Valuing Domestic Transport Infrastructure: A View from the Route Choice of Exporters*

Jingting Fan† Yi Lu‡ Wenlan Luo§

This version: January, 2020

Abstract

A key input to quantitative evaluations of transport infrastructure projects is their impact on transport costs. This paper proposes a new method of estimating this impact relying on the widely accessible customs data: by using the route choice of exporters. We combine our method with a spatial equilibrium model to study the aggregate effects of the massive expressway construction in China between 1999 and 2010. We find that the construction brings 5.6% welfare gains, implying a net return to investment of 180%. Our analysis also produces some intermediate output of independent interest, for example, a time-varying IV for city-sector export.

JEL codes: R13, R42, F14

1 Introduction

In 2016, the 47 member countries of the International Transport Forum—including OECD countries and China, among others—reported a total of over 850 billion euro investment in inland transport infrastructure (OECD, 2019). In China, the focus of this paper, the investment in inland transport infrastructure as a percent of GDP increased steadily from 2% in 2000 to 5% in recent years. China alone accounted for more than half of the total investment among the 47 countries. The sheer size of the investment in China and elsewhere has motivated many researchers to estimate the benefits of transport infrastructure. While earlier studies either adopt a measurement approach (e.g., Fogel, 1964) or a reduced-form approach (e.g., Banerjee et al., 2012), aided by new tools from international trade and spatial economics, a growing strand of literature is developing computational models to evaluate transport infrastructure through quantitative experiments.

*For helpful comments we thank Treb Allen, David Baqaee, Lorenzo Caliendo, Kerem Cosar, Fernando Parro, Nathaniel Young, and participants at 2018 Nankai University International Economics Workshop, 2019 Hong Kong University Globalization and Firm Dynamics Workshop, and Fudan University.

†Pennsylvania State University, PA, USA; jxf524@psu.edu

‡Tsinghua University, Beijing, China; luyi@sem.tsinghua.edu.cn

§Tsinghua University, Beijing, China; luowenlan@gmail.com

A key input into such quantitative exercises is the mapping from the travel distance/time along the transport network to trade cost. Two approaches of estimating this elasticity feature prominently in the literature.[1] The first approach is based on bilateral shipment data, such as the Commodity Flow Survey in the U.S. (Allen and Arkolakis, 2014, 2019). The second approach relies on price data. The idea is that under maintained assumptions on cost pass-through, the differences in the price of the same good across locations can be used to recover trade costs (e.g., Donaldson, 2018; Atkin and Donaldson, 2015; Asturias et al., 2018).[2]

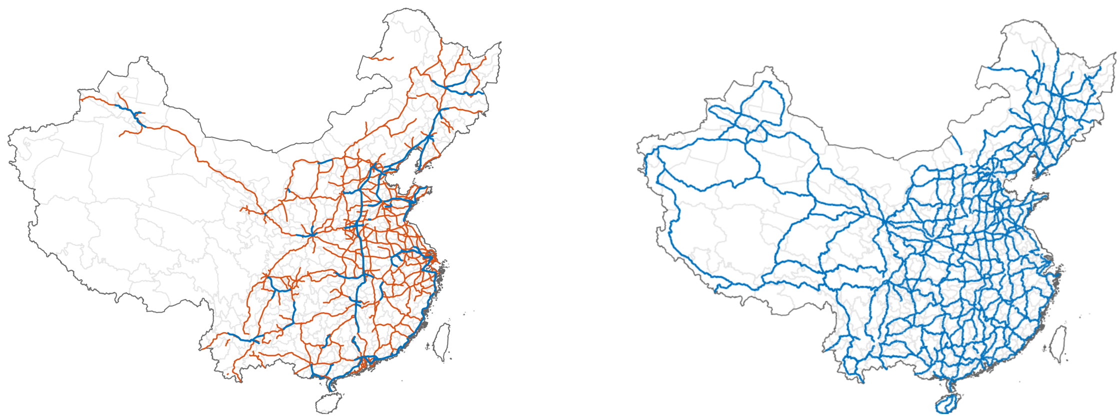
The data requirements of both approaches are quite demanding. Indeed, many countries do not collect or make accessible their versions of the Commodity Flow Survey; in the U.S., the surveys only started in 1993, when the inter-state highway system had been virtually completed. Perhaps for this reason, most existing studies using U.S. data or similar data from other countries rely on cross-sectional variations in bilateral road distance to estimate the trade cost elasticity. The price-based approach requires products to be homogeneous, so its application has been limited to agricultural commodities or goods identified by a unique producer or through bar codes.

This paper makes three contributions. First, we propose a methodology to estimate domestic trade costs using information contained in typical customs data. We estimate a routing model for structural parameters that determine the response of shipment to transport infrastructure, exploiting the *over-time* variations stemming from the expansion of the expressway network. Our design controls for bilateral fixed effects, which purge out other unobserved barriers to trade that are likely correlated with distance but unamenable to transport infrastructure. Second, we embed the estimate in a rich spatial equilibrium model with three ingredients—regional specialization, sector heterogeneity in trade costs, and intermediate inputs—to evaluate the effects of the $50,000$ kilometer expressways built in China between 1999 and 2010. Finally, we provide a second order characterization of the aggregate welfare effects of expressway networks and demonstrate its accuracy in our setting.

Our empirical design takes advantage of the increasingly available customs data. Like those of many other countries, the Chinese customs data contain the city of exporters and the ports from which they ship to foreign customers. Fractions of a city's export through different ports reflect, among potential confounding factors, the costs of transport routes through these ports. All else equal, if an inland city A ships most of its export via port B, then the routes passing through B likely incur lower costs than others. A direct application of this intuition to the data is subject to several sources of biases. First, the decision to export through a port might be driven by an unobserved connection with the port. Second, the total cost along an export route consists of costs along both its domestic and international segments. If the two components are negatively correlated, which would be the case if the data are generated by exporters minimizing the *total* cost, attributing the differences in port choice entirely to the domestic transport network exaggerates

[1] This elasticity governs how trade costs respond to travel time or distance and should be differentiated from trade elasticity, which governs how trade flows respond to trade costs.

[2] As an exception, Donaldson (2018) uses both price and shipment information to identify trade elasticity and the distance elasticity of trade cost, exploiting over-time variations from railroad construction.



(a) Expressway Network Expansion in China: 1999-2010 (b) Regular Road Network

Figure 1: Expressway and Regular Road Networks in China

Note: The left panel plots China's expressway network in 1999 (blue) and in 2010 (red); the right panel plots China's regular road network. Regular roads depicted are from 2007 and include 'national road' (or 'general highway') and 'provincial road'.

its importance.

We address both concerns by exploiting changes in bilateral trade costs resulting from the rapid expressway expansion in China between 1999 and 2010. As Figure 1a shows, over the decade, the expressway network grew from a few lines in the center and the southeast coast to covering most of the country, greatly supplementing China's existing regular road network, drawn in Figure 1b.[3] We estimate the relationship between the effective road distance between an inland city and a port, and the fraction of export of the city shipped through that port, controlling for city-port, city-time, and port-time fixed effects. We find that each additional hundred kilometer effective distance reduces the probability of exporting from a port by 15%. Not controlling for city-port fixed effects doubles this estimate. If a researcher had taken the same data and interpreted the cross-sectional estimate as trade cost elasticity, her estimate would have a significant upward bias.

A remaining concern is that the expressway network *expansion* might be endogenous to shipment between pairs of cities. We adopt an IV approach using the insight of Banerjee et al. (2012) and Faber (2014) that the expressways were planned to connect the major cities. We first construct the minimum-length network among all networks that connect the major cities. We then calculate the distance along the shortest path on this hypothetical network and use it as an IV for the distance on the actual network. This IV, together with the restriction of the sample to non-major cities, addresses the route endogeneity concern.

We embed the empirical design in a spatial equilibrium model consisting of Chinese prefecture

[3]'Expressway', or 'high-grade highway', refers to paved roads that are divided, fully enclosed, and not subject to traffic lights. 'Regular road' includes 'national' and 'provincial' roads, both of which have paved surfaces and are in general not enclosed. 'National road' is sometimes referred to as 'general highway'. In the rest of this paper, we use highway and expressway to refer to the enclosed road shown in Figure 1a. Between 1999 and 2010, most of the investment in inter-city road infrastructure was made to expressway. In fact, the network of the regular roads in 2010 is almost the same as that in 1999.

cities and the rest of the world (RoW). The model includes a routing block mapping road networks into trade costs. This routing block builds on Allen and Arkolakis (2019) but differs in two aspects: first, we allow flexible combinations of road segments from two co-existing networks, one of regular roads and one of expressways, in forming a route; second and more importantly, we allow trade costs to be higher for heavier sectors. We take advantage of the unit value information in the customs data, available for a wide range of narrowly defined products, to pin down the elasticity of sectoral trade cost in sectoral weight-to-value ratio. Our estimation finds this elasticity to be 0.3. The ad-valorem transport costs are 7.4% per hundred kilometer on regular roads and 5.5% on expressways, implying 25% cost savings from expressways. We parameterize the rest of the model to match the data on sectoral production and international trade, along with the average distance of domestic shipments. The former set of data determines regional productivity; the latter statistic pins down the level parameter in the domestic trade cost specification.

Armed with the parameterized model, we quantify the aggregate impacts of the expressway construction in China between 1999 and 2010. We find that the aggregate welfare gains from these expressways are around 5.6%. Deflated using the inflation rate for capital, the cumulative investment accounts for about 10% of the 2010 GDP. The sum of discounted gains far exceeds this cost and implies a net return of 180%. By reducing domestic trade frictions, the network increases domestic trade by 11% and export by 16% (both as shares of GDP). The latter accounts for around 25% of the actual increase in export intensity during this period. Finally, we assess the returns to the 14 mega projects that make up the backbone of the expressway network. The net returns to all these projects are positive, but also heterogeneous. Projects connecting the north and south of the country tend to generate higher returns, whereas projects connecting the hinterland to coastal ports like Shanghai and Fuzhou have the largest effects on export.

Restricted versions of the model without the three elements—-regional specialization, sector heterogeneity in trade costs, and intermediate inputs—predict significantly smaller welfare gains. Overlooking trade costs heterogeneity and intermediate inputs always underestimates gains because models without these two channels predict lower values of inter-regional trade, which to the first order determine the gains from trade cost reductions.[4] Overlooking regional specialization will predict a different spatial distribution of trade and generally has an ambiguous effect on the inferred welfare gains. In our setting, however, because the full model predicts higher fractions of trade along the routes that actually received expressway investment, it predicts larger gains.[5] When all three ingredients are omitted, the model inferred welfare gains are only a tenth of the actual gains, which turns the net return into negative. This stark difference highlights the importance of using the full model for evaluating large transport infrastructure projects.

Our evaluation are based on counterfactual experiments from the model which, while captur-

[4]From the international trade literature (Caliendo and Parro, 2015; Costinot and Rodríguez-Clare, 2014; Baqaee and Farhi, 2019a), it is well known that overlooking input-output linkages underestimates the gains from trade cost reductions because it infers lower trade over value added ratio. The role of sector heterogeneity in trade costs appears to be novel.

[5]In other words, according to the model without specialization, some of the investment was made in the wrong place, whereas in the full model the choice was wiser.

ing all general equilibrium interactions, is computationally costly and less suitable for comparison among a large number of proposed projects; an alternative approach that focuses on the first order effect (e.g. Hulten, 1978; Small, 2012) is convenient but overlooks endogenous responses of agents, which could be important especially under our estimate where addition of an expressway segment represents a large change in transport cost. In the last section of the paper, we propose a second order correction to the first order approach and demonstrate how it improves the accuracy in evaluating both individual expressway segments and large projects with many interacting segments. This characterization allows researchers to conveniently compare alternative expressway projects without the need to solve for full counterfactual equilibria.

Our paper is related to a number of different strands of literature. First, our work contributes to the growing research on spatial equilibrium models (e.g. Redding, 2016; Allen and Arkolakis, 2014; Caliendo et al., 2014; see Redding and Rossi-Hansberg, 2017 for a review). Within this literature, this paper is most closely related to a recent strand of work answering different questions on the effects of transport infrastructure projects, such as their impacts on regional development or growth (Fajgelbaum and Redding, 2014; Nagy, 2016; Cosar et al., 2019), market power of firms (Asturias et al., 2018), migration (Morten and Oliveira, 2018), distribution of activities within a city (Tsivanidis, 2018; Severen, 2018; Gu et al., 2018), match between buyers and suppliers (Xu, 2018), and the optimality of the transport network for the aggregate economy (Fajgelbaum and Schaal, 2019; Alder and Kondo, 2019).[6] We make two contributions to this line of research. First, we propose a new way of estimating the impacts of domestic transport infrastructure on trade costs using widely accessible customs data, which can be applied in other settings where domestic shipment or bar-code level price data are unavailable. Second, we characterize the second order effects for evaluating local and large projects (see also Baqaee and Farhi, 2019b). It offers a middle ground between the convenient but sometimes less accurate first order approach, and the more accurate but computationally intensive approach that solves full counterfactual equilibria.

At the core of our empirical analysis is the idea that export routes contain information on domestic transport infrastructure. We are not the first to recognize this. For example, Limao and Venables (2001) shows the importance of domestic infrastructure on export in a cross-country setting; Coşar and Demir (2016) and Martincus et al. (2017) use micro data to show road investment increases export. Instead, we combine the data and an equilibrium routing model to recover the deep parameters governing transport costs. Fan (2019) uses the gradient of city-level export in the city's distance to the nearest port as a subset of the moments identifying domestic trade costs. This paper is different in the explicit modeling of transport infrastructure and in the use of *route* of export—rather than export itself—for identification.

Finally, this paper adds to the rapidly growing literature on the spatial economy of China (Coşar and Fajgelbaum, 2016; Tombe and Zhu, 2019; Fan, 2019; Ma and Tang, 2019; Zi, 2016). In addition to carefully evaluating the expressway construction, which is important in its own right,

[6] A related literature estimates the effect of infrastructure on regional income/growth, using either heuristic or theory-based measures of exposure (see, e.g., Baum-Snow et al., 2016; Donaldson and Hornbeck, 2016; Hsu and Zhang, 2014).

our analyses draw some general lessons. First, the costs of moving goods across space are central to the predictions of many of these studies. Most existing work uses either railway shipments, which account for just 10% of total shipments and are available only at the provincial-pair level, a level too crude for studying transport infrastructure, or relies on regional input-output tables, which are imputed from the railway shipment data (see Zhang and Qi, 2012 for a description of procedures). Using new and more granular data, we generate predictions for domestic and international trade costs for 1999 and 2010, which can serve as inputs into future work in this area. Second, as a validation, we show that the model-predicted export growth in response to the exogenous component of the expressway expansion is strongly correlated with the actual growth in this period. Under suitable assumptions, the model-simulated export can serve as a time-varying IV for export at the city-sector level.[7]

The rest of the paper is organized as follows. Section 2 develops a routing model. Section 3 offers a first look at the data and provides some reduced-form estimates independent of the rest of the model. Sections 4 and 5 embed the routing block into a general equilibrium framework and bring in additional data to parameterize the model. Sections 6 and 7 perform counterfactual experiments and compare the results to alternative models and approaches. Section 8 concludes.

2 Route Choice on the Transport Network

In this section, we develop a routing model and derive a structural equation to take to the data.

2.1 From Route Cost to Trade Cost on a Single Network

We describe the model using the example illustrated in Figure 2. Each node represents a city, connected by edges that represent the transport network. We use ι_{od} to denote the travel cost along the edge $o \to d$. Costs along any edges are greater than 1 and symmetric: $\iota_{od} >= 1$ and $\iota_{od} = \iota_{do}, \forall o \neq d$. A path, or a route, is a set of inter-connected edges that links an origin to a destination; the cost it takes to travel along a path is the product of costs of the segments it is composed of. For example, $o \to k \to d$ forms a path from o to d; the cost along this path is $\iota_{ok} \cdot \iota_{kd}$.[8]

A truck driver going from o to d chooses among multiple feasible paths. There is a single one-edge path which costs ι_{od}; there are also two two-edge paths: $o \to k \to d$ and $o \to l \to d$. Drivers derive idiosyncratic dis-utility ν from each potential path, drawn from a Fréchet distribution with dispersion parameter θ and location parameter zero. The effective transport cost along a path is the product of the travel cost and the path-specific realization of ν. For example, the travel cost of $o \to k \to d$ is $\iota_{ok}\iota_{kd}\nu$.

[7] This IV would exploit the changes in access to foreign markets across regions driven by expressway construction and complements existing identification strategies in estimating the effects of export. For example, a strand of literature exploits variations across industries from the reductions in the level or uncertainty of exporting tariffs after the WTO accession of China to estimate the effects of export (see, e.g., Facchini et al., 2019; Tian, 2019).

[8] This multiplicative structure arises because the shipment cost along each segment is specified as ad-valorem. Allen and Arkolakis (2019) uses a similar structure and shows that it fits bilateral shipment data in the U.S. well.

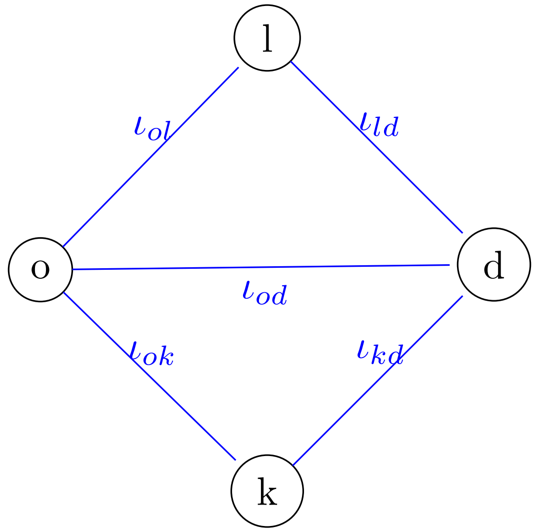


Figure 2: Routing on a Network: Four-Region Example

Note: This diagram illustrates the routing problem in a four-region example. In this example, except for nodes l and k, all other pairs of nodes are directly connected by a road segment.

The driver chooses the path that gives the lowest effective travel cost. If the three paths discussed above are the only options, the Fréchet assumption implies that the average effective trade cost between o and d across all possible realizations of $\nu(r)$ is:

$$\tau_{od,2} \equiv \mathbb{E}[\min_{r\in\{od,\ okd,\ old\}}\iota_r \cdot \nu(r)] = \Gamma(\frac{\theta-1}{\theta})\big([\iota_{od}]^{-\theta} + [\iota_{ol}\iota_{ld}]^{-\theta} + [\iota_{ok}\iota_{kd}]^{-\theta}\big)^{-\frac{1}{\theta}},\ o \neq d \qquad (1)$$

in which the subscript '2' in $\tau_{od,2}$ indicates that the choice is constrained to paths with two or fewer edges.

We derive the matrix representation for Equation (1). Consider the following adjacency matrix, with its elements being the $-\theta$ power of the cost between two adjacent nodes in the network.

$$\begin{array}{c c} & \begin{array}{cccc} o & l & d & k \end{array} \\ \begin{array}{c} o \\ l \\ d \\ k \end{array} & \begin{pmatrix} 0 & \iota_{ol}^{-\theta} & \iota_{od}^{-\theta} & \iota_{ok}^{-\theta} \\ \iota_{lo}^{-\theta} & 0 & \iota_{ld}^{-\theta} & 0 \\ \iota_{do}^{-\theta} & \iota_{dl}^{-\theta} & 0 & \iota_{dk}^{-\theta} \\ \iota_{ko}^{-\theta} & 0 & \iota_{kd}^{-\theta} & 0 \end{pmatrix} \end{array}$$

A zero in the matrix indicates that two cities are not directly connected by an edge, or that they are connected by an edge with an infinite transport cost.[9] Let $\mathbb{L}$ denote this matrix and $[\mathbb{L}_{(o,d)}]$ denote the (o,d) element of $\mathbb{L}$. The symmetry of transport costs implies $\mathbb{L} = \mathbb{L}'$.

Define $\mathbb{L}^2 \equiv \mathbb{L}\cdot\mathbb{L}$. The (o,d) element of $\mathbb{L}^2$, denoted by $[\mathbb{L}^2_{(o,d)}]$, equals $\sum_k \iota_{ok}^{-\theta}\cdot\iota_{kd}^{-\theta}$, which is

[9] We assume that the diagonal elements of the adjacency matrix are zero. Throughout the rest of this paper, we normalize the iceberg cost of trading within a city to be one.

the sum of ($-\theta$ power of) costs across all feasible paths with exactly two edges. We can now write Equation (1) as

$$\tau_{od,2} = \Gamma(\frac{\theta - 1}{\theta})\big([\mathbb{L}_{(o,d)}] + [\mathbb{L}^2_{(o,d)}]\big)^{-\frac{1}{\theta}},\ o \neq d \tag{2}$$

In addition to the three paths with two or fewer edges, the driver can in principle take a detour. For example, there are two three-edge paths from o to d: $o \to l \to o \to d$ and $o \to k \to o \to d$. The sum of the costs along these two three-edge paths is:

$$(\iota_{ol}\iota_{lo}\iota_{od})^{-\theta} + (\iota_{ok}\iota_{ko}\iota_{od})^{-\theta} = [\mathbb{L}^2_{(o,d)}] \cdot \iota_{o,d}^{-\theta} = [\mathbb{L}^3_{(o,d)}].$$

Therefore, if the driver is allowed to choose among all paths with three or fewer edges, the average travel cost between o and d across realizations of dis-utility shocks is:

$$\tau_{od,3} = \Gamma(\frac{\theta - 1}{\theta})\big([\mathbb{L}_{(o,d)}] + [\mathbb{L}^2_{(o,d)}] + [\mathbb{L}^3_{(o,d)}]\big)^{-\frac{1}{\theta}},\ o \neq d.$$

In larger networks with more nodes and edges, as drivers are free to take multiple detours, enumerating all possible paths is difficult. The above induction shows $[\mathbb{L}^n_{(o,d)}]$ represents the sum of all n-edge paths that goes from o to $d \neq o$, so the average transport costs across all possible paths is (assuming $(\mathbb{I} - \mathbb{L})$ is invertible):[10]

$$\tau_{od} \equiv \lim_{N\to\infty} \tau_{od,N} = \Gamma(\frac{\theta - 1}{\theta})\big(\sum_{i=1}^{\infty}[\mathbb{L}^i_{(o,d)}]\big)^{-\frac{1}{\theta}} = \Gamma(\frac{\theta - 1}{\theta})\big([(\mathbb{I} - \mathbb{L})^{-1}_{(o,d)}]\big)^{-\frac{1}{\theta}},\ o \neq d.$$

2.2 Combining Two Transport Networks

Our empirical application focuses on the expressway network expansion, which is an addition to—rather than the replacement of—the existing regular road network. As can be seen from Figure 1, by 2010, many adjacent cities are connected by expressways and regular roads at the same time. In these cases, both types of roads could be used.[11] Furthermore, drivers can combine the two networks to form a route to their own taste. For example, one person might prefer to go from o to l on regular roads, and then from l to d on expressways, while another person might choose expressways for both segments. We extend the previous probabilistic formulation of the transport problem to tractably accommodate these situations.

Let $\mathbb{H}$ and $\mathbb{L}$ denote the road matrix for expressways (H for High-speed) and regular roads (L for Low-speed), respectively, and let $(\iota^x_{od})^{-\theta}$, $x \in \{H,\ L\}$ be the (o,d) element of $\mathbb{H}$ and $\mathbb{L}$. Define $\mathbb{A}$ as the sum of the two matrices: $\mathbb{A} \equiv \mathbb{H} + \mathbb{L}$. As before, drivers choose among all possible paths subject to a path-specific idiosyncratic taste shock. But rather than being confined to either

[10] A sufficient condition for $(\mathbb{I} - \mathbb{L})$ to be invertible is that the spectral radius of $\mathbb{L}$ is less than one (Allen and Arkolakis, 2019). This will be case if the road network adjacency matrix is sparse and the routing elasticity θ is large, which holds in our structural estimates.

[11] Indeed, whereas expressways usually have a higher speed limit, they also charge more fees and are less flexible for entry and exit, so some drivers might prefer regular roads.

expressways or regular roads, the path can combine segments from both. In this case, the expected transport cost across the two one-edge path from o to d is:

$$\tau_{od,1} = \Gamma(\frac{\theta-1}{\theta})\big([\mathbb{H}_{(o,d)}] + [\mathbb{L}_{(o,d)}]\big)^{-\frac{1}{\theta}} = \Gamma(\frac{\theta-1}{\theta})\big([\mathbb{A}_{(o,d)}]\big)^{-\frac{1}{\theta}},$$

i.e., the average cost is simply the $-1/\theta$ power of the sum of the corresponding elements of the two networks. Similarly, if the driver were to choose among all possible paths with two or fewer edges, the average cost across realizations of the idiosyncratic shocks is

$$\begin{aligned}\tau_{od,2} &= \Gamma(\frac{\theta-1}{\theta})\big([\mathbb{H}_{(o,d)}] + [\mathbb{H}^2_{(o,d)}] + [\mathbb{L}_{(o,d)}] + [\mathbb{L}^2_{(o,d)}] + [(\mathbb{H}\cdot\mathbb{L})_{(o,d)}] + [(\mathbb{L}\cdot\mathbb{H})_{(o,d)}]\big)^{-\frac{1}{\theta}} \\ &= \Gamma(\frac{\theta-1}{\theta})\big([\mathbb{A}_{(o,d)}] + [\mathbb{A}^2_{(o,d)}]\big)^{-\frac{1}{\theta}}\end{aligned} \tag{3}$$

In the first line of Equation (3), $[(\mathbb{H}\cdot\mathbb{L})_{(o,d)}] = \sum_k (\iota^H_{ok}\iota^L_{kd})^{-\theta}$ is simply the sum of across all two-edge paths with the first segment being an expressway and the second being a regular road; analogously, $[(\mathbb{L}\cdot\mathbb{H})_{(o,d)}]$ is the sum across all paths with the first segment being a regular road and the second being an expressway. Equation (3) thus shows that to generalize Equation (2) to two networks, we can simply replace $\mathbb{L}$ with $\mathbb{A}$. More generally, we show in the appendix by induction that this result holds for when drivers are allowed to choose any combinations of regular roads and expressways with arbitrarily many edges. The expected trade costs across all such paths is

$$\tau_{od} \equiv \lim_{N\to\infty} \tau_{od,N} = \Gamma(\frac{\theta-1}{\theta})\big(\sum_{i=1}^{\infty}[\mathbb{A}^i_{(o,d)}]\big)^{-\frac{1}{\theta}} = \Gamma(\frac{\theta-1}{\theta})[\mathbb{B}_{(o,d)}]^{-\frac{1}{\theta}}, \text{ where } \mathbb{B} = (\mathbb{I}-\mathbb{A})^{-1}. \tag{4}$$

2.3 From Domestic Trade Cost to Port Choice of Exporters

A seller shipping to another location randomly meets with a driver and pays the expected transport cost before the idiosyncratic dis-utility shocks realize. This expected cost, given by Equation (4), is the trade cost between any two *domestic* locations, $o \neq d$. To use the export data to estimate domestic trade costs, we embed the routing block into an international shipment problem.

Imagine in an economy represented by Figure 3, an exporter from city o shipping one truckload of merchandises to foreign consumers. The total export cost consists of two components: cost from city o to one of the nation's ports l or d, denoted by $\tau_{ok}, k \in \{l, d\}$, and the cost from that port to the RoW, denoted by $\tau_{k,RoW}, k \in \{l, d\}$. To highlight that city o is not necessarily directly connected to port l or d, we indicate the two links using dotted lines.

The seller first decides from which port to ship the goods, taking the expected domestic transport cost as given. For each shipment, the seller receives a *port-specific* export taste shock, denoted by ν_F, drawn from a Fréchet distribution. This shock enters trade costs multiplicatively, so the international shipment cost from l to the RoW, for example, is $\tau_{l,RoW} \cdot \nu_F(l)$. Because the reason

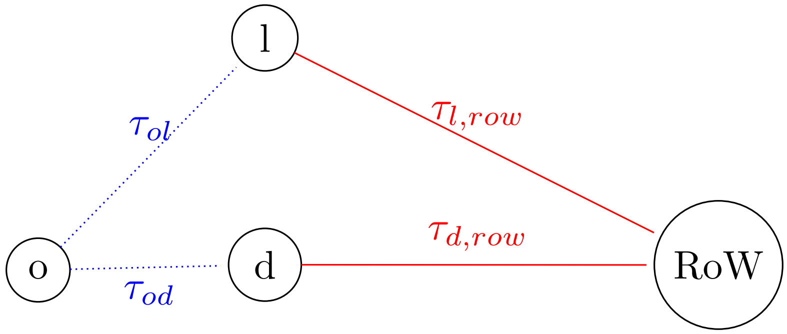


Figure 3: Port Choice of Exporters
Note: The diagram illustrates the choice of port through which to ship to the rest of world (RoW).

behind the idiosyncratic shock in international shipment might be different from that in domestic shipment, we allow the dispersion parameter of ν_F, θ_F, to be potentially different from θ.[12] The seller chooses $\min\{\tau_{ol}\tau_{l,RoW}\cdot\nu_F(l),\ \tau_{od}\tau_{d,RoW}\cdot\nu_F(d)\}$. Suppose port d is chosen, then, the seller randomly meets with a truck driver, who will find the minimum-cost route from o to d after his own realization of route-specific cost draws and charge the seller the expected cost.

The probability that the shipment is made via port d is:

$$\pi_{(o,RoW),d} = \frac{\tau_{od}^{-\theta_F}\cdot\tau_{d,RoW}^{-\theta_F}}{\sum_{\text{All ports k}}\tau_{ok}^{-\theta_F}\cdot\tau_{k,RoW}^{-\theta_F}}. \tag{5}$$

Equation (5) illustrates how the export data help identify domestic trade costs. All else equal, if port d is closer or better connected to city o through the domestic transport network (lower τ_{od}), exporters in city o will be more likely to ship via port d.

2.4 Parameterizing Road Network Matrices

In parameterizing road network matrices, as Allen and Arkolakis (2019) we assume that the travel cost between two adjacent cities k and l along the edge $k\to l$ is

$$\iota_{kl}^{x} = \exp(\kappa^x\cdot dist_{kl}), \tag{6}$$

in which $x\in\{H,L\}$ indicates the type of road, κ^x is the distance semi-elasticity of travel cost on an edge, and $dist_{kl}$ is the length of the edge connecting k and l. Matrix $\mathbb{A}$ is given by

$$[\mathbb{A}_{(k,l)}] = [\exp(-\theta\kappa^H\cdot dist_{kl}) + \exp(-\theta\kappa^L\cdot dist_{kl})]. \tag{7}$$

The structural parameters of the routing model are κ^H, κ^L, and θ. To bring out the connection

[12] While the heterogeneity in domestic shipment arises mainly from truck drivers' preference across routes, the choice of port likely depends on the routing of cargo ships, the export intermediary used, and the distance to the destination country, all of which we abstract from.

between the routing framework and our reduced-form analysis in the most straightforward manner, we consider the case of $\theta \to \infty$. In this limit, drivers' idiosyncratic utility draws play little role in the route choice, and the path that gives the lowest transport cost is always chosen.[13] Assuming $\kappa^H < \kappa^L$, then when both types of roads coexist between two adjacent cities, expressways will be chosen with probability one and the limit of route cost is $\lim_{\theta\to\infty}[\mathbb{A}_{(k,l)}]^{-1/\theta} = \exp(\kappa^H \cdot dist_{kl})$. The log of transport cost from o to d along a path of length $p : o \to m_1 \to m_2 \to ... \to m_{p-1} \to d$ is:

$$\sum_{i=0}^{p}[\mathbb{1}_{\left([\mathbb{H}_{(m_i,m_{i+1})}]=0,\ [\mathbb{L}_{(m_i,m_{i+1})}]>0\right)} \cdot \kappa^L + \mathbb{1}_{\left([\mathbb{H}_{(m_i,m_{i+1})}]>0\right)} \cdot \kappa^H] \cdot dist_{m_i,m_{i+1}}$$

in which $dist_{m_i,m_{i+1}}$ is the distance between node m_i and m_{i+1} (we label o as m_0 and d as m_p), and $\mathbb{1}$ is the indicator function. The log transport cost along any path is thus the sum of all segments weighted by whether the segment is regular only (κ^L), or contains expressway (κ^H).

The effective trade cost between o and d, τ_{od}, is the least costly path of all. Slightly abusing notation, we use $dist^E_{o\to d}$ and $dist^R_{o\to d}$ to denote the *total* length of expressways and regular roads along the shortest path, respectively. Then the trade cost between o and d is simply $\kappa^H dist^H_{o\to d} + \kappa^L dist^L_{o\to d}$. With this we can log transform Equation (5) to obtain:

$$\log(\pi_{(o,RoW),d}) = \beta_o + \beta_d - \theta_F(\kappa^H dist^H_{o\to d} + \kappa^L dist^L_{o\to d}). \tag{8}$$

In Equation (8), β_o and β_d capture characteristics of origin city o and port d, respectively. In the data, we observe the fraction of export shipments through each port d, $\pi_{(o,RoW),d}$. With measures of $dist^H$ and $dist^L$, we can estimate $\theta_F\kappa^H$ and $\theta_F\kappa^R$ directly. For transparency on identification, in the next section we provide direct evidence based on this specification; in Section 4, we relax the assumption of $\theta \to \infty$ and further extend the routing framework to accommodate non-road transport modes for quantitative exercises.

3 A First Look at the Data

3.1 Data and Measurements

The empirical analysis focuses on the change between 1999 and 2010. The massive expressway construction during this period provides variations in effective distance between cities and ports, which we will exploit. The following are our data sources.

Export routing. We measure the port choice of exporters using monthly transaction-level Chinese customs data. For each transaction, we observe the address of the exporter, the value and weight (when the unit of products is kg) of the shipment, and the customs office from which it is exported. We map the addresses of exporters and customs offices to prefecture cities, treating

[13] Although this limit case is intuitive, we use the probabilistic setup for quantification because it allows an analytical characterization of the second order effects of expressway construction.

the city of an exporter as the origin and the city of the customs office as the port.[14] The ideal measure for shipment is the number of trucks/containers. Without such information, we use the weight of shipments as a proxy and measure shipments from each origin city to the RoW through different Chinese ports at both the aggregate and the HS2 category level. Given that the variations in the expressway network are between 1999 and 2010, we construct a panel with two periods corresponding to the beginning and end of this decade.[15]

Transport network. We obtain the inter-city expressway maps for 1999 and 2010 from Baum-Snow et al. (2016), which digitized transport infrastructure for the entire mainland China from published maps.[16] We supplement their expressway maps with a map of regular roads for 2007 from the ACASIAN Data Center. Regular roads include 'National Road' and 'Provincial Road', which are paved, non-enclosed, non-divided roads, usually with two or four lanes. Because there are virtually no variations in regular roads during this period, we treat the regular road network as time-invariant.[17] The raw data are in the form of the coordinates of points on the roads. We convert each of the three maps into a matrix of cities (nodes) and links (edges) in two steps. Below we give an outline of the procedures; the details are described in the appendix.

In the first step, for each of the three road maps, we identify the list of cities (prefectures) connected to the network. A city is defined to be on a road network, or connected, if any segment of the road cuts through within 30 kilometer of the center of the city.

The second step focuses on cities on the road network and generates the adjacent matrix among them. For each connected city k, we check all geographically adjacent cities. If, say, a neighboring city l is also connected, then we draw a edge between node k and l and assign a value of ι_{kl} to the (k, l) element of the adjacent matrix; otherwise, k and l will not be connected by an edge, and the corresponding element on the matrix will be zero. ι_{kl} depends on the length of the edge, which is defined as the great circle distance between the two city centers. This effectively 'irons out' the road segments connecting each adjacent city. The result of this step is the matrix representation of each map. Figure A.1 in the appendix shows the original map (left panel) and the matrix representation (right panel).

We denote the matrices for the three maps $\mathbb{H}^{1999}$, $\mathbb{H}^{2010}$, and $\mathbb{L}$, respectively. For the reduced-form analysis in this section, in which we treat routes as perfect substitutes, the combined matrices are given by: $\mathbb{A}^{1999} = \max\{\mathbb{H}^{1999}, \mathbb{L}\}$ and $\mathbb{A}^{2010} = \max\{\mathbb{H}^{2010}, \mathbb{L}\}$. For quantitative analyses in the rest of this paper, we will use the theory-consistent definitions: $\mathbb{A}^{1999} = \mathbb{H}^{1999} + \mathbb{L}$ and

[14] It is possible for a shipment to be declared at the customs office in an inland city and sealed before it is shipped through a seaport to the RoW. To address this concern, our specifications focus on the set of customs cities that are seaports (see the appendix for the list of these cities).

[15] To reduce measurement errors we average between 2000-2001 for the beginning period and between 2009-2010 for the ending period. We do not have access to the 1999 customs data.

[16] Expressway is named 'high-grade highway' in their database, and refers to the same type of road as the 'National Trunk Highway System' studied in Faber (2014)

[17] Baum-Snow et al. (2016) also provides separate maps for 'general highway', which is of lower grade than 'high-grade highway', or expressway. The definition of 'general highway' is broad and generally includes 'national road', 'provincial road', and 'county road'. Because 'county road' is of much lower quality than 'national road' or 'provincial road', and because most inter-city transports rely on the latter two, we choose not to use Baum-Snow et al. (2016) to measure the regular road network.

$\mathbb{A}^{2010} = \mathbb{H}^{2010} + \mathbb{L}$.

Bilateral transport cost. With the combined networks constructed, we measure the lowest-transport cost between city pairs for 1999 and 2010. Among multiple paths connecting two cities composed of different combinations of regular roads and expressways, determining the least costly one requires the knowledge of $\frac{\kappa^H}{\kappa^L}$. We query the driving time between a random set of 2000 city pairs along expressways and regular roads separately on the Baidu Map, a Chinese search engine, and compare the average expected travel time of the two trips. Among these queries, the average speed on regular roads is about 55% of that on expressways, so we set $\frac{\kappa^H}{\kappa^L} = 0.5$. It means that each kilometer on expressways is equivalent to a half kilometer on regular roads. Under this assumption, we use the Dijkstra's algorithm to find the shortest path between each city pairs and measure the regular-road equivalent distance along the path.[18]

3.2 Expressway Construction and the Route Choice of Exporters

Our empirical exercises use various versions of the following specification, which comes out of Equation (8):

$$\ln(q^t_{(o,RoW),d}) = \beta_{od} + \beta^t_o + \beta^t_d + \gamma_1 \text{dist}^t_{od} + \epsilon^t_{od}. \tag{9}$$

The dependent variable of the specification, $q^t_{(o,RoW),d}$ is the total export from city o to the RoW through port city d in period t. β_{od}, β^t_o, β^t_d are fixed effects for city-port pair, city-time, and port-time, respectively. dist^t_{od} is the equivalent regular road distance along the shortest path. In some specifications, we will split dist^t_{od} into the expressway and regular road distance along the path, which will allow us to separately estimate their cost parameters.

The OLS estimator of specification (9) is subject to an obvious endogeneity concern. Cities closer to each other likely have stronger unobserved ties, which could attract shipments for reasons unrelated to the transport infrastructure. Through city-port fixed effects, we control for all time-invariant unobserved heterogeneity across pairs of cities. The identification thus comes from over-time changes in the effective distance along the shortest paths, resulting from the expansion of the expressway network. To the extent that the expressway construction might target specific cities and ports, this channel is captured by the city-time and port-time fixed effects.

The city-port fixed effects do not address the concern that new expressways might have been built to connect *pairs* of cities with *growing* unobserved economic ties. We adopt two strategies to alleviate this concern. The first is to exclude origin city o that is either a provincial capital city or otherwise had more than 5 million registered residents.[19] As discussed in Banerjee et al. (2012), the transport network in China was largely designed to link the major cities. With these cities excluded from the sample, our estimation exploits the increase in port access for the remaining,

[18] Alternatively, we can estimate κ^L and κ^H recursively using nonlinear least square as in Donaldson (2018): for a given level of κ^L and κ^H, find the shortest path between city pairs and generate bilateral shipping costs accordingly. Then search over the space of κ^L and κ^H to find the combination that minimizes some notions of prediction error—for example, the total deviation in the models' prediction on routing patterns from the data. We pursue a version of this exercise in the full structural estimation.

[19] We use the 2000 population census in defining this variable, according to which there were 55 large cities.

Figure 4: The Minimum Spanning Tree for the 2010 Expressway Network
Note: The minimum spanning tree is constructed as the network with the minimum total length that connects the 55 major cities that are on the actual 2010 expressway network.

smaller cities, which gained access because they were between major cities to be connected.

Second, in addition to excluding major cities, we adopt an IV strategy based on Faber (2014): to use an 'exogenous' hypothetical expressway network as an instrument for the actual network. Specifically, using the minimum-spanning tree algorithm, we first generate a minimum-length expressway network that connects all major cities that are on the actual network by 2010. We denote this matrix $\mathbb{H}^{2010,\text{hypothetical}}$ and define $\mathbb{A}^{2010,\text{ hypothetical}} = \max\{\mathbb{L}, \mathbb{H}^{2010,\text{hypothetical}}\}$. This represents the transport network configuration *if* the goal is to minimize the total length of expressway segments while still connecting the same set of major cities. From $\mathbb{A}^{2010,\text{ hypothetical}}$ we obtain a new measure of effective distance from cities to ports that is exogenous to *non-major* cities. This distance serves as an IV for the distance along the actual network in Equation (9). Figure 4 shows the hypothetical and actual networks.[20] The identification assumption is that non-major cities experienced an improvement in access to ports (and other cities) only because they were close to the minimum-length hypothetical network that connects the major cities.

3.3 Baseline results

Table 1 reports results from the benchmark specification. The dependent variable in all eight specifications is the log of total shipments (in weight) from city o to the RoW through port d. The specification in the first column includes only city, port, and time fixed effect, so the coefficient is identified off mostly cross-sectional variations. The point estimate suggests that each additional

[20]The expressway network was very sparse in 1999 (see Figure 1), so we use the distance along the 1999 actual network as an IV for itself. The IV is thus time varying and we can control for city pair fixed effects.

Table 1: Expressway and Routing of Export Shipments

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			OLS				PPML	
$dist_{o,d}$	-0.346***	-0.103***	-0.136***	-0.144***		-0.655***	-0.470***	
	(0.010)	(0.025)	(0.033)	(0.040)		(0.062)	(0.066)	
-on express					-0.082*			-0.286**
					(0.042)			(0.117)
-on regular					-0.148***			-0.488***
					(0.043)			(0.084)
Fixed Effects	o, d, t	od, t	od, ot, dt	od, ot, dt	od, ot, dt	ot, dt	od, ot, dt	od, ot, dt
Exclude major cities				yes	yes	yes	yes	yes
Observations	3625	2768	2738	2002	2002	2740	2002	2002
R^2	0.601	0.820	0.893	0.882	0.882	-	-	-

Notes: This table reports the regressions of export shipment through a port on the distance between the city and the port. The dependent variable is the log of total weight of goods exported in city o through port d to the RoW. The independent variables are the shortest equivalent regular road distance between city o and port d along the shortest path (Columns 1-4, Columns 6-7); and the separate length of expressways and regular roads along the shortest path (Columns 5 and 8). The specification of Columns 1 through 5 is ordinary least square; the specification of Columns 6 through 8 is Poisson Pseudo-Maximum Likelihood.
Standard errors are clustered at city-port level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

hundred kilometer equivalent regular road distance reduces the shipment through a port by 35%.

As discussed, the cross-sectional variations in routing are likely correlated with non-transport related barriers, such as information frictions, home biases, etc. To isolate the channel driven by transport infrastructure alone, in the second column, we include city-port fixed effects.[21] The point estimate shrinks by 70% and is rather precisely estimated. To rule out that the new expressway construction is correlated with city- or port-specific growth, in Column 3, we further control for city-time and port-time fixed effects. This leads to a modest increase in the coefficient.

In Column 4, our preferred specification, we further exclude all origin city o that are major cities. This restriction alleviates the concern that our finding is biased by the unobserved time-varying linkages between big cities and ports that also drove the expressway construction plan. The point estimate suggests that each additional hundred kilometer equivalent regular road distance decreases export through the port by 14.4%. This estimate is less than half of the coefficient in Column 1. This difference highlights the importance of isolating other unobserved bilateral linkages in identifying the role of transport infrastructure.

Our analysis so far focuses on the total effective distance, which is the weighted sum of the distance on expressways (weighted by 0.5) and regular roads. To investigate the relative costs of these two types of roads, in Column 5, we separate the total distance into the two subcomponents and estimate their respective costs. We find that the coefficient for the regular road is around -0.15, while the coefficient for the expressway is around -0.08—the former is close to twice as

[21] To the extent that that some of the non-transport barriers, such as information friction, also respond to additions to the transport network, it should and will be picked up by our estimation. What we would like to exclude through the addition of bilateral fixed effects is the component that does not respond to transport infrastructure.

large as the latter, consistent with our assumption in calculating the shortest path. The relative size of the coefficients also implies that for two adjacent cities that were already connected by regular roads, an additional expressway segment between them can reduce the trade cost significantly.

Columns 6 to 8 show that the general findings are robust when the Poisson Pseudo-Maximum Likelihood (PPML) method is used. Column 6 controls for city-time and port-time fixed effects and identifies mainly from the cross-sectional variations; Column 7 further controls for city-pair fixed effects. The coefficient shrinks by a statistically significant amount of 30%. Column 8 splits total effective distance into that for regular roads and expressways and finds the point estimate to be substantially larger for regular roads. Although the point estimates are generally larger with the PPML, the broad points stand.

It is instructive to discuss the implications of our results for domestic trade costs estimation in general. In a typical domestic trade setting without the routing block, researchers could have interpreted the shipments from origin cities to ports as trade flows. Specification (9) then corresponds to a gravity regression, with parameter γ_1 being the product of trade elasticity and the distance semi-elasticity of trade cost. For any given trade elasticity, the estimated $\hat{\gamma}_1$ maps one-to-one into the distance semi-elasticity of trade cost. Table 1 shows that in such settings, isolating the over-time variations in bilateral distance matters: not doing so would overestimate the key parameter by 100%.

3.4 IV and Additional Robustness

We conduct additional exercises to show the robustness of our estimates to the identification strategy and the choice of measurements. First, even though we have excluded major cities from our samples, it is still possible that an expressway zigzags locally to increase the accessibility of smaller cities. We adopt the IV strategy as described before to address this concern.

The first two columns of Table 2 report the IV estimates, controlling for the same set of fixed effects as in Columns 4 and 5 of Table 1. The high F-statistic in the first stage indicates relevance. The estimated coefficients for both the overall effective distance (Column 1) and the distance on expressways and regular roads (Column 2) are similar to those based on the OLS.

Another concern is that the results could be driven by changes in the sectoral compositions of city export. For example, if as cities gain access to ports, they also become more specialized in export-intensive industries, such as textile, and if for some reason, the export in the textile industry is concentrated among the ports that experienced disproportionate increases in expressway connectivity to the hinterland, then the correlation between the shipment share and the bilateral connectivity will be picked up by our regressions. Note that if the expressway expansion is truly exogenous to non-major cities, then this concern does not pose a threat to the IV estimate. Nevertheless, in Columns 3 through 5, we use shipments at the sectoral level for a robustness check. Column 3 includes city-port-sector, city-time, port-time, and sector-time fixed effects (letter 'i' in the row 'Fixed Effects' denote sectors). Column 4 further controls for city-sector-time and port-sector-time fixed effects. The point estimates in these specifications are both around -0.1, slightly

Table 2: IV and Sectoral-Level Regressions

	(1)	(2)	(3)	(4)	(5)
	Aggregate IV			Sectoral OLS	
dist	-0.156***		-0.092***	-0.110***	
	(0.050)		(0.030)	(0.037)	
-on express		-0.096			-0.088**
		(0.067)			(0.040)
-on regular		-0.164***			-0.120***
		(0.060)			(0.039)
Fixed Effects	*od, ot, dt*	*od, ot, dt*	*odi, ot, dt it*	*odi, oit, dit*	*odi, oit, dit*
Exclude major cities	yes	yes	yes	yes	yes
Observations	1926	1926	13006	11044	11044
R^2	-	-	0.839	0.896	0.896
First Stage KP-F statistic	1748.984	212.052	-	-	-

Notes: This Table reports the robustness analyses for Table 1. Columns 1 and 2 use the distance measures from the hypothetical minimum spanning tree as an IV for the distance measures on the actual network. Columns 3 to 5 use data at the HS2 category level.
Standard errors are clustered at city-port level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

smaller than in the benchmark specification. Finally, in Column 5, we estimate the transport costs separately for expressways and regular roads. We find shipping via expressways less costly compared to regular roads.

Additional robustness. In Appendix Table A.1, we show that PPML and IV specifications using sectoral data generate similar results. While not reporting the results here, we perform regressions with the value of shipments as the dependent variable, which is more standard in the international trade literature but not a theory-consistent measure when sectoral heterogeneity in transport costs are allowed. Results here are robust to this alternative.

4 Full Model

We now embed the routing block into a standard spatial equilibrium model, with costly trade and roundabout production (Eaton and Kortum, 2002; Caliendo and Parro, 2015). The main differences between our model and a standard spatial equilibrium model are that we connect trade costs to transport infrastructure through a routing block and that we allow for sector heterogeneity in domestic trade costs. Our exposition will be brief on the standard aspects and highlight the differences of our setting.

4.1 Preliminaries

There are N regions in the model, denoted by o and d, representing Chinese prefecture cities and the rest of the world (RoW). There are S sectors, denoted by i or j. Workers are immobile and consume a basket of sectoral final goods.[22] These final goods are non-tradable and aggregated

[22]We allow workers to be mobile in the sensitivity analysis in the appendix.

from tradable intermediate goods produced by different locations. The markets for labor, final goods, and intermediate goods are competitive.

4.2 Consumers

Consumers in location d choose their bundle of final goods for consumption to maximize their utility, given by the following:

$$U(C_d) = \prod_{j=1}^{S} [C_d^i]^{\alpha_d^i},$$

where C_d^i is the consumption of final goods in sector i, with price denoted by P_d^i. This preference gives an utility of $U_d = \frac{I_d}{P_d}$, where I_d is total income and $P_d = \prod_{i=1}^{S} [P_d^i / \alpha_d^i]^{\alpha_d^i}$ is the price index of the consumption basket.

4.3 Industry Final Good Production

There are representative sectoral final good producers for each industry i, location d. The task of the final good producers is to aggregate the intermediate goods in sector i produced in different locations into the sectoral final good. They have an Armington production technology, given by:

$$Q_d^i = \Big(\sum_o [\tilde{q}_{od}^i]^{\frac{\sigma-1}{\sigma}} \Big)^{\frac{\sigma}{\sigma-1}},$$

in which $\tilde{q}_{od}^i$ is the quantity of sector-i intermediate good from region o and Q_d^i is the quantity of final goods produced.

4.4 Intermediate Good Production and Trade

The representative intermediate good producers in sector i region d convert labor and sectoral final goods from different sectors into the intermediate goods using the following Cobb-Douglas technology:

$$q_d^i = T_d^i [l_d^i]^{\beta_d^i} \prod_{j=1}^{S} [m_d^{ij}]^{\gamma_d^{ij}}.$$

T_d^i is the location-sector specific productivity, which determines the specialization of a region. l_d^i and m_d^{ij} are inputs of labor and final goods from industry j, respectively; β_d^i and γ_d^{ij} are their respective shares: $\beta_d^i + \sum_j \gamma_d^{ij} = 1$. The unit production cost of sector-i intermediate good in region d is:

$$c_d^i = \frac{\kappa_d^i w_d^{\beta_d^i} \prod_{j=1}^{S} [P_d^j]^{\gamma_d^{ij}}}{T_d^i}, \tag{10}$$

where κ_d^i is a constant: $\kappa_d^i = [\beta_d^i]^{-\beta_d^i} \prod_{j=1}^S [\gamma_d^{ij}]^{-\gamma_d^{ij}}$.

The representative intermediate good producers sell their outputs to final good producers at their marginal costs, which include both the production cost and an iceberg trade cost, denoted by $\tilde{\tau}_{od}^i$. The price of the intermediate goods traded from region o to region d is thus $p_{od}^i = [c_o^i \tilde{\tau}_{od}^i]$. The price of final goods in region d sector i is:

$$P_d^i = \Big(\sum_o ([c_o^i \tilde{\tau}_{od}^i])^{1-\sigma} \Big)^{\frac{1}{1-\sigma}}.$$

And the value of trade flows from o to d in sector i is:

$$X_{od}^i = E_d^i \pi_{od}^i = E_d^i \frac{[p_{od}^i]^{1-\sigma}}{[P_d^i]^{1-\sigma}},$$

where E_d^i is the total expenditure on intermediate goods in sector i of region d and π_{od}^i is the trade share (the share of expenditure in sector i of region d spent on intermediate goods from region o).

The set of conditions characterizing the competitive equilibrium of the model are standard and hence delegated to Appendix B.2. Now we discuss how we model and parameterize $\tilde{\tau}_{od}^i$.

4.5 From Road Networks to Trade Costs

We construct the bilateral costs for domestic trade, $\tilde{\tau}_{od}^i$, by extending the routing framework in Section 2 to incorporate transport cost heterogeneity and an alternative (non-road) mode of transportation.

Sector heterogeneity in transport costs. We allow the ad-valorem equivalent trade cost of goods in sector i to depend on the 'heaviness' of the sector, measured by its weight-to-value ratio, h_i. Consider a seller looking to ship value y of sector i goods along road segment $k \to l$. The number of trucks needed for this task depends on the weight of the goods. Assuming each truck can load h_0 tons, the cost of shipment for this batch of goods on $k \to l$ is simply $\frac{yh_i}{h_0} \iota_{kl}^x$, in which ι_{kl}^x, $x \in \{H, L\}$ is defined in Equation (6). h_0 thus serves as a normalization parameter and determines the level of transport costs.

This setting imposes a linear relationship between trade cost and the weight of a shipment. More generally, the relationship needs not be linear. Indeed, using international shipment data on imports into the U.S., Hummels (2007) finds that the elasticity of ad-valorem shipping cost to weight-to-value ratio is around 0.4-0.5 for both sea-borne and air-borne shipments. We thus relax the linear assumption and specify the domestic segment of the ad-valorem trade cost for sector i, along route $k \to l$ as:

$$(\frac{h_i}{h_0})^\mu \iota_{kl}^x, \; x \in \{H, L\}, \tag{11}$$

in which μ determines the extent of sector heterogeneity in transport costs.

Now consider the trade cost from an origin city o to a destination d of goods in sector i, the iceberg trade cost between o and $d \neq o$ is:

$$\tau_{od}^{i} \equiv \lim_{N\to\infty} \tau_{od,N}^{i} = \Gamma(\frac{\theta-1}{\theta})\Big(\sum_{n=1}^{\infty}(\frac{h_i}{h_0})^{-\mu\theta}[\mathbb{A}_{(o,d)}^{n}]\Big)^{-\frac{1}{\theta}} = \Gamma(\frac{\theta-1}{\theta})(\frac{h_i}{h_0})^{\mu}[\mathbb{B}_{(o,d)}]^{-\frac{1}{\theta}}, \quad (12)$$
$$\text{where } \mathbb{B} = (\mathbb{I}-\mathbb{A})^{-1}.$$

That is, the sector heterogeneity in transport costs on any specific road segment, $k \to l$ translates tractably to the effective trade cost between two domestic locations, o and d.[23]

Alternative mode of transportation. While road transport is the dominant form of domestic shipments in China (road transport accounts for 76% of overall domestic shipments; see National Bureau of Statistics, 2010), there are also alternative modes of transportation via air, water, railway, and pipeline. Given the focus of our quantification on the expansion of the expressway network, we capture these alternative modes in a parsimonious way. Formally, we assume that between any origin city o and destination d, in addition to transport via the road network (which incurs an expected transport cost of τ_{od}^{i}), there is also an alternative mode with an expected cost of $\bar{\tau}_{od}^{i}$, given by

$$\bar{\tau}_{od}^{i} = (\frac{h_i}{h_0})^{\mu}\exp(\bar{\kappa}\cdot\overline{\text{dist}}_{od}),\ o \neq d \quad (13)$$

This cost specification differs from that of road transport in two aspects. First, $\bar{\kappa}$ can be different from κ^H and κ^L. Second, as $\bar{\tau}_{od}^{i}$ is meant to capture the average cost across *all* alternative modes including air transport, we specify it to be a function of the great circle distance $\overline{\text{dist}}_{od}$, determined solely by geography, so there is no routing within the alternative mode.[24]

With this additional mode, the full structure of the routing model works as follows. A seller from region o looking to ship a batch of goods to region d first decides whether to ship it via the road network or the alternative mode, taking the *average* iceberg transport costs for the two modes, τ_{od}^{i} and $\bar{\tau}_{od}^{i}$, as given. Each seller draws two i.i.d. multiplicative Fréchet cost shocks, denoted by ν_M, $M \in \{road, alt\}$, one for each mode, and chooses the mode with the lower effective cost: $\min\{\tau_{od}^{i}\nu_{\text{road}}, \bar{\tau}_{od}^{i}\nu_{\text{alt}}\}$.

If the seller chooses the alternative mode, then the good is directly shipped to the destination, incurring a cost of $\bar{\tau}_{od}^{i}\cdot\nu_{\text{alt}}$ for domestic destinations and $\tau_{RoW}^{i}\cdot\nu_{\text{alt}}$ for export.[25] If the seller

[23] It is possible that in addition to the *level* of trade cost, sectors also differ in the distance semi-elasticity, κ^x. We do not find evidence for heterogeneity along this dimension in the data.

[24] In reality, even after conditioning on the great circle distance between two cities, the cost of shipment on alternative modes can still differ according to the accessibility of direct flights, trains, and cargo ships between the origin and destination. Given the data limitation, we do not directly model these alternatives. Our counterfactual experiments should thus be viewed as keeping these alternatives as fixed. A related concern is whether by overlooking these alternatives, the estimate for the road transport cost will be biased. Our specification will control for bilateral fixed effects and identify key parameters from over-time changes. Under the assumption that when restricted to non-major cities, road network expansion is uncorrelated with *changes* in the cost of other means of transportation, the omission of the alternative mode will not bias our estimate. This assumption appears reasonable—while in addition to expressway, the country also invested heavily in the railway system, most of the investment went to the high-speed railway, which was for passengers but not merchandises.

[25] We assume the international segments of export cost are the same across locations and between the two modes.

chooses road transport, then the rest of the routing module plays out as described in Section 2. Specifically, if the final destination is in China, then the seller randomly meets with a truck driver and compensates the driver with the expected trade cost τ^i_{od}. If instead the final destination is the RoW, the seller first chooses a potential port to minimize $\tau^i_{od}\tau^i_{RoW} \cdot \nu_{\text{port}}$, in which ν_{port} is an idiosyncratic draw of the match quality between the seller and the port. Once the port choice is made, the seller then again meets randomly with the trucker driver, who decides the route from o to d.

Combining all these decisions, the expected trade cost between a domestic origin o and destination d for $o \neq d$ is the following:

$$\tilde{\tau}^i_{od} = \begin{cases} \Gamma(\dfrac{\theta_M - 1}{\theta_M})[(\bar{\tau}^i_{od})^{-\theta_M} + (\tau^i_{od})^{-\theta_M}]^{-1/\theta_M}, & \text{if } d \neq RoW \\ \Gamma(\dfrac{\theta_M - 1}{\theta_M}) \cdot \Big\{(\tau^i_{RoW})^{-\theta_M} + \Big(\Gamma(\dfrac{\theta_F - 1}{\theta_F})[\sum\limits_{\text{ports } k} (\tau^i_{ok}\tau^i_{RoW})^{-\theta_F}]^{-\frac{1}{\theta_F}}\Big)^{-\theta_M}\Big\}^{-1/\theta_M}, & \text{if } d = RoW \end{cases} \tag{14}$$

in which τ^i_{od} is given by Equation (12) and $\bar{\tau}^i_{od}$ given by Equation (13). We define the import cost when o is the RoW and d is a destination in China symmetrically.

4.6 Properties of the Routing Block

We describe some properties of the routing block for subsequent use. It is convenient to define the (k, l) elements of $\mathbb{A}$ and $\mathbb{B}$ as:

$$a_{kl} \equiv [\mathbb{A}_{(k,l)}] \text{ and } b_{kl} \equiv [\mathbb{B}_{(k,l)}].$$

Also define $\iota_{kl} \equiv a_{kl}^{-\frac{1}{\theta}}$. For two adjacent cities connected by both regular roads and expressways, $\iota_{kl} = [(\iota^H_{kl})^{-\theta} + (\iota^L_{kl})^{-\theta}]^{-\frac{1}{\theta}}$, which is the effective cost between k and l accounting for the coexistence of both regular roads and expressways as options. The percentage decrease in ι_{kl} from adding an expressway segment to two cities previously connected by a regular road is:

$$\begin{aligned} \Delta \log(\iota_{kl}) &= -\frac{1}{\theta}\Big(\log[\exp(-\theta\kappa^H dist_{kl}) + \exp(-\theta\kappa^L dist_{kl})] - \log[\exp(-\theta\kappa^L dist_{kl})]\Big) \\ &\approx (\kappa^H - \kappa^L) \cdot dist_{kl}. \end{aligned} \tag{15}$$

We can view a large project as a *collection* of expressway segments. Denote a large project by set $\mathbb{C}$. We write the second order approximation for the change in trade cost between two domestic

The structural estimation accommodates differences in international trade costs across ports through port-time fixed effects.

locations o and d due to expressway project $\mathbb{C}$ as:[26]

$$\Delta \log \tilde{\tau}_{od}^{i} \approx \sum_{kl \in \mathbb{C}} \frac{\partial \log \tilde{\tau}_{od}^{i}}{\partial \log \iota_{kl}} \cdot \Delta \log(\iota_{kl}) + \frac{1}{2} \sum_{kl \in \mathbb{C}} \sum_{k'l' \in \mathbb{C}} \frac{\partial^2 \log \tilde{\tau}_{od}^{i}}{\partial \log \iota_{kl}\, \partial \log \iota_{k'l'}} \Delta \log(\iota_{kl}) \Delta \log(\iota_{k'l'}). \tag{16}$$

The first term captures the direct effect of the project, which sums across the marginal effects of all individual segments $kl \in \mathbb{C}$. The elasticity of trade cost to ι_{kl} in the summand is:

$$\frac{\partial \log \tilde{\tau}_{od}^{i}}{\partial \log \iota_{kl}} = \frac{\partial \log \tilde{\tau}_{od}^{i}}{\partial \log \tau_{od}^{i}} \cdot \frac{\partial \log \tau_{od}^{i}}{\partial \log \iota_{od}} = \underbrace{\frac{(\tau_{od}^{i})^{-\theta_M}}{(\bar{\tau}_{od}^{i})^{-\theta_M} + (\tau_{od}^{i})^{-\theta_M}}}_{\equiv \pi_{od}^{road}} \cdot \underbrace{\frac{b_{ok} \cdot a_{kl} \cdot b_{ld}}{b_{od}}}_{\equiv \pi_{od}^{kl}}. \tag{17}$$

Equation (17) decomposes the partial derivative into the product of two elasticities: the first elasticity, π_{od}^{road}, equals the probability that a shipment from o to d is transported via the road network;[27] the second elasticity π_{od}^{kl} does not have a probability interpretation in general, but as $\theta \to \infty$, it converges to the probability that a shipment passes the edge $k \to l$ conditional on being transported via the road network (Allen and Arkolakis, 2019). Because our structural estimation will find θ to be fairly large (around 80), we adopt this interpretation in the rest of this paper, under which the product of the two terms is simply the unconditional probability of a shipment between o and d passing the edge $k \to l$.

Building on this intuition, the cross derivative term in Equation (16) can be written as $\frac{\partial \pi_{od}^{road} \pi_{od}^{kl}}{\partial \log \iota_{k'l'}}$, and it is the marginal effect of a change in $\iota_{k'l'}$ on the probability that the shipment between o and d uses the edge $k \to l$. The second term in Equation (16) thus captures the interaction between segments through re-optimization of the traffic. We characterize the second order derivative in the summand as

$$\frac{\partial^2 \log \tilde{\tau}_{od}^{i}}{\partial \log \iota_{kl} \partial \log \iota_{k'l'}} = \frac{\partial \pi_{od}^{road} \pi_{od}^{kl}}{\partial \log \iota_{k'l'}} = \pi_{od}^{road} \pi_{od}^{kl} \Big\{ -\theta [\mathbb{1}(kl = k'l') + \pi_{ok}^{k'l'} + \pi_{ld}^{k'l'} - \pi_{od}^{k'l'}] - \theta_M (1 - \pi_{od}^{road}) \pi_{od}^{k'l'} \Big\}, \tag{18}$$

where $\mathbb{1}(kl = k'l')$ is the indicator function that takes one if $kl = k'l'$ and zero otherwise.[28]

Consider first the own effect (when $kl = k'l'$). The first term in the curly bracket captures the impact on shipment over $k \to l$ through re-routing among the traffic within the road network. Because $1 + \pi_{ok}^{kl} + \pi_{ld}^{kl} - \pi_{od}^{kl}$ is always positive, this force contributes negatively: a *decrease* in the cost on edge $k \to l$ *increases* the share of shipment taking this edge. The second term in the bracket, $-\theta_M (1 - \pi_{od}^{road}) \pi_{od}^{k'l'}$, captures the response in the mode choice—more shipment will be made via road in response to a decrease in the edge cost. Both forces work in the same direction and imply

[26] When one of the two locations o or d is the RoW, analogous expressions can be derived.

[27] Notice that since the level of shipment cost, which is determined by the sectoral weight-to-value ratio, shifts τ_{od}^{i} and $\bar{\tau}_{od}^{i}$ proportionally, π_{od}^{road} does not vary across sectors.

[28] In the proof of Equation (18) in Appendix B.3, we show that this expression is symmetric in kl and $k'l'$, so the order of taking this cross derivative does not matter.

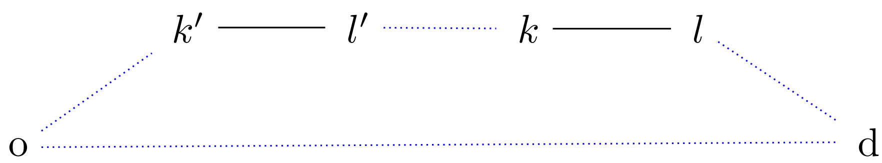


Figure 5: Interactions Between Segments

Note: The diagram illustrates a case in which expressway in $k' \rightarrow l'$ and $k \rightarrow l$ complement each other.

that as an expressway is added to $k \rightarrow l$, more trade flows will go through this edge.

Now consider the cross-derivative (when $kl \neq k'l'$). The response in the choice of mode has the same sign as before, but the first term in the bracket capturing the re-optimization of the ground traffic could be positive or negative, depending on the positions of $k' \rightarrow l'$ and $k \rightarrow l$ in the network. When $k' \rightarrow l'$ and $k \rightarrow l$ are on two different routes between o and d, shipments between o and k and between l and d are unlikely to pass through $k' \rightarrow l'$, so $\pi_{ok}^{k'l'}$ and $\pi_{ld}^{k'l'}$ are both small and $-\theta(\pi_{ok}^{k'l'} + \pi_{ld}^{k'l'} - \pi_{od}^{k'l'})$ is more likely to be positive. This corresponds to a situation in which a reduction in $\iota_{k'l'}$ draws ground traffic *away* from $k \rightarrow l$. On the other hand, if $k' \rightarrow l'$ is en route of $o \rightarrow k \rightarrow l \rightarrow d$, as in the example given in Figure 5, then the opposite can happen—reducing $\iota_{k'l'}$ increases the traffic passing through $k \rightarrow l$.[29] We summarize these results in Proposition 1.

Proposition 1. *Equations (15), (16), (17) and (18) characterize the effect of expressway project $\mathbb{C}$ on domestic trade costs up to the second order.*

4.7 Welfare Criteria

The competitive equilibrium of the model is Pareto efficient. We define the aggregate welfare to be the average log real income of regions, weighted by their value added:

$$W = \sum_d \frac{w_d L_d}{Y} \log(\frac{w_d}{P_d}), \tag{19}$$

where $Y = \sum_d w_d L_d$ is the total value added. In the absence of international trade, the desired allocations of a social planner with these weights replicate exactly those of the competitive equilibrium. We can therefore evoke an argument based on the social planner's optimization problem

[29] In the case illustrated, $\pi_{ld}^{k'l'}$ is close to zero and $\pi_{ok}^{k'l'}$ is close to one. The sum of the three terms is thus strictly positive as long as not all shipments between o and d go through the upper branch (i.e., $\pi_{od}^{k'l'} < 1$).

to express the welfare gains from an expressway project $\mathbb{C}$ as

$$\begin{aligned}
\Delta W &= \sum_i \sum_{od} \frac{dW}{d\log \tau^i_{od}} \Delta \log \tau^i_{od} + SO_T \qquad (20)\\
&= -\sum_i \sum_{o\neq d} \frac{X^i_{od}}{Y} \Delta \log \tau^i_{od} + SO_T\\
&= \underbrace{-\sum_i \sum_{o\neq d} \frac{X^i_{od}}{Y} \sum_{kl\in\mathbb{C}} \pi^{road}_{od} \pi^{kl}_{od} \Delta \log(\iota_{kl})}_{\text{FO effect}} + SO_R + HO_R + SO_T,
\end{aligned}$$

$$\text{where } SO_R = -\sum_i \sum_{o\neq d} \frac{X^i_{od}}{Y} \frac{1}{2} \sum_{kl\in\mathbb{C}} \sum_{k'l'\in\mathbb{C}} \frac{\partial^2 \log \tilde{\tau}^i_{od}}{\partial \log \iota_{kl}\, \partial \log \iota_{k'l'}} \Delta \log(\iota_{kl}) \Delta \log(\iota_{k'l'}),$$

and $\Delta \log \iota_{kl}$ and $\dfrac{\partial^2 \log \tilde{\tau}^i_{od}}{\partial \log \iota_{kl}\, \partial \log \iota_{k'l'}}$ are given by Equations (15) and (18).

The first line of the equation assumes that the resulting trade cost changes are known and expresses the welfare change as the sum of the first order effects of the changes in trade cost and a higher order residual (labeled SO_T). The second line applies the Envelope Theorem to the social planner's problem; the third line writes the change in trade cost, $\Delta \log \tau^i_{od}$, as the sum of three terms: the first order effect of project $\mathbb{C}$ on trade costs, the second order effect through changes in routing (SO_R), and a higher order residual in the routing block (HO_R).

The aggregate welfare gains from project $\mathbb{C}$ is therefore decomposed into four components. The first order component, labeled 'FO effect', is calculated under the assumption that in response to the new project, 1) neither the mode choice nor the routing patterns change; and 2) the trade patterns do not change. As discussed, for large θ, $\pi^{road}_{od} \cdot \pi^{kl}_{od}$ is the probability that trade between o and d is shipped via the edge $k \to l$. The 'FO effect' is thus the sum of cost-savings from trade flows on all the segments in $\mathbb{C}$. This is an insight dating back to Hulten (1978) and has been used in evaluating the global gains from trade (see e.g., Burstein and Cravino, 2015). The broad approach has also been embodied in the evaluations of transport projects.[30]

Looking beyond the first order effect, the SO_R term stands for the second order effect from the routing block and can be evaluated using Equation (18). It captures the substitution between modes and routes as a result of $\mathbb{C}$, assuming trade flows do not respond. Finally, HO_R captures the remaining approximation errors of $\Delta \log \tau^i_{od}$; and SO_T stands for the second and higher-order effect through changes in trade flows.[31] We summarize these results in Proposition 2:

[30] A widely used approach in transportation research in evaluating transportation programs is to focus on the value of travel time savings, which is the product of the time saved through the new transportation infrastructure and the value of time (see Small, 2012 for a recent survey). In our context, in which the competitive equilibrium is efficient, this method corresponds to the first order welfare gains. See Allen and Arkolakis (2019) for a method for models with inefficient competitive equilibria.

[31] We do not aim to characterize HO_R and SO_T—the first is of only the third order; the second, although potentially significant, tends to be small in our setting with Cobb-Douglas production functions for intermediate goods, which is one of the benchmark settings in most quantitative trade studies.

Proposition 2. *In a restricted model without international trade, the competitive equilibrium of the model coincides with the allocation of a social planner whose objective function is described by Equation (19). In this setting, the welfare gains from expressway project $\mathbb{C}$ is given by Equation (20).*

From the proposition, all we need for the FO effect is the structural elasticities (κ^H, κ^L) and the value of shipments passing through each edge $kl \in \mathbb{C}$. The latter is in principle measurable, so once the structural parameters are estimated in the next section, we can evaluate the first order welfare gains from the expressway expansion in China using Equation (20). Instead, we choose a different path, which is to parameterize the full model and conduct simulation exercises, because in this setting the first order approach faces two obstacles. The first is in measurement. Collecting the data on the value of shipments passing each of the hundreds of the segments involved in the expressway construction is conceptually straightforward but impractical.[32] We rely on the model, parameterized to match available statistics, to infer the value of shipments. Second, because the massive investment in expressways during the decade constitutes a large shock to both local linkages $(\Delta \log \iota_{kl})$ and the overall structure of the network, the second order effects through changes in routing could be important. After describing the main quantitative results, we will return to these two points in Sections 6 and 7.

5 Parameterization

In parameterizing the model, we adopt a two-step indirect inference approach. In the first step, without imposing the equilibrium conditions, we estimate two set of regressions to recover coefficients informative about the routing parameters (κ^H, κ^L, θ, θ_F). In the second step, we calibrate the full model to pin down their values and the remaining parameters in the equilibrium.

5.1 Export Routing Regression

The first specification is on the response in the route choice of exporters to the expressway network expansion. It is the structural version of the regressions in Section 2.

Consider an exporter from inland city o selling to customers in the RoW. From the routing block, conditioning on the goods being exported through a seaport, the probability that it goes through a particular port d, among all other ports, is given by Equation (5).[33] Taking log on both

[32] Because of the granularity and complex interactions among sectors and regions, predicting the traffic on an expressway segment ex-ante is difficult. After an expressway segment is completed, while it is straightforward to count the vehicles passing by, it is the *value* of the goods on these vehicles—which are much more difficult to estimate—not their weight, that matters for the aggregate welfare.

[33] Equation (5) still holds in the full model with the alternative mode because of the sample restriction—we focus on shipments from *interior* cities to the RoW through seaports. In our model, because goods exported via the alternative mode will be shipped directly to the RoW, by construction they are not in this sample. It is possible in the data, some goods are first shipped via the alternative mode (most likely the railways) to a seaport and then sent to the RoW. This is more likely for heavier and bulkier industries which are more dependent on railway for transportation, such as coal and wood. We perform a robustness in which we exclude these sectors from the regressions and find the results are similar.

sides of the equation and applying Equation (12) to obtain:

$$\log(\pi^i_{(o,RoW),d}) = \frac{\theta_F}{\theta}\log\left([\tilde{\mathbb{B}}(\kappa^H\theta,\kappa^L\theta)_{(o,d)}]\right) \underbrace{+\mu\log(\frac{h_i}{h_0}) - \theta_F\log(\tau^i_{d,RoW}) - \log(\sum_{\text{All ports k}}\tau_{ok}^{-\theta_F}\cdot\tau_{k,RoW}^{-\theta_F})}_{\text{fixed effects}} \tag{21}$$

On the right hand side of Equation (21), the overall international export costs $\tau^i_{d,RoW}$, the city-specific access to the international market, $(\sum_{\text{All ports k}}\tau_{ok}^{-\theta_F}\cdot\tau_{k,RoW}^{-\theta_F})$, and the 'heaviness' of sector i will be absorbed by fixed effects. The variations we exploit are changes in the (o,d) element of matrix $\mathbb{B}$ over time. Recall that $\mathbb{B}=(\mathbb{I}-\mathbb{A})^{-1}$ is entirely determined by $\mathbb{A}$. Because κ^H and κ^L enter $\mathbb{A}$ only multiplicatively with θ, we write$[\mathbb{B}_{(o,d)}]$ as $[\tilde{\mathbb{B}}(\kappa^H\theta,\kappa^L\theta)_{(o,d)}]$ to highlight its dependence on $\kappa^H\theta$, $\kappa^L\theta$.

We estimate Equation (21) for $\frac{\theta_F}{\theta}$, $\kappa^H\theta$, and $\kappa^L\theta$ by nonlinear least square without solving the equilibrium model. Informed by the reduced-form evidence, we focus on over-time variations and control for city-time, port-time, and city-port fixed effects. Formally, with the observed export route choices $\hat{\pi}^s_{(o,RoW),d,t}$ for time $t\in\{beginning, end\}$ in the data, we choose the structural parameters to minimize the following expression

$$\max_{\frac{\theta_F}{\theta},\kappa^H\theta,\kappa^L\theta,\boldsymbol{f}}\left[\frac{\theta_F}{\theta}\log\left([\tilde{\mathbb{B}}_t(\kappa^H\theta,\kappa^L\theta)_{(o,d)}]\right)+\boldsymbol{f}-\log(\hat{\pi}_{(o,RoW),d,t})\right]^2, \tag{22}$$

in which $\boldsymbol{f}$ is the full set of fixed effects. Given the large number of fixed effect included, this is a high-dimensional optimization problem which conventional optimization routines cannot handle. Note, however, that only $\kappa^H\theta$ and $\kappa^L\theta$ enter the objective function non-linearly, we can thus recast the original problem into a nested one as below:

$$\max_{\kappa^H\theta,\kappa^L\theta}\{\max_{\frac{\theta_F}{\theta},\boldsymbol{f}}\left[\frac{\theta_F}{\theta}\log\left([\tilde{\mathbb{B}}_t(\kappa^H\theta,\kappa^L\theta)_{(o,d)}]\right)+\boldsymbol{f}-\log(\hat{\pi}_{(o,RoW),d,t})\right]^2\}.$$

In the inner loop, given values for $\kappa^H\theta$ and $\kappa^L\theta$, we estimate $\boldsymbol{f}$ and $\frac{\theta_F}{\theta}$ using standard linear regressions. In the outer loop we search over the space of $\kappa^H\theta$ and $\kappa^L\theta$ to minimize the residual mean square error from the inner loop.

Table 3 reports the output of this exercise. While with routing information alone we cannot identify individual parameters yet, the estimates reveal their relative magnitudes. First, expressways are about a quarter less costly compared to regular road ($\frac{\kappa^H}{\kappa^L}=74\%$). This difference is broadly in accord with measurements based on actual speeds (see Section 2). This finding is reassuring, especially because here we do not impose that $\kappa^L>\kappa^H$. Second, the elasticity of substitution across routes (θ) is much larger than that across ports (θ_F). This appears reasonable, as the former is driven by the preference of drivers among routes, whereas the latter depends on the idiosyncratic shocks of sellers across ports, which could be related to business preferences.

Table 3: Estimates from the Routing Model

Parameter	$\kappa^H\theta$	$\kappa^L\theta$	$\frac{\theta_F}{\theta}$
Estimate	4.44	5.98	0.03

Notes: This table reports the estimates from Equation (22).

Note that the multiplicative nature of $(\frac{h_i}{h_0})^\mu$ implies that with the fixed effects we control for, μ is not identified from the domestic routing patterns alone. We next turn to the price data and discuss how it helps us identify μ and other structural parameters of the model.

5.2 Price Regressions

Consider a firm in sector i from an interior city o exporting to the RoW via a seaport d. Let the factory-gate price of the good be p^i_o. Under the assumption of complete pass-through, the average (across all route-specific draws) free-on-board price at port d is given by:

$$\begin{aligned} p^i_{(o,RoW),d} &= p^i_o \cdot \tau^i_{od} \\ &= p^i_o \cdot (\frac{h_i}{h_0})^\mu \cdot [\tilde{\mathbb{B}}(\kappa^H\theta, \kappa^L\theta)_{(o,d)}]^{-\frac{1}{\theta}},\ o \neq d \\ \implies \log(\frac{p^i_{(o,RoW),d}}{p^i_o}) &= \mu \log(h_i) - \frac{1}{\theta}\log\left([\tilde{\mathbb{B}}(\kappa^H\theta, \kappa^L\theta)_{(o,d)}]\right), \end{aligned} \tag{23}$$

where p^i_o is the producer price of a product. Equation (23) shows that variations in price ratios across sectors with different 'weight-to-value' ratios identify μ; assuming $\kappa^H\theta$ and $\kappa^L\theta$ are known, variations across city pairs with different distances identify θ.[34] With this intuition, we estimate the elasticity of price ratio with respect to h_i, and the semi-elasticity of price ratio with respect to the road distance between o and d. We then target these two estimates along with other empirical moments in the full calibration to pin down all routing parameters.[35] Because the two moments are estimated off different variations, we estimate them separately so more controls can be included.

Data. We measure prices as the unit values of exported goods from the customs data. Without the factory-gate price of each transaction, we construct the price ratio as follows. We restrict the sample to transactions with the origin city o being a seaport itself. For the goods produced in such city o, the average export price for when exporting directly from o, i.e., $p^i_{(o,RoW),o}$, is then a theory-consistent measure of the factory-gate price.

[34]We assume that the international trade cost, τ^i_{RoW}, is not included in the measured unit price. To the extend that it is included, our empirical specification will control for it.

[35]It is possible that the variations in price ratios might be driven by other reasons, such as quality differences, which could be attributed to trade costs. To avoid this problem, in estimating the two targets of the indirect inference, we will control for a rich set of fixed effects. We will use only the systematic variations of price ratios across ports and sectors—rather than the levels of price ratios—in quantification. The *level*, which is ultimately governed by h_0 will be pined down to match the average shipment distance in China. Alternatively, we can also estimate Equation (23) using non-linear least squares for the structural parameters directly as in 5.1. That approach, however, would require us to take the level of the price ratio more seriously.

The validity of this approach rests on the assumption that goods shipped directly from o to the RoW and goods shipped indirectly through a different city d are the same. To make this assumption reasonable, we take advantage of the details in the customs data and define each 'product' to be a combination of city, HS8 category, and destination country. For each such product, we calculate the average price of direct export transactions from origin city o to obtain $p^i_{(o,RoW),o}$. The log ratio between the price of the same product exported via a different city d and $p^i_{(o,RoW),o}$ is then the ad-valorem trade cost between o and d.

The narrow definition of a product addresses a few concerns in interpreting price ratios as trade costs. First, firms both export higher-quality goods and charge higher markups on these goods for destination countries with a higher income (see, e.g., Fan et al., 2015). Second, cities with a more skilled workforce tend to produce better products (Dingel, 2016). Conditioning on the same destination market and origin city avoids these two sources of biases. To alleviate other concerns, our empirical specifications further absorb remaining systematic variations in either qualities or markups across transactions through fixed effects; we also show that the results are similar if we focus on non-differentiated products, as classified in Rauch (1999), where such concerns are less important.

The drawback of using narrowly defined products is that there were not enough exports at the initial period to estimate the distance effect from over time variations, so we focus on cross-sectional regressions using the end-of-period data only.

Price-heaviness elasticity. Table 4 reports our estimate on the price-heaviness elasticity, with progressively more demanding fixed effects. The first four columns focus on the comparison of the log price differences across HS2 categories. The first and second columns control for city, port, and destination country fixed effects and city-port-country fixed effects, respectively. Even within a city-port-country cell, some firms might systematically set prices differently. To account for this possibility, Column 3 control for firm-port-country fixed effects. The point estimate increases from 0.16 to 0.27.

To the extent that the price ratio might still capture variations in qualities and markups despite our narrow definition of products and the set of fixed effects included, as long as they are not systematically correlated with the weight-to-value ratio, they will not affect our estimates. Nevertheless, Column 4 focuses only on the HS2 categories that are classified as non-differentiated goods (Rauch, 1999), which likely have a smaller scope for either quality differentiation or price discrimination. Reassuringly, although that the sample is only a tenth of the baseline sample, the point estimate remains broadly in line.

One further concern is that our measure of 'heaviness', the weight-to-value ratio, might capture other characteristics of a sector that correlates systematically with prices. In Columns (5) through (7), we estimate the specification using the weight-to-value ratio at the HS4 category level. This allows us to control for the HS2 fixed effects. The last column of Table 4 is our preferred specification, which is identified from within a city-HS2-port-country cell, whether heavier goods are relatively more expensive when exported through a different seaport than own city. The point

Table 4: Transport cost and weight-to-value ratio

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable	log price ratio				log price ratio		
Heaviness- HS2 Category	0.163***	0.161***	0.278***	0.199**			
	(0.056)	(0.056)	(0.086)	(0.089)			
Heaviness- HS4 Category					0.303***	0.362***	0.253***
					(0.044)	(0.050)	(0.043)
Fixed Effects	o, d, c	odc	fdc	fdc	fdc, i	$fdci$	$fdci$
Exclude major cities	yes	yes	yes	yes	yes	yes	yes
Exclude differentiated goods				yes			yes
Observations	1987140	1985946	1805563	190836	1805563	1126941	119077
R^2	0.063	0.074	0.375	0.481	0.417	0.596	0.639

Notes: This table reports the regressions of log price ratio on sector weight-to-value ratio, using data from 2010-2011. The dependent variable is the log of price ratio and is always computed by city-destination country-HS8 category; the independent variable is the log of the weight-to-value ratio at HS2 category level (Columns 1-4) and HS4 category level (Columns 5-7). Letters o, d, c, f, i stand for origin city, port, destination country, firm, and HS2 category fixed effects, respectively.

Standard errors are clustered at HS2 category level (Columns 1-4) or HS4 category level (Columns 5-7). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

estimate suggests that a one-percent increase in the weight-to-value ratio of a good increases the ad-valorem cost by around 0.3%.

Our estimate is at the lower range of Hummels (2007), which finds an elasticity of 0.4-0.5 for international shipment fees. The literature does not offer much guidance on this elasticity for domestic shipments. But the freight costs for domestic shipments documented in the literature is usually denoted linearly in weight (see Redding and Turner, 2015), which translates into an elasticity of one. To be conservative on the role of sector heterogeneity, we use 0.3 as the target in the calibration of the full model; we use an elasticity of one for sensitivity analyses.

Price-distance semi-elasticity. Table 5 reports the second set of price regressions focusing on the distance semi-elasticity. The independent variable is the effective distance along the shortest route from city o to port d, as defined in Section 2. Since we do not aim to identify μ in this regression, we can absorb the category characteristics in fixed effects. The first two columns use OLS and control for port-HS8-destination country and city-HS8-destination country fixed effects, respectively. The former set captures, within a HS8 category, the overall tendency of some ports or destination countries to be involved in the export of more pricey goods; the latter controls for the overall tendency of a city in producing pricey good for exporting to specific countries. The point estimate of the first column, which uses all categories, suggests that the price ratio increases by around 5.5% for an additional hundred kilometer equivalent regular road distance. The second column restricts to non-differentiated varieties for robustness and finds a similar coefficient.

To alleviate the concern about the endogeneity of the road network, Columns 3 and 4 replicate Columns 1 and 2 using the cross-sectional IV from the minimums-spanning tree. The point estimates are in the range of 0.05 to 0.06, statistically indistinguishable from the OLS estimates.

Table 5: Price Distance Regression

	(1)	(2)	(3)	(4)
	OLS		IV	
$dist_{od}$	0.055***	0.061***	0.053***	0.058***
	(0.013)	(0.022)	(0.012)	(0.021)
Fixed Effects	dci, oci	dci, oci	dci, oci	dci, oci
Exclude major cities	yes	yes	yes	yes
Exclude differentiated goods		yes		yes
Observations	1829372	232609	1829372	232609
R^2	0.323	0.340	-	-
First Stage KP-F statistic			1515.787	1156.297

Notes: This table reports the regressions of log price ratio on the distance between the origin city and the port. The dependent variable is the log of price ratio; the independent variable is the distance along the shortest path between city o and port d. Letters o, d, c, i stand for origin city, port, destination country, and HS-8 product fixed effects, respectively. Standard errors are clustered at city-port level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We use an estimate of 6% as the target for price-distance semi-elasticity in the calibration. While this estimate has no direct structural interpretation in our framework, under the assumption that $\theta \to \infty$, it means each additional 100 km increases the ad-valorem trade cost by 6%.

5.3 Model Parameterization

We parameterize the model to match the estimated moments described previously and additional features of the Chinese economy around 2010. Our calibration is at the prefecture-city level, with 323 prefectures and 25 sectors.[36]

Parameters calibrated externally. The upper panel of Table 6 describes the parameters and fundamentals of the regional economy calibrated externally. We assign the number of workers in each city based on the 2010 population census; we extract the employment in the RoW from the Penn World Table. We assume the sector shares in final consumption and intermediate production, $\{\alpha^i\}$ and $\{\gamma^{ij}\}$, and the labor share in production, $\{\beta^i\}$, are the same across regions and determine them based on the input-output table of China in 2007. We assign a value of 6 to the elasticity of substitution across goods from different regions, σ. θ_M governs the elasticity of substitution between different modes of transport. The existing estimates for this parameter range from $1 \sim 3$ in the earlier transportation literature (Abdelwahab, 1998) to 14 in the more recent Allen and Arkolakis (2019). We assign a value of 2.5 to θ_M for benchmark analyses and will conduct robustness checks with alternative values in the appendix.

Parameters determined in equilibrium. The remaining parameters, reported in the lower panel of Table 6, are determined jointly in equilibrium. The transport cost parameters along regular roads and expressways, κ^H and κ^L are pinned down together with the dispersion parameter for routing preference θ. In Section 5.1, we estimate Equation (21) and find that $\kappa^H \theta = 4.44$,

[36] The 25 sectors include one agriculture, four energy/mining, sixteen manufacturing, and four non-tradable sectors including utilities, construction, transportation, and others.

Table 6: Parameter Values

Parameters	Descriptions	Value	Targets/Source
Parameters calibrated externally			
$\beta^i, \gamma^{ij}, \alpha^j$	IO structure and consumption share	-	2007 IO table for China
L_d	Total employment	-	2010 Population Census
σ	Trade elasticity	6	
θ_M	Elasticity of substitution across modes	2.5	
Parameters calibrated in equilibrium			
θ	Routing elasticity	81.21	Estimates of Equations (21) and (23)
θ_F	Port choice elasticity	2.45	Estimates of Equations (21) and (23)
κ_H	Expressway route cost	0.055	Estimates of Equations (21) and (23)
κ_L	Regular route cost	0.074	Estimates of Equations (21) and (23)
h_0	Trade cost level	1.295	Average ground shipment distance: 177 km
$\bar{\kappa}$	Alternative mode cost	0.210	Share of non-road shipment: 0.24
μ	Cost-weight to value elasticity	0.3	Equation (23)
$\tau^i_{RoW}, \tau^{i'}_{RoW}$	Export and import costs	-	Sectoral export and import
T^i_d	Region-sector productivity	-	City-sector sales in 2008 Economic Census

$\kappa^L\theta = 5.98$, $\frac{\theta_F}{\theta} = 0.03$; we also show in Table 5 that empirically, each additional 100 km in distance leads to a 6% increases in log price ratio. We choose θ so that the price-distance semi-elasticity estimated using the simulated data from the equilibrium of the model is also 6%. This procedure determines $\theta = 81.21$, $\theta_F = 2.45$, $\kappa^H = 0.055$, $\kappa^L = 0.074$.

Parameter μ determines the variations of transport costs across sectors with different heaviness. We set μ to 0.3 based on Table 4. We determine the overall level of domestic trade cost h_0 by targeting the average shipment distance in China (National Bureau of Statistics, 2010), which is 177 kilometer. Under this target, the equilibrium inter-city trade as a fraction of total output is 44%. With all else determined, the remaining parameter of the routing model, the distance semi-elasticity for alternative modes (such as air transport), $\bar{\kappa}$, pins down the equilibrium share of shipment using roads versus other modes. About 76% of domestic shipment (by weight) is conducted by road transportation (National Bureau of Statistics, 2010). We choose $\bar{\kappa}$ so that the model generates the same ratio.

We use the sectoral trade costs between the port city and the RoW to target the sectoral import and export as shares of domestic GDP.[37] Finally, we use the region-sector productivity parameters, $\{T^i_d\}_{d \neq RoW}$, to match the sectoral output shares of each prefecture city, constructed from the 2008 China Economic Census.[38] We calibrate $\{T^i_{RoW}\}$ such that the model implied ratios between the sectoral output of China and that of the RoW match the data.

Figure 6 plots values of shipment flows between pairs of adjacent cities. Darker colors indicate

[37] To match the import and export shares, our calibration takes into account *exogenous* international trade surpluses of China. After the calibration, we solve for a baseline equilibrium without trade imbalances. All the counterfactual experiments will then be compared against this baseline equilibrium. Throughout the rest of the paper we also refer to this as the calibrated equilibrium.

[38] We use the economic census to construct the production shares within the manufacturing and service sectors. We use the industry employment shares from the 2010 population census for agriculture and mining sectors that the economic census does not cover.

Figure 6: Model Predicted Shipment Flows

Note: This figure plots the value of road shipments between cities, normalized by the GDP of China. When two adjacent cities are connected by both regular roads and expressways, the value plotted is the sum across the two.

higher intensities. With large heterogeneity in importance across segments, standing out from the map are a few corridors that connect the most important economic centers of China. The northeastern corridor surrounding the Bohai Bay, linking Beijing and Tianjin to centers of heavy industrial sectors such as Dalian, Shenyang, and Changchun; the corridor between Beijing and the Yangtze River Delta in the southeast, the most economically prosperous area of China; finally, the corridor connecting the northwest to the center of China.

Zooming into local areas, the three busiest segments on the entire map are between Wuxi and Changzhou, between Suzhou and Nanjing, and between Taizhou and Suzhou, all of which are in the Yangtze River Delta. This is in accord with the popular press coverage that frequently dubs the expressway between Nanjing and Shanghai, which all the three segments belong to, as the busiest expressway in China.

5.4 Model Validation

We validate the model by comparing some of its 'out-of-sample' predictions to the data.

Transport hubs. Because of their central locations in the transport network, some cities become 'hubs' that shipments to other places go through. To validate the model, we can compare the model-inferred shipments passing a city to its empirical counterpart, sourced from the 2010 yearbook for transportation.[39] Table 7 reports the regression of the log shipment in the data on

[39]The data is aggregated by city and produced by the National Bureau of Statistics from surveys of firms in the logistics industry. The data series appear inconsistently defined over time, with frequent abrupt changes from one year

Table 7: Predicting City Shipment

	(1)	(2)	(3)
Log(shipment), model	0.365***	0.208***	0.185***
	(0.041)	(0.038)	(0.046)
Log(employment)		0.594***	0.584***
		(0.059)	(0.064)
Observations	239	239	233
Fixed Effects	no	no	prov
R^2	0.236	0.490	0.633

Notes: The dependent variable is the log city shipment in the data (2010); the independent variable is the log city shipment in the model. Robust standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

the model prediction. The first column shows the raw correlation. The second column controls for city employment. The coefficient is still significant and meaningful. This suggests that the model prediction correlates with the data not only because of the usual gravity force that predicts more trade for bigger cities, but also because it captures the traffic passing by. The third column further shows that including the provincial fixed effect does not change the estimate. This implies that the prediction power comes from the network connections of a city shaped by the routing model, rather than the broad location of the city.

Expressway and Export Growth. In the second validation test, we compare the model-predicted export growth led by the expressway network expansion to the actual export growth in the data. This is a joint test of two hypotheses: 1) whether the expressway expansion as large as the one seen in China over the decade led to differential growth of exports across cities; 2) when fed into the expressway expansion, whether the model can generate the changes in trade patterns in the data. This comparison is out-of-sample, because in the calibration we absorb the level of export through city-time fixed effects and use only the information from the patterns of routing.

In implementing this exercise, we feed in the 1999 expressway network to the model and solve a counterfactual equilibrium holding all other parameters at the calibrated values. We treat the export generated from this counterfactual equilibrium as the model export in 1999. We then compare the export at the city-sector level between the model and the data for 1999 and 2010. Table 8 reports the results. The dependent variable is the log export in the data and the independent variable is its model counterpart. The first column controls for the time fixed effects to look at cross-sectional predictions. The second column controls for sector-time and city-sector fixed effects, so the comparison is on export growth within a city-sector cell. The point estimate is around 0.7 and highly statistically significant. The third column excludes major cities from the sample and the point estimate remains similar.

Importantly, all these regression models have a F statistic above the rule-of-thumb for bounding biases in IV estimates. Under the assumption that the road networks affect city export only

to another, so we do not use the time dimension of the data.

Table 8: Predicting Export Growth

	(1)	(2)	(3)
Log(export), model	0.362***	0.685***	0.622***
	(0.045)	(0.159)	(0.165)
Fixed Effects	t	oi, it	oi, it
Exclude major cities	no	no	yes
Observations	8544	8544	6648
R^2	0.311	0.876	0.858
F-statistic	63.831	18.651	14.269

Notes: The dependent variable is the log city-sector export in the data; the independent variable is the log city-sector export in the model. Letters t, o, i, in the 'Fixed Effects' stand for time, city, and sector (two-digit) fixed effects, respectively. Standard errors (clustered by city) in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

through improving the access of a city to ports, the model predictions can serve as an IV for export at the city-industry level. A growing literature has examined the impacts of Chinese export on its domestic economy). One IV commonly used in this literature is the tariff cuts after the WTO accession, which is valid assuming the pre-WTO tariffs are exogenous (Tian, 2019). The IV based on our model predictions vary across regions and over time, and is valid under a different set of assumptions from existing studies.

6 The Impacts of the Expressway Expansion

6.1 Benchmark Results

Armed with the parameterized model, we examine the aggregate impacts of the expressway construction. We calculate the percentage change in a number of statistics as the economy moves from an equilibrium with the 1999 expressway network to the one with 2010 expressway network. Table 9 reports the result. The aggregate welfare, defined by Equation (19), increases by 5.6%. To put this number into perspective, the welfare relevant aggregate TFP of China grew by 36% between 1999 and 2010 (Penn World Table 9.0, see Feenstra et al., 2015). Through the lens of our model, the reductions in domestic transport costs brought about by the expressway network expansion account for around 16% of the increase.

The expanding expressway network has a large impact on the patterns of domestic and international trade. The domestic trade over GDP share increases by 11%. Because interior regions ship their goods to the RoW through ports, the expressway expansion also affected international trade. It is tempting to think that lower domestic trade costs will always encourage international trade, but the theoretical prediction is ambiguous.[40] It turns out that in our setting, the reductions in domestic trade cost lead to a 16% increase in international trade. In the data, the export over

[40]On the one hand, interior regions will trade more with the RoW because of the improved access; on the other hand, the coastal regions might be diverted to trade more intensively with the interior, leading to a decline in the aggregate international trade.

Table 9: The Effects of the Expressway Expansion, 1999-2010

Change in	Value
Aggregate welfare	0.056
Log(Domestic trade / GDP)	0.113
Log(Exports / GDP)	0.157
Std Log(real wage) across regions	-0.0288

Note: Changes in model statistics are calculated by comparing the calibrated equilibrium and a counterfactual equilibrium with the 1999 expressway network.

GDP ratio increased by 70%, from 18% in 1999 to 32% in 2008, before it plummeted during the great trade collapse. About a quarter of this 70% increase in export intensity could be explained by the expansion in the domestic expressway network.

By connecting previously remote areas to the network, the expressway generates distributional effects. The real wage inequality across regions, measured by the standard deviation of log real wages, decreases by around 3%. This change, however, represents only a modest decrease (around 5%) from the large income dispersion in the baseline economy.

6.2 The Role of International Trade, Sector Heterogeneity, and Input-output Linkages

Our benchmark model differs from those used in the growing literature quantifying the impacts of transportation infrastructure (see, e.g., Asturias et al., 2018; Fajgelbaum and Schaal, 2019; Allen and Arkolakis, 2019) in three aspects. First, our structural estimation exploits changes in the route choice of exporters resulting from the domestic expressway network expansion, which naturally implies that the expressways have the added benefit of reducing import and export costs for the hinterland; second, with sector level information on production and export prices, we allow for regions to differ in sector specializations and sectors to differ in trade costs; third, we incorporate intermediate inputs.

In this subsection, we argue that because these ingredients allow us to infer the distributions of shipments among different routes and the shipment values more accurately, they are important for the quantitative results. To make this point, we parameterize a series of restricted models and comparing the inferred welfare gains in these models to the baseline results. For transparency, throughout this subsection we recalibrate only the trade cost level parameter, h_0, to match the average domestic shipment distance, and city-sector productivity $\{T_d^i\}$ to match sales by either city or city-sector, depending on the restriction on the model. We keep other structural parameters in the routing problem as in the benchmark.

Domestic transport costs in international trade. The second column of Table 10 is the result from a model without international trade, i.e., with $\tau_{RoW}^i = \infty$, $\forall i$. The inferred gains from expressway construction in this model is about 7% (or 0.4 p.p.) smaller than in the baseline model (reproduced in Column 1).

We can understand the difference by inspecting the first order effect on the aggregate welfare

Table 10: Welfare Gains in Alternative Models, Matching Average Ground Distance

	Baseline	Model (2)	Model (3)	Model (4)	Model (5)
International trade	✓				
Regional specialization	✓	✓			
Trade cost heterogeneity	✓	✓	✓		
Intermediate input	✓	✓	✓	✓	
Welfare gains	5.64%	5.27%	4.54%	3.18%	0.74%

Note: For each alternative model, city-sector productivity $\{T_d^i\}$ and the level of transport cost h_0 are recalibrated to match the same city-sector sales (or city-level sales, depending on whether regional specialization is incorporated in the calibration) and the same average domestic ground shipment distance.

of *one* expressway segment, $\Delta \log \iota_{kl}$. From Equation (20), the equation below holds in Model (2):

$$\Delta W \approx -\sum_i \sum_{o \neq d} \frac{X_{od}^i}{Y} \cdot \pi_{od}^{road} \pi_{od}^{kl} \cdot \Delta \log \iota_{kl},\ o \neq \text{RoW},\ d \neq \text{RoW}. \tag{24}$$

Equation (24) does not hold in the full model because the social planer that replicates the competitive equilibrium would place positive Pareto weights on the RoW. However, if we view trading with the RoW as a reduced-form production function, then the first order domestic welfare gains in the full model are given by an extended version of Equation (24) that allows o or d to be the RoW. By matching the average shipment distance for goods within China, both the full model and the model without international trade generate similar X_{od}^i, $o \neq$ RoW, $d \neq$ RoW, so they predict similar cost savings from *domestic* trade. Through the lens of the full model, however, these are only part of the benefits—the improvements in domestic infrastructure reduce the cost for the importers and exporters from the hinterland. Because part of these additional cost savings will accrue to the Chinese economy, overlooking this component leads to smaller inferred gains.

Regional specialization. In the data, Chinese regions specialize in different broad sectors. For example, the manufacturing share in value added averages around 50% in the southeastern region that encompasses Shanghai, Jiangsu, Fujian, Zhejiang, and Guangdong provinces, but only 20-25% in Xinjiang and Qinghai autonomous regions in the northwest and Heilongjiang province in the northeast; on the other hand, the energy share averages around 14% in the latter group but only less than 1% along the southeastern coast. How valuable is the information on specialization?

To answer this question we recalibrate a model without specialization. Specifically, we assume all sectors within a region have the same productivity, i.e., $T_o^i = T_o^j = T_o, \forall o, j, i$, and pin down $\{T_o\}$ by matching the total sales of each city in the data. The input-output structure is kept the same as in the baseline model. We assume there is no international trade, so Equation (24) remains a valid characterization of the first order effects. Column 3 of Table 10 reports that the inferred gains in this model are 14% smaller than an otherwise similar model with regional specialization (Column 2).

Patterns of regional specialization matters because they contain information for the distribution of trade flows across pairs of domestic partners. Because of the strong spatial clustering

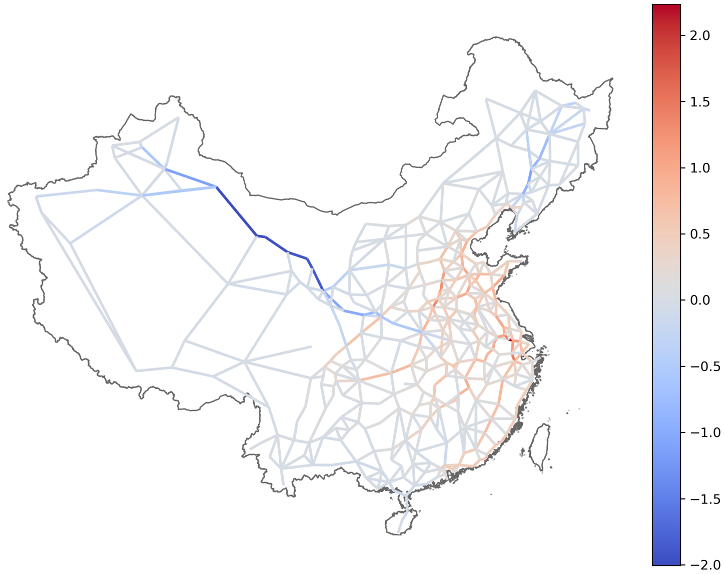


Figure 7: Differences in Shipment Value Shares, 'No Specialization' Minus 'Baseline'
Note: The numbers are the differences in shipment value/GDP between Model (2) and Model (3). The values plotted include both expressway and regular road shipments. Cold colors indicate that there is less shipment in Model (3) than in Model (2).

of production, the calibrated productivity in the full model has a spatial correlation, too. As a result, regions tend to trade with partners that are far away. When comparative advantages are eliminated, the spatial clustering also disappears, so inter-city trade in the restricted model shifts towards partners that are closer to each other. Although both models are calibrated to generate the same average shipment distance, this simple statistic does not capture all these patterns. Indeed, Figure 7 plots the change in shipment intensities between city pairs from Model (2) to Model (3). The segments that see the biggest decrease in inferred shipments are the ones connecting the northwest and northeast—the energy producing area—to the center of the country with a heavy manufacturing presence; the segments that see an increase in inferred shipments are the ones connecting regions within the center and the east of China. As a result, Model (3) infers higher gains for expressway segments in the center of the country and lower gains for projects connecting the center to the northeast and northwest—regions with very different comparative advantages. Whether the model underestimates or overestimates the return to a specific project thus depends crucially on where a project is. For the actual projects built during the decade, the balance comes down to an underestimation of the welfare gains by 14%.

Transportation intensity. The comprehensive price information from the customs data allows us to incorporate sector heterogeneity in transport costs. To demonstrate its relevance, we set $\mu = 0$ and then calibrate the model to match both city-level sales and the average shipment distance. We then conduct the same exercise as before. Under the assumption of homogeneous transport cost across sectors, the inferred gains are down from Model (3) by two-fifth to 3.18%.

At first glance, this might seem surprising, as with a large enough number of regions and road segments, the law of large numbers should have kicked in and the heterogeneity in transport intensity across sectors could be washed out. The reason why sector heterogeneity is not simply washed out is, when calibrated to match the same average shipment distance, Model (3) infers systematically higher *values* of shipment compared to Model (4). More specifically, with heterogeneity in trade costs, for the same level of inter-city shipment, Model (3) will predict a higher fraction of them in lighter sectors (with lower weight-to-value ratios) because they incur lower shipping costs in Model (3) but not in Model (4). Because the welfare gains are, to the first order, proportional to the value of goods but not their weights, the model with sector transport intensities predicts larger welfare gains.

Intermediate inputs. In the final comparison, we further shut down intermediate inputs in production by assuming the labor shares (β^i) are 1 in all industries. The welfare gains inferred by this model decline by three-quarters to around 0.7%. This difference can be understood by inspecting Equation (25).

$$\frac{X^i_{od}}{Y} = \frac{X^i_{od}}{\sum_i \sum_{o,d} X^i_{od}} \cdot \frac{\sum_i \sum_{o,d} X^i_{od}}{Y}. \tag{25}$$

For a simple example, assume that all regions o and d are symmetric, with positive but symmetric inter-regional transport costs. When calibrated to match the average shipment distance, Models (4) and (5) generate the same trade intensity, i.e., $\frac{X^i_{od}}{\sum_i \sum_{o,d} X^i_{od}}$. However, in the model without intermediate inputs, the overall absorption $\sum_i \sum_{o,d} X^i_{od}$ is equal to the GDP, whereas in the model with intermediate inputs, the overall absorption is several (around three in our calibration) times of the GDP. As a result, the inferred value of $\frac{\sum_i \sum_{o,d} X^i_{od}}{Y}$ is too small in the model without intermediate inputs. By assuming away intermediate inputs, the restricted model overlooks that goods are traded multiple times on the road, which amplifies the gains from reductions in transport cost.[41]

6.3 Cost-Benefit Analysis

The above comparisons underscore the importance of incorporating all necessary model ingredients. We now use the full model to conduct a cost-benefit analysis of the overall expressway network expansion and a few mega projects.

To this end, we collect the total investments on the expressway network during 1999-2010 and

[41] Although it is well known that the inferred gains from international trade are larger when intermediate goods are introduced (Caliendo and Parro, 2015; Costinot and Rodríguez-Clare, 2014), we show that for the evaluation of domestic infrastructure projects, this insights matters at least as much, if not more. In recent work, Baqaee and Farhi (2019a) shows that if the true underlying model is one with intermediate goods, and the researcher specifies a model without intermediate goods, then calibrating the specified model to match trade over GDP ratio (as opposed to the theory-consistent target under this model, trade over absorption/production) gives a better approximation to the true gains from trade. In our setting, this approach (one that changes the target, but not the model) runs into two practical difficulties. First, reliable inter-regional trade data are lacking, so we cannot directly measure trade/value added at the regional level. Second, even when the data are available, at the micro level, this measure could be easily above one, which a model without input-output linkages cannot accommodate. In our baseline economy, for example, this ratio is around 1.45 for the tradable sector as a whole.

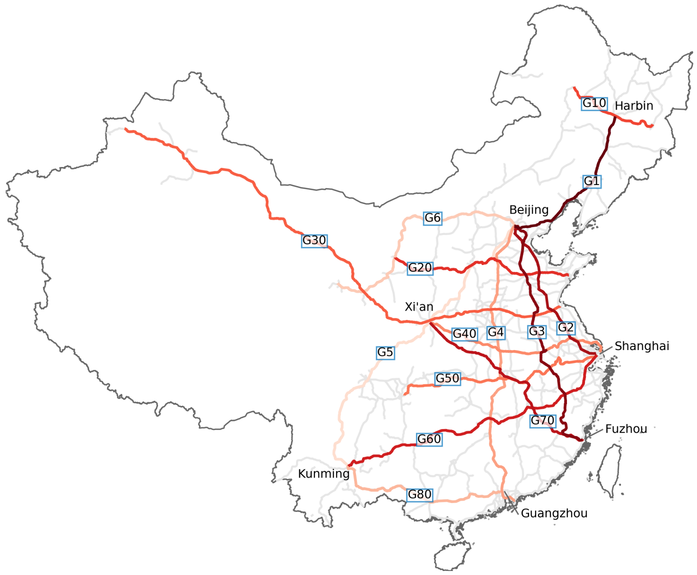


Figure 8: Mega Expressway Projects in China

Note: Projects with higher returns are plotted with darker colors. Some segments were completed by 1999 (most of G4 and G10); the newly built segments of all selected projects during 1999-2010 account for 43.7% of the total length built during this period.

infer the investment on individual projects. The raw data are from the yearly bulletin of road and waterway transport development (Ministry of Transport of the People's Republic of China, 2000-2010). We deflate it yearly using the inflation rate for capital investment. Measured in 2010 price, the cumulative investment in inter-city highway projects during the decade is around 570 billion USD, accounting for about 10% of the 2010 GDP. To compare this cost to discounted future benefits, we assume the annual depreciation rate for expressways is around 10.5% as in Bai et al. (2006). Given that the expressway network is planned by the central government, whose opportunity cost is to direct investment elsewhere, a natural choice for the discount rate is the return to capital in the overall Chinese economy. We assume this return is around 10%.[42]

Assuming all investment expenditures are made in 2010, then, the discounted future welfare gains ($5.6\% \times [1 + 0.8 + 0.8^2 ...]$) is around 28% of the 2010 GDP, implying a net return of about 180%: even after taking into account the high opportunity cost in a growing economy like China, the expressway investment generates a huge net return. In comparison, if we had used a simple one sector model for the evaluation as in most existing quantitative studies, our conclusion would have been that the investment led to 63% net losses ($\approx \frac{0.74\%}{0.2} / 10\% - 1$).

A few mega projects more than 1000 kilometers long form the backbone of the entire expressway network. Figure 10 plots 14 such projects. Some of them connect the north to the south. For example, G1 connects Beijing to the Northeast, passing through industrial centers such as Shenyang, Changchun, and ending at Harbin. G2-G5, on the other hand, connect Beijing to the

[42] Bai et al. (2006) finds that between 1998 and 2005, the return to capital is around 20%, a level that seems unsustainable especially given the secular stagnation in much of the developed world. To be conservative we chose to use 10%.

Table 11: Costs and Benefits of 14 Mega Projects

ID	Length (km)	Cost as % GDP	Cost per km (million)	Welfare Gains (%)	Net return to investment	% Change in dom. trade/GDP	% Change in Export/GDP
G1	1533.61	0.30	77.71	0.52	772.65%	0.60	0.94
G2	1768.29	0.38	85.94	0.45	497.92%	0.16	1.28
G3	2513.38	0.54	85.53	0.79	636.11%	0.65	4.37
G4	2924.88	0.65	89.14	0.40	204.45%	0.46	1.12
G5	2829.75	0.73	103.16	0.26	79.63%	0.26	0.51
G6	2095.37	0.38	72.26	0.17	118.86%	0.12	0.54
G10	891.73	0.15	67.25	0.12	286.87%	0.16	0.68
G20	1688.68	0.31	74.08	0.25	295.70%	0.30	0.73
G30	4356.49	0.85	78.04	0.63	270.24%	1.26	0.77
G40	1727.03	0.34	78.43	0.22	223.44%	0.46	0.93
G50	1936.36	0.38	79.61	0.26	234.74%	0.58	1.06
G60	2662.22	0.48	72.99	0.54	453.53%	1.08	2.13
G70	1706.35	0.38	89.62	0.43	465.28%	0.52	3.49
G80	1378.30	0.30	88.62	0.15	142.51%	0.23	0.83
Total	30012.46	6.16	-	5.16	-	6.84	19.37

Note: Each row corresponds to a counterfactual experiment by removing from the 2010 expressway network a mega expressway project referred by 'ID'. The statistics are calculated by comparing the benchmark equilibrium and the counterfactual equilibrium.

South with Shanghai (G2), Fuzhou (G3), Guangzhou (G4), and Kunming (G5), respectively. A few others connect the coastal areas to the center and the west of the country. G40, for example, links Shaanxi province, an important coal producing region, with Shanghai.

We evaluate costs and benefits for each of these projects. In the absence of a consistently defined cost measure for individual projects, we follow Faber (2014) and adopt a formula based on the engineering literature linking the *relative* construction cost of a segment to the average slope of the terrain and whether it passes water or wetland areas. We use the formula to evaluate all the segments constructed between 1999 and 2010 as a function of an unknown *level* coefficient and choose this coefficient so that the total cost of these segments is equal to the aggregate investment (10% of GDP). Appendix A.3 provides more details. The output of this procedure is the cost estimate for each of the 14 mega projects.

The third column of Table 11 reports the cost per kilometer for these projects. The most expensive project is G5, which passes through the rugged terrains in the southeast. Stretching across the flat northeastern plain in the other end of the country, G10 costs the least per kilometer. The average cost across all projects constructed in this period is around 80 million *yuan* per kilometer. This number is in the same ballpark as the best available evidence we can find.[43]

We evaluate the benefit of each project. Columns 4 and 5 report the per-period welfare gains and the net return to investment of these projects. All of them generate large returns. We indicate higher net gains using darker color in Figure 10. The projects with the highest returns are north-

[43]Most construction costs we can find online are for projects completed well before 2010. The website `http://news.roadcost.com/News/20120216/180.html` (in Chinese) discloses an audit report of expressway projects in Fujian province in 2011 Quarter one, according to which on average the cost is 80 million per kilometer.

south expressway lines (G1, G2, and G3). G5, running from Xi'an to Kunming, generates the lowest return, in part due to its hefty price tag. The last two columns report the change in domestic and international trade as a share of GDP after each project is completed. The projects that had the biggest impact on domestic trade are G30 and G60, which stretch across the vast central China, likely because they connect areas with different comparative advantages. On the other hand, the projects that had the largest impacts on export are G3, G60, and G70—roads that connect interior China to ports like Shanghai and Fuzhou.

The last row of the table reports the sum of each column. Despite that in terms of monetary investment, these project account only about 60% of the investment made to expressway during the decade, the sum of the marginal gains from these projects are around 5.2%, more than 90% of the gains from the entire network. And their collective impacts on export over GDP is 19%, higher than the effect of the entire network. That the sum of marginal effects of individual projects—assuming all existing projects have been built—is higher than the aggregate effect hints at significant complementarity between projects. This result highlights the importance of taking into account the interaction among regions and transport infrastructures in welfare evaluation.

To summarize, large return heterogeneity notwithstanding, until 2010 the expressway network in China was worth every penny of the investment. More recently, there has been a heated discussion among the popular press on whether China 'over-invested' in transport infrastructure. We should note that our finding does not necessarily apply to the latest wave of investment. Indeed, as the major population centers have been connected, building roads in the more mountainous areas, usually with redistributive motives, might incur higher costs while generating smaller returns.

7 First Order Measurement and Second Order Correction

From Proposition 2, all we need for inferring the first order welfare gains from a transport projects is the structural elasticities and the value of shipments along each segment. We chose to evaluate the full model with all three ingredients in part because it allows us to infer the value of shipments in GDP accurately, as demonstrated in Section 6.2. In this section, we further argue that because the expressway expansion represents a large shock to both individual links and the overall network, even with perfect data on shipment flows, the first order approach introduces a significant bias, and that the second order correction to the routing problem in Proposition 2 improves the accuracy materially. This second order characterization thus offers a convenient way to evaluate and compare many large hypothetical expressway projects. This could be useful, for example, when researchers and practitioners need to identify the best choice among a large sets of potential expressway configurations.

7.1 First Order and Nonlinear Effects for a Local Segment

We start with a local project that adds an expressway segment from k to l. We shut down international trade in all experiments in this section so Equation (20) holds. And it specializes to:

$$\Delta W = \overbrace{\underbrace{\sum_i \sum_{o,d} -\frac{X^i_{od}}{Y} \pi^{road}_{od} \pi^{kl}_{od} \Delta \log \iota_{kl}}_{\text{FO effect on shipments of } k \to l} \underbrace{- \frac{1}{2} \sum_i \sum_{o,d} \frac{X^i_{od}}{Y} \cdot \frac{\partial^2 \log \tilde{\tau}^i_{od}}{(\partial \log \iota_{kl})^2} (\Delta \log \iota_{kl})^2}_{\text{SO effect from routing}} + HO_R}^{\text{FO Trade}} + SO_T, \tag{26}$$

where the second order effect from routing is given by Equation (18). Note first that because $\theta >> \theta_M$, the sign of $\frac{\partial^2 \log \tilde{\tau}^i_{od}}{(\partial \log \iota_{kl})^2}$ depends mainly on the first term of Equation (18), which is negative. Intuitively, when re-optimization is allowed, when an expressway segment is removed, trucks originally taking the expressway from k to l might opt out to other routes. Starting from an equilibrium with expressway $k \to l$ present and inferring the welfare gains using the first order effect will thus overestimate the losses due to the removal of the expressway.[44]

Because with the parameterized full model, we can directly calculate the change in trade cost, $\Delta \log \tau^i_{od}$, we are also able to evaluate the sum of the first three terms, which captures the full effects of routing on trade costs. We label this sum 'FO Trade' in Equation (26). It stands for the inferred first order effect from trade directly, if the changes in realized effective trade costs are measured. The difference between FO trade and the prediction of the full model is the higher order equilibrium effects from the changes in trade cost, labeled SO_T.

We use Equation (26) to evaluate the quantitative significance of various forces. We consider the top 100 pairs of adjacent cities connected by both regular roads and expressways with the highest values of shipment. We conduct 100 experiments, in each of which we remove one of the expressway segment and calculate the aggregate welfare gains from that individual segment. We then compare the full effect of each segment to different approximations assuming all data at the calibrated equilibrium are observed perfectly.

Figure 9 plots the results. The horizontal is the full welfare effect. The range of welfare gains is between 0.007% to 0.06%. Even within the busiest subset of the road network, some segments are eight times as much important as others—a Zipf's law for inter-city traffic. The vertical axis presents predictions for welfare gains based on various approximations. The circles denote the first order cost savings, calculated directly using the value of actual shipments on each segment. They correspond to the first term in Equation (26). As anticipated, all circles lie above the 45 degree line, indicating overestimation of the losses from an expressway removal. The biases are smaller in percent for segments that are less busy, with the average absolute percentage error being around 27%.

The diamonds further incorporate the second order effect from re-routing among road net-

[44] Similarly, starting from an equilibrium and inferring the gains from the addition of an expressway segment ex-ante will underestimate the gains.

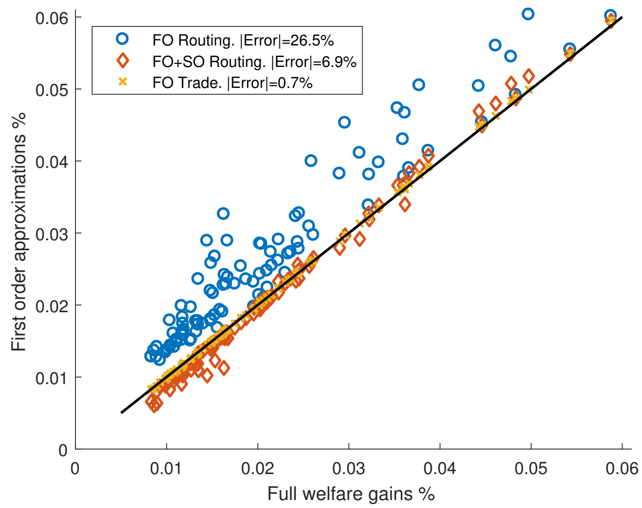


Figure 9: Nonlinear Welfare Gains v.s First Order Approximations

Note: Each point corresponds to an experiment with one expressway segment removed. The sample segments are the top 100 busiest segments in the baseline equilibrium. 'Error' reported in the legend is the mean absolute value of the percentage difference between each approximation and the full nonlinear effect.

works, assuming trade flows do not change. They correspond to 'SO effect from routing' in Equation (26). The approximation is now much more accurate, centering tightly around the 45 degree line. On average, the mean absolute percentage error of this approximation falls by three quarters to 7%. Finally, the crosses incorporate all responses in routing (corresponding to 'FO Trade' in Equation (26)). This further improves the quality of the approximation, reducing the mean absolute error to less than 1%.

7.2 Interactions Between Expressway Segments

When we move from analyzing individual segment $k \to l$ to a large expressway project with many segments $\mathbb{C}$, in addition to the approximation error documented in Figure 9, the first order approximation also misses the interaction between segments. Formally, we can decompose SO_R in Equation (20) into the sum of two terms. The first term captures the second order routing effect from self; the second captures the cross-derivatives between different segments, as below:

$$SO_R = -\frac{1}{2}\sum_i \sum_{o \neq d} \frac{X^i_{od}}{Y} \sum_{kl \in \mathbb{C}} [\underbrace{\frac{\partial^2 \log \tilde{\tau}^i_{od}}{(\partial \log \iota_{kl})^2}(\Delta \log(\iota_{kl}))^2}_{\text{Own } SO_R} + \underbrace{\sum_{k'l' \in \mathbb{C}, k'l' \neq kl} \frac{\partial^2 \log \tilde{\tau}^i_{od}}{\partial \log \iota_{kl}\, \partial \log \iota_{k'l'}} \Delta \log(\iota_{kl}) \Delta \log(\iota_{k'l'})}_{\text{Cross } SO_R}]. \quad (27)$$

While the sign of 'Own SO_R' is always negative, the sign of 'Cross SO_R' is ambiguous and depends on the network structure, as discussed in Section 4.6. We illustrate two cases in Figure 10. Consider one of the busiest expressway segments, the edge between Laiwu and Linyi, colored solid black in the map. The colors of other edges indicate their cross-derivative term with the edge between Laiwu and Linyi. Cold colors indicate that the cross-derivative is negative, in which

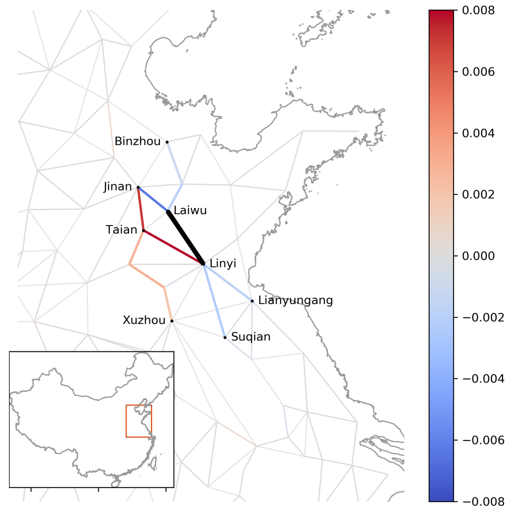


Figure 10: Complementarities and Substitutions between Segments: An Example

Note: The selected road segment is from Laiwu to Linyi, colored black. The map shows the cross derivative between each segment and the selected one (Laiwu to Linyi). Warm colors indicate that the cross derivative is positive, suggesting that an expressway between Laiwu and Linyi would draw traffic away from that segment. Cold colors indicate the opposite. Numbers are in percentage points of aggregate welfare.

case a new expressway between Laiwu and Linyi will increase the traffic on the segment, as they jointly form a longer route. For example, the segment between Jinan and Laiwu.[45] On the other hand, warm colors indicate that a segment is a substitute to the expressway between Laiwu and Linyi. For example, the route from Jinan to Xuzhou.

For a systematic evaluation of various channels, we decompose Equation (20) as follows:

$$\underbrace{\Delta W}_{0.032} = \underbrace{\text{FO effect}}_{138\%} + \underbrace{\text{Own SO}_R}_{-43\%} + \underbrace{\text{Cross SO}_R}_{9\%} + \underbrace{\text{HO}_R + \text{SO}_T}_{-5\%}. \tag{28}$$

The first term on the right side of the equation is simply the sum of the first order effects of individual segments. The second and third correspond to the two second order effects from routing, which we calculate using Equation (27). We treat HO_R and SO_T as a residual term, capturing the remaining difference between the full nonlinear effect and the analytical predictions.

The importance of the cross-derivative force crucially depends on the segments being jointly evaluated. As an example, we calculate the nonlinear welfare gains of the segments constructed between 1999 and 2010. Because the first order approximation can be evaluated only when k and l are connected by both regular roads and expressways (otherwise the edge cost increases to infinity when the expressway is removed), in this experiment we focus on expressways between cities also connected by regular roads. In total, these expressway segments generate 3.2% welfare gains. The first order effects based on ex-post traffic overestimate this number by 38%. The own

[45]The cross-derivative being negative means that the second order effect has the opposite sign of the first order effect. So for an ex-post evaluation of the welfare losses from the removal of an expressway, it adjusts down the inferred first order effect. The opposite is true when the cross-derivative is positive.

substitution effects more than correct this bias; the cross-substitution effects add another 10% to the inferred welfare gains. The overall positive cross-routing effects mask some segments that strongly complement each other and others that compete for traffic.

Once both second order effects are included, what is left for the approximation error is mere 5%. Equation (20) thus gives a second order correction that can be used to evaluate the gains from large projects, which could be any combinations of edges, without having to solve for the counterfactual equilibrium.

8 Conclusion

This paper proposes a method to evaluate the effect of transport infrastructure on domestic trade costs that circumvents the lack of reliable domestic trade data in many countries—by using the route choice of exporters. We combine this method and a spatial equilibrium model to study the aggregate welfare effects of the $50,000$ kilometer expressway construction taking place between 1999 and 2010 in China. We find a large positive net return to the investment and shows that overlooking the model ingredients can lead to the opposite conclusion. We further provide a second order characterization of the aggregate welfare gains from expressway networks. Once the full model is parameterized, this characterization allows convenient evaluations of expressway networks with a higher accuracy than the commonly used first order approach, without the need to solve for full counterfactual equilibria.

The reason why having the full model is important is that in the absence of direct measurements of domestic trade, we rely on the model structure and other data in inferring the value of inter-city shipments. The information used in our evaluations (e.g., sectoral production and the full structure of the transport network) might not be accessible in some empirical settings; investigating the minimum sufficient information to infer welfare gains from large projects, especially ex-ante, is thus a valuable venue for future research. Throughout the paper, we also abstract from a number of interesting channels. For example, the effects of expressways on technological diffusion and the dynamic responses through regional growth. Incorporating these channels might yield additional insights on the role of transport infrastructure.

References

Abdelwahab, Walid M, "Elasticities of mode choice probabilities and market elasticities of demand: evidence from a simultaneous mode choice/shipment-size freight transport model," *Transportation Research Part E: Logistics and Transportation Review*, 1998, *34* (4), 257–266.

Alder, Simon and Illenin Kondo, "Political Distortions and Infrastructure Networks in China: A Quantitative Spatial Equilibrium Analysis," *Working Paper*, 2019.

Allen, Treb and Costas Arkolakis, "Trade and the Topography of the Spatial Economy," *The Quarterly Journal of Economics*, 2014, *129* (3), 1085–1140.

— and — , "The welfare effects of transportation infrastructure improvements," *NBER Working Paper No. 25487*, 2019.

Asturias, Jose, Manuel García-Santana, and Roberto Ramos, "Competition and the welfare gains from transportation infrastructure: Evidence from the Golden Quadrilateral of India," *Journal of the European Economic Association*, 2018.

Atkin, David and Dave Donaldson, "Who's getting globalized? The size and implications of intra-national trade costs," *NBER Working Paper No. 21439*, 2015.

Bai, Chong-En, Chang-Tai Hsieh, and Yingyi Qian, "The Return to Capital in China," *Brookings Papers on Economic Activity*, 2006, *2006* (2), 61–88.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian, "On the Road: Access to Transportation Infrastructure and Economic Growth in China," Technical Report 2012.

Baqaee, David and Emmanuel Farhi, "Networks, barriers, and trade," *NBER Working Paper No. 26108*, 2019.

Baqaee, David Rezza and Emmanuel Farhi, "The macroeconomic impact of microeconomic shocks: beyond Hulten's Theorem," *Econometrica*, 2019, *87* (4), 1155–1203.

Baum-Snow, Nathaniel, J Vernon Henderson, Matthew A Turner, Qinghua Zhang, and Loren Brandt, *Highways, market access and urban growth in China*, SERC, Spatial Economics Research Centre, 2016.

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

Caliendo, Lorenzo and Fernando Parro, "Estimates of the Trade and Welfare Effects of NAFTA," *The Review of Economic Studies*, 2015, *82* (1), 1–44.

— , — , Esteban Rossi-Hansberg, and Pierre-Daniel Sarte, "The impact of regional and sectoral productivity changes on the US economy," *The Review of Economic Studies*, 2014.

Coşar, A Kerem and Banu Demir, "Domestic road infrastructure and international trade: Evidence from Turkey," *Journal of Development Economics*, 2016, *118*, 232–244.

— and Pablo D Fajgelbaum, "Internal geography, international trade, and regional specialization," *American Economic Journal: Microeconomics*, 2016, *8* (1), 24–56.

Cosar, A Kerem, Banu Demir, Devaki Ghose, and Nathaniel Young, "Road Capacity, Domestic Trade and Regional Outcomes," *Mimeo*, 2019.

Costinot, Arnaud and Andrés Rodríguez-Clare, "Trade theory with numbers: Quantifying the consequences of globalization," in "Handbook of international economics," Vol. 4, Elsevier, 2014, pp. 197–261.

Dingel, Jonathan I, "The determinants of quality specialization," *The Review of Economic Studies*, 2016, *84* (4), 1551–1582.

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

— and Richard Hornbeck, "Railroads and American economic growth: A "market access" approach," *The Quarterly Journal of Economics*, 2016, *131* (2), 799–858.

Eaton, Jonathan and Samuel Kortum, "Technology, geography, and trade," *Econometrica*, 2002, *70* (5), 1741–1779.

Faber, Benjamin, "Trade integration, market size, and industrialization: evidence from China's National Trunk Highway System," *Review of Economic Studies*, 2014, *81* (3), 1046–1070.

Facchini, Giovanni, Maggie Y Liu, Anna Maria Mayda, and Minghai Zhou, "China's "Great Migration": The impact of the reduction in trade policy uncertainty," *Journal of International Economics*, 2019.

Fajgelbaum, Pablo and Stephen J. Redding, "External Integration, Structural Transformation and Economic Development: Evidence from Argentina 1870-1914," *NBER Working Paper No. w20217*, 2014.

Fajgelbaum, Pablo D and Edouard Schaal, "Optimal transport networks in spatial equilibrium," *NBER Working Paper No. 23200*, 2019.

Fan, Haichao, Yao Amber Li, and Stephen R Yeaple, "Trade liberalization, quality, and export prices," *Review of Economics and Statistics*, 2015, *97* (5), 1033–1051.

Fan, Jingting, "Internal geography, labor mobility, and the distributional impacts of trade," *American Economic Journal: Macroeconomics*, 2019, *11* (3), 252–88.

Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer, "The Next Generation of the Penn World Table," *American Economic Review*, 2015, *105* (10), 3150–3182.

Fogel, Robert William, *Railroads and American economic growth*, Johns Hopkins Press Baltimore, 1964.

Gu, Yizhen, Junfu Jiang Chang Zhang, and Ben Zou, "Subways and Road Congestion," *Working Paper*, 2018.

Hsu, Wen-Tai and Hongliang Zhang, "The fundamental law of highway congestion revisited: Evidence from national expressways in Japan," *Journal of Urban Economics*, 2014, *81*, 65–76.

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

Hummels, David, "Transportation costs and international trade in the second era of globalization," *Journal of Economic perspectives*, 2007, *21* (3), 131–154.

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

Ma, Lin and Yang Tang, "Geography, trade, and internal migration in china," *Journal of Urban Economics*, 2019.

Martincus, Christian Volpe, Jerónimo Carballo, and Ana Cusolito, "Roads, exports and employment: Evidence from a developing country," *Journal of Development Economics*, 2017, *125*, 21–39.

Ministry of Transport of the People's Republic of China, "The annual bulletin of road and water transport development," 2000-2010.

Morten, Melanie and Jaqueline Oliveira, "The effects of roads on trade and migration: Evidence from a planned capital city," *NBER Working Paper. Cambridge, MA, USA: National Bureau of Economic Research*, 2018.

Nagy, Dávid Krisztián, "City location and economic development," *Working Paper*, 2016.

National Bureau of Statistics, *China Transportation Statistical Yearbook*, Zhongguo Tongji Chubanshe,Beijing, 2010.

OECD, "Infrastructure investment (indicator)," *doi: 10.1787/b06ce3ad-en (Accessed on 13 November 2019)*, 2019.

Rauch, James E, "Networks versus markets in international trade," *Journal of international Economics*, 1999, *48* (1), 7–35.

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

— and Esteban Rossi-Hansberg, "Quantitative spatial economics," *Annual Review of Economics*, 2017, *9*, 21–58.

— and Matthew A Turner, "Transportation costs and the spatial organization of economic activity," in "Handbook of regional and urban economics," Vol. 5, Elsevier, 2015, pp. 1339–1398.

Severen, Christopher, "Commuting, Labor, and Housing Market Effects of Mass Transportation: Welfare and Identification," *FRB of Philadelphia Working Paper 18-14,*, 2018.

Small, Kenneth A, "Valuation of travel time," *Economics of transportation*, 2012, *1* (1-2), 2–14.

Tian, Yuan, "International Trade Liberalization and Domestic Institutional Reform: Effects of WTO Accession on Chinese Internal Migration Policy," *Working Paper*, 2019.

Tombe, Trevor and Xiaodong Zhu, "Trade, migration, and productivity: A quantitative analysis of china," *American Economic Review*, 2019, *109* (5), 1843–72.

Tsivanidis, Nick, "The Aggregate and Distributional Effects of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio," *Working Paper*, 2018.

Xu, Mingzhi, "Riding on the New Silk Road: Quantifying the Welfare Gains from High-Speed Railways," *Working Paper*, 2018.

Zhang, Yaxiong and Suchang Qi, "2007 China multi-regional input-output model," 2012.

Zi, Yuan, "Trade liberalization and the great labor reallocation," *Working Paper*, 2016.

Appendix For Online Publication

Valuing Domestic Transport Infrastructure: A View from the Route Choice of Exporters

Jingting Fan
Pennsylvania State University

Yi Lu
Tsinghua University

Wenlan Luo
Tsinghua University

Contents

A Data and Empirics

A.1 Defining City Coordinates

We define the location of a county by its center of mass using the county geographic information in the 2010 census. We weight the coordinates of all counties making up a prefecture city by their population to calculate an average coordinate, which we then define as the location of a prefecture city. For four provincial-level cities, Beijing, Shanghai, Tianjin, and Chongqing, we generate the coordinates by weighting the coordinates of their urban sub-divisions (districts). We exclude the rural sub-divisions in these provincial-level cities because the rural areas are large and have a disproportionate impact on the measured economic center.

A.2 Constructing Network Graphs

Our raw data consist of the geographic coordinates of city centers and the line strings of road networks (1999 and 2010 expressway, and 2007 regular road which is treated as time-invariant). To combine the data and the model, we first need to generate a network of cities with the roads connecting them. We do so separately for each of the three maps, according to the following procedures.

- **Define connected cities.** First, each city is defined as 'connected' in a map, if the center of the city is within the 30 kilometer radius of any roads on a map. Practically, it means measuring whether any of the coordinates characterizing roads from a map are within 30 kilometer of the city center.

- **Define connections between cities.** We 're-base' the coordinates of 'connected' cities to the nearest coordinates of the road network. For each pair of connected cities, we search for the shortest path between them along the road network using the Dijkstra's algorithm. If the shortest path between two cities does not pass through 30 kilometer radius of another city, we define the pair to be 'directly connected'.

- **Construct the graph.** We construct the graph in which cities and roads are nodes and edges as follows. We start with a collections of nodes. We draw a edge between two cities, if they are found to be 'directly connected' in the previous step. We define the length of the edge to be the great-circle distance between the two city centers. This effectively 'irons out' the local curvatures in constructing the network, which helps us eliminate measurement errors when comparing the expressway networks across two periods.[1] Note that in this graph, each city can only be connected to another city through its adjacent cities.

The left panel of Figure A.1 is the original digital maps for expressways and regular roads. The right panel is their network representation, which is the output of this step. They correspond to the network structure underlying $\mathbb{H}^{1999}$, $\mathbb{H}^{2010}$, and $\mathbb{L}$.

[1]The two expressway maps are digitized from the projection of published hard-copy maps, which introduce measurement errors that change the exact locations of roads.

(a) 1999 Expressway Map

(b) 1999 Expressway Network

(c) 2010 Expressway Map

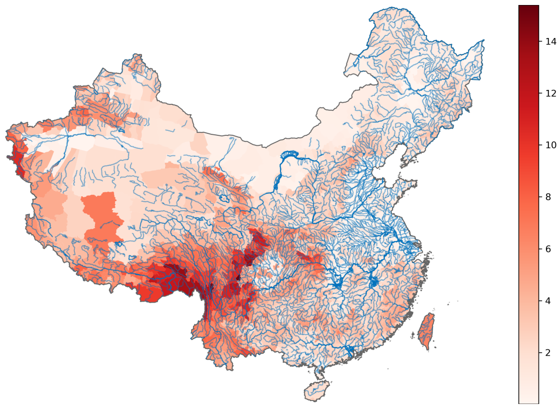
(d) 2010 Expressway Network

(e) Regular Road Map

(f) Regular Road Network

Figure A.1: From Road Maps to Road Networks

Note: Two cities are defined as connected on a road network if the shortest path connecting them on a road network does not pass a third city. The distance between two connected cities is then calculated as the greater circle distance between their city centers.



(a) Slopes and Water Areas

(b) Slopes of Expressway Segments

Figure A.2: Geography and Expressway Construction Costs

Note: The left figure plots the slope of land and the geographic distribution of water areas. The right panel plots the expressways in 2010, indicating using color the average slope for each 10-kilometer segment.

A.3 Backing Out Construction Costs for All Segments

We first cut expressways into 10-km segments. We calculate the average slope of each segment and determine whether the segment passes water.[2] We calculate the *relative* construction cost of segment i following a simple function from the transport engineering literature:

$$cost_i = 1 + slope_i + 25 \times PassWater_i,$$

which is similar to the one used Faber (2014), except that we abstract from the measure of existing buildings due to the lack of data. According to this formula, the cost of constructing a segment passing water costs 26 times as much as on a dry plain. The *level* of the construction cost is determined such that the total cost of the newly constructed segments from 1999 to 2010 is 9.92% of the 2010 GDP.

The total cost (9.92% of the 2010 GDP) is 3983 billion 2010 CNY. The total dry-plain equivalent distance of all roads constructed during this period is $453,447$ kilomter, so each dry-plain equivalent kilometer of expressway costs about 8.85 million 2010 CNY. The total length of expressway actually constructed during this period is $49,760$ km, so the average cost for each kilometer is around 80 million 2010 CNY. This cost is much higher than the dry-plain equivalent cost, reflecting that most of the projects during this decade pass rugged or water areas.

A.4 The Lists of Port and Major Cities

List of port cities: Tianjin, Dalian, Shanghai, Ningbo, Fuzhou, Xiamen, Qingdao, Guangzhou, Shenzhen, Zhuhai, Shantou.

List of major cities: Beijing, Tianjin, Shijiazhuang, Tangshan, Handan, Xingtai, Baoding, Cangzhou, Shenyang, Dalian, Changchun, Haerbin, Shanghai, Xuzhou, Suzhou, Nantong, Yancheng, Hangzhou, Wenzhou, Fuyang, Suzhou, Liuan, Quanzhou, Ganzhou, Jinan, Qingdao, Yantai, Weifang, Jining, Linyi, Liaocheng, Heze, Zhengzhou, Luoyang, Xinxiang, Nanyang, Shangqiu, Xinyang, Zhoukou, Zhumadian,

[2] 23.3% of the segments pass water areas.

Table A.1: Route Choice at the Sectoral Level

	(1)	(2)	(3)	(4)
	Sectoral PPML		Sectoral IV	
$dist_{ij,t}$	-0.387***		-0.107**	
	(0.074)		(0.046)	
-on express		-0.258***		-0.084
		(0.079)		(0.055)
-on regular		-0.393***		-0.119**
		(0.079)		(0.053)
Fixed Effects	*odi, oit, dit*	*odi, oit, dit*	*odi, oit, dit*	*odi, oit, dit*
Exclude major cities	yes	yes	yes	yes
Observations	11044	11044	10808	10808
First Stage KP-F statistic			1007.661	141.529

Notes: This table reports the robustness exercises of results in Table 2 to sectoral-level data. Sectors are defined at HS-2 level. All regressions control for the sector fixed effect and its interaction with other fixed effects. Columns 1 and 2 replicate Columns 4 and 5 of Table 2 using PPML; Columns 3 and 4 use the IV specification and replicate the first two columns of Table 2. Standard errors are clustered at city-port level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Wuhan, Huanggang, Changsha, Hengyang, Shaoyang, Changde, Guangzhou, Zhanjiang, Munidiqu, Chongqing, Chengdu, Nanchong, Zunyi, Bijiediqu, Xi'an.

A.5 Additional Results on Reduced-Form Regressions

Table A.1 reports additional robustness exercises for results reported in Table 2. All regressions use HS2 level data and exclude major cities from the sample. The first two columns use the PPML and replicate the results in Columns 4 and 5 of Table 2. The third and fourth columns use the IV based on the minimum-spanning tree and replicate the first two columns of Table 2. In all specifications, we control for the combinations of HS2 sector fixed effects with other fixed effects in Table 2.

B Model

B.1 Additional Properties of the Routing Block

Combining two routing matrices into one. We prove that when truck drivers choose from two networks (regular roads $\mathbb{L}$ and expressways $\mathbb{H}$), the average cost is as if drivers choose from one single combined network: $\mathbb{A} = \mathbb{L} + \mathbb{H}$.

We prove this by induction. First, consider the average cost of going from o to d among all routes with only one edge.

$$\tau_{od,1} = \Gamma(\frac{\theta-1}{\theta})\big([\mathbb{L}_{(o,d)}] + [\mathbb{H}_{(o,d)}]\big)^{-\frac{1}{\theta}} = \Gamma(\frac{\theta-1}{\theta})\big([\mathbb{A}_{(o,d)}]\big)^{-\frac{1}{\theta}}.$$

Note also that if the (o,d) element of both $\mathbb{L}$ and $\mathbb{H}$ are zero, then $\tau_{od,1} = \infty$, meaning there is no feasible one-edge path from o to d.

Assuming that the sum of (the $-\theta$ exponent of) cost from o to d across all paths with exactly N steps is $[\mathbb{A}^N_{(o,d)}]$, then the sum across all paths with exactly $N+1$ steps is:

$$[(\mathbb{A}^N \cdot \mathbb{H} + \mathbb{A}^N \cdot \mathbb{L})_{(o,d)}].$$

The first part sums across all the paths that first gets to an adjacent city of d in exactly N steps and then goes on to d through an expressway; the second part sums across all the paths that gets to an adjacent city of d in N steps and then goes on to d through a regular road.

The above expression equals exactly $[\mathbb{A}^{N+1}_{(o,d)}]$. In other words, $[\mathbb{A}^{N+1}_{(o,d)}]$ is the sum across all paths goes from o to d in exactly $N+1$ steps. The average cost across all paths is then:

$$\tau_{od} = \lim_{N\to\infty} \tau_{od,N} = \Gamma(\frac{\theta-1}{\theta})\big(\sum_{i=1}^{\infty}[\mathbb{A}^i]_{(o,d)}\big)^{-\frac{1}{\theta}} = \Gamma(\frac{\theta-1}{\theta})\mathbb{B}^{-\theta}_{(o,d)},$$

where $\mathbb{B} \equiv \big(\mathbb{I} - \mathbb{A}\big)^{-1}$, and $\mathbb{A} \equiv \mathbb{L} + \mathbb{H}$.

B.2 Definition of Equilibrium

We define the competitive equilibrium as a set of prices and quantities that satisfy a number of conditions as below.

Definition 1. *Given fundamentals $\{\tilde{\tau}^i_{od}, L_d, T^i_d\}$[3], a competitive equilibrium is: (1) consumption of sectoral final goods C^i_d, labor allocation l^i_d, uses of sectoral final goods as intermediate input m^{ij}_d, production of sectoral final goods Q^i_d, intermediate goods traded $\tilde{q}^i_{od}$, production of intermediate goods q^i_d; (2) prices of final goods P^i_d, import prices of intermediate goods p^i_{od}, costs of input bundles for producing intermediate goods $\hat{c}^i_o$, wages w_d, s.t.*

- *Consumers' optimization conditions hold:*

$$\alpha^i_d w_d L_d = P^i_d C^i_d. \tag{B.1}$$

- *Intermediate goods producers' optimization conditions hold:*

$$q^i_d = z^i_d [l^i_d]^{\beta^i_d} \prod_{j=1}^{S} [m^{ij}_d]^{\gamma^{ij}_d}$$

$$\hat{c}^j_o = \kappa^i_d w_d^{\beta^i_d} \prod_{j=1}^{S} [P^j_d]^{\gamma^{ij}_d}$$

$$P^j_o m^{ij}_o = \hat{c}^j_o \gamma^{ij}_o [l^i_o]^{\beta^i_o} \prod_k [m^{ik}_o]^{\gamma^{ik}_o} \tag{B.2}$$

$$w_d l^i_o = \hat{c}^j_o \beta^i_o [l^i_o]^{\beta^i_o} \prod_k [m^{ik}_o]^{\gamma^{ik}_o} \tag{B.3}$$

$$p^i_{od} = [\hat{c}^i_o \tilde{\tau}^i_{od}]/T^i_o,$$

where $\kappa^i_d = (\beta^i_d)^{-\beta^i_d} \prod_{j=1}^{S} (\gamma^{ij}_d)^{-\gamma^{ij}_d}$.

[3] The solution to the transport mode choice and the drivers' routing problem have been incorporated through the trade cost matrix $\tilde{\tau}^i_{od}$.

- *Final goods producers' optimization conditions hold:*

$$
\begin{aligned}
Q_d^i &= \Big(\sum_o [\tilde{q}_{od}^i]^{\frac{\sigma-1}{\sigma}}\Big)^{\frac{\sigma}{\sigma-1}} \\
\frac{\tilde{q}_{od}^i}{\tilde{q}_{o'd}^i} &= [\frac{p_{od}^i}{p_{o'd}^i}]^{-\sigma} \\
P_d^i &= \Big(\sum_o [p_{od}^i]^{1-\sigma}\Big)^{\frac{1}{1-\sigma}}.
\end{aligned}
\tag{B.4}
$$

- *Markets clear for labor, final goods, and intermediate goods:*

$$
\begin{aligned}
\sum_i l_d^i &= L_d && \textit{(Labor markets clear)} \\
\sum_d \tilde{\tau}_{od}^i \tilde{q}_{od}^i &= q_o^i && \textit{(Intermediates goods markets clear)} \\
\sum_i m_d^{ij} + C_d^j &= Q_d^j && \textit{(Final goods market clear).}
\end{aligned}
$$

B.3 Proof of Propositions

We first state a lemma characterizing the first order effect of the segment shipment cost on the trade cost via the road network. The lemma is one of the key results behind the tractability of the routing block in Allen and Arkolakis (2019), which we build on. We refer readers to the proof there.

Lemma B.1. *The entries of $\mathbb{A}$ and of its Leontief inverse $\mathbb{B} \equiv (\mathbb{I} - \mathbb{A})^{-1}$ satisfy*

$$
\frac{\partial \log b_{od}}{\partial \log a_{kl}} = \frac{b_{ok} \cdot a_{kl} \cdot b_{ld}}{b_{od}},
$$

where a_{kl} and b_{od} are the (k,l) and (o,d) elements of $\mathbb{A}$ and $\mathbb{B}$ respectively.

We now apply Lemma B.1 to prove Proposition 1.

Proof of Proposition 1.

Proof. Consider

$$
\begin{aligned}
&\log[\exp(-\theta\kappa^H dist_{kl}) + \exp(-\theta\kappa^L dist_{kl})] \\
&= -\theta\kappa^H dist_{kl} + \log\Big(1 + \exp(-\theta(\kappa^L - \kappa^H) dist_{kl})\Big) \\
&= -\theta\kappa^H dist_{kl} + o\Big(\exp(-\theta(\kappa^L - \kappa^H) dist_{kl})\Big) \\
&\approx -\theta\kappa^H dist_{kl}
\end{aligned}
$$

where the approximation on the last line applies that $\kappa^L > \kappa^H$ and θ is large. Therefore we have

$$
\begin{aligned}
\Delta \log(\iota_{kl}) &= -\frac{1}{\theta}\Big(\log[\exp(-\theta\kappa^H dist_{kl}) + \exp(-\theta\kappa^L dist_{kl})] - \log[\exp(-\theta\kappa^L dist_{kl})]\Big) \\
&\approx (\kappa^H - \kappa^L) dist_{kl},
\end{aligned}
$$

which proves Equation (16).

Equation (17) directly applies the second-order Taylor expansion w.r.t. $(\log(\iota_{kl}))_{kl \in \mathbb{C}}$.

For Equation (18), recall that $\tau^i_{od} = (\frac{h_i}{h_0})^\mu b_{od}^{-1/\theta}$ and $\iota_{kl} = a_{kl}^{-1/\theta}$. Applying Lemma B.1, we have

$$\pi^{kl}_{od} \equiv \frac{\partial \log \tau^i_{od}}{\partial \log \iota_{kl}} = \frac{\partial \log b_{od}}{\partial \log a_{kl}} = \frac{b_{ok} \cdot a_{kl} \cdot b_{ld}}{b_{od}}.$$

Then from

$$\tilde{\tau}^i_{od} = \Gamma(\frac{\theta_M - 1}{\theta_M})[(\bar{\tau}^i_{od})^{-\theta_M} + (\tau^i_{od})^{-\theta_M}]^{-1/\theta_M}$$

we have

$$\pi^{road}_{od} \equiv \frac{\partial \log \tilde{\tau}^i_{od}}{\partial \log \tau^i_{od}} = \frac{(\tau^i_{od})^{-\theta_M}}{(\bar{\tau}^i_{od})^{-\theta_M} + (\tau^i_{od})^{-\theta_M}}.$$

For Equation (19), first it can be shown that

$$\frac{\partial \log(\pi^{road}_{od})}{\partial \log \tau^i_{od}} = -\theta_M(1 - \pi^{road}_{od}).$$

Combining with

$$\frac{\partial \log \tau^i_{od}}{\partial \log \iota_{k'l'}} = \pi^{k'l'}_{od}$$

we have

$$\frac{\partial \log(\pi^{road}_{od})}{\partial \log \iota_{k'l'}} = -\theta_M(1 - \pi^{road}_{od})\pi^{k'l'}_{od}. \tag{B.5}$$

Next start with

$$\log \pi^{kl}_{od} = \log b_{ok} + \log a_{kl} + \log b_{ld} - \log b_{od}$$

we get

$$\frac{\partial \log \pi^{kl}_{od}}{\partial \log \iota_{k'l'}} = -\theta(\pi^{k'l'}_{ok} + \pi^{k'l'}_{ld} - \pi^{k'l'}_{od} + \mathbb{1}(kl = k'l')), \tag{B.6}$$

which applies Lemma B.1 to each of the terms. Combining (B.5) and (B.6) we arrive at

$$\frac{\partial(\pi^{road}_{od}\pi^{kl}_{od})}{\partial \log \iota_{k'l'}} = \pi^{road}_{od}\pi^{kl}_{od}[-\theta_M(1 - \pi^{Road}_{od})\pi^{k'l'}_{od} - \theta(\pi^{k'l'}_{ok} + \pi^{k'l'}_{ld} - \pi^{k'l'}_{od} + \mathbb{1}(kl = k'l'))].$$

This proves Equation (19). To see that the expression is symmetric in kl and $k'l'$, consider

$$
\begin{aligned}
\pi_{od}^{kl}\pi_{ok}^{k'l'} &= \frac{b_{ok}a_{kl}b_{ld}}{b_{od}}\frac{b_{ok'}a_{k'l'}b_{l'k}}{b_{ok}} \\
&= \frac{b_{ok'}a_{k'l'}b_{l'k}a_{kl}b_{ld}}{b_{od}} \\
&= \frac{b_{ok'}a_{k'l'}b_{l'd}}{b_{od}}\frac{b_{l'k}a_{kl}b_{ld}}{b_{l'd}} \\
&= \pi_{od}^{k'l'}\pi_{l'd}^{kl}
\end{aligned}
$$

Similarly, $\pi_{od}^{kl}\pi_{ld}^{k'l'} = \pi_{od}^{k'l'}\pi_{ok'}^{kl}$. Therefore,

$$
-\theta\pi_{od}^{kl}(\pi_{ok}^{k'l'} + \pi_{ld}^{k'l'} - \pi_{od}^{k'l'} + \mathbb{1}(kl = k'l')) = -\theta\pi_{od}^{k'l'}(\pi_{ok'}^{kl} + \pi_{l'd}^{kl} - \pi_{od}^{kl} + \mathbb{1}(kl = k'l')).
$$

And we have

$$
\begin{aligned}
\frac{\partial(\pi_{od}^{road}\pi_{od}^{kl})}{\partial\log\iota_{k'l'}} &= \pi_{od}^{road}[-\theta_M(1-\pi_{od}^{Road})\pi_{od}^{k'l'}\pi_{od}^{kl} - \theta\pi_{od}^{kl}(\pi_{ok}^{k'l'} + \pi_{ld}^{k'l'} - \pi_{od}^{k'l'} + \mathbb{1}(kl = k'l'))] \\
&= \pi_{od}^{road}[-\theta_M(1-\pi_{od}^{Road})\pi_{od}^{kl}\pi_{od}^{k'l'} - \theta\pi_{od}^{k'l'}(\pi_{ok'}^{kl} + \pi_{l'd}^{kl} - \pi_{od}^{kl} + \mathbb{1}(kl = k'l'))] \\
&= \frac{\partial(\pi_{od}^{road}\pi_{od}^{k'l'})}{\partial\log\iota_{kl}}.
\end{aligned}
$$

□

Proof of Proposition 2. We now establish the First Welfare Theorem with the appropriate choice of Pareto weights for the social planner's problem. The Pareto weights, which are equal to the regional value added in the corresponding competitive equilibrium, are used to evaluate the aggregate welfare and to calculate the change of welfare between equilibria. The proof can be viewed as an application of the standard equivalence result in an Arrow-Debreu equilibrium with production, viewing trade just as another form of production technology—to convert from origin goods to destination goods—with the efficiency determined by the inverse of the trade cost.

Lemma B.2. *The allocations in the competitive equilibrium can be replicated by the solution to the following social planner's problem*

$$
W = \max_{l_d^i, C_d^i, m_d^{ij}, q_d^i, Q_d^i, \tilde{q}_{od}^i} \sum_d \omega_d \log\left(\prod_{i=1}^{S}[C_d^j]^{\alpha_d^i}\right) \tag{B.7}
$$

subject to

$$q^i_d = T^i_d [l^i_d]^{\beta^i_d} \prod_{j=1}^{S} [m^{ij}_d]^{\gamma^{ij}_d} \qquad \textit{(Production of intermediate goods)}$$

$$Q^i_d = \Big(\sum_o [\tilde{q}^i_{od}]^{\frac{\sigma-1}{\sigma}}\Big)^{\frac{\sigma}{\sigma-1}} \qquad \textit{(Production of final goods)}$$

$$\sum_i l^i_d = L_d \qquad \textit{(Resource constraints for labor markets)}$$

$$\sum_d \tilde{\tau}^i_{od} \tilde{q}^i_{od} = q^i_o \qquad \textit{(Resource constraints for intermediates goods)}$$

$$\sum_i m^{ij}_d + C^j_d = Q^j_d \qquad \textit{(Resource constraints for final goods),}$$

where the Pareto weights are given by $\omega_d = \frac{w_d L_d}{Y}$*, with* w_d *being the equilibrium nominal wage of location d and* $Y = \sum_d w_d L_d$ *being the aggregate nominal GDP under the competitive equilibrium.*

Proof. The Lagrangian (with multipliers λ^j_d, μ^i_o, v^i_d, n_d)for the planner's problem is[4]

$$\mathcal{L} = \sum_d \omega_d \log\Big(\prod_{i=1}^{S}[C^i_d]^{\alpha^i_d}\Big) + \sum_{d,j}\lambda^j_d\Big(Q^j_d - C^j_d - \sum_i m^{ij}_d\Big) + \sum_{o,i}\mu^i_o\Big(T^i_o[l^i_o]^{\beta^i_o}\prod_{j=1}^{S}[m^{ij}_o]^{\gamma^{ij}_o} - \sum_d \tilde{\tau}^i_{od}\tilde{q}^i_{od}\Big)$$
$$+ \sum_{d,i} v^i_d\Big((\sum_o[\tilde{q}^i_{od}]^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}} - Q^i_d\Big) + \sum_d n_d\Big(L_d - \sum_i l^i_d\Big) \tag{B.8}$$

The first order conditions (FOCs) for the planner's problem thus reads

$$\{C^i_d\} : \frac{\omega_d \alpha^i_d}{C^i_d} = \lambda^i_d \tag{B.9}$$

$$\{m^{ij}_o\} : -\lambda^j_o + \mu^i_o T^i_o \gamma^{ij}_o [m^{ij}_o]^{-1} [l^i_o]^{\beta^i_o} \prod_j [m^{ij}_o]^{\gamma^{ij}_o} = 0 \tag{B.10}$$

$$\{Q^i_d\} : \lambda^i_d - v^i_d = 0 \tag{B.11}$$

$$\{\tilde{q}^i_{od}\} : v^i_d \{[\tilde{q}^i_{od}]^{\frac{\sigma-1}{\sigma}-1}(\sum_{\tilde{o}}[\tilde{q}^i_{\tilde{o}d}]^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}-1}\} - \mu^i_o \tilde{\tau}^i_{od} = 0 \tag{B.12}$$

$$\{l^i_o\} : -n_o + \mu^i_o T^i_o \beta^i_o [l^i_o]^{-1} [l^i_o]^{\beta^i_o} \prod_j [m^{ij}_o]^{\gamma^{ij}_o} = 0. \tag{B.13}$$

The resource constraints and technology constraints in the competitive equilibrium and the social planner's problem agree. We now construct the Lagrangian multipliers from the prices and allocations in the competitive equilibrium as below

$$\lambda^i_d = \frac{P^i_d}{Y},\ \mu^i_o = \frac{\hat{c}^i_o}{T^i_o}\frac{1}{Y},\ v^i_d = \frac{P^i_d}{Y},\ n_o = \frac{w_o}{Y}. \tag{B.14}$$

[4] We have combined the "production of intermediate goods" and "resource constraints for intermediates goods" to

$$\sum_d \tau^i_{od}\tilde{q}^i_{od} = T^i_o[l^i_o]^{\beta^i_o}\prod_{j=1}^{S}[m^{ij}_o]^{\gamma^{ij}_o}.$$

We now verity that the FOCs of the planner's problem hold under these multipliers and the Pareto weights

$$\omega_d = \frac{w_d L_d}{Y}.$$

Plug ω_d and λ_d^i into (B.9) we arrive at (B.9)⇔(B.1). Plug λ_d^i and μ_o^i into (B.10) we arrive at (B.10)⇔(B.2). (B.11) is implied by the definitions of λ_d^i and v_d^i. Plug μ_o^i and v_d^i into (B.12) we have

$$\begin{aligned}
(B.12) \Leftrightarrow \frac{P_d^i}{Y}\Big(\sum_{\tilde{o}}[\tilde{q}_{\tilde{o}d}^i]^{\frac{\sigma-1}{\sigma}}\Big)^{\frac{\sigma}{\sigma-1}} \frac{[\tilde{q}_{od}^i]^{\frac{\sigma-1}{\sigma}}}{\sum_{\tilde{o}}[\tilde{q}_{\tilde{o}d}^i]^{\frac{\sigma-1}{\sigma}}} &= \frac{\hat{c}_o^i}{T_o^i}\frac{1}{Y}\tilde{\tau}_{od}^i\tilde{q}_{od}^i \\
\Leftrightarrow \frac{[\tilde{q}_{od}^i]^{\frac{\sigma-1}{\sigma}}}{\sum_{\tilde{o}}[\tilde{q}_{\tilde{o}d}^i]^{\frac{\sigma-1}{\sigma}}} &= \frac{p_{od}^i\tilde{q}_{od}^i}{P_d^i Q_d^i} \\
\Leftrightarrow \frac{[\tilde{q}_{od}^i]^{\frac{\sigma-1}{\sigma}}}{\sum_{\tilde{o}}[\tilde{q}_{\tilde{o}d}^i]^{\frac{\sigma-1}{\sigma}}} &= \frac{p_{od}^i\tilde{q}_{od}^i}{\sum_{\tilde{o}} p_{\tilde{o}d}^i\tilde{q}_{\tilde{o}d}^i} \Leftrightarrow (B.4).
\end{aligned}$$

Finally, plug n_o and μ_o^i into (B.13) we arrive at (B.13)⇔(B.3). We have thus verified that the competitive equilibrium can be replicated by the solution to the social planner's problem with the choice of Pareto weights ω_d. □

Based on Lemma B.2 we prove a version of the Hulten's theorem (Hulten, 1978) to associate the marginal gains in welfare after a reduction in trade cost to the observed trade flows under the corresponding competitive equilibrium.

Lemma B.3. *With the social planner's welfare function defined in (B.7), we have*

$$\frac{dW}{d\log\tilde{\tau}_{od}^i} = -\frac{X_{od}^i}{Y},$$

where $Y = \sum_d w_d L_d$ is the total income and X_{od}^i is the value of trade flows from o to d in sector i under the corresponding competitive equilibrium.

Proof. The envelope theorem implies that at the solution to the social planner's problem

$$\frac{dW}{d\tilde{\tau}_{od}^i} = \frac{\partial\mathcal{L}}{\partial\tilde{\tau}_{od}^i},$$

where $\mathcal{L}$ is the Lagrangian defined in (B.8). Therefore, we have

$$\frac{dW}{d\tilde{\tau}_{od}^i} = -\mu_o^i q_{od}^i,$$

where μ_o^i is the Lagrangian multiplier and is associated with equilibrium objects through (B.14), restated here

$$\mu_o^i = \frac{\hat{c}_o^i}{T_o^i}\frac{1}{Y} = \frac{p_{od}^i\tilde{\tau}_{od}^i}{Y}.$$

Therefore

$$\frac{dW}{d\tilde{\tau}_{od}^i} = -\frac{p_{od}^i q_{od}^i \tilde{\tau}_{od}^i}{Y}$$
$$\Rightarrow \frac{dW}{d\log\tilde{\tau}_{od}^i} = -\frac{X_{od}^i}{Y}.$$

□

Proof of Proposition 2.

Proof. Combining Lemma B.2, Lemma B.3, and Proposition 1, we have proved Proposition 2. □

C Quantification

C.1 Numerical Implementation

Solve the competitive equilibria. We describe the design of the algorithm that makes it possible to load the most intensive part of the computation to GPUs. This enables us to solve equilibria robustly and efficiently, despite the size of the problem (our benchmark model has 323 regions and 25 sectors).[5] The large size of the problem also renders a well-known approach to solve/calibrate this type of model—Mathematical Programming with Equilibrium Constraint (Su and Judd, 2012)—less effective as the Jacobian matrix is a dense matrix with $(323 \times 25)^2$ entries. Our algorithm falls back to a fixed point algorithm described below.

With E_d^i being the total expenditure on intermediate goods in sector i of region d, the minimal system of equations that can be used to solve the equilibrium is[6]

$$E_o^j = \alpha^j w_o L_o + \sum_i \gamma_o^{ij} \sum_d \pi_{od}^i E_d^i$$
$$w_o L_o = \sum_i \beta_o^i [\sum_d \pi_{od}^i E_d^i]$$
$$P_d^i = \Big(\sum_o [p_{od}^i]^{1-\sigma}\Big)^{\frac{1}{1-\sigma}}, \tag{C.1}$$

for unknowns (E_d^i, w_o, P_d^i), where p_{od}^i and π_{od}^i can be viewed as intermediate variables and can be evaluated according to

$$p_{od}^i = [\kappa_d^i w_d^{\beta_d^i} \prod_{j=1}^{S} [P_d^j]^{\gamma_d^{ij}} \tilde{\tau}_{od}^i] / T_o^i$$
$$\pi_{od}^i = \frac{[p_{od}^i]^{1-\sigma}}{[P_d^i]^{1-\sigma}}. \tag{C.2}$$

[5] For example, to estimate the model with indirect inference, we need to solve the equilibria numerous times. And because of the sequential nature of many global optimization routines, paralleling this step is not straightforward, so speed is important.

[6] We describe the algorithm setting the exogenous deficits to zero. The model with exogenous deficits can be solved similarly.

We design a nested fixed point algorithm according to the strength of the hardware. A key observation is that given π^i_{od}, the first two equations of (C.1) give a (dense) system of linear equations for E^i_d and $w_d L_d$, for which GPUs are designed to solve efficiently. Based on this observation we design the nested fixed-point algorithm below:

Algorithm 1 Nested fixed-point algorithm for solving the competitive equilibrium using GPUs

1 Guess $(w_{d,Old}, P^i_{d,Old})$
2 Set flag_converged to **false**
while flag_converged is false **do**
  3 Construct π^i_{od} according to (C.2) based on $(w_{d,Old}, P^i_{d,Old})$
  4 Solve the system of linear equations for E^i_d and $w_d L_d$ (with GPUs)
  5 Construct p^i_{od}, P^i_d according to (C.2) and (C.1)
  6 Set flag_converged to **true** if distance between (w_d, P^i_d) and $(w_{d,Old}, P^i_{d,Old})$ is small enough
  7 Update $(w_{d,Old}, P^i_{d,Old})$ according to (w_d, P^i_d)
end while

The step of solving the system of linear equations (line 4 in the algorithm) takes more than 90% of the computation time in our benchmark model. Starting from an initial guess with uniform entries in (w_d, P^i_d), the benchmark equilibrium can be solved (under the convergence criterion of $1e-6$ in log difference) within a minute with a GTX1080Ti video card, compared to around 10 minutes with 2*Intel Xeon CPU E5-2650 v4.

Calibrate city-sector productivities T^i_d. The indirect inference estimation proceeds in a nested manner. In the inner loop, given other model parameters, we calibrate T^i_d for tradable sectors i such that the sectoral sales ratios between each city and the RoW in the model agree with those in the data. To do this, we treat sales ratios as observables, and solve T^i_d to generate the observable sales ratios while respecting the equilibrium conditions. Specifically, the minimal system of equations for calibrating T^i_d to match sales ratios M^i_d[7] while respecting the equilibrium conditions are

$$\begin{aligned} M^j_o &= \sum_d \frac{[\hat{c}^j_o / T^j_o \cdot \tilde{\tau}^j_{od}]^{1-\sigma}}{\sum_{\tilde{o}}[\hat{c}^j_{\tilde{o}} / T^j_{\tilde{o}} \cdot \tilde{\tau}^j_{\tilde{o}d}]^{1-\sigma}} \left(\alpha^j_d I_d + \sum_i \gamma^{ij}_d M^i_d\right) \text{ for } j \text{ a tradable sector} \\ w_d L_d &= \sum_i \beta^i_d M^i_d \\ P^i_d &= \left(\sum_o [p^i_{od}]^{1-\sigma}\right)^{\frac{1}{1-\sigma}}, \end{aligned} \tag{C.3}$$

for unknowns $\left((T^i_d)_{i \text{ tradable}}, P^i_d, w_d\right)$[8], where $I_d, \hat{c}^i_d, p^i_{od}$ are intermediate variables and evaluated accord-

[7] M^i_d in the model is the total sales of intermediate goods from sector i of region d and is linked to E^i_d defined before through $M^i_o = \sum_d E^i_d \pi^i_{od}$.

[8] Notice the system of Equation (C.3) is homogeneous of degree one in T^i_d for any given i. That is, fixing i, scaling up T^i_d by the same factor scales nominal price and wage proportionally but does not affect real allocations. Therefore we normalize $T^i_d = 1$ for a chosen region d for all i.

ing to

$$I_d = \sum_i \beta_d^i w_d L_d + D_d$$
$$\hat{c}_d^i = \kappa_d^i w_d^{\beta_d^i} \prod_{j=1}^{S} [P_d^j]^{\gamma_d^{ij}}$$
$$p_{od}^i = [\hat{c}_o^i \tilde{\tau}_{od}^i] / T_o^i,$$

with D_d being the exogenous trade deficits which are necessary to match the aggregate import and export shares. We use a similar iterative nested fixed point algorithm to Algorithm 1 to implement this.

Calibrate remaining model parameters. With the inner loop inverting T_d^i to match M_d^i exactly, in the outer loop we search over other parameters to target the rest of the moments. These parameters include the sectoral international export and import cost τ_{RoW}^i and $\tau_{RoW}^{i'}$, routing elasticity θ, trade cost level h_0, alternative mode cost $\bar{\kappa}$, and the cost- weight-to-value elasticity μ. Since the number of parameters is equal to the number of moments, calibrating these parameters is to solve the system of equations such that the model moments are equal to their data counterparts listed in Table 6. We solve the system of equations using an iterative procedure based on a line search method. The equations are solved such that the maximum distance between the data moments and the model moments is less than 1%, and the maximum difference in the inner loop is smaller than $1e-5$.

C.2 Sensitivity Analyses

Table C.1: Sensitivity Analyses

Change in	(1) High Heterogeneity in Sectoral Trans. Cost	(2) High Substitution across Trans. Modes	(3) External Economy of Scale	(4) Expressway Expansion with Free Migration
Aggregate welfare	0.068	0.054	0.054	0.081
Log(Domestic trade / GDP)	0.085	0.099	0.165	0.278
Log(Exports / GDP)	0.165	0.164	0.208	0.277
Std Log(real wage) across regions	-0.027	-0.026	-0.022	-

Note: The counterfactual experiments are calculated by changing the expressway network from 1999 to 2010, alternative models in (1)-(3) are recalibrated to match the same targets as in Table 6.

We conduct a number of exercises to assess the sensitivity of the baseline results to alternative assumptions. We focus on four scenarios. The first is on the sector heterogeneity of transport costs. Instead of 0.3 in the baseline calibration, we now set μ to 1, which corresponds to a linear relationship of iceberg cost on weight-to-value ratio. Our second robustness allows for industry-level agglomeration. Specifically, we set $T_d^i = \overline{T}_d^i [l_d^i]^\chi$, with $\chi = 0.13$, around the median estimate of Bartelme et al. (2018). This assumption implies an external increasing return to scale to specialization. The third robustness check increases the elasticity of substitution between road transportation and the outside mode, θ_M, from the benchmark value 2.5 to 14.2, an estimate by Allen and Arkolakis (2014). Finally, given the recent push for an overhaul of the Hukou system in China, we assume there is a reform that makes labor perfectly mobile. We then evaluate the gains from the expressway projects after the reform.

The first column of Table C.1 shows that as expected, when sector heterogeneity in transport costs is more important, the inferred welfare gains are larger. When the elasticity of substitution between transport modes increases, the inferred welfare gains are slightly smaller. This is because after expressways are removed, traders can switch to the alternative mode more easily and incur less losses. Adding external economies of scale at the industry level has very impacts on the aggregate welfare gains but affects domestic and international trade significantly. Finally, the welfare gains increases significantly with free labor mobility.

References

Allen, Treb and Costas Arkolakis, "Trade and the Topography of the Spatial Economy," *The Quarterly Journal of Economics*, 2014, *129* (3), 1085–1140.

_ and _ , "The welfare effects of transportation infrastructure improvements," *NBER Working Paper No. 25487*, 2019.

Bartelme, Dominick, Arnaud Costinot, Dave Donaldson, and Andres Rodriguez-Clare, "Economies of Scale and Industrial Policy: A View from Trade," *NBER Working Paper*, 2018.

Faber, Benjamin, "Trade integration, market size, and industrialization: evidence from China's National Trunk Highway System," *Review of Economic Studies*, 2014, *81* (3), 1046–1070.

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

Su, Che-Lin and Kenneth L. Judd, "Constrained optimization approaches to estimation of structural models," *Econometrica*, 2012, *80* (5), 2213–2230.