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

Jingting Fan Yi Lu Wenlan Luo

This version: December, 2019

Abstract

The choice of the port through which an inland exporter ships goods to foreign countries depends on the trade cost between the firm and the port and reacts to additions to the domestic transport infrastructure network. Building on this insight we develop a new method of estimating domestic trade cost that uses increasingly accessible customs data. We apply our method and a spatial equilibrium model with three ingredients—sector regional specialization, sector heterogeneity in trade cost, and intermediate inputs—to study the aggregate effect of the expressway network expansion in China between 1999 and 2010. Counterfactual experiments show that expressway construction brings 5% welfare gains and can account for about a quarter of export growth during this period. The net return to investment in the projects is around 170% and tends to be higher for lines connecting the north to the south. Each of the three ingredients is important and their omission can turn the return into negative. Our analysis also produces some intermediate outputs of independent interest: for example, a time-varying IV for city-sector export.

1 Introduction

In 2016, the 47 member countries of the International Transport Forum—including OECD countries and China, among others—report 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 increases steadily from 2% in 2000 to 5 % in recent years, accounting for more than half of the total investment among the 47 countries. The sheer size of the investment in China and elsewhere has motivated considerable research measuring the benefits from 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

*Fan: Pennsylvania State University (jxf524@psu.edu); Lu and Luo, Tsinghua University (luowenlan@sem.tsinghua.edu). For helpful comments we thank Treb Allen, David Baqaee, Lorenzo Caliendo, Kerem Cosar, Fernando Parro and Nathaniel Young. We also thank seminar participants at 2018 Nankai University International Economics Workshop, 2019 Hong Kong University Globalization and Firm Dynamics Workshop, Fudan University for helpful comments.

tools from international trade and spatial economics, a growing strand of literature is developing computational models to evaluate transport infrastructure through quantitative experiments.

A key input into such quantitative exercises is the mapping from 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 the price data. The idea is that, under maintained assumptions on cost pass-through, the differences in price across locations of the same good can be used to recover trade cost (e.g., Donaldson, 2018; Atkin and Donaldson, 2015; Asturias et al., 2018).[2]

The data requirement 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, existing studies using the U.S. data or similar data from other countries rely mostly on cross-sectional variations in estimating 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 two contributions. First, we propose a methodology to estimate domestic trade costs using information contained in typical customs data. We estimate a routing model structurally for key parameters that determines the response of shipment to transport infrastructure, exploiting the *over-time* variations steaming 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 spatial equilibrium model with regional specialization, input-output linkages, and sector heterogeneity in trade costs and use it to evaluate the effects of expressway network expansion in China between 1999 and 2010. Our main findings are that these expressway projects generate large positive net returns and collectively account for around 20% of export growth in China over this period. We further show that welfare evaluation using simpler models or an alternative approach focusing on the first order effect can lead to incorrect assessments.

Our empirical design takes advantage of the increasingly available customs data. As those of many other countries, the Chinese customs data contain the city of exporters and the port from which they ship to foreign customers. Fractions of a city's export through different ports reflect, among potential confounding factors, the relative 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 less trade costs than other routes. A direction 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 trade cozst consists of

[1]This elasticity governs how trade cost 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 rare exception, Donaldson (2018) uses both price and shipment information to identify trade elasticity and the distance elasticity of trade cost, exploiting overtime variations from construction of railroad.

(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 road network depicted is from 2007 and includes 'national road' (or 'general highway') and 'provincial road'.

costs along the domestic and the international segments of the route. If the two components are negatively correlated (which they should if exporters choose the route to minimize the *total* cost), attributing choice probability differences entirely to the domestic transport network exaggerates its importance.

We address both concerns by exploiting changes in bilateral trade costs resulting from the impressive 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, shown in Figure 1b.[3] We estimate the relationship between the effective distance between an inland city and a port on the transport network, and the fraction of export of the city shipped through the port, controlling for city-port, city-time, and port-time fixed effects. We find that each additional 100 km effective distance reduces the probability of exporting from a port by 15%. Not controlling for city-port fixed effects doubles this estimate.

A remaining concern is that the expressway network *expansion* might be endogenous to shipment between pairs of cities. Building on the insight of Banerjee et al. (2012) and Faber (2014) that the expressways were planed to connect the major cities, we use the distance along the shortest path generated from a hypothetical network that minimizes the total network length while still connects the major cities, as an IV for the distance on the actual network. The use of this IV, together with the restriction of sample to non-major cities, addresses the route endogeneity

[3]'Expressway', or 'high-grade highway', refers to paved roads that are divided, fully enclosed, and not subject 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.

concern.

We embed the empirical design in a spatial equilibrium model (Eaton and Kortum, 2002 and Caliendo and Parro, 2015), consisting of Chinese prefecture cities and the rest of the world (RoW), and enriched to include a routing block that maps road networks into trade costs. The routing block extends Allen and Arkolakis (2019) to accommodate the following: first, we allow flexible combination of edges from two co-existing networks, regular road and expressway, in forming a route; second, we incorporate an 'outside mode' of transportation as an imperfect substitute to road transport; third and most importantly, we allow the trade costs to be different across sectors and depend on the average weight-to-value ratio of a sector. This is feasible because the customs data provide unit price for a wide range of narrowly defined products, which enables us to use price variations to pin down the importance of sector heterogeneity in transport cost. Our estimates imply a 25% trade cost savings on expressway relative to regular road and an elasticity of 0.3 for weight-to-value ratio in trade costs. We parameterize the rest of the model to match the data on sector production, international trade, which determines regional productivity, and average shipment distance, which pins down the level of domestic trade costs.

Armed with the parameterized model, we quantify the aggregate impacts of new expressway built in China between 1999 and 2010. We find that the aggregate welfare gains from the network expansion is around 5%. The sum of discounted gains far exceeds the total investment into the projects (around 10% of 2010 GDP) and implies about 170% net return. By reducing domestic frictions, the network increases inter-regional trade share in GDP by 11% and export share in GDP by 16%. The latter accounts for around 20% 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. Expressway lines that connect the north and south of the country tend to generate higher returns, whereas lines that connects the hinterland to big ports like Shanghai and Fuzhou has the biggest effect on export.

We show that 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 underestimate gains because model without these two channels predict lower *value* of inter-regional trade, which to the first order determine the size of the gains from trade cost reduction.[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 a higher fraction of trade along the routes that received more expressway investment, it predicts larger gains.[5] When all three ingredients are omitted, the model infers welfare gains that

[4]From the international trade literature (Costinot and Rodríguez-Clare, 2014; Baqaee and Farhi, 2019a), it is well known that overlooking input-output linkages underestimates the gains from trade cost because it infers lower trade over value added ratio. The role of sector heterogeneity in trade costs appear 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 not as wrong.

are about one tenth of the full model and turns the net return into negative. This stark difference highlights the importance of using the full model and sectoral data for assessing values of large transport infrastructure projects.

As is well known in both the broad macroeconomic setting (Hulten, 1978) and in evaluation of transport infrastructure projects (Small, 2012; Allen and Arkolakis, 2019), in efficient models, if the *value* of goods being shipped on *each* segment of an expressway is known, then estimating the first-order aggregate welfare gains from that expressway does not require solving for the full equilibrium. In the domestic transport setting, where the *value* of shipments passing through a road is generally unavailable ex-ante and difficult to measure ex-post, using the full model is important precisely because it uses all other information available in inferring the value. In the last section of the paper, we show that even if the value of shipment is known, in our setting, where an addition segment of expressway represent large (around 25%) change in route cost, the quality of the first-order approximation is poor. For the welfare effect of an individual segment, the first order approximation underestimates the true losses from a removal of the expressway, because it fails to take into account that drivers can re-optimize and switch to another route and that firms can also change their trade behavior in response to the increase in trade cost. Focusing on the 200 busiest expressway segment in China, the first order effect overestimate the losses from removal of the expressway by 30% on average, most of which is due to driver re-routing. We propose a second-order correction term, which can reduces the measurement errors by two-thirds, but this correction term requires additional knowledge on the patterns of routing.

When looking beyond local segments and to evaluate large projects that build multiple segments at once, in addition to the local approximation error, the first-order approximation also misses interaction between different segments, which could be either complements or substitutes depending on their positions in the network. We analytically characterize the interaction terms and show that in our setting, the segments built during this period tend to be complements, and the difference between first order effect and the aggregate effects are quite large. Our characterization provides a way to evaluate large projects without having to solve for the counterfactual equilibrium.

This paper contributes to studies on the effects of large transport infrastructure projects. Research in this literature falls into two broad categories. The first is to parameterize a quantitative model and simulate the model for evaluation. The literature has studied the impacts of roads (Allen and Arkolakis, 2019; Asturias et al., 2018; Morten and Oliveira, 2018; Van Leemput, 2016; Alder and Kondo, 2019; Fajgelbaum and Schaal, 2017; Cosar et al., 2019), railroads (Fajgelbaum and Redding, 2014; Donaldson, 2018; Nagy et al., 2016; Xu, 2018), and urban transit (Severen, 2018; Tsivanidis, 2018). The second aims to empirically estimate the effect of infrastructure on regional income/growth, using either heuristic or theory-based measures of exposures (see, e.g., Banerjee et al., 2012; Faber, 2014; Storeygard, 2016; Baum-Snow and Kahn, 2000; Baum-Snow et al., 2016; Donaldson and Hornbeck, 2016; Alder, 2016). In addition to carefully evaluating the impacts of the massive investment in expressway in China (around 600 billion USD), which is important in its own right, our analysis draws general lessons. Our methodology of estimating domestic trade costs can be used in other countries, where domestic shipment or bar-code level price data are unavailable; the message on the importance of regional specialization and sector heterogeneity in trade cost likely applies to other settings as well. Finally, we characterize and demonstrate the importance of higher-order effects for evaluating local and large projects, contributing to a growing macroeconomic literature that emphasizes non-linearity (see, e.g., Baqaee and Farhi, 2019b).

At the core of our empirical analysis is the idea that export route contains information on domestic transport infrastructure. We are not the first to recognize this. Limao and Venables (2001) shows the importance of domestic infrastructure on export in a cross-country setting, whereas Coşar and Demir (2016) focuses on micro data from Turkey and shows that regions with a higher stock of high-quality roads export more. Instead, we combine the data with equilibrium model with routing and estimate the model structurally to infer the deep parameters governing transport costs. In our earlier work, Fan (2019) uses the gradient of city-level export with respect to the city's distance to the nearest port as a subset of the moments identifying domestic trade costs. The difference of this article is that it uses the *route* of export, as opposed to exporting itself for identification, and that it explicitly models and quantify the effects of transport infrastructure.

More broadly, this paper adds to the rapidly growing quantitative spatial economics literature (see Redding and Rossi-Hansberg, 2017 for a recent review), particularly those focusing on China (Fan, 2019; Tombe and Zhu, 2019; Ma and Tang, 2019; Zi, 2016). The costs of moving goods across space is central to the predictions of these studies. Most of the current research on China uses either railway shipments, which account for only 10% of the shipment and are available only at the provincial-pair level, a level too crude for studies on transport infrastructure, or rely on the regional input-output table, often imputed from the railway shipment data (see Zhang and Qi, 2012). Using a new and more micro data source, our analysis generates predictions for domestic and international trade costs for 1999 and 2010, which can serve as inputs into future work in this area. We also 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 assumption, the model-simulated export can serve as a *time-varying* instrument for export at city-sector level that exploits changes in access to foreign market from transport infrastructure change, which complements existing identification strategies in estimating the effects of export.

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 through C.2 perform counterfactual experiments in the baseline model and compare the results to alternative models and approaches. Section 8 concludes.

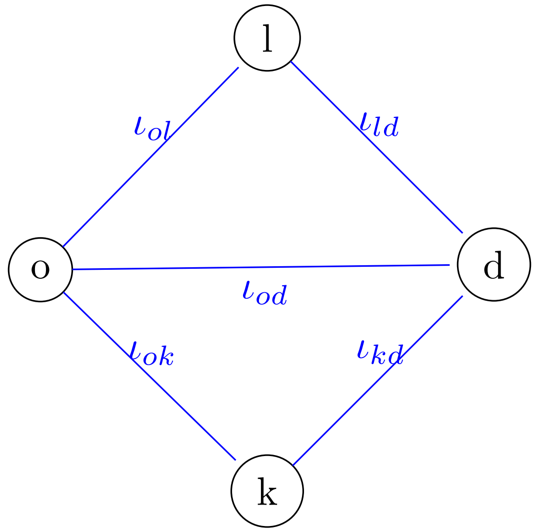


Figure 2: Routing on a Network: Four-Region Example
Note: This diagram illustrates the routing problem in a four-region example. Except for nodes l and k, all other pairs of nodes are directly connected by a road segment.

2 Route Choice on the Transport Network

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

2.1 From Route Cost to Trade Cost on a Single Network

We describe the machinery of the model using a four-region example, illustrated in Figure 2. Each node (o, l, k, or d) 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: $\iota_{od} >= 1$ & $\iota_{od} = \iota_{do}, \forall o \neq d$. 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 the path is the product of costs of the segments it is made of. For example, $o \to k \to d$ forms a path from o to d; the cost along this path is $\iota_{ok} \cdot \iota_{kd}$.

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 $o \to l \to d$. Following Allen and Arkolakis (2019), we allow drivers to derive idiosyncratic dis-utility ν from each potential path, drawn from a Frechet distribution with dispersion parameter θ and location parameter 1. 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 $o \to l \to d$ is $\iota_{ol}\iota_{ld}\nu$.

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

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

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

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

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

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 infinitely high transport cost.[6] Use $\mathbb{L}$ to denote this matrix and $[\mathbb{L}_{(o,d)}]$ to denote the (o,d) element of $\mathbb{L}$. 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 the sum of costs across all feasible paths with two edges. We can 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:

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

Therefore, if the driver is allowed to choose among all paths with three or fewer edges, the expected travel cost between o and d, before the realization 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 $d \notin o$, so the average transport costs across all possible

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

paths is:

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

2.2 Combining Two Transport Networks

Our empirical application focuses on the expressway network expansion, which is an addition to the existing regular road network. We can see from Figure 1 that after the completion of the project, many adjacent cities are connected by expressway and regular road at the same time. In these cases, both types of roads could be used.[7] Furthermore, drivers can combine the two networks to their own taste to form a route. For example, one person might prefer to go from B to C on regular road, and then from C to D on expressway, while another person might choose expressway for both segments. We extend the probabilistic formulation of the transport problem in Allen and Arkolakis (2019) to tractably accommodate these situations.

Let $\mathbb{H}$ and $\mathbb{L}$ denote the road matrix for expressway (H for High-speed) and regular road (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: $\mathbb{A} = \mathbb{H} + \mathbb{L}$. As before, drivers choose among all possible paths subject to a path-specific idiosyncratic taste shock. But rather than being confined to expressway or regular road, 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 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 costs across possible realization 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 expressway and the second being regular road; analogously, $[(\mathbb{L}\cdot\mathbb{H})_{(o,d)}]$ is the sum across all paths with the first segment being regular road and the second being 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 results holds for when drivers are allow to choose any possible combinations of regular road and expressway segments with arbitrarily many edges. The expected trade costs

[7] Indeed, whereas expressways usually have a higher speed limit, they also charge more fees and are more inflexible for entry and exit, which might make some drivers prefer regular roads.

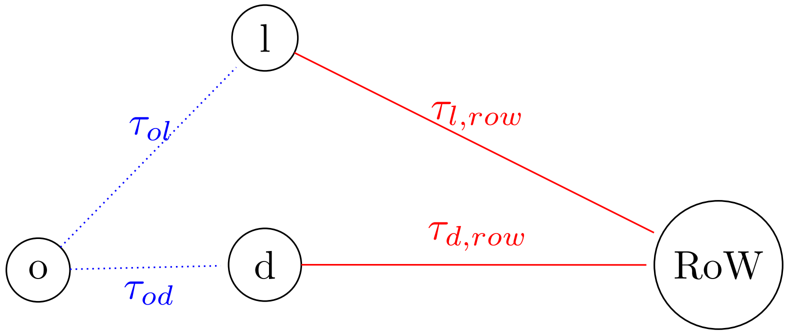


Figure 3: Port Choice of Exporters
Note: The diagram illustrates the choice of port through which to ship to the RoW.

across *all* possible paths is

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

2.3 From Domestic Trade Cost to Port Choice of Exporters

A seller shipping to another location randomly meet with a driver and pays the expected transport cost before the idiosyncratic dis-utility shocks realize. This expected cost, given by Equation (4), is therefore 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 truck-load 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 ν_F, drawn from a Frechet distribution. This shock enters trade cost multiplicatively, so the international shipment cost from l to the RoW, for example, is $\tau_{l,row} \cdot \nu_F(l)$. Given that the routing pattern in international shipment might be different from domestic, we assume that the dispersion parameter of ν_F is $\theta_F \neq \theta$.[8] The seller chooses among $\min\{\tau_{ol}\tau_{l,RoW} \cdot \nu_F(l),\ \tau_{od}\tau_{d,RoW} \cdot \nu_F(d)\}$. Suppose port d is chosen, then the seller randomly meets with a truck driver, who will then find the minimum-cost route from o to d, given his own realization of route-specific cost draw, and charge the seller for the cost.

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

The expected export cost to the RoW is

$$\tau_{o,row} = \Gamma(\frac{\theta_F - 1}{\theta_F})(\sum_{\text{All ports k}} \tau_{ok}^{-\theta_F} \cdot \tau_{k,RoW}^{-\theta_F})^{-\frac{1}{\theta_F}}. \tag{5}$$

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

$$\pi_{(o,row),d} = \frac{\tau_{od}^{-\theta_F} \cdot \tau_{d,RoW}^{-\theta_F}}{\sum_{\text{All ports k}} \tau_{ok}^{-\theta_F} \cdot \tau_{k,RoW}^{-\theta_F}}. \tag{6}$$

Equation (6) illustrates how export data help identify domestic trade costs. All else equal, if port k is closer or better connected to city o through domestic infrastructure (lower τ_{ok}), exporters in city o will be more likely to ship via port k.

2.4 Parameterizing Road Network Matrices

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

$$\iota_{kl}^{x} = \exp(\kappa^x \cdot dist_{kl}), \tag{7}$$

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

$$[\mathbb{A}_{(k,l)}] = [\exp(-\theta\kappa^H \cdot dist_{kl}) + \exp(-\theta\kappa^L \cdot dist_{kl})]. \tag{8}$$

The structural parameters of the routing model is κ^H, κ^L, and θ. To bring out the connection 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, as in Donaldson (2018). Assuming $\kappa^H < \kappa^L$, then between two adjacent cities, when both types of roads coexist, expressway will be chosen with probability one and the limit route cost is $\lim_{\theta\to\infty}[\mathbb{A}_{(k,l)}]^{-1/\theta} = \exp(\kappa^H \cdot dist_{kl})$. The log transport cost from o to d along a path of length $p : o \to m_1 \to m_1 \to ... \to m_{p-1} \to d$ is:

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

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, respectively). 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 simply 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 (6) to obtain:

$$\log(\pi_{(o,RoW),d}) = \beta_o + \beta_d - \theta_F(\kappa^H dist^H_{o\to d} + \kappa^L dist^L_{o\to d}). \tag{9}$$

In Equation (9), β_o and β_d capture characteristics of the origin city o and port d, respectively. In the data, we observe the fraction of export shipments through each port d, $\pi_{(d,RoW),d}$. 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 extend the routing framework to allow for sector-specific transport costs and alternative modes of transportation for the parameterization of the full model.

3 A First Look at the Data

This section takes a first look at the data and shows that the port choice of exporters responded to changes in the domestic transport network.

3.1 Data and Measurements

The empirical analysis focuses on the change between 1999 and 2010, a decade that witnessed great expansion of the expressway network. As shown in Figure 1a, by 2010, most of the populous center and eastern China has direct access to the expressway network. The massive construction provides exogenous variations we will use to estimate trade costs on both regular road and expressway.

Export routing. We measure the port choice of exporters using the monthly transaction-level Chinese customs data. For each transaction, we observe the address of the exporter, the value and weight (when the unit of measurement 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 the city of an exporter as the origin city and the city of the customs office as the port.[9] The ideal measure for shipment is the number of trucks/containers shipped through different ports. In the absence of this information, we use the weight of merchandise as a proxy. We construct the shipment from each origin city to the RoW through different Chinese ports at both aggregate and HS2 category level. Given that the variations in expressway network are between 1999 and 2010, we construct a panel with two periods corresponding to the beginning and end of the decade.[10]

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

[10]The beginning of period data are average across 2000 and 2001; the end-of-period data are average between 2009

Transport network. We obtain inter-city expressway map for 1999 and 2010 from Baum-Snow et al. (2016), which digitized transport infrastructure for the entire mainland China from published maps.[11] 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, two or four-lane roads. Because there are virtually no variations in regular roads during this period, we treat the regular road network as time-invariant.[12] The raw data are in the form of the coordinates of a series of points on the road network. We convert each of the three maps into a matrix of cities (nodes) and links (edges). The procedures are similar to those in Fajgelbaum and Schaal (2017) and Allen and Arkolakis (2019), so we give an outline here and describe the details in the appendix. We proceed in three steps.

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 km of the center of the city, which is defined as the population weighted average location of centers of the counties making up 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 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 cities. The result of this step is the matrix representation of each map. Figure 9 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 analysis in the rest of this paper, we will use the following definition: $\mathbb{A}^{1999} = \mathbb{H}^{1999} + \mathbb{L}$ and $\mathbb{A}^{2010} = \mathbb{H}^{2010} + \mathbb{L}$.

Bilateral transport cost. With the combined networks constructed, we measure the lowest-transport cost between city pairs for 1999 and 2010. In this step, we need the relative size of κ^H and κ^R: among multiple paths connecting two cities composed of different compositions of regular road and expressway length, which one is the least costly depends on $\frac{\kappa^H}{\kappa^L}$. We query the driving time between a random set of 2000 city pairs along expressway and regular road

and 2010. We do not have access to the customs data for 1999.

[11]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)

[12]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' appears to include '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 transport rely on the latter two, we choose not to use Baum-Snow et al. (2016) to measure the regular road network.

on the Baidu Map, a Chinese search engine, and compare the average anticipated travel time of the two trips. Among these queries the average speed on regular road is about 55% of that on expressways, so we set $\frac{\kappa^H}{\kappa^L} = 0.5$, which means each kilometer on expressway is equivalent to half kilometer on regular road. Under this assumption, we use the Dijkstra's algorithm to find the shortest path between any city pairs and measure the regular-road equivalent distance along the path. .[13]

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 (9):

$$\ln(q^t_{(o,RoW),d}) = \beta_{od} + \beta^t_o + \beta^t_d + \gamma_1 \text{dist}^t_{od} + \epsilon^t_{od}. \tag{10}$$

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 a set of fixed effects for city-port pair, city-time, and port-time, respectively. dist^t_{od} is the regular road-equivalent distance in kilometer along the shortest path. In some specifications, we will split dist^t_{od} into the highway and regular road distance along the path, which will allow us to separately estimate their cost parameter.

The OLS estimator of specification (10) is subject to the obvious endogeneity concern. Cities closer to each other likely have stronger unobserved ties, which could attract shipment for reasons having nothing to do with domestic infrastructure. Through city-port fixed effects, which control for all time-invariant unobserved heterogeneity across pairs of cities, the identification of our baseline specification comes from over-time change in the distance along the shortest path, 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 effect cannot address the concern that newly built expressway might serve to connect specific *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 with more than 5 million registered residents in 2000 (by Hukou).[14] As discussed in Banerjee et al. (2012), the transport network in China was largely designed to connected the major cities. Once we exclude these, our estimation exploits the changes in access to port for the remaining, smaller cities, which took place simply because these cities happened to be between major cities to be connected.

Second, in addition to excluding major cities, we adopt an IV strategy based on Faber (2014) and use an 'exogenous' hypothetical expressway network as an instrument for the actual net-

[13] 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 notation of prediction error—for example, the deviation in the models' prediction on routing patterns from the data. We pursue a version of this exercise in the full structural estimation.

[14] By this definition, there are in total 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 on the 2010 expressway network.

work. Specifically, using the minimum-spanning tree algorithm, we first generate a minimum-distance expressway network that connects all major cities 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 connecting the same set of major cities. Using $\mathbb{A}^{2010,\text{ hypothetical}}$ in place of the actual network, $\mathbb{A}^{2010}$, when measuring bilateral shortest distance, we thus have a new measure of bilateral distance that is exogenous to specific pair of *non-major* cities. This distance serves as an IV for the distance along the actual network in Equation (10). Figure 4 shows the hypothetical network and actual network.[15] The identifying assumption is that non-major cities experienced an improvement in access to ports (and other cities) only because they were close to the *minimum-distance* hypothetical expressway network that connects the major cities.

3.3 Baseline results

Table 1 reports results from the benchmark specification. The dependent variable in all seven specifications are the log of total shipment (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 of mostly cross-sectional variations. The point estimate suggests that each additional hundred kilometer regular road-equivalent distance reduces the probability a port is chosen for export by 35%.

[15] In 1999 the expressway network is very sparse (see Figure 1), so we use the distance along the 1999 actual network as an IV for itself.

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 choice probability 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 effective shortest distance between city o and port d along the shortest path (Columns 1-4, Columns 6-7); and the length of expressway and regular road 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

As discussed, the cross-sectional variations in routing are likely correlated with on 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, so identification comes from the network expansion.[16] The point estimate shrinks by 70% and is rather precisely estimated. To rule out the possibility that transportation infrastructure improvement is correlated with city- or port-specific growth, in Column 3, we further control for city-time and port-time fixed effects. These additional controls lead to a modest increase in the estimated 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 unobserved time-varying linkages between big cities and ports which drove the expressway construction plan. The point estimate suggests that each additional hundred km regular road-equivalent 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 distance on expressway (weighted by 0.5) and regular road. To investigate the relative costs of these two types of roads, in Column 5, we separate the total effective distance into the two components and estimate their respective costs. We find that the coefficient for regular road is

[16]It is possible that some of the non-transport barriers, such as information friction, also respond to addition to the transport network. Our estimate will pick up this effect, which it should because we would like to take into account this channel into the welfare evaluation. What we would like to exclude through the addition of bilateral fixed effect is the component that does not respond to transport infrastructure.

around -0.15, while the coefficient for highway is around -0.08—the former is close to twice as large as the latter, consistent with our assumption in calculating the shortest path. This relative distance also implies that for two adjacent cities that were already connected by regular road, an addition expressway segment between them can reduce the trade cost significantly.

Columns 6 to 8 show that the general findings are robust when we use the Poisson Pseudo-Maximum Likelihood method. Column 6 controls for city-year and port-year fixed effects and identifies the cross-sectional estimate; Column 7 further control for city-pair fixed effects. The coefficient shrinks by a statistically significant amount of 30%. Column 7 split total effective distance into for regular road and expressway, and find the point estimate to be substantially larger for regular roads. Both results collaborate our finding from the OLS.

3.4 IV and Additional Robustness

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

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 first stage F indicates relevance. The point estimates for both the overall effective distance (Column 1) and the distance on expressway and regular road (Column 2) are similar to that based on the OLS, although the estimates are less precise.

A different, but related 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 the ports, they also become more specialized in export-oriented industries, such as textile, and if for some reason, export in the textile industry is concentrated among the ports that experienced disproportionate increase in expressway connectivity to the hinterland, then the correlation between shipment share and bilateral connectivity will be picked by our regressions. Note that if the expressway network 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 sectoral level shipment for robustness. Column 3 includes city-port-sector, city-time, port-time, and sector-time fixed effect. 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 smaller than in the benchmark specification. Finally, in Column 5, we estimate the transport costs separately for expressway and regular road, and find shipping via expressway is less costly compared to regular roads.

Additional robustness. In Appendix Table 13, we show that PPML and IV specifications using sectoral data generate similar results. While not reported, we perform regressions with total value of shipment as the dependent variable, which is more standard in the international trade literature, but is not the theory-consistent measure when sector heterogeneity in transport

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 robustness analysis of 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 HS2 category level.
Standard errors are clustered at city-port level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

intensities 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 standard aspects and highlight the differences in our setting.

4.1 Preliminaries

There are N regions in the model, denoted by o and d, representing Chinese prefectures cities and the RoW. Sectors are denoted by i or j. Workers are immobile and consume a basket of sector final goods. Final goods are non-tradable and aggregated from tradable intermediate goods produced by different locations.

4.2 Consumers

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

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

where C_d^i is final good in sector i, whose price is 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.

4.3 Industry Final Good Production

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

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

in which q_{od}^i is the quantity of the sector-i intermediate good produced in region o and Q_d^i is the quantity of final .

4.4 Intermediate Good Production and Trade

A representative intermediate good producer in sector i region d convert labor and industry final goods from different sectors into the intermediate goods using the following 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 input from labor and final goods from industry j, respectively; and β_d^i and γ_d^{ij} are their respective shares: $\beta_d^i + \sum_j \gamma_d^{ij} = 1$.

The unit cost of sector-i intermediate good from region d is:

$$c_d^i = \frac{\kappa_d^i w_d^{\beta_d^i} \prod_{j=1}^{S} [P_d^j]^{\gamma_d^{ij}}}{T_d^i}, \tag{11}$$

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 output to final goods producers at their marginal costs, which include both production cost and an iceberg trade cost, denoted $\tilde{\tau}_{od}^i$. The price of intermedaite goods produced in region o at region d is $p_{od}^i = [\kappa_o^i \tilde{\tau}_{od}^i]$. The price index of final goods in region d sector i is therefore:

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

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

$$X^i_{od} = X^i_d \frac{[p^i_{od}]^{1-\sigma}}{[P^i_d]^{1-\sigma}} = X^i_d \pi^i_{od},$$

with π^i_{od} being the trade shares.

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

4.5 From Road Network to Trade Costs

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

Sector heterogeneity in transport costs. We allow the ad-valorem equivalent trade cost of good from a sector i to depend on '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 that each truck can load h_0 tons, the cost of shipment for this batch of goods on $k \to l$ is simply $\frac{yh_i}{h_0}\iota^x_{kl}$, in which $\iota^x_{kl},\ x \in \{H, L\}$ is defined in Equation (7), and h_0 is a scaler that determines the level of the overall domestic transport cost.

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

$$(\frac{h_i}{h_0})^{\mu}\iota^x_{kl},\ x \in \{H, L\} \tag{12}$$

in which μ determines the extent of sector heterogeneity in transport costs.[17]

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

$$\tau^i_{od} \equiv \lim_{N\to\infty} \tau^i_{od,N} = \Big(\sum_{n=1}^{\infty}(\frac{h_i}{h_0})^{-\mu\theta}[\mathbb{A}^n_{(o,d)}]\Big)^{-\frac{1}{\theta}} = (\frac{h_i}{h_0})^{\mu}[\mathbb{B}_{(o,d)}]^{-\frac{1}{\theta}},\ \text{where } \mathbb{B} = (\mathbb{I} - \mathbb{A})^{-1}. \tag{13}$$

The sector heterogeneity in transport cost on any specific road segment, $k \to l$ translates tractably

[17] In the price quotes from actual shipment companies are usually proportional to the weight of the shipment: see Limao and Venables (2001) for international cargo shipment cost and Redding and Turner (2015) for freight cost within the U.S. The linear functional form thus corresponds to a strict interpretation of domestic trade cost as freight cost. We adopt a more flexible functional form in part to allow transport cost to be interpreted as capturing other trade cost associated with weight but not necessarily perfectly linear in weight.

to the effective trade cost between two domestic locations, o and d.[18]

Alternative mode of transportation. While road transport is the dominant form of domestic shipment in China (road transport account for 78% of overall domestic shipment; 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 theses alternative modes in a simple way. Formally, we assume that in addition to transport via the road network, which incurs an expected transport cost of τ^i_{od}, between any origin city o and destination d, there is also an alternative mode with an expected cost of $\bar{\tau}^i_{od}$, given by

$$\bar{\tau}^i_{od} = (\frac{h_i}{\bar{h}_0})^{\mu} \exp(\bar{\kappa} \cdot \overline{\text{dist}}_{od}),\ o \neq d \tag{14}$$

in which $\overline{\text{dist}}_{od}$ is the great circle distance between o and d. This specification mode differs from road transport costs in two aspects. First, the structural parameters: $\bar{\kappa}$, which governs the distance elasticity, is allowed to differ from κ^H and κ^L. Second, as $\bar{\tau}^i_{od}$ is meant to capture the cost associated with *all* alternative modes to road transportation including air transport, we specify it to be a function of the great circle distance, which is determined by geography, as opposed to effective distance under any particular choice of mode.[19]

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 good to region d first decides whether to ship it via road transportation or the alternative mode, with the *average* iceberg transport costs for the two modes being τ^i_{od} and $\bar{\tau}^i_{od}$, respectively. Each seller draws two i.i.d. Frechet costs shock, denoted $\nu_m,\ m \in \{road, alt\}$, one for each mode, and chooses the mode with the lower effective cost: $\min\{\tau^i_{od}\nu_{\text{road}}, \bar{\tau}^i_{od}\nu_{\text{alt}}\}$.

If the seller chooses the alternative mode, then the good is directly shipped to the destination, incurring a cost of $\bar{\tau}^i_{od} \cdot \nu_{\text{alt}}$ for domestic destination and τ^i_{RoW} for export.[20] If the seller chooses road transport, then the rest of the routing module plays out as described in Section 2. If the final destination is in China, then the seller randomly meet with a truck driver with his or her own idiosyncratic draws and compensate the driver for route cost and the idiosyncratic dis-utility

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

[19] In reality, even after conditioning on the great circle distance between two cities, the cost of shipment on alternative modes clearly can still differ according to the accessibility of direct flights, train, and cargo ships in between. Given the data limitation, we do not directly model these alternatives. Our counterfactual experiments should thus be viewed as keeping these alternative 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 overtime 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 structural estimates will not be biased due to omission of alternative mode. This assumption appears reasonable—while in addition to expressway, the country also invested heavily in the railway system, most of the investment went to high-speed railway, which was for passengers not merchandises.

[20] We assume the international segment of the export is the same across locations and the two modes. The structural estimation controls for port-time fixed effect and thus accommodates differences in international trade costs across cities.

draw. The expected trade cost is τ^i_{od}. If instead the final destination is the RoW, the seller first chooses a potential port j, given the expected domestic transport costs from o to j, and a port-specific idiosyncratic shock ν_{port}, to minimize $\tau^i_{od}\tau^i_{RoW} \cdot \nu_{\text{port}}$, in which ν_{port} is an idiosyncratic draw of match quality between the seller and the port. Once the port choice is made, the seller then again meet randomly with the trucker driver, who decides the route from o to d.

Combining all these decisions, the expected trade cost between an origin o and destination d is the following:

$$\tilde{\tau}^i_{od} = \begin{cases} \Gamma(\dfrac{\theta_m - 1}{\theta_m})[(\bar{\tau}^i_{od})^{-\theta_m} + (\tau^i_{od})^{-\theta_m}]^{-1/\theta_m}, & \text{if } d \neq RoW \\ \Gamma(\dfrac{\theta_m - 1}{\theta_m}) \cdot \tau^i_{RoW} \cdot \left(1 + \Gamma(\dfrac{\theta_F - 1}{\theta_F})^{-\theta_M}[\sum\limits_{\text{All ports k}} (\tau^i_{ok})^{-\theta_F}]^{\frac{\theta_M}{\theta_F}}\right)^{-1/\theta_m}, & \text{if } d = RoW \end{cases} \tag{15}$$

in which τ^i_{od} is given by Equation (13) and $\bar{\tau}^i_{od}$ given by Equation (14).

5 Parameterization

This section parameterize the model. We adopt two-step indirect inference. In the first step, without imposing the equilibrium conditions, we estimate two specifications to recover coefficients that are direct combinations of the routing parameters or otherwise are informative about their values. In the second step, we calibrate the full model to pin down in equilibrium the level of all parameters in the routing block, as well as remaining parameters from the rest of the model.

Among the structural parameters of the routing block, κ^H, κ^L, and $\bar{\kappa}$ govern how fast transport cost increases with distance on different types of transport networks; θ, θ_M, and θ_F characterize the elasticity of substitution across different routes on the road network, between road and other means of transportation, and across ports in exporting; μ governs the sector heterogeneity in transport costs; finally, h_0 and $\bar{h}_0$ govern the overall level of inter-regional transport cost. In the firs step, we estimate two specifications to retrieve two sets of information to determine their value.

5.1 Export Routing Regression

The fist specification is on the response in the route choice of exporters to expressway network expansion and is the structural version of 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 (6).[21] Taking log on both

[21] Equation (6) 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 exported goods on 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

sides of the equation to obtain:

$$\log(\pi^i_{(o,RoW),d}) = \frac{\theta_F}{\theta}\log\left([\tilde{\mathbb{B}}(\{dist_{jl},\forall j,l\},\kappa^E\theta,\kappa^R\theta)_{(o,d)}]\right) \underbrace{+\mu\log(\frac{h_i}{h_0}) - \theta_F\log(\tau^i_{RoW}) - \theta_F\log(\tau^i_{o,RoW})}_{\text{fixed effects}}. \tag{16}$$

On the right hand side of Equation (16), the overall international export costs τ^i_{RoW}, the city-specific access to the international market, $\tau^i_{o,RoW}$, and the 'heaviness' of sector i will be absorbed by fixed effects. The variations we will exploit are changes in the (o,d) element of matrix $\mathbb{B}$ overtime. Recall that $\mathbb{B} = (\mathbb{I} - \mathbb{A})^{-1}$ is entirely determined by $\mathbb{A}$. Because κ^E and κ^R enter $\mathbb{A}$ only multiplicatively with θ, we write$[\mathbb{B}_{(o,d)}]$ as $[\tilde{\mathbb{B}}(\{dist_{jl},\forall j,l\},\kappa^E\theta,\kappa^R\theta)_{(o,d)}]$ to highlight its dependence on $\kappa^E\theta$, $\kappa^R\theta$, as well as the whole network, summarized in $\{dist_{jl},\forall j,l\}$.

We can estimate Equation (16) without solving the full model using nonlinear least square for $\frac{\theta_F}{\theta}$, $\kappa^E\theta$, and $\kappa^R\theta$. In principle, we can do so using cross-sectional variations alone. Our reduced-form analysis shows that such estimates are biased, so we focus on over-time variations and control for city-time, port-time, city-port fixed effects. This full set of fixed effects also ensures that alternative confounding factors such as shift in specialization will not drive our estimate. Formally, with the observed export route choices $\hat{\pi}^s_{(o,RoW),d,t}$ for year $t \in \{99\ 00, 10\ 11\}$ in the data, we choose the structural parameters to minimize the following expression

$$\max_{\frac{\theta_F}{\theta},\kappa^E\theta,\kappa^R\theta,\boldsymbol{f}} \left[\frac{\theta_F}{\theta}\log\left([\tilde{\mathbb{B}}(\{dist_{jl},\forall j,l\},\kappa^E\theta,\kappa^R\theta)_{(o,d)}]\right) + \boldsymbol{f} - \log(\hat{\pi}_{(o,RoW),d,t})\right]^2, \tag{17}$$

where $\boldsymbol{f}$ is the full set of fixed effects. Given the number of fixed effect included, this is a high-dimensional optimization problem which conventional optimization routines cannot handle. Note, however, that only $\kappa^E\theta$ and $\kappa^R\theta$ enters the objective function non-linearly, we can thus recast the original problem into the following nested problem:

$$\max_{\kappa^R\theta,\kappa^E\theta}\ \max_{\frac{\theta_F}{\theta},\boldsymbol{f}} \left[\frac{\theta_F}{\theta}\log\left([\tilde{\mathbb{B}}(\{dist_{jl},\forall j,l\},\kappa^E\theta,\kappa^R\theta)_{(o,d)}]\right) + \boldsymbol{f} - \log(\hat{\pi}_{(o,RoW),d,t})\right]^2.$$

In the inner loop, given values for $\kappa^R\theta$ and $\kappa^E\theta$, we estimate $\boldsymbol{f}$ and $\frac{\theta_F}{\theta}$ using standard linear regressions. In the outer loop we search over the space of $\kappa^R\theta$ and $\kappa^E\theta$ to minimize the mean square error in 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 magnitude. First, the expressway is about a quarter less costly compared to the regular road ($\frac{\kappa^H}{\kappa^L} = 73\%$). This difference

goods are first shipped via the alternative mode (most likely via the railway) 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.

Table 3: Estimates from the Routing Model

Parameter	Estimate
$\kappa^H\theta$	4.44
$\kappa^L\theta$	6.07
$\frac{\theta_F}{\theta}$	0.02

is broadly in accord with measurements based on actual speed alone (see Section 2). This finding is reassuring, especially because here we do not impose that $\kappa^R > \kappa^E$. Second, the elasticity of substitution across routes (θ) is much larger than across ports (θ_F). This appears reasonable, as the former is driven by preference of drivers among routes whereas the latter depends on the idiosyncratic preference of seller across ports, which could be related to unobserved business connections.

As Equation (17) indicates, the route choices alone does not contain enough information to separate κ^R, κ^R, θ, and θ_F. Moreover, the multiplicative nature of $(\frac{h_i}{h_0})^\mu$ implies that μ is not identified from domestic routing patterns alone—in fact, the level of shipping cost, $(\frac{h_i}{h_0})^\mu$, does not affect the relative probability of exporting through different ports from an interior city. We next turn to the price data and discuss how it helps us identify these and additional 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_o^i. Under the assumption of complete pass-through, the average (across all route-specific draws) f.o.b price at port d is given by:

$$\begin{aligned} p^i_{(o,RoW),d} &= p^i_o \cdot \tau^i_{od} && (18)\\ &= p^i_o \cdot (\frac{h_i}{h_0})^\mu \cdot [\tilde{\mathbb{B}}(\{dist_{jl}, \forall j,l\}, \kappa^E\theta, \kappa^R\theta)_{(o,d)}]^{-\frac{1}{\theta}},\ o \neq d. \\ \implies \log(\frac{p^i_{(o,RoW),d}}{p^i_o}) &= \mu \log(h_i) - \frac{1}{\theta}\log\left([\tilde{\mathbb{B}}(\{dist_{jl}, \forall j,l\}, \kappa^E\theta, \kappa^R\theta)_{(o,d)}]\right), \end{aligned}$$

Equation (18) shows that variations in price ratio across sectors with different 'weight-to-value' identify μ; assuming $\kappa^E\theta$ and $\kappa^R\theta$ are known, variations across city pairs with different distances identify θ.[22] Building on this intuition, we estimate the elasticity of price ratio with respect to the h_i, and the semi-elasticity of price ratio with respect to the distance between o and d along the road network. We then target these two estimates along with other empirical moments in the

[22]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.

full calibration to pin down all routing parameters.[23]

Data. We measure price as the unit value of exported goods from the transaction-level 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 port city itself. For the goods produced in these port cities, denoted 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.

The validity of this approach rests on the assumption that goods exported directly from o to the RoW and goods shipped indirectly through a different city d are the same. To make this assumption realistic, 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 production 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 equivalent trade cost.

Defining a product to be a combination of origin city, HS8 category, and destination country addresses a few concerns in interpreting price ratios as trade costs. First, the literature has documented that firms export both higher-quality goods and charge higher markup on these goods for destination countries with higher income (see, e.g., Simonovska, 2015; Fan et al., 2015). Second, recent research has also documented that cities with a more skilled workforce tend to produce better products (Dingel, 2016). Conditioning on the same destination market and origin city avoid these two sources of biases. To alleviate other concerns, our empirical specifications include additional fixed effects that absorb remaining systematic variations in either quality or markup across transactions; 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 for us to use overtime variations. The main set of regressions we will rely on is thus cross-sectional in 2010-2011. In appendix, we use a crude measure of product (HS2 categories) to show that the results hold broadly when exploiting overtime variations from the road network expansion between 1999 and 2010. Finally, because the two moments are estimated off different sources of variations, we will estimate them separately, so more controls can be added in both specifications.

Price-heaviness elasticity. Table 4 reports our estimate on the weight-to-value ratio elasticity, with progressively more demanding fixed effects. The first four columns focus on comparison of the log price differences across HS2-level weight-to-value ratios. The first and second columns control for city, port, and destination country fixed effects and city-port-country fixed effects,

[23] It is possible that the variations in price ratios might be driven by other reasons, such as quality difference in goods, 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 levels of the price ratio—in quantification. The *level*, which is ultimately governed by h_0 will be pined down to match the overall level of domestic shipment. Alternatively, we can also estimate Equation (18) using non-linear least square 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.

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 heaviness, 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 HS category fixed effects, respectively.

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

respectively. Even within a city-port-country cell, some firms might systematically price differently. To account for this possibility, Column 3 control for firm-port-country fixed effects. The point estimate increased somewhat to 0.27 and is precisely estimated.

The set of fixed effects and the narrow definition of a product allows us to control for many plausible alternative explanations. To the extent that the price ratio might still capture variations in quality and markup, as long as they are not systematically correlated with weight-to-value ratio, it will not affect our estimate. Nevertheless, Column 4 focuses only on the HS2 categories that are classified as non-differentiated goods (Rauch, 1999), which supposedly have a smaller scope for either quality differentiation or price discrimination. Reassuringly, the point estimate remains broadly similar, despite that the sample is only a tenth of the baseline sample.

One reasonable concern is that our measure of 'heaviness', the weight-to-value ratio, might capture other characteristics of a sector that correlates systematically with prices. To alleviate this concern, in Columns 5 through 7, we estimate the specification using HS4 category-level measure. This allows us to control for 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 estimate suggests that a one-percent increase in the weight-to-value ratio of a good increases the ad-valorem shipping cost by 0.25%.

To compare this estimate to the literature, using international shipping fees of inbound goods to the U.S., Hummels (2007) finds the elasticity to be around 0.45 for air-freight and 0.4 for ship freight. When it comes to domestic shipment, the literature does not offer much guidance on this elasticity. But domestic freight cost, documented in the literature, is usually denoted linearly in

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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

weight (see Redding and Turner, 2015), which translates into an elasticity of one. Our estimate is thus in the lower range of the literature. 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 robustness.

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 category characteristics in fixed effects. The first two columns use OLS and control for port-HS8 8-destination country and city-HS8-destination country fixed effects, respectively. The former set captures, within a HS8 category, the overall tendency of some ports of destination countries to be involved in 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% as as additional 100 km is added to the regular-road equivalent distance between the city and the port. The second column restrict to non-differentiated varieties for robustness. This restriction significantly reduces the sample size but the point estimate remains similar.

To alleviate the concern about endogeneity of the road network, Columns 3 and 4 replicate Columns 1 and 2 but use cross-sectional IV from the minimums-spanning tree. The point estimates are in the range of 0.05 to 0.06, and statistically indistinguishable from the OLS estimates. 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 can be interpreted as each additional 100 km increases the ad-valorem trade cost by 6%.

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
L_d	Total employment	-	2010 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 (16) and (18)
θ_F	Port choice elasticity	2.5	Estimates of Equations (16) and (18)
κ_H	Expressway route cost	0.055	Estimates of Equations (16) and (18)
κ_L	Regular route cost	0.074	Estimates of Equations (16) and (18)
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 (18)
τ^i_{RoW}, $\tau^{i'}_{RoW}$	Export and import costs	-	Sectoral export and import
T^i_d	Region-sector productivity	-	City-sector sales

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.

Parameters calibrated externally. Panel A of Table 6 describes the parameters and fundamentals of regional economy calibrated externally outside the equilibrium. 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 determine the sector shares in final consumption and intermediate production, $\{\alpha^i\}$ and $\{\gamma^{ij}\}$, and the labor share in production, $\{\beta^i\}$, based on the input-output table of China for 2007. We assign a value of six to the elasticity of substitution across goods from different regions, σ. Finally, θ_M governs the elasticity of substitution between different modes of transport. The transportation literature has estimated this parameter using the Commodity Flow Survey from the U.S. and found it to fall between 1.5 to 3. We assign a value of 2.5 to θ_M for benchmark analysis and will conduct robustness with alternative values.

Parameters determined in equilibrium. The remaining parameters, reported in Panel B of Table 6, are determined jointly and in equilibrium. The transport cost parameter along regular roads and expressways, κ^E and κ^R are pin down together with the dispersion parameter for routing preference θ. In Section 5.1, we estimate Equation (9) and find that $\kappa^H\theta = 4.47$, $\kappa^L\theta = 6.06$, $\frac{\theta_F}{\theta} = 0.02$; 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 κ^H so that the price-distance semi-elasticity estimated using the simulated data from the equilibrium of the model is also 6%. This procedure determines $\theta = 81.21$, $\theta_F = 2.5$, $\kappa^H = 0.055$, $\kappa^L = 0.074$.

Parameter μ determines the variations of transport costs across sectors with different heaviness. Empirically, we estimate the reduced-form elasticity to be in the range of 0.24 to 0.3; we

Figure 5: Model Predicted Shipment Flows
Note: This figure plots the value of shipment between cities, normalized by the GDP of China.

set μ to 0.3. 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 km. With all these determined, the final 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 others. The annual statistics of transportation reports that about 76% of shipment (by weight) is conducted by road transportation. We choose $\bar{\kappa}$ so that in equilibrium, the model generates the same ratio.

We use the trade costs from the port city and the RoW to pin down the international trade, by targeting import and export as a share of GDP.[24] Finally, we use region-sector productivity parameters, $\{T_d^i\}$ to match the sector production shares (sectors are at two-digit level), constructed from the population census and survey of industrial firms in 2010.[25] We assume all sectors in the RoW have the same productivity and calibrate this productivity bo match the share of China in the world GDP. This also serves as a normalization of sector productivity.

Figure 5 plots the value of shipment flows between all pairs of adjacent cities normalized by the GDP of China. These shipments contain trade flows not only between the two cities at the two ends of the road, but also between other cities that passes the segment. Darker colors indicate higher intensities. While different segments are clearly highly heterogeneous in their

[24]Our calibration takes into account exogenous international trade surpluses of China. After 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.

[25]We use the industrial survey for shares within the manufacturing sectors. Between manufacturing and non-manufacturing sectors, we use employment share from the population census.

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

Robust Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

importance, standing out from the map are a few corridors that connect important economic centers of China. The first is the northeast corridor surrounding the Bohai Bay, which links Beijing and Tianjin to centers of heavy industrial sectors such as Shenyang, Changchun, and Dalian. The second is the corridor between Beijing and the southeast coast, an area encompassing the most economically prosperous areas of China, the Yangtze River Delta. Finally, the corridor that connect the northwest to the center of China is also important.

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 Economic Zone. This is in accord with coverage in the popular press that dub 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 location in the transport network, some cities become as a 'hub' that shipments to other places pass by. To validate the model, we can compare the model-inferred city shipments (all shipments passing the node) to its empirical counterpart, sourced from the 2010 yearbook for transportation.[26] Table 7 reports the regression of the log shipment in the data on 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 indicates the model prediction correlates with the data not just because the usual gravity force, which predicts more trade for bigger cities, but also because it captures the traffic passing by. The third column further shows that including provincial fixed effect do not change the estimate. This implies that the prediction power comes from the model of network connections within a city, rather than the rough location of a city.

[26]The data is aggregated by city; the National Bureau of Statistics survey firms in a city and use their reported shipment to produce this statistic.

Table 8: Predicting Export Growth

	(1)	(2)	(3)
Log(export), model	0.106***	0.109***	0.104***
	(0.017)	(0.015)	(0.016)
Fixed Effects	*t*	*oi*, *it*	*oi*, *it*
Exclude major cities	no	no	yes
Observations	8287	7544	5820
R^2	0.129	0.885	0.870
F-statistic	37.401	49.668	42.973

Standard errors (clustered by city) in parenthesis.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Expressway and Export Growth. In the second validation tests, we compare the model-predicted export growth as a result of expressway network expansion to the actual export growth in the data. This is a joint test of two hypothesis: 1) whether the expressway expansion as large as the one seen in China over this decade led to differential growth of export across cities; 2) when fed into the expressway expansion, the prediction of the model match the data. This comparison is 'out-of-sample' because in quantification, we absorb the export through fixed effects and only identify from the routing.

In implementing this exercise, we feed into the actual road network in 1999 and solve for the counterfactual equilibrium holding all other fundamentals of the model at the calibrated level. We then compare the export at city-sector level between the model and the data. Table 8 reports the results. The dependent variable is the log export in the data. The independent variable is the log export in the model. The first column control for time fixed effect to look at cross-sectional predictions. The second column control for sector-time and city-sector fixed effects, so the comparison is on export growth within a city-sector. The point estimate is 0.1 and highly statistically significant. Given the expressways were built to connect major cities, one might be concerned that the export growth in the data are driven not by changing access due to expression expansion. The third column exclude major cities from the sample and the point estimate remains similar.

Importantly, all these regression models have a F statistic that is well above the rule-of-thumb for bounding biases in IV estimates. Under the assumption that the road networks affect city export only through improving the access of a city to ports, the model predictions can serve as an IV for export at city-industry level. An important literature has examined the impacts of exports on the Chinese economy and elsewhere. One commonly used IV in this literature is the tariff before the WTO accession, which is not time varying and subject to potential endogeneity concern. The IV based on our model predictions are time varying, and valid under a different set of assumptions than existing studies. This IV is an independent contribution of this paper.

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 from the baseline equilibrium with the 2010 express network to a counterfactual equilibrium with the 1999 express network.

6 The Aggregate and Regional Impacts of Expressway Expansion

6.1 Benchmark Results

Armed with the parameterized model, we examine the aggregate impacts of the expressway construction. We solve for an equilibrium with the 1999 expressway network and then calculate the percentage change in relevant objects from this counterfactual equilibrium to the baseline economy.

Table 9 reports the result. We use value added-weighted real income across cities to measure the aggregate welfare of China.[27] The expressway expansion increased the aggregate welfare by 5.7%. 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 cost brought about by the expressway network accounted for around 16% of the increase.

The expanding expressway network also had a large impact on the pattern of domestic and international trade. The domestic trade as a share of GDP increased by 11%. Because interior regions ship their goods to the RoW through ports, expressway 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. 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 interior, leading to an decline in the aggregate international trade. It turns out that in our setting, the reductions in domestic trade cost lead to a 16% increase in international trade. In the data, export as a share of GDP increased by 70%, from 18% in 1999 to 32% in 2008, before it plummeted over the great trade collapse. About a quarter of this 70% increase in export intensity could be explained by the expansion in domestic expressway network.

By connecting previously remote areas to the network, the expressway generates distributional effects. The initially less connected regions benefit through the disproportionate increase in access to other markets. This also lead to a decrease in real wage inequality across regions,

[27] As shown in the appendix, when international trade is shut down, this measure corresponds to the objective function of a social planner whose allocation coincides with the competitive equilibrium.

measured by the standard deviation of log real wage—0.03. This change, however, represents only a modest decrease (around 5%) from the large income dispersion in the base 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, 2017; Allen and Arkolakis, 2019) in three aspects. First, our structural estimation exploits changes in route choice of exporters in response to domestic road network expansion, which naturally implies that transport infrastructure investment reduces trade cost not only for trade between domestic partners but also for trade between the hinterland and foreign countries; 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. To understand the importance of these ingredients, we parameterize a series of restricted models and compare the inferred welfare gains through 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 model. Other structural parameters in the routing problem are kept 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 $\tau_{j,RoW} = \infty$, $\forall j$. The inferred gains from expressway construction in this model is about 10% (or 0.5 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 of a marginal reduction in the route cost along edge $m \to n$, denoted by ι_{mn}. We show in the appendix that, because the model without international trade is efficient, its competitive equilibrium coincides with the solution to the problem of a social planer with the Pareto weights equal to the income share of a region. Applying the envelop theorem to the social planer's problem gives the change in aggregate welfare W in response to a change in ι_{mn} as:

$$\begin{aligned} \frac{dW}{d\log \iota_{mn}} &= \sum_i \sum_{o \neq d} \frac{dW}{d\log \tau_{od}^i} \cdot \frac{d\log \tau_{od}^i}{d\log \iota_{mn}},\ o \neq \mathrm{RoW},\ d \neq RoW \\ &= -\sum_i \sum_{o \neq d} \frac{X_{od}^i}{Y} \cdot \pi_{od}^{mn},\ o \neq \mathrm{RoW},\ d \neq RoW. \end{aligned} \tag{19}$$

The first line applies the chain rule to express the elasticity as the product of two components: the marginal effect of trade cost on the aggregate welfare, and the marginal effect of route cost on trade cost. Because a reduction in the cost of the edge $m \to n$ can affect the trade costs between any pair of cities, the welfare effect sums over all sectors and city pairs. The second equality replaces $\frac{dW}{d\log \tau_{od}^i}$ with the ratio between value of trade flow and aggregate welfare. Intuitively, to

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 $\bar{\omega}$ are recalibrated to match the same city-sector sales (or city-level sales, depending on whether targeting city specializations) and the same average ground transportation distance.

the first order, the size of the benefit from a reduction in τ_{od}^i is the direct cost savings, which is proportional to X_{od}^i. This insight dates back to Hulten (1978) and has been used in evaluating the global gains from trade (Burstein and Cravino, 2012). The second equality also expresses the elasticity of τ_{od}^i to ι_{mn} as π_{od}^{mn} (Allen and Arkolakis, 2019). When θ is large enough, π_{od}^{mn} can be interpreted as the probability of trade flow from o to d passing the road $m \to n$. Taken together, Equation (19) suggests that, to the first-order, the welfare gains is simplify the value of all shipments passing through segment $m \to n$ as a fraction of aggregate welfare.[28]

Equation (19) does not hold in the full model because the social planer that replicates the competitive equilibrium places positive Pareto weights on the RoW. Loosely speaking, however, if we view the RoW as a reduced-form production function, then the first-order domestic welfare gains in the full model is given by an extended version of Equation (19) 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 generates 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—improvement in domestic infrastructures reduce the cost when firms from interior cities trade with the RoW. 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 encompassing 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 later group but only less than 1% in the southeast. How important is using the information on specialization for welfare evaluation?

To answer this question we re-calibrate a model without specialization. Specifically, we assume all sectors within a region have the same productivity, i.e., $T_o^i = T_o^j = T_o, \forall o, j$, and pin

[28] 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 a 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 exactly to the first order welfare gains.

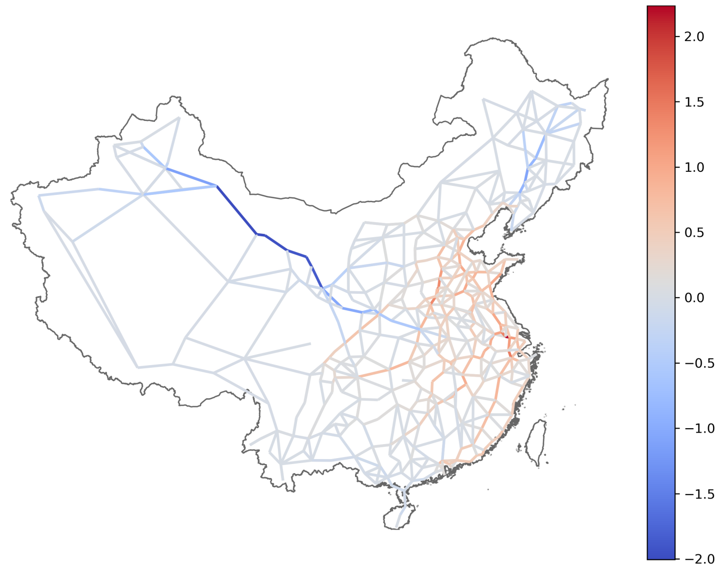


Figure 6: Differences in Shipment Value Shares, 'No Specialization' Minus 'Baseline'
Note: The number indicates the difference in shipment value/GDP. A segment with cold colors indicate that there is less shipment in the model with no specialization compared to the baseline.

down $\{T_o\}$ by matching the total sales of each city in the data. The input-output structure are same as in the baseline model. To make the comparison as clear as possible we assume there is no international trade, so Equation (19) remains valid. Column 3 of Table 10 reports that the inferred gains in this model are 15% smaller than an otherwise similar model with regional specialization (Column 2).

Patterns of regional specialization is important because they contain information for the distribution of trade flows across pairs of domestic partners. Because of the strong spatial clustering 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 the comparative advantages are eliminated, the spatial clustering also disappear. As a result, 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 statistics does not capture all trade patterns. Indeed, Figure 6 plots the change in shipment intensities between city pairs as we move from Model (2) to Model (3). The segments that see the biggest decrease in shipment are the ones that connect the northwest and northeast to the central areas; the segments that see more 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 project connecting the center to the northeast and northwest, regions with very different comparative advantages. Whether it underestimates or overestimates the return to a specific project thus depends crucially on where a project is. Under the actual network built made during the decade, the balance comes down to underestimating the welfare gains by 15%.

Transportation intensity. The comprehensive price information from the customs data allows us to incorporate sector heterogeneity in transport costs. To demonstrate the relevance of this channel, we set $\mu = 0$ and then calibrate the model to match both city-level sales and 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 intuition why sector heterogeneity is not simply washed out is, when calibrated to match the same average shipment distance, Model 3 infers systematically higher value of shipment compared to Model 3. More specifically, with trade cost heterogeneity, for the same level of inter-city shipment, Model 3 will predict a higher fraction of them coming from lighter sector (with higher value-to-weight ratio) because they incur lower shipping costs in Model 3 but not in Model 4. Because the welfare gains are proportional to the value of the goods, not their weights, the model with sector transport intensities predict larger welfare gains.

Intermediate inputs. In the final comparison, we further shut down intermediate inputs in production by assuming labor shares (β_l) is 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 (20).

$$\frac{X_{od}^i}{Y} = \frac{X_{od}^i}{\sum_i \sum_{o,d} X_{od}^i} \cdot \frac{\sum_i \sum_{o,d} X_{od}^i}{Y}. \tag{20}$$

Assuming, for simplicity, that all regions o and d are symmetric, with positive but symmetric inter-regional transport costs across sectors. When calibrated to match the average shipment distance, Models (4) and (5) generate the same trade intensity, i.e., $\frac{X_{od}^i}{\sum_i \sum_{o,d} X_{od}^i}$. However, in the model without intermediate inputs, the overall absorption $\sum_i \sum_{o,d} X_{od}^i$ is equal to the GDP, whereas in the model with intermediate inputs, the overall absorption is several (three in our calibration) times of the GDP. As a result, the inferred value of $\frac{\sum_i \sum_{o,d} X_{od}^i}{Y}$ is too small in the model without intermediate inputs: by assuming away intermediate inputs, the alternative model overlooks that goods are traded multiple times on the road, which amplifies the gains from reductions in transport cost.[29]

[29] 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 evaluating 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 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-provincial trade data is lacking, so we cannot directly measure trade/value added at regional level. Second, even when the data is available, at 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.

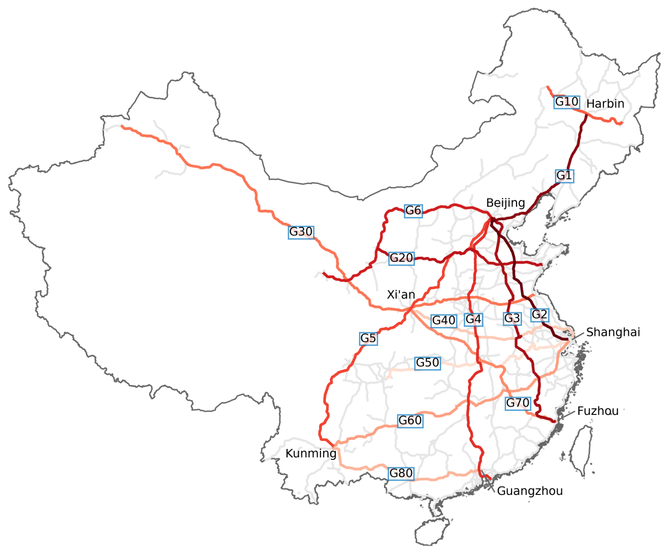


Figure 7: Mega Expressway Projects in China

Note: Projects with higher returns are plotted with darker colors. Some segments were completed by 1999 (most of G1 and G2); the newly built segments of these projects during 1999-2010 contribute to 43.7% of the total length expanded and 47.5% of the total cost of the entire expressway expansion during 1999-2010.

6.3 Cost-Benefit Analysis

The above model comparisons underscore the importance of incorporating all necessary 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 expressway network during 1999-2010 and infer the investment on individual projects.

We collect the investment on expressway projects from the yearly bulletin of road and waterway transport development published by the department of transportation and deflate yearly investment expenditures using the inflation rate for capital accumulation. Measured in 2010 price, the cumulative investment in inter-city highway projects 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 expressway is around 8%, the depreciation rate used for structures in Bai and Qian (2010). 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. Bai et al. (2006) finds that between 1998 and 2005, the return to capital is around 20%. This return seems unsustainable especially given the secular stagnation in much of the developed world. We stay conservative and assume the aggregate return to capital in China is around 10% in 2010.

Assuming all investment expenditures are made in 2010, then, the discounted future welfare gains ($5.2\% \times [1 + \frac{1-8\%}{1+10\%} + (\frac{1-8\%}{1+10\%})^2 ...]$) is around 27% of the 2010 GDP, implying a net return of about 170%: even taking into account the high opportunity cost in a growing economy like China, the expressway projects generated huge positive net return. In comparison, if we had

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	792.28%	0.60	0.94
G2	1768.29	0.38	85.94	0.45	511.37%	0.16	1.28
G3	2513.38	0.54	85.53	0.79	652.67%	0.65	4.37
G4	2924.88	0.65	89.14	0.40	211.30%	0.46	1.12
G5	2829.75	0.73	103.16	0.26	83.67%	0.26	0.51
G6	2095.37	0.38	72.26	0.17	123.78%	0.12	0.54
G10	891.73	0.15	67.25	0.12	295.57%	0.16	0.68
G20	1688.68	0.31	74.08	0.25	304.61%	0.30	0.73
G30	4356.49	0.85	78.04	0.63	278.57%	1.26	0.77
G40	1727.03	0.34	78.43	0.22	230.71%	0.46	0.93
G50	1936.36	0.38	79.61	0.26	242.27%	0.58	1.06
G60	2662.22	0.48	72.99	0.54	465.98%	1.08	2.13
G70	1706.35	0.38	89.62	0.43	478.00%	0.52	3.49
G80	1378.30	0.30	88.62	0.15	147.96%	0.23	0.83
Total	30012.46	6.16	-	5.16	-	6.84	19.37

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

used the simple one sector model for evaluation, as most of existing quantitative assessments of transport infrastructure do, our conclusion would have been that the investment led to 74% net losses.

A few mega projects with length more than 1000 km form the backbone of the entire expressway network. Figure 7 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, G3, and G4, on the other hand, connect Beijing to the 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) by adopting a formula based on the engineering literature which links the *relative* construction cost of a segment to the average slop of the terrain and whether it contains water or wetland areas. We then use the formula to evaluate all the segments constructed between 1999 and 2010 and choose the level coefficient in the formula so that the total cost of these segments is equal to the aggregate investment (10% of GDP).[30] Once the level coefficient of the formula is determined this way, we use it to evaluate the cost of the 14 projects. The appendix provides more details on this procedure.

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 southeast. Stretching across

[30] Among these projects, G1 and G10 had been largely completed by 1999. In calculating the total construction costs we exclude segments in these two projects.

the flat northeast plain, G10 in the other end of the country costs the least per km. The average cost across all projects constructed in this period around 80 million *yuan* per kilometer. This number is in the same ball park as best available evidence we can find.[31]

We evaluate the benefit of each project by removing it from the 2010 transport network and calculating the difference in aggregate welfare between the new equilibrium and the baseline economy. Columns 4 and 5 report the flow welfare gains and the net return to investment of these projects. Clearly, all these mega projects generate large returns. We indicate higher net gains using darker color in Figure 1a. The projects with the highest returns are north-south expressway lines (G1, G2, and G3). G5, running from Xi'an to Kunming, generates the lowest return, in part due to its large construction costs. The last two columns report the change in domestic and international trade as a share of GDP after each project is completed. The project that had the biggest impact on domestic trade are G30 and G60, which stretch across the vast central China. They had a larger impact on domestic trade than projects of comparable length 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 tan 90% of the gains from the entire network; 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 results highlights the importance of taking into account interaction among regions and transport infrastructures in welfare evaluation.

To summarize, large return heterogeneity notwithstanding, the expressway network in China was worth every penny of the investment at least until 2010. More recently, there has been 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 might incur higher costs while generate smaller returns.

7 First-Order Measurement and Second-Order Correction

In efficient (closed economy) macroeconomic settings, the Domar weight associated with a firm summarizes, up to the first order, the aggregate welfare effect of a shock to that firm (Hulten, 1978). This insight has also been used in the evaluating of transit projects, which measures the cost savings using the first-order approach. We chose to evaluate the full model with all three

[31]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 km.

ingredients in part because it allows us to infer the value of shipment in GDP accurately, as discussed in Section 6.2. This motivation is especially relevant in the domestic transport setting, in which directly measuring the value of shipment passing a route is difficult.[32] In this section, we further argue that because investment in expressway represent a large shock to both local linkages and the overall network, even if we had the perfect data, the first-order based approach do not work well and a second-order correction we develop improve the accuracy significantly.

7.1 First-Order and Nonlinear Effects for a Local Segment

We start with a local project that connect two adjacent cities, m and n, with an expressway segment. For local methods we assume the two cities are already connected by regular road so the expressway is an addition to the network.[33] Denote the (m, n) element of the combined road matrix $\mathbb{A}$ as $\iota_{mn}^{-\theta}$, then the percentage change in ι_{mn} is

$$\begin{aligned}\Delta \log(\iota_{mn}) &= -\frac{1}{\theta}\Big(\log[\exp(-\theta\kappa^H dist_{mn}) + \exp(-\theta\kappa^L dist_{mn})] - \log[\exp(-\theta\kappa^L dist_{mn})]\Big) \\ &\approx (\kappa^H - \kappa^L)\cdot dist_{mn},\end{aligned} \tag{21}$$

which is negative. We can expressway the change in welfare in response to this expressway segment as:

$$\begin{aligned}\Delta W &= \sum_i \sum_{o,d} \frac{dW}{d\log\tau_{od}^i}\Delta\log\tau_{od}^i + HO_T \\ &= \sum_i \sum_{o,d} -\frac{X_{od}^i}{Y}\Big(\pi_{od,mn}^i\cdot(\kappa_H-\kappa_L)\text{dist}_{mn} + HO_R\Big) + HO_T \\ &= \underbrace{\sum_i \sum_{o,d} -\frac{X_{od}^i}{Y}\pi_{od,mn}^i(\kappa_H-\kappa_L)\text{dist}_{mn}}_{\text{FO effect on shipments on m} \to \text{n}} - \underbrace{\sum_i \sum_{o,d}\frac{X_{od}^i}{Y}\cdot HO_R}_{\text{HO effect from routing}} + \underbrace{HO_T}_{\text{HO effect from trade}}.\end{aligned} \tag{22}$$

The first line of the equation expresses ΔW as the first order effect through changes in trade costs plus higher order effects. The first order effect captures the welfare gains under the assumption that intermediate and final good producers do not change their purchasing and export behavior. HO_T then captures the effects associated with re-optimization of producers. The second and third line then further decompose the FO effect from trade costs change into two components. The first is the FO effect from the change in $\Delta\log(\iota_{mn})$, which captures direct cost-savings from the expressway addition under the assumption that not only the intermediate and final good producers do not change their behavior, but also truck drivers keep the same routes. The

[32]Because of the granularity and complex interaction 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 vehicles using the that segment, it is the *value* of these shipments—which are much more difficult to estimate—not their weight, that matter for the aggregate welfare.

[33]Most pairs of adjacent cities were already connected by regular road, as Figure 1 shows.

second term, $\sum_i \sum_{o,d} \frac{X^i_{od}}{Y} \cdot HO_R$ then captures the additional effect, under the assumption that truck drivers re-optimize and producers re-optimize their mode choice—and thus generating a different trade cost matrix—but that producers do not react to the new trade costs by choosing a different bundles of inputs. The errors from the first order approximation is the sum of the second and third terms.

A general characterization of the two higher order effects, from the re-optimization of drivers and firms, respectively, is difficult, we examine only the *second* order effect through the routing decision of drivers and show that it captures most of the approximation errors in the first-order approach. Formally, in the appendix we show that the HO effect from driver re-optimization is

$$\begin{aligned} HO_R &= \frac{1}{2}\frac{d\pi^{mn}_{od}}{d\log\iota_{mn}} \cdot [(\kappa^H - \kappa^L)d_{mn}]^2 + o_R(2) \\ &= \underbrace{-\theta\pi^{mn}_{od}[1 + \pi^{mn}_{om} + \pi^{mn}_{nd} - \pi^{mn}_{od}] \cdot [(\kappa^H - \kappa^L)d_{mn}]^2}_{\text{Second order effect from re-optimization of trucks}(SO_R)} + o_R(2). \end{aligned}$$

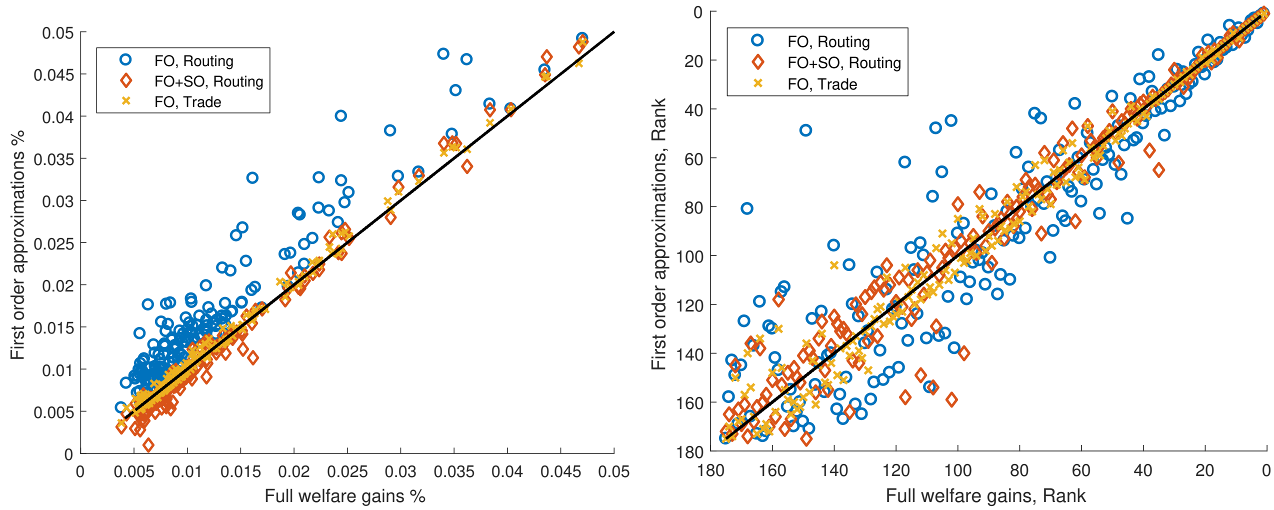
Observe that the second order effect from driver routing is negative, i.e., taking into account SO_R implies a larger increase in the aggregate welfare for a decrease in route cost and a smaller welfare loss for an increase in route cost. Intuitively, when reoptimization is allowed, trucks taking the expressway between $m \to n$ might opt out to other routes. The additional costs will therefore will be lower than if they are forced to take the regular road between $m \to n$. The residual $o_R(2)$ contains both the switching behaviors between different modes of transport, and the third- or higher-order effects of re-routing within a mode.[34] With this decomposition, we can write Equation (22) as:

$$\Delta W = \overbrace{\underbrace{\underbrace{\sum_i \sum_{o,d} -\frac{X^i_{od}}{Y}\pi^i_{od,mn}(\kappa_H - \kappa_L)\text{dist}_{mn}}_{\text{FO effect}} - \sum_i \sum_{o,d} \frac{X^i_{od}}{Y} \cdot SO_R}_{\text{SO allowing for truck re-routing}} + \sum_i \sum_{o,d} X^i_{od} \cdot o_R(2)}^{\text{FO effect from trade}} + HO_T. \tag{23}$$

We use this equation to evaluate the quantitative significance of various forces. To this end, we consider the top 200 segments (pairs of adjacent cities) in shipment in China. By 2010, all 200 pairs were already connected by an expressway, and 175 were also connected by a regular road. We conduct 200 experiments, in which we remove one of the expressway segment at a time and calculate the aggregate welfare gains from that individual segment. In doing these exercises, we assume there is no international trade so the first order approximation based on the social planner's problem is correct. We then compare the full effect of each segment to different approximations, assuming we have the perform knowledge from the simulation of the full model.

Figure 8a plots results. The horizontal is the full welfare effect. The range of welfare gains are

[34] Because the elasticity of substitution between modes are small (2.5 in calibration), the former, the response through changes of mode is very small. For this reason, while this force is a second order effect, we attribute it in $o_R(2)$.



(a) Welfare Gains, Levels

(b) Welfare Gains, Ranks

Figure 8: Nonlinear Welfare Gains v.s First-order Approximations

Note: Each point corresponds an experiment by removing one expressway segment at a time. The sample segments are the ones with top 200 shipment flows in the baseline model, among which 25 linkages with no regular roads are dropped since the FO and SO cannot be evaluated.

Table 12: Error in First-order Approximations of Welfare Gains

	Mean Error	Mean Absolute Error
FO, Routing	29.8%	29.8%
FO+SO, Routing	-7.1%	10.1%
FO, Trade	1.2%	4.4%

Note: The statistics are computed across the 175 experiments removing one expressway segment at a time. *Error*: log(approximation)-log(non-linear welfare gains). Mean Error: the mean of *Error*. Mean Absolute Error: the mean of $|Error|$. Std Error: the standard deviation of *Error*.

between 0.005% to 0.05%. Even within the busiest subset of the road network, some segments are ten times as much important than others—a Zipf's law for inter-city traffic. The vertical axis are predictions based on various approximations. The blue circles denote the first order cost savings, calculated directly using the actual shipment on each segment. It corresponds to the 'FO effect' in Equation (23) and captures the total welfare gains assuming once expressway segment $m \to n$ is removed, all traffic originally on it switch to the regular road between $m \to n$ but not other routes. Because for the 25 pairs without a regular road, the first order approximation is invalid (see Equation 21), we calculate this for the remaining 175 cities only. As anticipated, all circles lie above the 45 degree line, indicating that the first-order approach overestimates the welfare losses from the removal of the expressways. The biases are smaller in percent for segments that are less busy. Table 12 shows that the average log difference between the FO and the full effect is 30%.

The diamonds further incorporates the second order effect from re-routing among road networks, assuming trade flows and the choice of mode do not change. They correspond to 'SO from routing' in Equation (23). The prediction is now much improved, centered tightly around the 45 degree line, although for the segments with less traffic, this approximation underestimates the welfare gains. On average, the mean absolute error of this approximation is 10.1%. Finally, the crosses incorporates all responses in routing (corresponding to 'FO effect from trade' in Equation 23). This further improves the quality of the approximation, reducing the mean absolute error to 4.4%.

To identifying worthwhile investment projects, sometimes the ranking of project returns matter more than their level. Figure 8b plots the ranking based on the full equilibrium effect against the ranking under different approximations. Because of the large dispersion in the importance of the traffic, especially at the very top, the quality of first-order approximation is reasonably good for the busiest segments. At the lower and middle range, the relationship between the two ranks are nosier. As we incorporate second order effect on routing, the relationship become tighter; once all higher order routing effects are incorporated, the predictions are centered tightly around the 45 degree line, with only small deviations at the lower end of the distribution.

To summarize, we demonstrate that the ex-post first-order approximation tends to overestimate the gains from addition of an expressway—because routes passing different cities are highly substitutable and because addition of an expressway segment represents a large shock to the local road cost, drivers' responses through re-routing are large enough to undermine the quality of the first-order approximation. Empirically, this bias is smaller for segments with less traffic and averages around 30% for the top 200 busiest segments of the country. We propose a second-order correction that amend this first-order effect, and show that this correction can reduce the mean absolute error by two thirds, form 30% to 10%. In principle, the resulting change in trade costs from the expressway removal can lead to change in firms' trade decision, leading to higher order effect from trade. In our setting with Cobb-Douglas production function and a reasonable trade elasticity (6), which is a common baseline model in the quantitative trade literature, this channels accounted for only 4.4% of the total welfare gains, and is only around 15% (4.4/30) of the approximation error in the first-order approach.

7.2 Interaction Between Segments in the Network

We extend the first-order approximation and second-order correction to large projects that build multiple segments at the same time.

$$\Delta W = \underbrace{\sum_{mn} \frac{dW}{d\log(\iota_{mn})} \Delta \log(\iota_{mn})}_{\text{FO effect}} + \underbrace{\frac{1}{2} \sum_{mn} \sum_{m'n'} \frac{d \frac{dW}{d\log(\iota_{mn})}}{d\log(\iota_{m'n'})} \Delta \log(\iota_{mn}) \Delta \log(\iota_{m'n'})}_{\text{Interaction between segments}} + o(2). \tag{24}$$

The first term on the right side of the equation is simply the sum of the first-order effect of individual segments. This summation corresponds to what we would have obtained, if we conduct the ex-post evaluation using the first-order approach. The second term captures the interaction between different segments—the marginal effect of the improvement along segment $m \to n$ depends on the cost along the segment $m' \to n'$. The third term on the right hand side captures all higher order effects.

Theoretically, the second term on the right hand side could be positive or negative. For example, if two segments are on two routes that are close substitutes, then the interaction would imply that the marginal gains of improving one segment decreases on the quality of the other. The opposite is equally plausible. If, for example, two segments are on the same route between two partners trading intensively with one another, then the investment that reduces costs on one segment draws more traffic into the route, thus increasing the return from investing in other segments on the route. The second order term can be characterized as follows:

$$\begin{aligned} \frac{d\frac{dW}{d\log(\iota_{mn})}}{d\log(\iota_{m'n'})} &= -\sum_i \sum_{o \neq d} \frac{X^i_{od}}{Y} \frac{d\pi^{mn}_{od}}{d\log(\iota_{m'n'})} \\ &= -\sum_i \sum_{o \neq d} \frac{X^i_{od}}{Y} \cdot (-\theta) \cdot \pi^{mn}_{od} [\pi^{m'n'}_{om} + \pi^{m'n'}_{nd} - \pi^{m'n'}_{od}]. \end{aligned} \tag{25}$$

We use this to decompose the welfare gains from simultaneous constructions of multiple segments, given by Equation (24). Further, we can decompose the interaction effects into 'own' ($mn = m'n'$) versus 'cross' ($mn \neq m'n'$). We use the experiments on the busiest segments for this experiment, decomposing the full nonlinear effects of removing all 175 expressway segments at the same time into various components, given below:

$$\underbrace{\text{Full}}_{0.023} = \underbrace{\text{FO}}_{128\%} + \underbrace{\text{Own SO}_R}_{-32\%} + \underbrace{\text{Cross SO}_R}_{10\%} + \underbrace{\text{residual}}_{-7\%}.$$

The welfare gains from constructing all the 175 segments at once is 2.3%. The first-order effect based on ex-post traffic overestimate this number by 28%. The own substitute term more than corrects this bias; the cross second order effect adds another 10% to the welfare gains. Note also these cross second order effect masks segments that are strongly complementary—those on the same trade route—and those that are highly substitutable. All inclusive, what is left for the approximation error is mere 7%. Equations (24) and (25) thus give 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 infer the effect of transport infrastructure on domestic trade cost—by using the route choice of exporters, which are measurable in increasingly accessible customs data and circumvents the lack of reliable domestic trade data in many countries. We combine this method and a spatial equilibrium model to study the aggregate welfare effects of the $60,000$ kilometer expressway construction taking place between 1999 and 2010 in China. We find the overall welfare gains is around 5% and the net return to be around 170%. We further show that the three key ingredients incorporated in the model, regional specialization, sector heterogeneity in transport, and intermediate inputs, are important for this evaluation and their omission will lead to the conclusion of negative aggregate return.

The reason why having the ingredients is important is that in the absence of reliable data on domestic trade, we rely on model structure in inferring the value of inter-city shipments on the road network. We further argue that because of the interaction between segments and re-optimization of drivers' behavior, for both the overall projects and for addition of an expressway segment between two adjacent cities, the first-order based approach to transport infrastructure evaluation will lead to significant biases. We propose a second order correction that significantly reduces the biases, and can be used flexible measure welfare gains from large projects.

A separate contribution of the paper is to provide an estimate of transportation cost along expressways and regular roads, using over-time variation from China's great expansion of the highway network, which could be an input into future quantitative studies of spatial equilibrium models on China. Moreover, we show that the model-predicted export growth in response to exogenous component of the expressway network expansion predicts well the actual export growth in the data. Under some assumption, this provides and IV for studies of export on city outcomes.

References

Alder, Simon, "Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development," 2016.

_ and Illenin Kondo, "Political Distortions and Infrastructure Networks in China: A Quantitative Spatial Equilibrium Analysis," 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," 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," 2015.

Bai, Chong-En and Yingyi Qian, "Infrastructure development in China: the cases of electricity, highways, and railways," *Journal of Comparative Economics*, 2010, *38* (1), 34–51.

—, Chang-Tai Hsieh, and Yingyi Qian, "The return to capital in China," Technical Report, National Bureau of Economic Research 2006.

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

Baqaee, David and Emmanuel Farhi, "Networks, barriers, and trade," 2019.

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

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

Baum-Snow, Nathaniel and Matthew E Kahn, "The effects of new public projects to expand urban rail transit," *Journal of Public Economics*, 2000, *77* (2), 241–263.

—, 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," Working Paper 17767, National Bureau of Economic Research January 2012.

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

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

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.

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," Technical Report, National Bureau of Economic Research 2017.

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.

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.

Leemput, Eva Van, "A passage to India: Quantifying internal and external barriers to trade," *FRB International Finance Discussion Paper*, 2016, (1185).

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.

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 et al., "City location and economic development," *Princeton University, mimeograph*, 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 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," 2018.

Simonovska, Ina, "Income differences and prices of tradables: Insights from an online retailer," *The Review of Economic Studies*, 2015, *82* (4), 1612–1656.

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

Smith, G., *China expressway: connecting people and markets for equitable development*, The World Bank, 2007.

Storeygard, Adam, "Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa," *The Review of economic studies*, 2016, *83* (3), 1263–1295.

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

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," *Unpublished manuscript*, 2018.

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

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

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

A Data and Empirics

A.1 Defining City Coordinates

Starting with the county geo information in the 2010 census, we first define the location of a county by its center of mass. We then weight the coordinates of all counties making up a city by their population to calculate an average coordinate, which we define as the location of a prefecture city. Four the four provincial-level city, Beijing, Shanghai, Tianjin, and Chongqing, we proceed slightly differently. We treat Beijing, Shanghai, Tianjin, each as a city, and aggregates up one set of coordinates for each of them based on the coordinates of its districts (urban sub-units) and counties (rural sub-units). We treat Chongqing differently as two separate cities because geographically it is too large to view as one city. We treat all the urban areas of Chongqing (District) as a city, and all the rural areas (county) as a city. For each of the two locations, we calculate the weighted average coordinates across sub-units.

A.2 Constructing Network Graphs

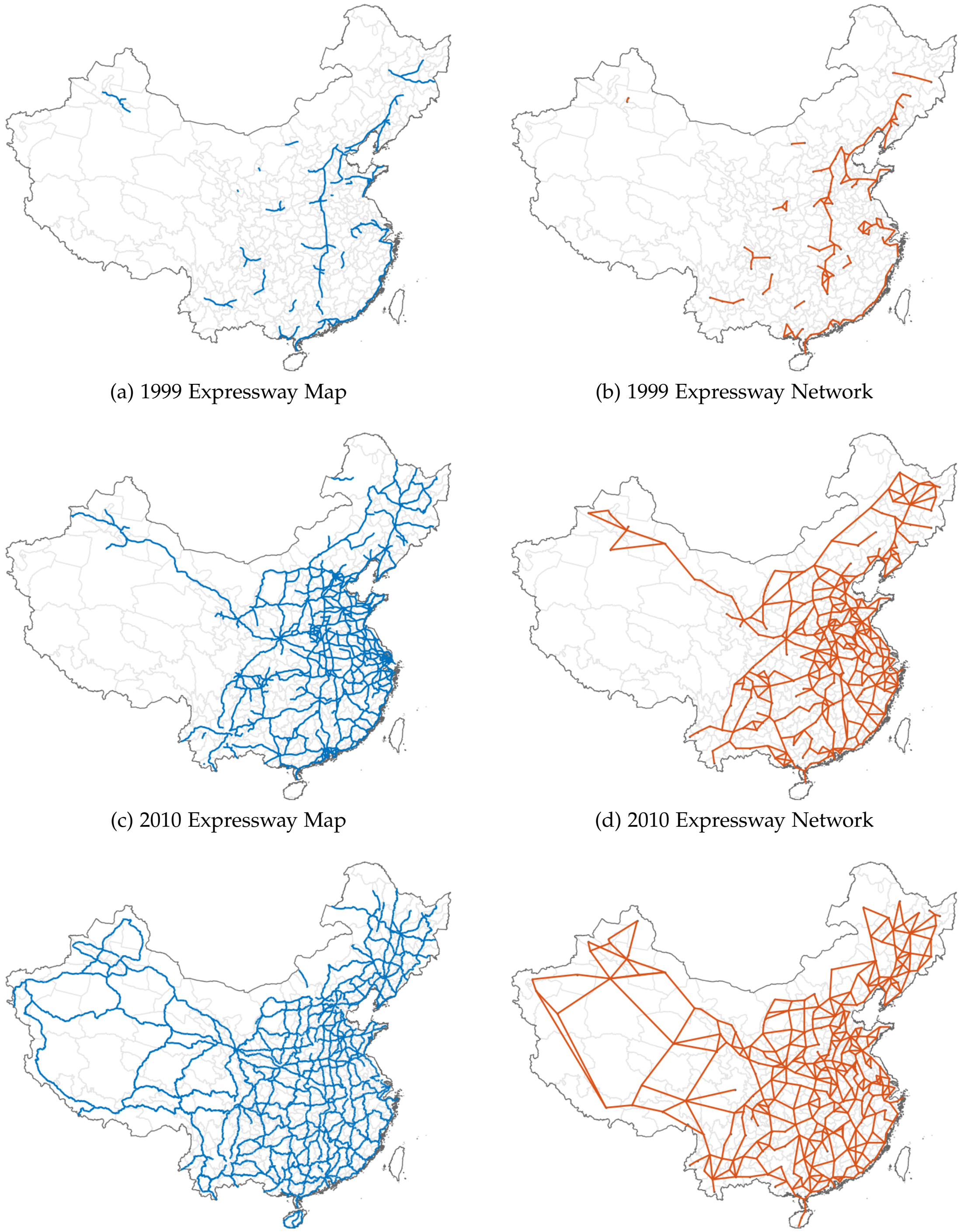
Our raw data consist of geographic coordinates of the center of a city and line string of the road networks (1999 and 2010 expressway, and 2007 regular road which is treated as time-invariant). To combine the data and model, we first need to generate a network of with the roads that link them. We do 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 with 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 roads on the map, using the Dijkstra's algorithm. If the shortest path between two cities do not pass through 30 kilometer radius of another city, we define these two cities two be 'directly connected', meaning they are adjacent and connected by a road.
- **Construct the graph.** We construct the graph in which cities and roads are graphs 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 'iron out' the local curvatures in constructing the network, which helps us eliminate measurement

when comparing the expressway networks across two periods.[35] Note that in this graph, each city can only be connected to another city through its adjacent cities.

The left panel of Figure 9 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 corresponds to the network structure underlying $\mathbb{H}^{1999}$, $\mathbb{H}^{2010}$, and $\mathbb{L}$.

[35]The two expressway maps are digitized from projection of published hard-copy maps, which create measurement errors that changes the location of the same road.



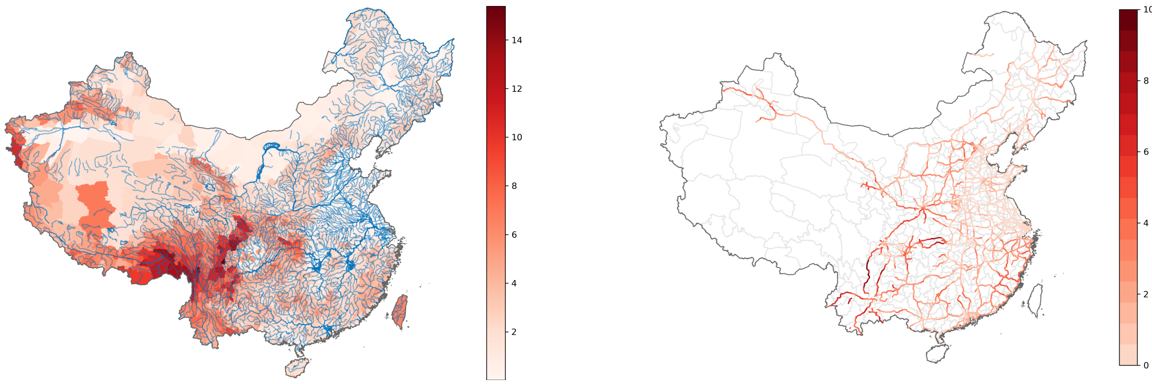
(a) 1999 Expressway Map (b) 1999 Expressway Network

(c) 2010 Expressway Map (d) 2010 Expressway Network

(e) Regular Road Map (f) Regular Road Network

Figure 9: From Road Maps to Road Networks

Note: Two cities are defined as connected on a road network if they are adjacent and the shortest path that connects them on a road network does not pass a third city. The distance between two connected cities is then calculated as the distance along the respective road network.



(a) Slope Gradients and Water Areas

(b) Slope Gradients of Expressway Segments

Figure 10: Geography and Expressway Construction Costs

Note: The left figure plots the slope of land and the geographic distribution of wetland and water areas. The right panel plots the expressways in 2010, idicating 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 gradient of each segment and determine whether the segment passes water or wetland areas.[36] 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 similar to the ones used Faber (2014), except that we abstract from the measure of existing building due to the lack of data. 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, based on the estimation by Smith (2007) and the national statistics.

The total cost during this period is 9.92% of the 2010 GDP, or 3983 billion 2010 CNY. The total dry-plain equivalent distance of all roads constructed during this period is 453447 kilomter, so each dry-plain equivalent kilometer of expressway costs about 8.85 million 2010 CNY. The total actual length of expressway constructed during this period is 49760 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 are not on dry plains.

A.4 Additional Results on Reduced-Form Regressions

Table 13 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. The third and fourth columns use the IV based on the minimum-spanning tree.

[36] 23.3% of the segments pass water areas.

Table 13: 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 sector fixed effects 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.
Standard errors are clustered at city-port level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 14: Price Distance Regression

	(1)	(2)	(3)	(4)
	OLS	OLS	OLS	IV
dist	0.032***	0.012*	0.065**	0.113***
	(0.007)	(0.006)	(0.028)	(0.031)
Fixed Effects				
Observations	10459	9827	9820	9820
R^2	0.114	0.721	0.740	0.001
IV				279.454

Notes: This Table reports robustness analysis of results in Table 5.
Standard errors are clustered at city-port level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 14 reports additional results of the price-distance regression reported in Table 5, using over-time variations but a broader definition of product (HS 2 category).

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.

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

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 edges with exactly N segments is $[A^N_{(od)}]$, then the sum across all edges with exactly $N+1$ segments is:

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

The first part the sum across all the paths that gets to an adjacent city of d and then goes to d through the final a final exprssway segment; the second part is the sum across all the path that goes from o to an adjacent city to d in exactly N steps, and then goes to d in the final edge through a regular road segment.

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

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

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

B.2 Definition of Equilibrium

The competitive equilibrium can be defined using prices and expenditure shares as equilibrium objects (as in e.g. Caliendo and Parro (2015)). We also provide a definition based on prices and quantities to facilitate establishing the equivalence of the competitive equilibrium and the solution to a social planner's problem described in subsection B.3.

Definition 1. *Given fundamentals* $\{\tau^i_{od}, L_d, T^i_d\}$[37]*, a competitive equilibrium is: (1) consumption allocation* c^i_d*, labor allocation* l^i_d*, uses of final goods as input* m^{ij}_d*, production of 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* κ^i_o*, wages* w_d*, s.t.*

- *Consumers' optimization conditions hold:*

$$\alpha^i_d w_d L_d = P^i_d c^i_d. \tag{26}$$

[37] The driver's routing solution in equilibrium has been taken incorporated and captured by the vector of trade cost τ^i_{od}.

- *Intermediate goods producers' optimization conditions hold:*

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

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

$$P^j_o m^{ij}_o = \kappa^j_o \gamma^{ij}_o [l^i_o]^{\beta^i_o} \prod_k [m^{ik}_o]^{\gamma^{ik}_o} \tag{27}$$

$$w_d l^i_o = \kappa^j_o \beta^i_o [l^i_o]^{\beta^i_o} \prod_k [m^{ik}_o]^{\gamma^{ik}_o} \tag{28}$$

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

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

- *Final goods producers' optimization conditions hold:*

$$Q^i_d = \left(\sum_o [\tilde{q}^i_{od}]^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$$

$$\frac{\tilde{q}^i_{od}}{\tilde{q}^i_{\tilde{o}d}} = [\frac{p^i_{od}}{p^i_{\tilde{o}d}}]^{-\sigma} \tag{29}$$

$$P^i_d = \left(\sum_o [p^i_{od}]^{1-\sigma} \right)^{\frac{1}{1-\sigma}}.$$

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

$$\sum_i l^i_d = L_d \qquad \text{(Labor markets clear)}$$

$$\sum_d \tau^i_{od} \tilde{q}^i_{od} = q^i_o \qquad \text{(Intermediates goods markets clear)}$$

$$\sum_i m^{ij}_d + c^j_d = Q^j_d \qquad \text{(Final goods market clear).}$$

B.3 Welfare Criteria

We now establish the First Welfare Theorem with the appropriate choice of Pareto weights for the social planner's problem. The Pareto weights are used to evaluate the aggregate welfare and to calculate the change of welfare between equilibria. As we show below, the change in the aggregate welfare defined here can be interpreted as the average change in log real income across regions weighted by the region's initial value added. The proof can be viewed as an application of the standard equivalence result in an Arrow–Debreu equilibrium with production, if we view 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 1. *The allocations in the competitive equilibrium can be replicated by the solution to the following social planner's problem*

$$W = \max_{l_d^i, c_d^i, m_d^{ij}, q_d^i, Q_d^i, \tilde{q}_{od}^i} \sum_d \omega_d \log \Big(\prod_{i=1}^{S} [c_d^j]^{\alpha_d^i} \Big) \tag{30}$$

subject to

$$\begin{aligned}
q_d^i = T_d^i [l_d^i]^{\beta_d^i} \prod_{j=1}^{S} [m_d^{ij}]^{\gamma_d^{ij}} & \qquad \textit{(Production of intermediate goods)} \\
Q_d^i = \Big(\sum_o [\tilde{q}_{od}^i]^{\frac{\sigma-1}{\sigma}} \Big)^{\frac{\sigma}{\sigma-1}} & \qquad \textit{(Production of final goods)} \\
\sum_i l_d^i = L_d & \qquad \textit{(Resource constraints for labor markets)} \\
\sum_d \tau_{od}^i \tilde{q}_{od}^i = q_o^i & \qquad \textit{(Resource constraints for intermediates goods)} \\
\sum_i m_d^{ij} + c_d^j = Q_d^j & \qquad \textit{(Resource constraints for final goods),}
\end{aligned}$$

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

Proof. The Lagrangian for the planner's problem is[38]

$$\begin{aligned}
\mathcal{L} = & \sum_d \omega_d \log \Big(\prod_{i=1}^{S} [c_d^i]^{\alpha_d^i} \Big) + \sum_{d,j} \lambda_d^j \Big(Q_d^j - c_d^j - \sum_i m_d^{ij} \Big) + \sum_{o,i} \mu_o^i \Big(T_o^i [l_o^i]^{\beta_o^i} \prod_{j=1}^{S} [m_o^{ij}]^{\gamma_o^{ij}} - \sum_d \tau_{od}^i \tilde{q}_{od}^i \Big) \\
& + \sum_{d,i} v_d^i \Big((\sum_o [\tilde{q}_{od}^i]^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}} - Q_d^i \Big) + \sum_d n_d \Big(L_d - \sum_i l_d^i \Big)
\end{aligned} \tag{31}$$

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

$$\sum_d \tau_{od}^i \tilde{q}_{od}^i = T_o^i [l_o^i]^{\beta_o^i} \prod_{j=1}^{S} [m_o^{ij}]^{\gamma_o^{ij}}$$

We have reindexed variables accordingly for the convenience of expositions.

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

$$\{c_d^i\} : \frac{\omega_d \alpha_d^i}{c_d^i} = \lambda_d^i \tag{32}$$

$$\{m_o^{ij}\} : -\lambda_o^j + \mu_o^i T_o^i \gamma_o^{ij} [m_o^{ij}]^{-1} [l_o^i]^{\beta_o^i} \prod_j [m_o^{ij}]^{\gamma_o^{ij}} = 0 \tag{33}$$

$$\{Q_d^i\} : \lambda_d^i - v_d^i = 0 \tag{34}$$

$$\{\tilde{q}_{od}^i\} : v_d^i \{[\tilde{q}_{od}^i]^{\frac{\sigma-1}{\sigma}-1} (\sum_{\tilde{o}} [\tilde{q}_{\tilde{o}d}^i]^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}-1}\} - \mu_o^i \tau_{od}^i = 0 \tag{35}$$

$$\{l_o^i\} : -n_o + \mu_o^i T_o^i \beta_o^i [l_o^i]^{-1} [l_o^i]^{\beta_o^i} \prod_j [m_o^{ij}]^{\gamma_o^{ij}} = 0. \tag{36}$$

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

$$\lambda_d^i = \frac{P_d^i}{Y},\ \mu_o^i = \frac{\kappa_o^i}{T_o^i} \frac{1}{Y},\ v_d^i = \frac{P_d^i}{Y},\ n_o = \frac{w_o}{Y}. \tag{37}$$

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 (26) we arrive at (32). Plug λ_d^i and μ_o^i into (27) we arrive at (33). (34) is implied by construction of λ_d^i and v_d^i. Plug μ_o^i and v_d^i into (35) we have

$$\begin{aligned}
(35) \Leftrightarrow \frac{P_d^i}{Y} (\sum_{\tilde{o}} [\tilde{q}_{\tilde{o}d}^i]^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}} \frac{[\tilde{q}_{od}^i]^{\frac{\sigma-1}{\sigma}}}{\sum_{\tilde{o}} [\tilde{q}_{\tilde{o}d}^i]^{\frac{\sigma-1}{\sigma}}} &= \frac{\kappa_o^i}{z_o^i} \frac{1}{Y} \tau_{od}^i \tilde{q}_{od}^i \\
\Leftrightarrow \frac{[\tilde{q}_{od}^i]^{\frac{\sigma-1}{\sigma}}}{\sum_{\tilde{o}} [\tilde{q}_{\tilde{o}d}^i]^{\frac{\sigma-1}{\sigma}}} &= \frac{p_{od}^i \tilde{q}_{od}^i}{P_d^i Q_d^i} \\
\Leftrightarrow \frac{[\tilde{q}_{od}^i]^{\frac{\sigma-1}{\sigma}}}{\sum_{\tilde{o}} [\tilde{q}_{\tilde{o}d}^i]^{\frac{\sigma-1}{\sigma}}} = \frac{p_{od}^i \tilde{q}_{od}^i}{\sum_{\tilde{o}} p_{\tilde{o}d}^i \tilde{q}_{\tilde{o}d}^i} &\Leftrightarrow (29).
\end{aligned}$$

Finally, plug n_o into (28) we arrive at (36). 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 1 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. With the interpretation that trade is another form of production technology, this result simply says that the marginal effect of improvements in an exporter's efficiency on aggregate welfare is the exporter's 'sales' ratio.

Lemma 2. *With the social planner's welfare function defined in (30), we have*

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

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

Proof. At the solution to the social planner's problem

$$\frac{dW}{d\tau_{od}^i} = \frac{d\mathcal{L}}{d\tau_{od}^i}$$

where $\mathcal{L}$ is the Lagrangian defined in (31). Thus the envelope theorem implies that at the solution,

$$\frac{dW}{d\tau_{od}^i} = -\mu_o^i q_{od}^i,$$

where μ_o^i is the Lagrangian multiplier and is associated with equilibrium objects through (37), restated here

$$\mu_o^i = \frac{\kappa_o^i}{T_o^i}\frac{1}{Y} = \frac{p_{od}^i \tau_{od}^i}{Y}.$$

Therefore

$$\frac{dW}{d\tau_{od}^i} = -\frac{p_{od}^i q_{od}^i \tau_{od}^i}{Y}$$

$$\Rightarrow \frac{dW}{d\log\tau_{od}^i} = -\frac{X_{od}^i}{Y}.$$

□

C Quantification

C.1 Numerical Implementation

Although the solution methods for this class of model have become standard, we describe the design of the algorithm that makes it possible to load the most intensive part of the computation to a GPU. This enables us to solve equilibria robustly and efficiently, despite the size of the problem (our benchmark model has 323 prefecture cities plus one RoW, and 21 two-digit tradable sectors plus 4 non-tradable sectors).[39] The large size of the problem also renders a well-known

[39]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. To

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 $(324 \times 25)^2$ entries. Our algorithm falls back to a fixed point algorithm described below.

The minimal system of equations that can be used to solve the equilibrium is[40]

$$\begin{aligned} E_o^j &= \alpha^j w_o L_o + \sum_i \gamma_o^{ij} \sum_d \pi_{od}^i E_d^i \\ w_o L_o &= \sum_i \beta_o^i [\sum_d \pi_{od}^i E_d^i] \\ P_d^i &= \Big(\sum_o [p_{od}^i]^{1-\sigma} \Big)^{\frac{1}{1-\sigma}}, \end{aligned} \tag{38}$$

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

$$\begin{aligned} p_{od}^i &= [\mathbb{C}_d^i w_d^{\beta_d^i} \prod_{j=1}^{S} [P_d^j]^{\gamma_d^{ij}} \tau_{od}^i] / T_o^i \\ \pi_{od}^i &= \frac{[p_{od}^i]^{1-\sigma}}{[P_d^i]^{1-\sigma}}. \end{aligned} \tag{39}$$

We design a nested fixed point algorithm according to the strength of the hardware. A key observation is that given π_{od}^i, the first two equations of (38) give a (dense) system of linear equations for E_d^i and $w_d L_d$, for which GPUs are designed to solve very 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 $(w_{d,Old}, P_{d,Old}^i)$
2 Set flag_converged to **false**
while flag_converged is false **do**
  3 Construct π_{od}^i according to (39) based on $(w_{d,Old}, P_{d,Old}^i)$
  4 Solve the system of linear equations for E_d^i and $w_d L_d$ (with GPUs)
  5 Construct p_{od}^i, P_d^i according to (39) and (38)
  6 Set flag_converged to **true** if distance between (w_d, P_d^i) and $(w_{d,Old}, P_{d,Old}^i)$ is small enough

  7 Update $(w_{d,Old}, P_{d,Old}^i)$ according to (w_d, P_d^i)
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_d^i), the benchmark equilibrium can be solved (under the convergence

evaluate the non-linear effects of the improvements in the 200 busiest segments one at a time, we need to solve 200 counter-factual equilibria with high accuracy (this could be parallelized so speed is less of a concern).

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

Table 15: Robustness Checks

Change in	(1) High Heterogeneity in Sectoral Transport Cost	(2) External Economy of Scale	(3) Free Migration
Aggregate welfare	0.068	0.054	0.048
Log(Domestic trade / GDP)	0.085	0.165	0.096
Log(Exports / GDP)	0.165	0.208	0.090
Std Log(real wage) across regions	-0.027	-0.022	-

Note: The counterfactual equilibrium is solved by replacing the 2010 expressway network with the 1999 expressway network from an alternative benchmark. For all three set of experiments, the alternative benchmark is recalibrated to match the same targets as in Table 6.

criterion of $1e-6$ in log difference) within a minute with a GTX1080Ti GPU, compared to around 10 minutes with 2*Intel Xeon CPU E5-2650 v4.

C.2 Sensitivity Analysis

We conduct a number of exercises to assess the robustness of the baseline results. We focus on three alternative assumptions. The first is on the sector heterogeneity of transport costs. Instead of 0.3, 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 $\chi = 0.13$ estimated by Bartelme et al. (2018), so there is increasing payoff to specialization. Finally, we consider a benchmark with free mobility to give a bound on the effect of mobile labor.

Table 15 shows that as expected, when we increase sector heterogeneity in transport cost, the inferred welfare gains are larger. Adding external economies of scale at the industry level or making workers freely mobile reduces the gains somewhat, but the overall findings are similar.