Econometrica, Vol. 88, No. 4 (July, 2020), 1411–1452

OPTIMAL TRANSPORT NETWORKS IN SPATIAL EQUILIBRIUM

PABLO D. FAJGELBAUM
UCLA and NBER

EDOUARD SCHAAL
CREI, Universitat Pompeu Fabra, Barcelona GSE, and CEPR

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 the road networks of European countries.

KEYWORDS: Transport network, spatial equilibrium, economic geography.

1. INTRODUCTION

TRADE COSTS ARE a ubiquitous force in international trade and economic geography, as they shape the spatial distributions of prices, real incomes, and trade flows. Transport infrastructure is a key determinant of trade costs.[1] Governments, international organizations, and private companies routinely invest large amounts of resources to improve transport networks within and across countries. How should these investments be allocated? Are observed transport networks suboptimal, and if so, how important are these inefficiencies?

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.

Pablo D. Fajgelbaum: pfajgelbaum@econ.ucla.edu
Edouard Schaal: eschaal@crei.cat

We thank four anonymous referees. We thank Costas Arkolakis, Dave Donaldson, and Jonathan Eaton for their discussions. We thank Thomas Chaney, Arnaud Costinot, Don Davis, Manuel Garcia Santana, Cecile Gaubert, Gueorgui Kambourov, Carlos Llano, Ezra Oberfield, Stephen Redding, Esteban Rossi-Hansberg, and Jonathan Vogel for helpful comments. We also thank Dorian Henricot and Cristiano Mantovani for superb research assistance. 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]See Limao and Venables (2001) and Atkin and Donaldson (2015). For a review of various determinants of trade costs, see Anderson and Van Wincoop (2004).

 https://doi.org/10.3982/ECTA15213

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 either for a fixed spatial distribution of the primary factors (as in international trade models) or for labor to be perfectly 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 optimize in the space of equilibrium prices applying the numerical methods typically used for optimal transport problems.

While strong enough congestion in transport guarantees the convexity of the planner's problem and enables the use of a duality approach, our framework can also be used when congestion is weak or absent—a case that implies increasing returns in the transport technology. We numerically approximate the global solution in non-convex cases by combining the duality approach to obtain the optimal flows 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 one would expect in the presence of economies of scale, such as more concentration in fewer links and a larger amount of zeros.

The framework has enough flexibility to be matched to data on actual transport networks. 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 assuming Pigouvian taxes to correct congestion externalities. 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. We calibrate the productivity and the endowment of non-traded goods such that the model reproduces the

observed population and value added at a high spatial resolution separately for each of the 24 countries in our data. We construct a measure of the road infrastructure linking any two contiguous cells in the data and entertain different assumptions on labor mobility and on the returns to infrastructure, encompassing both convex and non-convex cases. 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 for how building costs vary with observable geographic features.

Our counterfactuals in the benchmark parameterization with convex costs imply that, across countries, the average welfare gain from an optimal 50% expansion in the observed road networks and the average welfare loss from road misallocation are on average 2% and range between 0.1% and 7%. The optimal expansion or reallocation of roads reduces regional inequalities in real consumption, reflecting that optimal infrastructure investments reduce dispersion in the marginal utility of consumption of traded commodities. 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. We conclude with an exercise involving multiple countries in Western Europe, which highlights the importance of trade across borders and international coordination in infrastructure policy.

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 Supplemental Material Appendix (Fajgelbaum and Schaal (2020)).

2. RELATION TO THE LITERATURE

A quantitative literature in international trade and spatial economics 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 multi-country competitive equilibrium. A standard approach to study the gains from market integration is to fit these models 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] We develop a different approach to implementing counterfactuals. We first pinpoint the best set of infrastructure investments in a transport network, and then ask what would happen if trade costs were to change in the way implied by the efficient transport network.

Recent studies undertake counterfactuals with respect to the cost of shipping across specific links in models where traders choose least-cost routes to ship goods.[3] Allen and Arkolakis (2014) measured the aggregate effect of the U.S. highway system, Donaldson

[2]Costinot and Rodríguez-Clare (2013) reviewed the quantitative gravity literature on changes in trade costs focused on measuring gains from international trade. Redding and Rossi-Hansberg (2017) reviewed 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]Chaney (2014a) studied 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).

and Hornbeck (2016) calculated the historical impact of railroads on the U.S. economy, and Redding (2016) compared the impact of infrastructure changes in models with varying degrees of increasing returns. Alder (2019) simulated counterfactual transport networks in India, Nagy (2016) studied how the development of U.S. railways affected city formation, and Sotelo (2016) simulated 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, Parro, Rossi-Hansberg, and Sarte (2017), and Ramondo, Rodríguez-Clare, and Saborío-Rodríguez (2016).

Some papers feature an optimization over transport networks. Alder (2019) applied a heuristic algorithm that adds or removes links in a specific order based on their contribution to net aggregate income.[4] Felbermayr and Tarasov (2015) studied optimal infrastructure investments by competing planners in an Armington model where locations are arranged on a line. Allen and Arkolakis (2019) computed the welfare gradient with respect to reductions in the cost of shipping across specific links in an Armington model with spillovers.[5] Their approach is suitable to compute the first-order welfare impact of infrastructure investments around an initially observed allocation.

We solve instead a global optimization over the space of networks in a neoclassical framework. Both our model and the studies cited above include an optimal transport problem, defined as the trader's problem of choosing least-cost routes across pairs of locations.[6] In most of the studies cited above, the optimal transport problem does not include congestion and can therefore be solved independently from the general equilibrium. In our context, congestion in transport renders the infrastructure investment problem convex, enabling the search for the global optimum. The least-cost route optimization from the applications of the gravity trade models discussed before corresponds to the solution of our optimal transport problem in the special case without congestion.

As mentioned in the Introduction, the planner's subproblem of choosing how to ship goods given demand, supply, and infrastructure formally defines an optimal transport problem. Optimal transport problems were studied early on by Monge (1781) and Kantorovich (1942).[7] Because we analyze the optimal route problem instead of the direct assignment of sources to destinations, our approach is more closely related to optimal flow problems on a network as in Chapter 8 of Galichon (2016).[8] Our problem differs from this literature in two important aspects. First, in our model, consumption and production are endogenous because they respond to standard general-equilibrium forces. Instead, the aforementioned optimal flow problems map sources with fixed supply to sinks with fixed demand.[9] Second, our ultimate focus is on the 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, Caselles, and Morel (2009).

[4]See Section B of the Supplementary Material for a comparison with the Alder (2019) algorithm.

[5]Swisher IV (2015) and Trew (2016) allowed for endogenous transport costs in different historical contexts.

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

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

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

[9]See Beckmann (1952) for an early continuous-space example of such an optimal transport problem in economics. See Carlier (2010) and Ekeland (2010) for lecture notes on optimal transport and its connection to economics.

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, 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.[10]

A large body of research estimates how actual changes in transport costs impact economic activity. For instance, Fernald (1999) estimated the impact of road expansion on productivity across U.S. industries; Chandra and Thompson (2000), Baum-Snow (2007), and Duranton, Morrow, and Turner (2014) estimated the impact of the U.S. highways on various regional economic outcomes; Donaldson (2018) estimated the impact of access to railways in India; and Faber (2014) estimated the impact of connecting regions to the expressway system in China.[11] Our application measures the aggregate country-level welfare gains from optimally expanding current road networks in European countries. In the counterfactuals, we inspect the relationship between optimal infrastructure investment and population growth across regions.

As we apply the model to measure the potential losses from misallocation of roads, the paper is broadly related to studies of the aggregate effects of misallocation such as Restuccia and Rogerson (2008) and Hsieh and Klenow (2009). Desmet and Rossi-Hansberg (2013), Brandt, Tombe, and Zhu (2013), Asturias, García-Santana, and Ramos (2019), Fajgelbaum, Morales, Suárez Serrato, and Zidar (2018), and Hsieh and Moretti (2019), among others, focused on geographical misallocation arising from frictions or spatial policies.

3. MODEL

3.1. *Environment*

Preferences

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

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

where the utility function U is homothetic and concave in both of its arguments. In location j, per capita consumption of traded goods is $c_j = C_j/L_j$, where C_j is the aggregate demand of the traded goods bundle in location j.

[10] A network-design and planning literature in operations research studies related network-design problems in telecommunications and transport without embedding them in general-equilibrium spatial models. See Ahuja, Magnanti, and Orlin (1989).

[11] See also Coşar and Demir (2016) and Martincus, Carballo, and Cusolito (2017) for empirical studies of how infrastructure investments impact international shipments.

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 (e.g., a CES aggregator),

$$C_j = D_j\big(D_j^1, \ldots, D_j^N\big), \tag{2}$$

where D_j^n is sector n's output used in location j.

Production

The supply side corresponds to a standard neoclassical economy. In addition to labor, there is a fixed supply $\mathbf{V}_j = (V_j^1, \ldots, V_j^M)'$ 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

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

where L_j^n is the number of workers, $\mathbf{V}_j^n = (V_j^{1n}, \ldots, V_j^{Mn})'$ is the quantity of other primary factors, and $\mathbf{X}_j^n = (X_j^{1n}, \ldots, X_j^{Nn})$ is the quantity of each sector's output allocated to the production of sector n in location j. The production function F_j^n is either neoclassical (constant returns to scale, increasing and concave in all its arguments) or a constant (endowment economy).

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 only be shipped through connected locations; that is, 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 interpretation is that j is a geographic unit such as county, $\mathcal{N}(j)$ are its bordering counties, and shipments are done by land. More generally, the neighbors in the model 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 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. We adopt two alternative specifications for transport costs: iceberg costs without congestion across commodities, and a case with cross-goods congestion. For simplicity of exposition, we discuss the former case here. In Section 3.7, we discuss congestion across goods, which will also be the benchmark specification in our applications. Appendix A of the Supplemental Material (Fajgelbaum and Schaal (2020)) presents the model in a general case encompassing both formulations.

Transporting each unit of good n from j to k requires τ_{jk}^n units of the good n itself, so that $1 + \tau_{jk}^n$ corresponds to the iceberg cost. 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}:

$$\tau_{jk}^n = \tau_{jk}\left(Q_{jk}^n, I_{jk}\right). \tag{4}$$

The per-unit cost of shipping is increasing in the quantity of commodities shipped:

$$\frac{\partial \tau_{jk}(Q, I)}{\partial Q} \geq 0. \tag{5}$$

This assumption allows for decreasing returns in the shipping sector. We refer to these decreasing returns as congestion, with the understanding that this concept encapsulates several real-world forces whereby an increase in shipping activity leads to higher marginal transport costs. These forces may include higher travel times or road damage, as well as decreasing returns to scale in transportation due to land-intensive fixed factors such as warehousing or specialized inputs. In short, the more that is shipped, the higher the per-unit shipping cost.[12]

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

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

The transport technology $\tau_{jk}(\cdot)$ is allowed to vary by jk, capturing variation in shipping costs across links 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; for example, if elevation is higher in j than k and it is cheaper to drive downhill, then $\tau_{jk}(Q, I) < \tau_{kj}(Q, I)$.

Flow Constraint

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

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

The left-hand side of this inequality is location j's consumption D_j^n of good n, intermediate-input use $X_j^{nn'}$ by each sector n', exports to neighbors Q_{jk}^n, and inputs to the transport sector $\tau_{jk}^n Q_{jk}^n$. These flows are bounded by the local production Y_j^n and imports from neighbors Q_{ij}^n. In standard minimum-cost flow problems, this restriction is known as the conservation of flows constraint.

[12]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 studied road damage externalities, whereby the road damage caused by one vehicle increases the operating costs of subsequent vehicles. Maibach et al. (2013) listed 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 let P_j^n be the multiplier of this constraint. This multiplier reflects society's valuation of a marginal unit of good n in location j. In the decentralized allocation, this multiplier will equal the price of good n in location j. Therefore, we refer to P_j^n as the price of good n in location j.

Network-Building Technology

We define the transport network as the distribution of infrastructure $\{I_{jk}\}_{j\in\mathcal{J},k\in\mathcal{N}(j)}$. The network-design problem will determine this distribution. We assume that building infrastructure requires a resource ("concrete" or "asphalt") in fixed aggregate supply K, which can be freely shipped across locations and cannot be used for other purposes. This assumption represents a situation where an amount of resources has been sunk into network-building but must still be allocated across the network. When characterizing the planner's problem, it will lead to the intuitive property that the opportunity cost of building infrastructure in any location is only foregoing infrastructure elsewhere.

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

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

We allow the network-design problem to take place when some lower bound for infrastructure $\underline{I}_{jk}$ is already in place. We also allow for an upper bound $\overline{I}_{jk}$ to how much can be built, possibly representing geographic constraints on the capacity to build on a specific link. While the graph $(\mathcal{J},\mathcal{E})$ is undirected, the infrastructure matrix $\{I_{jk}\}$ defines a weighted directed graph, as there is no need to impose symmetry in investments or costs between connected locations.

While both the transport technology $\tau_{jk}(Q,I)$ in (4) and the infrastructure building cost δ_{jk}^I may vary across links jk, each type of variation reflects different forces. Variation in $\tau_{jk}(Q,I)$ by jk captures how features of the terrain impact per-unit shipping costs given quantity shipped and infrastructure, whereas δ_{jk}^I captures the marginal cost of setting up infrastructure. In the planner's problem below, δ_{jk}^I will not impact the allocation other than through infrastructure I_{jk}.

3.2. *Planner's Problem*

We solve the problem of a utilitarian social planner who maximizes welfare under two extreme scenarios: labor is either immobile or freely mobile. In the former case, we let ω_j be the planner's weight attached to each worker located in region j. We define each problem in turn.

DEFINITION 1: The planner's problem with immobile labor is

$$W = \max_{\substack{c_j,h_j,\{I_{jk}\}_{k\in\mathcal{N}(j)},\\ \{D_j^n,L_j^n,\mathbf{V}_j^n,\mathbf{X}_j^n,\{Q_{jk}^n\}_{k\in\mathcal{N}(j)}\}_n}} \sum_j \omega_j L_j U(c_j,h_j)$$

subject to:

(i) availability of traded commodities,

$$c_j L_j \leq D_j\left(D_j^1, \ldots, D_j^N\right) \quad \text{for all } j;$$

and availability of non-traded commodities,

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

(ii) the balanced-flows constraint,

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

(iii) the network-building constraint,

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

subject to a pre-existing network,

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

(iv) local labor-market clearing,

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

and local factor market clearing for the remaining factors,

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

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

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

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

DEFINITION 2: The planner's problem with labor mobility is

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

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

(vi) free labor mobility,

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

(vii) aggregate labor-market clearing,

$$\sum_j L_j = L.$$

This formulation restricts the planner's problem to allocations satisfying utility equalization across locations, a condition that must hold in the competitive allocation. Since U is strictly increasing, restriction (vi) implies that the planner will allocate $u = U(c_j, h_j)$ 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.[13] When labor mobility is allowed, the model nests urban economics model with a single homogeneous tradable good 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 quantitative extensions with increasing returns (Allen and Arkolakis (2014), Redding (2016)). In Section 3.7, we discuss how to implement some cases with increasing returns and provide some examples.

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

$$W = \max_{I_{jk}} \max_{Q^n_{jk}} \max_{\{c_j, h_j, D^n_j, L^n_j, \mathbf{v}^n_j, \mathbf{x}^n_j\}} \sum_j \omega_j L_j U(c_j, h_j)$$

subject to the constraints. The innermost maximization problem 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. We now discuss some intuitive features of the solution to the optimal flows subproblem over Q^n_{jk} and the network design problem over I_{jk}.

Optimal Flows

The optimal flow problem that determines the gross flows Q^n_{jk} combines an optimal transport problem—how to map production sources to destinations—and a least-cost route problem with congestion. Under the assumption that domestic absorption D^n_j and production Y^n_j are taken as given, this problem 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 shares 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

[13]The 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^j_j is region j's output in the differentiated commodity that (only) region j produces. 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.

to Proposition 4 below. The first-order condition (A.1) from the planner's problem in Supplemental Material Appendix A.1 gives the following equilibrium price differential for commodity n between j and $k \in \mathcal{N}(j)$:

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

Condition (8) is a no-arbitrage condition: the price differential between a location and its neighbors must be less than or equal to the marginal transport cost. From the planner's perspective, this marginal cost takes into account the diminishing returns due to congestion. In the absence of congestion, $\partial \tau_{jk}^n / \partial Q_{jk} = 0$, 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_{jk}^n as a function of the price differential, as long as the right-hand side can be inverted. The inversion is possible if the total transport cost $Q_{jk}^n \tau_{jk}^n$ 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 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 despite having 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. In the example, the good is produced in the location at the origin (light gray circle) and demanded in ten locations (dark gray circles). This example uses the functional form for τ_{jk} in (10), which implies that there are some ship-

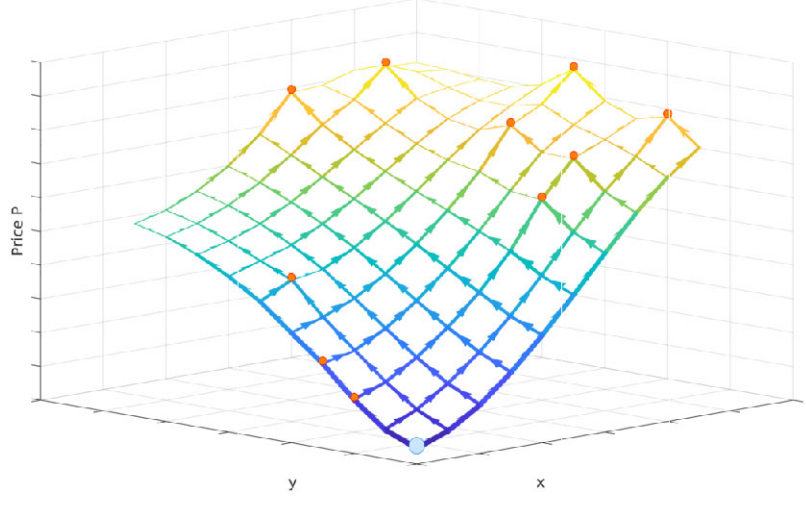


FIGURE 1.—Example of optimal flows as a function of the price field. *Notes:* The figure shows an example of optimal flows in a 15×15 square network with uniform infrastructure across links and one good produced at the origin (light gray circle) and consumed in 10 other locations (dark gray circles). The price in each location is indicated by the z-axis coordinate, and corresponds to a solution of the optimal flow problem given production, consumption, and population. The density of flows is represented by the thickness of links and their direction is indicated by the arrows.

ments 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 to that point. The optimal flows follow the price gradient according to equation (8) under equality. The consumption locations are local peaks of the price field as long as they 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 $j \in \mathcal{J}$ and $k \in \mathcal{N}(j)$. Letting P_K be the multiplier of the network-building constraint (iii) (in other words, the shadow price of asphalt), and as long as the (possibly infinite) upper bound $\overline{I}_{jk}$ is not binding, the planner's choice for I_{jk} implies

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

with equality if there is actual investment, $I_{jk} > \underline{I}_{jk}$. This condition compares the marginal cost and benefits from investing on the link jk. The left-hand side is the opportunity cost of building an extra unit of infrastructure along jk, equal to the marginal value of the scarce resource K in the economy (the multiplier P_K of the network-building constraint (7)) times the rate δ^I_{jk} at which that resource translates to infrastructure. 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 inputs in the transport technology.[14]

Importantly, the network investment problem inherits the properties that make the optimal transport problem tractable. Substituting the solution for Q^n_{jk} as a function of the price differentials P^n_k/P^n_j into (9) implies that the optimal infrastructure I_{jk} between locations j and k is only a function of prices in each location. Hence, rather than searching in the very large space of all networks, this condition allows us to solve for the optimal investment link by link given the smaller set of all prices. Similar properties can be attained in a case with cross-goods congestion, as described in Supplemental Material Appendix A.1.

[14] Various papers measure the first-order impact of changes in bilateral trade costs on world welfare (Atkeson and Burstein (2010) Burstein and Cravino (2015), Lai, Fan, and Qi (2015), Allen, Arkolakis, and Takahashi (2019)) or in trade costs in specific links of a transport network on country-level welfare (Allen and Arkolakis (2019)) 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.

3.3. *Properties*

Convexity

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

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

The first result establishes that the joint optimal allocation and optimal transport subproblems, taking the infrastructure network $\{I_{jk}\}$ as given, define a convex problem for which strong duality holds under the mild requirement that the transport technology $Q\tau_{jk}(Q, I_{jk})$ is (weakly) convex in Q. This property ensures that our specific way of introducing an optimal transport problem into a general neoclassical economy is tractable. Specifically, it guarantees the existence of Lagrange multipliers that implement the optimal allocation and transport subproblems and ensures the sufficiency of the Karush–Kuhn–Tucker (KKT) conditions, in turn allowing us to apply a duality approach to solve the model numerically—an approach which, as discussed in Section 3.6, 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 still 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.[15] This condition restricts how congestion in shipping and the returns to infrastructure enter in the transport technology in each link through $\tau_{jk}(Q, I)$. In the absence of congestion (i.e., if $\partial\tau_{jk}/\partial Q = 0$), convexity fails unless τ_{jk} is a constant. The intuition for this convexity requirement is that the model features two complementarity forces between infrastructure investments and commodity shipments: the higher the investment in a link, the lower the transport costs and the higher the flows. In turn, higher shipments lead to more congestion and to more incentives to develop its infrastructure. The global convexity of the transport cost function ensures that the congestion forces eventually dominate and that the solution to the investment problem is interior and stable. Section C of the Supplementary Material develops this point more formally.

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

Log-Linear Parameterization of Transport Costs

A convenient parameterization of (4) is the constant-elasticity transport technology,

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

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

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:

$$Q^{n}_{jk} = \left[\frac{1}{1+\beta} \frac{I^{\gamma}_{jk}}{\delta^{\tau}_{jk}} \max\left\{ \frac{P^{n}_{k}}{P^{n}_{j}} - 1, 0 \right\} \right]^{\frac{1}{\beta}}. \tag{11}$$

This solution naturally implies that better infrastructure is associated with higher flows given prices and geographic trade frictions. It also shows that the total flows fall with congestion β and increase with the average price differentials. Third, using the log-linear transport technology (10), the optimal level of infrastructure is

$$I_{jk} = \min\left[\max\left(I^{*}_{jk}, \underline{I}_{jk}\right), \overline{I}_{jk}\right], \tag{12}$$

where I^{*}_{jk} is the optimal infrastructure (9) arising from the unconstrained optimal network problem ($\underline{I}_{jk} = 0$ and $\overline{I}_{jk} = \infty$),

$$I^{*}_{jk} = \left[\frac{\gamma}{P_K} \frac{\delta^{\tau}_{jk}}{\delta^{I}_{jk}} \left(\sum_{n} P^{n}_{j} \left(Q^{n}_{jk}\right)^{1+\beta} \right) \right]^{\frac{1}{1+\gamma}}. \tag{13}$$

Given the prices at origin, the optimal infrastructure increases with 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 δ^{τ}_{jk}, reflecting that the optimal investments offset geographic trade frictions, and decreases with δ^{I}_{jk}, reflecting that the investment is smaller where it is more costly to build. 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, $P^{n}_{j} \neq P^{n}_{k}$ for any n.

Combining (11) with (13), we reach an explicit characterization of the optimal infrastructure in each link as a function of prices, elasticities, and geographic frictions:

$$I^{*}_{jk} = \left[\frac{\gamma}{P_K \delta^{I}_{jk} \left(\delta^{\tau}_{jk}\right)^{\frac{1}{\beta}}} \left(\frac{1}{1+\beta} \sum_{n: P^{n}_{k} > P^{n}_{j}} P^{n}_{j} \left(\frac{P^{n}_{k}}{P^{n}_{j}} - 1 \right)^{\frac{1+\beta}{\beta}} \right) \right]^{\frac{\beta}{\beta-\gamma}}, \tag{14}$$

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* (12).

Under a general formulation of the transport technology $\tau_{jk}(Q, I)$ and in the absence of a pre-existing network ($\underline{I}_{jk} = 0$), the solution to the full planner's problem may feature no infrastructure (and therefore no trade) in some links, even if prices vary between the nodes connected by those links. However, when the transport technology takes the log-linear form (10), this possibility arises if and only if there are no incentives to trade ($P_j^n = P_k^n$ for all n) due to the Inada condition on I_{jk} in the transport technology (10) and 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 tractable 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. Focusing on the log-linear specification (10) introduced above, such non-convexities arise when the transport technology features economies of scale, $\gamma > \beta$. We now 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, Colaiori, Flammini, Maritan, and Rinaldo (2000), Bernot, Caselles, and Morel (2009)).

PROPOSITION 3: *In the absence of a pre-existing network* (*i.e*., $\underline{I}_{jk} = 0$, $\overline{I}_{jk} = \infty$), *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 without loops. Intuitively, under the conditions of the proposition, it is always better to remove alternative paths linking pairs of nodes and concentrate infrastructure investments in fewer links. As a result, in the optimal network a single path connects any two locations, a defining characteristic of a tree. This property is guaranteed to hold when there is only one source for one commodity. In the general case, it may still be optimal to maintain loops, but the incentives to concentrate flows on fewer but larger routes remain. In Section 4, we present examples with multiple goods and multiple production locations where, if $\gamma > \beta$, the optimal network is sparser and concentrated on fewer links relative to cases with $\gamma \leq \beta$.

3.4. *Decentralized Allocation Given the Network*

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

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

As long as there is congestion in shipping, the traders will engage in an inefficient amount of shipping. We assume that the market allocation features policies that correct this externality. Specifically, the shipments of commodity n over link jk are subject to ad valorem taxes $\varepsilon_{Q,jkn}^{\tau}\tau_{jk}^n$ on their value at j. Consider then a trader purchasing good n at location o and delivering it to location d. This company maximizes profits by optimizing over the route $r = (j_0, \ldots, j_\rho) \in \mathcal{R}_{od}$, where $j_0, \ldots, j_r$ is a sequence of nodes from o to d and $\mathcal{R}_{od}$ is the set of all such routes. The optimal route r_{od}^n maximizes the per-unit profits:

$$\pi_{od}^n = \max_{r=(j_0,\ldots,j_\rho)\in\mathcal{R}_{od}} p_d^n - \underbrace{p_o^n}_{\text{Sourcing Costs}} - \underbrace{\sum_{k=0}^{\rho-1} p_{j_k}^n \tau_{j_k j_{k+1}}^n}_{\text{Transport costs}} - \underbrace{\sum_{k=0}^{\rho-1} p_{j_k}^n t_{j_k j_{k+1}}^n}_{\text{Taxes}}, \tag{15}$$

where p_j^n 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^n and obtains the price p_d^n. In addition, shippers must pay the transport costs $p_{j_k}^n \tau_{j_k j_{k+1}}^n$ as well as the "toll" $p_{j_k}^n t_{j_k j_{k+1}}^n$ on each segment. 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 $\sum_{j=1}^J t_j L_j = \Pi$, where Π is an aggregate portfolio including ownership of fixed factors and government transfers. Hence, workers are rebated all tax revenues and own 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 they are standard, we relegate the definitions of the competitive allocation with and without labor mobility to Definition 3 in the Supplemental Material. Using that definition, we establish that the welfare theorems given the transport network hold.

PROPOSITION 4—First and Second Welfare Theorems: *If the tax on shipments of product n from j to k is*

$$t_{jk}^n = \varepsilon_{Q,jkn}^{\tau}\tau_{jk}^n,$$

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

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

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

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

These results are useful for bringing our model to the data. 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. We note that the optimal allocation can be equivalently implemented by per-unit toll $\theta_{jk}^n = p_{jk}^n \varepsilon_{Q,jkn}^{\tau} \tau_{jk}^n$.

3.5. *Decentralization of Network Investments*

We now discuss a market structure that efficiently decentralizes the infrastructure investments. Consider a decentralized allocation as in Definition 3, including tolls. Suppose that, in addition, I_{jk} is endogenously determined by a link-specific builder who is granted the right to build infrastructure and receives in exchange a per-unit toll θ_{jk}^n. Builders can purchase the "asphalt" K at a price p_K, the stock of K may be initially owned by the government or by private individuals, and the price p_K adjusts such that the market for K clears. The builders will solve the problem

$$\max_{I_{jk}} \sum_n \theta_{jk}^n Q_{jk}^n(I_{jk}) - p_K \delta_{jk}^I I_{jk},$$

where $Q_{jk}^n(I_{jk})$ is the quantity of good n consistent with zero profits of shipping companies on the link jk given infrastructure I_{jk} and prices. The builders internalize that, by adding infrastructure, they can increase the flow of goods through their link, but we assume that they do not internalize general-equilibrium impacts on commodity prices. Now, the specific transfers t_j exhaust a portfolio Π which, in addition to fixed factors and government transfers, also includes ownership of K and net profits of builders.

We then obtain the following result.

PROPOSITION 5: *If the global convexity condition of Proposition 1 is satisfied and the toll* θ_{jk}^n *is consistent with the optimal Pigouvian tax* ($\theta_{jk}^n = P_j^n \varepsilon_{Q,jkn}^{\tau} \tau_{jk}^n$), *then the decentralized infrastructure choice implements the optimal network investment*.

This result echoes the self-financing theorem of Mohring and Harwitz (1962) who showed that revenues from optimal congestion taxes are sufficient to cover capital costs of roads when the transport cost can be expressed as a function of the ratio Q/I. Our result is not restricted to the case in which the transport cost function is homogeneous of degree 0. The global convexity condition of Proposition 1 ensures the sufficiency of the first-order conditions in implementing the optimal allocation. In this general case, however, lump-sum transfers to builders may be needed to ensure participation.

3.6. *Numerical Implementation*

In this section, we broadly discuss our numerical implementation and relegate details to Section B of the Supplementary Material.[16]

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 nonlinear equations with many unknowns. 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)).[17]

Our problem can be tackled numerically using two equally valid approaches. The first one is to feed the numerical solver the *primal* problem, meaning the full planner's problem exactly as written in Definition 1. Specifically, letting $\mathcal{L}$ be the Lagrangian of the planner's problem as a function of the controls $\mathbf{x} = (c_j, h_j, D_j^n, L_j^n, \mathbf{V}_j^n, Q_{jk}^n, \ldots)$ and the multipliers $\boldsymbol{\lambda} = (P_j^n, \ldots)$, the primal consists of solving the saddle-point problem

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

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

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

In our context, the convexity of the full planner's problem without labor mobility ensures that the dual coincides with the primal (Proposition 1), that is, 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 (13), as well as those from the neoclassical allocation problem, to express the control variables as functions of the multipliers, $\mathbf{x}(\lambda)$. The remaining minimization problem, $\inf_{\lambda \geq 0} \mathcal{L}(\mathbf{x}(\lambda), \boldsymbol{\lambda})$, is a convex minimization problem over fewer variables, subject to non-negativity constraints only.

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

[16] A Matlab toolbox implementing our model with detailed documentation and examples is available on the authors' websites.

[17] 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)).

a guess on the network investment I_{jk}, we solve for the optimum over c_j, h_j, D_j^n, L_j^n, $\mathbf{V}_j^n$, and Q_{jk}^n, and then use the optimal network investment condition (9) to obtain a new guess over I_{jk}, and then repeat until convergence. We then refine the solution using a simulated annealing method that perturbs the local optimum and gradually reaches better solutions. Appendix B of the Supplemental Material provides 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 incorporate this feature while preserving the convexity of the problem is to assume that the per-unit cost τ_{jk}^n is denominated in units of the bundle of traded goods aggregated through $D_j(\cdot)$ rather that in units of the good itself. We assume that transporting each unit of good n from j to $k \in \mathcal{N}(n)$ requires

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

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

$$T_j = \sum_{k \in \mathcal{N}(j)} \tau_{jk}(Q_{jk}, I_{jk}) Q_{jk}.$$

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.

Appendices A.1 and A.2 of the Supplemental Material present the definition and first-order conditions of the planner's problem in a general case encompassing iceberg costs with own-good congestion, as well as this alternative formulation with congestion across goods. As it is more realistic, 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 $Y_j^n = F_j^n(L_j^n, \mathbf{V}_j^n, \mathbf{X}_j^n; L_j)$, 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 entering through utility, $U(c_j, h_j; L_j)$. Similarly, without the Pigouvian taxes t_{jk}^n correcting the congestion externality in shipping, the market allocation is inefficient. In these cases, it is in principle 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 τ_{jk}^n on Q_{jk}^n. For example, in the case of production spillovers, the fictitious planner's problem is defined as in Definition 2 under the assumption that the vector of aggregate population levels $\overline{\mathbf{L}} = \{\overline{L}_j\}$

in $Y_j^n = F_j^n(L_j^n, \mathbf{V}_j^n, \mathbf{X}_j^n; \overline{L}_j)$ is taken as given.[18] As long as $F_j^n(\cdot)$ is neoclassical given $\overline{L}_j$, the statement in part (i) of Proposition 1, establishing convexity of the planner's problem given the network, remains the same given $\overline{L}_j$. However, this approach requires solving an additional loop imposing that the vector of population $\mathbf{L} = \{L_j\}$ that solves the fictitious planner's problem coincides with the aggregate distribution $\overline{\mathbf{L}}$ taken as given by the planner. Every distribution of population $\overline{\mathbf{L}}$ satisfying this fixed point problem corresponds to a market allocation and vice versa.[19] An important caveat, however, is that we can no longer establish the general convexity of the problem corresponding to part (ii) of Proposition 1. Section D of the Supplementary Material explains how to implement these cases along with a method to derive the optimal 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. All the figures can be found in Section A of the Supplementary Material. 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, $U = (c^\alpha h^{1-\alpha})^{1-\rho}/(1-\rho)$ 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, $\delta_{jk}^\tau = \delta_{jk}^I = Distance_{jk}$.

Figure A.3 presents a network with 9×9 locations uniformly distributed in a square, each connected to eight 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.

[18]The fictitious planner problem is defined exactly as in Definition 2 with $U(c_j, h_j; \overline{L}_j)$ in the case of consumption externalities, taking $\overline{L}_j$ as given, or with $\tau_{jk}(\overline{Q_{jk}^n}, I_{jk})$ in the case of congestion externalities, taking the shipments $\overline{\mathbf{Q}} = \{\overline{Q_{jk}^n}\}_{j,k,n}$ as given.

[19]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. For example, see Allen and Arkolakis (2014) for conditions that lead to existence and uniqueness in an Armington model with labor mobility and size spillovers.

Figure A.4 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 (13). 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.5 in the Supplementary Material 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 10 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 Supplementary Material Section B involving simulated annealing. The bottom panel compares the non-convex network before and after the annealing refinement. The refined network economizes on the number of links, leading to a welfare increase but preserving the no-loops property.

4.2. *Many Sectors, Labor Mobility, and Non-Convexity*

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

Figure A.6 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.7 shows the optimal network with annealing in the non-convex 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.6 and A.7 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.8 shows 20 cities randomly allocated in a space where each location is connected to eight 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 (13), the optimal infrastructure in a given link depends on the link-specific building cost δ^I_{jk}. In panel (a), we show the optimal network under the assumption that the cost of building infrastructure is proportional to the Euclidean distance:

$$\delta^I_{jk} = \delta_0 Distance_{jk}^{\delta_1}. \tag{17}$$

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

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

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

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

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

$$\delta^I_{jk} = \delta_0 Distance_{jk}^{\delta_1}\left(1 + |\Delta Elevation|_{jk}\right)^{\delta_2} \delta_3^{CrossingRiver_{jk}} \delta_4^{AlongRiver_{jk}}. \tag{19}$$

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

In panel (d), we assume instead that no dry patch exists but 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 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 via ground transport.

Finally, panel (f) shows the non-convex case, $\gamma = 2 > \beta$, implemented through the combination of first-order conditions and simulated annealing described in Section 3.6. Now, a unique route links any two cities and fewer roads are 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.

5.1. *Data*

Sources

We combine geocoded data on road networks, population, and income across European countries. The road network data are from EuroGlobalMap (EGM) by EuroGeographics.[20] The data set 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.[21] For example, the French road network is represented by 38,668 segments of active roads connecting 159,258 geographic points with a total length of about 130,000 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 are available. This set includes rich and poor countries, as well as geographically large and small. Table A.1 in Supplemental Material Appendix B 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 data set 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.[22] On average across the countries in our data, national roads 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

[20]This product includes Intellectual Property from the 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 license are available at https://eurogeographics.org/services/opendata/topographic-data/.

[21]We 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."

[22]For example, 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.

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 data sets correspond to the year 2005. The GPW population data are reported for 30 arc-second cells (approximately 1 kilometer) and the G-Econ value-added data are 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 data sets and the coarser resolution of the G-Econ data set, 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 geographic units in the G-Econ data set coincide with ours. We allocate population to each 0.5-degree cell by aggregating the smaller cells in GPW and we 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.[23] We denote the population and value added observed in each cell j of each country by L_j^{obs} and GDP_j^{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 is available, 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.[24]

Underlying Graph

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

To define the set of nodes $\mathcal{J}$ 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.[25] We define the observed population and income of each node $j \in \mathcal{J}$ to be equal to the total income GDP_j^{obs} and the population L_j^{obs} of the cell that contains it.

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

[23] 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 fewer than 20 cells, we use 0.25 arc-degree cells (Luxembourg and Cyprus).

[24] NUTS (Nomenclature of Territorial Units for Statistics) is a standard developed by the European Union to divide the territory. 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.

[25] On average across countries, the relocation across all cells within a country is 5.3 km.

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:

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

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.[26] We follow these steps as long as the cheapest path does not stray from the cells containing j and k.[27] 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 $s \in S$.

We define the observed measure of infrastructure I_{jk}^{obs} for each $j \in \mathcal{J}$ and $k \in \mathcal{N}(j)$ by letting $I_{jk}^{\text{obs}} = \text{lanes}_{jk}$ for national roads and $I_{jk}^{\text{obs}} = \text{lanes}_{jk} \times \kappa$ for non-national roads, where $\kappa < 1$ captures the smaller cost of non-national roads. We set $\kappa = 1/5$, 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). We impose $I_{jk}^{\text{obs}} = I_{kj}^{\text{obs}}$, implying that infrastructure applies equally in either direction. In sum, we construct the observed infrastructure I_{jk}^{obs} as the average number of national road lanes over the path from j to k on the actual road network, if a direct path exists.[28]

[26] This step does not use the model. In this step, for each pair of nodes $j \in \mathcal{J}$ and $k \in \mathcal{N}(j)$ we ask: what are the average characteristics (number of lanes and type of road) of the existing route connecting these two locations in the real world? To do this, we must choose some route connecting the pair of locations in the real world. We use the cheapest-route criterion as a way to choose this route. The cheapest path is constructed by weighting each segment s by its road user cost based on data from Combes and Lafourcade (2005) and other sources. See Supplemental Material Appendix B for details.

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

[28] Across connected nodes in the discretized network, there is a correlation of 0.67 between I_{jk}^{obs} and the speed on the quickest path according to GoogleMaps.

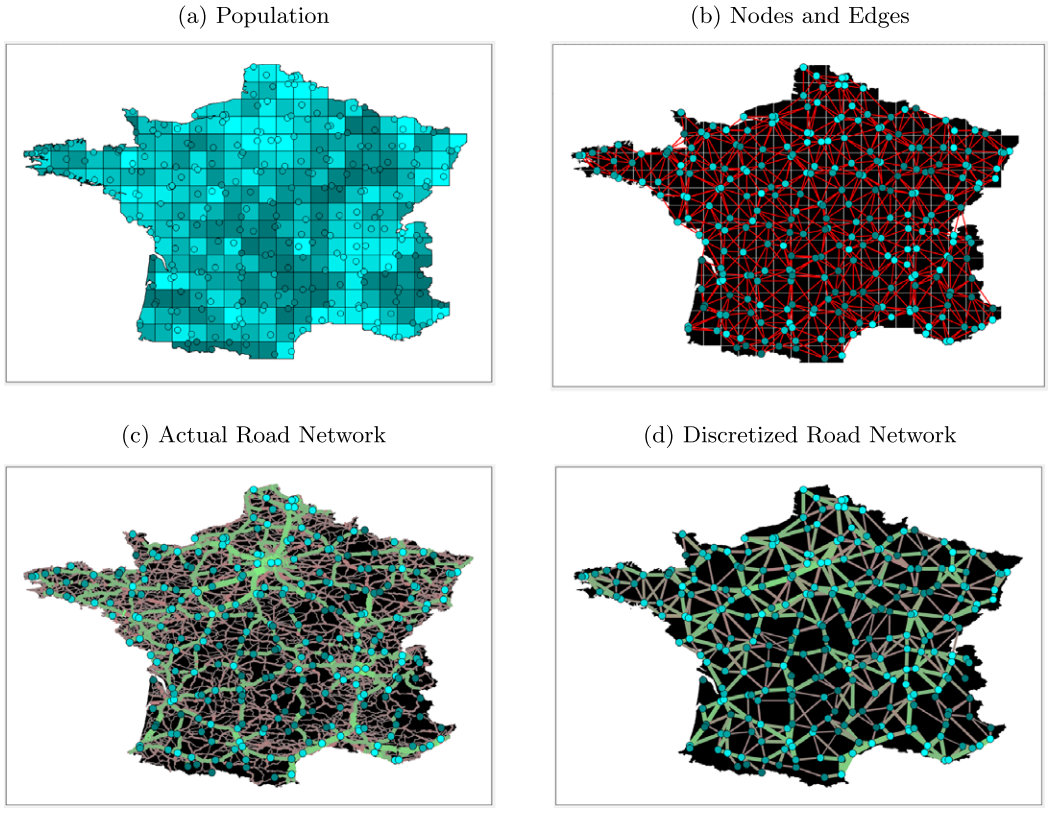


FIGURE 2.—Discretization of the French road network. *Notes:* Panel (a) shows total population from GPW aggregated into 0.5 arc-degree (approximately 50 km) cells. Panel (b) shows the nodes $\mathcal{J}$ corresponding to the population centroids of each cell in panel (a), reallocated to their closest point on the actual road network, and the edges $\mathcal{E}$ corresponding to all the vertical and diagonal links between cells. Panel (c) shows the centroids and the actual road network. Light gray segments correspond to national roads, dark gray 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 of lanes on the cheapest path between each pair of nodes on the road network. The shade varies according to the fraction of the shortest path traveled on a national road.

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 figure 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, and the edges of the underlying graph. The (c) panels show the centroids and the full road network. Light gray segments correspond to national roads and dark gray segments correspond to other roads. 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, lanes_{jk}, and a level of brightness depending on the likelihood of using a national road, nat_{jk}. The width and color scale are the same as in panel (c). When no direct link from j to k is identified by our procedure, no edge is shown. The resulting discretized networks on the baseline grids clearly mirror the actual road networks for both countries,

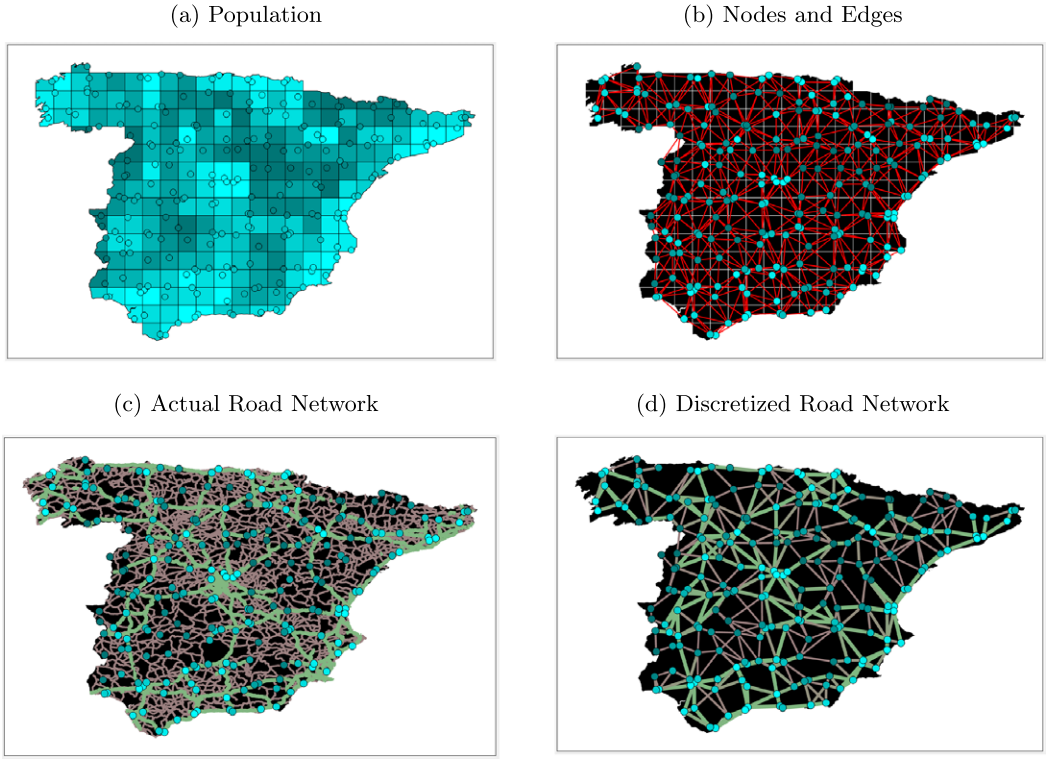


FIGURE 3.—Discretization of the Spanish road network. *Notes:* Panel (a) shows total population from GPW aggregated into 0.5 arc-degree (approximately 50 km) cells. Panel (b) shows the nodes $\mathcal{J}$ corresponding to the population centroids of each cell in panel (a), reallocated to their closest point on the actual road network, and the edges $\mathcal{E}$ corresponding to all the vertical and diagonal links between cells. Panel (c) shows the centroids and the actual road network. Light gray segments correspond to national roads, dark gray 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 of lanes on the cheapest path between each pair of nodes on the road network. The shade ranges according to the fraction of the shortest path traveled on a national road.

but they are now expressed in terms of the nodes and edges of our model and therefore allow us to quantify it.

5.2. *Parameterization*

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

Preferences and Technologies

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

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

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

$$C_j = \left(\sum_{n=1}^{N} \left(C_j^n \right)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \tag{20}$$

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

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

We 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 σ.

Labor Mobility

We undertake the country-by-country analysis of misallocation for the cases in which labor is fixed and in which it is perfectly mobile. In this way, we accommodate that internal rates of labor mobility may be different across countries. In the absence of data to discipline this assumption, we opt for reporting the results in these polar cases. For the multi-country application of the last section, we allow in addition for a partial-mobility case where labor is mobile within countries but not across countries.

Transport Technology

We adopt the log-linear transport technology (10) with cross-good congestion as described in Section 3.7. We must parameterize the congestion parameter β and the returns to infrastructure parameter γ. Ideally, we would like to set these parameters to elasticities of total trade costs with respect to trade flows and infrastructure. Since such elasticities are not readily available, we narrow the focus to the impact of shipping time on trade costs. Several studies point to a relevant value of time in international shipping.[29] Since the majority of inland shipments in the EU are done via road, they likely include goods that are time-sensitive.[30]

We assume that: (i) trade costs are a linear function of shipping time; (ii) shipping speed is a log-linear function of the number of vehicles and road lane kilometers; and (iii) the number of vehicles is a linear function of the quantity shipped. As shown in Supplemental Material Appendix B, under these assumptions we can calibrate β and γ to match the

[29] Anderson and Van Wincoop (2004) calculated 9-percent tax equivalent of the average ocean shipping time cost in the U.S. over the second half of the 20th century. Hummels and Schaur (2013) quantified that one additional day in transit is equivalent to 0.6 to 2.1 percent tariff, and Djankov, Freund, and Pham (2010) argued that each additional day of delay is equivalent to a country distancing 70 km from its trade partner. Firth (2017) showed that the time delays caused by congestion of railroads impact firm-level outcomes in India, and Brancaccio, Kalouptsidi, and Papageorgiou (2019) argued, using a structurally estimated search model of ships and exporters, that congestion in ports leads to costly delays for exporters.

[30] Seventy-five percent 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.

empirical relationship between speed, roads, and vehicles estimated by Couture, Duranton, and Turner (2018) in U.S. data. Their estimates imply $\beta = 0.13$ and $\gamma = 0.10$, suggesting decreasing returns to scale.[31] We use these parameters as a benchmark and also implement the analysis in a case with increasing returns. Specifically, we consider a higher value of γ such that the ratio between γ and β is the mirroring case, $\gamma/\beta = 0.13/0.10$ for $\beta = 0.13$.

Productivities, Endowments, and Geographic Frictions

We must impose values for the productivities z_j^n and the endowment of non-traded services H_j. In the case with perfect labor mobility, we interpret $(L_j^{\text{obs}}, GDP_j^{\text{obs}})$ 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_j^n, 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_j^n, normalizing non-tradeable consumption per capita H_j/L_j^{obs} to 1 and setting the planner's weights $\omega_j = 1$ everywhere.

Since our data only include 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 homogeneous good is produced by all the remaining cells. We assume 10 different sectors and explore the robustness to alternative values of N. 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. We choose each location's productivity (and supply of non-traded goods, when allowing for labor mobility) such that, taking the observed network I_{jk}^{obs} as given, the planner's solution to the optimal allocation and optimal flows problems from Definition 2 reproduces the observed value added (and population, when allowing for labor mobility).[32]

We must also determine the values of the geographic trade frictions δ_{jk}^{τ} entering in the transport technology (10). We assume that all goods have the same weight ($m_n = 1$) and that frictions depend on distance, $\delta_{jk}^{\tau} = \delta_0^{\tau} dist_{jk}$. To calibrate δ_0^{τ}, we target the level of intra-regional trade in Spain, where regional-level trade data are available. We estimate δ_0^{τ} jointly with the other fundamentals so that the model matches the 44% share

[31] Couture, Duranton, and Turner (2018) found decreasing returns to scale across their specifications, although the difference between the two parameters is typically small (see Tables 5 and 6 of their paper). They referred to this result as suggesting "modest decreasing returns to scale." Their preferred estimate, used for our calibration, is column (6) of Table 5 of their paper.

[32] To compute GDP, we invoke the second welfare theorem from Proposition 4 to recover the prices in the calibrated allocation as the multipliers of the various constraints in the planner's problem. In the solution of the planner's problem, each location's value added in tradable and non-tradable sectors is $P_j^{n(j)} z_j L_j + P_j^H H_j$, where $n(j)$ denotes the good produced by location j, P_j^n is the price of good n in location j (i.e., multiplier of the flows constraint for good n in j in the planner's problem), and P_j^H is the price of non-traded services in sector j (i.e., the multiplier of the availability of non-traded goods constraint in the planner's problem). Value added in the transport sector is not attached to specific nodes and often corresponds to links connecting near empty locations. For accounting purposes, we allocate the national value added in the transport sector proportionally to value added in other sectors, so that regional variation in measured GDP is driven by goods and services.

of intra-regional trade in intra-national trade among tradable sectors across the 15 continental level-2 NUTS regions of Spain from 2001 to 2005, according to Spain's C-Intereg Dataset (Llano, Esteban, Pérez, and Pulido (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.[33] We then use this value of δ_0^τ in the remaining countries when calibrating fundamentals.

Figure A.1 in Supplemental Material Appendix B 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.[34]

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 shares 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 parameterize the cost of infrastructure along each edge, δ_{jk}^I. 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. In this case, the observed network, I_{jk}^{obs}, is consistent with the planner's first-order condition for I_{jk} in (13) under the assumption that $\underline{I}_{jk} = 0$. Imposing symmetry on that first-order condition, we then recover the cost of infrastructure as a function of outcomes from the calibrated model (see Supplemental Material Appendix A.2). We refer to this measure as the "FOC-based" measure of building costs, $\delta_{jk}^{I,\text{FOC}}$.

Our second approach is agnostic about whether the observed network results from any sort of optimization, but takes a stand about how the building costs depend on geographic features. Specifically, we rely on data from Collier, Kirchberger, and Söderbom (2016), who estimated highway building costs from World Bank infrastructure investment projects across the world, and then related these costs to a host of geographic and non-geographic frictions.[35] We assume that δ_{jk}^I is a function of two geographic features included in their study, distance and ruggedness of the terrain, and refer to this building-cost measure as the "geographic" measure, $\delta_{jk}^{I,\text{GEO}}$. We interpret an improvement to the connection between any pair of nodes in our counterfactuals as an infrastructure project.

[33]The model does not make a prediction for bilateral 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%.

[34]Across the 24 countries, the average internal trade share is 38%, with a standard deviation of 12%. We calibrate 0.00156 in the benchmark convex case with mobile labor and 0.00164 in the non-convex case.

[35]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 (21) correspond to the average of the coefficients over the distance dummy and the ruggedness index across the six specifications in Tables 4 and 5 of their paper.

In our notation, their estimates imply

$$\ln\left(\frac{\delta_{jk}^{I,\text{GEO}}}{dist_{jk}}\right) = \ln\left(\delta_0^I\right) - 0.11 \times (dist_{jk} > 50 \text{ km}) + 0.12 \times \ln(rugged_{jk}), \tag{21}$$

where $dist_{jk}$ is the distance between j and k and $rugged_{jk}$ is the average ruggedness over locations j and k, constructed as detailed in Supplemental Material Appendix B. Hence, 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.

These steps give two alternative measurements of δ_{jk}^I up to scale in each country. We set $K = 1$ and choose δ_0^I to satisfy the network-building constraint with equality in each country.

5.3. *Model-Implied Trade Flows and Congestion*

We check the model predictions for bilateral trade flows. We use bilateral trade data across Spanish regions defined at the level-2 NUTS subdivision from Spain's C-Intereg Dataset. Excluding islands, this gives 15 Spanish regions. Panel (a) of Figure 4 shows observed 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 hollow circles. The model implied bilateral flows have a correlation of 0.79 with the data.

Another way to assess the implied trade flows 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 is often applied within countries (Allen and Arkolakis (2014)). A common approach is to parameterize bilateral trade costs as a function of geographic frictions. In our parameterization, 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 β and γ. We can compare the relationship between trade flows and distance implied by the model using the previous aggregation to the level-2 NUTS subdivision in Spain. Panel (b) of Figure 4 shows the bilateral import share among level-2 NUTS and the log of distance in the model and in the data, after controlling for exporter fixed effects.[36] A linear regression yields elasticities 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.[37]

We examine the congestion taxes needed to implement the allocation. Due to the log-linear specification of the transport technology, Proposition 4 implies that the taxes are a fraction $\beta = 13\%$ of the transport costs in every link. Given the total transport costs,

[36]We run, in both model-generated and observed data, the regression $\ln(\lambda_{jk}^{\text{NUTS}}) = \delta \ln(dist_{jk}) + \psi_j + \varepsilon_{jk}$, where λ_{jk} is the import share and $dist_{jk}$ 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. Distance is computed between geographic centroids. We exclude zero flows and flows to the own region.

[37]The results are similar in calibrations that assign differentiated products to the largest locations in the country, with correlations around 0.8 between model-based nonzero bilateral trade flows and the data. They are also similar in calibrations that assign one good to each region but assume away the homogeneous product. The relationship between trade and distance is very similar for France, suggesting that the key gravity properties are dictated by elasticity parameters rather than by the distribution of fundamentals.

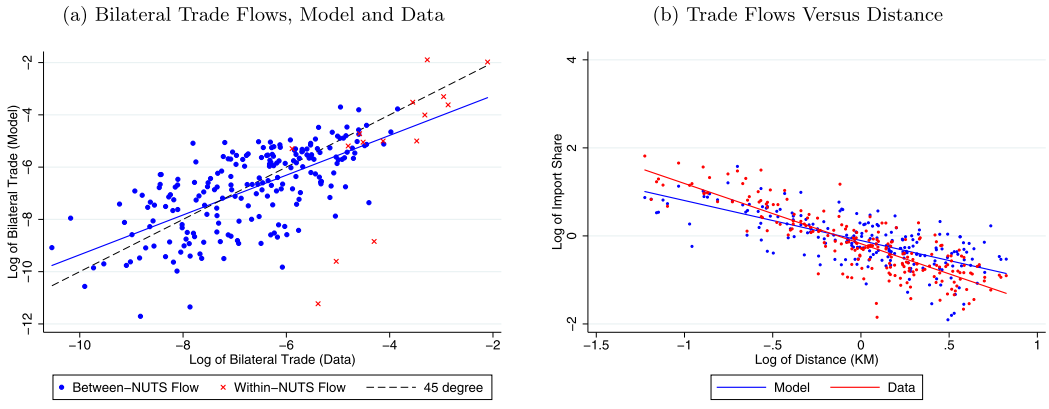


FIGURE 4.—Actual and predicted trade. *Note:* The left panel shows the model-based flows across level-2 NUTS regions of continental Spain against the data in the calibration assigning a differentiated product to the largest city within each level-2 NUTS region. The right panel shows the log of the model-implied and observed import shares against distance as residuals from exporter fixed effects. Linear regression slope (robust SE) is: −0.908 (0.069) in the model and −1.368 (0.058) in the data. Trade flows to the same region appear as hollow circles in the first panel.

we obtain numerically that the mean ad valorem tax across locations is about 0.6% in the cases with and without labor mobility. The total taxes paid represent 0.3% of GDP in both cases.[38]

5.4. *Optimal Expansion and Reallocation*

We simulate two types of counterfactuals. First, we compute the aggregate gains from an optimal expansion of the observed road network within each country. 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}. In the notation of restriction (iii) in Definitions 1 and 2, this means that $\underline{I}_{jk} = I_{jk}^{\text{obs}}$. Second, we compute the losses due to misallocation of current roads within each country. 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}. In the notation of restriction (iii) in Definitions 1 and 2, this means that $\underline{I}_{jk} = 0$. We set the upper bound on infrastructure $\overline{I}_{jk}$ to be 50% above the largest level of infrastructure observed in each country.

In short, the first "optimal expansion" counterfactual amounts to optimally expanding the network on top of what is already observed, while 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.

[38]In simulations of the calibration for Spain where we randomly increase capacity in random links, we find an elasticity of quantity flows Q_{jk} to infrastructure I_{jk} of 0.503 (0.0202) in the benchmark calibration ($\gamma < \beta$), 0.799 (0.0330) in a calibration that imposes $\gamma = \beta$, and 1.296 (0.0395) in the non-convex calibration ($\gamma > \beta$). Duranton and Turner (2011) reported IV estimates of the relationship between total vehicles-kilometers traveled and road capacity across U.S. cities between 0.68 and 1.33 (in their Table 6), with many of these estimates close to 1.

We implement the optimal expansion under the two measures of building costs, the FOC-based measure $\delta_{jk}^{I,\text{FOC}}$ and the geographic measure $\delta_{jk}^{I,\text{GEO}}$. The optimal reallocation is only meaningful under the geographic measure, since the observed network is optimal by construction under $\delta_{jk}^{I,\text{FOC}}$. We implement each of these counterfactuals for each of the two values of γ, assuming both fixed and mobile labor, separately for each of the 24 countries. We recalibrate the model for each value of γ, assumption on labor mobility, and country.

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 $\delta^{I,\text{GEO}}$. Panels (a) and (b) show the optimal expansion and panels (c) and (d) show the optimal reallocation under labor mobility. Panels (e) and (f) reproduce the optimal reallocation assuming that labor is not mobile. 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 higher reallocation are brighter. In turn, with labor mobility, lighter nodes denote higher population increase.

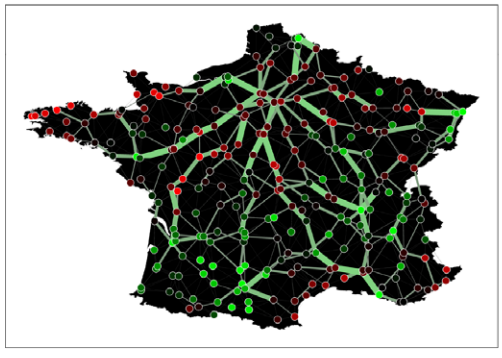
In the optimal reallocation counterfactual, we observe positive investments radiating away from some areas with higher economic activity in the case of France, but a more dispersed investment pattern in Spain. As we compare panels (a) and (b) with panels (c) and (d), we recognize similar investment patterns in the optimal reallocation and expansion counterfactuals within each country: the links identified as having too much infrastructure, shown in dark gray 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.

In the cases of optimal expansion and optimal reallocation, the 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.[39] 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.

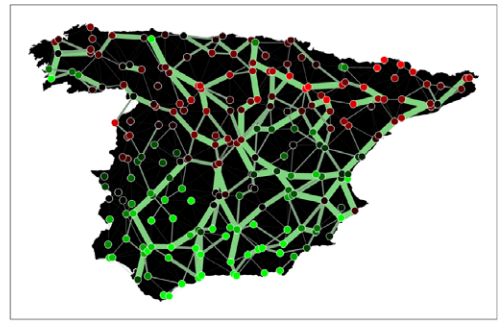
We inspect, across the 24 countries, how a few typically observable regional outcomes map to infrastructure investment. Panel (a) of Table I reports results from regressions of infrastructure growth on each location's initial population and tradable income per capita. We report here results corresponding to the case with labor mobility and the benchmark parameterization of γ and β, but the qualitative features we discuss are similar under alternative specifications. Optimal road investments are directed to locations with initially lower levels of infrastructure, reflecting decreasing returns to infrastructure at the link level. 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 24–39% under geographic measure of trade costs and only 4% under the FOC measure).

[39]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 allocations.

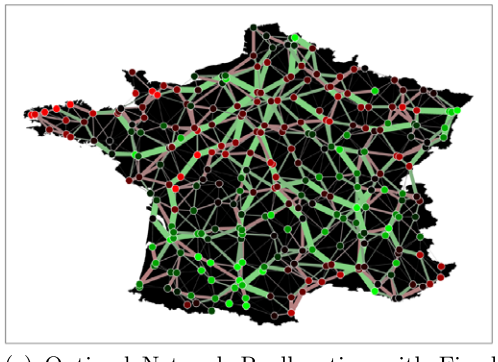
(a) Optimal Network Expansion with Labor Mobility, France



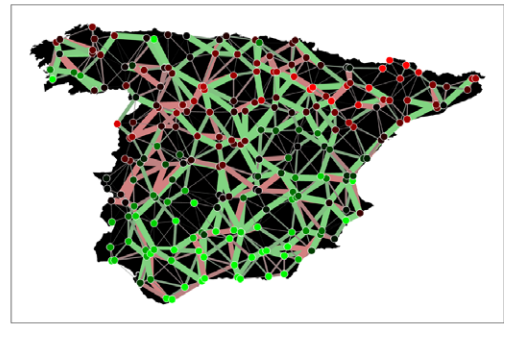
(b) Optimal Network Expansion with Labor Mobility, Spain



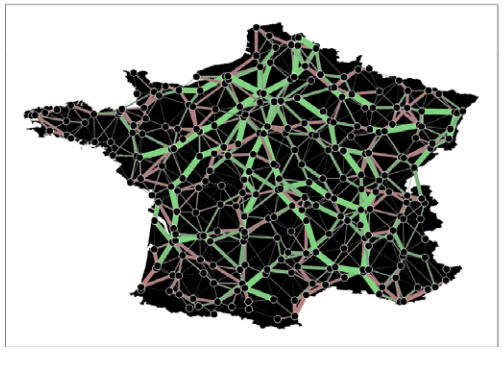
(c) Optimal Network Reallocation with Labor Mobility, France



(d) Optimal Network Reallocation with Labor Mobility, Spain



(e) Optimal Network Reallocation with Fixed Labor, France



(f) Optimal Network Reallocation with Fixed Labor, Spain

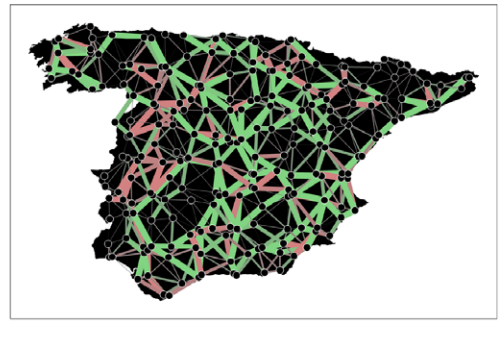


FIGURE 5.—Optimal network reallocation and expansion: Spain and France. *Notes:* All counterfactuals use the geographic measure of building cost, $\delta^{I,\mathrm{GEO}}$ and the benchmark parameterization of β and γ. Graph nodes aggregate data from 0.5 arc-degree (approximately 50 km) cells. The width and brightness of each link is proportional to the difference between the optimal counterfactual network and the observed network, $I^{*}_{jk} - I^{\mathrm{obs}}_{jk}$, 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. Dark links represent negative investment. With labor mobility, brighter nodes represent larger population increase.

TABLE I

OPTIMAL INFRASTRUCTURE INVESTMENT, POPULATION GROWTH, AND LOCAL CHARACTERISTICS[a]

	(1) Reallocation	(2) Expansion (GEO)	(3) Expansion (FOC)
	Panel A: Dependent Variable: Infrastructure Growth		
Population	0.343	0.125	0.002
Tradable Income per Capita	0.151	0.071	0.007
Infrastructure	−0.418	−0.235	−0.010
Observations	868	868	868
Adjusted R-squared	0.29	0.24	0.04
	Panel B: Dependent Variable: Population Growth		
Population	−0.001	−0.000	−0.000
Tradable Income per Capita	0.008	0.008	0.001
Consumption per Capita	−0.061	−0.060	−0.008
Infrastructure	−0.001	−0.001	−0.000
Infrastructure Growth	0.002	0.002	−0.001
Differentiated Producer	0.010	0.010	0.002
Observations	868	868	868
Adjusted R-squared	0.53	0.54	0.80

[a]Each column corresponds to a different regression pooling all locations across the 24 countries in the benchmark parameterization of γ and β, assuming mobile labor and $N = 10$. All regressions include country fixed effects. Standard errors are clustered at the country level. 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 is 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.

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. The handful of variables in the regression explain between 50% and 80% of the population changes. Infrastructure growth in a location has a positive impact on population in the counterfactuals that use the geographic measure of trade costs. The magnitude of the effect and its explanatory power on the distribution of population changes is small, reflecting that growth in a location depends on investments in other locations.

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 spa-

tial 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 II shows the average welfare gain for each counterfactual across the 24 countries in our data. Tables A.2 and A.3 in Supplemental Material Appendix B show the results for each country with fixed and mobile labor, respectively. In the benchmark parameterization of γ and β, using the geographic measure of building costs, we find average welfare gains across countries of 1.7%–1.8%. 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 γ, with the average welfare gains increasing to between 2.4% and 2.9% under geographic measure of building costs. These effects vary considerably across countries, ranging from around 0.1% to 8%. There is no clear relationship between misallocation and country size or income. Some Eastern European countries such as Georgia, Lithuania, and Latvia appear with relatively high misallocation in the benchmark case (3.0% to 3.6% relative to a mean of 2.4%), and so do Denmark (7.8%), France (3.4%), and Spain (4.7%). Belgium, Luxembourg, and Macedonia appear as the least misallocated countries.

This distribution of welfare gains across countries is, in general, stable regardless of the parameterization of γ, the assumption on labor mobility, the parameterization of the building costs δ^I, or the type of counterfactual. For example, across the parameterizations of γ and labor mobility, the correlation between the gains from optimally expanding the network under the two measures of building costs, $\delta^{I,\text{GEO}}$ and $\delta^{I,\text{FOC}}$, is between 0.76 and 0.92. 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 are robust across these cases.

Alternative Assumptions

The analysis was implemented assuming $N = 10$ sectors. We also implement the calibration and counterfactuals assuming different numbers of sectors. Table A.4 in Supplemental Material Appendix B reports the coefficients from column (2) of Table I corresponding to the optimal expansion under calibrations that assume $N = 5$ or $N = 15$. In

TABLE II

AVERAGE WELFARE GAINS ACROSS COUNTRIES[a]

	Returns to Scale			
	Benchmark		Non-Convex	
	Labor		Labor	
	Fixed	Mobile	Fixed	Mobile
Optimal Reallocation				
$\delta = \delta^{I,\text{GEO}}$	1.7%	1.8%	2.4%	2.5%
Optimal Expansion				
$\delta = \delta^{I,\text{GEO}}$	1.7%	1.8%	2.8%	2.9%
$\delta = \delta^{I,\text{FOC}}$	0.3%	0.3%	0.9%	1.3%

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

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 II 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 γ. 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. The correlation in welfare gains across countries between the benchmark case and this alternative allocation is above 0.8 across assumptions of labor mobility, number of goods, and type of counterfactual.

Finally, Table A.6 replicates the benchmark case under the assumption of no congestion across goods. We find very similar average welfare effects in the two cases.

5.5. *Application to Multiple Countries Within Europe*

Our previous applications considered each country in isolation. We now implement the analysis for a region of western Europe.[40] Supplemental Material Appendix Figure A.2 shows the baseline map and the discretized network for this connected set of countries. We assume that each country produces a country-specific differentiated product, in addition to a homogeneous good, and use the same parameters as in the benchmark. We recalibrate the fundamentals assuming that the five largest locations in terms of observed population within each country produce the differentiated product of that country, while the remaining locations produce the homogeneous product. We now 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. The counterfactuals highlight the 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, reflecting the European planner's incentives to deal with international trade. The European planner prioritizes two corridors connecting Spain to the north of Portugal and the center of Spain to France, whereas the Spain-level planner 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 panel (a) of Figure 5, but now we see optimal investments in the connection with Spain, as well as investments in the northeast to connect with neighboring countries.

The optimal network expansion is very similar in the three cases of labor mobility and the welfare gains are close to 2.5% 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 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. The network includes roads,

[40]We include 11 countries: Austria, Belgium, Switzerland, Germany, Denmark, Spain, France, Italy, Luxembourg, Netherlands, and Portugal. We use 1 degree by 1 degree cells, resulting in 261 cells.

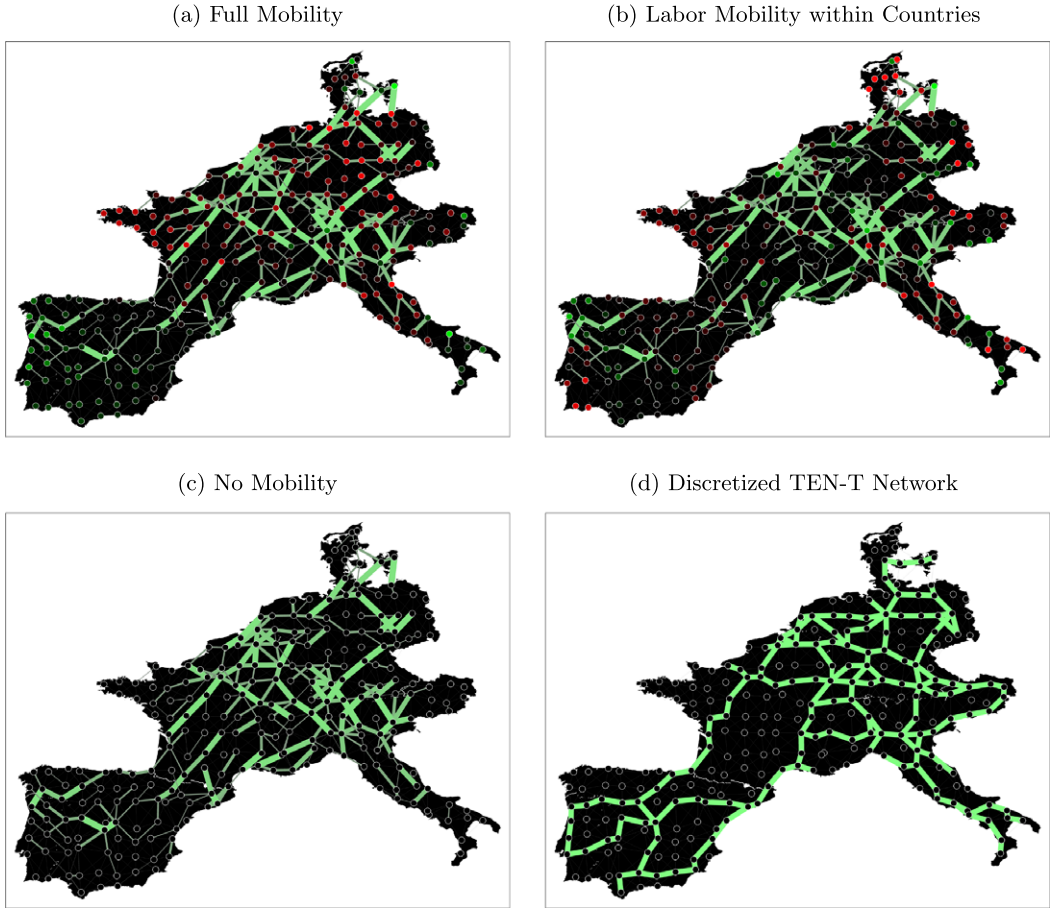


FIGURE 6.—Optimal network expansion: Europe. *Notes:* All counterfactuals use the geographic measure of building costs, $\delta^{I,\text{GEO}}$ and the benchmark parameterization of β and γ. The width and brightness of each link is proportional to the difference between the optimal counterfactual network and the observed network, $I^*_{jk} - I^{\text{obs}}_{jk}$, for each link $jk \in \mathcal{E}$ shown in panel (b) of Figure A.2. The color scale is the same as in Figure 2. Dark links represent negative investment. With labor mobility, brighter nodes represent larger population increase. Panel (d) shows a discretized version of the TEN-T Core Network Corridors of the Trans-European Transport Network, based on information available at http://ec.europa.eu/transport/infrastructure/tentec/tentec-portal/site/en/maps.html.

air, and inland waterways. The TEN-T defines a core network of "strategic importance" for future investments based on criteria such as eliminating bottlenecks or following suggestions from member states.[41] Panel (d) of Figure 6 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 southwest of France, north of Spain, and Portugal; and the connection between Germany and Denmark. However, we also see some

[41]See https://ec.europa.eu/transport/themes/infrastructure/ten-t-guidelines/maps_en. The planning guidelines are mentioned in the European Commission working document, available in https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52013SC0542&from=EN.

differences, as the solution to our planning problem does not identify the need to invest in roads connecting the southeast of France to the south of Spain and Portugal.

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 across commodities, and an optimal network design. It nests standard 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.

Our approach using a global planner is particularly well suited for environments where agglomeration spillovers may not be too strong. We have also shown how in principle the model could be used in cases with spillovers. Due to current limitations in computing power and given the level of geographic detail that we handle, we have only applied the model to cases with a limited number of commodities.

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. We have refrained from identifying sources of misallocation, but it would be interesting to know the role of regional characteristics such as institutional quality. 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 past investments corresponding to dated economic fundamentals. We have also 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, Chaney, Coşar, and Hortaçsu (2019), who used the prediction of gravity models to infer the location of cities, the model could be used to infer the location of roads in historical data. The framework might also be used to construct instruments for investments in transport infrastructure.

Finally, a number of forces such as commuting or dynamic adjustment were left out of our analysis. We believe these are all interesting avenues to pursue in future research.

REFERENCES

AHUJA, R. K., T. L. MAGNANTI, AND J. B. ORLIN (1989): "Chapter IV Network Flows," in *Handbooks in Operations Research and Management Science*, Vol. 1, 211–369. [1415]

ALDER, S. (2019): "Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development," Manuscript, University of North Carolina, Chapel Hill. [1414]

ALLEN, T., AND C. ARKOLAKIS (2014): "Trade and the Topography of the Spatial Economy," *The Quarterly Journal of Economics*, 129, 1085–1139. [1413,1420,1430,1441]

——— (2019): "The Welfare Effects of Transportation Infrastructure Improvements," Manuscript, Dartmouth and Yale. [1414,1422]
ALLEN, T., C. ARKOLAKIS, AND Y. TAKAHASHI (2019): "Universal Gravity," *Journal of Political Economy*. [1422]
ANDERSON, J. E., AND E. VAN WINCOOP (2003): "Gravity With Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 93 (1), 170–192. [1413,1420]
——— (2004): "Trade Costs," *Journal of Economic Literature*, 42 (3), 691–751. [1411,1438]
ARROW, K. J., AND A. C. ENTHOVEN (1961): "Quasi-Concave Programming," *Econometrica: Journal of the Econometric Society*, 29, 779–800. [1423]
ASTURIAS, J., M. GARCÍA-SANTANA, AND R. RAMOS (2019): "Competition and the Welfare Gains From Transportation Infrastructure: Evidence From the Golden Quadrilateral of India," *Journal of the European Economic Association*, 17 (6), 1881–1940. [1415]
ATKESON, A., AND A. T. BURSTEIN (2010): "Innovation, Firm Dynamics, and International Trade," *Journal of Political Economy*, 118 (3), 433–484. [1422]
ATKIN, D., AND D. DONALDSON (2015): "Who's Getting Globalized? The Size and Implications of Intra-National Trade Costs," Technical Report, National Bureau of Economic Research. [1411]
BANAVAR, J. R., F. COLAIORI, A. FLAMMINI, A. MARITAN, AND A. RINALDO (2000): "Topology of the Fittest Transportation Network," *Physical Review Letters*, 84 (20), 4745. [1425]
BARJAMOVIC, G., T. CHANEY, K. A. COŞAR, AND A. HORTAÇSU (2019): "Trade, Merchants, and the Lost Cities of the Bronze Age," *The Quarterly Journal of Economics*, 134 (3), 1455–1503. [1449]
BARTELME, D. (2015): "Trade Costs and Economic Geography: Evidence From the U.S.," Technical Report, University of Michigan. [1414]
BAUM-SNOW, N. (2007): "Did Highways Cause Suburbanization?," *The Quarterly Journal of Economics*, 122, 775–805. [1415]
BECKMANN, M. (1952): "A Continuous Model of Transportation," *Econometrica: Journal of the Econometric Society*, 20, 643–660. [1414]
BERNOT, M., V. CASELLES, AND J.-M. MOREL (2009): *Optimal Transportation Networks: Models and Theory*, Vol. 1955. Springer Science & Business Media. [1414,1425]
BERTSEKAS, D. P. (1998): *Network Optimization: Continuous and Discrete Models*. CiteSeer. [1414,1420]
BOYD, S., AND L. VANDENBERGHE (2004): *Convex Optimization*. Cambridge University Press. [1428]
BRANCACCIO, G., M. KALOUPTSIDI, AND T. PAPAGEORGIOU (2019): "Geography, Transportation, and Endogenous Trade Costs," *Econometrica*. [1438]
BRANDT, L., T. TOMBE, AND X. ZHU (2013): "Factor Market Distortions Across Time, Space and Sectors in China," *Review of Economic Dynamics*, 16 (1), 39–58. [1415]
BURSTEIN, A., AND J. CRAVINO (2015): "Measured Aggregate Gains From International Trade," *American Economic Journal: Macroeconomics*, 7 (2), 181–218. [1422]
CALIENDO, L., F. PARRO, E. ROSSI-HANSBERG, AND P.-D. SARTE (2017): "The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy," *The Review of Economic Studies*, 85 (4), 2024–2096. [1414]
CARLIER, G. (2010): "Optimal Transportation and Economic Applications," Technical Report, IMA. [1414]
CHANDRA, A., AND E. THOMPSON (2000): "Does Public Infrastructure Affect Economic Activity?: Evidence From the Rural Interstate Highway System," *Regional Science and Urban Economics*, 30 (4), 457–490. [1415]
CHANEY, T. (2014a): "The Network Structure of International Trade," *The American Economic Review*, 104 (11), 3600–3634. [1413]
——— (2014b): "Networks in International Trade," in *The Oxford Handbook of the Economics of Networks*. [1413]
COLLIER, P., M. KIRCHBERGER, AND M. SÖDERBOM (2016): "The Cost of Road Infrastructure in Low- and Middle-Income Countries," *The World Bank Economic Review*, 30 (3), 522–548. [1440]
COMBES, P.-P., AND M. LAFOURCADE (2005): "Transport Costs: Measures, Determinants, and Regional Policy Implications for France," *Journal of Economic Geography*, 5 (3), 319–349. [1435]
COŞAR, A. K., AND B. DEMIR (2016): "Domestic Road Infrastructure and International Trade: Evidence From Turkey," *Journal of Development Economics*, 118, 232–244. [1415]
COSTINOT, A., AND A. RODRÍGUEZ-CLARE (2013): "Trade Theory With Numbers: Quantifying the Consequences of Globalization," in *Handbook of International Economics*, Vol. 4. [1413]
COUTURE, V., G. DURANTON, AND M. A. TURNER (2018): *Speed. Review of Economics and Statistics*, 100 (4), 725–739. [1439]
DESMET, K., AND E. ROSSI-HANSBERG (2013): "Urban Accounting and Welfare," *American Economic Review*, 103 (6), 2296–2327. [1415]
DJANKOV, S., C. FREUND, AND C. S. PHAM (2010): "Trading on Time," *The Review of Economics and Statistics*, 92 (1), 166–173. [1438]

DOLL, C., H. VAN ESSEN et al. (2008): "Road Infrastructure Cost and Revenue in Europe," Produced within the study Internalisation Measures and Policies for all external cost of Transport (IMPACT) P Deliverable 2. [1435]

DONALDSON, D. (2015): "The Gains From Market Integration," *Annual Review of Economics*, 7 (1), 619–647. [1413]

——— (2018): "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, 108 (4–5), 899–934. [1415]

DONALDSON, D., AND R. HORNBECK (2016): "Railroads and American Economic Growth: A Market Access Approach," *The Quarterly Journal of Economics*, qjw002. [1413,1414]

DURANTON, G., AND M. A. TURNER (2011): "The Fundamental Law of Road Congestion: Evidence From US Cities," *The American Economic Review*, 101 (6), 2616–2652. [1442]

DURANTON, G., P. M. MORROW, AND M. A. TURNER (2014): "Roads and Trade: Evidence From the US," *The Review of Economic Studies*, 81 (2), 681–724. [1415]

EATON, J., AND S. KORTUM (2002): "Technology, Geography, and Trade," *Econometrica*, 70 (5), 1741–1779. [1413]

EKELAND, I. (2010): "Notes on Optimal Transportation," *Economic Theory*, 42 (2), 437–459. [1414]

FABER, B. (2014): "Trade Integration, Market Size, and Industrialization: Evidence From China's National Trunk Highway System," *The Review of Economic Studies*, rdu010. [1415]

FAJGELBAUM, P. D., E. MORALES, J. C. SUÁREZ SERRATO, AND O. ZIDAR (2018): "State Taxes and Spatial Misallocation," *The Review of Economic Studies*, 86 (1), 333–376. [1415]

FAJGELBAUM, P. D, AND E. SCHAAL (2020): "Supplement to 'Optimal Transport Networks in Spatial Equilibrium'," *Econometrica Supplemental Material*, 88, https://doi.org/10.3982/ECTA15213. [1413,1416]

FELBERMAYR, G. J., AND A. TARASOV (2015): "Trade and the Spatial Distribution of Transport Infrastructure." [1414]

FERNALD, J. G. (1999): "Roads to Prosperity? Assessing the Link Between Public Capital and Productivity," *American Economic Review*, 89, 619–638. [1415]

FIRTH, J. (2017): "I've Been Waiting on the Railroad: The Effects of Congestion on Firm Production." [1438]

GALICHON, A. (2016): *Optimal Transport Methods in Economics*. Princeton University Press. [1414,1420]

HEAD, K., AND T. MAYER (2014): "Gravity Equations: Workhorse, Toolkit, and Cookbook," in *Handbook of International Economics*, Vol. 4. [1438,1441]

HELPMAN, E. (1998): "The Size of Regions: Transport and Housing as Factors in Agglomeration," in *Topics in Public Economics*, ed. by D. Pines, E. Sadka, AND I. Zilcha. Cambridge: Cambridge University Press, 33–54. [1420]

HSIEH, C.-T., AND P. J. KLENOW (2009): "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 124 (4), 1403–1448. [1415]

HSIEH, C.-T., AND E. MORETTI (2019): "Housing Constraints and Spatial Misallocation," *American Economic Journal: Macroeconomics*, 11 (2), 1–39. [1415]

HUMMELS, D. L., AND G. SCHAUR (2013): "Time as a Trade Barrier," *American Economic Review*, 103 (7), 2935–2959. [1438]

KANTOROVICH, L. V. (1942): "On the Translocation of Masses," *Dokl. Akad. Nauk SSSR*, 37, 199–201. [1414]

KRUGMAN, P. (1991): "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99 (3), 483–499. [1420]

LAI, E. L.-C., H. FAN, AND H. S. QI (2015): "Global Gains From Reduction of Trade Costs." [1422]

LIMAO, N., AND A. J. VENABLES (2001): "Infrastructure, Geographical Disadvantage, Transport Costs, and Trade," *The World Bank Economic Review*, 15 (3), 451–479. [1411]

LLANO, C., A. ESTEBAN, J. PÉREZ, AND A. PULIDO (2010): "Opening the Interregional Trade Black Box: The c-Intereg Database for the Spanish Economy (1995–2005)," *International Regional Science Review*. [1440]

MAIBACH, M., C. SCHREYER, D. SUTTER, H. VAN ESSEN, B. BOON, R. SMOKERS, A. SCHROTEN, C. DOLL, B. PAWLOWSKA, AND M. BAK (2013): "Handbook on Estimation of External Costs in the Transport Sector." [1417]

MARTINCUS, C. V., J. CARBALLO, AND A. CUSOLITO (2017): "Roads, Exports and Employment: Evidence From a Developing Country," *Journal of Development Economics*, 125, 21–39. [1415]

MOHRING, H., AND M. HARWITZ (1962): "Highway Benefits: An Analytical Framework," Technical Report, Northwestern University. [1427]

MONGE, G. (1781): *Mémoire sur la théorie des déblais et des remblais*. De l'Imprimerie Royale. [1414]

NAGY, D. (2016): "City Location and Economic Development," Manuscript, CREI. [1414]

NESTEROV, Y., AND A. NEMIROVSKII (1994): *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM. [1428]

NEWBERY, D. M. (1988): "Road Damage Externalities and Road User Charges," *Econometrica: Journal of the Econometric Society*, 56, 295–316. [1417]
RAMONDO, N., A. RODRÍGUEZ-CLARE, AND M. SABORÍO-RODRÍGUEZ (2016): "Trade, Domestic Frictions, and Scale Effects," *American Economic Review*, 106 (10), 3159–3184. [1414]
REDDING, S. J. (2016): "Goods Trade, Factor Mobility and Welfare," *Journal of International Economics*, 101, 148–167. [1414,1420]
REDDING, S. J., AND E. ROSSI-HANSBERG (2017): "Quantitative Spatial Economics," *Annual Review of Economics*, 9, 21–58. [1413]
REDDING, S. J., AND M. A. TURNER (2015): "Transportation Costs and the Spatial Organization of Economic Activity," in *Handbook of Regional and Urban Economics*, Vol. 5. [1413]
RESTUCCIA, D., AND R. ROGERSON (2008): "Policy Distortions and Aggregate Productivity With Heterogeneous Establishments," *Review of Economic Dynamics*, 11 (4), 707–720. [1415]
ROBACK, J. (1982): "Wages, Rents, and the Quality of Life," *The Journal of Political Economy*, 90, 1257–1278. [1420]
SANTAMBROGIO, F. (2015): *Optimal Transport for Applied Mathematicians*. NY: Birkäuser. [1420]
SOTELO, S. (2016): "Domestic Trade Frictions and Agriculture," Manuscript, University of Michigan. [1414]
SWISHER IV, S. (2015): "Reassessing Railroads and Growth: Accounting for Transport Network Endogeneity," Technical Report. [1414]
TREW, A. W. (2016): "Endogenous Infrastructure Development and Spatial Takeoff." [1414]
VILLANI, C. (2003): *Topics in Optimal Transportation*, Vol. 58. American Mathematical Soc. [1414]
WINSTON, C. (1985): "Conceptual Developments in the Economics of Transportation: An Interpretive Survey," *Journal of Economic Literature*, 23 (1), 57–94. [1417]

Co-editor Liran Einav handled this manuscript.

Manuscript received 27 March, 2017; final version accepted 9 December, 2019; available online 5 February, 2020.