{j,k} \zeta^Q_{jk} Q_{jk}, \quad \zeta^Q_{jk} \geq 0. \tag{A.5}
\]

\[\Box\]

The source of nonconvexities is the term \( \left[ \sum_{j,k} \left( \frac{\delta_{jk}^I}{\delta_{jk}^Q} \right)^{\gamma+1} \left( P_j Q^{1+\beta}_{jk} \right)^{\gamma+1} \right]^{-\gamma+1} \), which is convex when \( \beta \geq \gamma \), but neither convex nor concave when \( \gamma > \beta \). Let us now assume that \( (c^*, Q^*) \) with \( c^* = (c_{j1}^*, \ldots, c_{j\gamma-1}^*)^T \) and \( Q^* = \]

\(^{58}\)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 \( \omega_j \) with the Lagrange multipliers of the constraints \( L_j u \leq L_j U(c_j, h_j) \).

\(^{59}\)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.
\((Q^*_n)_{k \in N(j)}\) is a local optimum, i.e., it satisfies the FOCs and SOCs of the Lagrangian \((A.5)\). We are going to show that the graph associated with \(Q^*\) is a tree. Define the (undirected) graph associated to \(Q^*\) as the tuple \((I, \mathcal{E}^*)\) such that \(\mathcal{E}^* \subset \mathcal{E}\) is a subset of the edges of the underlying geography such that

\[
\mathcal{E}^* = \{ \{j, k\} \in \mathcal{E} \mid Q^*_{jk} > 0 \}.
\]

Note that since \(I_{jk}\) is non-zero whenever \(Q^*_{jk} > 0\) or \(Q^*_{kj} > 0\), the support of graph \((I, \mathcal{E}^*)\) coincides with that of the transport network \(\{I_{jk}\}\). After this preparatory work, we now refer the reader to Proposition 6 in Appendix G which establishes that \(\mathcal{E}^*\) is a tree.

**Definition 3.** The decentralized equilibrium without labor mobility consists of quantities \(Q^*(p_\pi^n, p_j^D, p_j^H)\) and factor prices \(w_j, \{r^n_m\}_m\) in each location \(j\) such that:

(i)(a) consumers optimize:

\[
\arg\max_{\hat{c}_j, \hat{h}_j} U(\hat{c}_j, \hat{h}_j) = p_j^D \hat{c}_j + p_j^H \hat{h}_j = c_j \equiv w_j + t_j,
\]

where \(c_j\) are expenditures per worker in \(j\) and where \(p_j^D\) is the price index associated with \(D_j (D_j^\tau, .., D_j^N)\) at prices \(p_j^H\) and \(t_j\) is a transfer per worker located in \(j\). The set of transfers satisfy

\[
\sum_j t_j L_j = \Pi
\]

where \(\Pi\) adds up the aggregate returns to the portfolio of fixed factors and the government tax revenue,

\[
\Pi = \sum_j p_j^H H_j + \sum_j \sum_m r_j^n V_j^m + \sum_j \sum_{k \in N(j)} n_j^m \sum_n Q^*_{jk}^n;
\]

(i)(b) firms optimize:

\[
\{L_j^n, V_j^n, X_j^n\} = \arg\max_{L_j^n, V_j^n, X_j^n} \left( \sum_j p_j^n L_j^n - \sum_m r_j^n V_j^m; \right)
\]

(i)(c) the transport companies optimize,

\[
\pi_{ad} = \max_{r = (j_0, \ldots, j_{\rho}) \in R_{ad}} \left( p_d - p_o - \sum_{k = 0}^{\rho-1} \left( \chi p_j^n r_{j_k,j_{k+1}} + (1 - \chi) D_j^n r_{j_k,j_{k+1}} \right) - \sum_{k = 0}^{\rho-1} p_{j_{k+1},j_{k+1}}^{n} ; \right)
\]

for all \((a, d) \in J^2\), where \(R_{ad} = \{(j_0, \ldots, j_{\rho}) \in J^{\rho+1} \mid j_0 = a, j_{\rho} = d, j_{\rho+1} \in \mathcal{N} (j_{\rho}) \} \) for all \(0 \leq k < \rho\) is the set of routes from \(o\) to \(d\), and there is free entry to delivering products from every source to every destination: \(\pi_{ad} \leq 0\) for all \((a, d) \in J^2\), if good \(n\) is shipped from \(o\) to \(d\).

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

\[
\{D_j^n\} = \arg\max_{D_j^n} D_j^n \left( \left\{ \hat{D}_j^n \right\} - \sum_j p_j^n \hat{D}_j^n; \right)
\]

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 tax on shipments of product \(n\) from \(j\) to \(k\), denominated in the same unit as transport costs, is \(t_{jk}^n = \chi m^n \hat{e}_{Q^*jk} (Q^*_{jk}, I_{jk}) + (1 - \chi) \hat{e}_{Q^*jn} r_{jk}^n (Q^*_{jk}, I_{jk}) \) where
This condition is consistent with the first-order condition \( \omega \) equals the multiplier on the balanced-flows constraint in the planner’s allocation, \( \beta \). 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., \( t_j = \frac{P_j}{\bar{P}} \). 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, \( \beta \).

**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 \( k \in \mathcal{N}(j) \),

\[
 p^j_k \leq p^j_j + (1 - \chi) p^j_j (\tau^j_k + \epsilon^j_k) + \chi p^j_j (m^j \tau^j_k + \epsilon^j_k) , \quad \text{if } Q^j_{jk} > 0.
\]

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:

\[
 \begin{align*}
 [c_j] & \quad \left( \frac{1}{\lambda_j} \right) U^j_e (c_j, h_j) = p^j_D \\
 [h_j] & \quad \left( \frac{1}{\lambda_j} \right) U^j_H (c_j, h_j) = p^j_H \\
 [\bar{D}^j_n] & \quad p^j_j \frac{\partial D^j_P}{\partial D^j_P} = p^j_n \\
 \left[ \bar{L}^j_n \right] & \quad \frac{\partial Y^j_n}{\partial L^j_n} P^j_n \leq w_j , \quad \text{if } L^j_n > 0 \\
 \left[ V^{mn}_j \right] & \quad \frac{\partial Y^j_{mn}}{\partial V^{mn}_j} P^j_n \leq r^j_{mn} , \quad \text{if } V^{mn}_j > 0.
\end{align*}
\]

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 \( \omega_j \) 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 \( U = G (U_0 (c, h)) \), where \( U_0 \) is homogeneous of degree 1. Then, the planner’s allocation coincide with the market’s under weights

\[
 \omega_j = \frac{c_j}{G' (U_0 (c_j, h_j)) U_0 (c_j, h_j)}
\]

where \( c_j \) is the expenditure per worker and \( c_j, h_j \) are the consumption per worker of the traded and non-traded goods in the market allocation. If \( U \) is homogeneous of degree one, then \( \omega_j = P^U_j \), where \( P^U_j \) is the price index associated with \( U(c_j, h_j) \) at the market equilibrium prices \( p^D_j, p^H_j \). In the opposite direction, given arbitrary weights \( \omega_j \), the market allocation implements the planner’s under the transfers \( t_j = p^D_j c_j + p^H_j h_j - W_j \) constructed using the quantities \( \{c_j, h_j\} \) from the planner’s allocation and the multipliers \( \{p^D_j, p^H_j\} \) 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:

\[
 p^D_j c_j + p^H_j h_j = W_j - W^L.
\]

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 \( e_j = w_j + \text{Constant} \) for all \( j \). The only transfer scheme delivering the same
and we now evaluate each term separately. Consider the generic deviation term which simplifies to 

\[ r_{j,l} \]

for a given route transfer per capita is \( t_j = \frac{n}{T} \).

**Equivalence of Least Cost Routes.** We want to establish that any route used in the planner’s problem is a solution to (i)(c) in Definition 3 under the proposed tax scheme. Fix good \( n \). We introduce the following notation: for a given route \( r = (j_0, \ldots, j_p) \in \mathcal{R}_{od} \), we denote by \( f_{r}^{n} \) the matrix of flows

\[
f_{r}^{n}(j,k) = \begin{cases} 1 & \text{if } \exists l, 0 \leq l \leq \rho - 1 \text{ and } j = j_{l}, k = j_{l+1} \smallskip \text{otherwise} \\
0 & \end{cases}
\]

Consider an optimal route from \( o \) to \( d \), \( r^{*} = (j_{0}^{*}, \ldots, j_{p}^{*}) \in \mathcal{R}_{od}, \) i.e., such that \( Q_{j_{l}j_{l+1}}^{n} = 0 \) at the optimum of the planner’s problem \( (r_{j_{l}j_{l+1}}^{Q} = 0) \). We now consider redirecting a marginal amount of goods \( \varepsilon > 0 \) from \( r^{*} \) to some other route \( r = (j_0, \ldots, j_p) \in \mathcal{R}_{od} \). In other words, denoting \( Q^{n} = (Q_{jk}^{n})_{j \in \mathcal{J}, k \in \mathcal{N}(j)} \), we consider the perturbation \( Q^{n} + \varepsilon f_{r}^{n} - \varepsilon f_{r}^{n} \). The first-order effect of the deviation around the optimum must reduce the Lagrangian 

\[ \mathcal{L}(Q^{n} + \varepsilon f_{r}^{n} - \varepsilon f_{r}^{n}) - \mathcal{L}(Q^{n}) \leq 0. \]

To translate this condition as a minimum cost route problem, we decompose the first-order impact on the Lagrangian as follows

\[ \mathcal{L}(Q^{n} + \varepsilon f_{r}^{n} - \varepsilon f_{r}^{n}) - \mathcal{L}(Q^{n}) = \mathcal{L}(Q^{n} + \varepsilon f_{r}^{n}) - \mathcal{L}(Q^{n}) - (\mathcal{L}(Q^{n} + \varepsilon f_{r}^{n} - \varepsilon f_{r}^{n}) - \mathcal{L}(Q^{n})) + o(\varepsilon), \]

and we now evaluate each term separately. Consider the generic deviation term

\[
\mathcal{L}(Q^{n} + \varepsilon f_{r}^{n}) - \mathcal{L}(Q^{n}) = \sum_{l=0}^{\rho-1} (-P_{j_{l}}^{n} + P_{j_{l}+1}^{n} + \zeta_{j_{l}j_{l+1}}^{r, n} - \chi P_{j_{l}}^{D} m^{n} \tau_{j_{l}j_{l+1}}^{n} + \varepsilon (1 - \chi) P_{j_{l}}^{n} \tau_{j_{l}j_{l+1}}^{n} + \varepsilon \left( \tau_{j_{l}j_{l+1}}^{n} + \zeta_{j_{l}j_{l+1}}^{r, n} \frac{\partial \tau_{j_{l}j_{l+1}}^{n}}{\partial Q_{j_{l}j_{l+1}}^{n}} \right) \right) \varepsilon + o(\varepsilon), \]

which simplifies to

\[
\mathcal{L}(Q^{n} + \varepsilon f_{r}^{n}) - \mathcal{L}(Q^{n}) = \left[ P_{o}^{n} - P_{a}^{n} + \sum_{l=0}^{\rho-1} \zeta_{j_{l}j_{l+1}}^{r, n} - \sum_{l=0}^{\rho-1} \chi P_{j_{l}}^{D} m^{n} \tau_{j_{l}j_{l+1}}^{n} + (1 - \chi) P_{j_{l}}^{n} \tau_{j_{l}j_{l+1}}^{n} \right] \varepsilon + o(\varepsilon). \]

Using the definition of the optimal tax, we have that

\[ t_{j_{l}j_{l+1}}^{n} = \chi Q_{j_{l}j_{l+1}}^{n} m^{n} \frac{\partial \tau_{j_{l}j_{l+1}}^{n}}{\partial Q_{j_{l}j_{l+1}}^{n}} + (1 - \chi) Q_{j_{l}j_{l+1}}^{n} \frac{\partial \tau_{j_{l}j_{l+1}}^{n}}{\partial Q_{j_{l}j_{l+1}}^{n}}. \]

Substituting into the deviation term, we obtain

\[
\mathcal{L}(Q^{n} + \varepsilon f_{r}^{n}) - \mathcal{L}(Q^{n}) = \left[ P_{o}^{n} - P_{a}^{n} - \sum_{l=0}^{\rho-1} \chi P_{j_{l}}^{D} m^{n} \tau_{j_{l}j_{l+1}}^{n} + (1 - \chi) P_{j_{l}}^{n} \tau_{j_{l}j_{l+1}}^{n} \right] \varepsilon + o(\varepsilon). \]

By assumption, the total deviation \( Q^{n} + \varepsilon f_{r}^{n} - \varepsilon f_{r}^{n} \) has a negative impact on the Lagrangian for the feasible deviation
\[ \varepsilon > 0 \text{ and that } \zeta^Q_{j_{i+1}k} = 0, \text{ we get:} \]

\[
P^*_n = \sum_{l=0}^{\rho^*} \left( \chi P_{j_l}^D m^n \tau_{j_l,j_{l+1}} + (1 - \chi) P_{j_l}^o \tau_{j_l,j_{l+1}}^n \right) + \sum_{l=0}^{\rho^*} P_{j_l}^n t_{j_l,j_{l+1}}^n \]

\[
\leq P^o_n + \sum_{l=0}^{\rho^*} \left( \chi P_{j_l}^D m^n \tau_{j_l,j_{l+1}} + (1 - \chi) P_{j_l}^o \tau_{j_l,j_{l+1}}^n \right) + \sum_{l=0}^{\rho^*} \left( \chi P_{j_l}^D + (1 - \chi) P_{j_l}^o \right) t_{j_l,j_{l+1}}^n - \sum_{l=0}^{\rho^*} \zeta^Q_{j_l,j_{l+1}} n
\]

\[
\leq P^o_n + \sum_{l=0}^{\rho^*} \left( \chi P_{j_l}^D m^n \tau_{j_l,j_{l+1}} + (1 - \chi) P_{j_l}^o \tau_{j_l,j_{l+1}}^n \right) + \sum_{l=0}^{\rho^*} \left( \chi P_{j_l}^D + (1 - \chi) P_{j_l}^o \right) t_{j_l,j_{l+1}}^n.
\]

Hence, the optimal route \( \tau^* \) is solution to the least-cost route problem

\[
\min_{r=(j_0, \ldots, j_p) \in \mathcal{R}_{od}} P^o_n + \sum_{l=0}^{\rho^*} \left( \chi P_{j_l}^D m^n \tau_{j_l,j_{l+1}} + (1 - \chi) P_{j_l}^o \tau_{j_l,j_{l+1}}^n \right) + \sum_{l=0}^{\rho^*} \left( \chi P_{j_l}^D + (1 - \chi) P_{j_l}^o \right) t_{j_l,j_{l+1}}^n,
\]

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

Note, finally, that the minimum cost route problem in the case of own-good congestion (\( \chi = 0 \)) is equivalent to

\[
\min_{r=(j_0, \ldots, j_p) \in \mathcal{R}_{od}} P^o_n \prod_{l=0}^{\rho^*} \left( 1 + \tau_{j_l,j_{l+1}}^n \right) + \sum_{l=0}^{\rho^*} P_{j_l}^n t_{j_l,j_{l+1}}^n \prod_{k=l}^{\rho^*} \left( 1 + \tau_{j_k,j_{k+1}}^n \right)
\]

since \( P^o_{j_{i+1}} = \left( 1 + \tau_{j_{i+1}}^n \right) P^o_{j_i} + P^o_{j_{i+1}} t_{j_{i+1}}^n \) along any used path.

\[ \square \]

**Proposition 5.** If the transport technology is convex (\( \beta \geq \gamma \)) and if the toll is set equal to the optimal Pigovian tax, \( \theta^n_{jk} = P^n_j \delta^\gamma_{jk} \), then the decentralized infrastructure choice implements the optimal network investment along with specific transfers \( t_j \).

**Proof.** For simplicity, we focus on the case with own-good congestion. The case with cross-good congestion is a straightforward extension. We consider the problem of a regulated monopoly on link \( jk \) allowed by law to charge a fixed toll cost \( \theta^n_{jk} \) per-unit of good \( n \) shipped. The monopolist can purchase asphalt from the government at price \( pK \). Assume the government forbids entry on unused links or sets a price too low for entry, allowing us to focus on links used at the social optimum (\( Q^n_{jk} > 0 \) for some \( n, I_{jk} > 0 \)).

Free-entry of shipping companies on link \( jk \) yields

\[
P^n_k \leq P^n_j (1 + \tau^n_{jk}) + \theta^n_{jk}, = \text{if } Q^n_{jk} > 0.
\]

Under the assumption that the transport technology takes the log-linear form \( \tau^\gamma_{jk} \left( Q^n_{jk}, I_{jk} \right) = \delta^\gamma_{jk} \left( Q^n_{jk} \right) \delta^\gamma_{jk} I_{jk} \), one can derive the demand for transport

\[
Q^n_{jk} = \left( P^n_k - P^n_j - \theta^n_{jk} \delta^\gamma_{jk} I_{jk} \right) ^{\frac{1}{\gamma}} \text{ for } P^n_k \geq P^n_j + \theta^n_{jk}.
\]

The monopoly solves the following profit maximization problem

\[
\max_{I_{jk}} \sum_n Q^n_{jk} \theta^n_{jk} - pK \delta^\gamma_{jk} I_{jk}
\]

subject to \( Q^n_{jk} = \left( P^n_k - P^n_j - \theta^n_{jk} \delta^\gamma_{jk} I_{jk} \right) ^{\frac{1}{\gamma}}, \forall n. \)

Under the condition that guarantee the convexity of our planning problem, \( \gamma \leq \beta \), this problem is well defined and
constrained. The first-order condition over infrastructure is

\[
\frac{\gamma}{\beta} \sum_n \left( \frac{P^n - P^n - \theta^n_k}{P^n} I^n_k \right) \frac{\partial \theta^n_k}{\partial j_k} = p_K \delta^n_k.
\]

Assume now that the toll cost is set exactly to the Pigovian congestion tax from Proposition 4,

\[
\theta^n_{jk} = P^n t^n_{jk} = \beta P^n \tau^n_{jk} (Q^n_{jk}, I^n_{jk}).
\]

Under the Pigovian tax, the commodity flows correspond are optimal, \(Q^n_{jk}\). Substituting \(Q^n_{jk}\) into the equation, we obtain

\[
\frac{\gamma}{\beta} \sum_n Q^n_{jk} \theta^n_{jk} = p_K \delta^n_{jk},
\]

which further simplifies to

\[
\frac{\gamma}{\beta} \sum_n \delta^n_{jk} P^n (Q^n_{jk})^{1+\beta} I^n_{jk}^{-\gamma} = p_K \delta^n_{jk}.
\]

One may now conclude by recognizing that this is equation is identical to equation (12) when the price of asphalt \(p_K\) is set equal to the value of the Lagrange multiplier \(P^n\). Note, finally, that the operating company makes profits in equilibrium. These profits should be rebated to consumers as part of the general transfers \(t_j\) as in Proposition (4) to make sure that each budget constraint is satisfied in equilibrium.

\[\blacksquare\]

### A.4 Appendix to Section 3.6 (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 3.6, 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, \(\tau^n_{jk} = \delta^n_{jk} (Q^n_{jk})^{\beta} I^n_{jk}\); (own-good congestion, \(\chi = 0\)) or \(\tau^n_{jk} = m^n \delta^n_{jk} (Q^n_{jk})^{\beta} I^n_{jk}\); with \(Q_{jk} = \sum_{n=1}^N m^n Q^n_{jk}\) (cross-good congestion, \(\chi = 1\)); ii) labor is the sole production factor, \(F^n_j \left(L^n_j\right) = z^n_j \left(L^n_j\right)^{\alpha}\); and iii) \(C^n\) is a CES aggregator with elasticity of substitution \(\sigma\). We consider the case with immobile labor.\(^{60}\)

We write the Lagrangian of the problem

\[
\mathcal{L} = \sum_j \omega_j L_j \left(c_j, h_j\right) - \sum_j \sum_{P^n_j D_j} \left[ c_j L_j + \chi \sum_{k \in N(j)} \delta^n_{jk} \left( \sum_{n=1}^N m^n Q^n_{jk} \right)^{1+\beta} I^n_{jk}^{-\gamma} - \left( \sum_n (D^n_j)^{\frac{\beta-1}{\beta}} \right)^{\frac{\sigma}{\beta-1}} \right]
\]

\[
- \sum_j \sum_n P^n_j \left[ D_j^n + \sum_{k \in N(j)} \left( Q^n_{jk} + (1-\chi) \delta^n_{jk} (Q^n_{jk})^{1+\beta} I^n_{jk}^{-\gamma} \right) - z^n_j (L^n_j)^{\alpha} - \sum_{i \in N(j)} Q^n_{ij} \right]
\]

\[
- \sum_j W_j \left[ \sum_n L^n_j - L_j \right] + \sum_{j, k, n} \zeta^n_{jkn} Q^n_{jk} + \sum_{j, n} \zeta^n_{jnn} L^n_j + \sum_{j, n} \zeta^n_{jnn} D^n_j + \sum_j \zeta^n_i c_j.
\]

\(^{60}\)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.
Recall that the dual problem consists of solving

\[
\inf_{\lambda \geq 0} \sup_{x} \mathcal{L}(x, \lambda).
\]

We start by expressing our control variables \( x = (c_j, D^j_n, Q^n_{jk}, L^n_j) \) as functions of the Lagrange multipliers \( \lambda = (P^D_j, P^n_j, W^Q_j, \xi^n_{jkn}, \xi^L_j, \xi^n_C, \xi^L_C, \xi^Q_j) \). Using the optimality conditions, one obtains the following expressions:

\[
c_j = U^{-1}_c \left( \omega_j^{-1} \left( \sum_{n'} \left( P^n_{j'} \right)^{1-a} \right)^{1-a}, h_j \right)
\]

\[
D^j_n = \left[ \frac{P^n_j}{\left( \sum_{n'} \left( P^n_{j'} \right)^{1-a} \right)^{1-a}} \right]^{1-a} \left( c_j L_j + \chi \sum_{k \in N(j)} \delta^n_{jk} \left( \sum_{n=1}^N m^n Q^n_{jk} \right)^{1+\beta} I^{-\gamma}_{jk} \right)
\]

\[
L^n_j = \frac{\left( P^n_j \right)^{1-a}}{\sum_{n'} \left( P^n_{j'} \right)^{1-a}} L_j.
\]

In the case of own-good congestion (\( \chi = 0 \)), the flows can be easily inverted,

\[
Q^n_{jk} = \left[ \frac{1}{1+\beta} \delta^n_{jk} \max \left( \frac{P^n_k}{P^n_j} - 1, 0 \right) \right]^{1/\beta},
\]

but in the case of cross-good congestion, one first needs to invert the aggregate flows \( Q_{jk} = \sum_{n=1}^N m^n Q^n_{jk} \), which is sufficient to evaluate the Lagrangian at any point,\(^{61}\)

\[
Q_{jk} = \max_{n} \left( \frac{P^n_k - P^n_j}{(1+\beta) P^n_j m^n \delta^n_{jk} I^{-\gamma}_{jk}} \right)^{1/\beta}.
\]

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 \( \lambda = (P^n_j)_{j,n} \). We may now compute the inner part of the saddle-point problem:\(^{62}\)

\[
\mathcal{L}(x(\lambda), \lambda) = \sum_j \omega_j L_j U(c_j(\lambda), h_j) - \sum_j P^D_j \left( c_j(\lambda) L_j + \chi \sum_{k \in N(j)} \delta^n_{jk} Q^n_{jk}^{1+\beta}(\lambda) I^{-\gamma}_{jk} \right)
\]

\[
- \sum_j \sum_n P^n_j \left[ D^n_j(\lambda) + \sum_{k \in N(j)} \left( Q^n_{jk}(\lambda) + (1-\chi) \delta^n_{jk} (Q^n_{jk}(\lambda))^{1+\beta} I^{-\gamma}_{jk} \right) - z^n_j (L_j(\lambda))^a - \sum_{i \in N(j)} Q^n_{ij} \right].
\]

The dual problem then consists of the simple unconstrained, convex\(^{63}\) minimization problem in \( J \times N \) unknowns:

\[
\min_{\lambda, x} \mathcal{L}(x(\lambda), \lambda).
\]

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

---

\(^{61}\)The specific values \( Q^n_{jk} \) can be recovered at the end of the optimization by inverting the linear system (for given multipliers \( \lambda \)) which corresponds to the balanced-flows constraints, the constraints \( Q_{jk}(\lambda) = \sum_{n=1}^N m^n Q^n_{jk} \) and the complementary slackness conditions, \( \xi^n_{jkn} Q^n_{jk} = 0 \).

\(^{62}\)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.

\(^{63}\)Dual problems are always convex, by construction, even when the primal problem is not.
can be described as follows: Simulated annealing is a very popular probabilistic method to search for the global optimum in a reasonable amount of time. Unfortunately, there exists to our knowledge few global optimization methods that would guarantee convergence to a global maximum in a reasonable amount of time.

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 optimum 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 nonconvex cases

$$\nabla Z(\mathbf{x}(\lambda), \lambda) = - \begin{pmatrix} C^0_j(\lambda) + \sum_{k \in \mathcal{N}(j)} (Q^0_j(\lambda) + \delta^0_{jk}(Q^0_j(\lambda))^{1+\beta} \Gamma_{jk}^{-\alpha}) - \varepsilon \theta_j(L^0_j(\lambda)) - \sum_{i \in \mathcal{N}(j)} Q^0_{ij}(\lambda) \\ \vdots \\ \vdots \end{pmatrix}.$$ 

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. 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 $l := 1$. Guess some initial level of infrastructure $\{I^{(1)}_{jk}\}$ that satisfies the network building constraint.
2. Given the network $\{I^{(l)}_{jk}\}$, solve for $(c_j, D^0_j, Q^0_{jk}, L^0_j)$ using a duality approach.
3. Given the flows $Q^0_{jk}$ and the prices $P^0_j$, get a new guess $I^{(l+1)}_{jk} = \left[ \frac{\gamma \delta_{jk} \left( P^0_j Q^{1+\beta}_{jk} \right)} {\sum_{n} \frac{\delta_{jk} \left( P^0_j (Q^0_{jk})^{1+\beta} \right)}} \right]^{1/\gamma}$ for $\chi = 1$ (or $I^{(l+1)}_{jk} = \left[ \frac{\gamma \delta_{jk} \left( \sum_{n} P^0_j (Q^0_{jk})^{1+\beta} \right)} {\sum_{n} \frac{\delta_{jk} \left( P^0_j Q^{1+\beta}_{jk} \right)}} \right]^{1/\gamma}$ for $\chi = 0$) and set $P_K$ such that $\sum \delta_{jk} I^{(l+1)}_{jk} = K$.
4. If $\sum_{j,k} |I^{(l+1)}_{jk} - I^{(l)}_{jk}| \leq \varepsilon$, then we have converged to a potential candidate for a local optimum. If not, set $l := l + 1$ and go back to (2).

1. Let $l := 1$. Set the initial network $\{I^{(1)}_{jk}\}$ 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 $\{\hat{I}_{jk}\}$ by perturbing $\{I^{(l)}_{jk}\}$ (see below). [Optional: deepen the network.] Compute the corresponding optimal allocation and transport $\{c_j, D^0_j, Q^0_{jk}, L^0_j\}$. Compute associated welfare $\hat{v}$.
3. Accept the new network, i.e., set $I^{(l+1)}_{jk} = \hat{v}$ 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, $\{I^{(l+1)}_{jk}\} = \{I^{(l)}_{jk}\}$ and $v^{(l+1)} = v^{(l)}$. 

$^64$ Techniques such as the branch-and-bound method are guaranteed to converge to the global optimum, but remain heavy to implement and computationally intensive.
4. Stop when $T < T_{\text{min}}$. Otherwise set $l := l + 1$ and $T := \rho T$ and return to (2),

where $\rho < 1$ 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 $\{ \hat{I}_{jk} \}$ 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 propose the two different perturbation methods that have consistently produced the best results throughout our simulations.

**Algorithm #1 “Rebranching”**. This simple algorithm 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 $I \subset J$ for “random rebranching” or set $I = J$ for “deterministic rebranching”.
2. For each $j \in I$, identify $m(j)$ as the neighbor with the lowest price index for the bundle of tradable goods, $m(j) = \arg\min_{k \in N(j)} P_{nk}^0$, and $n(j)$ as the “parent” with the highest level of infrastructure, $n(j) = \arg\max_{k \in N(j)|P_{nk} \leq P_{nj}^{I_{kj}}} I_{kj}$.
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.

**Algorithm #2 “Hybrid Alder-FS”**. This algorithm attempts to implement the spirit of the algorithm proposed in Alder (2019), while exploiting the continuity of infrastructure investments and differentiability of the problem. The algorithm can be described as follows:

1. Compute the welfare gradient with respect to infrastructure investments and some given price of asphalt $P_K$. Delete the 5% links that correspond to the lowest elements of the gradient.
2. Compute the new economic allocation given the new network.
3. Compute the new welfare gradient and add a predefined quantity of asphalt to the link that has the highest element in the welfare gradient among inactive links.
4. Rescale the network so that the total quantity of asphalt is used and make sure that the network is connected (otherwise add the most beneficial link).
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 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.4: 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.5: 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.6: The Optimal Transport Network under Alternative Building Costs

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.
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 $I_{jk}$ involves the cheapest path $\mathcal{P}(j,k)$ for all $j \in J$ 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

$$
\text{length}_s \chi_{\text{lane}} \chi_{\text{nat}} \chi_{\text{paved}} \chi_{\text{median}},
$$

where $\text{length}_s$ is the length of the segment, $\chi_{\text{lane}}$ is the number of lanes, $\chi_{\text{nat}}$ equals 1 if the segment belongs to a national road, $\chi_{\text{paved}}$ equals 1 if the segment is paved, and $\chi_{\text{median}}$ equals 1 if the segment has a median. We parametrize $\chi_{\text{lane}}, \chi_{\text{nat}}, \chi_{\text{paved}},$ and $\chi_{\text{median}}$ 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. 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 lanes on top of 2 lanes, a 125% increase in the number of lanes, reduces costs by 25%, implying an elasticity of user costs with respect to number of lanes of $\chi_{\text{lane}} = 0.25$. 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 $\chi_{\text{nat}} = 1.07$. According to Figueroa et al. (2013), road user costs are 35% higher on gravel relative to paved roads, implying $\chi_{\text{paved}} = 1.35$, and according to Tay and Churchill (2007), adding a median increases speed by 5%, implying $\chi_{\text{median}} = 1.05$.

Calibration of $\beta$ and $\gamma$ To parametrize $\beta$ and $\gamma$ we assume that: i) the trade cost is a linear function of shipping time, $\tau_{jk} = a^\tau_{jk} \text{dist}_{jk} / S_{jk}$; ii) shipping speed is a log-linear function of the number of vehicles and road lane kilometers,

$$
S_{jk} = a^S_{jk} I_{jk} V_{jk}^{-\beta};
$$

(A.7)

and iii) the number of vehicles is a linear function of the quantity of goods that is shipped, $V_{jk} = a^V_{jk} Q_{jk}$. These assumptions are consistent with the functional form 10 for $\tau_{jk}(Q,I)$. To recover the parameters $\gamma$ and $\beta$, one would ideally like to estimate the relationship between speed, roads, and vehicles in (A.7) across links. This relationship is estimated by Couture et al. (2018). Equation (2) in their paper assumes a log-linear relationship between speed, roads, and vehicle travel time, defined as vehicle-kilometers (i.e., vehicles times distance) over speed. To measure roads they use the log of lane-kilometers of interstate highways in the U.S., which corresponds to our measure of $I_{jk}$. Assuming that their estimates would hold at the level of a connection between populated areas in our data, equation (2) of their paper can be written, in our notation:

$$
\ln S_{jk} = \alpha^{\text{CDT}} \ln I_{jk} - \theta^{\text{CDT}} \ln \left( \frac{V_{jk} \times \text{dist}_{jk}}{S_{jk}} \right) + \varepsilon_{jk},
$$

(A.8)

where we use $\alpha^{\text{CDT}}$ and $\theta^{\text{CDT}}$ to refer to $\alpha$ and $\beta$ in their paper. These parameters translate to ours as follows: $\alpha^{\text{CDT}} \equiv 1 / \gamma$ and $\theta^{\text{CDT}} \equiv \beta / \gamma$. When $\alpha^{\text{CDT}} < \theta^{\text{CDT}}$ there are decreasing returns to scale in the provision of vehicle kilometers traveled. Couture et al. (2018) find decreasing returns to scale ($\alpha^{\text{CDT}} < \theta^{\text{CDT}} \rightarrow \gamma < \beta$) across all their specification (Tables 5 and 6 of their paper) using a variety of OLS and IV approaches. Their preferred estimate (column 6 of table 5) yields $\alpha^{\text{CDT}} = 0.09$ and $\theta^{\text{CDT}} = 0.13$, implying $\gamma = 0.10$ and $\beta = 0.13$.

\footnote{In their notation, vehicle travel time (number of vehicles times time of travel) is $VTT_i = \frac{VKT_i}{S_i}$, where $VKT_i$ is vehicle kilometers (number of vehicles times distance) and $S_i$ is speed (distance over time of travel). In our notation, therefore, $VKT_i = V_j \times \text{dist}_{jk}$.}
<table>
<thead>
<tr>
<th>Country</th>
<th>Code</th>
<th>Length (Km.)</th>
<th>Number of Segments</th>
<th>Average Lanes per Km.</th>
<th>Number of Cells</th>
<th>Length (Km.)</th>
<th>Average Infrastructure Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>AT</td>
<td>17229</td>
<td>6156</td>
<td>2.36</td>
<td>46</td>
<td>9777</td>
<td>1.53</td>
</tr>
<tr>
<td>Belgium</td>
<td>BE</td>
<td>19683</td>
<td>10472</td>
<td>2.49</td>
<td>21</td>
<td>3535</td>
<td>2.12</td>
</tr>
<tr>
<td>Switzerland</td>
<td>CHLI</td>
<td>14451</td>
<td>11043</td>
<td>2.26</td>
<td>24</td>
<td>5307</td>
<td>1.71</td>
</tr>
<tr>
<td>Cyprus</td>
<td>CY</td>
<td>2818</td>
<td>947</td>
<td>2.29</td>
<td>20</td>
<td>2077</td>
<td>0.69</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>CZ</td>
<td>28651</td>
<td>10183</td>
<td>2.20</td>
<td>48</td>
<td>10593</td>
<td>1.01</td>
</tr>
<tr>
<td>Germany</td>
<td>DE</td>
<td>115173</td>
<td>66423</td>
<td>2.42</td>
<td>54</td>
<td>27227</td>
<td>2.61</td>
</tr>
<tr>
<td>Denmark</td>
<td>DK</td>
<td>15358</td>
<td>6297</td>
<td>2.20</td>
<td>30</td>
<td>5418</td>
<td>0.99</td>
</tr>
<tr>
<td>Spain</td>
<td>ES</td>
<td>101990</td>
<td>18048</td>
<td>2.39</td>
<td>61</td>
<td>30686</td>
<td>1.80</td>
</tr>
<tr>
<td>Finland</td>
<td>FI</td>
<td>70394</td>
<td>9221</td>
<td>2.04</td>
<td>70</td>
<td>34057</td>
<td>0.28</td>
</tr>
<tr>
<td>France</td>
<td>FR</td>
<td>128787</td>
<td>38668</td>
<td>2.05</td>
<td>74</td>
<td>38754</td>
<td>2.12</td>
</tr>
<tr>
<td>Georgia</td>
<td>GE</td>
<td>28682</td>
<td>9009</td>
<td>1.95</td>
<td>39</td>
<td>8420</td>
<td>0.34</td>
</tr>
<tr>
<td>Hungary</td>
<td>HU</td>
<td>32728</td>
<td>9235</td>
<td>2.10</td>
<td>50</td>
<td>11572</td>
<td>0.95</td>
</tr>
<tr>
<td>Ireland</td>
<td>IE</td>
<td>24952</td>
<td>4144</td>
<td>2.10</td>
<td>47</td>
<td>10671</td>
<td>0.63</td>
</tr>
<tr>
<td>Italy</td>
<td>IT</td>
<td>77586</td>
<td>44155</td>
<td>2.32</td>
<td>36</td>
<td>14742</td>
<td>2.18</td>
</tr>
<tr>
<td>Lithuania</td>
<td>LT</td>
<td>10682</td>
<td>1586</td>
<td>2.39</td>
<td>43</td>
<td>9408</td>
<td>0.59</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>LU</td>
<td>1778</td>
<td>863</td>
<td>2.31</td>
<td>8</td>
<td>574</td>
<td>0.73</td>
</tr>
<tr>
<td>Latvia</td>
<td>LV</td>
<td>11495</td>
<td>2102</td>
<td>2.03</td>
<td>47</td>
<td>9857</td>
<td>0.33</td>
</tr>
<tr>
<td>Moldova</td>
<td>MD</td>
<td>8527</td>
<td>1457</td>
<td>2.21</td>
<td>20</td>
<td>4359</td>
<td>0.38</td>
</tr>
<tr>
<td>Macedonia</td>
<td>MK</td>
<td>5578</td>
<td>908</td>
<td>2.15</td>
<td>13</td>
<td>2205</td>
<td>0.27</td>
</tr>
<tr>
<td>Northern Ireland</td>
<td>ND</td>
<td>7087</td>
<td>1888</td>
<td>2.18</td>
<td>12</td>
<td>1780</td>
<td>0.67</td>
</tr>
<tr>
<td>Netherlands</td>
<td>NL</td>
<td>14331</td>
<td>8383</td>
<td>2.68</td>
<td>20</td>
<td>3884</td>
<td>2.55</td>
</tr>
<tr>
<td>Portugal</td>
<td>PT</td>
<td>15034</td>
<td>4933</td>
<td>2.10</td>
<td>43</td>
<td>11116</td>
<td>1.49</td>
</tr>
<tr>
<td>Slovenia</td>
<td>SI</td>
<td>7801</td>
<td>2441</td>
<td>2.19</td>
<td>12</td>
<td>1525</td>
<td>1.25</td>
</tr>
<tr>
<td>Slovakia</td>
<td>SK</td>
<td>11406</td>
<td>2601</td>
<td>2.18</td>
<td>30</td>
<td>5880</td>
<td>1.04</td>
</tr>
</tbody>
</table>

Table A.1: Summary Statistics of Actual and Discretized Road Network by Country

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

70
Figure A.7: 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

Notes: All the figures pool the 868 cells across the 24 countries in the convex case of the parameters for the calibration with 10 differentiated goods. Similar relationships hold for the non-convex case. In the Panels (c) to (e), log-productivity and log-endowment of the non-traded good per capita are demeaned within each country.
Figure A.8: Actual and Predicted Trade

Between-NUTS Flow  Within-NUTS Flow

45 degree

Correlation: .76

Note: the panels show the model-based flows across level-2 NUTS regions of continental Spain against the data including all the non-zero flows in the model, across calibrations that assign a different number \( N \) of differentiated products to the largest \( N \) cities in the country. Panel (b) corresponds to the case with 10 goods that appears in the benchmark calibration of the model.
Figure A.9: Actual and Predicted Trade

(a) Import Shares vs. Distance, Model and Data

(b) Import Shares vs. Distance with FEs, model and Data

(c) Model Based Import Shares vs. Distance, Spain and France

(d) Model Based Import Shares vs. Distance with FEs, Spain and France
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>3.4%</td>
<td>1.0%</td>
<td>3.0%</td>
<td>1.3%</td>
<td>0.2%</td>
<td>1.3%</td>
</tr>
<tr>
<td>Belgium</td>
<td>1.0%</td>
<td>0.4%</td>
<td>0.8%</td>
<td>0.6%</td>
<td>0.2%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Switzerland</td>
<td>1.7%</td>
<td>0.5%</td>
<td>1.4%</td>
<td>1.0%</td>
<td>0.2%</td>
<td>1.0%</td>
</tr>
<tr>
<td>Cyprus</td>
<td>1.4%</td>
<td>0.6%</td>
<td>1.2%</td>
<td>1.0%</td>
<td>0.2%</td>
<td>1.0%</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>1.5%</td>
<td>0.5%</td>
<td>1.2%</td>
<td>1.1%</td>
<td>0.2%</td>
<td>1.1%</td>
</tr>
<tr>
<td>Germany</td>
<td>2.6%</td>
<td>0.9%</td>
<td>2.2%</td>
<td>1.9%</td>
<td>0.3%</td>
<td>1.7%</td>
</tr>
<tr>
<td>Denmark</td>
<td>8.3%</td>
<td>1.1%</td>
<td>7.8%</td>
<td>4.4%</td>
<td>0.4%</td>
<td>4.4%</td>
</tr>
<tr>
<td>Spain</td>
<td>5.6%</td>
<td>1.7%</td>
<td>4.7%</td>
<td>3.7%</td>
<td>0.5%</td>
<td>3.5%</td>
</tr>
<tr>
<td>Finland</td>
<td>5.2%</td>
<td>1.1%</td>
<td>4.7%</td>
<td>2.6%</td>
<td>0.4%</td>
<td>2.3%</td>
</tr>
<tr>
<td>France</td>
<td>4.0%</td>
<td>1.0%</td>
<td>3.4%</td>
<td>2.7%</td>
<td>0.4%</td>
<td>2.5%</td>
</tr>
<tr>
<td>Georgia</td>
<td>3.4%</td>
<td>1.4%</td>
<td>3.1%</td>
<td>2.2%</td>
<td>0.3%</td>
<td>2.2%</td>
</tr>
<tr>
<td>Hungary</td>
<td>3.0%</td>
<td>0.7%</td>
<td>2.7%</td>
<td>1.7%</td>
<td>0.2%</td>
<td>1.7%</td>
</tr>
<tr>
<td>Ireland</td>
<td>2.7%</td>
<td>1.2%</td>
<td>2.3%</td>
<td>1.8%</td>
<td>0.3%</td>
<td>1.7%</td>
</tr>
<tr>
<td>Italy</td>
<td>2.5%</td>
<td>1.5%</td>
<td>1.9%</td>
<td>1.6%</td>
<td>0.4%</td>
<td>1.3%</td>
</tr>
<tr>
<td>Lithuania</td>
<td>3.4%</td>
<td>1.0%</td>
<td>3.0%</td>
<td>2.4%</td>
<td>0.3%</td>
<td>2.3%</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>0.2%</td>
<td>0.2%</td>
<td>0.1%</td>
<td>0.2%</td>
<td>0.1%</td>
<td>0.1%</td>
</tr>
<tr>
<td>Latvia</td>
<td>3.9%</td>
<td>1.0%</td>
<td>3.5%</td>
<td>2.5%</td>
<td>0.3%</td>
<td>2.7%</td>
</tr>
<tr>
<td>Moldova</td>
<td>1.8%</td>
<td>0.5%</td>
<td>1.6%</td>
<td>1.0%</td>
<td>0.2%</td>
<td>0.9%</td>
</tr>
<tr>
<td>Macedonia</td>
<td>1.0%</td>
<td>0.5%</td>
<td>0.9%</td>
<td>0.5%</td>
<td>0.1%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Northern Ireland</td>
<td>1.2%</td>
<td>0.5%</td>
<td>1.0%</td>
<td>0.9%</td>
<td>0.1%</td>
<td>0.8%</td>
</tr>
<tr>
<td>Netherlands</td>
<td>1.6%</td>
<td>0.5%</td>
<td>1.4%</td>
<td>1.2%</td>
<td>0.2%</td>
<td>1.1%</td>
</tr>
<tr>
<td>Portugal</td>
<td>2.6%</td>
<td>0.9%</td>
<td>2.1%</td>
<td>1.4%</td>
<td>0.3%</td>
<td>1.2%</td>
</tr>
<tr>
<td>Slovenia</td>
<td>1.8%</td>
<td>0.6%</td>
<td>1.6%</td>
<td>1.4%</td>
<td>0.2%</td>
<td>1.4%</td>
</tr>
<tr>
<td>Slovakia</td>
<td>2.7%</td>
<td>1.2%</td>
<td>2.3%</td>
<td>1.9%</td>
<td>0.3%</td>
<td>2.0%</td>
</tr>
<tr>
<td>Average</td>
<td>2.8%</td>
<td>0.9%</td>
<td>2.4%</td>
<td>1.7%</td>
<td>0.3%</td>
<td>1.6%</td>
</tr>
</tbody>
</table>

Table A.2: Welfare Gains From Optimal Reallocation or Expansion of Current Networks, Fixed Labor
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>3.0%</td>
<td>1.1%</td>
<td>2.6%</td>
<td>2.1%</td>
<td>0.3%</td>
<td>2.1%</td>
</tr>
<tr>
<td>Belgium</td>
<td>0.8%</td>
<td>0.3%</td>
<td>0.7%</td>
<td>0.6%</td>
<td>0.1%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Cyprus</td>
<td>1.4%</td>
<td>0.5%</td>
<td>1.2%</td>
<td>0.9%</td>
<td>0.2%</td>
<td>0.9%</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>1.5%</td>
<td>0.7%</td>
<td>1.2%</td>
<td>0.9%</td>
<td>0.2%</td>
<td>0.9%</td>
</tr>
<tr>
<td>Denmark</td>
<td>7.0%</td>
<td>1.0%</td>
<td>6.6%</td>
<td>3.7%</td>
<td>0.3%</td>
<td>3.7%</td>
</tr>
<tr>
<td>Finland</td>
<td>4.0%</td>
<td>2.0%</td>
<td>3.4%</td>
<td>2.5%</td>
<td>0.4%</td>
<td>2.2%</td>
</tr>
<tr>
<td>France</td>
<td>3.4%</td>
<td>1.7%</td>
<td>2.6%</td>
<td>3.0%</td>
<td>0.5%</td>
<td>2.8%</td>
</tr>
<tr>
<td>Georgia</td>
<td>3.1%</td>
<td>1.6%</td>
<td>2.6%</td>
<td>2.1%</td>
<td>0.3%</td>
<td>2.2%</td>
</tr>
<tr>
<td>Germany</td>
<td>2.5%</td>
<td>0.9%</td>
<td>2.1%</td>
<td>1.7%</td>
<td>0.4%</td>
<td>1.5%</td>
</tr>
<tr>
<td>Hungary</td>
<td>2.2%</td>
<td>0.7%</td>
<td>1.9%</td>
<td>1.6%</td>
<td>0.2%</td>
<td>1.6%</td>
</tr>
<tr>
<td>Ireland</td>
<td>1.6%</td>
<td>1.9%</td>
<td>1.0%</td>
<td>1.4%</td>
<td>0.2%</td>
<td>1.3%</td>
</tr>
<tr>
<td>Italy</td>
<td>1.4%</td>
<td>0.6%</td>
<td>0.9%</td>
<td>1.3%</td>
<td>0.6%</td>
<td>1.0%</td>
</tr>
<tr>
<td>Latvia</td>
<td>3.6%</td>
<td>1.3%</td>
<td>3.2%</td>
<td>2.3%</td>
<td>0.3%</td>
<td>2.4%</td>
</tr>
<tr>
<td>Lithuania</td>
<td>3.0%</td>
<td>1.7%</td>
<td>2.5%</td>
<td>2.3%</td>
<td>0.3%</td>
<td>2.3%</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>0.2%</td>
<td>0.2%</td>
<td>0.1%</td>
<td>0.2%</td>
<td>0.1%</td>
<td>0.1%</td>
</tr>
<tr>
<td>Macedonia</td>
<td>0.5%</td>
<td>0.3%</td>
<td>0.4%</td>
<td>0.4%</td>
<td>0.1%</td>
<td>0.4%</td>
</tr>
<tr>
<td>Moldova</td>
<td>1.8%</td>
<td>0.7%</td>
<td>1.6%</td>
<td>1.0%</td>
<td>0.2%</td>
<td>0.9%</td>
</tr>
<tr>
<td>Netherlands</td>
<td>1.5%</td>
<td>0.6%</td>
<td>1.3%</td>
<td>1.1%</td>
<td>0.2%</td>
<td>1.1%</td>
</tr>
<tr>
<td>Northern Ireland</td>
<td>0.3%</td>
<td>0.6%</td>
<td>0.4%</td>
<td>0.5%</td>
<td>0.1%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Portugal</td>
<td>2.2%</td>
<td>0.9%</td>
<td>1.8%</td>
<td>1.2%</td>
<td>0.3%</td>
<td>1.1%</td>
</tr>
<tr>
<td>Slovakia</td>
<td>1.3%</td>
<td>1.4%</td>
<td>0.7%</td>
<td>1.5%</td>
<td>0.3%</td>
<td>1.5%</td>
</tr>
<tr>
<td>Slovenia</td>
<td>1.7%</td>
<td>0.8%</td>
<td>1.5%</td>
<td>1.3%</td>
<td>0.1%</td>
<td>1.4%</td>
</tr>
<tr>
<td>Spain</td>
<td>4.4%</td>
<td>2.5%</td>
<td>3.9%</td>
<td>3.6%</td>
<td>0.5%</td>
<td>3.4%</td>
</tr>
<tr>
<td>Switzerland</td>
<td>1.7%</td>
<td>0.5%</td>
<td>1.5%</td>
<td>1.0%</td>
<td>0.2%</td>
<td>0.9%</td>
</tr>
<tr>
<td>Average</td>
<td>2.2%</td>
<td>1.0%</td>
<td>1.9%</td>
<td>1.6%</td>
<td>0.3%</td>
<td>1.5%</td>
</tr>
</tbody>
</table>

Table A.3: Welfare Gains From Optimal Reallocation or Expansion of Current Networks, Mobile Labor
Table A.4: Optimal Infrastructure Investment, Population Growth and Local Characteristics for Different Number of Sectors

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=5</td>
<td>(1)</td>
<td>(2)</td>
<td>N=10</td>
<td>(3)</td>
<td>N=15</td>
<td>(4)</td>
</tr>
<tr>
<td>Population</td>
<td>0.112***</td>
<td>-0.001**</td>
<td>0.136***</td>
<td>-0.001**</td>
<td>0.131***</td>
<td>-0.001</td>
<td></td>
</tr>
<tr>
<td>Tradeable Income per Capita</td>
<td>0.178***</td>
<td>0.022***</td>
<td>0.164***</td>
<td>0.020***</td>
<td>0.186***</td>
<td>0.030***</td>
<td></td>
</tr>
<tr>
<td>Infrastructure</td>
<td>-0.211***</td>
<td>-0.001**</td>
<td>-0.242***</td>
<td>-0.001**</td>
<td>-0.245***</td>
<td>-0.001</td>
<td></td>
</tr>
<tr>
<td>Consumption per Capita</td>
<td>-0.064***</td>
<td></td>
<td>-0.062***</td>
<td></td>
<td>-0.078***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infrastructure Growth</td>
<td>-0.000</td>
<td></td>
<td>0.000</td>
<td></td>
<td>0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Differentiated Producer</td>
<td>0.001***</td>
<td></td>
<td>0.001***</td>
<td></td>
<td>0.009***</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Each column corresponds to a different regression pooling all 868 locations in the optimal expansion counterfactual across the 24 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: population growth is defined as $\Delta \ln L_j$, where $\Delta x$ denotes the difference between variable $x$ in the counterfactual and in the calibrated allocation. Investment growth is defined as the difference over the average, $\Delta I_j / (\frac{1}{2} (I_j + \Delta I_j))$, where total infrastructure at the node level defined as $I_j = \sum_{k \in \mathcal{N}(j)} I_{jk}$. Independent variables 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 calibrated model. Differentiated producer is a dummy for whether the location is a producer of differentiated goods in the calibration.

Table A.5: Average Welfare Gains Across Countries for Different Number and Allocation of Differentiated Goods

<table>
<thead>
<tr>
<th>Allocation of goods</th>
<th>Benchmark</th>
<th>Within NUTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Sectors</td>
<td>N=10</td>
<td>N=15</td>
</tr>
<tr>
<td>Labor:</td>
<td>Fixed</td>
<td>Mobile</td>
</tr>
<tr>
<td>Optimal Reallocation</td>
<td>$\delta = \delta^{I,GEO}$</td>
<td>1.6%</td>
</tr>
<tr>
<td>Optimal Expansion</td>
<td>$\delta = \delta^{I,GEO}$</td>
<td>1.7%</td>
</tr>
<tr>
<td></td>
<td>$\delta = \delta^{I,FOC}$</td>
<td>0.3%</td>
</tr>
</tbody>
</table>

Each element of the table shows the average welfare gain in the corresponding counterfactual across the 24 countries for the convex case.
Figure A.10: Discretization of the European 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 1 arc-degree (approximately 100 km) cells. Panel (b) shows the nodes $J$ corresponding to the population centroids of each cell in Panel (a), reallocated to their closest point on the actual road network, and the edges $E$ 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.
Figure A.11: TEN-TEC Core Network

The figure shows an image clipped from the TEN-T Core Network Corridors of the Trans-European Transport Network, available in http://ec.europa.eu/transport/infrastructure/tentec/tentec-portal/site/maps_upload/SchematicA0_EUcorridor_map.pdf, original figure with Copyright European Union, 1995-2019. The colors correspond to different core corridors defined by the European Commission.
Supplementary Material
(Online only)
D Supplementary Material I: Intuition behind global convexity

We propose in this section an alternative interpretation of the model that allows us to clarify the meaning of the convexity condition on the transport technology. Focusing on a single link, we show that the infrastructure investment and shipping decisions can be viewed as the outcome of a game between (1) a decentralized planning agency that chooses how much infrastructure to invest taking the price of commodities and asphalt as given and (2) a shipping company that chooses how much to ship taking infrastructure and prices as given.

With this alternative interpretation, it becomes clear that the model features a “race” between infrastructure investments and congestion. More precisely, due to the complementarity between infrastructure and trade, an investment in infrastructure in a given link calls for an increase in the flows of commodities on that same link. These additional flows encourage, in turn, further investments. We show that the convexity condition on the transport technology determines whether this race is stable or divergent. In other words, when the transport technology is convex, each additional increase in investment or flow in the sequence described above becomes smaller over time until convergence to a non-trivial interior point. In the nonconvex case, the race is unstable with explosive or collapsing dynamics.

For clarity, we focus on the single good case. The multiple good case is a straightforward extension.

Game description

We focus on a link \((o,d)\) from origin \(o\) to destination \(d\). We first consider the problem of the shipping company. While prices and infrastructure are taken as given, we assume for simplicity that the shipping company fully internalizes the congestion externality.\(^{66}\)

The problem faced by the shipping company is

\[
\max_{Q \geq 0} P_d Q - P_o Q (1 + \tau(Q, I)) ,
\]

where \(P_i, i = o, d\), is the price in each location. The first-order condition is

\[
P_d - P_o - P_o \frac{\partial (Q \tau)}{\partial Q} \leq 0 \text{ with equality if } Q > 0.
\]

\(^{(A.9)}\)

Note that this first-order condition is identical to equation (8) from the full planning problem. Denote \(Q^*(I)\) the optimal amount of commodity flow \(Q\) that solves equation \((A.9)\).\(^{67}\)

Example. For \(\tau(Q, I) = \delta^\tau Q^\beta I^{-\gamma}\), we have \(Q^*(I) = \left( \frac{1}{1+\beta} \frac{P_d-P_o}{P_o} \right)^\frac{1}{\beta} I^\frac{1}{\beta} \) for \(P_d \geq P_o, 0 \text{ otherwise}\).

Consider now the problem of a myopic decentralized agency. It is myopic in the sense that it takes the trade flow \(Q\) as given and does not internalize the fact that \(Q\) responds to infrastructure in equilibrium, and decentralized in the sense that it solely focuses on investing in link \((o,d)\) taking as given the price \(P_K\) of asphalt (that is, the Lagrange multiplier of the asphalt budget constraint in the full planning problem) and the price of commodities. The objective of the planning agency is to minimize the sum of the market value of transport costs and the cost of asphalt. We write the problem as follows

\[
\min_{I \geq 0} P_o Q \tau(Q, I) + P_K \delta^I I ,
\]

where \(\delta^I\) is the marginal cost of investing infrastructure in the link. We obtain the following first-order condition

\[
P_o \frac{\partial (Q \tau)}{\partial I} + P_K \geq 0 \text{ with equality if } I > 0.
\]

\(^{(A.10)}\)

\(^{66}\)We know from section 3 that this is equivalent to having the shipping company not internalize the congestion externality while facing an appropriately defined pigovian tax.

\(^{67}\)Under the fairly weak requirement that \(Q \tau\) is convex in \(Q\) and satisfies some Inada conditions \(\lim_{Q \to 0} \frac{\partial (Q \tau)}{\partial Q} = 0\) and \(\lim_{Q \to \infty} \frac{\partial (Q \tau)}{\partial Q} = \infty\), the solution \(Q^*(I)\) exists and is unique.
Note again that this condition is identical to equation (9) in the main text. We denote \(I^* (Q)\) the optimal infrastructure investment that solves equation \((A.10)\).

**Example.** For \(\tau (Q, I) = \delta^* Q^{\beta} I^{-\gamma}\), we have \(I^* (Q) = \left(\frac{\delta^* Q}{\gamma} P_0\right)^{\frac{1}{\gamma-1}} Q^{\frac{\beta+1}{\gamma-1}}\).

**Equilibria and stability** Figure A.12 graphically represents the two best response functions \(Q^* (I)\) and \(I^* (Q)\) for the convex and nonconvex cases using the functional form \(\tau (Q, I) = \delta^* Q^{\beta} I^{-\gamma}\). As the figure illustrates, the game admits two Nash equilibria in both cases: \(i\) (i) the null equilibrium, in which no infrastructure is invested and no commodities are traded through the link, \(ii\) an interior equilibrium with \(I > 0\) and \(Q > 0\).

The essential difference between the two cases is that the curve \(Q^* (I)\) crosses \(I^* (Q)\) from above in the convex case \((\beta > \gamma)\), while the opposite is observed in the nonconvex case \((\gamma > \beta)\). This feature implies distinct stability properties for the two equilibria in both cases:\(^{70}\)

- In the convex case, only the interior equilibrium is stable. A small increase (resp. decrease) in infrastructure prompts an increase (resp. decrease) in \(Q\), but the response is limited due to the strong congestion forces. In turn, the moderate increase in flows justifies reducing (resp. increasing) the level of infrastructure back to the interior equilibrium.

- In the nonconvex case, the interior equilibrium is no longer stable but the null equilibrium is. A small increase (resp. decrease) in infrastructure triggers a sizeable response in \(Q\) due to the weak congestion effects, which in turn justifies increasing (resp. reducing) the level of infrastructure even further. This explains why the nonconvex optimum features many links with zero infrastructures and a few links sharing all resources.

The above discussion shows that the convexity property on the transport technology controls the level of strategic complementarities between the planning agency and the private economy. In the nonconvex case, complementarities are strong: a deviation by one player invites an even greater deviation from the other. The convex case corresponds to weak complementarities: a deviation by one player is accompanied by a deviation by the other player in the same direction (they remain complements) but the response is sufficiently weak to push the economy back to the interior equilibrium.

We now formally establish the equivalence between stability of the interior equilibrium and global convexity of the function \(Q^* (Q, I)\). The interior equilibrium is stable when \(Q^* (I)\) crosses \((I^*)^{-1} (I)\) from above, that is

\[
\frac{\partial Q^*}{\partial I} \leq \left(\frac{\partial I^*}{\partial Q}\right)^{-1}.
\]

(A.11)

Differentiating the first-order conditions (A.9) and (A.10), we have \(\frac{\partial^2 Q^*}{\partial Q^2} = -\frac{\partial^2 Q^*}{\partial Q \partial I} / \frac{\partial^2 Q^*}{\partial I^2}\) and \(\frac{\partial^2 I^*}{\partial Q^2} = -\frac{\partial^2 Q^*}{\partial Q \partial I} / \frac{\partial^2 Q^*}{\partial I^2}\).

The stability condition (A.11) is thus equivalent to

\[
\frac{\partial^2 (Q^*)}{\partial Q^2} \geq -1.
\]

(A.12)

Under the assumption that \(Q^*\) is convex in \(I\) and satisfies the Inada conditions \(\lim_{I \to 0} \frac{\partial Q^*}{\partial I} = -\infty\) and \(\lim_{I \to \infty} \frac{\partial Q^*}{\partial I} = 0\), \(I^* (Q)\) exists and is unique.\(^{68}\)

The reason why the game features two equilibria even in the convex case is that the planning agency is myopic and takes the commodity flow \(Q\) as given. In the full model and in the convex case, the social planner internalizes the response in \(Q\) and never chooses the null equilibrium under the suitable Inada conditions.\(^{69}\)

Different concepts of stability may apply here. The most intuitive one, commonly referred to as “best response dynamics” in reference to the analysis of oligopolies in Cournot (1838) and Edgeworth (1925), is the one represented on the graph. Under this approach, an equilibrium is stable if it is the limit of a sequence of static games in which every players expects the other player to act as in the previous round and optimizes accordingly for initial conditions in a neighborhood of the equilibrium. More formally, our equilibria are also continuously stable in the sense of Eshel (1983).
We finally note that (A.12) is equivalent to \( \det (\text{Hess} (Q \tau)) \geq 0 \), which along with the maintained assumption that \( Q \tau \) is convex in each individual argument, \( \frac{\partial^2 (Q \tau)}{\partial Q \partial I}^2 \geq 0 \) and \( \frac{\partial^2 (Q \tau)}{\partial Q \partial I}^2 \geq 0 \), is equivalent to requiring that the total transport costs \( Q \tau (Q, I) \) is jointly convex over \( Q \) and \( I \).

**Link to the Fundamental Law of Road Congestion**  Convexity of the transport technology imposes restrictions on the elasticity of \( Q \tau \) with respect to \( Q \) and \( I \). We discuss in this subsection how this relates to the “Fundamental Law of Road Congestion” from Duranton and Turner (2011). This law is an empirical relationship that states that the number of vehicle-kilometers traveled in US cities tends to increase one for one with interstate highways. We interpret this law as meaning that \( \frac{d \log Q^* (I)}{d \log I} \approx 1 \). Differentiating the first-order condition (A.9), we have

\[
\frac{\partial^2 (Q \tau)}{(\partial Q)^2} dQ + \frac{\partial^2 (Q \tau)}{\partial Q \partial I} dI = d \left( \frac{P_d}{P_o} \right),
\]

allowing for potential general equilibrium effects \( d \left( \frac{P_d}{P_o} \right) \neq 0 \). Hence, the relative change in traded commodities with respect to a change in infrastructure is

\[
\frac{d \log Q^* (I)}{d \log I} = \frac{I}{Q^* (I)} \frac{dQ^* (I)}{dI} = \frac{I}{Q^* (I)} \frac{\partial^2 (Q \tau)}{(\partial Q)^2} + \frac{I}{Q^* (I)} \frac{\partial^2 (Q \tau)}{\partial Q \partial I} \frac{d(P_d/P_o)}{dI},
\]

where \( \varepsilon_{PE} \) is the partial equilibrium elasticity of \( Q \) with respect to infrastructure \( I \), while the second term captures a general equilibrium correction due to the potential change in relative prices at origin and destination.

**Example.** Using our preferred functional form, \( \tau (Q, I) = \delta^\alpha Q^\beta I^{-\gamma} \), the elasticity \( \varepsilon_{PE} = \gamma/\beta \), so this relationship becomes

\[
\frac{d \log Q^* (I)}{d \log I} = \frac{\gamma}{\beta} + \frac{I}{Q^* (I)} \frac{d(P_d/P_o)}{dI} \frac{\partial^2 (Q \tau)}{(\partial Q)^2}.
\]

As the above example suggests, our interpretation of the Fundamental Law depends on the strength and sign of the general equilibrium effects. On the one hand, if general equilibrium effects are weak or already controlled for in the empirical regression, then the fundamental law implies that \( \gamma/\beta \approx 1 \), which is at the margin of the convex case. On the other hand, if general equilibrium effects are non-negligible, theory does not immediately tell us in which way the correction should go. For instance, the reduction in transport costs may lead to a decline in the relative price \( P_d/P_o \). Conversely, due to the redirection of commodity flows to the link \((o, d)\) or changes in demand and supply at origin or destination, the relative price \( P_d/P_o \) may increase.
Figure A.12: Nash equilibria of the infrastructure investment game

(a) Convex case $\beta \geq \gamma$

(b) Nonconvex case $\gamma > \beta$

Note: Continuously stable equilibria indicated by a black dot, unstable equilibria by a white dot.
E Supplementary Material II: Dealing with Inefficiencies

In this section, we show how to extend our approach to deal with certain cases with externalities in which Pigovian taxes are not available or incorrectly set. We provide an example with partially internalized congestion externality in transport as well as agglomeration/congestion externalities in production and amenities.

E.1 Environment

We modify the environment by introducing agglomeration externalities or spillovers in production
\[ Y^n_j = F(L^n_j, V^n_j, X^n_j; \bar{T}_j) \]
where \( \bar{T}_j \) captures the production spillovers. \( \bar{T}_j \) is taken as given by firms but is required to satisfy in equilibrium
\[ \bar{T}_j = \sum_n L^n_j. \]
We continue assuming that \( F^n_j(\cdot) \) has constant returns to scale in its first three arguments, and is increasing and concave in each. Similarly, we allow for externalities in amenities by assuming that utility is given by
\[ U(c_j, h_j; \bar{T}_j). \]

Finally, we assume that the congestion externalities in transport are only partially internalized by shipping companies. We focus on the convex case studied in our numerical exercise with mobile labor and cross-good congestion, but the analysis can be easily applied to the other cases. Specifically, we assume that the transport technology is given by
\[ \tau^n_{jk}(Q^n_{jk}, I^n_{jk}; \bar{Q}_{jk}) = m^n \delta^n_{jk}(\bar{Q}_{jk})^{\beta-\tilde{\beta}} (Q^n_{jk})^{\tilde{\beta}} I^n_{jk}^{1-\gamma} \beta \geq \gamma, \beta \geq \tilde{\beta} \geq 0, \]
where \( Q^n_{jk} = \sum_{n=1}^N m^n Q^n_{nj} \). In this specification, the total congestion externalities amount to a total exponent of \( \beta \), but shipping companies only internalize them up to exponent \( \tilde{\beta} \), taking the aggregate flow \( \bar{Q}_{jk} \) as given.

In equilibrium, the perceived flow on a given link \( Q^n_{jk} \) is required to satisfy
\[ Q^n_{jk} = \bar{Q}_{jk}. \]
Existence and uniqueness in an environment with congestion and agglomeration externalities often requires to bound the strength of the externalities. Allen and Arkolakis (2014) propose such conditions in the special case of the Armington model with labor mobility and no congestion in transport. Providing such conditions in the general case is, however, beyond our scope for the current paper.

E.2 Fictitious planning approach

Solving for the economic allocation \( \{c^n_j, L^n_j, \ldots\} \) and trade flows \( \{Q^n_{jk}\} \) given an infrastructure network \( \{I^n_{jk}\} \) is an important step in the optimization. Following the method described in the main text, it is often convenient to solve for the allocation and flows jointly using a planning approach. Unfortunately, the presence of externalities does not allow us to invoke the First Welfare Theorem directly. We nonetheless propose a fictitious planning approach in which a “fictitious planner” does not internalize the effect of his decisions on \( T_j \) and \( \bar{Q}_{jk} \).

The fictitious planning problem for our economy can be characterized as the solution of a fixed point problem. We first define the problem faced by the fictitious planner given \( \{L_j, \bar{Q}_{jk}\} \):
\[ W \left( \{T_j, \bar{Q}_{jk}\}_{j \in J, k \in N(j)}; \{I^n_{jk}\} \right) = \max_{u, c_j, h_j, L_j, D^n_j, L^n_j, V^n_j, X^n_j, Q^n_{jk}, Q^n_{jk}} u \tag{A.13} \]
\[ \text{subject to } L^n_j, V^n_j, X^n_j, Q^n_{jk}, Q^n_{jk} \]

\[ \text{This example is akin to having congestion taxes set to a fraction } \frac{\tilde{\beta}}{\beta} \text{ of the efficient pigovian taxes relative to unit transport costs.} \]
subject to (i) availability of traded commodities,
\[ c_j L_j + \sum_{k \in N(j)} \tau_{jk} (Q_{jk}, I_{jk}; \overrightarrow{Q}_{jk}) Q_{jk} \leq D_j \left( D_1^j, \ldots, D_N^j \right) \] for all \( j \),
where \( Q_{jk} = \sum_{n=1}^{N} m^n Q_{jk}^n \), and non-traded commodities,
\[ h_j L_j \leq H_j \] for all \( j \);
(ii) the balanced-flows constraint,
\[ D_j^n + \sum_{n'} X_j^{n'n'} + \sum_{k \in N(j)} Q_{jk}^n \leq F_j^n (L_j^n, V_j^n, X_j^n, T_j) + \sum_{i \in N(j)} Q_{ij}^n \] for all \( j, n \);
(iii) local factor market clearing,
\[ \sum L_j^n \leq L_j \] for all \( j \);
and
\[ \sum V_j^{mn} \leq V_j^n \] for all \( j, m \);
(iv) free labor mobility,
\[ L_j u \leq L_j U (c_j, h_j; \overrightarrow{T}_j) \] for all \( j \);
(v) aggregate labor market clearing,
\[ \sum L_j = L \]
and (vi) non-negativity constraints on consumption, flows and factors.

For a given \( \overrightarrow{T}_j \) and \( \overrightarrow{Q}_{jk} \), the above fictitious planning problem is a well-behaved convex problem under the same conditions as in the main text and can be efficiently solved using a convex solver. Imposing that \( L_j = L_j \) and \( Q_{jk} = Q_{jk} \) in equilibrium, the first-order conditions of this planning problem are identical to that of the competitive equilibrium.\(^{72}\) We are now ready to define a solution to the fictitious planning problem.

**Definition 4.** An allocation \( \Omega = \{u, c_j, h_j, L_j, D_j^n, L_j^n, V_j^n, X_j^n, Q_{jk}, Q_{jk}^n, I_{jk} \}_{j \in J, k \in N(j), 1 \leq n \leq N} \) is a solution to the fictitious planning problem of this economy for a given infrastructure network \( \{I_{jk}\} \) if it is a solution of the problem (A.13),
\[ \Omega \in \text{argmax} \ W (\{\overrightarrow{T}_j, \overrightarrow{Q}_{jk}\}; \{I_{jk}\}) \]
and satisfies
\[ \overrightarrow{T}_j = L_j \] for all \( j \),
\[ \overrightarrow{Q}_{jk} = Q_{jk} \] for all \( j, k \in N(j) \).

A fictitious planning allocation is thus the solution of a fixed point problem over \( \overrightarrow{T}_j \) and \( \overrightarrow{Q}_{jk} \), which can be solved using a simple iterative procedure.\(^{73}\)

### E.3 Infrastructure gradient

After solving for the economic allocation and trade flows given a certain infrastructure network \( \{I_{jk}\} \), we are now ready to compute the welfare gradient with respect to infrastructure investments from the point of view of an overall
\(^{72}\)In cases where uniqueness is guaranteed, there is a perfect equivalence between the competitive equilibrium and the fictitious planning economy. In cases for which uniqueness is not guaranteed, the fictitious planning approach remains a tractable way to compute candidate equilibria.

\(^{73}\)In practice, we start with a guess on \( \{\overrightarrow{T}_j, \overrightarrow{Q}_{jk}\} \), solve problem (A.13), update our guess and repeat until convergence. We do not have any proof guaranteeing the convergence of this procedure, but our experience has shown that it converges most of the time for reasonable externality parameters.
planner that internalizes the externalities. To achieve this, we use the fact the competitive equilibrium is locally characterized as the solution to the fictitious planner’s first-order conditions. We then differentiate these first-order conditions to evaluate the local impact of infrastructure changes.

We consider the case studied in our numerical exercise from section 5 with (i) labor as the only factor of production, (ii) $D_j$ is a CES aggregator with elasticity of substitution $\sigma$, (iii) a unique good produced in each location (there is at most a unique $n(j)$ such that $\varepsilon_j^{n(j)} > 0$ for each $j$) and the production function is given by $Y_j^n = z_{jn}^l T_{nj}$ where $\varepsilon$ governs the production externality and and (iv) partially internalized transport externality ($0 \leq \beta < \beta^*$. In that case, the fictitious planning problem described in (A.13) simplifies to

$$\Omega = \left(u, c_j, h_j, L_j, D_j^n, Q_{jk}^n \right) \ u$$

subject to

$$L_j u \leq L_j U \left(c_j, h_j; T_j \right), \forall j \ [\times \omega_j] \ \text{(A.14)}$$

$$c_j L_j + \sum_{k \in N(j)} \delta_{jk} Q_{jk}^n \left( \sum_m m^n Q_{jk}^n \right)^{1+\beta} I_j^\gamma \leq \left( \sum_m \left( D_j^n \right)^{\varepsilon_{jn}(1+\beta)} \right)^{\frac{1}{\varepsilon_{jn}(1+\beta)}}, \forall j \ [\times P_j^D] \ \text{(A.15)}$$

$$h_j L_j \leq H_j, \forall j \ \ [\times P_j^H] \ \text{(A.16)}$$

$$D_j^n + \sum_{k \in N(j)} Q_{jk}^n \leq z_{jn}^l L_j T_j + \sum_{k \in N(j)} Q_{jk}^n, \forall (j, n) \ \ [\times P_j^n] \ \text{(A.17)}$$

$$\sum_j L_j = 1 \ [\times W] \ \text{(A.18)}$$

where the variables in brackets denote the associated Lagrange multipliers. Constraints (A.14)-(A.18) will bind in equilibrium and define a system of $2 + 3J + J \times N$ equations that we summarize as

$$G_{cons} (\Omega; Q_{jk}, T_j, I_{jk}) = 0.$$
Imposing $\overline{Q}_{jk} = Q_{jk}$ and $\overline{L}_j = L_j$, the constraints and first-order conditions together define the system

$$G(\Omega; Q_{jk}, I_{jk}) = \left( \begin{array}{c} G_{\text{cons}}(\Omega; Q_{jk}, I_{jk}) \\ G_{\text{FOC}}(\Omega; Q_{jk}, I_{jk}) \end{array} \right) = 0$$

of $2 + 7J + 2J \times N + 4 \times n\text{links} \times N$ equations in $2 + 4J + 2J \times N + 4 \times n\text{links} \times N$ unknowns (more precisely, the allocation $u$, $c_j$, $h_j$, $L_j$, $D^n_j$, $Q^n_{jk}$ and the multipliers $\omega_j$, $P^D_j$, $P^H_j$, $P^n_j$, $W$, $\nu^n_{jk}$, $\xi_j$).

By differentiating our equilibrium conditions around a solution to the fictitious planning problem described above, we can now compute how each element of the allocation $\Omega$ is affected by a change in infrastructure. Using the implicit function theorem, we obtain

$$J_I(\Omega) = -\left[ J_{\Omega}(G) \right]^{-1} J_I(G),$$

where $J_X(F)$ denotes the jacobian of some function $F$ with respect to some vector of variables $X$.\(^{\text{74}}\)

Since welfare $u$ is one of the variables in the allocation $\Omega$, the gradient of welfare with respect to a change in infrastructure $I_{jk}$ assuming a shadow value $P_K$ for asphalt/concrete is simply given by

$$\frac{du}{dI_{jk}} = -\delta_{jk}P_K.$$ 

Provided with this welfare gradient, it is possible to optimize on the infrastructure network using a simple gradient-descent algorithm. There are, however, two new caveats in comparison to the efficient case covered in the main text: i) the computations can be quite time-consuming since the fictitious planning problem needs to be solved at every step of the optimization, ii) Proposition 1 on the convexity of the overall planning problem no longer applies and the resulting network has no guarantee to be the global optimum.

### E.4 Numerical comparison

To quantitatively evaluate the importance of inefficiencies, we return to the benchmark case of Spain in our main quantitative exercise.\(^{\text{75}}\) We put aside agglomeration externalities in production and consumption to focus on partially internalized congestion in transport. The model parameters are those estimated in the calibration and we simply vary the perceived congestion parameter $\tilde{\beta}$ while keeping the overall level of congestion $\beta$ constant. We consider three cases $\tilde{\beta} = 0.25\beta$, $0.5\beta$ and $0.75\beta$ which correspond to having the Pigovian taxes set to 25%, 50% and 75% of what would be their optimal level relative to transport costs.

Table A.6 reports some summary statistics. Unsurprisingly, overall welfare is lower the less shipping companies internalize congestion (lower $\tilde{\beta}$). The difference is nonetheless quite small with welfare differences in the order of $10^{-4}$ in consumption equivalent. Similarly, we find that the resulting equilibrium trade flows $Q_{jk}$ and infrastructure gradients are almost identical with correlations close to 1 between the efficient and inefficient allocations.

<table>
<thead>
<tr>
<th>$\tilde{\beta}$</th>
<th>Welfare</th>
<th>corr($Q_{\text{efficient}}^{\text{jk}}$, $Q_{\text{inefficient}}^{\text{jk}}$)</th>
<th>corr($dW/dI_{\text{efficient}}$, $dW/dI_{\text{inefficient}}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficient</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.75$\beta$</td>
<td>-0.0048%</td>
<td>0.9999</td>
<td>0.9975</td>
</tr>
<tr>
<td>0.5$\beta$</td>
<td>-0.0198%</td>
<td>0.9998</td>
<td>0.9948</td>
</tr>
<tr>
<td>0.25$\beta$</td>
<td>-0.0462%</td>
<td>0.9996</td>
<td>0.9917</td>
</tr>
</tbody>
</table>

Notes: $dW/dI_{jk}$ denotes the welfare gradient. Welfare is stated in percentage deviation from the efficient benchmark in consumption equivalent.

Table A.6: Statistical comparison between efficient and inefficient allocations

---

\(^{\text{74}}\)Executing this step by hand can be cumbersome. For that purpose, we use an autodifferentiation software in our numerical exercise.

\(^{\text{75}}\)More precisely, we consider the case $\beta = 0.13$, $\gamma = 0.10$ with labor mobility and cross-good congestion.
Despite the similarities, a pattern emerges between the various allocations. Since shipping companies fail to internalize their impact on congestion, we find that the inefficient allocation tends to feature larger trade flows the less the externality is internalized in comparison to the efficient one (Figure A.13). The left panel of Figure A.15 shows that the difference in flows are especially important on large axes around Madrid, Barcelona and Valencia. The discrepancies between the welfare gradients in infrastructure shows no such systematic pattern, as Figure A.14 illustrates. The right panel of Figure A.15 shows nonetheless a tendency for the planner to invest infrastructure, in the inefficient case, in alternative routes around the most congested centers in order to alleviate congestion on the main axes. This is most easily seen in the cases of Madrid and Valencia.

Figure A.13: Comparison of trade flows $Q_{jk}$ between efficient and inefficient cases

Figure A.14: Comparison of infrastructure gradients between efficient and inefficient cases
Notes: The differences are plotted in green when higher in the inefficient allocation and in red when higher in the efficient allocation. The figure displays the case $\tilde{\beta} = 0.5\beta$.

Figure A.15: Comparison of infrastructure gradients between efficient and inefficient cases

F Supplementary Material III: Numerical evaluation of the non-convex algorithm

We propose to evaluate our simulated annealing approach in the nonconvex case in a range of random economies for which one can compute or approximate the global optimum.

F.1 Brute force approach

An important caveat when running this performance evaluation is that infrastructure investments are continuous, so that the space of feasible networks is infinite. There is, unfortunately, no readily available brute force algorithm that guarantees to find the global optimum in that case. To circumvent this issue, we propose to use a brute force algorithm to explore every single discrete combination of bilateral links (on/off), which we then use as initial conditions to iterate on the first-order conditions that characterize optimal infrastructure investment in our model. While this approach offers no guarantee to find the global optimum 100% of the time, we view it as our best attempt at finding it.\footnote{Thanks to its extensive search in the space of initial conditions, this algorithm should visit most local optima. Our simulations comforted this view since none of the other algorithms that we tried ever defeated this brute force approach.}

Brute force can, however, only be used for small economies with a limited number of locations. For instance, an economy with $n$ locations has $n(n - 1)/2$ possible links after imposing symmetry and ruling out links from a node to itself. This means that there are $2^{n(n-1)/2}$ potential combinations to explore. As Table A.7 illustrates, the number of combinations explodes rapidly, so the scope for using brute force is limited.

F.2 Alder (2019)

In this evaluation, we also compare the performance of our algorithm to that described in Alder (2019), who proposes another heuristic algorithm to find the globally optimal network in a spatial economic model. Since the
Table A.7: Number of network configurations to explore

<table>
<thead>
<tr>
<th>Nb. of locations $n$</th>
<th>Nb. of networks</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>64</td>
</tr>
<tr>
<td>5</td>
<td>1,024</td>
</tr>
<tr>
<td>6</td>
<td>32,768</td>
</tr>
<tr>
<td>7</td>
<td>2,097,152</td>
</tr>
<tr>
<td>8</td>
<td>268,435,456</td>
</tr>
</tbody>
</table>

Alder algorithm is discrete in nature (locations are connected or not, and there is no intensive aspect to infrastructure investments), there is no unique and obvious way to perform the comparison. We propose two approaches.

In the first approach, which we refer to as “Alder”, we implement a simple version of the Alder (2019) algorithm in our model. Specifically, we restrict infrastructure investments to be discrete (0 or $I$), where $I$ is uniform across links and chosen such that the asphalt budget constraint is satisfied with equality. Under this constraint, we implement every step of the Alder algorithm, that is: 1) starting from full network, 2) compute welfare by removing each link one by one, 3) remove the 5% least beneficial links, 4) add the most beneficial link, 5) make sure that the network is connected and iterate over (2)-(5) until no further step improves welfare. As a final step, we use the outcome of the Alder algorithm as initial condition in the first-order condition iteration from our model, making sure that welfare keeps improving along the way. We obtain a continuous version of the optimal network that we can directly compare to ours.

A caveat in the first approach is that the first part of the algorithm is heavily constrained by the discreteness of investments, which distorts the search for the optimum. We thus devise a new algorithm that combines aspects from Alder (2019) with appealing features from our continuous-investment model, such as the use of differentiation or the ability of the simulated annealing approach to escape local optima. We refer to it as “Hybrid Alder-FS”. To be more precise, we modify our simulated annealing algorithm so that the stage in which we perturb the network is done in the spirit of the Alder algorithm. Specifically, the network is perturbed in the following way: 1) compute the welfare gradient with respect to infrastructure investment, 2) remove the 5% links that have to the lowest element in the welfare gradient, 3) recompute the welfare and welfare gradient in the network after removal, 4) add the most beneficial link that has the highest element in the welfare gradient among nonexisting links, 5) make sure that the network is connected. The rest of the algorithm is kept identical to ours. In particular, the algorithm always iterates on the first-order conditions associated to infrastructure investment before deciding whether to accept the perturbation or not. In that sense, the intensive dimension of infrastructure investment is preserved throughout and the algorithm is not artificially constrained by the discreteness assumption.

Note finally that the original Alder (2019) algorithm is designed to maximize aggregate income net of the dollar value of infrastructure investments. In this comparison, we change the objective function so that it maximizes welfare instead subject to a fixed asphalt budget.

F.3 Simulated annealing

There are many ways one can perturb the network in each loop of the simulated annealing method. To explore the role of the perturbation stage, we experiment with various ways to perturb the network: 1) purely random perturbations, 2) deterministic rebranching, and 3) random rebranching. “Rebranching” refers to the algorithm that we describe in Appendix A.4. In the deterministic version, all nodes are selected for rebranching every loop (i.e., reconnected to their best neighbors), while they are picked at random in the random version. This comparison allows us to evaluate the importance of randomness in the perturbation process as a way to escape local optima.
F.4 Small random economies

To perform the algorithm comparison, we draw random geographies for a given number of replications $N_{\text{reps}}$. Specifically, for each geography, we draw $n$ locations from a uniform distribution over $[0, 1] \times [0, 1]$. The underlying network is assumed to be fully connected. We parametrize the transport cost parameters $\delta_{jk}$ and network building costs $\delta_{jk}$ to be equal to the euclidean distance between each pair of locations. To ease comparison with Alder (2019) who uses a gravity model without labor mobility, we study an Armington version of our model in which each location produces its variety with fixed labor. The rest of the model is otherwise identical to the one described in section 3. In particular, there is congestion in the transport technology, that is $\beta > 0$ using technology (10). We focus exclusively on nonconvex cases with $\gamma > \beta$.\footnote{In the convex case $\beta \geq \gamma$, a simple iteration over our model’s first-order conditions or descent along the gradient suffice to find the global optimum quickly, so the algorithm comparison is irrelevant.}

Table A.8 below reports the performance of the various numerical approaches for several numbers of locations, replications and allowing for uniform or random fundamentals. Due to the computation time in the brute force algorithm, we must restrict ourselves to 6 locations at most.

Before starting the individual comparison of each algorithm, it should be noted that the welfare losses are overall quite small with numbers in the magnitude of $10^{-4}$ in consumption equivalent. This suggests that, all things considered, each of these algorithms perform quite well for small economies. Looking at their individual performance, we find that the simple Alder algorithm is the least effective method with a welfare loss of about 0.02-0.03% in consumption equivalent compared to the brute force algorithm in most cases. A potential reason, that we posited before, is that the discreteness constraint heavily distorts the search for the optimum. Turning to the other algorithms, our findings suggest that the random version of our “rebranching” algorithm and the Hybrid Alder-FS algorithm get the best results with average welfare losses of about 0.02% or less. The Hybrid Alder-FS algorithm does slightly better in our simulations with 5 locations, but the random rebranching algorithm has the edge in the 6 locations case. In terms of computation time, the Hybrid algorithm requires slightly more time than the FS algorithm due to an additional evaluation of the economic equilibrium during the updating process. The Alder algorithm is the fastest algorithm given the small amount of locations, but the difference will quickly revert when we increase their number, as we will shortly see.

F.5 Large random economies

A potential concern in our performance evaluation so far is that we have focused on extremely small economies with limited spatial complexity. We now try to extend our comparison to economies with more locations. Unfortunately, we can no longer use the brute force algorithm, but we can still compare the performance of the Alder and the various simulated annealing algorithms. In this last exercise, we focus on the version of our model that we used for the European road exercise in section 5. Namely, we consider $w \times h$ rectangular networks in which all location are connected to their immediate neighbors along the horizontal, vertical and diagonals. Since we study economies with a large number of locations, we limit the number of traded goods to 2. As we did in the European exercise, we assume that one of the goods is a homogeneous “agricultural” good that can be produced in any location. The other good is a differentiated variety that can only be produced in a number of locations $N_{\text{cities}}$ drawn uniformly among all locations. The rest of the model is otherwise identical as in the previous section.

The results of this exercise, in Table A.9 below, are reported in terms of the welfare gains with respect to the local optimum that results from iterating over our model’s first-order conditions. Our findings suggest this time larger differences between the algorithms as spatial complexity increases. While the Hybrid Alder-FS algorithm seems to perform quite well overall (and definitely better than random perturbations), our “rebranching” algorithm, in its deterministic and random versions, yields the best results in all cases. The random version performs the best with welfare improvements ranging between 0.01% and 0.1%. The simple (discrete) Alder algorithm yields the worst results in this setup, most likely for the reason advanced before.
Table A.8: Comparison of the various algorithms for small economies

<table>
<thead>
<tr>
<th>#locations</th>
<th>$N_{\text{reps}}$</th>
<th>Alder</th>
<th>Random</th>
<th>Rebranching (deterministic)</th>
<th>Rebranching (random)</th>
<th>Hybrid Alder-FS</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>100</td>
<td>-0.0360%</td>
<td>-0.0036%</td>
<td>-0.0036%</td>
<td>-0.0020%</td>
<td>-0.0020%</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>-0.1281%</td>
<td>-0.0046%</td>
<td>-0.0047%</td>
<td>-0.0043%</td>
<td>-0.0032%</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>-0.0216%</td>
<td>-0.0016%</td>
<td>-0.0019%</td>
<td>0.0004%</td>
<td>-0.0019%</td>
</tr>
</tbody>
</table>

Notes: This table reports welfare in percentage consumption equivalent compared to the brute force result using the network produced by each optimization method. Time is the computation time for a single replication. The parameters chosen for the simulation are $\beta = 1$, $\gamma = 2$, $K = 1$, $\sigma = 5$, uniform amenities, productivity and population, no cross-good congestion, Cobb-Douglas utility with share $\alpha = 0.5$ on traded goods.
Table A.9: Comparison of the various algorithms for large economies

<table>
<thead>
<tr>
<th>#locations</th>
<th>N cities</th>
<th>N reps</th>
<th>Alder</th>
<th>Random</th>
<th>Rebranching (deterministic)</th>
<th>Rebranching (random)</th>
<th>Hybrid Alder-FS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>10</td>
<td>100</td>
<td>-0.0272%</td>
<td>0.0000%</td>
<td>0.0122%</td>
<td>0.0153%</td>
<td>0.0035%</td>
</tr>
<tr>
<td>64</td>
<td>20</td>
<td>100</td>
<td>-0.0133%</td>
<td>0.0000%</td>
<td>0.0378%</td>
<td>0.0431%</td>
<td>0.0017%</td>
</tr>
<tr>
<td>64</td>
<td>40</td>
<td>100</td>
<td>-0.0068%</td>
<td>0.0003%</td>
<td>0.0057%</td>
<td>0.0061%</td>
<td>0.0012%</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>100</td>
<td>-0.0033%</td>
<td>0.0011%</td>
<td>0.0858%</td>
<td>0.0954%</td>
<td>0.0024%</td>
</tr>
<tr>
<td>100</td>
<td>40</td>
<td>100</td>
<td>-0.0062%</td>
<td>0.0015%</td>
<td>0.0259%</td>
<td>0.0303%</td>
<td>0.0018%</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>100</td>
<td>-0.0048%</td>
<td>0.0005%</td>
<td>0.0042%</td>
<td>0.0061%</td>
<td>0.0008%</td>
</tr>
</tbody>
</table>

Notes: Welfare is reported in percentage consumption equivalent compared to the outcome of our model without simulated annealing, in other words the benchmark is obtained by iterating over the first-order conditions of our model starting from a full network. The parameters chosen for the simulation are $\beta = 1$, $\gamma = 2$, $K = 1$, $\sigma = 5$, uniform amenities, productivity and population, no cross-good congestion, Cobb-Douglas utility with share $\alpha = 0.5$ on traded goods.
G Supplementary Material IV: Proposition 3

G.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 $1 \leq k, l \leq n - 1$ 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.

G.2 Propositions and Lemmas

Proposition 6. $E^*$ 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, $E^*$ must be connected. It remains to show that $E^*$ cannot have simple cycles. We proceed by contradiction. Assume there exists a simple cycle $p = (i_1, \ldots, i_n)$. Figure A.16 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 $(c^*, Q^*)$ 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 $(c^*, Q^*)$ 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 $E^*$ is a tree.

![Figure A.16: Different types of simple cycles](image)

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

Proof. Case (i) in Figure A.16 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

$$Q_{jk}^* = \begin{cases} Q_{jk}^* - \varepsilon & \text{if } \exists l, 1 \leq l \leq n - 1, \text{ such that } j = i_l \text{ and } k = i_{l+1} \\ Q_{jk}^* & \text{elsewhere} \end{cases}$$
If \( \varepsilon \leq \min_{1 \leq k \leq n-1} Q_{ik,i_{k+1}} \), then \((c_j, Q'_j)\) 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 \( \varepsilon \) is strictly greater than 0 (recall that \( P_j = u_\varepsilon (c_j, h_j) > 0 \)), contradicting the assumption that \((e^*, Q^*)\) is a local optimum.

\[ \square \]

**Lemma 2.** For every node \( a \in I \) distinct from the productive center \( 0 \in I \) 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 \( Q_{ik,i_{k+1}} > 0 \) for \( 1 \leq k \leq n - 1 \).

**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 \( 1 \leq k, l \leq n \) and \( k \neq l \) and such that \( Q_{ik,i_{k-1}} > 0 \). 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,

\[
c_a L_a = \sum_{k \in N(a)} Q_{ka} - \sum_{k \in N(a)} Q_{ak} \left[ 1 + \delta_{ak} \frac{Q_{ka}^0}{P_{ak}} \right] > 0,
\]

tells us that location \( a \) must be a net recipient of goods from its neighbors. Hence, there exists \( k \in N(a) \) 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 \( i_1 = a, i_2 \neq i_1 \neq 0 \) and \( Q_{i_1,i_2} > 0 \). Assume now that \( n \geq 2 \) 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 \( 1 \leq k \neq l \leq n \) and \( k \neq l \) such that \( Q_{ik,i_{k-1}} > 0 \). Consider location \( i_n \). The balanced flow constraint at \( i_n \) tells us that

\[
c_{i_n} L_{i_n} = \sum_{k \in N(i_n)} Q_{ki_n} - \sum_{k \in N(i_n)} Q_{i_n k} \left[ 1 + \delta_{i_n k} \frac{Q_{i_n k}^0}{P_{i_n k}} \right] > 0.
\]

Since we know by recursion hypothesis that \( Q_{i_n,i_{n-1}} > 0 \), then there exists a \( k \in N(i_n) \) such that \( Q_{ki_n} > 0 \). We know that \( k \neq i_l \) for all \( 1 \leq l \leq n \) because otherwise there would exist 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 \( 1 \leq k \neq l \leq n+1 \) and \( k \neq l \) such that \( Q_{i_k,i_{k-1}} > 0 \).

We conclude as follows. Since \( I \) 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 exist a simple path \( p \) of size \( n < |I| \) with \( p = (i_1, \ldots, i_n) \) such that \( i_1 = a, i_n = 0 \) and \( Q_{i_k,i_{k+1}} > 0 \) for \( 1 \leq k \leq n - 1 \). By construction, the path \( \tilde{p} = (i_n, i_{n-1}, \ldots, i_1) \) proves the statement.

\[ \square \]

**Lemma 3.** Assume there exists a simple cycle \( p = (i_1, \ldots, i_n) \). Then, there exists \( (a,b) \in I^2 \), \( a \neq b \), such that there exists two distinct simple paths from \( a \) to \( b \), \( p_1 = (i_{k_1}^1)_{1 \leq k \leq n_1} \) and \( p_2 = (i_{k_2}^2)_{1 \leq k \leq n_2} \) with \( i_1^1 = i_1^2 = a \) and \( i_{n_1}^1 = i_{n_2}^2 = b \), such that the flows are strictly positive from \( a \) to \( b \) along both paths, i.e., \( Q_{i_k,i_{k+1}}^1 > 0 \) 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.16.

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 \( \hat{Q}_{ik,i_{k-1}} = Q'_{ik,i_{k-1}} - Q'_{i_{k+1},ik} \) the net flow from \( ik \) to \( i_{k+1} \) for \( 0 \leq k \leq n \), which can be either strictly positive or strictly negative. We know from Lemma 1 that the net flows \( \hat{Q}_{ik,i_{k+1}} \) cannot have the same sign, otherwise we would have a cycle with circular flow, violating the local optimality condition of \((e^*, Q^*)\). Hence, there must exist \( 1 \leq k \leq n \) such that \( \hat{Q}_{ik,i_{k+1}} > 0 \) and \( \hat{Q}_{ik,i_{k+1}} < 0 \). 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.16. Set \( a = 0 \) and \( b = i_k \). We know from Lemma 2 that there exists a path \( p_1 = (j_1^1, \ldots, j_{n_1}^1) \) such that \( j_1^1 = a = 0, j_{n_1}^1 = i_{k-1} \) and \( Q'_{j_1^1,j_{n_1}^1} > 0 \) for all \( 1 \leq l \leq n_1 \). Similarly, there exists a path \( p_2 = (j_1^2, \ldots, j_{n_2}^2) \) such that \( j_1^2 = a = 0, j_{n_2}^2 = i_{k+1} \) and \( Q'_{j_1^2,j_{n_2}^2} > 0 \) for all \( 1 \leq l \leq n_2 \). We now argue
that the paths $\tilde{p}_1 = (j_1^1, \ldots, j_{n_1}^1, b)$ and $\tilde{p}_2 = (j_1^2, \ldots, j_{n_2}^2, b)$ are two distinct simple paths from $a$ to $b$ with strictly positive flows. By construction, we know that $\tilde{Q}_{j_1^1, i_k} > 0$ and $\tilde{Q}_{j_1^2, i_k} > 0$ 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 $1 \leq l \leq n_1$ such that $j_l^1 = b$. If this was the case, then $(b, j^1_{l+1}, \ldots, j^1_{n_1}, b)$ 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., $p_1 = (i_1^1)_{1 \leq k \leq n_1}$ and $p_2 = (i_1^2)_{1 \leq k \leq n_2}$ with $i_1^1 = i_1^2 = a$ and $i_{n_1} = i_{n_2} = b$, such that $Q_{i_k, i_{k+1}} > 0$ 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.16 cannot exist. Assume by contradiction that such a cycle exists and that $p_1 \neq p_2$. Note that we can assume WLOG that $i_k^1 \neq i_k^2$ for all $1 < k < n_1$ and $1 < l < n_2$. To see this, let $\tilde{k} = \min \{ |k| i_k^1 \neq i_k^2 \}$ and $\tilde{k}_1 = \min \{ |k| l \tilde{k} > \tilde{k}_2, i_k^1 = i_k^2 \}$ and $\tilde{k}_2$ be such that $i_{\tilde{k}_1} = i_{\tilde{k}_2}$. By construction, the path $p_1 = (i_{\tilde{k}_1}^1, \ldots, i_{\tilde{k}_2}^1)$ and $p_2 = (i_{\tilde{k}_1}^2, \ldots, i_{\tilde{k}_2}^2)$ are two paths such that $i_k^1 = i_k^2$, $i_{\tilde{k}_1}^1 = i_{\tilde{k}_1}^2$, and $i_k^1 \neq i_k^2$ for all $k < \tilde{k}_1 < \tilde{k}_2$ and $k < l < \tilde{k}_2$.

![Figure A.17: Redirecting the flows to one branch](image)

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.17 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 $(e^*, Q^*)$. Consider the allocation $Q^\varepsilon = \{Q^\varepsilon_{jk}\}$ for $\varepsilon \in \mathbb{R}$ such that

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

In other words, $Q^\varepsilon_{jk}$ corresponds to the pattern of flows $Q_{jk}^*$ where a volume $\varepsilon$ of the flows going through path 2 are redirected through path 1. By construction, $Q^\varepsilon_{jk}$ 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.5) unchanged except through the term $\left[ \sum_{k=1}^{n_1} \mathcal{R}_{\tilde{k}+1} \left( \sum_{i_k^1} \mathcal{P}_{\tilde{k}+1} \tilde{k} \delta_{i_k^1} \tilde{k} \mathcal{Q}_{i_k^1, i_{k+1}^1} \right) \right]$ where $\delta_{i_k^1} \tilde{k} = \delta_{i_k^1} / \delta_{i_k^1} \tilde{k}$.

Consider the derivative of the Lagrangian with respect to $\varepsilon$:

$$\frac{\partial \mathcal{L}^\varepsilon}{\partial \varepsilon} = -(1 + \beta) \left[ \sum_{j,k} \delta_{j,k}^\varepsilon \left( \mathcal{P}_{j,k} Q_{j,k}^{1+\beta} \right) \right] \mathcal{R}_{\tilde{k}+1} \left[ \sum_{1 \leq k \leq n_1-1} \delta_{i_k^1} \tilde{k} \mathcal{P}_{i_k^1} \tilde{k} \mathcal{Q}_{i_k^1, i_{k+1}^1} - \sum_{1 \leq k \leq n_2-1} \delta_{i_k^1} \tilde{k} \mathcal{P}_{i_k^1} \tilde{k} \mathcal{Q}_{i_k^1, i_{k+1}^1} \right]$$

96
which satisfies \( \frac{\partial L}{\partial \varepsilon} = 0 \) by assumption (local optimum). Let us examine the second order condition:

\[
\frac{\partial^2 L}{\partial \varepsilon^2} = -(1 + \beta) \left( \frac{1 + \beta}{1 + \gamma} - 1 \right) \left[ \sum_{j,k} \delta_{jk}^{\gamma+1} \left( P_j Q_{jk}^{1+\gamma} \right)^{\frac{1}{\gamma+1}} \right]^\gamma 
\times 
\left[ \sum_{1 \leq k \leq n_1 - 1} \delta_{jk}^{\frac{\gamma}{\gamma+1}} P_{jk}^{\frac{1+\beta}{1+\gamma}} Q_{jk}^{\frac{1+\gamma+\beta}{1+\gamma}} - \sum_{1 \leq k \leq n_2 - 1} \delta_{jk}^{\frac{\gamma}{\gamma+1}} P_{jk}^{\frac{1+\beta}{1+\gamma}} Q_{jk}^{\frac{1+\gamma+\beta}{1+\gamma}} \right] 
- (1 + \beta)^2 \frac{\gamma}{\gamma+1} \left[ \sum_{j,k} \delta_{jk}^{\frac{\gamma}{\gamma+1}} \left( P_j Q_{jk}^{1+\gamma} \right)^{\frac{1}{\gamma+1}} \right]^{\gamma-1} 
\times 
\left[ \sum_{1 \leq k \leq n_1 - 1} \delta_{jk}^{\frac{\gamma}{\gamma+1}} P_{jk}^{\frac{1+\beta}{1+\gamma}} Q_{jk}^{\frac{1+\gamma+\beta}{1+\gamma}} - \sum_{1 \leq k \leq n_2 - 1} \delta_{jk}^{\frac{\gamma}{\gamma+1}} P_{jk}^{\frac{1+\beta}{1+\gamma}} Q_{jk}^{\frac{1+\gamma+\beta}{1+\gamma}} \right]^{2}.
\]

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

\[\Box\]