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

Jingting Fan† Yi Lu‡ Wenlan Luo§

This version: June, 2021

Abstract

A key input to quantitative evaluations of transport infrastructure projects is their impact on transport costs. We propose a new method of estimating this impact relying on widely accessible customs data: by using the route choice of exporters. We combine our method with a spatial equilibrium model to study the effects of the massive expressway construction in China between 1999 and 2010. We find transport costs are 20% lower on expressways than on regular roads. The expressways construction increases aggregate exports by 10% and domestic trade by 14%. It generates 5.1% welfare gains, implying a 150% net return to investment.

JEL codes: R13, R42, F14

*We thank the editor and four anonymous referees, whose comments significantly improved the paper. We also benefit from helpful comments by Treb Allen, Costas Arkolakis, David Baqaee, Lorenzo Caliendo, Kerem Cosar, Fernando Parro, Nathaniel Young, and participants at the 2018 Nankai University International Economics Workshop, the 2019 Hong Kong University Globalization and Firm Dynamics Workshop, the SHUFE-ISER Online Trade Workshop, Fudan University, the National School of Development at Peking University, and the University of Tokyo. We thank Pin Sun for excellent research assistance and Jingjing Chen for help in accessing the data. All errors are our own.

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

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

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

1 Introduction

In 2016, the 47 member countries of the International Transport Forum—including OECD countries, China, and others—invested over 850 billion euros in inland transport infrastructure (OECD, 2019). In China, the focus of this paper, investments in inland transport infrastructure increased steadily from 2% of GDP in 2000 to 5% in recent years. China alone accounted for more than half of investments made by the above 47 countries. The sheer size of investments in China and elsewhere has renewed the interest of academics and policy makers in understanding the returns to infrastructure projects. While earlier studies either perform a measurement exercise (e.g., Fogel, 1964) or adopt a reduced-form approach (e.g., Banerjee et al., 2020), aided by new tools from international trade and spatial economics, a growing strand of the literature develops quantitative models to evaluate transport infrastructure through counterfactual experiments.

A key input into such quantitative exercises is the mapping from distance along the transport network to trade cost.[1] Two approaches of estimating this mapping feature prominently in the literature. The first uses shipment data, such as the Commodity Flow Survey in the U.S. (Allen and Arkolakis, 2014, 2019). The second relies on price data, the idea being, given assumptions on cost pass-through, variations in the price of the same good across locations identify trade costs (e.g., Donaldson, 2018; Asturias et al., 2018).

The data requirements of both approaches are quite demanding. Indeed, many countries do not collect or make accessible their versions of the Commodity Flow Survey; in the U.S., the survey started in 1993, when the inter-state highway system had been virtually completed.[2] Perhaps for this reason, most studies using the U.S. data or similar data from other countries rely on cross-sectional variation for estimation. The second approach—the price approach—requires products to be homogeneous, so its application

[1]When trade costs are specified as a log linear function of distance, this mapping is governed solely by trade cost elasticity. This elasticity should be differentiated from trade elasticity, which governs how trade flows respond to trade costs.

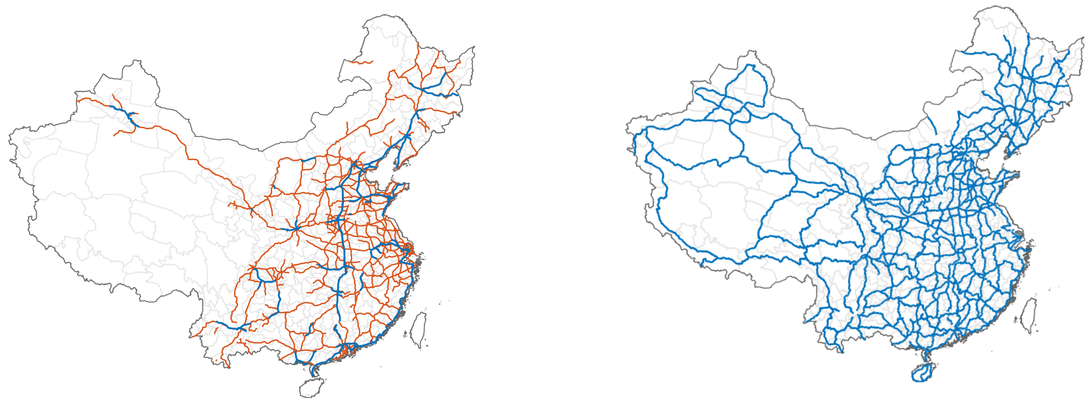
[2]See Hillberry and Hummels (2008) for a pioneering study using this survey. A predecessor of the Commodity Flow Survey was conducted in 1963, but the micro data are not yet easily accessible.

has been limited to agricultural commodities or goods identified through bar codes or by their unique producers.

This paper makes two contributions. First, we propose a methodology for estimating domestic trade costs using information contained in typical customs data. We estimate a routing model for structural parameters governing the response of exporters' port choice to the domestic transport network, exploiting the *over-time* variation stemming from the expansion of the expressway network. Second, we embed these estimates in a spatial equilibrium model with regional comparative advantage, input-output linkages, and sector heterogeneity in trade costs and use it to evaluate the return to the fifty thousand kilometers (km) of expressways built in China between 1999 and 2010. We find that the investment generates large positive net returns. Evaluation based on simpler models or an alternative approach focusing on the first order effect can lead to biased assessments.

Our empirical design takes advantage of the increasingly available customs data. Like those of many other countries, the Chinese customs data contain the city of exporters and the port from which they ship to foreign customers. The fractions of a city's exports that go through different ports reflect, among potential confounding factors, the costs of transport routes through these ports. All else equal, if an inland city A ships most of its exports via port B, then the routes passing through B likely incur lower costs than others. A direct application of this intuition to the data is subject to several sources of biases. First, the decision to export through a port might be driven by an unobserved connection with the port, which could be correlated with distance but will not respond to expressway construction. Second, the total cost along an export route consists of costs along both its domestic and international segments. If the two components are negatively correlated, which would be the case if the data are generated by exporters minimizing the *total* cost, attributing port choices entirely to domestic transport costs exaggerates their importance.

We address both concerns by exploiting changes in bilateral trade costs resulting from the rapid expressway expansion in China between 1999 and 2010. As Figure 1a shows,



(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 networks in 1999 (blue) and in 2010 (red); the right panel plots China's regular road network in 2007. Regular roads include 'national roads' and 'provincial roads'.

over this decade, the expressway network grew from a few lines in the center and the southeast coast to covering most of the country, greatly supplementing China's existing regular road network, drawn in Figure 1b.[3] Controlling for city-port, city-time, and port-time fixed effects, we estimate that each additional 100 km of road distance reduces the probability a port is chosen for shipment by 15.7%. Not controlling for city-port fixed effects doubles this estimate. This finding is robust when we exclude from sample major cities, which serve as the nodes of the expressway network, and when we use a hypothetical network that minimizes total length as an instrumental variable, so it is unlikely to be biased due to endogenous placement of expressways.

We embed the empirical design in a spatial equilibrium model consisting of Chinese cities and the rest of the world (RoW). The model includes a routing block mapping road networks into trade costs, which builds on Allen and Arkolakis (2019) but differs in two aspects: first, it allows flexible combinations of regular and expressway segments

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

in forming a route; second, it allows trade costs to be higher for heavier sectors. We use unit values of shipments from the customs data, available for a wide range of narrowly defined products, to estimate the elasticity of sectoral trade cost in the weight-to-value ratio. Our estimation implies 20% trade cost savings on expressways compared to regular roads of equivalent length and a 0.3 elasticity of trade cost in weight-to-value ratio.

Through counterfactual experiments, we find that expressways built during the decade bring 5.1% aggregate welfare gains to China. The sum of discounted gains far exceeds project investment (approximately 10% of 2010 GDP) and implies a *net* return of 150%. Restricted versions of the model without the three key elements—-regional comparative advantage, heterogeneous trade costs, and intermediate inputs—predict significantly smaller welfare gains because they infer either too little domestic trade or an incorrect distribution of shipments on the road network. If all three ingredients are omitted, the model infers welfare gains that are only 17% of the actual gains, implying a negative return on investment.

In the final section of the paper, we take advantage of the model's tractability to analytically derive the gains from transport infrastructure improvements for China. To the first order, the welfare gains are simply total savings in trade costs of goods being transported on the affected road segments, net of the savings passed on to the RoW. This result connects with the 'social savings' approach in evaluating transport projects (see, e.g. Small, 2012), which can be viewed as a first-order approximation to welfare gains for closed economies. However, despite being transparent, this approximation is inaccurate: it fails to take into account that drivers can reoptimize and switch routes when an expressway segment is built. Moreover, in evaluating large projects with multiple segments, it overlooks the potential complementarity or substitution between segments.

We find that such biases average 21% of the actual effects across the 100 busiest expressway segments in China and amount to 46% for large projects that consist of many segments. We propose a second-order correction that can be evaluated after the model has been parameterized. This term captures the rerouting of drivers as well as the interactions among segments, and reduces the average approximation errors to less than 7%. Our formula thus offers a way to evaluate large projects accurately, without having to solve for counterfactual equilibria. This could be especially useful in applications where comparisons among many large projects are needed (e.g., Fajgelbaum and Schaal, 2020).

This paper contributes to the literature on the effects of transport infrastructure projects.[4] Beyond estimates of the return to the Chinese expressways, our analysis draws general lessons. Our method 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 comparative advantage and heterogeneous trade costs likely applies to other settings as well. Finally, we characterize and demonstrate the importance of second-order effects for evaluating large projects, contributing to a growing agenda in macroeconomics that emphasizes nonlinearity (e.g., Baqaee and Farhi, 2019).

Central to our analysis is the idea that export routes contain information on domestic trade costs. We are not the first to recognize this. For example, Limao and Venables (2001) shows the importance of domestic infrastructure for exports in a cross-country setting; Coşar and Demir (2016) and Martincus et al. (2017) show that road construction increases exports with micro data; Sequeira and Djankov (2014) shows that exporters choose ports to avoid corruption of border officials, a different form of trade costs. Different from existing work, we combine export data with a routing model to infer structural parameters governing transport costs and use a rich general equilibrium model for counterfactuals.

Finally, this paper adds to the rapidly growing literature on quantitative spatial economics (see Redding and Rossi-Hansberg, 2017 for a review), particularly the strand focusing on China (Fan, 2019; Tombe and Zhu, 2019; Ma and Tang, 2019). Domestic trade

[4] The literature on infrastructure primarily uses two approaches. The first involves quantitative exercises via simulations. Studies using this approach have investigated the impacts of roads (e.g., Morten and Oliveira, 2018; Alder and Kondo, 2019; Cosar et al., 2019), railroads (Fajgelbaum and Redding, 2014; Xu, 2018), and urban transit (Tsivanidis, 2018; Severen, 2018). The second approach estimates the treatment effect of infrastructure on regional income/growth, using either heuristic or theory-based measures of treatment (see, e.g., He et al., 2020; Baum-Snow et al., 2020).

costs are central to the predictions of these studies. Most current work on China either uses railway shipments, which account for only 10% of total shipments and are available only at the provincial-pair level, a level too crude for studying transport infrastructure, or relies on regional input-output tables imputed from railway shipments (see Zhang and Qi, 2012 for the imputation procedure). Using new and more granular data, our analysis generates predictions for domestic and international trade costs for 1999 and 2010, which can serve as input into future work in this area. We also show that the model-predicted export growth in response to expressway expansion is strongly correlated with the actual growth in this period. Under suitable assumptions, model-simulated exports can serve as a *time-varying* IV for exports at the city-sector level.[5]

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

2 First Look at the Data

Our empirical investigation focuses on the long-run change between 2000 and 2010, a period of rapid expressway buildup in China. This section introduces the data and illustrates the variation that our structural estimation will exploit.

2.1 Data and the Sample

Export routing. We measure exporters' port choice using transaction-level customs data. For each transaction, we observe the address of the exporter, the value and weight (if the unit of output is kilogram) of the shipment, and the customs office from which it

[5]This IV would exploit the changes in access to foreign markets driven by expressway construction and complement the existing identification strategies in estimating the effects of export. A strand of literature exploits the variation from the reductions in the level or uncertainty of export tariffs *across industries* after the WTO accession to estimate the effects of export (see, e.g., Facchini et al., 2019; Tian, 2019). The source of variation in our model-based IV is *across regions*.

was exported. We map the addresses of exporters and customs offices to prefecture cities, treating the city of an exporter as the origin and the city of the customs office as the port.[6]

We aggregate transactions to obtain the aggregate and sectoral total shipments from each origin city to the RoW through different Chinese ports. In the baseline analysis, we follow a tradition in the international trade literature and use the value of export as a proxy for shipments. In Appendix A.8, we show that the results are robust if the weight of shipment is used instead. Given the focus on long run changes, we construct a panel with two periods corresponding to the beginning and the end of the decade.[7]

Transport network. We obtain intercity expressway maps for 1999 and 2010 from Baum-Snow et al. (2020), who digitized China's transport infrastructure from hard-copy maps. We supplement these expressway maps with a map of regular roads from the ACASIAN Data Center for 2007.[8] Since there was virtually no variation in the regular road network during this period, we treat it as time-invariant.

We calculate the distances on the road transport network between cities and ports for both 1999 and 2010. There are many feasible paths between any pair of cities. For the reduced-form analyses below, we assume that the least-cost path is always taken and that its length is the effective distance between two cities. Because paths vary in their compositions of regular roads and expressways, identifying the least-cost path requires us to take a stand on the relative cost of the two road types. To this end, we query the driving times

[6]It is possible for a shipment to be declared at the customs office in an inland city and sealed before it is shipped to the rest of the world, either directly through ground or air transport, or indirectly via a seaport. In the latter case, the city of the customs office is not the point of goods' exit from China. To rule out this scenario, our specifications focus on customs locations that are seaports (see the appendix for the list of these customs). The exports processed through these customs offices account for 90% of the total exports of China and for 82% of the total exports from all non-port cities.

[7]The beginning period's data are the average across 2000 and 2001; the end period's data are the average between 2010 and 2011. We do not have access to customs data for 1999.

[8]Regular roads in the ACASIAN database include 'national roads' and 'provincial roads', which are paved, non-enclosed, non-divided roads, usually with two or four lanes. Baum-Snow et al. (2020) also provides separate maps for 'general highways', which are of lower grade than 'high-grade highways', or expressways. The definition of 'general highways' is broad and generally includes 'national roads', 'provincial roads', and 'county roads'. Because a 'county road' is of much lower quality than a 'national road' or a 'provincial road', and because most intercity transports rely on the latter two, we choose not to use Baum-Snow et al. (2020) to measure the regular road network.

between a random set of 2000 city pairs along expressways and regular roads separately on Baidu Maps, a Chinese search engine, and compare the average expected travel time of the two trips. Among these queries, the average speed on regular roads is approximately 55% of that on expressways, so we set the cost of traveling 1 km on expressways to be equivalent to that of traveling 0.5 km on regular roads. We then use the Dijkstra's algorithm to find the least-cost path for each city pair.

Let $dist_{od}^{t}$ be the regular-road-equivalent length of the shortest path between o and d at time t, and $dist_{od}^{L,t}$ and $dist_{od}^{H,t}$ be the lengths of regular-road and expressway segments on this path, respectively. Then, $dist_{od}^{t} = dist_{od}^{L,t} + 0.5 \times dist_{od}^{H,t}$.

2.2 Descriptive Statistics

Table 1 reports the summary statistics of key variables. Each observation is an export route—a pair of an exporting city and a seaport. In the initial period, the average export volume per route was $165 million. A decade later, this amount increased to approximately a billion, mirroring the five-fold growth in China's exports over this period.

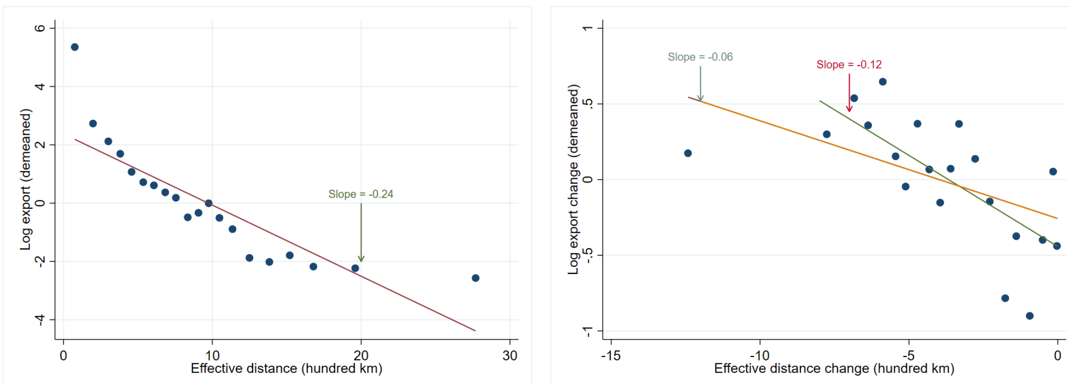
Accompanying the dramatic export growth were improvements in cities' access to ports due to expressway construction. The average total length of export routes decreased by 15% from 2,040 km to 1,724 km—with a denser expressway network, exporters in the hinterland now no longer needed to take detours for express access to ports. The length of expressway segments in these routes increased from 1,293 to 1,688 km. Factoring that expressways are less costly than regular roads on a per-km basis, the growing composition of expressways further reduced the effective, or *regular-equivalent*, length of routes, which decreased by 37% from 1,393 to 880 km.

Binned scatter plots in Figure 2 illustrate the relationship between export routes and domestic transport costs. The left panel plots the cross-sectional relationship between the value of exports via a route and the effective length of that route, pooling data from both periods. There is a strong negative correlation between the two variables, with a slope of -0.24. If we were to interpret this relationship as causal, this slope would imply that a

Table 1: Descriptive Statistics

	2000-2001		2010-2011	
Route-level variables	mean	std	mean	std
Exports	165	2051	990	12130
Total length	20.40	11.60	17.24	10.45
Length of expressway segments	12.93	7.71	16.88	9.72
Effective (regular-equivalent) length	13.93	9.00	8.80	5.77

Notes: This table reports summary statistics at the level of export routes. An export route is defined as a city-port pair. Exports are measured in million USD; distance is measured in units of 100 km.



(a) Cross-sectional variation

(b) Over-time variation

Figure 2: Exports and Route Length

Notes: The figures are binned scatter-plots that show the relationship between the logarithm of shipment value on an export route and the regular-equivalent length of that route. The left panel plots the cross-sectional relationship; the right panel plots the over-time relationship. In the right panel, the slope of the fitted line is -0.06 if all observations are included, and is -0.12 if the leftmost 5% are excluded.

100 km increase in the effective distance reduces the shipment value by 24%.

Of course, this negative correlation could be driven by factors other than the route distance between the city and the port. For example, regions closer to each other might have stronger cultural and ethnic ties, and are also more likely to be connected through common business (e.g., export intermediary and logistic) networks. All of these factors could contribute to higher shipment values. While some of these connections might react to expressway expansions, others were formed historically and unamenable to road construction. The cross-sectional relationship thus overestimates how the road network affects shipment values.[9] Figure 2b plots the relationship between the changes in the

[9] In the setting of international trade, Feyrer (2009) makes a similar point. Using the variation from the

two variables from 2000 to 2010. As the two periods are a decade apart, the changes likely capture most of the long-run effect of road construction. The best linear fit to the changes has a slope of -0.06, much lower than the cross-sectional slope.[10] This difference demonstrates that a significant part of the cross-sectional negative relationship is likely due to factors that do not respond to the transport network.

While Figure 2b is informative about the variation in the data, it does not show a causal relationship. City- and port- specific shocks could confound the slope estimate; roads connecting certain pairs of cities could have been built with the goal of increasing exports. We perform regression analyses to address these concerns.

2.3 Empirical Specification

We estimate variants of the following specification, which will be micro founded by the structural model developed in later sections:

$$\ln(v^t_{(o,RoW),d}) = \beta_{od} + \beta^t_o + \tilde{\beta}^t_d + \gamma_1 dist^t_{od} + \epsilon^t_{od}. \tag{1}$$

The dependent variable, $v^t_{(o,RoW),d}$, denotes exports from city o to the RoW through port city d in period t. Variables β_{od}, β^t_o, $\tilde{\beta}^t_d$ are city-port pair, city-time, and port-time fixed effects, respectively. $dist^t_{od}$ is the effective distance between o and d along the least-cost-path on the period-t network. In some specifications, we replace it with $dist^{H,t}_{od}$ and $dist^{L,t}_{od}$ to estimate the separate effects of expressway and regular road segments.

As alluded to before, the OLS estimate of specification (1) is subject to a few endogeneity concerns. First, expressways might have been built to promote the economic growth of specific ports or regions. We control for city-time and port-time fixed effects, which would capture export growth driven by city- or port-specific shocks that also de-

closure and reopening of the Suez Canal, he shows that the estimated distance elasticity is half of that based on the cross-sectional variation only.

[10] If pairs for which the effective distance decreased by more than 1,000 km are excluded, the slope becomes -0.12, still significantly smaller than that implied by cross-sectional data. Cities with a more than 1000 km decrease in the effective distance to seaports are all in the northwestern Xinjiang Autonomous Region, which exports more via ground transport following the Silk Road. Our regression analyses will have city fixed effects, so the estimate will not be sensitive to whether those cities are included.

Table 2: Expressway and Routing of Export Shipments

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Effective Route Length and Exports						By Type of Road		
	OLS				IV Reduced Form	2SLS	OLS	IV Reduced Form	2SLS
$dist^t_{od}$	-0.341***	-0.384***	-0.157***	-0.174***		-0.170***			
	(0.011)	(0.011)	(0.037)	(0.045)		(0.058)			
-on express							-0.088**		-0.162**
							(0.038)		(0.068)
-on regular							-0.174***		-0.198***
							(0.045)		(0.063)
IV $dist^t_{od}$					-0.198***				
					(0.068)				
-IV express								-0.147***	
								(0.053)	
-IV regular								-0.230***	
								(0.073)	
Fixed Effects	o, d, t	ot, dt	od, ot, dt	od, ot, dt	od, ot, dt	od, ot, dt	od, ot, dt	od, ot, dt	od, ot, dt
Exclude Major Cities				yes	yes	yes	yes	yes	yes
Observations	3668	3660	2838	2068	2038	2038	2068	2038	2038
R^2	0.646	0.709	0.906	0.897	0.897	0.020	0.897	0.897	0.015
First Stage K-P F stat						1400.799			170.204

Notes: This table reports the regressions of export shipment through a port on the distance between the city and the port. The outcome variable is the logarithm of the total value of goods exported by city o through port d to the RoW. In Columns 1-4, the explanatory variable is the regular-equivalent length of the shortest path between city o and port d. Columns 5 and 6 show the reduced-form and 2SLS estimates using the minimum-spanning network IV. Column 7 separates the total length of the shortest path into those of expressways and regular roads. Columns 8 and 9 show the reduced-form and 2SLS estimates using the minimum-spanning network IV.

Standard errors are clustered at the city-port level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

termined expressway construction. Second, and perhaps more importantly, cities closer to each other likely have lower barriers of other sorts, such as information frictions and home biases, which could increase export volume for reasons unrelated to transport infrastructure. Through city-port fixed effects, we control for all time-invariant unobserved heterogeneity across pairs of cities. The identification thus comes from over-time changes in effective distance resulting from the *expansion* of the expressway network.[11]

2.4 Expressway Construction and the Route Choice of Exporters

Table 2 reports the baseline results. The specification in the first column includes only city, port, and time fixed effects, so the coefficient is identified off mostly cross-sectional variation. The point estimate is -0.341. The second column further includes city-time and port-time fixed effects, which leads to a modest increase in the estimated coefficient. In

[11] To the extent that some of non-transport barriers, such as information friction, also respond to additions to the transport network, such responses should and will be picked up by our estimate, which uses a specification that focuses on long-run effects. What we would like to exclude through the addition of bilateral fixed effects is the components that do not respond to transport infrastructure.

the third column, we include city-port fixed effects to focus on the long-run changes. The point estimate shrinks by 60%, in accordance with the patterns documented in Figure 2. The coefficient implies that each additional 100 km of effective distance reduce the exports through a port by 15.7%.

Expressway network endogeneity concerns. A remaining concern is that new expressways might have been built to connect specific *pairs* of cities with *growing* economic ties, which might be correlated with export growth. We use two strategies to alleviate this concern. The first is to exclude each origin city o that is either a provincial capital or had more than 5 million registered residents.[12] As discussed in Banerjee et al. (2020), transport networks in China were largely designed to link major cities. With these cities excluded, our estimation exploits the increase in port access for the remaining, smaller cities, which gained access because they were between major cities to be connected. Column 4 reports the results of this specification, which is also our preferred specification. The point estimate is -0.174, similar to when major cities are included.[13]

Second, in addition to excluding major cities, we adopt an IV strategy based on Faber (2014): to use an 'exogenous' hypothetical expressway network as an instrument for the actual network. Specifically, using a minimum-spanning tree algorithm, we first generate the expressway network with the smallest total length that connects all major cities on the actual network by 2010. This network would be optimal *if* the goal were to minimize the total length of expressways while still connecting the same set of major cities. As Appendix Figure A.3 shows, this hypothetical network spans the same regions as the actual one, but has no curvature and is far sparser. We use it instead of the actual expressway network in calculating the shortest-path distance between cities for 2010, denoted by $dist_{od}^{IV,2010}$, which then serves as an instrumental variable for $dist_{od}^{2010}$.[14] The identifi-

[12]According to the 2000 population census, 55 cities are classified as major cities.

[13]Using a similar specification, Coşar and Demir (2016) finds that upgrading a carriageway to an expressway reduces trade costs over a trip of 820 km by about 27% (p. 222). Converting our baseline estimate of 0.174 into a similar object would imply a cost saving of 18% over the same distance, so our estimate is in line with the existing evidence. Appendix A.6 presents the underlying calculation.

[14]We use the distance along the actual 1999 network as an IV for itself. The IV is time-varying, and

cation assumption is that nonmajor cities experienced an improvement in access to ports only because they were close to the hypothetical network that connected the major cities. Column 5 reports the reduced-form results obtained using this instrument, showing a coefficient of -0.198. Column 6 presents the two-stage-least-squares (2SLS) estimate. The high first-stage F-statistic indicates relevance. The regression coefficient is close to the OLS estimate and is approximately half the size of the coefficient in Column 1.

The stability of results across Columns 3-6 shows that our estimate is unlikely to be biased due to an endogenous placement of expressways. The over-time estimates being consistently smaller than cross-sectional ones has implications for evaluating the impacts of transport projects. In typical domestic trade models, researchers could have interpreted shipments from origin cities to ports as trade flows. Specification (1) then corresponds to a gravity regression, with γ_1 being the product of trade elasticity and the distance semi-elasticity of trade cost. Given the trade elasticity, γ_1 maps one-to-one into the effect of distance on trade costs. Using cross-sectional variation only thus overestimates this effect by 100%, which in turn would exaggerate the welfare gains from the improvements of domestic transport infrastructure.[15]

Relative costs of expressways and regular roads. Thus far, we have focused on estimating the coefficient for regular-equivalent distance ($dist_{od}^{t}$). To further investigate the separate effects of the two types of roads, Column 7 splits $dist_{od}^{t}$ into two components: distance along expressway segments and that along regular road segments. The coefficient is -0.17 for regular roads and -0.09 for expressways—the former are most costly than the latter, as one would expect. Columns 8 and 9 report the IV reduced-form and 2SLS estimates. Both approaches result in a larger coefficient for regular road distance,

---

we can thus control for city pair fixed effects. Our specification effectively uses the *differences* in bilateral distance between the actual 1999 network and the hypothetical 2010 network as an instrumental variable for the actual *changes* in distance over this period.

[15]The model developed in the next section explicitly incorporates the port choice of exporters. According to the model, the estimated coefficient γ_1 is the product of two structural parameters: the elasticity of substitution between ports (instead of the trade elasticity), and the distance semi-elasticity of domestic trade costs. However, the same point remains valid: for a given port elasticity, a larger γ_1 leads to larger inferred welfare impacts.

although the distinction becomes smaller for the 2SLS estimate.

Summary of additional results and robustness. In Appendix A, we report a number of additional results. First, we examine the channel through which the responses in shipment occur. We find that the responses are driven by the growth of exports at the city-port level, rather than the rerouting of a city's exports among different ports. Second, we visualize the bilateral variation exploited on the map and show that our regressions use both the variation between broad geographic regions and the variation within a region. Third, we conduct a host of robustness analyses, which show that all results hold when we use sectoral level data and control for the full set of sector-related fixed effects, when the specification is PPML, an often used alternative to linear models in estimating trade flows, and when we measure shipments by weight instead of value.

In summary, the reduced-form results provide robust evidence that the port choice of exporters responds to domestic infrastructure and that it is crucial to estimate such responses using the over-time variation. In the rest of this paper, we develop a structural model of exporter route choice and use it to (1) estimate the key structural parameters of export and domestic trade routing and (2) conduct counterfactual experiments.

3 Route Choice on the Transport Network

We start by describing the routing block of our model, which extends that of Allen and Arkolakis (2019) to accommodate two coexisting networks.

3.1 From Networks to Road Costs

Consider a four-region economy, illustrated in Figure 3. Nodes (o, l, k, d) represent cities connected by links that represent a transport network. Curly dashed lines represent regular roads, and straight solid lines expressways. We use ι^x_{kd}, $x \in \{H, L\}$ to denote the travel cost for edge $k \to d$, where H and L indicate expressways and regular roads, respectively. Costs along any edges are greater than 1 and are symmetric: $\iota^x_{kd} \geq 1$, $\iota^x_{kd} =$

$\iota^x_{dk}, \forall k \neq d, x \in \{H, L\}$. We assume that for any adjacent cities k and l,

$$\iota^x_{kl} = \exp(\kappa^x \cdot dist^x_{kl}), x \in \{H, L\}, \tag{2}$$

in which $dist^x_{kl}$ is the length of the edge of road type x connecting k and l, and κ^x is the corresponding distance semi-elasticity.

A path, or a route, is a set of connected edges that links an origin to a destination; the cost of traveling on a path is the product of costs of all edges forming that path. For example, $o \xrightarrow{L} k \xrightarrow{H} d$ is a path from o to d with its first leg on a regular road and the second leg on an expressway; the cost along this path is $\iota^L_{ok} \cdot \iota^H_{kd}$.

Truckers going from o to d face multiple options. There are two direct paths, by an expressway or a regular road, costing ι^H_{od} or ι^L_{od} each. Truckers derive from each path an idiosyncratic disutility ν drawn independently from the Fréchet distribution with a dispersion parameter θ. The *effective* travel cost along a path is the product of the *fundamental* cost and a path-specific realization of ν. For example, the effective cost of $o \xrightarrow{H} d$ is $\iota^H_{od} \cdot \nu$.

If the two direct paths are the only options between o and d, the Fréchet assumption implies that the expected travel cost across all possible paths r is:

$$\tau_{od,1} \equiv \mathbb{E}[\min_{r \in \{o \xrightarrow{L} d,\ o \xrightarrow{H} d\}} \iota_r \cdot \nu(r)] = \gamma_\theta \big([\iota^H_{od}]^{-\theta} + [\iota^L_{od}]^{-\theta}\big)^{-\frac{1}{\theta}}, \tag{3}$$

where $\gamma_\theta \equiv \Gamma(\frac{\theta-1}{\theta})$ is a constant, and subscript '1' in $\tau_{od,1}$ denotes that the choice is among paths with one edge.

Truckers can take detours. In the above example, there are three two-edge paths from o to d: $o \xrightarrow{L} k \xrightarrow{H} d$, $o \xrightarrow{L} k \xrightarrow{L} d$, and $o \xrightarrow{L} l \xrightarrow{H} d$. If constrained to choose among paths with two or fewer edges, the expected trade cost is:

$$\tau_{od,2} = \gamma_\theta \big([\iota^H_{od}]^{-\theta} + [\iota^L_{od}]^{-\theta} + [\iota^L_{ok}\iota^H_{kd}]^{-\theta} + [\iota^L_{ok}\iota^L_{kd}]^{-\theta} + [\iota^L_{ol}\iota^H_{ld}]^{-\theta}\big)^{-\frac{1}{\theta}}. \tag{4}$$

We derive a matrix representation for the expected travel costs. Let $\mathbb{L}$ and $\mathbb{H}$ be the adjacent matrices corresponding to regular and expressway networks, respectively:

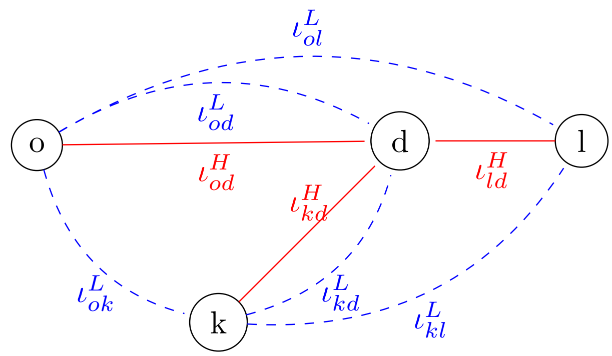


Figure 3: Routing on a Network: a Four-Region Example

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

The nonzero elements in $\mathbb{L}$ and $\mathbb{H}$ are the -θth power of the cost between two adjacent nodes in these networks. Zeros indicate that two cities are not directly connected by an edge.[16] Let $\mathbb{A}$ be the sum of the two matrices, $\mathbb{A} \equiv \mathbb{H} + \mathbb{L}$, and let $[\mathbb{X}_{(o,d)}]$ denote the od-th element of matrix $\mathbb{X}$. Equation (3) becomes:

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

Furthermore, define $\mathbb{A}^2 \equiv \mathbb{A} \cdot \mathbb{A}$. The od-th element of $\mathbb{A}^2$ is

$$[\mathbb{A}^2_{(o,d)}] = \sum_{x,x' \in \{H,L\}} \sum_{k} (\iota_{ok}^{x} \cdot \iota_{kd}^{x'})^{-\theta} = (\iota_{ok}^{L}\iota_{kd}^{H})^{-\theta} + (\iota_{ok}^{L}\iota_{kd}^{L})^{-\theta} + (\iota_{ol}^{L}\iota_{ld}^{H})^{-\theta},$$

i.e., the sum of the $-\theta$th power of the cost across all two-edge paths. The matrix representation of Equation (4) is:

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

[16]Equivalently, they may be regarded as connected by an edge with an infinite transport cost. We assume that the diagonal elements of the adjacency matrix are zero. Correspondingly, throughout the rest of this paper, we normalize the iceberg cost of trading within a city to one.

In principle, truckers can take more costly detours (e.g., $o \xrightarrow{L} k \xrightarrow{L} l \xrightarrow{H} d$) or even revisit a stop (e.g., $o \xrightarrow{L} d \xrightarrow{H} l \xrightarrow{H} d$).[17] For larger networks, as truckers freely take multiple detours combining expressway and regular road segments, enumerating all possible paths becomes a complex combinatorial problem. In Appendix B.1, we show by mathematical induction that the sum across all paths between o and d with *exactly* N edges is $[\mathbb{A}^N_{(o,d)}]$, which implies that the expected cost across all possible paths is:

$$\tau_{od} \equiv \lim_{N\to\infty} \tau_{od,N} = \gamma_\theta \Big(\sum_{n=1}^{\infty} [\mathbb{A}^n_{(o,d)}] \Big)^{-\frac{1}{\theta}} = \gamma_\theta \big([\mathbb{B}_{(o,d)}] \big)^{-\frac{1}{\theta}}, \text{ for } o \neq d, \tag{5}$$

where $\mathbb{B} \equiv (\mathbb{I} - \mathbb{A})^{-1}$.[18] Equation (5) expresses transport costs as a differentiable function of the structure of the transport network. This feature will enable us to characterize the first and second order welfare effects of expressway projects.

3.2 From Road costs to Trade Costs

The routing block gives the expected costs for truckers between domestic locations. We now build on it to obtain sector-specific trade costs for domestic and international shipments.

Sector heterogeneity in trade costs. Trade costs have an iceberg form and vary across sectors depending on the 'heaviness' of a sector, measured by its weight-to-value ratio, h_i. Let $\iota^{i,x}_{kl}$ be the edge cost of sector i between k and l on road type $x \in \{H, L\}$, specified as:

$$\iota^{i,x}_{kl} = \left(\frac{h_i}{h_0} \right)^{\mu} \cdot \iota^{x}_{kl},$$

where ι^x_{kl} is defined in Equation (2), h_0 is a scaler that shifts the overall level of trade costs, and μ is an elasticity that governs how trade costs vary with the weight of shipments.[19]

[17] As θ increases, the probability of routes with repeated stops being chosen approaches zero. We characterize this probability in Appendix B.2 and shows that given our structural estimates, routes with repeated stops are chosen with negligible probabilities.

[18] Suppose there is only one expressway network, then $\mathbb{A} = \mathbb{H}$, in which case our model collapses to the one-network model of Allen and Arkolakis (2019). A sufficient condition for $(\mathbb{I} - \mathbb{A})$ to be invertible is that the spectral radius of $\mathbb{A}$ is less than one (Allen and Arkolakis, 2019). This will be the case if the road network adjacency matrix is sparse and the routing elasticity θ is large; these conditions hold for our empirical estimate.

[19] An alternative is to assume that sectors differ in the distance-cost semi elasticity, κ^H_i and κ^L_i. However, we did not find support for this specification in the data.

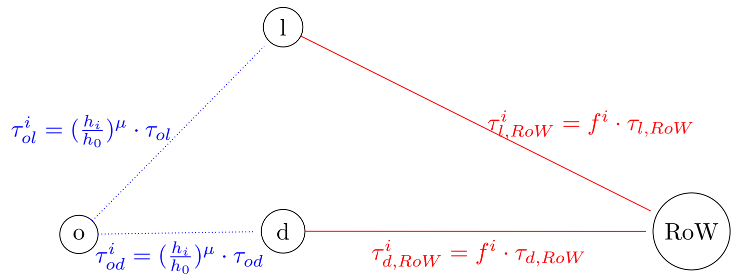


Figure 4: Port Choice of Exporters

Note: The diagram illustrates the choice of a port through which to ship to the RoW. l and d are ports. Dashed lines ol and od indicate that city o might be connected to ports l and d only indirectly via road networks.

In the limit case of $\mu = 1$, this specification implies that trade costs increase linearly in the weight of a shipment.[20] More generally, the relationship between trade costs and weights needs not be linear—using data on U.S. imports, Hummels (2007) estimates that the elasticity of the ad-valorem shipping cost in the weight-to-value ratio is 0.4-0.5 for both seaborne and airborne shipments. We allow for a nonlinear relationship and will estimate μ.

The specification of ι^i_{kl} also implies that the trade cost between o and d in sector i is:

$$\tau^i_{od} \equiv \lim_{N\to\infty} \tau^i_{od,N} = \gamma_\theta \Big(\sum_{n=1}^{\infty} (\frac{h_i}{h_0})^{-\mu\theta} [\mathbb{A}^n_{(o,d)}] \Big)^{-\frac{1}{\theta}} = \left(\frac{h_i}{h_0}\right)^\mu \cdot \tau_{od}, \tag{6}$$

i.e., τ^i_{od} is $(\frac{h_i}{h_0})^\mu$ multiplied by τ_{od} characterized by Equation (5).

Decisions of exporters. To use the customs data for estimation, we now embed the domestic routing problem into a port choice problem of exporters. Consider in an economy represented by Figure 4, an exporter from city o looks to send a truckload of merchandise of sector i to foreign buyers. The total export cost has two components: a domestic component between o and one of the nation's ports l or d, denoted by τ^i_{ok}, $k \in \{l, d\}$, and an international component $\tau^i_{k,RoW}$, which is the product of a sector-specific cost, f^i, and the overall access of port $k \in \{l, d\}$ to the RoW, $\tau_{k,RoW}$.

[20] To see this, 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 $\tilde{h}_0$ tons, the cost of shipping for this batch of goods on $k \to l$ is simply $\frac{yh_i}{\tilde{h}_0}\iota^x_{kl}$, where $\frac{yh_i}{\tilde{h}_0}$ is the number of trucks needed. Redefining $\frac{y}{\tilde{h}_0} = \frac{1}{h_0}$ gives $(\frac{h_i}{h_0})\iota^x_{kl}$ as the cost.

The exporter first choose the port from which to ship the goods, taking the expected domestic trade cost as given. Each seller receives a *port-specific* export cost shock, denoted by $(\nu_F(k))_{k\in\{l,d\}}$, drawn from a Fréchet distribution. This shock enters trade costs multiplicatively, so the international shipment cost from l to the RoW, for example, is $\tau^i_{l,RoW}\cdot\nu_F(l)$. Because the source of idiosyncratic shocks for international shipments might differ from that of shocks for domestic shipments, we allow the dispersion parameter of $\nu_F(k)$, θ_F, to differ from θ.[21] The seller chooses $\min\{\tau^i_{ol}\tau_{l,RoW}\cdot\nu_F(l),\ \tau^i_{od}\tau_{d,RoW}\cdot\nu_F(d)\}$. Suppose that port d is chosen, then, the seller randomly meets with a trucker, who will charge the expected cost for the domestic leg and finds the least-cost route from o to d given his own taste shocks. The expected export cost faced by an exporter is thus:

$$\tau^i_{o,RoW}=\Gamma(\frac{\theta_F-1}{\theta_F})[\sum_{k\in\text{ports}}(\tau^i_{ok}\cdot\tau^i_{k,RoW})^{-\theta_F}]^{-\frac{1}{\theta_F}}. \tag{7}$$

The probability that the export from city o is shipped via port d is:

$$\pi^i_{(o,RoW),d}=\frac{(\tau^i_{od}\cdot\tau^i_{d,RoW})^{-\theta_F}}{\sum_{k\in\text{ports}}(\tau^i_{ok}\cdot\tau^i_{k,RoW})^{-\theta_F}}.$$

This equation illustrates how the export data identify domestic trade costs. All else equal, if port d is better connected to city o through the domestic transport network (lower τ^i_{od}), more shipments from city o will go through d. Noting that the sector shifters enter the choice probability multiplicatively, we obtain

$$\pi^i_{(o,RoW),d}=\pi_{(o,RoW),d}=\frac{(\tau_{od}\cdot\tau_{d,RoW})^{-\theta_F}}{\sum_{k\in\text{ports}}(\tau_{ok}\cdot\tau_{k,RoW})^{-\theta_F}}. \tag{8}$$

That is, sector shifters do not affect the patterns of port choice.[22] To identify the importance of sector heterogeneity, we will use the price information in the customs data.

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

[22]Sector heterogeneity affects the level of trade costs and hence the level of trade across sectors. However, among inter-regional shipments, it does not affect the probability of a port being chosen.

4 Estimating Domestic Trade Costs Using Customs Data

4.1 Port Choice of Exporters

Our structural estimation proceeds in two steps. In the first, we estimate three composite parameters, $\kappa^H\theta$, $\kappa^L\theta$, and $\frac{\theta_F}{\theta}$ from the port choice of exporters. To this end, we introduce time superscript t into Equation (8), substitute Equation (5) for τ_{od}, and apply the logarithmic transformation to obtain

$$\log(\pi^t_{(o,RoW),d}) = c + \frac{\theta_F}{\theta}\log\left(\tilde{\mathbb{B}}^t(\kappa^H\theta, \kappa^L\theta)_{(o,d)}\right) - \theta_F \log(\tau^t_{d,RoW}) - \log(\sum_{k\in\text{ports}}(\tau^t_{ok}\cdot\tau^t_{k,RoW})^{-\theta_F}),$$

where c is a constant. $\log(\tau^t_{d,RoW})$ and $\log(\sum_{k\in\text{ports}}(\tau^t_{ok}\cdot\tau^t_{k,RoW})^{-\theta_F})$ on the right side of the equation represent the export cost shifter specific to port d and the access of city o to the RoW through all ports, respectively. $[\tilde{\mathbb{B}}^t(\kappa^H\theta, \kappa^L\theta)_{(o,d)}]$ is the od-th element of matrix $\mathbb{B}^t$. Given the road network in period t, $\mathbb{B}^t$ depends on κ^H and κ^L only through their products with θ.[23] We write it as $\tilde{\mathbb{B}}^t(\kappa^H\theta, \kappa^L\theta)$ to highlight this dependence.

The above equation gives the model-predicted shares of export shipments via different ports. To estimate the routing parameters, we minimize the deviations between the model and the data using nonlinear least squares, interpreting these deviations as measurement errors. Let $\log(\hat{\pi}^t_{(o,RoW),d})$ be the share of export shipment from city o via port d in the data. We solve the following:

$$\min_{\frac{\theta_F}{\theta},\ \kappa^H\theta,\ \kappa^L\theta,\ fe} \quad \sum_{o,d,t}\left[\frac{\theta_F}{\theta}\log\left(\tilde{\mathbb{B}}^t(\kappa^H\theta, \kappa^L\theta)_{(o,d)}\right) + fe - \log(\hat{\pi}^t_{(o,RoW),d})\right]^2, \tag{9}$$

in which fe is a set of fixed effects. To account for port-specific cost shifters, $\log(\tau^t_{d,RoW})$, and city-specific access to the RoW, $\log(\sum_{k\in\text{ports}}(\tau^t_{ok}\cdot\tau^t_{k,RoW})^{-\theta_F})$, both unobserved, we include city-time and port-time fixed effects. Motivated by the reduced-form findings, we also control for city-port fixed effects, so the source of variation is the change in $[\tilde{\mathbb{B}}^t(\kappa^H\theta, \kappa^L\theta)_{(o,d)}]$ due to expressway network expansion.

Although (9) is a high-dimensional optimization problem, note that only $\kappa^H\theta$ and $\kappa^L\theta$

[23] Recall that $[\mathbb{A}_{(o,d)}] = [{\iota^H_{od}}^{-\theta} + {\iota^L_{od}}^{-\theta}] = [\exp(-\theta\kappa^H\cdot dist^{H,t}_{od}) + \exp(-\theta\kappa^L\cdot dist^{L,t}_{od})]$, i.e., given the network structure, $\mathbb{A}$ is determined solely by $\kappa^H\theta$ and $\kappa^L\theta$. This is also true for $\mathbb{B}$, the Leontief inverse of $\mathbb{A}$.

Table 3: Estimates of the Routing Model

	Value	s.e.	Median	p10	p90
Panel A: routing data					
$\kappa^L\theta$	4.68	1.90	4.67	4.26	6.18
$\kappa^H\theta$	3.78	1.08	3.77	3.36	4.83
θ_F/θ	0.06	0.03	0.05	0.03	0.09
Panel B: price data					
θ	111.52	35.41	111.05	103.49	127.31
μ	0.29	0.04	0.29	0.23	0.35

Notes: Inference for $\kappa^L\theta$, $\kappa^H\theta$, θ_F/θ, and θ is based on 200 cluster-bootstrapped samples, each constructed by sampling with replacement at the city level. Inference for μ is based on its asymptotic distribution after estimating Equation (10) with OLS.

enter the objective function nonlinearly via $\tilde{\mathbb{B}}^t(\kappa^H\theta, \kappa^L\theta)$; hence, the original problem can be cast into a nested one. In the inner nest, given the values of $\kappa^H\theta$ and $\kappa^L\theta$, the problem is linear in $\frac{\theta_F}{\theta}$ and the fixed effects, and can be estimated using the OLS. In the outer nest, we search over the space of $\kappa^H\theta$ and $\kappa^L\theta$ to minimize the residual mean square errors of the OLS estimated in the inner nest. This approach also makes it possible to estimate the equation hundreds of times in bootstrapping.

Appendix C.1 discusses in detail how the three composite parameters are identified and describes the inference procedures. Panel A of Table 3 reports their point estimates and distribution statistics. Although parameters are not separately identified by port choices alone, their relative values are identified. Two observations stand out. First, $\frac{\kappa_H}{\kappa_L} \approx 0.8$, i.e., expressways save 20% of shipping costs relative to regular roads. Second, θ is an order of magnitude larger than θ_F. This appears reasonable given that port choices likely depend on export intermediaries used, which makes ports less substitutable than routes.

4.2 Price Regressions

The second step of our structural estimation uses price data to identify θ—which would separate the composite parameters estimated in the first step—and μ. Consider a firm in sector i from an interior city o exporting to the RoW via port d. Let the factory-gate price of the good be p_o^i. Assuming a perfect pass-through (the trade model developed in the next section will satisfy this assumption), the average free-on-board price at port d across all route-specific draws is given by:

$$p^i_{(o,RoW),d} = p^i_o \cdot \tau^i_{od} = p^i_o \cdot \left(\frac{h_i}{h_0}\right)^{\mu} \cdot \gamma_\theta \cdot [\tilde{\mathbb{B}}(\kappa^H\theta, \kappa^L\theta)_{(o,d)}]^{-\frac{1}{\theta}}$$
$$\Longrightarrow \log\left(\frac{p^i_{(o,RoW),d}}{p^i_o}\right) = constant + \mu \log(h_i) - \frac{1}{\theta}\log\left(\tilde{\mathbb{B}}(\kappa^H\theta, \kappa^L\theta)_{(o,d)}\right). \tag{10}$$

Equation (10) shows that the variation in price ratios across sectors with different 'weight-to-value' ratios identifies μ; once $\kappa^H\theta$ and $\kappa^L\theta$ have been estimated, the variation in $\tilde{\mathbb{B}}(\kappa^H\theta, \kappa^L\theta)_{(o,d)}$ across pairs of cities due to the structure of the road networks identifies θ.

We use unit values of exported goods from the transaction-level customs data to construct price ratios. Without observing the factory-gate price of each transaction, we restrict the sample to transactions with the origin city o being a port itself. For the goods produced in such city o, their average price when they are exported directly from o, $[p^i_{(o,RoW),o}]$, is then a theory-consistent measure of the factory-gate price.

The validity of this approach rests on the assumption that goods shipped directly from o to the RoW and goods shipped indirectly through another port d are comparable. For such an assumption to be valid, we use rich information from the customs data and define each product to be a combination of a city, an HS-8 category, and a destination country. We calculate the average price of direct export transactions from city o for each of these products to obtain $p^i_{(o,RoW),o}$. We then construct the dependent variable as a ratio of the price of the same product exported via another port d and $p^i_{(o,RoW),o}$.

The narrow definition of a product addresses the leading concerns in interpreting price ratios as trade costs. First, firms both export higher-quality goods and charge higher markups on these goods for destination countries with higher incomes (Fan et al., 2015). Second, cities with a more skilled workforce tend to produce better products (Dingel, 2016). Conditioning on the same destination market and origin city avoids these two sources of bias. To further alleviate these concerns, our empirical specifications absorb the remaining systematic differences in either quality or markup across cities and ports through fixed effects; we also show that similar results are obtained if we focus on nondifferentiated products, as classified in Rauch (1999), for which 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 estimate θ from over-time variation; thus, we focus on cross-sectional regressions using the end-of-period data only.

Since μ and θ are identified of different variations, we estimate them separately, which allows for more flexible controls. Specifically, we identify μ from comparisons among the same pairs of cities o and d, whether heavier goods have larger price gaps. Our most demanding specification controls for firm-port-destination country-HS2 category fixed effects and identifies μ using variation in h_i across HS4 categories. On the other hand, identification of θ uses how the price gap increases in the effective distance between o and d. We control for city-destination country-HS8 and port-destination country-HS8 fixed effects, and use the IV constructed from the minimum-spanning expressway network.

Appendix C.1 provides additional discussion of identification of μ and θ and demonstrates the robustness of the estimates to different sample restrictions and controls. Panel B of Table 3 reports estimates from the preferred specifications. We estimate that $\mu = 0.29$, which means that a 1% increase in the weight-to-value ratio increases the ad-valorem shipping cost by around 0.3%. This estimate is at the lower end of the estimate of Hummels (2007) in the setting of international shipping costs (0.4-0.5). The literature does not offer much guidance on this elasticity for domestic shipments, but the freight costs of domestic shipments documented in the literature are usually denoted linearly in weight (Redding and Turner, 2015), which translates into an elasticity of 1. To be conservative on the role of sector heterogeneity, we use 0.29 as the baseline and an elasticity of 1 for sensitivity analyses.

We obtain a point estimate of $\theta = 111.5$, implying that different routes leading to the same port are highly substitutable. Plugging this into Panel A of Table 4 gives point estimates of $\kappa^H = 0.034$ and $\kappa^L = 0.042$, which mean that additional 100 km on expressways and regular roads increases trade costs by 3.4% and 4.2%, respectively.

5 The Full Model

We embed the routing decision into a spatial equilibrium model, with costly trade and input-output linkages (Caliendo and Parro, 2015). The model will be used to perform general equilibrium counterfactual experiments, and will pin down the level and distribution of shipment flows, which, as we show in Section 7, enable a second-order approximation of the welfare gains from expressway projects.

5.1 Spatial Equilibrium Model

Environment. There are N regions, denoted by o or d, representing Chinese prefecture cities (CHN) and the RoW. There are S sectors, denoted by i or j. Domestic consumers are freely mobile across cities, and consume land and a basket of sectoral final goods. The number of consumers of the RoW is fixed. Sectoral final goods are non-tradable and aggregated from tradable intermediate goods produced by different locations. Land is in fixed supply. All markets are perfectly competitive.

Consumers. Consumers in region d maximize the following utility:

$$U_d = B_d [H_d]^{\alpha^0} \cdot \prod_{i=1}^{S} [C_d^i]^{\alpha^i},$$

where C_d^i is the consumption of final goods in sector i, with price denoted by P_d^i. H_d is the consumption of land, with price denoted by R_d. B_d is the amenity of region d. α^i are shares of land and sectoral final goods: $\sum_{i=0}^{S} \alpha^i = 1$. This preference gives an indirect utility of $U_d = B_d \frac{I_d}{P_d}$, where I_d is the total income and $P_d = \left(\frac{R_d}{\alpha^0}\right)^{\alpha^0} \cdot \prod_{i=1}^{S} \left(\frac{P_d^i}{\alpha^i}\right)^{\alpha^i}$ is the price of the consumption basket. Domestic consumers choose a location to maximize their utilities, so in equilibrium $U_d = U_{d'}, \forall d, d' \in$ China.

Land market. Region d is endowed with $\overline{H}_d$ amount of land. Let the equilibrium number of consumers in region d be L_d. The land market clearing condition is

$$H_d L_d = \overline{H}_d, \forall d.$$

Domestic land is owned by the national government, which collects rents and rebates them to domestic consumers via a lump-sum transfer, Tr. Let w_d be the wage in city d. The total income of a consumer in city d is $I_d = w_d + Tr$, and Tr is given by government budget balance: $\sum_{d \in CHN} R_d H_d L_d = Tr \cdot \sum_{d \in CHN} L_d$.

Industry final good production. In each industry i and region d, a representative final good producer aggregates intermediate goods in sector i from different locations into sectoral final goods. Intermediate goods produced in different locations are distinct from each other. Let $\tilde{q}^i_{od}$ be the quantity of intermediate goods of sector-i from region o, the quantity of final goods produced, Q^i_d, is

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

where σ is the elasticity of substitution across goods from different regions.

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

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

where T^i_d is the location-sector specific productivity shaping the specialization of a region. l^i_d and m^{ij}_d are inputs of labor and final goods from industry j, respectively; β^i and γ^{ij} are their respective shares: $\beta^i + \sum_j \gamma^{ij} = 1$. The unit production cost of sector-i intermediate goods in region d is thus:

$$c^i_d = \frac{\kappa^i w_d^{\beta^i} \prod_{j=1}^{S} [P^j_d]^{\gamma^{ij}}}{T^i_d}, \text{ where } \kappa^i \text{ is a constant: } \kappa^i = [\beta^i]^{-\beta^i} \prod_{j=1}^{S} [\gamma^{ij}]^{-\gamma^{ij}}.$$

The representative intermediate good producers sell their output to final good producers at marginal costs, which consist of production costs and iceberg trade costs, $\tilde{\tau}^i_{od}$, specified below. The price of intermediate goods sold from o to d is thus $p^i_{od} = c^i_o \tilde{\tau}^i_{od}$. The price of final goods in region d sector i is:

$$P^i_d = \left(\sum_o (c^i_o \tilde{\tau}^i_{od})^{1-\sigma} \right)^{\frac{1}{1-\sigma}}.$$

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

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

where E^i_d is the total expenditure on intermediate goods in sector i of region d and π^i_{od} is the share of expenditure of region d spent on sector-i intermediate goods from region o.

Other details of the model and equilibrium conditions are standard and hence delegated to Appendix B.3. Now we discuss how we model the trade costs, $\tilde{\tau}^i_{od}$.

5.2 Incorporating Alternative Transport Modes

In Section 3, we have developed a routing model that gives rise to trade costs along the road network. While ground transport is the dominant form of transport in China accounting for 76% of all domestic shipment (National Bureau of Statistics, 2010), alternative modes of transport via air, water, railways, and pipelines might still be relevant for counterfactual experiments, as improvements in road infrastructure might draw traffic away from other modes. We capture these alternatives parsimoniously by assuming that between any two domestic regions, in addition to the road network (with an expected cost of τ^i_{od} given in Section 3.2), there is an alternative transport mode with an expected cost, $\bar{\tau}^i_{od}$, specified as

$$\bar{\tau}^i_{od} = \left(\frac{h_i}{h_0} \right)^{\mu} \exp(\bar{\kappa} \cdot \overline{\text{dist}}_{od}),\ o \neq d, \tag{11}$$

where $\bar{\kappa} > 0$ is a parameter to be estimated. We specify $\bar{\tau}^i_{od}$ as a function of the great circle distance between o and d, $\overline{\text{dist}}_{od}$, because it is meant to capture the average cost among all alternative modes, including air transport.[24]

[24] Even after conditioning on the great circle distance between two cities, the cost of shipping via alternative modes might still differ according to the accessibility of direct flights and trains. Given the data limitations, we do not directly model these alternatives. Our counterfactual experiments should thus be

With this additional mode, the full structure of the routing model is as follows. For a seller from region o looking to ship a batch of goods, when the destination is a domestic region, the seller decides whether to ship goods via ground transport or the alternative mode. The seller draws two independent mode-specific shocks from a Fréchet distribution with dispersion parameter θ_M, denoted by ν_M, $M \in \{\text{road}, \text{alt}\}$, 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 ground transport is chosen, the seller randomly meets with a trucker and pays the expected trade cost along the road network, τ^i_{od}; otherwise he faces the cost for the alternative mode, $\bar{\tau}^i_{od}$. When the destination is the RoW, the seller first chooses a port d, taking into account the realization of port-specific shocks, before choosing the mode of transport, ground or the alternative, from o to d.[25]

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

$$\tilde{\tau}^i_{od} = \begin{cases} \Gamma(\dfrac{\theta_M - 1}{\theta_M})[(\bar{\tau}^i_{od})^{-\theta_M} + (\tau^i_{od})^{-\theta_M}]^{-\frac{1}{\theta_M}}, & \text{if } d \in CHN, \\ \Gamma(\dfrac{\theta_F - 1}{\theta_F}) \cdot [\sum\limits_{\text{ports } k} (\tilde{\tau}^i_{ok} \cdot \tau^i_{k,RoW})^{-\theta_F}]^{-\frac{1}{\theta_F}}, & \text{if } d = RoW, \end{cases} \tag{12}$$

where $\tilde{\tau}^i_{ok}$ in the second line of the equation is given by the first line for domestic port city k. We construct import costs of Chinese cities from the RoW in the same form.

6 Quantification

6.1 Parameterization

As summarized in Panel A of Table 4, the key structural parameters of the routing model have been estimated in Section 4. This section parameterizes the rest of the model

viewed as keeping these alternatives fixed.

[25]It is possible that in the data, as in the model, some goods are first shipped via the alternative mode (most likely by train) to a seaport and then sent to the RoW. One concern is that not excluding such transshipments could bias our estimates. Since transshipment from railways to ports is more likely for heavier and bulkier industries which are more dependent on railway for transportation, such as coal and wood, we exclude these two categories from the reduced-form estimation and obtain essentially the same result. In a related finding, robustness exercises reported in Appendix A.6 also show that focusing on within-industry variation gives similar results, which suggests that possible transshipment in some industries is unlikely to bias our estimates.

Table 4: Parameter Values

Parameters	Descriptions	Value	s.e.	Targets/Source
A. Estimated Routing Parameters				
θ	Routing elasticity	111.5	35.4	Estimates of Equations (9) and (10)
θ_F	Port choice elasticity	6.35	3.33	
κ^H	Expressway route cost	0.034	0.002	
κ^L	Regular route cost	0.042	0.008	
μ	Cost-weight to value elasticity	0.29	0.04	
B. Remaining parameters: from external sources				
$\beta^i, \gamma^{ij}, \alpha^j$	IO structure and consumption share	-	-	China Input Output Table (2007)
σ	Trade elasticity	6	-	
θ_M	Elasticity of substitution across modes	2.5	-	
C. Remaining parameters: estimated jointly				
h_0	Trade cost level	1.260	0.015	Average ground shipment distance: 177 km
$\bar{\kappa}$	Alternative mode cost	0.163	0.001	Share of non-road shipment: 0.24
τ^i_{RoW}	International trade costs	-	-	Sectoral export and import
T^i_d	Region-sector productivity	-	-	City-sector sales (2008 Economic Census)
B_d	Amenities	-	-	Pop. dist. (2010 Pop. Census)
$\overline{H}_d$	Land supply shifter	-	-	Rent (2005 mini Census)

and conducts counterfactual exercises.

Parameters assigned directly. Panel B of Table 4 describes the parameters and fundamentals of the economy that are assigned directly. We determine sector shares in final consumption and intermediate production, $\{\alpha^i\}$ and $\{\gamma^{ij}\}$, and the labor shares in production, $\{\beta^i\}$, based on the 2007 Input-Output Table of China. We set the elasticity of substitution across goods from different regions, σ, to be 6, implying a trade elasticity of 5, which is in the range of estimates in the literature (see, e.g., Simonovska and Waugh, 2014). Finally, θ_M governs the elasticity of substitution between different modes of transport. Existing estimates of θ_M range from 1 to 3 in the earlier transportation literature (Abdelwahab, 1998) to 14 in a more recent study by Allen and Arkolakis (2019). We assign $\theta_M = 2.5$ as the baseline and conduct sensitivity checks with alternative values.

Parameters estimated jointly. The remaining parameters, reported in Panel C, are estimated jointly, with their standard errors generated through bootstrapping of parameters in Panel A.[26] We determine the overall level of domestic trade cost h_0 by targeting

[26]The standard errors are generated by recalibrating the model each time we randomly sample parameters in Panel A from their joint distribution. The inference procedure is described in Appendix C.2. In this process, we treat parameters in Panel B as fixed, as those either are aggregate moments (IO shares), which have no sampling errors, or are taken from the literature. We explore how results vary with these

the average shipping distance in China, which is 177 km (National Bureau of Statistics, 2010). The distance semi-elasticity for the alternative mode, $\bar{\kappa}$, pins down the equilibrium share of shipment using roads versus the other mode. Approximately 76% of domestic shipments are via ground transport (National Bureau of Statistics, 2010). Matching this target gives $\bar{\kappa} = 0.163$.

We assume that export and import costs vary by sector but not by port: $\tau^i_{k,RoW} = \tau^i_{RoW}, \forall k \in CHN$. These parameters are then pinned down by matching the sectoral imports and exports as shares of domestic output.[27] The model has 323 prefecture cities and 25 sectors, 4 of which are non-tradable. We pin down the region-sector productivity parameters $\{T^i_d\}_{d \neq RoW}$ by matching the sectoral output shares of each prefecture city, constructed from the 2008 economic census. We calibrate $\{T^i_{RoW}\}$ so that the ratios of the sectoral output of China and that of the RoW match the data. Finally, we determine the amenities of cities by matching the population distribution, calculated from the 2010 census. City population, together with rental rates, pins down the land supply shifters, $\{\overline{H}_d\}$.

Figure 5 plots the value of shipment flows from the calibrated equilibrium. Darker colors indicate higher intensities. Standing out from the map are a few corridors that connect the most important economic centers of China. The first is the northeast corridor surrounding the Bohai Bay, which links Beijing and Tianjin to clusters of heavy industries such as Dalian, Shenyang, and Changchun. The second is the corridor between Beijing and the southeast coast, an area encompassing the most prosperous areas of China, the Yangtze River Delta. Finally, the corridor that connects the northwest to the center of China is also important.

Summary of validation exercises. To verify that the model, disciplined by the customs

parameters in Appendix C.6.

[27] To match the import and export shares, our calibration accounts for *exogenous* international trade surpluses of China. After the model has been calibrated, 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 baseline as the calibrated equilibrium.

Figure 5: Model-Predicted Shipment Flows

Note: This figure plots the value of road shipments (the sum of shipments via expressways and regular roads) between directly connected cities. Value are in percentage points of Chinese GDP.

data, indeed matches the patterns of shipment and trade, we conduct a few validation exercises. First, we show that despite the key structural parameters of the model being estimated from *within* variation in export routing patterns, the calibrated model fits the *level* of city exports well. Second, the model-predicted city-level export *growth* due to the expressway expansion fits that in the data. In addition to being an out-of-sample test of the model, the predicted export growth also provides an IV for city-level export growth due to exogenous expressway expansion. Third, we obtain bilateral truck flows in 2019 and compare the model-implied bilateral shipments to truck flows. Finally, we relate the model-implied shipments that *pass through* each city to the data. Appendix C.3 reports details of these validation exercises. Together, these exercises demonstrate that our approach closely matches the patterns of China's domestic and international trade, which are the first order determinants of the inferred welfare gains.

6.2 Counterfactuals: Impacts of the Expressway Network Expansion

Main results. We examine the impacts of the expressway construction through counterfactual experiments. Table 5 reports the differences between the calibrated equilib-

Table 5: Impacts of the Expressway Network Expansion, 1999-2010

Change in	Value	s.e.	Median	p10	p90
Aggregate welfare (%)	0.051	0.025	0.052	0.022	0.096
Log(Domestic trade)	0.136	0.052	0.136	0.068	0.230
Log(Exports)	0.097	0.080	0.108	0.035	0.219

Note: The table reports (the negative of) changes in model statistics as the economy moves from the calibrated equilibrium with the 2010 expressway network to the one with the 1999 expressway network. Inferences of these statistics are generated through counterfactual experiments based on 200 recalibrated models with parameters in Panel A of Table 4 sampled from their joint distribution.

rium with the 2010 expressway network and the counterfactual equilibrium with the 1999 expressway network. The comparison suggests that the aggregate welfare of China increased by 5.1% because of the expressway construction. To put this number into perspective, the welfare relevant TFP of China grew by 36% from 1999 to 2010 (Penn World Table 9.0, see Feenstra et al., 2015). Through the lens of our model, about 14% of this increase can be attributed to domestic expressway network expansion.

The expressway construction also has had large impacts on both domestic and international trade. With more connected domestic markets, trade within China increased by 13.6%. Because the hinterland ship its exports to ports via ground transport, expressway expansion also affects international trade. It is tempting to think that lower domestic shipping costs will encourage international trade, but the theoretical prediction is ambiguous.[28] It turns out that in the model, the net effect is a 9.7% increase in international trade.

Role of three model ingredients. Our model differs from those used in the growing literature quantifying the impacts of transport infrastructure (e.g., Asturias et al., 2018; Fajgelbaum and Schaal, 2020; Allen and Arkolakis, 2019) in three aspects. First, our structural estimation exploits changes in the route choice of exporters resulting from the domestic expressway network expansion, which naturally implies that the network expansion reduces trade costs not only for trade between domestic partners but also for

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

that between the hinterland and foreign countries; second, with sector level information on production and export prices, we allow regions to differ in sector specializations and sectors to differ in trade costs; third, we incorporate intermediate inputs.

Because these ingredients allow us to more accurately infer the value of intercity shipments and the distribution of these shipments among different routes, they are important for the quantitative results. To illustrate the roles of these ingredients, in Appendix C.4, we parameterize a series of restricted models with fewer ingredients and calculate the welfare gains in these alternative models. We find decreasing gains as these ingredients are eliminated one by one. When all three ingredients are removed—so the model is reduced to a bare-bone single-sector spatial equilibrium model—the inferred welfare gain is about 0.89%, around 17% of the baseline result.

Sensitivity to alternative setups and parameters. In Appendix C.6, we report results for alternative values of external parameters and results for models with (1) frictional instead of freely mobile labor and (2) industry-level external economies of scale. We show that adopting these alternative assumptions does not materially affect the inferred welfare gains from the expressway expansion.

6.3 Cost-Benefit Analysis

We evaluate the return to investment for both the overall expressway network expansion and a few mega-projects.

Overall expressway expansion. We calculate the total investment in the expressway network during 1999-2010. The raw data are from the Yearly Bulletin of Road and Waterway Transport Development (Ministry of Transport of the People's Republic of China, 2000-2010). Converted to the 2010 price using the price index for capital, the cumulative investment in intercity expressway projects during the decade was 570 billion USD, or about 10% of the 2010 GDP. To compare this cost to discounted future benefits, we assume that the annual depreciation rate for expressways and the discount rate are both

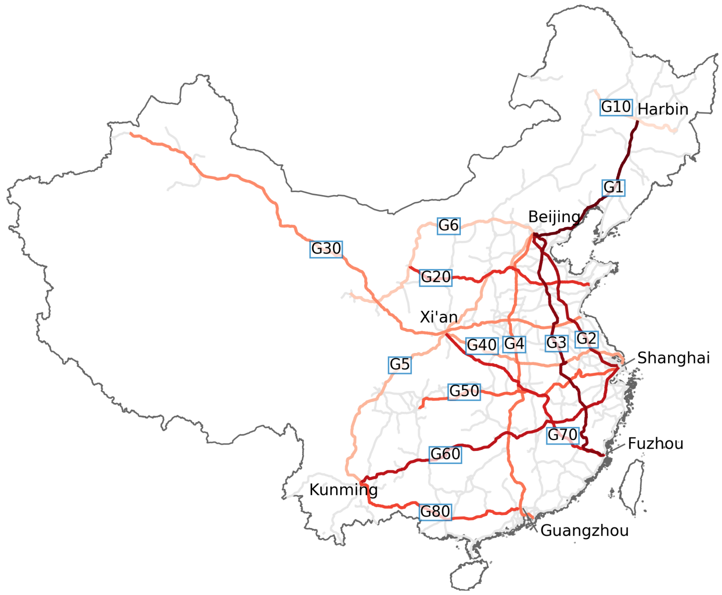


Figure 6: Expressway Mega-projects in China

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

10%.[29]

Assuming that all expenditures were incurred in 2010, then, the discounted future welfare gains ($\frac{5.1}{0.1+0.1}$) are around 25% of the 2010 GDP, implying a net return of about 150%: even after taking into account the high opportunity cost in a growing economy such as that of China, the expressway investment generates a large net return. In comparison, if we had used a simple one sector model for the evaluation, as done in most existing quantitative studies, our conclusion would have been that the investment led to a 55% net loss ($\approx \frac{0.89\%}{0.2}/10\% - 1$).

Return to backbone projects. Fourteen mega-projects, shown in Figure 6, form the backbone of the entire network. We evaluate the cost and benefit for each of them. In the absence of a consistently defined cost measure for individual projects, we follow Faber (2014) and adopt a formula based on the engineering literature linking the *relative* construction cost of a segment to whether it passes water or wetland areas and the average

[29] The choice of the depreciation rate follows Bai et al. (2006). For the discount rate, a natural candidate appears to be the return to capital in the overall Chinese economy, given that the expressway was planned by the central government, whose opportunity cost is to direct investment elsewhere. Bai et al. (2006) finds that between 1998 and 2005, the return to capital is around 20%, a level that seems unsustainable especially given the secular stagnation in much of the developed world. To be conservative, we assume it to be 10%.

Table 6: Costs and Benefits of 14 Mega-projects

ID	Length (km)	Cost as % GDP	Cost per km (million RMB)	Welfare Gains (%)	Net return to investment	% Change in dom. trade	% Change in Exports
G1	1533.61	0.30	77.71	0.40	567.19%	1.16	0.56
G2	1768.29	0.38	85.94	0.29	284.82%	0.89	0.88
G3	2513.38	0.54	85.53	0.49	354.10%	1.05	1.86
G4	2924.88	0.65	89.14	0.32	149.50%	0.82	0.44
G5	2829.75	0.73	103.16	0.20	38.04%	0.57	0.00
G6	2095.37	0.38	72.26	0.08	3.87%	0.25	0.03
G10	891.73	0.15	67.25	0.02	-22.92%	0.09	0.02
G20	1688.68	0.31	74.08	0.19	204.16%	0.54	0.33
G30	4356.49	0.85	78.04	0.39	129.32%	1.34	-0.10
G40	1727.03	0.34	78.43	0.12	75.26%	0.34	0.19
G50	1936.36	0.38	79.61	0.22	180.97%	0.56	0.28
G60	2662.22	0.48	72.99	0.35	258.44%	0.67	0.46
G70	1706.35	0.38	89.62	0.24	217.86%	0.36	1.06
G80	1378.30	0.30	88.62	0.17	185.05%	0.11	0.48
Total	30012.46	6.16	-	3.47	-	8.76	6.48

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

slope of the terrain. We use this formula to evaluate all segments constructed between 1999 and 2010 as a function of an unknown *level* coefficient and determine this coefficient so that the total cost of these segments equals the aggregate investment (10% of 2010 GDP). Appendix A.4 provides more details.

The output of this procedure, reported in Table 6, Column 3, is the estimated cost of each of these projects. The most expensive project per kilometer is G5, which passes through the rugged terrain in the southeast. Stretching across the flat northeastern plain at the other end of the country, G10 costs the least per kilometer. The average cost across all projects constructed in this period is around 80 million *yuan* per kilometer. This number is in the same ballpark as the best directly available evidence.[30]

Columns 4 and 5 report per-period welfare gains and the net return to investment for each project. Most projects generate positive net returns. Projects with the highest returns are north-south expressway lines (G1, G2, and G3). G10, a small project passing

[30]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 the first quarter of 2011. According to it, the average construction cost was 80 million yuan per kilometer.

through the less prosperous northeastern China, is the only one that loses money. The last two columns report the impact of a project on domestic and international trade. The project that had the biggest impact on domestic trade is G30. This is likely because it stretches across China's center to the northwest, connecting areas with very different specializations, and has no major competing routes. On the other hand, the projects that have had the largest impacts on exports are G2, G3 and G70—roads that connect northern and central China to southeastern ports like Shanghai and Fuzhou.

In summary, our evaluation suggests that, return heterogeneity notwithstanding, until 2010 the expressway network in China was worth every penny invested.[31]

Comparison to existing studies. At least three other strategies have been used to evaluate the return to expressway investment in China. The first directly measures the capital value added of the transportation sector (Bai and Qian, 2010); the second estimates a regional production function, in which transport infrastructure is one of the inputs (Fan and Chan-Kang, 2005); the third relies on quantitative simulations, as we do here, but does not use domestic or international trade to discipline the model (Roberts et al., 2012). We compare these approaches and explain their differences in Appendix C.5. While each study obtains different numbers, our estimate is generally of the same order of magnitude as those in existing studies. This further supports using customs data as a promising source of information for understanding the impacts of domestic transport infrastructure.

7 First Order Measurements and Second Order Corrections

We have evaluated infrastructure projects through counterfactual experiments. However, in applications that require comparing a large number of proposed transport projects, solving for many counterfactual experiments in the fully fledged model could be computationally demanding. An alternative 'social cost saving' approach, dating back to Fogel

[31] More recently, there has been a heated discussion in the popular press on whether China 'overinvested' in transport infrastructure. We note that our finding does not necessarily apply to the latest wave of investment. Indeed, as major population centers have been connected, building roads in the more mountainous areas, usually with redistributive motives, might incur higher costs while generating smaller returns.

(1964), is to evaluate the total savings in transport cost based on observed shipments. This approach is transparent and easy to implement, but its accuracy is less clear. In this section, we derive a theory-based formula for the welfare gains, taking advantage of the tractability of the routing model. This formula is as transparent and interpretable as the 'social cost saving' approach, but improves on the accuracy of the latter.

7.1 Analytical Characterization of Welfare Gains

Individual expressway segment and edge cost. We first characterize the change in trade costs from adding an expressway segment between adjacent cities. Suppose that two adjacent cities, k and l, are connected by both an expressway and a regular road. Let ι_{kl} be the expected edge cost between k and l, $\iota_{kl} = \Gamma(\frac{\theta}{\theta-1})[(\iota_{kl}^H)^{-\theta} + (\iota_{kl}^L)^{-\theta}]^{-\frac{1}{\theta}}$. Using the definition of ι_{kl}^H and ι_{kl}^L, the *increase* in ι_{kl} from *removing* expressway $k \xrightarrow{H} l$ is:

$$\begin{aligned}\Delta \log(\iota_{kl}) &= -\frac{1}{\theta}\Big(\log[\exp(-\theta\kappa^L dist_{kl}^L)] - \log[\exp(-\theta\kappa^H dist_{kl}^H) + \exp(-\theta\kappa^L dist_{kl}^L)]\Big) \quad (13)\\ &\approx (\kappa^L - \kappa^H)\cdot dist_{kl}^L + \kappa^H(dist_{kl}^L - dist_{kl}^H),\end{aligned}$$

in which the first term captures the cost increase holding the road length constant and the second term captures that expressways—using more bridges and tunnels—tend to be straighter than regular roads.

Expressway projects and trade cost. We can view a large project as a *collection* of expressway segments. Denote a large project by set $\mathbb{C}$. The second order approximation for the change in trade cost between two domestic locations o and d in response to $\mathbb{C}$ is:[32]

$$\Delta \log \tilde{\tau}_{od}^i \approx \sum_{kl\in\mathbb{C}} \frac{\partial \log \tilde{\tau}_{od}^i}{\partial \log \iota_{kl}} \Delta \log(\iota_{kl}) + \frac{1}{2}\sum_{kl\in\mathbb{C}}\sum_{k'l'\in\mathbb{C}} \frac{\partial^2 \log \tilde{\tau}_{od}^i}{\partial \log \iota_{kl}\,\partial \log \iota_{k'l'}} \Delta \log(\iota_{kl})\Delta \log(\iota_{k'l'}) \quad (14)$$

The first term adds up the first order effect on $\tilde{\tau}_{od}^i$ of all individual segments $kl \in \mathbb{C}$. The second term captures the second order effect. It includes each segment's own second order effect (when $kl = k'l'$) and the interactions among different segments (when $kl \neq k'l'$). The lemma below characterizes the partial derivatives in Equation (14).

[32] Analogous expressions can be derived for when one of the two locations o or d is the RoW.

Lemma 1. *Let π_{od}^{road} be the fraction of shipment between o and d that uses ground transport, and π_{od}^{kl} be the fraction of ground-transported shipment between o and d that passes edge kl. Then when θ is large,*

$$\frac{\partial \log \tilde{\tau}_{od}^{i}}{\partial \log \iota_{kl}} \approx \pi_{od}^{road} \pi_{od}^{kl}, \tag{15}$$

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

This lemma follows from a result in Allen and Arkolakis (2019) and is proven in Appendix B.4. Since we find θ to be fairly large (≈ 111.5), the premise of the Lemma applies. The first part of Equation (15) establishes that the marginal impact of an edge on trade cost between two cities is approximately the fraction of trade between the two cities being transported via edge kl. The second part of the equation builds on the first part by noting that $\frac{\partial^2 \log \tilde{\tau}_{od}^{i}}{\partial \log \iota_{kl} \partial \log \iota_{k'l'}} = \frac{\partial (\pi_{od}^{road} \pi_{od}^{kl})}{\partial \log \iota_{k'l'}}$, i.e., the cross-derivative captures how edge $k'l'$ affects the importance of kl for the shipment between o and d. As discussed in Appendix B.5, this term could be either positive or negative, depending on whether $k'l'$ and kl are complementary or substitutable edges on the road network.

Once the model has been parameterized, all variables in Equations (13) and (15) are known. We can then directly evaluate the change in the trade cost between any pairs, $\Delta \log(\tilde{\tau}_{od}^{i})$, in Equation (14). This circumvents the need to search for the shortest paths between cities when a new segment is built. As important, differentiability carries over to the trade model so welfare changes could also be expressed as observables, as we show in Proposition 1.

Proposition 1. *Let W be the utility of Chinese consumers in the competitive equilibrium. The effect of an infrastructure project $\mathbb{C}$ on W is:*

$$\Delta \log W = \underbrace{-\sum_{o,d,i} \sum_{kl \in \mathbb{C}} \Big(\frac{X_{od}^{i}}{Y} - \frac{\Lambda_{o}^{i}}{Y} \mathbb{1}_{d=RoW} \Big) \cdot \frac{\partial \log \tilde{\tau}_{od}^{i}}{\partial \log \iota_{kl}} \Delta \log(\iota_{kl})}_{FO_R} + SO_R + \underbrace{HO_R + ToT + SO_T}_{Residual}, \tag{16}$$

where Y is the domestic GDP, Λ_{o}^{i} is the exposure of the RoW consumer to goods in sector i *produced in city o,*[33] *and SO_R is:*

$$SO_R = -\frac{1}{2}\sum_{o,d,i}\left(\frac{X_{od}^i}{Y} - \frac{\Lambda_o^i}{Y}\mathbb{1}_{d=RoW}\right)\sum_{kl\in\mathbb{C}}\sum_{k'l'\in\mathbb{C}}\frac{\partial^2 \log\tilde{\tau}_{od}^i}{\partial\log\iota_{kl}\,\partial\log\iota_{k'l'}}\Delta\log(\iota_{kl})\Delta\log(\iota_{k'l'}),$$

with $\frac{\partial\log\tilde{\tau}_{od}^i}{\partial\log\iota_{kl}}$ *and* $\frac{\partial^2\log\tilde{\tau}_{od}^i}{\partial\log\iota_{kl}\,\partial\log\iota_{k'l'}}$ *given by Equations (15).*

We delegate the proof to the appendix and explain here what Equation (16) entails. In the first component, the summation over $\frac{X_{od}^i}{Y}\frac{\partial\log\tilde{\tau}_{od}^i}{\partial\log\iota_{kl}}\Delta\log(\iota_{kl}) \approx \frac{X_{od}^i}{Y}\pi_{od}^{road}\pi_{od}^{kl}\Delta\log(\iota_{kl})$ is simply the sum of cost savings on all trade flows shipped via the road segments in $\mathbb{C}$, assuming that trade flows (X_{od}^i), transport mode choices (π_{od}^{road}), and route choices (π_{od}^{kl}) do not adjust in response. This term captures the first order cost savings from the project for the *world economy*. Part of these cost savings are passed on to the RoW. Adjusting for these spillovers, captured by $\frac{\Lambda_o^i}{Y}\mathbb{1}_{d=RoW}\cdot\pi_{od}^{road}\pi_{od}^{kl}\Delta\log(\iota_{kl})$, the net effect is the first order gains from the infrastructure project for the domestic economy. We label it FO_R.

An approach pioneered by Fogel (1964) and used widely in transportation economics (see Small, 2012 for a recent survey) is to focus on the value of travel time or transport cost savings, calculated as the product of the time saved through the new transportation infrastructure and the value of time. Proposition 1 makes clear the connection between this approach and the current model: it approximates the first order effect of transportation projects in a closed-economy, when neither trade nor traffic patterns respond.

Beyond the first order effect, SO_R stands for the second order effect from the routing block. It consists of two components. First, the own effect (when $k'l' = kl$). Note that $\frac{\partial^2\log\tilde{\tau}_{od}^i}{\partial\log\iota_{kl}\partial\log\iota_{kl}} = \frac{\partial(\pi_{od}^{road}\pi_{od}^{kl})}{\partial\log\iota_{kl}} < 0$. This term captures that in response to a reduction in ι_{kl}, a higher fraction of trade flows between o and d would be rerouted to pass edge kl. Second, the cross derivative (when $k'l' \neq kl$), which captures the impact of edge $k'l'$ on shipping via kl. For large projects with many segments, omitting the interaction among shipments can either underestimate or overestimate the welfare impact. Since SO_R reflects the response in routing patterns to transport projects, it tends to be larger if routes are

[33] Λ_o^i can be expressed as a function of observables only, characterized in the proof of the proposition.

more substitutable (when θ is larger) and for larger changes to the transport network. Since both conditions are met in our setting, SO_R could be quantitatively significant.

In addition to FO_R and SO_R, the full welfare effect also includes the following: HO_R, a residual term capturing trade cost changes not embodied in $FO_R + SO_R$ (i.e., the approximation error in Equation (14)); a terms-of-trade effect (ToT) reflecting that for a large economy like China, the domestic expressway network expansion can affect the relative wages between China and the RoW; and last but not the least, a second order effect due to response in trade flows both within China and between China and the RoW (SO_T). While ToT and SO_T are only known after the counterfactual equilibrium has been solved for, FO_R and SO_R can be evaluated ex ante using the data from the baseline equilibrium. Proposition 1 thus provides an approximation to the welfare gains without solving counterfactual equilibria. In the remaining of this section, we demonstrate the importance and adequacy of the second order correction, SO_R, in this approximation.[34]

7.2 Evaluating the Quality of Approximations

Welfare gains from individual expressway segments. We first evaluate the quality of approximations for individual expressway segments by removing one segment at a time from the network. In this case, the term FO_R captures the cost savings that accrue to China of shipments transported on that segment. FO_R, however, likely overstates the welfare impacts. Intuitively, when reoptimization is allowed, in response to the *removal* of an expressway segment, some of the shipments originally passing over the segment may be rerouted. Holding shipment flows on the segment unchanged, therefore, inflates the loss from removing the expressway. Similarly, starting from an equilibrium without the expressway and inferring the gains from the *addition* of expressways will underestimate the gains. The effect of such route reoptimization can be partly captured by the second order correction, SO_R. We also gauge the importance of the remaining high order effects

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

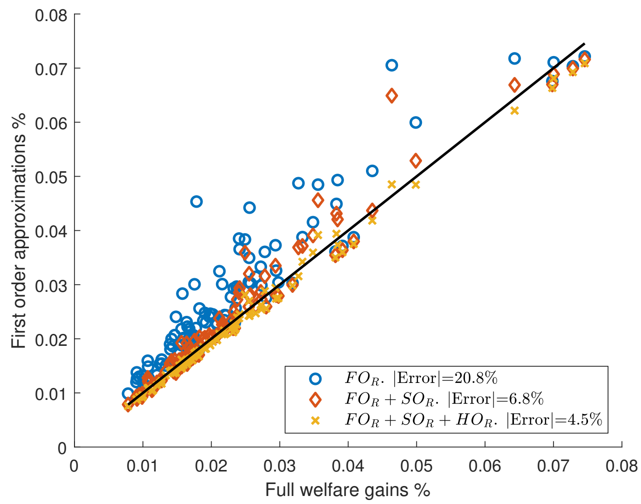


Figure 7: Nonlinear Welfare Gains v.s Formula Approximations

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

by evaluating the welfare effects under the actual changes of trade costs that take into account HO_R.

We consider the top 100 expressway segments ranked by shipment value. In each experiment, we remove one segment and calculate the effects of that segment on the welfare of China. The horizontal axis of Figure 7 plots the full effect calculated by solving the counterfactual equilibrium. The vertical axis shows the results of various approximations. The circles denote the first order cost savings, FO_R. As anticipated, most circles are above the 45 degree line, indicating that FO_R overestimates the loss from removing an expressway segment. The biases average around 21% and are larger for busier segments.

The diamonds further incorporate the second-order effect from rerouting, SO_R. This results in a substantial improvement in the quality of the approximation, with the diamonds centering closely around the 45 degree line. The mean absolute error is reduced by two thirds to 7%. Finally, the crosses are based on the actual changes of trade costs and incorporate all responses in routing. This improves the quality of the approximation marginally by 2.3%, suggesting that the remaining higher order effects in the routing block, HO_R, are generally unimportant. Left out of this approximation are the terms-of-trade effects and the second order effects from changes in trade costs. Even for a large economy like China, approximation errors from these forces average to only about 4.5%.

Large projects and interaction between segments. When analyzing large expressway projects with many segments, in addition to the approximation errors illustrated in Figure 7, the first order approximation also misses the interaction between segments. Formally, we can decompose SO_R in Equation (16) into the own and cross second order effects:

$$[\underbrace{\sum_{kl\in\mathbb{C}}\frac{\partial^2\log\tilde{\tau}^i_{od}}{(\partial\log\iota_{kl})^2}(\Delta\log(\iota_{kl}))^2}_{\text{Own SO}_R}+\underbrace{\sum_{k'l'\in\mathbb{C},k'l'\neq kl}\frac{\partial^2\log\tilde{\tau}^i_{od}}{\partial\log\iota_{kl}\,\partial\log\iota_{k'l'}}\Delta\log(\iota_{kl})\Delta\log(\iota_{k'l'})}_{\text{Cross SO}_R}].$$

The 'Own SO_R' term is exactly the SO_R correction for individual segments in Figure 7. The 'Cross SO_R', on the other hand, reflects interactions between different segments of a project. As illustrated through an example in Appendix B.5, depending on whether kl and $k'l'$ fall on the same long route or on competing routes, the 'Cross SO_R' could be either positive or negative.

To gauge the accuracy of the first and second order approximations for large projects, we calculate the nonlinear welfare gains of expressway segments constructed during the decade between adjacent cities also connected by regular roads.[35] In total, these expressway segments generate 2.2% welfare gains, which can be decomposed as follows:

$$\underbrace{\Delta\log W}_{0.022}=\underbrace{\text{FO effect}}_{146\%}+\underbrace{\text{Own SO}_R}_{-58\%}+\underbrace{\text{Cross SO}_R}_{8\%}+\overbrace{\underbrace{\text{ToT}}_{-6\%}+\underbrace{\text{HO}_R+\text{SO}_T}_{10\%}}^{\text{Approximation error}}. \tag{17}$$

The first order effect overestimates $\Delta\log W$ by 46%. This bias is more than entirely corrected by the own second order effect, which adjust the gains downward by 58%. The cross-substitution effect, which could work in both directions, adds 8% to the welfare gains in net. Once both second order effects have been included, what is left as an approximation error, consisting of the terms of trade effects and higher order effects from routing and trade costs, is merely 4% of the total gains.

[35] We focus on pairs of cities also connected by regular roads, because otherwise the edge cost increases to infinity when the expressway is removed, so local approximations would not apply.

This decomposition shows that the formula proposed in Proposition 1—in addition to being intuitive and theory consistent—performs well for major projects in a large open economy like China. It offers an intuitive and flexible alternative to the often computationally demanding counterfactual experiments in evaluating transport projects.

8 Conclusion

This paper proposes a method for evaluating the effect of transport infrastructure improvements on domestic trade costs that circumvents the lack of reliable domestic trade data in many countries—by using information on the route choice of exporters contained in typical customs data. We combine this method and a spatial equilibrium model to study the welfare gains from the $50,000$ km of expressways constructed between 1999 and 2010 in China. We find around 5.1% overall welfare gains from these projects and a net return to investment of 150%. Overlooking the three key ingredients in the model—regional comparative advantage, heterogeneous trade costs, and intermediate inputs—will lead to the conclusion of a negative aggregate return.

Because of the interactions between segments and rerouting on the transport network, evaluations of both local and large projects based on the first-order effects are inaccurate. Taking advantage of the model's tractability, we propose a second-order correction that reduces the biases, which allows convenient and accurate evaluation of transport projects.

References

Abdelwahab, Walid M, "Elasticities of Mode Choice Probabilities and Market Elasticities of Demand: Evidence from a Simultaneous Mode Choice/Shipment-size Freight Transport Model," *Transportation Research Part E: Logistics and Transportation Review*, 1998, *34* (4), 257–266.

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

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

_ and _ , "The Welfare Effects of Transportation Infrastructure Improvements," *NBER Working Paper No. 25487*, 2019.

Asturias, Jose, Manuel García-Santana, and Roberto Ramos, "Competition and the Welfare Gains from Transportation Infrastructure: Evidence from the Golden Quadrilateral of India," *Journal of the European Economic Association*, 2018.

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," *Brookings Papers on Economic Activity*, 2006, *2006* (2), 61–88.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian, "On the Road: Access to Transportation Infrastructure and Economic Growth in China," *Journal of Development Economics*, 2020, p. 102442.

Baqaee, David Rezza and Emmanuel Farhi, "The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem," *Econometrica*, 2019, *87* (4), 1155–1203.

Baum-Snow, Nathaniel, J Vernon Henderson, Matthew A Turner, Qinghua Zhang, and Loren Brandt, "Does Investment in National Highways Help or Hurt Hinterland City Growth?," *Journal of Urban Economics*, 2020, *115*, 103124.

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," *Working Paper*, 2019.

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.

Faber, Benjamin, "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System," *Review of Economic Studies*, 2014, *81* (3), 1046–1070.

Facchini, Giovanni, Maggie Y Liu, Anna Maria Mayda, and Minghai Zhou, "China's "Great Migration": The Impact of the Reduction in Trade Policy Uncertainty," *Journal of International Economics*, 2019.

Fajgelbaum, Pablo and Stephen J. Redding, "External Integration, Structural Transformation and Economic Development: Evidence from Argentina 1870-1914," *NBER Working Paper No. 20217*, 2014.

Fajgelbaum, Pablo D and Edouard Schaal, "Optimal Transport Networks in Spatial Equilibrium," *Econometrica*, 2020, *88* (4), 1411–1452.

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.

Fan, Shenggen and Connie Chan-Kang, *Road Development, Economic Growth, and Poverty Reduction in China*, Vol. 12, Intl Food Policy Res Inst, 2005.

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.

Feyrer, James, "Distance, Trade, and Income- The 1967 to 1975 Closing of the Suez Canal as a Natural Experiment," *NBER Working Paper No. 15557*, 2009.

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

He, Guojun, Yang Xie, and Bing Zhang, "Expressways, GDP, and the environment: The case of China," *Journal of Development Economics*, 2020, p. 102485.

Hillberry, Russell and David Hummels, "Trade Responses to Geographic Frictions: A Decomposition Using Micro-Data," *European Economic Review*, 2008, *52* (3), 527–550.

Hummels, David, "Transportation Costs and International Trade in the Second Era of Globalization," *Journal of Economic Perspectives*, 2007, *21* (3), 131–154.

Limao, Nuno and Anthony J Venables, "Infrastructure, Geographical Disadvantage, Transport Costs, and Trade," *The World Bank Economic Review*, 2001, *15* (3), 451–479.

Ma, Lin and Yang Tang, "Geography, Trade, and Internal Migration in China," *Journal of Urban Economics*, 2019.

Martincus, Christian Volpe, Jerónimo Carballo, and Ana Cusolito, "Roads, Exports and Employment: Evidence from a Developing Country," *Journal of Development Economics*, 2017, *125*, 21–39.

Ministry of Transport of the People's Republic of China, "The Annual Bulletin of Road and Water Transport Development," 2000-2010.

Morten, Melanie and Jaqueline Oliveira, "The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City," *NBER Working Paper No. 22158*, 2018.

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.

Roberts, Mark, Uwe Deichmann, Bernard Fingleton, and Tuo Shi, “Evaluating China’s Road to Prosperity: A New Economic Geography Aapproach,” *Regional Science and Urban Economics*, 2012, *42* (4), 580–594.

Sequeira, Sandra and Simeon Djankov, “Corruption and Firm Behavior: Evidence from African Ports,” *Journal of International Economics*, 2014, *94* (2), 277–294.

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

Simonovska, Ina and Michael E Waugh, “The Elasticity of Trade: Estimates and Evidence,” *Journal of International Economics*, 2014, *92* (1), 34–50.

Small, Kenneth A, “Valuation of Travel Time,” *Economics of Transportation*, 2012, *1* (1-2), 2–14.

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

Tombe, Trevor and Xiaodong Zhu, “Trade, Migration, and Productivity: A Quantitative Analysis of China,” *American Economic Review*, 2019, *109* (5), 1843–72.

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

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

Zhang, Yaxiong and Suchang Qi, *2007 China Multi-Regional Input-Output Model*, China Statistics Press, Beijing, China, 2012.

Appendix For Online Publication

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

Jingting Fan
Pennsylvania State University

Yi Lu
Tsinghua University

Wenlan Luo
Tsinghua University

Contents

A Data and Empirics

A.1 Constructing Coordinates of Ports and Origin Cities

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

In mapping the location of exporters to these coordinates, we use the origin city of export shipments from the customs data. An alternative definition is to use the registered address of the exporters. Using the former, instead of the latter, avoids potential measurement errors for the export of multi-plant firms.

A.2 Constructing Distance between Cities for Reduced-form Analyses

The raw road maps are in the form of line strings. For reduced-form analysis, we use the following procedure to find the shortest path between all pairs of cities for both 1999 and 2010.

In the first step, we split the entire main land China into $2km \times 2km$ squares. We define a square as 'on regular roads', if it intersects with any segments of regular roads. We define two *adjacent* squares as connected by regular roads, if both of them are 'on regular roads'. A path consisting of only regular roads is then a chain of connected squares, starting from the one containing the coordinates of the origin, ending with the one containing the coordinates of the destination.

We process the expressway networks in a similar way. We then overlay the two processed networks (regular roads and expressways). A path on this joint network is a set of connected squares, with each two adjacent squares connected by *either* regular roads *or* expressways. A square that is both 'on regular roads' and 'on expressways' is thus viewed as an intersection of regular roads and expressways, from which a trucker can switch from one type of road to the other.

Between any pair of cities, there could be many paths. In the second step, we search for the least-cost path. To this end, we assume each km along a regular road is twice as costly as a km on expressways and calculate the total regular road-equivalent length of all paths. We then use the Dijkstra's algorithm to find the path with the lowest length.

We do the above for both 1999 and 2010 road networks, generating the time-varying distances, $dist_{od}^t$.

A.3 Constructing City Network Graphs for the Routing Model

Our routing model treats individual cities as nodes in a network, connected by roads. Before the structural estimation, we prepare the data so that they are consistent with this model. To this end, we apply the following procedures separately to each of the three maps (expressways in 1999 and 2010, and regular roads in 2007 which are treated as time invariant).

- **Define connected cities.** In the first step, we identify the list of cities (prefectures) connected to the network. We define cities as 'connected' in a map, if the center of the city is within the 50 km radius of any roads on a map. Practically, it means measuring whether any of the coordinates characterizing roads from a map are within 50 km 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 on the road network using the Dijkstra's algorithm. If the shortest path between two cities does not pass through another city, we define the pair to be 'directly connected'.

- **Construct the graph.** We construct the graph in which cities are the nodes and roads form the edges, through the following procedure. We draw an 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 length of the shortest path between the two cities.[1]

The left panel of Figure A.1 is the original digital maps. The right panel overlays their network representation, which is the output of the above process. Again, even though the edges are drawn as straight lines in the right panel, the length we assign to each edge is the length of the actual road.

We transform the right panel of Figure A.1 into adjacent matrices, $\mathbb{H}^{1999}$, $\mathbb{H}^{2010}$, and $\mathbb{L}$, respectively, for structural estimation. Element (k, l) in a matrix will be $\iota_{kl}^{-\theta}$, if cities k and l are adjacent and 'directly connected' in the road network represented by that matrix; otherwise (k, l) will be zero.

A.4 Backing Out Segment-Specific Road Construction Cost

We first cut expressways into 10-km segments. For each such segment, we check if it passes water and calculate the average slope of its terrains.[2] We calculate the *relative* construction cost of segment i following a simple function from the transport engineering literature:

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

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

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

Figure A.2 shows the geographic features of China, which determine the cost estimates.

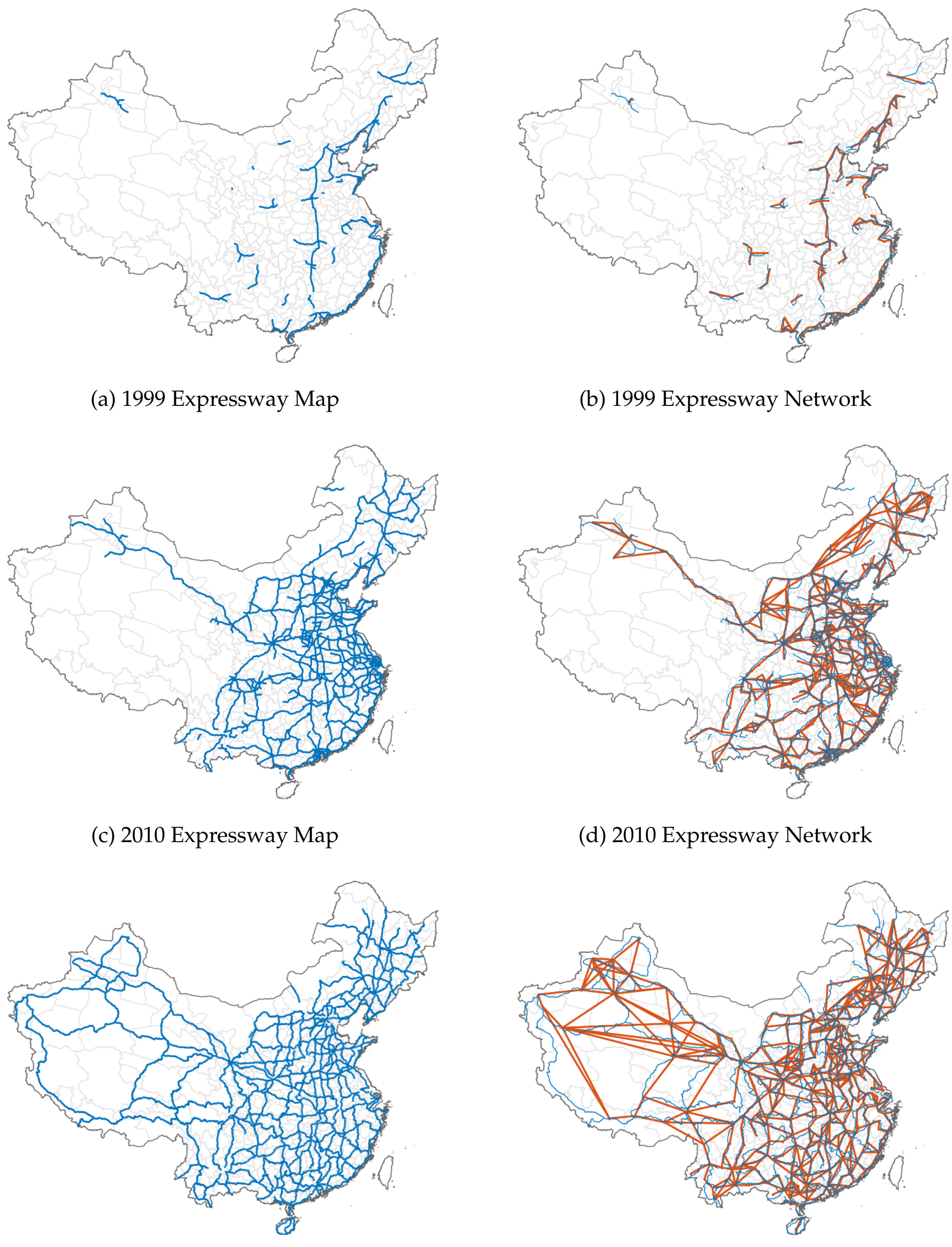
A.5 The Lists of Ports and Major Cities

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

List of major cities: Beijing, Tianjin, Shijiazhuang, Tangshan, Handan, Xingtai, Baoding, Cangzhou, Shenyang, Dalian, Changchun, Haerbin, Shanghai, Xuzhou, Suzhou, Nantong, Yancheng, Hangzhou,

[1]The two expressway maps are digitized from the projection of published hard-copy maps, which introduce measurement errors that change the exact locations of roads. The same road might therefore has slightly different measured lengths from 1999 and 2010 expressway maps. We inspect all segments with less 5% change in length to rule out measurement errors.

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



(a) 1999 Expressway Map

(b) 1999 Expressway Network

(c) 2010 Expressway Map

(d) 2010 Expressway Network

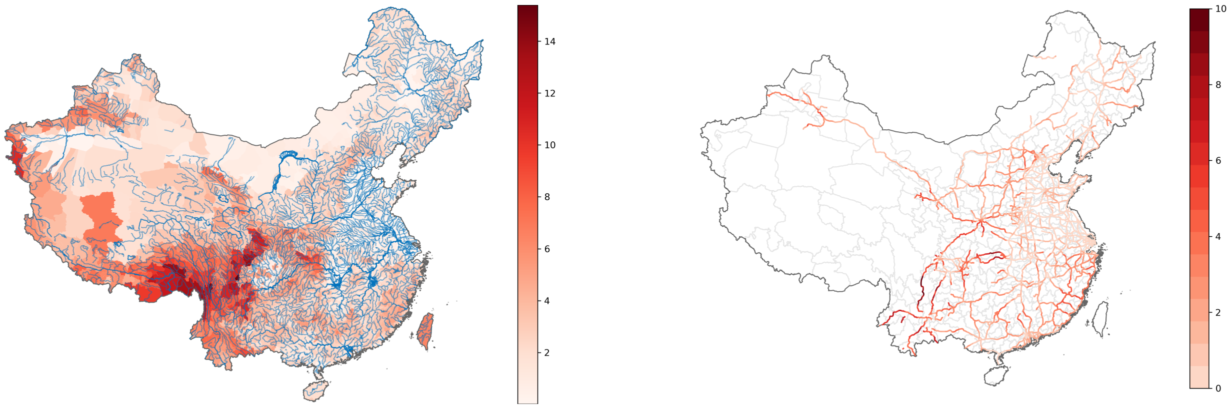
(e) Regular Road Map

(f) Regular Road Network

Figure A.1: From Road Maps to Road Networks

Note: A city is defined as 'connected' on a road network, if the center of the city is within the 50 km radius of any roads of the network. Two cities are defined as connected on a road network if the shortest path connecting them on the road network does not pass a third city. The distance between two connected cities is then defined as the *road length* of the shortest path between them. The left three panels plot the maps of the three road networks. The right three figures overlay the connected city pairs; each solid line segment corresponds to a pair of connected cities.

Wenzhou, Fuyang, Suzhou, Liuan, Quanzhou, Ganzhou, Jinan, Qingdao, Yantai, Weifang, Jining, Linyi, Liaocheng, Heze, Zhengzhou, Luoyang, Xinxiang, Nanyang, Shangqiu, Xinyang, Zhoukou, Zhumadian,



(a) Slopes and Water Areas

(b) Slopes of Expressway Segments

Figure A.2: Geography and Expressway Construction Costs

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

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

A.6 Additional Analyses: Channels and Sectoral Level Results

This subsection compares our reduced-form estimates to the literature. It inspect the channels and presents additional robustness results.

Comparing the estimate to the literature. The closest empirical setting to ours is Coşar and Demir (2016), who estimate the impacts of regional road capacity on trade. In a semi-elasticity specification (p. 240), they find that upgrading from carriageway to expressways lead to a "reduction of travel costs around 27% on an average stretch of 820 km." Take our estimate from Column 4 of Table 2 (with a coefficient of 0.174), under the assumption that expressways are on average twice as fast as regular roads, our finding implies that the upgrade reduces the coefficient by $0.174/2 = 0.087$. This coefficient is the product of an elasticity and the percentage difference in trade cost between regular roads and expressways. Using the elasticity of 4 used in Coşar and Demir (2016), our baseline estimate implies that upgrading each hundred km reduces trade cost by $0.087/4 = 2.2\%$. An upgrade of 820 km would therefore reduce trade cost by around 18%. This estimate is lower than Coşar and Demir (2016), likely because many regular roads in China have two or more lanes, so the marginal gains from upgrading to expressway are not as important as the upgrade from single-lane carriageway in Turkey. Nevertheless, the two estimates are in the same order of magnitude and their confidence intervals overlay.

Export growth versus rerouting. By controlling for city-time fixed effects, our baseline estimate uses two sources of variation: organic growth in shipments over an existing route, and rerouting of city export through competing ports. Notice both forces reflect the change in route choice due to the change in domestic shipment cost, and are precisely the forces used to infer the trade cost elasticity that is of interest. To gauge the relative importance of the two forces, we use two complementary approaches that rely on different assumptions.

The first approach is to aggregate the export of a city into a few groups of ports based on the geographic location of ports. The idea is that, if an improvement in connection between city o and port

Table A.1: Understanding Channels and Results from Sectoral Data

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Aggregate Data			Sectoral Data					
	Growth v.s. Rerouting			Baseline			Growth v.s. Rerouting		
$dist^t_{od}$	-0.226***	-0.166***	-0.157***	-0.373***	-0.138***		-0.183***	-0.149***	-0.149***
	(0.050)	(0.041)	(0.041)	(0.017)	(0.044)		(0.044)	(0.039)	(0.039)
- on express 5						-0.075*			
						(0.039)			
- on regular						-0.137***			
						(0.044)			
log(export of o through $d' \neq d$)		0.171**						-0.015	
		(0.073)						(0.026)	
Specification	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Fixed Effects	og_d, ot, g_dt	od, p_ot, dt	od, p_ot, dt	oti, dti	odi, oti, dti	odi, oti, dti	og_di, oti, gti	odi, p_oti, dti	odi, p_oti, dti
Exclude Major Cities yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	1048	2082	2082	20946	11758	11758	7060	12372	12372
R^2	0.932	0.870	0.869	0.593	0.896	0.896	0.926	0.852	0.852

Notes: The dependent variable is the log of total value of goods exported in city o through port d to the RoW. Columns 1 through 3 use aggregate data to explore whether response in export is due to organic growth or rerouting between ports (see text in Appendix A.6 for explanation). Columns 4 through 9 show the results are similar if we use data at the (2 digit) sectoral level. Columns 4 through 6 show that with sectoral data, controlling for city-port fixed effect also halves the cross-sectional estimate, and that expressways are less costly than regular roads. Columns 7 through 9 replicate Columns 1 through 3 using sectoral data. In 'Fixed Effects,' o stands for exporting city, d stands for port city, t stands for time, p_o stands for province of city o, g_d stands for geographic group of port d, i stands for sector.
Standard errors are clustered at city-port level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

d increases the export through d mostly by drawing in export of o through *other ports near* d, then by estimating the regressions at city-port group level, part of the rerouting effect would cancel out and the point estimate would be mainly about the export growth effect. To implement this approach, we group all ports into one of the three based on their geographic locations: North, Central, and South. We then aggregate the data to city-port group-time level, and estimate a similar panel-fixed effect specification as the baseline, controlling for city-port group, city-time, and port group-time fixed effects. Column 1 of Table A.1 finds the point estimate to be -0.226. If any, this is larger (although not statistically significantly) than the baseline estimate of -0.174 (Column 4 of Table 2), suggesting that it is growth of export in a city-port pair, rather than substitution between ports, that drives the baseline estimate.

The second approach seeks to directly control for export through other ports. Specifically, if the rerouting force is strong, then an improvement in the connection between city o and port d likely reduces the export of city o through all other ports. This implies that the export of city o through other ports ($d' \neq d$) is negatively correlated with the improving access between city o and port d, which in turn implies that if we control for city o's export via other ports, the estimated coefficient for distance will shrink. Columns 2 and 3 of Table A.1 implement this test. Column 2 includes export of a city through other ports as a control. Notice that $\sum_d v_{od}$ (total export of city o) is co-linear with $\sum_{d' \neq d} v_{od'}$ (the sum of export through other ports) if city-time fixed effects are included, and the model is not identified. Therefore, we control for province-time fixed effects instead, aiming to capture the overall export growth of a region that might be correlated with road connection.[3] The identifying assumption is that, to the extent that expressway expansion could be endogenous to the overall prospect of export growth in a city,

[3]More precisely, because we use the log of export, rather than the level of export, we can still include city-time fixed effects; however, the identification comes only from the difference between log and linear function forms.

once the major cities are excluded, such correlation is similar across smaller cities within a province and captured by the province-time fixed effects. We find that the coefficient for export through other ports is positive—inconsistent with a strong rerouting force. More importantly, the coefficient for regular-equivalent distance is -0.166, similar to the baseline estimate of -0.174. To rule out that such similarity is a coincidence under a different set of fixed effects from the baseline, Column 3 includes the province-time fixed effects as in Column 2 but excludes from independent variables the export through other ports. The estimated coefficient for regular-equivalent distance under this specification is similar.

While each of these two exercises requires stronger assumption than baseline specification—the first on the working of rerouting by geographic regions, the second on the identifying assumption—the consistent conclusion in both suggest that our finding is likely primarily driven by export growth, rather than rerouting.[4]

Robustness using sectoral level data. A remaining concern of the baseline specification is that it might be driven by changes in the sectoral composition of city export. For example, if as cities gain access to ports, they also become more specialized in export-intensive industries, such as textile, and if for some reason, exports in the textile industry is concentrated among the ports that experienced disproportionate increases in expressway connectivity to the hinterland, then the correlation between the shipment share and the bilateral connectivity will be picked up by our regressions. We note that if the expressway expansion is truly exogenous to non-major cities, then this concern does not pose a threat to the IV estimate. Nevertheless, in Columns 4 through 6, we use shipment value at the sectoral level for a robustness check. Column 4 includes city-time-sector, port-time-sector (letter 'i' in the row 'Fixed Effects' denote sectors), and Column 5 further add city-time-sector fixed effects. They show that, as in the baseline regressions, using over-time variation estimates a much smaller coefficient compared to using cross-sectional variation. Column 6 further confirms that expressways are less costly than regular roads.

Finally, we examine the importance of growth versus rerouting using sectoral data. This is useful because if rerouting takes place within a sector (i.e., exporters of cloth used to go through Shanghai, now switch to Guangzhou), then the previous exercises using more aggregate data might be too blunt to detect such patterns. Columns 7 through 9 revisit the exercises in Columns 1 through 3 of Table A.1. Column 7 uses data at city-port group-sector level, and finds slightly larger estimate than the baseline estimator of 0.138. Columns 8 and 9 further show that include export of a city through other ports do not have a big impact on the distance coefficient. Together, these results corroborate the earlier finding that differential route-specific export growth accounts for most of the estimated effects.

A.7 PPML Specifications and Comparison to OLS

An alternative specification used in estimating the trade or shipment elasticity is Poisson Pseudo-Maximum Likelihood, which can address biases arising from heteroskedasticity in the error term. In this appendix, we show that our main points are robust if PPML is used.

Table A.2 reports the results. Columns 1 through 4 vary the set of fixed effects included as in the first four columns of Table 2. Importantly, the difference between Columns 2 and 3 confirms that once city-

[4]A caveat in this conclusion is that during the sample period, China's export increased by five folds, and this could be part of the reason we find most effects were on the export growth margin.

Table A.2: Robustness using PPML

	(1)	(2)	(3)	(4)	(5)
$dist^t_{od}$	-0.670***	-0.706***	-0.350***	-0.488***	
	(0.039)	(0.042)	(0.054)	(0.077)	
-on express					-0.279***
					(0.096)
-on regular					-0.481***
					(0.071)
Fixed Effects	o, d, t	ot, dt	od, ot, dt	od, ot, dt	od, ot, dt
Exclude Major Cities				Yes	Yes
Observations	3668	3660	2838	2068	2038

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

port fixed effects are controlled for, the estimate for distance shrinks by half.[5] Column 5 further splits the total regular-equivalent distance into the length of regular road and expressway segments along the shortest path between o and d. The estimated coefficient for regular roads is higher than that for expressways, consistent with findings from Table 2.

PPML and OLS comparison. The PPML specifications reported above generally produce larger estimates than linear regressions. To see what leads to this result, consider the PPML specification, in which the data generating process is assumed to be $E(v^t_{od}|dist^t_{od}) = \exp(\gamma_1 dist^t_{od} + \beta^t_o + \tilde{\beta}^t_d + \beta_{od})$. In this specification, β^t_o and $\tilde{\beta}^t_d$ are city-time and port-time fixed effects, respectively. Under the assumption that v^t_{od} follows a Poisson distribution, the first order condition with respect to the estimator, $\hat{\gamma}_1$, to maximize the pseudo-likelihood function is:

$$\sum_{o,d,t} dist^t_{od} \cdot (v^t_{od} - \hat{v}^t_{od}) = 0,$$

i.e., the choice of $\hat{\gamma}_1$ is as if minimizing the distance-weighted sum of the *level* difference between v^t_{od} and its predicted values, $\hat{v}^t_{od}$. In an OLS specification, in contrast, the coefficient is chosen to satisfy the following first order condition, in order to minimize the sum of the errors in log values:

$$\sum_{o,d,t} dist^t_{od} \cdot (\log(v^t_{od}) - \log(\hat{v}^t_{od})) = 0.$$

Comparison between the two first order conditions shows that, because PPML minimizes the level difference whereas the OLS minimizes the percentage difference, the PPML effectively places more weights on observations with larger export values. As shown in Figure 2a of the text, the distance gradient for export is larger among the city-port pairs that are particularly close to each other. Since these city-port

[5]The regression sample shrinks as Column 3 requires a city-port pair to have non-zero values in both periods. Cross-sectional regressions using the same sample as in Column 3 gives an estimate of -0.715, so the difference between Columns 2 and 3 is not due to the change in sample, but rather the difference between the cross-sectional and within-pair estimators.

Table A.3: Alternative PPML Specifications

	(1)	(2)	(3)	(4)	(5)	(6)
	Replicate baseline		Exclude dist<100 km		Include zeros	
$dist_{od}^{t}$	-0.706***	-0.350***	-0.423***	-0.239***	-0.714***	-0.321***
	(0.042)	(0.054)	(0.032)	(0.039)	(0.043)	(0.050)
Observations	3660	2838	3542	2694	5852	4328
Fixed Effects	ot, dt	od, ot, dt	ot, dt	od, ot, dt	ot, dt	od, ot, dt

Notes: This table reports additional results using PPML. The dependent variable is total value of export from city o to the RoW through port d. Columns 1 and 2 reproduce Columns 2 and 3 of Table A.2 for ease of comparison; Columns 3 and 4 exclude city-port pairs that are less than 100 km apart; Columns 5 and 6 include observations with zero export values.
Standard errors are clustered at city-port level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

pairs are the ones with the highest export volume, PPML results in a larger estimated distance effect. Columns 3 and 4 of Table A.3 exclude city-port pairs that are less than 100 km apart. They show that, first, excluding these observations indeed brings the PPML estimate much closer to the OLS estimates. Second, that the overt-time estimate is significantly smaller than cross-sectional estimate continues to hold.

The role of zeros. To keep the specification consistent with OLS, we have excluded zeros in the PPML specifications. Columns 5 and 6 of Table A.3 show that this choice does not affect our estimates materially: including zeros increases the number of observations by around 50%, but the estimated coefficients are similar to those reported in Columns 1 and 2 of the table.

A.8 Measuring Shipment using Weights

Our estimation has used the value of goods to measure shipment, which is a commonly used measure in the trade literature. Below we show that all results are similar if we measure shipment by their weight, which is a theory-consistent measure when heterogeneity in transport costs is allowed.

Table A.4 replicates Table 2 with the log of shipment weight as the dependent variable. Although the coefficients change slightly compared to the baseline, the main points are robust: 1) the cross-sectional specifications overestimate the distance effect by as much as 100%; 2) the IV estimates are quantitatively similar to the OLS estimates; and 3) when both are included, the coefficient for regular roads is bigger than that for expressways.

A.9 Illustration of the IV

Figure A.3 shows the hypothetical expressway network used to construct the IV and illustrates the variation underlying the first stage regression. In the left panel, the blue lines indicate the actual expressways in 2010; the red lines indicate the minimum-length hypothetical expressways that connect all major cities. Like the actual network in 2010, the hypothetical network covers the entire country, but it consists of mostly straight lines, and is far less dense.

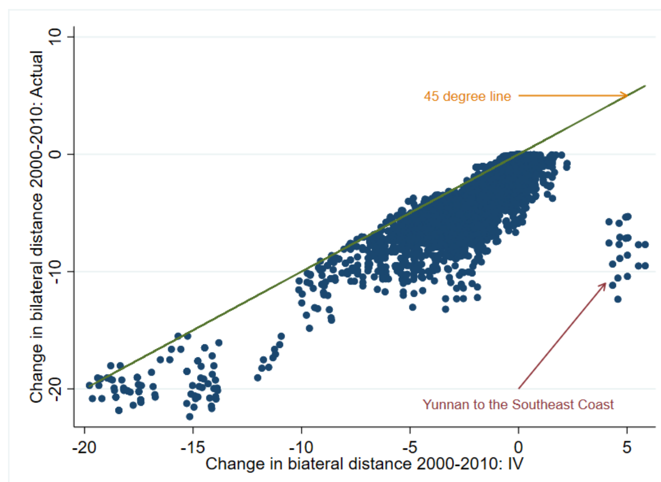
We use the hypothetical network to construct the IV for $dist_{od}^{2010}$ and use $dist_{od}^{2000}$ as an IV for itself. Given that the panel has exactly two periods, with city-port fixed effects controlled for, the first stage of the two-stage least square is essentially regressing the change in actual bilateral distance on the change predicted by the IV. Panel b of Figure A.3 illustrates the correlation between the two changes. The distance changes predicted by the IV are strongly correlated with the actual changes. The dots are

Table A.4: Robustness with Weight of Shipment as the Dependent Variable

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Effective Route Length and Export						By Type of Road		
	OLS				IV Reduced Form	2SLS	OLS	IV Reduced Form	2SLS
$dist_{od}^{t}$	-0.363***	-0.412***	-0.191***	-0.217***		-0.231***			
	(0.011)	(0.012)	(0.042)	(0.052)		(0.067)			
-on express							-0.088**		-0.163**
							(0.044)		(0.080)
-on regular							-0.215***		-0.248***
							(0.052)		(0.074)
IV $dist_{od}^{t}$					-0.268***				
					(0.079)				
-IV express								-0.161**	
								(0.063)	
-IV regular								-0.285***	
								(0.086)	
Fixed Effects	o, d, t	ot, dt	od, ot, dt	od, ot, dt	od, ot, dt	od, ot, dt	od, ot, dt	od, ot, dt	od, ot, dt
Exclude Major Cities				yes	yes	yes	yes	yes	yes
Observations	3612	3603	2786	2024	1996	1996	2024	1996	1996
R^2	0.606	0.680	0.893	0.884	0.884	0.023	0.884	0.884	0.018
First Stage K-P F Stat						1356.045			163.977

Notes: This table replicates Table 2 using the log of weight of export as the dependent variable. See notes under Table 2 for more information. Standard errors are clustered at city-port level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

(a) The Minimum Spanning 2010 Expressway Network

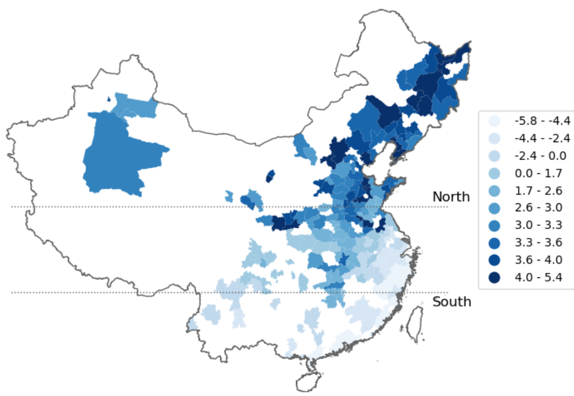


(b) First Stage in Differences

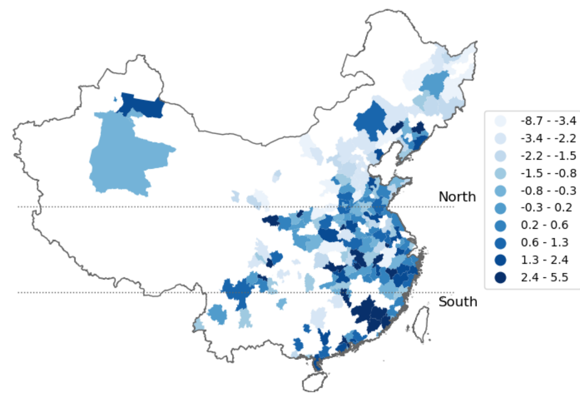
Figure A.3: Illustration of the IV

Note: The left panel overlays the minimum spanning tree (red) on the 2010 expressway network (blue); the right panel plots the first-stage regression.

mostly below the 45-degree line, reflecting that the minimum-spanning network is more sparse than the actual network. Finally, as annotated in the figure, the IV predicts that a small groups of cities—in the southwestern Yunnan Province—to have increased in distance to ports. This happens because the minimum-spanning network removes one road along the southern border of China that connects Yunnan to Guangzhou–according to the minimum-spanning algorithm, this link should not have been built.



(a) Relative Change in Port Access: North Minus South



(b) Relative Growth in Export: North Minus South

Figure A.4: Relative Growth in Export: North Minus South

Note: The figure covers only cities exporting through both groups of ports in both periods.

A.10 Visualizing Variation by Geographic Regions

Figure 2 presented in the text does not show the geographic dimension of the variation. In Figure A.4 we further show how each city' differential improvements in access to ports affect their choice of ports.

The left panel plots the *relative* reduction in cities' distances to two *groups* of ports due to the expressway construction during the decade. Focusing on the difference between the two port groups reduces the dimensionality of data from route-level to city-level, so the variation can be shown on a map. The first group comprises ports in the north (above the upper dotted line), while the second group comprises ports in the southeastern China (below the lower dotted line). For each city o, we first calculate its average regular-equivalent distance to ports in each of the two groups, denoted by $dist^t_{o,\text{North}}$ and $dist^t_{o,\text{South}}$, respectively. The left panel of Figure A.4 plots $(dist^{2010}_{o,\text{North}} - dist^{2000}_{o,\text{North}}) - (dist^{2010}_{o,\text{South}} - dist^{2000}_{o,\text{South}})$ by city. Dark colors indicate that an exporting city experienced a larger decrease in its distance to southern ports than to northern ports; light colors indicate the opposite. Northern cities tend to have much improved access to the ports in the south. Southern cities, which were already well connected to the southeastern coast before the massive expressway construction, experienced more substantial decreases in their distance to northern ports.

The right panel shows the relative growth in the export of city o through northern and southern ports, defined as $(\log(v^{2010}_{o,\text{North}}) - \log(v^{2000}_{o,\text{North}})) - (\log(v^{2010}_{o,\text{South}}) - \log(v^{2000}_{o,\text{South}}))$. Positive values indicate more rapid growth in export through northern ports than through southern ports; negative values indicate the opposite. The figure shows that southern cities saw more rapid export growth through northern ports than through southern ports. The two panels show that cities with dark colors on the right panel tend to have light colors in the left panel—the cities experiencing larger improvements in access to a port group also export more through that group, suggestive evidence that export routing responds to domestic transport infrastructure improvements.

The contrast between the two panels in Figure A.4 highlights that the variation among individual

Table A.5: Result from Within-Region Variation Only

	(1)	(2)	(3)	(4)	(5)	(6)
	Aggregate Data			Sectoral Data		
$dist_{od}^{t}$	-0.471***	-0.173***		-0.501***	-0.187***	
	(0.024)	(0.060)		(0.028)	(0.055)	
-on express			-0.058			-0.080*
			(0.047)			(0.043)
-on regular			-0.208***			-0.193***
			(0.066)			(0.055)
Fixed Effects	$ot, dt, r_o g_d t$	$ot, dt, r_o g_d t, od$	$ot, dt, r_o g_d t, od$	$oit, dit, r_o g_d t$	$oit, dit, r_o g_d t, odi$	$oit, dit, r_o g_d t, odi$
Observations	2752	2068	2068	20903	11682	11682
R^2	0.706	0.898	0.879	0.630	0.901	0.901

Notes: This table shows that using only variation within geographic regions gives similar point estimates and also leads to the conclusion that the cross-sectional estimate is larger than the over-time estimate. All specifications are estimated using OLS. In 'Fixed Effects', o, d, and t stand for origin city, port, and time, respectively; r_o stands for the big region that city o belong to and g_d stands for the geographic group of port d; i stands for sector. Standard errors are clustered at city-port level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

cities in their access to different groups of ports is one source of identification. If this is the only source of variation exploited, however, a potential concern is that the increasing connectedness between southern cities and northern ports, and between northern cities and southern ports, could be driven by other macroeconomic trends that increased overall connectedness between broad geographic regions within China. The IV might not fully eliminate this concern as the minimum spanning network is also designed to connect broad geographic regions.

To address this potential concern, we show in Appendix Table A.5 that there is also variation among cities and ports *within* broad regions, which our baseline estimate also exploits. To show this, all specifications in Table A.5 also control for $r(o) - g(d) - t$ fixed effects, in which $r(o)$ stands for the geographic region that city o belongs to, $g(d)$ stands for the group a port d belongs to, and t stands for time. The grouping of cities into regions is based on a geographic classification that splits China into seven regions: north, northeast, east, central, south, southwest, and northwest, each containing on average 5 provinces. The grouping of ports into geographic groups is the same as in Figure A.4, in which there are north, central, and south three port groups. These region-port group-time fixed effects absorb all variation from changes in overall connectedness between broad geographic regions in China. Once they are controlled for, identification comes from the variation in distance to port *within* pairs of broad geographic regions, arising from changes in local expressway connections due to the expansion.

Columns 1 through 3 use the aggregate data and show that focusing entirely on local variation tells a consistent story: that moving from cross-sectional to over-time estimate significantly reduces the point estimate, and that regular roads are more costly than expressways. The coefficients are also around the same magnitudes as the baseline estimate. Columns 4 through 6 use sectoral level data with fixed effects related to sectors controlled for. The conclusions are similar.

Figure A.4 and Table A.5 together demonstrate the two sources of variation exploited in the empirical exercise: one from changes in the connectedness between broad geographic regions; one from local variation in expressway access within individual geographic regions.

B Model

B.1 Deriving Equation (5)

We prove that when truck drivers choose from two networks (regular roads represented by $\mathbb{L}$ and expressways represented by $\mathbb{H}$), the expected trade cost can be derived based on the combined adjacency matrix $\mathbb{A} = \mathbb{L} + \mathbb{H}$. Moreover, the expected trade cost is given by Equation (5).

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

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

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

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

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

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

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

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

B.2 The Probability of Repeated Segments in a Route

The routing model in principle allows routes to have repeated segments. For example, a trucker going from Shanghai to Beijing would choose an Shanghai-Nanjing-Shanghai-Beijing trip with a positive probability. This choice is strictly dominated absent trucker heterogeneity, because it involves repeated use of the Shanghai-Nanjing segment.

Proposition B.1 derives the probability that a segment is being used more than once in a trip, conditional on it is being used. It shows that at the estimated range of our parameters, the probability of such event happening is very small.

Proposition B.1. *Denote $\pi^{kl,(n)}_{od}$ the fraction of ground-transported shipment between o and d that passes edge kl for n time(s). Then,*

$$\pi^{kl,(n)}_{od} = \frac{\tilde{b}_{ok}(a_{kl}\tilde{b}_{lk})^{n-1}a_{kl}\tilde{b}_{ld}}{b_{od}}, \forall n \geq 1,$$

where $\tilde{b}_{od}$ is the od-th element of $\tilde{\mathbb{B}}$, with $\tilde{\mathbb{B}} = (\mathbb{I} - \tilde{\mathbb{A}})^{-1}$ and $\tilde{\mathbb{A}}$ being equal to $\mathbb{A}$ except that the kl-th element is *set to zero, and a_{kl} is the kl-th element of $\mathbb{A}$.*

Proof. Denote P_{od} the set of all paths from o to d. Denote $P_{od,K}^{kl,(n)}$ the set of paths from o to d of K steps that passes edge kl for n time(s). Then, under the assumption that the path-specific dis-utility follows the Fréchet distribution, $\pi_{od}^{kl,(n)}$ satisfies

$$\pi_{od}^{kl,(n)} = \frac{1}{b_{od}} \sum_{K=0}^{\infty} \sum_{p \in P_{od,K}^{kl,(n)}} r_p,$$

where $p = (p_0 = o, p_1, ..., p_{K-1}, p_K = d)$ denotes a path of K steps, with its k-th element being the k-th node of the path, and

$$r_p \equiv \prod_{k=1}^{K} a_{p_{k-1},p_k},$$

with a_{kl} being the kl-th element of $\mathbb{A}$ and b_{od} being the od-th element of $\mathbb{B}$. Now consider $\pi_{od}^{kl,(1)}$, it can be written as

$$\pi_{od}^{kl,(1)} = \frac{1}{b_{od}} \sum_{K=0}^{\infty} \sum_{B_1=0}^{K-1} \sum_{B_2=0}^{K-1} \mathbb{1}_{B_1+B_2=K-1} \times (\sum_{p \in P_{ok,B_1}^{kl,not}} r_p) \times a_{kl} \times (\sum_{q \in P_{ld,B_2}^{kl,not}} r_q)$$

where $\mathbb{1}_{B_1+B_2=K-1}$ is an indicator function that takes one if and only if $B_1 + B_2 = K - 1$, and $P_{od,B}^{kl,not}$ denotes the set of paths from o to d of step B that *does not* pass edge kl. Presenting the summation in a compact form, we can write $\pi_{od}^{kl,(1)}$ as

$$\pi_{od}^{kl,(1)} = \frac{\tilde{b}_{ok} a_{kl} \tilde{b}_{ld}}{b_{od}}, \tag{B.1}$$

where $\tilde{b}_{od}$ is the od-th matrix of $\tilde{\mathbb{B}}$, with $\tilde{\mathbb{B}} = (\mathbb{I} - \tilde{\mathbb{A}})^{-1}$ and $\tilde{\mathbb{A}}$ being equal to $\mathbb{A}$ except that the kl-th element is set to zero. Intuitively, the numerator of (B.1) enumerates r_p for all paths p that first take an arbitrary number of steps to go from o to k without passing kl, then pass kl, and next take an arbitrary number of steps to go from l to d without passing kl.

Similarly, we can write $\pi_{od}^{kl,(2)}$ as

$$\begin{aligned}\pi_{od}^{kl,(2)} &= \frac{1}{b_{od}} \sum_{K=0}^{\infty} \sum_{B_1=0}^{K-2} \sum_{B_2=0}^{K-2} \sum_{B_3=0}^{K-2} \mathbb{1}_{B_1+B_2+B_3=K-2} \times (\sum_{p \in P_{ok,B_1}^{kl,not}} r_p) \times a_{kl} \times (\sum_{q \in P_{lk,B_2}^{kl,not}} r_q) \times a_{kl} \times (\sum_{s \in P_{ld,B_3}^{kl,not}} r_s) \\ &= \frac{\tilde{b}_{ok} a_{kl} \tilde{b}_{lk} a_{kl} \tilde{b}_{ld}}{b_{od}}.\end{aligned}$$

Similarly, we can show that

$$\pi_{od}^{kl,(n)} = \frac{\tilde{b}_{ok} (a_{kl} \tilde{b}_{lk})^{n-1} a_{kl} \tilde{b}_{ld}}{b_{od}}, \forall n \geq 1.$$

□

We now apply Proposition B.1 to calculate the probability of passing an edge kl more than once, conditional on passing kl, denoted by Q_{od}^{kl},

$$
\begin{aligned}
Q_{od}^{kl} &= \frac{\sum_{m=2}^{\infty} \pi_{od}^{kl,(m)}}{\sum_{m=1}^{\infty} \pi_{od}^{kl,(m)}} \\
&= \frac{\sum_{m=2}^{\infty} (a_{kl}\tilde{b}_{lk})^{m-1}}{\sum_{m=1}^{\infty} (a_{kl}\tilde{b}_{lk})^{m-1}} = a_{kl}\tilde{b}_{lk} \equiv Q^{kl},
\end{aligned}
$$

which is irrelevant of od.

We calculate $Q^{kl}, \forall kl$ that forms an edge (i.e., $a_{kl} > 0$). Table B.1 reports the distribution of Q^{kl} across all edges. As shown, at the calibrated $\theta = 111.5$, the mean of Q^{kl} is 0.3%, the median is less than 0.1%, and the 95th percentile is only around 1.2%. This shows that the likelihood for passing an edge more than once is rather low. Other things equal, increasing θ further lowers the likelihood for repetitive passing and lowering θ increases the likelihood. But overall, the likelihood remains quite low for the range of θ estimated.

Table B.1: Conditional Probability of Passing an Edge More than Once

θ	Min	Max	Mean	Median	p95	p99	p99.9
80	0.000	0.768	0.017	0.000	0.080	0.352	0.767
111.5	0.000	0.332	0.003	0.000	0.012	0.054	0.324
200	0.000	0.045	0.000	0.000	0.000	0.002	0.043

B.3 Definition of Equilibrium

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

Definition 1. *Given fundamentals* $\{\tilde{\tau}_{od}^i, T_d^i, B_d, \overline{H}_d, L_{CHN}, L_{RoW}\}$,[6] *a competitive equilibrium is: consumer utility* U_d*, consumption of land* H_d *and sectoral final goods* C_d^i*, labor allocations* L_d *and* l_d^i*, quantities of sectoral final goods used as intermediate input* m_d^{ij}*, quantities of sectoral final goods produced* Q_d^i*, quantity of intermediate goods traded* $\tilde{q}_{od}^i$*, quantity of intermediate goods produced* q_d^i*, lump-sum transfers for domestic regions and RoW* (Tr, Tr_{RoW})*, rental prices of land* R_d*, prices of final goods* P_d^i*, import prices of intermediate goods* p_{od}^i*, unit production costs of intermediate goods* c_o^i*, and wages* w_d*, s.t.*

- *Consumers' optimality conditions hold:*

$$
\begin{aligned}
U_d &= B_d [H_d]^{\alpha^0} \prod_{i=1}^{S} [C_d^i]^{\alpha^i}, \\
\alpha^0 I_d &= R_d H_d, \\
\alpha^i I_d &= P_d^i C_d^i,
\end{aligned}
$$

where $I_d = w_d + Tr$, $\forall d \in CHN$ *and* $I_{RoW} = w_{RoW} + Tr_{RoW}$.

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

- *Intermediate good producers' optimality conditions hold:*

$$\begin{aligned}
q_d^i &= T_d^i [l_d^i]^{\beta^i} \prod_{j=1}^{S} [m_d^{ij}]^{\gamma^{ij}},\\
c_d^i &= \kappa^i w_d^{\beta^i} \prod_{j=1}^{S} [P_d^j]^{\gamma^{ij}} / T_d^i,\\
P_d^j m_d^{ij} &= \gamma^{ij} c_d^i q_d^i,\\
w_d l_d^i &= \beta^i c_d^i q_d^i,\\
p_{od}^i &= c_o^i \tilde{\tau}_{od}^i,
\end{aligned}$$

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

- *Final good producers' optimality conditions hold:*

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

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

$$\begin{aligned}
\sum_i l_d^i &= L_d, && \textit{(Labor markets clear)}\\
H_d L_d &= \overline{H}_d, && \textit{(Land markets clear)}\\
\sum_d \tilde{\tau}_{od}^i \tilde{q}_{od}^i &= q_o^i, && \textit{(Intermediate good markets clear)}\\
\sum_i m_d^{ij} + C_d^j L_d &= Q_d^j. && \textit{(Final good markets clear).}
\end{aligned}$$

- *Rents from land are rebated via lump-sum transfers:*

$$\sum_{d \in CHN} R_d H_d L_d = Tr \cdot L_{CHN},$$
$$R_{RoW} H_{RoW} L_{RoW} = Tr_{RoW} \cdot L_{RoW}.$$

- *Domestic workers are mobile:*

$$U_d = U_{d'}, \forall d, d' \in CHN.$$
$$\sum_{d \in CHN} L_d = L_{CHN}.$$

We also state the definitions of other equilibrium objects used in the main text and the appendix, that can be written *as functions of the equilibrium objects defined above.*

- *The total expenditure on intermediate goods in sector i of region d*

$$E_d^i \equiv P_d^i Q_d^i.$$

- *The value of trade flows from o to d in sector i*

$$X_{od}^i \equiv p_{od}^i \tilde{q}_{od}^i = E_d^i \pi_{od}^i,$$

where $\pi_{od}^i \equiv [\frac{p_{od}^i}{P_d^i}]^{1-\sigma}$.

B.4 Proof of Lemma 1 and Proposition 1

We first state a lemma characterizing the first order effect of the segment shipment cost on the trade cost.

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

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

where a_{kl} *and* b_{od} *are the kl-th and od-th elements of* $\mathbb{A}$ *and* $\mathbb{B}$*, respectively.*

Proof. Apply the the formula for the derivative of the inverse of a matrix we have

$$\begin{aligned}\frac{\partial \mathbb{B}}{\partial \log a_{kl}} &= -(\mathbb{I} - \mathbb{A})^{-1} \frac{\partial (\mathbb{I} - \mathbb{A})}{\partial \log a_{kl}} (\mathbb{I} - \mathbb{A})^{-1} \\ &= \mathbb{B}(\mathbb{E}_{kl} \circ \mathbb{A})\mathbb{B},\end{aligned}$$

where $\mathbb{E}_{kl}$ is a matrix of the same size as $\mathbb{A}$, with the kl-th element being one and other elements being zero. Therefore,

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

□

Denote $\tilde{\pi}_{od}^{kl} \equiv \frac{b_{ok} \cdot a_{kl} \cdot b_{ld}}{b_{od}}$, we prove the following lemma.

Lemma B.2. *With* $\tilde{\pi}_{od}^{kl}$ *defined above, we have the following*

(1) $$\frac{\partial \log \tilde{\tau}_{od}^i}{\partial \log \iota_{kl}} = \pi_{od}^{road} \tilde{\pi}_{od}^{kl},$$

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

Proof. First, from

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

we have

$$\frac{\partial \log \tilde{\tau}^i_{od}}{\partial \log \tau^i_{od}} = \pi^{road}_{od} = \frac{(\tau^i_{od})^{-\theta_M}}{(\bar{\tau}^i_{od})^{-\theta_M} + (\tau^i_{od})^{-\theta_M}}. \tag{B.2}$$

Next, recall that $\tau^i_{od} = (\frac{h_i}{h_0})^{\mu} b^{-1/\theta}_{od}$ and $\iota_{kl} = a^{-1/\theta}_{kl}$. Applying Lemma B.1, we have

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

This proves part (1).

Now from Equation (B.2), we have

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

Combining with

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

we have

$$\frac{\partial \log(\pi^{road}_{od})}{\partial \log \iota_{k'l'}} = -\theta_M(1 - \pi^{road}_{od})\tilde{\pi}^{k'l'}_{od}. \tag{B.3}$$

Next start with the definition of $\tilde{\pi}^{kl}_{od}$ we have,

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

Take derivative with respect to $\log \iota_{k'l'}$ and apply Lemma B.1, we have

$$\frac{\partial \log \tilde{\pi}^{kl}_{od}}{\partial \log \iota_{k'l'}} = -\theta(\tilde{\pi}^{k'l'}_{ok} + \tilde{\pi}^{k'l'}_{ld} - \tilde{\pi}^{k'l'}_{od} + \mathbb{1}(kl = k'l')). \tag{B.4}$$

Combining (B.3) and (B.4) we arrive at

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

This proves part (2). □

We are now ready to prove Lemma 1.

Proof for Lemma 1.

Proof. With Lemma B.1, it suffices to prove that $\tilde{\pi}_{od}^{kl}$ converges to the fraction of ground-transported shipments between o and d that passes edge kl, denoted by π_{od}^{kl}, as θ goes to infinity. In fact, we prove a stronger result below:

$$\lim_{\theta\to\infty} \frac{\pi_{od}^{kl} - \tilde{\pi}_{od}^{kl}}{\pi_{od}^{kl}} = 0, \text{ if } \pi_{od}^{kl} > 0.$$

That is, not only the level error but also the relative error converges to zero.[7]

First, under the assumption that the path-specific dis-utility follows the Fréchet distribution, π_{od}^{kl} satisfies

$$\pi_{od}^{kl} = \frac{1}{b_{od}} \sum_{K=0}^{\infty} \sum_{p\in P_{od,K}^{kl}} \prod_{k=1}^{K} a_{p_{k-1},p_k}, \tag{B.5}$$

where $p = (p_0 = o, p_1, ..., p_{K-1}, p_K = d)$ denotes a path of K steps, with its k-th element being the k-th node of the path, and $P_{od,K}^{kl}$ denotes the set of paths of K steps that pass edge kl. We partition $P_{od,K}^{kl}$ into two disjoint sets: $\overline{P}_{od,K}^{kl}$ and $\tilde{\mathcal{P}}_{od,K}^{kl}$, where $\overline{P}_{od,K}^{kl}$ denotes the subsets of paths that pass edge kl *only once*, and $\tilde{\mathcal{P}}_{od,K}^{kl}$ denotes the subsets of paths that pass edge kl *more than once*. Given any $\bar{p} \in \tilde{\mathcal{P}}_{od,\bar{K}}^{kl}$, there exists $K' < \bar{K}$ and $p' \in \overline{P}_{od,K'}^{kl}$ for which p' is the path removing any loops in $\bar{p}$ that involve multiple passes of kl. Denote $\bar{a} = \max_{kl} a_{kl}$, then $\lim_{\theta\to\infty} \bar{a} = 0$ since $a_{kl} = \exp(-\kappa^L \theta \text{dist}_{kl}^L) + \exp(-\kappa^H \theta \text{dist}_{kl}^H)$. Therefore,

$$\frac{\prod_{k=1}^{\bar{K}} a_{\bar{p}_{k-1},\bar{p}_k}}{\sum_{K=0}^{\infty} \sum_{p\in P_{od,K}^{kl}} \prod_{k=1}^{K} a_{p_{k-1},p_k}} \le \frac{\prod_{k=1}^{\bar{K}} a_{\bar{p}_{k-1},\bar{p}_k}}{\prod_{k=1}^{K'} a_{p'_{k-1},p'_k}} \le \bar{a} \to 0, \tag{B.6}$$

as $\theta \to \infty$, where the first inequality shrinks the positive denominator, and the second inequality applies that p' is a path removing loops in $\bar{p}$ (i.e., $\bar{p}$ contains all segments in p' and additional detoured segments). Since $\cup_{\bar{K}=0}^{\infty} \tilde{\mathcal{P}}_{od,\bar{K}}^{kl}$ is a countable set, multiply (B.6) by $\bar{K}$ and sum over all $\bar{p} \in \cup_{\bar{K}=0}^{\infty} \tilde{\mathcal{P}}_{od,\bar{K}}^{kl}$, we have

$$\lim_{\theta\to\infty} \sum_{\bar{K}=0}^{\infty} \sum_{\bar{p}\in\tilde{\mathcal{P}}_{od,\bar{K}}^{kl}} \frac{\bar{K} \cdot \prod_{k=1}^{\bar{K}} a_{\bar{p}_{k-1},\bar{p}_k}}{\sum_{K=0}^{\infty} \sum_{p\in P_{od,K}^{kl}} \prod_{k=1}^{K} a_{p_{k-1},p_k}} = 0. \tag{B.7}$$

Now consider the summation

$$\varsigma_{od}^{kl} \equiv \sum_{K=0}^{\infty} \sum_{B=0}^{K-1} \left(\sum_{p\in P_{ok,B}} \prod_{k=1}^{B} a_{p_{k-1},p_k} \right) \times a_{kl} \times \left(\sum_{q\in P_{kd,K-B-1}} \prod_{k=1}^{K-B-1} a_{q_{k-1},q_k} \right).$$

Then the cost of a path $\bar{p}$ of $\bar{K}$ steps that passes edge kl for $n \le \bar{K}$ time(s), $\prod_{k=1}^{\bar{K}} a_{\bar{p}_{k-1},\bar{p}_k}$, will appear in the

[7] The stronger result allows $\pi_{od}^{kl} \to 0$ as $\theta \to \infty$.

above summation for exactly n times. Therefore,

$$\frac{\varsigma_{od}^{kl}}{b_{od}} \geq \pi_{od}^{kl} = \frac{1}{b_{od}} \sum_{K=0}^{\infty} \sum_{p \in P_{od,K}^{kl}} \prod_{k=1}^{K} a_{p_{k-1},p_k}$$
$$\geq \frac{1}{b_{od}} \left\{ \varsigma_{od}^{kl} - \sum_{\bar{K}=0}^{\infty} \sum_{\bar{p} \in \tilde{\mathcal{P}}_{od,\bar{K}}^{kl}} \bar{K} \cdot \prod_{k=1}^{\bar{K}} a_{\bar{p}_{k-1},\bar{p}_k} \right\},$$

where the first inequality applies that paths in $\tilde{\mathcal{P}}_{od,\bar{K}}^{kl}$ appear in the summation for more than one time, and the second inequality applies that paths in $\tilde{\mathcal{P}}_{od,\bar{K}}^{kl}$ appear in the summation no more than $\bar{K}$ times.

Notice that

$$\frac{\varsigma_{od}^{kl}}{b_{od}} = \frac{1}{b_{od}} \sum_{K=0}^{\infty} \sum_{B=0}^{K-1} \left(\sum_{p_1=1}^{N} \cdots \sum_{p_{B-1}=1}^{N} a_{o,p_1} \times \ldots \times a_{p_{B-1},k} \right) \times a_{kl} \times \left(\sum_{q_1=1}^{N} \cdots \sum_{q_{K-B-1}=1}^{N} a_{l,q_1} \times \ldots \times a_{q_{K-B-1},d} \right)$$
$$= \frac{1}{b_{od}} \sum_{K=0}^{\infty} \sum_{B=0}^{K-1} \mathbb{A}_{ok}^{B} \times a_{kl} \times \mathbb{A}_{ld}^{K-B-1} = \frac{b_{ok} a_{kl} b_{ld}}{b_{od}} = \tilde{\pi}_{od}^{kl}.$$

Therefore,

$$\tilde{\pi}_{od}^{kl} \geq \pi_{od}^{kl} \geq \tilde{\pi}_{od}^{kl} - \frac{\sum_{\bar{K}=0}^{\infty} \sum_{\bar{p} \in \tilde{\mathcal{P}}_{od,\bar{K}}^{kl}} \bar{K} \cdot \prod_{k=1}^{\bar{K}} a_{\bar{p}_{k-1},\bar{p}_k}}{b_{od}}$$
$$\Rightarrow 0 \geq \frac{\pi_{od}^{kl} - \tilde{\pi}_{od}^{kl}}{\pi_{od}^{kl}} \geq -\frac{\sum_{\bar{K}=0}^{\infty} \sum_{\bar{p} \in \tilde{\mathcal{P}}_{od,\bar{K}}^{kl}} \bar{K} \cdot \prod_{k=1}^{\bar{K}} a_{\bar{p}_{k-1},\bar{p}_k}}{\sum_{K=0}^{\infty} \sum_{p \in P_{od,K}^{kl}} \prod_{k=1}^{K} a_{p_{k-1},p_k}}$$
$$\Rightarrow 0 \geq \lim_{\theta \to \infty} \frac{\pi_{od}^{kl} - \tilde{\pi}_{od}^{kl}}{\pi_{od}^{kl}} \geq -\lim_{\theta \to \infty} \frac{\sum_{\bar{K}=0}^{\infty} \sum_{\bar{p} \in \tilde{\mathcal{P}}_{od,\bar{K}}^{kl}} \bar{K} \cdot \prod_{k=1}^{\bar{K}} a_{\bar{p}_{k-1},\bar{p}_k}}{\sum_{K=0}^{\infty} \sum_{p \in P_{od,K}^{kl}} \prod_{k=1}^{K} a_{p_{k-1},p_k}} = 0,$$

where the second row applies (B.5) and that $\pi_{od}^{kl} > 0$, and the last equality applies (B.7). This completes the proof for $\lim_{\theta \to \infty} \frac{\pi_{od}^{kl} - \tilde{\pi}_{od}^{kl}}{\pi_{od}^{kl}} = 0$. □

To prove Proposition 1, we first prove the following series of lemmas. Define world welfare $\log \overline{W} \equiv \log W + \omega_{RoW} \log U_{RoW}$ with the Pareto weight $\omega_{RoW} = Y_{RoW}/Y$, where Y_{RoW} and Y are the RoW GDP and domestic GDP under the competitive equilibrium, respectively. Define Ω the expanded input-output matrix evaluated at the competitive equilibrium, encompassing all domestic and foreign final good producers, intermediate good producers, and traders.[8] Specifically, the kj-th entry of Ω is defined as

$$\Omega_{kj} \equiv \frac{\hat{p}_j \hat{q}_{kj}}{S_k},$$

where $\hat{p}_j$ is the price of good j, $\hat{q}_{kj}$ is the quantity of good j used by sector k, and S_k is the total revenue in sector k. Stack the productivities of all final-good producers, intermediate good producers, and traders

[8] Traders from place o to place d in sector i competitively convert intermediate goods of (o, i) to intermediate goods of (d, i). With such interpretations the iceberg trade costs are the inverse of the productivities of traders.

into a vector denoted by A, and the corresponding price into a vector denoted by $\hat{p}$. Define χ_j the total sales of sector j as the share of domestic GDP:

$$\chi_j \equiv \frac{\sum_k \hat{p}_j \hat{q}_{kj}}{Y}.$$

Lemma B.3 below associates the first-order effect of sectoral productivity on welfare with sectoral sales share, a result extending Hulten (1978) with international trade and domestic mobile labor.

Lemma B.3. *With $\overline{W}$, A_j and χ_j defined above,*

$$\frac{\mathrm{d}\log \overline{W}}{\mathrm{d}\log A_j} = \chi_j.$$

Proof. By Shephard's lemma,

$$\mathrm{d}\log(\hat{p}) = \Omega \mathrm{d}\log(\hat{p}) + \beta \mathrm{d}\log(w) - \mathrm{d}\log(A),$$

where β is the vector stacking the labor shares of all expanded sectors. So we have

$$\mathrm{d}\log(\hat{p}) = (\mathbb{I} - \Omega)^{-1}\Big(\beta \mathrm{d}\log(w) - \mathrm{d}\log(A)\Big). \tag{B.8}$$

Starting from the consumer utility in region d, $U_d = B_d \frac{I_d}{P_d}$, we have

$$\mathrm{d}\log U_d = \mathrm{d}\log(I_d) - \sum_{i=1}^{S} \alpha^i \mathrm{d}\log(P_d^i) - \alpha^0 \mathrm{d}\log(R_d), \forall d.$$

Combining the optimal housing expenditure and the housing market clearing conditions:

$$\alpha^0 I_d L_d = R_d \overline{H}_d, \forall d,$$

we have

$$\mathrm{d}\log U_d = (1-\alpha^0)\mathrm{d}\log I_d - \sum_{i=1}^{S} \alpha^i \mathrm{d}\log(P_d^i) - \alpha^0 \mathrm{d}\log(L_d). \tag{B.9}$$

Define Ξ_d a column vector, of which the k-th entry equals α^i if the k-th producer among all extended producers is the final good producer of sector i of region d. That is, Ξ_d maps final good producers of region d to the index of extended producers. With such definitions, (B.9) can be written as

$$\mathrm{d}\log U_d = (1-\alpha^0)\mathrm{d}\log I_d - \Xi_d' \mathrm{d}\log(\hat{p}) - \alpha^0 \mathrm{d}\log(L_d).$$

Plugging in (B.8) we have

$$\mathrm{d}\log U_d = (1-\alpha^0)\mathrm{d}\log I_d - \Xi_d'(\mathbb{I} - \Omega)^{-1}\Big(\beta \mathrm{d}\log(w) - \mathrm{d}\log(A)\Big) - \alpha^0 \mathrm{d}\log(L_d). \tag{B.10}$$

With domestic GDP $Y = \sum_{d \neq RoW} I_d L_d$ as the numeraire, multiply (B.10) by $I_d \cdot L_d$ both sides and sum over d, and notice $U_d = W, \forall d \neq RoW$ by the domestic welfare equalization condition, we have

$$\mathrm{d}\log W + \frac{Y_{RoW}}{Y}\mathrm{d}\log U_{RoW} = \sum_d \left[(1-\alpha^0)I_d L_d \mathrm{d}\log I_d - I_d L_d \Xi_d'(\mathbb{I}-\Omega)^{-1}\Big(\beta \mathrm{d}\log(w) - \mathrm{d}\log(A)\Big)\right]$$
$$- \alpha^0 \sum_d \mathrm{d}\log(L_d) I_d L_d. \tag{B.11}$$

The labor market clearing conditions imply:

$$\sum_d w_d L_d \mathrm{d}\log w_d - \sum_d I_d L_d \Xi_d'(\mathbb{I}-\Omega)^{-1}\beta \mathrm{d}\log(w) = 0. \tag{B.12a}$$

The normalization $Y = \sum_{d \neq RoW} I_d L_d = 1$ implies

$$\sum_{d \neq RoW} [I_d L_d \mathrm{d}\log I_d + \mathrm{d}\log(L_d) I_d L_d] = 0. \tag{B.12b}$$

The determination of domestic transfer and normalization implies

$$Tr = \frac{\alpha^0 \sum_{d \neq RoW} I_d L_d}{\sum_{d \neq RoW} L_d} \Rightarrow \mathrm{d}Tr = 0$$
$$\Rightarrow \mathrm{d}I_d = \mathrm{d}(w_d + Tr) = \mathrm{d}w_d, \forall d \neq ROW$$
$$\Rightarrow I_d L_d \mathrm{d}\log I_d = w_d L_d \mathrm{d}\log w_d, \forall d \neq ROW. \tag{B.12c}$$

The determination of RoW transfer implies

$$Tr_{RoW} = \alpha^0 I_{RoW} L_{RoW}, \quad I_{RoW} = Tr_{RoW} + w_{RoW}$$
$$\Rightarrow I_{RoW} = \frac{w_{RoW}}{1-\alpha^0} \Rightarrow \mathrm{d}I_{RoW} = \frac{1}{1-\alpha^0}\mathrm{d}w_{RoW}. \tag{B.12d}$$

Plugging (B.12a)-(B.12d) and $\mathrm{d}L_{RoW} = 0$ to (B.11) we have

$$\mathrm{d}\log W + \frac{Y_{RoW}}{Y}\mathrm{d}\log U_{RoW} = \sum_d I_d L_d \Xi_d'(\mathbb{I}-\Omega)^{-1}\mathrm{d}\log(A)$$
$$= \sum_j \chi_j \mathrm{d}\log(A_j).$$

□

Notice that the productivity of a trader from o to d in sector i is the inverse of the trade cost from o to d in sector i. And the sales of the trader is the sales of intermediate goods from o to d in sector i. Therefore, we have

Corollary 1. *With $\overline{W}$ defined above,*

$$\frac{\mathrm{d}\log \overline{W}}{\mathrm{d}\log \tilde{\tau}_{od}^i} = -\frac{X_{od}^i}{Y}.$$

We next characterize the exposure of RoW consumption to RoW import prices.

Lemma B.4. *Assume* $\mathrm{d}\log T^i_{RoW} = 0$ *(i.e., there is no change in RoW productivity). Then*

$$\mathrm{d}\log U_{RoW} = -\sum_{o,i} \frac{\Lambda^i_o}{Y_{RoW}} \mathrm{d}\log(p^i_{o,RoW}/Y_{RoW}),$$

where $\Lambda^i_o = Y_{RoW}[\alpha' \left(\mathbb{I} - \hat{\Omega}\right)^{-1}]_i \pi^i_{o,RoW}$*, in which* $\alpha' = (\alpha^1, \alpha^2, ..., \alpha^S), \hat{\Omega}_{ij} = \gamma^{ij}\pi^i_{RoW,RoW}$*, and* $[x]_i$ *is the i-th element of row vector x.*

Proof. Denote $\hat{x} \equiv x/Y_{RoW}$, for $x = (P^i_{RoW}, c^i_{Row}, p^i_{o,RoW}, w_{RoW}, R_{RoW}, I_{RoW})$. Since $Y_{RoW} = I_{RoW}L_{RoW} = \frac{w_{RoW}L_{RoW}}{1-\alpha^0}$, and L_{RoW} is fixed, we have

$$\mathrm{d}\log \hat{w}_{RoW} = \mathrm{d}\log \hat{I}_{RoW} = 0.$$

By Shephard's lemma,

$$\mathrm{d}\log \hat{P}^i_{RoW} = \pi^i_{RoW,RoW}\mathrm{d}\log \hat{c}^i_{RoW} + \sum_{o'} \pi^i_{o',RoW}\mathrm{d}\log \hat{p}^i_{o',RoW}, \tag{B.13}$$

and since $\mathrm{d}\log T^i_{RoW} = 0$, we have

$$\mathrm{d}\log \hat{c}^i_{RoW} = \sum_j \gamma^{ij}\mathrm{d}\log \hat{P}^j_{RoW} + \beta^i \mathrm{d}\log \hat{w}_{RoW}. \tag{B.14}$$

Plugging (B.14) into (B.13) and applying $\mathrm{d}\log \hat{w}_{RoW} = 0$, in matrix form we have

$$\mathrm{d}\log \hat{P}_{RoW} = (\mathbb{I} - \hat{\Omega})^{-1}\Pi, \tag{B.15}$$

where $\log \hat{P}_{RoW} = (\log \hat{P}^i_{RoW}, ..., \log \hat{P}^S_{RoW})'$, and Π is a column vector with $\Pi_i = \sum_{o'} \pi^i_{o',RoW}\mathrm{d}\log \hat{p}^i_{o',RoW}$. Plug (B.15) to

$$\mathrm{d}\log U_{RoW} = \mathrm{d}\log \hat{I}_{RoW} - \alpha^0 \mathrm{d}\log \hat{R}_{RoW} - \alpha'\mathrm{d}\log \hat{P}_{RoW},$$

and note that $\mathrm{d}\log(\hat{R}_{RoW}) = 0$ since $R_{RoW}\overline{H}_{RoW} = \alpha^0 Y_{RoW}$ and $\overline{H}_{RoW}$ is fixed, we have the desired result. □

Proof of Proposition 1

Proof. Combine Corollary 1 and Lemma B.4 we arrive at

$$\frac{\mathrm{d}\log W}{\mathrm{d}\log \tilde{\tau}^i_{od}} = -\frac{X^i_{od}}{Y} - \frac{Y_{RoW}}{Y}\frac{\mathrm{d}\log U_{RoW}}{\mathrm{d}\log \tilde{\tau}^i_{od}},$$

where

$$\frac{\mathrm{d}\log U_{RoW}}{\mathrm{d}\log \tilde{\tau}^i_{od}} = -\sum_{o',i'} \frac{\Lambda^{i'}_{o'}}{Y_{RoW}} \frac{\mathrm{d}\log[p^{i'}_{o',RoW}/Y_{RoW}]}{\mathrm{d}\log \tilde{\tau}^i_{od}}.$$

Notice that $p^{i'}_{o',RoW} = \tilde{\tau}^{i'}_{o',RoW} \cdot c^{i'}_{o'}$, so

$$\frac{\mathrm{d}\log[p^{i'}_{o',RoW}/Y_{RoW}]}{\mathrm{d}\log\tilde{\tau}^{i}_{od}} = \mathbb{1}(i'=i, o'=o, d=RoW) + \frac{\mathrm{d}\log[c^{i'}_{o'}/Y_{RoW}]}{\mathrm{d}\log\tilde{\tau}^{i}_{od}},$$

and

$$\frac{\mathrm{d}\log U_{RoW}}{\mathrm{d}\log\tilde{\tau}^{i}_{od}} = -\sum_{o',i'}\frac{\Lambda^{i'}_{o'}}{Y_{RoW}}\frac{\mathrm{d}\log[p^{i'}_{o',RoW}/Y_{RoW}]}{\mathrm{d}\log\tilde{\tau}^{i}_{od}} = -\frac{\Lambda^{i}_{o}}{Y_{RoW}}\mathbb{1}_{d=RoW} - \sum_{o',i'}\frac{\Lambda^{i'}_{o'}}{Y_{RoW}}\frac{\mathrm{d}\log[c^{i'}_{o'}/Y_{RoW}]}{\mathrm{d}\log\tilde{\tau}^{i}_{od}}.$$

Therefore,

$$\frac{\mathrm{d}\log W}{\mathrm{d}\log\tilde{\tau}^{i}_{od}} = -\Big(\frac{X^{i}_{od}}{Y} - \frac{\Lambda^{i}_{o}}{Y}\mathbb{1}_{d=RoW}\Big) + \sum_{o',i'}\frac{\Lambda^{i'}_{o'}}{Y}\frac{\mathrm{d}\log[c^{i'}_{o'}/Y_{RoW}]}{\mathrm{d}\log\tilde{\tau}^{i}_{od}}.$$

Apply the first order Taylor expansion of $\log W$ with respect to $(\log\tilde{\tau}^{i}_{od})_{o,d,i}$, we have

$$\Delta\log W = -\sum_{o,d,i}\Big(\frac{X^{i}_{od}}{Y} - \frac{\Lambda^{i}_{o}}{Y}\mathbb{1}_{d=RoW}\Big)\Delta\log\tilde{\tau}^{i}_{od} + TOT + HO_T, \tag{B.16}$$

where $TOT = \sum_{o,d,i}\Big(\sum_{o',i'}\frac{\Lambda^{i'}_{o'}}{Y}\frac{\mathrm{d}\log[c^{i'}_{o'}/Y_{RoW}]}{\mathrm{d}\log\tilde{\tau}^{i}_{od}}\Big)\Delta\log\tilde{\tau}^{i}_{od}$, is the first order terms-of-trade effects, and HO_T is the higher order effect of trade costs on welfare. Further apply Taylor expansion of $\Delta\log\tilde{\tau}^{i}_{od}$ with respect to $(\Delta\log\iota_{kl})_{kl\in\mathbb{C}}$, we have

$$\Delta\log\tilde{\tau}^{i}_{od} = \sum_{kl\in\mathbb{C}}\frac{\partial\log\tilde{\tau}^{i}_{od}}{\partial\log\iota_{kl}}\Delta\log(\iota_{kl}) + \frac{1}{2}\sum_{kl\in\mathbb{C}}\sum_{k'l'\in\mathbb{C}}\frac{\partial^2\log\tilde{\tau}^{i}_{od}}{\partial\log\iota_{kl}\,\partial\log\iota_{k'l'}}\Delta\log(\iota_{kl})\Delta\log(\iota_{k'l'}) + \widetilde{HO}_R, \tag{B.17}$$

where $\widetilde{HO}_R$ is the effect of route costs on trade costs beyond the second order effect. Plugging (B.17) to (B.16), we have the desired result. □

B.5 Interaction Among Routes

We characterize the interaction among different routes and illustrate it through an example in the calibrated model.

We can view a large project as a *collection* of expressway segments. Proposition 1 shows that the second order effect from rerouting, denoted by SO_R, is

$$SO_R = -\frac{1}{2}\sum_{o,d,i}\Big(\frac{X^{i}_{od}}{Y} - \frac{\Lambda^{i}_{o}}{Y}\mathbb{1}_{d=RoW}\Big)\sum_{kl\in\mathbb{C}}\sum_{k'l'\in\mathbb{C}}\frac{\partial^2\log\tilde{\tau}^{i}_{od}}{\partial\log\iota_{kl}\,\partial\log\iota_{k'l'}}\Delta\log(\iota_{kl})\Delta\log(\iota_{k'l'}),$$

in which $\sum_{kl\in\mathbb{C}}\sum_{k'l'\in\mathbb{C}}\frac{\partial^2\log\tilde{\tau}^{i}_{od}}{\partial\log\iota_{kl}\,\partial\log\iota_{k'l'}}$ captures the own second-order effect (when $kl = k'l'$) as well as potential complementary and substitution among different routes (when $kl \neq k'l'$).

To see what the interaction effect entials, we use Lemma 1 to write the cross-derivative term in the summand as:

$$\frac{\partial^2 \log \tilde{\tau}^i_{od}}{\partial \log \iota_{kl} \partial \log \iota_{k'l'}} \approx \pi^{road}_{od} \pi^{kl}_{od} \Big\{ \underbrace{-\theta[\mathbb{1}(kl = k'l') + \pi^{k'l'}_{ok} + \pi^{k'l'}_{ld} - \pi^{k'l'}_{od}]}_{\text{Rerouting of ground traffic}} - \underbrace{\theta_M (1 - \pi^{road}_{od}) \pi^{k'l'}_{od}}_{\text{mode switch}} \Big\}, \tag{B.18}$$

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

Consider first the own effect (when $kl = k'l'$). The first term in the curly bracket captures the impact on shipment over $k \to l$ through rerouting *within* the road network. With $1 + \pi^{kl}_{ok} + \pi^{kl}_{ld} - \pi^{kl}_{od} > 0$, this force contributes negatively: a *decrease* in the cost on edge $k \to l$ *increases* the share of shipments taking this edge. The second term in the bracket captures the response in the mode choice—more shipment will be made via road in response to a decrease in the edge cost. Both forces work in the same direction and imply that as an expressway is added to $k \to l$, more trade flows will go through this edge.

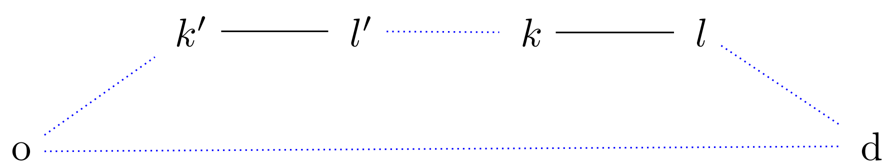


Figure B.1: Interactions Between Segments

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

Now consider the cross-derivative (when $kl \neq k'l'$). The response in mode choice has the same sign as before, but the first term in the bracket capturing the re-optimization of the ground traffic could be positive or negative, depending on the positions of $k' \to l'$ and $k \to l$ in the network. When $k' \to l'$ and $k \to l$ are on competing routes between o and d, shipments between o and k and between l and d are unlikely to pass through $k' \to l'$, so $\pi^{k'l'}_{ok}$ and $\pi^{k'l'}_{ld}$ are both small and $-\theta(\pi^{k'l'}_{ok} + \pi^{k'l'}_{ld} - \pi^{k'l'}_{od})$ is more likely to be positive. In such cases a reduction in $\iota_{k'l'}$ draws ground traffic *away* from $k \to l$. On the other hand, if $k' \to l'$ is en route of $o \to k \to l \to d$, as in the example given in Figure B.1, then the opposite can happen—reducing $\iota_{k'l'}$ increases the traffic passing through $k \to l$.[9]

To see the importance of own- and cross-routing effects, we can decompose SO_R into two terms.

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

While the sign of 'Own SO_R' is always negative, the sign of 'Cross SO_R' is ambiguous as discussed above. We illustrate two cases in Figure B.2 using the parameterized model. Consider one of the busiest expressway segments, the one between Laiwu and Linyi, colored solid black in the map. The colors of other edges indicate their cross derivative term with the one between Laiwu and Linyi. Cold colors indicate that the cross derivative is negative, in which case a new expressway between Laiwu and Linyi

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

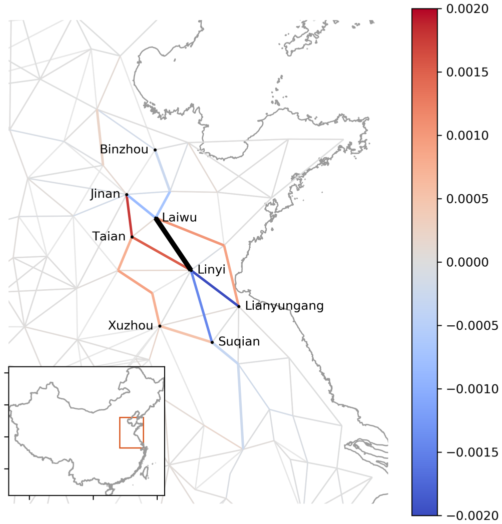


Figure B.2: Complementarity and Substitution between Segments: An Example

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

will increase the traffic on the segment. For example, the segment between Jinan and Laiwu.[10] On the other hand, warm colors indicate that a segment is a substitute to the expressway between Laiwu and Linyi. For example, the route from Jinan to Xuzhou. Importantly, most of these rerouting are traffic from cities to the north of Laiwu, to cities to the south of Linyi. This suggests when evaluating expressways, it is necessary to consider not only the direct trade between two cities connected by a segment, but also traffic that is merely passing by.

The importance of the cross-derivative force crucially depends on the segments being jointly evaluated. Equation (17) in the text provides a decomposition focusing on the 100 busiest expressway segments. It shows that the own second order effect accounts for -58% of the effect, whereas the cross-second order effect accounts for 8%. This implies that the expressway segments tend to be complementary.

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

C Quantification

C.1 Identification of Structural Parameters

Composite parameters that enter the route model. To understand how the composite parameters governing route choices—$\kappa^H\theta$, $\kappa^L\theta$ and $\frac{\theta_F}{\theta}$—are identified, it is useful to consider a limit case with $\theta = \infty$, under which the effective trade cost between o and d, τ_{od}, is simply the cost of the least-cost path. Slightly abusing notation, we use $dist_{od}^{H,t}$ and $dist_{od}^{L,t}$ to denote the *total* length of expressways and regular roads along the least-cost path at time t, respectively. Then the trade cost between o and d is $\kappa^H dist_{od}^{H,t} + \kappa^L dist_{od}^{L,t}$. With this we can write the structural routing equation as:

$$\log(\pi^t_{(o,RoW),d}) = -\theta_F(\kappa^H dist_{od}^{H,t} + \kappa^L dist_{od}^{L,t}) + \text{fixed effects}.$$

The above equation provides a micro-foundation for the reduced-form specification in Section 2. It also conveys two points on identification. First, route choices can identify the relative costs between the two types of roads, $\frac{\kappa^H}{\kappa^L}$. The identification comes from changes in compositions of regular roads and expressways in a route. Second, route choices can only identify *products of parameters*, $\theta_F\kappa^H$, and $\theta_F\kappa^L$. Intuitively, port choices reflect the combined effect of two forces: the marginal cost of additional distance; the marginal impact of cost on port choices. This is reminiscent of an result in gravity estimation that trade cost and trade elasticity cannot be separately identified using trade flows alone. Research in international trade has used price data (Eaton and Kortum, 2002) to overcome this problem. In the same spirit, we use the unit value information contained in the customs data, a step that we will return to below.

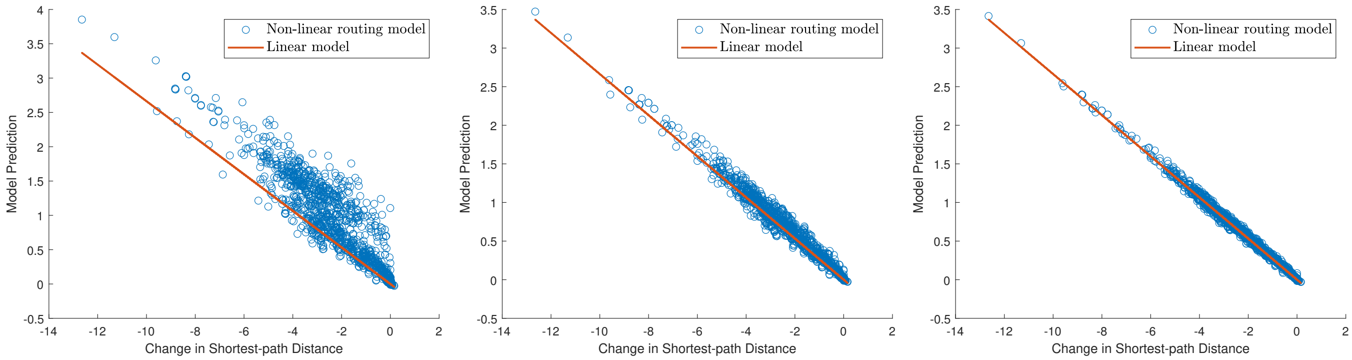
Moving from the above limit case back to our specification, *suppose we have already estimated θ_F using the* unit value information, what identifies θ? We can decompose the routing pattern from o to d implied by the model as below:

$$\begin{aligned}
\log(\pi^t_{(o,RoW),d}) &= \frac{\theta_F}{\theta}\log(\tilde{\mathbb{B}}^t_{(o,d)}) + \text{fixed effects} \\
&= -\theta_F\kappa^L \underbrace{\left(\frac{\kappa^H}{\kappa^L} dist_{od}^{H,t} + dist_{od}^{L,t}\right)}_{\text{regular-equivalent distance}} - \underbrace{\left[\frac{\theta_F}{\theta}\log(\tilde{\mathbb{B}}^t_{(o,d)}) - \theta_F\kappa^L\left(\frac{\kappa^H}{\kappa^L} dist_{od}^{H,t} + dist_{od}^{L,t}\right)\right]}_{\text{deviation}} + \text{fixed effects},
\end{aligned}$$

i.e., $\log(\pi^t_{(o,RoW),d})$ encompasses the effect of the regular-equivalent length of the least-cost path, and a deviation term. Given estimated (κ^L, κ^H), the deviation term is a function of θ, and summarizes collective impacts of all routes that are inferior to the least-cost one. For a given network structure, θ affects the relative importance of the inferior routes. When θ is infinite, only the least-cost path matters, so the deviation term has no impact on routing; when θ is small, routes that are slightly inferior will be used more often and the deviation term will have more explanatory power for port choices.

Figure C.1 illustrates this intuition. In each panel, the horizontal axis is the change in the effective length of the least-cost path after the expressway construction.[11] The vertical axis denotes the change in $\log(\pi^t_{(o,RoW),d})$ predicted by the model. If all circles fall on the prediction by the linear component, it means that all truckers choose the least-cost paths and the deviation term defined above matters very

[11]The regular-road equivalent distance is based on the estimated value of (κ^L, κ^H).



(a) $\theta = 0.75 \times \theta^*$ (b) $\theta = \theta^*$, the calibrated value (c) $\theta = 1.5 \times \theta^*$

Figure C.1: Model Prediction with Varying θ

Note: The horizontal axis is the change in regular-equivalent distance between city pairs due to the expressway construction; the vertical axis is the model-predicted change in log shipments. As θ increases, changes in predicted shipments become closer to linearly correlated with changes in shortest distance.

little in the model; if dots are more spread out, it means that beyond the least-cost path, the structure of the entire network, captured by the deviation term, also matters. As expected, the figures show that as we increase θ from the estimated value (θ^*), the circles fall more tightly around the prediction by the linear component; on the other hand, the deviation from the linear prediction increases as θ decreases.

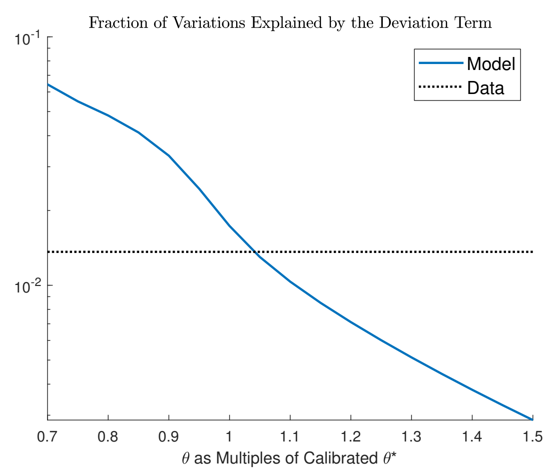


Figure C.2: Predictions of the Nonlinear Model Beyond the Shortest-path Distance

Note: The figure reports the fraction of variation in log export shipment explained by the deviation term, among total variation explained by the model. The deviation term is calculated as the difference between the prediction of the linear model and that of the nonlinear model under the calibrated θ. See the text for details.

This discussion makes clear that, *conditional on* θ_F, θ can be identified by the relevance of the structure of the road network beyond shortest-path distance in explaining changes in port choices. To see what the data tell us about this importance, we calculate the fraction of variation in log export shipment explained by the deviation term, among total variation explained by the road network structure.[12] The dashed line in Figure C.2 plots this fraction. As shown, the deviation term explains a non-zero fraction of data variation, so θ is not infinite. The solid line plots the implication of the model for this object as θ

[12] To do this, we estimate Equation (9) with only the length of the least-cost path and the fully nonlinear structure separately, and then calculate the percentage change of the residual sum of squared errors.

Table C.1: 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	0.184***	0.184***	0.289***	0.187**			
	(0.062)	(0.062)	(0.089)	(0.087)			
Heaviness- HS4 Category					0.292***	0.352***	0.229***
					(0.044)	(0.047)	(0.030)
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	3362494	3361110	3127626	236027	3127625	2017990	142835
R^2	0.058	0.069	0.330	0.427	0.374	0.570	0.582

Notes: This table reports the regressions of the log price ratio on log sector-level weight-to-value ratio. The dependent variable is the log of price ratio and is always computed within a city-destination country-HS8 category; the explanatory variable is the log of the weight-to-value ratio at HS2 category level (Columns 1-4) and HS4 category level (Columns 5-7). Letters o, d, c, f, i stand for origin city, port, destination country, firm, and HS2 category fixed effects, respectively.
Standard errors are clustered at HS2 category level (Columns 1-4) or HS4 category level (Columns 5-7).
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

varies. At around our estimated θ^*, the model generates about the same prediction as in the data.[13] This argument identifies θ only conditional on θ^F, i.e., it identifies $\frac{\theta_F}{\theta}$. Below we use price variation contained in the customs data to identify the *level* of θ.

Price-heaviness elasticity μ. We estimate Equation (10) to identify μ and θ. Since these two parameters are identified from different variations, we estimate them separately, so more flexible controls can be included. Table C.1 reports our estimate of μ, the elasticity of the trade cost with respect to the weight-to-value ratio of a sector. The first four columns focus on the comparison of the log price differences across HS2 categories, with progressively more demanding fixed effects. The first and second columns control for city, port, and destination country fixed effects and city-port-country fixed effects, respectively. The estimated coefficient is around 0.184. Even within a city-port-country cell, some firms might systematically set prices differently. To account for this possibility, Column 3 control for firm-port-country fixed effects. The point estimate increases somewhat to 0.29 and is precisely estimated.

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

One further concern is that our measure of 'heaviness', the weight-to-value ratio, might capture other characteristics of a sector that correlate systematically with prices. In Columns 5 through 7, we estimate the specification using the weight-to-value ratio at the HS4 category level. This allows us to control for

[13]The discussion here aims at visualizing the data patterns that identify θ. The intersection in the figure is not exactly at θ^* because θ^* is not chosen to target this fraction, but rather estimated via the nonlinear least square specified in Section 4.

Table C.2: Price Distance Regression

	(1)	(2)	(3)	(4)	(5)
	OLS		2SLS		Structural
$dist_{od}$	0.050***	0.057***	0.049***	0.058***	
	(0.011)	(0.012)	(0.011)	(0.012)	
$\log(\tilde{B}(\widehat{\kappa^H\theta}, \widehat{\kappa^L\theta})_{(o,d)})$					-0.0090***
					(0.0022)
Fixed Effects	dci, oci	dci, oci	dci, oci	dci, oci	dci, oci
Exclude major cities	yes	yes	yes	yes	yes
Exclude differentiated goods		yes		yes	yes
Observations	3156133	279165	3156133	279165	279158
R^2	0.335	0.351	-	-	-
First Stage KP-F statistic			1191.648	935.979	1090.070

Notes: This table reports the regressions of the log price ratio on the distance between the origin city and the port (Columns 1-4) or the output of the routing model (Column 5). Letters o, d, c, i stand for origin city, port, destination country, and HS-8 product fixed effects, respectively.

Standard errors in Columns 1 through 4 are clustered at city-port level. Standard error in Column 5 is generated through bootstrapping. See Appendix C.2 for details.

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

the HS2 fixed effects. Column 5 of Table C.1 is our preferred specification, which is identified from within a city-port-country and HS2 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 around 0.3%.

Route elasticity θ. Table C.2 reports our estimates for θ. Since we do not aim to identify μ in this regression, we can absorb the category characteristics in fixed effects. We present results for two sets of regressors. The first is for the distance between city o and port d, which illustrates the variation in the data that identifies θ more transparently . The second is for $\log(\tilde{B}(\widehat{\kappa^H\theta}, \widehat{\kappa^L\theta})_{(o,d)})$, which allows us to estimate θ directly.

The first two columns use OLS and control for port-HS8-destination country and city-HS8-destination country fixed effects, respectively. The former set of fixed effects captures, within a HS8 category, the overall tendency of some ports or destination countries to be involved in the export of more pricey goods; the latter controls for the overall tendency of a city in producing pricey good for exporting to specific countries. The point estimate of the first column, which uses all categories, suggests that the price ratio increases by around 5% as an additional hundred km regular-road equivalent distance is added. The second column restricts to non-differentiated varieties for robustness. This restriction significantly reduces the sample size but the point estimate remains similar. To alleviate the concern about the endogeneity of the road network, Columns 3 and 4 estimate a 2SLS specification using the IV generated from the minimum-spanning tree. The point estimates are in the range of 0.05 to 0.06, statistically indistinguishable from the OLS estimates.

These reduced-form results show that: the price of goods is more expensive when a port is further apart from the origin city. Importantly, the point estimate is robust across specifications with different fixed effects, sample restrictions, and the use of IV.

The variation exploited in these estimates can identify θ. Specifically, recall that from Equation (10), $-\frac{1}{\theta}$ is exactly the elasticity of log price ratio with respect to $\log(\tilde{B}(\kappa^H\theta, \kappa^L\theta)_{(o,d)})$. Having estimated $\kappa^H\theta$

and $\kappa^L\theta$ in the previous step, we can plug in their values and construct $\log(\tilde{B}(\widehat{\kappa^H\theta}, \widehat{\kappa^L\theta})_{(o,d)})$. Column 5 in Table C.2 uses the same specification as in Column 4, but has $\log(\tilde{B}(\widehat{\kappa^H\theta}, \widehat{\kappa^L\theta})_{(o,d)})$ as the dependent variable. The point estimate of $-\frac{1}{\theta} \approx -0.0090$ translates into $\theta \approx 111$, implying that different routes connecting the same pair of cities are highly substitutable. This estimate is close to the estimate of Allen and Arkolakis (2019) using routing information of domestic shipments.

C.2 Inference of Structural Parameters

Table 3 in the text reports point estimates and distributional statistics of the key structural parameters. Panel C of Table 4 reports statistics for additional parameters determined jointly in calibration. This section explains how we draw statistical inference for each of these parameters.

Step 1. The model does not incorporate structural errors at city-port level, so we assume that the source of uncertainty is due to measurement errors and use bootstrap to infer the size of uncertainty. For parameters in Panel A of Table 3, we resample with replacement by city 200 times. Each time we draw a sample, we estimate the nonlinear routing problem described in Equation (9) to obtain a new set of estimated $(\kappa^H\theta, \kappa^L\theta, \frac{\theta_F}{\theta})$. We then calculate from these 200 repetitions the distributional statistics of the composite parameters. With bootstrapping at the city-level, the standard errors calculated capture the potential correlation between different city-port pairs, so they tend to be more conservative than city-port level clustering in reduced-form analyses.

Step 2. For the inference of θ, on top of measurement error of the price data, the errors due to generated regressors ($\log(\tilde{B}(\widehat{\kappa^H\theta}, \widehat{\kappa^L\theta})_{(o,d)})$) also need to be taken into account. We therefore use a joint bootstrap procedure. Specifically, from each bootstrap sample in Step 1, we have obtained one estimate for $\widehat{\kappa^H\theta}, \widehat{\kappa^L\theta}$. We use the corresponding ($\log(\tilde{B}(\widehat{\kappa^H\theta}, \widehat{\kappa^L\theta})_{(o,d)})$) on a bootstrapped price sample for the regression reported in Column 5 of Table C.2. We obtain the distributional statistics of θ by repeating this procedure 200 times. As Column 5 shows, the standard error generated this way is similar in magnitude to those for the linear regressions (Columns 1 through 4) calculated using asymptotic theories.

To draw inference for μ, which is estimated using linear regressions, we use directly the standard error in Column 5 of Table C.1.

Step 3. The uncertainty in these structural parameters affect the estimation of the full model. Panel A of Table 4 summarizes the parameters estimated using micro data and their standard errors. Panel B is the parameters from external sources, which we take as given. Panel C is the parameters estimated jointly. To take into account the uncertainty of parameters in Panel A, we draw 200 realizations of Panel A parameters from their joint distribution; for each of these draws, we calibrate the remaining parameters to match the same targets. Reported in Panel C of Table A are the standard errors of calibrated h_0 and $\bar{\kappa}$. The standard errors tend to be small, reflecting that they are mostly determined by their own targets, rather than the parameters in Panel A.[14]

Three comments are in order on this procedure. First, when generating the joint distribution of the parameters in Panel A, we take into account that some of these parameters are estimated jointly and thus correlated. In bootstrap, we consider three sets of parameters. (1) The composite parameters estimated from the port choices, $\kappa^H\theta$, $\kappa^L\theta$, and $\frac{\theta^F}{\theta}$. (2) For each draw of these parameters, we estimate the linear regression in Column 5 of Table C.2 for θ, which, together with the three composite parameters,

[14]Note that in this procedure, each time we will also have different values for other parameters in Panel C such as T_d^i. We omit these parameters and their standard errors from the Table.

gives us all four parameters for the routing model. (3) We randomly draw a μ from its own asymptotic distribution. Each calibration of the model then uses one realization of these three sets of parameters.

Second, in evaluating the impacts of the expressway construction, we repeat the counterfactual experiment for each of the 200 model calibrations. The distributional statistics calculated in Table 5 are from these 200 counterfactual experiments.

Finally, in this entire procedure we take the parameters from external sources, such as the trade elasticity and the IO table parameters, as given. These parameters are either from the aggregate data, or estimated by a other studies with no consistent ways for statistical inference. We conduct sensitivity analyses to show how results vary with these parameters in Section C.6 of this appendix.

C.3 Model Validation

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

Expressway and Export Growth. Given our use of export data in estimation, we first assess how the model fits city-level exports in the data. This comparison is out-of-sample, because in calibration we absorb the level of export through city-port fixed effects and use only the within-variation from the patterns of routing. Figure (C.3) plots the model-implied city export against the data. To ensure that the fit is not due to city size, the plots are for residuals from a regression that controls for city-level employment. The figure shows that the model closely matches the export observed in the data well.

In the second validation test, we compare the model-predicted export growth led by the expressway network expansion to the actual export growth in the data. This is a joint test of two hypotheses: 1) whether the expressway expansion as large as the one seen in China over the decade led to differential growth of exports across cities; 2) when fed into the expressway expansion, whether the model can generate the changes in trade patterns in the data.

To implement this exercise, we feed in the 1999 expressway network to the model and solve a counterfactual equilibrium holding all other parameters at the calibrated values. We treat the export generated from this counterfactual equilibrium as the model export around 2000. We then compare the export at the city-sector level between the model and the data for 2000 and 2010. Table C.3 reports the results. The dependent variable is the log export in the data and the independent variable is its model counterpart. The first column presents the result based on a cross-sectional specification. The second and third columns control for sector-time and city-sector fixed effects, so the comparison is on export growth within a city-sector cell. The point estimates are highly statistically significant in both cases.

Importantly, all these regression models have a F statistic above the rule-of-thumb for bounding biases in IV estimates. Under the assumption that 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 the city-industry level. A growing literature has examined the impacts of Chinese export on its domestic economy. One IV commonly used in this literature is variation in tariffs due to the WTO accession, which is valid by assuming that the pre-WTO tariffs are exogenous (Facchini et al., 2019; Tian, 2019). The IV based on our model predictions vary across regions and over time and is valid under a different set of assumptions from existing studies. It could be of use for future research in this area.

Comparison with truck flow data. Our second set of validation exercises aims to show that the customs data capture the variation in domestic shipment well. Specifically, we obtain the number of bilateral truck movements between pairs of Chinese cities in 2019, collected by a digital logistic platform,

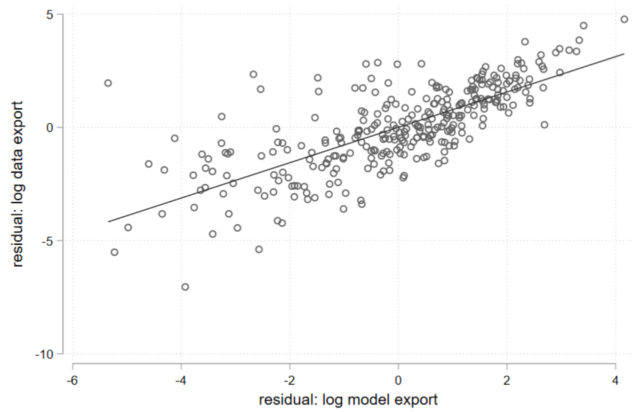


Figure C.3: City-level export: Model versus Data

Note: The figure plots model-implied city export against the data, netting out employment at the city level.

Table C.3: Predicting Export Growth

	(1)	(2)	(3)
Log(export), model	0.465***	0.953***	0.871***
	(0.048)	(0.189)	(0.199)
Fixed Effects	t	oi, it	oi, it
Exclude major cities	no	no	yes
Observations	8472	8472	6576
R^2	0.333	0.878	0.860
F-statistic	92.706	25.332	19.223

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

G7, which helps logistic companies and transportation firms to manage 1.3 million trucks in China.[15] Since the dataset is available only for long after expressways were built, we can not use it to estimate the specification exploiting over-time variation, but we can still estimate a cross-sectional specification. We find an estimate -0.432 (Column 1 of Table C.4) off the cross-sectional specification, which is very close to the baseline estimate of -0.384 (Column 2, Table 2) using customs data. This exercise shows that at least when the cross-sectional variation is used, customs data and domestic shipment data give similar estimates on the shipment-distance semi-elasticity.

We also compare the model-implied bilateral shipment to the data. Column 2 of Table C.4 shows that the model-implied shipment flows are highly correlated with truck flows, with a linear regression coefficient of 1.3. Of course, the raw correlation between the two variables could be driven by the size of origin and destination cities. Column 3 controls for the origin and destination fixed effects and shows that doing so does not diminish the importance of model-implied shipments in explaining the data.

Figure C.4 visualizes the close relationship between the two variables. After netting out origin and

[15]The company provides services to logistic companies. Among their main services is the management of an in-truck camera that monitors risky driving behaviors (such as drowsy driving and DUI). This in-truck device records the trip made by truckers. We do not observe whether a truck is loaded or not, so the measure of shipment is symmetric.

Table C.4: Validation: Data Truck Flow v.s Model Shipment Flow

Dependent variable	Log data truck flow		
	(1)	(2)	(3)
Effective distance	-0.432***		
	(0.003)		
Log model shipment flow		1.281***	1.668***
		(0.007)	(0.011)
Observations	54057	54057	54057
o and d fixed effects	yes	no	yes
R^2	0.627	0.435	0.597

Notes: The dependent variable is log of number truck flows between city pairs in the data (2019); the independent variable is the regular-road equivalent distance, and the log shipment flow between city pairs predicted by the model, calibrated to match the 2010 Chinese economy. Robust standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

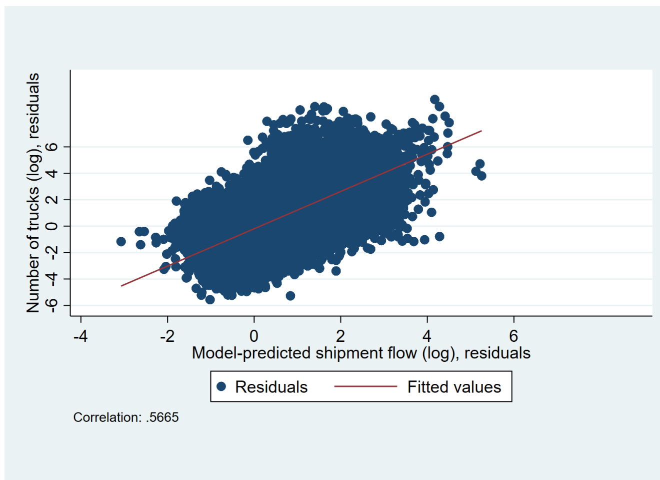


Figure C.4: Data Truck Flows v.s Model Shipment Flows

Notes: The figure plots the residual correlation between truck flows and model implied shipment flows after netting out origin and destination fixed effects.

destination fixed effects, the residual correlation between them is around 0.57. This test is 'out-of-sample' in two senses. First, our model is estimated using the export data, whereas truck flows capture mainly domestic shipment. Second, our estimation uses over-time variation between 1999 and 2010, whereas truck flows are from 2019, after a decade of rapid growth and transformation of economic landscape in China. Despite this, the fit is comparable to the 'in-sample' fit of models that are estimated to match domestic trade flows. For example, Allen and Arkolakis (2019) (Figure 2) shows a residual correlation of 0.60. This suggests that the customs data, combined with our routing model, can capture the bilateral shipment in the data well.

Beyond looking at the bilateral correlation, we also examine whether the model generates the pattern of shipment over different distances as in the data. The model implies that the value of shipment decreases in bilateral distance significantly when the distance is below 200 miles, and then the decrease becomes more gradual. This is consistent with what we find using the truck flow data. Interestingly, both the model-implied shipment flows and truck flows capture the salient features of domestic shipment documented in Hillberry and Hummels (2008) based on the U.S. data. These results are available

Table C.5: Correlation with Shipment

	(1)	(2)	(3)
Log(shipment), model	0.314***	0.177***	0.176***
	(0.040)	(0.035)	(0.041)
Log(employment)		0.594***	0.587***
		(0.059)	(0.062)
Observations	240	240	234
Fixed Effects	no	no	prov
R^2	0.234	0.488	0.636

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

upon request.

Transport hubs. As a final validation, we examine the model's prediction on shipment by city. Because of their central locations in the transport network, some cities become 'hubs' that shipments to other places go through. To validate the model, we can compare the model-inferred shipment that passes a city to its empirical counterpart, sourced from the 2010 yearbook for transportation.[16] Table C.5 reports the regression of 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 suggests that the model prediction correlates with the data not only due to usual gravity forces, which predicts more trade for bigger cities, but also because it captures the traffic passing by. The third column further shows that including province fixed effects does not change the estimate. This implies that the prediction power comes from the network connections of a city shaped by the routing model, rather than the broad location of the city.

C.4 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, 2020; Allen and Arkolakis, 2019) in three aspects. First, our structural estimation exploits changes in the route choice of exporters resulting from the domestic expressway network expansion, which naturally implies that the network expansion reduces trade costs 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 sectoral specializations and sectors to differ in trade costs; third, we incorporate intermediate inputs.

This section shows that because these ingredients allow us to infer the distribution of shipment among different routes and the shipment values more accurately, they are important for the quantitative results. We parameterize a series of restricted models and compare the inferred welfare gains in these models to the baseline results. For transparency, throughout this subsection we recalibrate only the trade cost level parameter h_0, the amenity $\{B_d\}$, and the city-sector productivity $\{T_d^i\}$, to match the av-

[16] The data are aggregated by city; the National Bureau of Statistics surveys firms in the logistics industry to produce this statistic. The data series appear inconsistently defined over time, with frequent abrupt changes from one year to another, so we do not use the time dimension of the data.

Table C.6: Welfare Gains in Alternative Models, Matching Average Shipment Distance

	Baseline	Model (2)	Model (3)	Model (4)	Model (5)
International trade	✓				
Trade cost heterogeneity	✓	✓			
Regional specialization	✓	✓	✓		
Intermediate Input	✓	✓	✓	✓	
Welfare gains	5.10%	4.47%	4.29%	3.36%	0.89%

Note: For each alternative model, the trade cost level parameter h_0, amenity $\{B_d\}$, and city-sector productivity $\{T_d^i\}$ are recalibrated to match the same average domestic ground shipment distance, population distribution, and city-sector sales (or city-level sales, depending on whether regional specialization is allowed).

erage domestic shipment distance, the population distribution, and the sales by either city or city-sector, depending on the restriction on the model. We keep other structural parameters in the routing model as in the benchmark. Below reports the changes in results as we sequentially eliminate the elements in the model.

Domestic transport costs in international trade. Column 2 of Table C.6 is the result from a model without international trade, i.e., with $\tau^i_{RoW} = \infty$, $\forall i$. The inferred gains from expressway construction in this model is about 12% (or 0.63 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 expressway segments. As we show in Proposition 1, the welfare gains of trade cost reductions can be approximated by total cost savings on trade flows on segments directly affected, adjusted for the savings that are passed on the RoW. By matching the average shipment distance for goods within China, both the full model and the restricted model without international trade generate similar predictions for domestic trade flows, so they predict similar cost savings from domestic trade. Through the lens of the full model, however, these are only part of the benefits—the improvements in domestic infrastructure reduce the trade costs for the importers and exporters from the hinterland. Because part of these additional cost savings will accrue to the Chinese economy, overlooking international trade leads to smaller inferred gains.

Transportation intensity. In the second experiment, we set $\mu = 0$ and then recalibrate the model to match other moments. We then conduct the same exercise as before. Under the assumption of homogeneous transport cost across sectors, the inferred gains are down from Model (2) to 4.29%.

At first glance, this might seem surprising, as with a large enough number of regions and road segments, the law of large numbers should have kicked in and the heterogeneity in transport intensity across sectors could be washed out. The reason why sector heterogeneity is not simply washed out is, when calibrated to match the same average shipment distance, Model (2) infers systematically higher values of shipment compared to Model (3). More specifically, with heterogeneity in trade costs, for the same level of inter-city shipments, Model (2) will predict a higher fraction of them in lighter sectors (with lower weight-to-value ratios) because they incur lower shipping costs in Model (2) but not in Model (3). Because the welfare gains are, to the first order, proportional to the value of goods but not their weights, the model with sector transport intensities predicts larger welfare gains.

Regional comparative advantage. Chinese regions specialize in different broad sectors, e.g., manufacturing and service in the Southeast versus energy in the Northwest. To understand the importance of

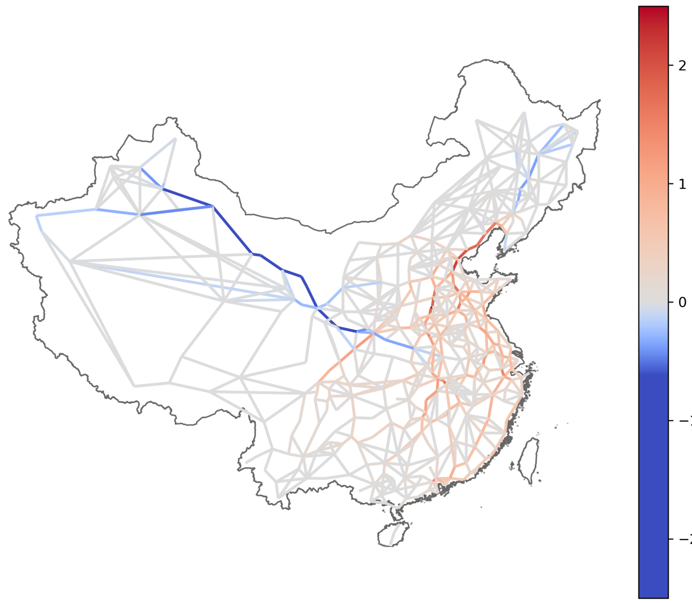


Figure C.5: Differences in Shipment Value Shares, 'No Specialization' Minus 'Baseline'

Note: The numbers are the differences in shipment value over GDP between a model with no specialization and the baseline. The values plotted include both expressway and regular road shipments. Cold colors indicate that there is less shipment in the model with no specialization than the baseline.

accounting for regional productivity differences, Column 4 reports the result from a recalibrated model without specialization. Specifically, we assume all sectors within a region have the same productivity, i.e., $T_o^i = T_o^j = T_o, \forall o, i, j$, and pin down $\{T_o\}$ by matching the total sales of each city in the data. The input-output structure is kept the same as in the baseline model. The inferred gains in this model are 22% smaller than an otherwise similar model with regional specialization (Column 3).

Patterns of regional specialization matter 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 comparative advantages are eliminated, the spatial clustering also disappears, so inter-city trade in the restricted model shifts towards partners that are closer to each other. Although both models are calibrated to generate the same average shipment distance, this simple statistic does not capture all these patterns. Indeed, Figure C.5 plots the change in shipment intensities between city pairs from Model (3) to Model (4). The segments that see the biggest decrease in inferred shipments are the ones connecting the northwest and northeast—the energy producing area—to the center of the country with a heavy manufacturing presence; the segments that see an increase in inferred shipments are the ones connecting regions within the center and the east of China. As a result, Model (4) infers higher gains for expressway segments in the center of the country and lower gains for projects connecting the center to the northeast and northwest—regions with very different comparative advantages. Whether it underestimates or overestimates the return to a specific project thus depends crucially on where a project is. For the actual projects built during the decade, the balance comes down to an underestimation of the welfare gains by 22%.

Intermediate inputs. In the final comparison, we further shut down intermediate inputs in production by assuming the labor shares (β^i) are one and sectoral shares (γ^{ij}) are zero in all industries. The welfare gains inferred by this model decline by three-quarters to around 0.9%. This difference can be understood by inspecting Equation (C.1).

$$\frac{X^i_{od}}{Y} = \frac{X^i_{od}}{\sum_i \sum_{o,d} X^i_{od}} \cdot \frac{\sum_i \sum_{o,d} X^i_{od}}{Y}. \tag{C.1}$$

For a simple example, assume that all regions o and d are symmetric, with positive but symmetric inter-regional transport costs. When calibrated to match the average shipment distance, Models (4) and (5) generate the same trade intensity, i.e., $\frac{X^i_{od}}{\sum_i \sum_{o,d} X^i_{od}}$. However, in the model without intermediate inputs, the overall absorption $\sum_i \sum_{o,d} X^i_{od}$ is equal to the GDP, whereas in the model with intermediate inputs, the overall absorption is several (around three in our calibration) times of the GDP. As a result, the inferred value of $\frac{\sum_i \sum_{o,d} X^i_{od}}{Y}$ is too small in the model without intermediate inputs. As we show in Proposition 1, by and large, the overall gains from the expressway construction are determined by $\frac{X^i_{od}}{Y}$. By assuming away intermediate inputs, the restricted model overlooks that goods are traded multiple times on the road, which amplifies the gains from the reduction in transport costs.[17]

To summarize, when restricted to a bare-bone one-sector model used in most of the literature, the welfare gains is only a small fraction of the baseline result.

C.5 Comparison to Existing Evaluations Using Other Approaches

We compare our assessment of the welfare impacts to existing evaluations by academia and policy institutions using three alternative approaches.

The first approach, which is also what we adopt in this paper, is to rely on quantitative models for simulations. The obstacle faced by this approach is the lack of reliable domestic trade data for disciplining the importance of transport infrastructure for trade and welfare. Roberts et al. (2012) uses a one-sector new economic geography model. The model expresses regional wages as a function of market access, which in turn depends on trade elasticity and transport costs. Cross-sectional wage variation can then be used to discipline trade and transport cost elasticity. Roberts et al. (2012) finds that the static welfare gains from the expressway network to be around 6%, slightly higher than our estimate.[18]

The second approach directly measures the return to expressway investment in the transport, logistic, and postal service sectors. The idea is that if transport infrastructure affects the aggregate economy

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

[18]In light of our finding that input-output linkages and sector heterogeneity amplify the welfare gains, the larger gains in Roberts et al. (2012) might be surprising. The reason for this finding is that, instead of targeting trade flows, Roberts et al. (2012) targets wage dispersions, under the assumption that the observed wage dispersions are entirely due to differences in market access arising from trade costs. Large wage disparities in the data are thus interpreted as large trade frictions, which in turn imply large gains from infrastructure investment. Compared with our evaluation, which allows regional productivity differences, Roberts et al. (2012) imply too low a volume of trade, but too large trade cost reductions led by transport infrastructure. These two forces turn out to bring their calculated gains similar to ours.

only through these sectors, then its return would be captured in the capital value added of these sectors. Return measured this way might be lower than the true social return of infrastructure for two reasons. First, China's transport infrastructure is likely under priced,[19] so the capital return to companies in charge of expressway operations might be lower than the social value of investment. Second, the transport, logistic, and postal service sectors are not the only sectors using the expressways. For example, transportation of goods and services by residents or manufacturing firms benefit from the expressway expansion, but their benefits do not necessarily show up in the value added of the transportation sector. Using sectoral value added data, Bai and Qian (2010) finds that the gross per-period return to expressway investment to be around 25-30%. This is smaller than the 51% static gross return to capital implied by our approach (5.1% welfare gains divided by 10% GDP capital investment), but in the same order of magnitude.

The third approach is to estimate a provincial-level production function, with provincial GDP being the output and expressway investment being one of many inputs. Given challenges to identification and that the studies generally use different measures of infrastructure, the literature has not reached a consensus.[20] For example, Shi and Huang (2014) finds that after 2001, investment in a broad category of transport infrastructures offers a lower gross return than private capital, but the estimate include different kinds of infrastructure so it is hard to compare this number to ours. On the other hand, focusing on roads, Fan and Chan-Kang (2005) estimates that each additional km of 'high-quality' roads generates 32% static gross return. Their focus, 'high-quality' roads, includes multi-lane paved roads that are not expressways. In addition, such an approach identifies only the different effects across regions while overlooks the general equilibrium effects, which likely improve welfare in all regions. These differences might explain why their estimate is smaller, but the conventional confidence intervals of their estimate covers our baseline estimate of a 51% return.

Overall, comparing with findings from existing studies using different approaches, our estimate appears reasonable.

C.6 Sensitivity Analyses

Table C.7: Sensitivity Analyses

Change in	(1) High Heterogeneity in Sectoral Trans. Cost	(2) High Substitution across Trans. Mode	(3) External Economy of Scale	(4) Immobile Labor	(5) Mobile Labor + Migration Costs	(6) $\sigma = 9$
Aggregate welfare (%)	0.053	0.047	0.048	0.052	0.052	0.037
Log(Domestic trade)	0.117	0.119	0.144	0.140	0.138	0.120
Log(Exports)	0.097	0.102	0.097	0.113	0.108	0.133

Note: The table reports (the minus of) changes in model statistics as the economy moves from the calibrated equilibrium with the 2010 expressway network to the one with the 1999 expressway network. Alternative models in (1)-(6) are recalibrated to match the same targets as in Table 4.

We conduct a number of exercises to assess the sensitivity of the baseline results to alternative as-

[19]Many highway management companies are on the government support since they cannot self sustain operations.

[20]A challenge to this approach is that at the provincial level, stock of expressways might be endogenous to GDP growth. To circumvent the identification challenge, more recent empirical works focus on counties or prefecture cities, at which level it is possible to use 'exogenous' placement of expressways (see, e.g., Banerjee et al., 2020, Faber, 2014, and Baum-Snow et al., 2020). Because this approach overlooks the general equilibrium effects, and often estimates a coefficient associated with a dummy indicating whether being connected to expressways, it is difficult to convert such estimates to overall benefits of the macro economy.

sumptions. We focus on five scenarios. The first is on the sector heterogeneity of transport costs. Instead of 0.3 in the baseline calibration, we now set μ to 1, which corresponds to a linear relationship of the iceberg trade cost in the weight-to-value ratio. The second robustness check increases the elasticity of substitution between road transportation and the outside mode, θ_M, from the benchmark value 2.5 to 14.2, an estimate by Allen and Arkolakis (2014). Our third robustness allows for industry-level agglomeration. Specifically, we set $T_d^i = \overline{T}_d^i [l_d^i]^\chi$ with $\chi > 0$. This assumption implies an external increasing return to scale to specialization. The estimates for χ in the literature, as surveyed in Combes and Gobillon (2015), range from 0.02 to 0.13. We set $\chi = 0.075$, which lies in the mid-range of the estimates. Fourth, given the hukou reform in China that reduces migration frictions was gradual during this period, we conduct two exercises—one with immobile labor, the other with partially mobile labor. The additional model ingredients with immobile labor or with migration frictions are presented at the end of this subsection. Finally, we experiment with a higher elasticity of substitution for σ.

The first column of Table C.7 shows that when sectoral heterogeneity in transport costs is more important, the inferred welfare gains are slightly larger. The second column shows that when the elasticity of substitution between transport modes increases, the inferred welfare gains are slightly smaller. This is because after expressways are removed, traders can switch to the alternative mode more easily and incur smaller losses. Adding external economies of scale at the industry level leads to a modest decrease in the inferred gains. Fourth, the welfare gains increase slightly if domestic workers are completely immobile (Column 4) or subject to migration frictions (Column 5), once the models are recalibrated to match the same targets. Finally, when we the elasticity of substitution between goods produced in different regions is higher, the inferred welfare gains is smaller—intuitively, as goods become more substitutable, gains from inter-regional trade are smaller, so infrastructure improvements also have a smaller welfare effect. Nevertheless, this implies a net return to investment of 85%, so the conclusion that expressway investment during 1999-2010 generated high return remains robust. The impacts of expressway expansion on domestic trade and export, on the other hand, is in similar magnitudes.

Model with immobile labor. With the assumption of immobile labor, the numbers of workers in domestic cities, $\{L_d\}_{d \in CHN}$, are fixed. The competitive equilibrium with immobile labor can be defined by including $\{L_d\}_{d \in CHN}$ as additional fundamentals, and removing the free labor mobility condition from Definition 1. The aggregate welfare is defined as the income weighted average consumer utility of domestic regions:

$$\log W \equiv \sum_{d \in CHN} \omega_d \log U_d,$$

where $\omega_d = \frac{I_d L_d}{Y}$, with (I_d, Y) evaluated at the calibrated equilibrium. This definition ensures that Proposition 1 still applies, so it is comparable to the welfare defined in the benchmark model with free labor mobility.

Model with mobile labor subject to migration costs. Assume domestic cities are endowed with initial numbers of workers $\overline{L}_o$. The utility a worker ς from city o would obtain by migrating to region d is:

$$\frac{U_d}{d_{od}} \epsilon_d(\varsigma),$$

where U_d is the utility living in city d that is specified in Subsection 5.1, d_{od} is the iceberg migration cost for migrating from o to d, and $\epsilon_d(\varsigma)$ is an idiosyncratic preference shock that is drawn from the Fréchet distribution and assumed to be i.i.d. across o, d, ς. Under the optimal migration decision, the fraction of workers from city o that migrate to city d is thus

$$\pi^e_{od} = \frac{\left(\frac{U_d}{d_{od}}\right)^{\theta_e}}{\sum_{d'}\left(\frac{U_{d'}}{d_{od'}}\right)^{\theta_e}},$$

which implies that the total number of workers in region d satisfies

$$L_d = \sum_o \overline{L}_o \pi^e_{od}. \tag{C.2}$$

The competitive equilibrium with migration costs can be defined by including $(\overline{L}_o, d_{od})$ as additional fundamentals, and replacing the free labor mobility condition from Definition 1 with (C.2). For the calibration of the parameters that enter the migration model block, we set the elasticity of migration $\theta_e = 1.5$, taken from the estimate in Tombe and Zhu (2019). We specify the migration cost d_{od} as a function of geographic and cultural distances, and use the estimates of d_{od} from Fan (2019) for China. We recalibrate the amenities of domestic cities $\{B_d\}_{d\in CHN}$, along with other parameters, such that the equilibrium domestic labor allocations $\{L_d\}_{d\in CHN}$ match the population distribution from the 2010 Census—the same target used in the benchmark calibration. The aggregate welfare is defined as the income weighted average consumer utility of domestic regions:

$$\log W \equiv \sum_{d\in CHN} \omega_d \log U_d,$$

where $\omega_d = \frac{I_d L_d}{Y}$, with (I_d, L_d, Y) evaluated at the calibrated equilibrium. This definition ensures that the aggregate welfare agrees with the model with mobile labor if migration costs are set to zero ($d_{od} = 1, \forall o, d$), or agrees with the model with immobile labor if migration costs are set to infinity.

C.7 Numerical Implementation

Solve the competitive equilibria. We describe the design of the algorithm that makes it possible to load the most intensive part of the computation to GPUs. This enables us to solve equilibria robustly and efficiently, despite the size of the problem (our benchmark model has 323 regions and 25 sectors).[21] The large size of the problem also renders a well-known approach to solve/calibrate this type of model—Mathematical Programming with Equilibrium Constraint (Su and Judd, 2012)—less effective as the Jacobian matrix is a dense matrix with $(323 \times 25)^2$ entries. Our algorithm falls back to a fixed point algorithm described below.

With E^i_d being the total expenditure on intermediate goods in sector i of region d, the minimal system

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

of equations that can be used to solve the equilibrium is[22]

$$
\begin{aligned}
E_o^j &= \alpha^j (w_o + Tr_o) L_o + \sum_i \gamma^{ij} \sum_d \pi_{od}^i E_d^i, \\
w_o L_o &= \sum_i \beta^i [\sum_d \pi_{od}^i E_d^i], \\
P_d^i &= \Big(\sum_o [p_{od}^i]^{1-\sigma} \Big)^{\frac{1}{1-\sigma}},
\end{aligned}
\tag{C.3}
$$

for unknowns (E_d^i, w_o, P_d^i), where $(p_{od}^i, \pi_{od}^i, L_d, Tr_o)$ are auxiliary variables and are evaluated according to[23]

$$
\begin{aligned}
p_{od}^i &= [\kappa^i w_o^{\beta^i} \prod_{j=1}^{S} [P_o^j]^{\gamma^{ij}} \tilde{\tau}_{od}^i] / T_o^i, \\
\pi_{od}^i &= \frac{[p_{od}^i]^{1-\sigma}}{[P_d^i]^{1-\sigma}}, \\
L_d &= \frac{B_d^{\frac{1}{\alpha^0}} \overline{H}_d \left[\frac{(w_d + Tr_d)^{1-\alpha^0}}{\prod_{i=1}^{S} (P_d^i)^{\alpha^i}} \right]^{\frac{1}{\alpha^0}}}{\sum_{d'} B_{d'}^{\frac{1}{\alpha^0}} \overline{H}_{d'} \left[\frac{(w_{d'} + Tr_{d'})^{1-\alpha^0}}{\prod_{i=1}^{S} (P_{d'}^i)^{\alpha^i}} \right]^{\frac{1}{\alpha^0}}} L_{CHN}, \forall d \in CHN, \\
Tr_o &= \frac{\alpha^0}{1-\alpha^0} \frac{1}{L_{CHN}} \sum_{d \in CHN} w_d L_d, \forall o \in CHN, \\
Tr_{ROW} &= \frac{\alpha^0}{1-\alpha^0} w_{RoW}.
\end{aligned}
\tag{C.4}
$$

We design a nested fixed point algorithm according to the strength of the hardware. A key observation is that given (π_{od}^i, L_d, Tr_o), the first two equations of (C.3) give a (dense) system of linear equations for E_d^i and $w_d L_d$, for which GPUs are designed to solve efficiently. Based on this observation we design the nested fixed-point algorithm below:

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

---

1 Guess $(w_{d,Old}, P_{d,Old}^i, Tr_{d,Old})$
2 Set flag_converged to **false**
while flag_converged is false **do**
  3 Construct (π_{od}^i, L_d) according to (C.4) based on $(w_{d,Old}, P_{d,Old}^i, Tr_{d,Old})$
  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, Tr_d according to (C.4) and (C.3)
  6 Set flag_converged to **true** if distance between (w_d, P_d^i, Tr_d) and $(w_{d,Old}, P_{d,Old}^i, Tr_{d,Old})$ is small enough
  7 Update $(w_{d,Old}, P_{d,Old}^i, Tr_{d,Old})$ according to (w_d, P_d^i, Tr_d)
end while

The step of solving the system of linear equations (line 4 in the algorithm) takes more than 90% of

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

[23] To see the determination of L_d, combine the consumer utility at the optimal choice $U_d \propto B_d \frac{w_d + Tr}{R_d^{\alpha^0} \prod_{i=1}^{S} (P_d^i)^{\alpha^i}}$, the land market clearing condition $R_d \overline{H}_d = \alpha^0 L_d (w_d + Tr)$, and the free mobility condition $U_d = U_{d'}, \forall d, d' \in CHN$.

the computation time in our benchmark model. Starting from an initial guess with uniform entries in (w_d, P_d^i, Tr_d), the benchmark equilibrium can be solved (under the convergence criterion of $1e-6$ in log difference) within a minute with a GTX1080Ti video card, compared to around 10 minutes with 2*Intel Xeon CPU E5-2650 v4.

Calibrate city-sector productivities T_d^i. The indirect inference estimation proceeds in a nested manner. In the inner loop, given other model parameters, we calibrate T_d^i for tradable sectors i such that the sectoral sales ratios between each city and the RoW in the model agree with those in the data. To do this, we treat sales ratios as observables, and solve T_d^i to generate the observable sales ratios while respecting the equilibrium conditions. Specifically, the minimal system of equations for calibrating T_d^i to match the sales ratios M_d^i[24] while respecting the equilibrium conditions are

$$\begin{aligned} M_o^j &= \sum_d \frac{[c_o^j \cdot \tilde{\tau}_{od}^j]^{1-\sigma}}{\sum_{\tilde{o}}[c_{\tilde{o}}^j \cdot \tilde{\tau}_{\tilde{o}d}^j]^{1-\sigma}} \Big(\alpha^j I_d + \sum_i \gamma^{ij} M_d^i\Big) \text{ for all tradable sector } j, \\ w_d L_d &= \sum_i \beta^i M_d^i, \\ P_d^i &= \Big(\sum_o [p_{od}^i]^{1-\sigma}\Big)^{\frac{1}{1-\sigma}}, \end{aligned} \tag{C.5}$$

for unknowns $\Big((T_d^i)_{i \text{ tradable}}, P_d^i, w_d\Big)$[25], where $(I_d, c_d^i, p_{od}^i, Tr_o)$ are auxiliary variables and evaluated according to

$$\begin{aligned} I_d &= (w_d + Tr_d)L_d + D_d, \\ c_d^i &= \kappa^i w_d^{\beta^i} \prod_{j=1}^{S} [P_d^j]^{\gamma^{ij}} / T_d^i, \\ p_{od}^i &= c_o^i \tilde{\tau}_{od}^i, \\ Tr_o &= \frac{\alpha^0}{1-\alpha^0} \frac{1}{L_{CHN}} \sum_{d \in CHN} w_d L_d, \forall o \in CHN, \\ Tr_{ROW} &= \frac{\alpha^0}{1-\alpha^0} w_{RoW}, \end{aligned}$$

with D_d being the exogenous trade deficits which are necessary to match the aggregate import and export shares. The above procedure is done by taking the targeted regional labor L_d as given. After the calibration, the relative amenities B_d are backed out combining the following: (1) the consumer utility at the optimal choice $U_d \propto B_d \frac{w_d + Tr_d}{R_d^{\alpha^0} \prod_{i=1}^{S} (P_d^i)^{\alpha^i}}$, (2) the land market clearing condition $R_d \overline{H}_d = \alpha^0 L_d (w_d + Tr)$, and (3) the free mobility condition $U_d = U_{d'}, \forall d, d' \in CHN$.

Calibrate remaining model parameters. With the inner loop inverting T_d^i to match M_d^i exactly, in the outer loop we search over other parameters to target the rest of the moments. These parameters include the sectoral international trade costs τ_{RoW}^i, the trade cost level parameter h_0, and the alternative mode

[24] M_d^i in the model is the total sales of intermediate goods from sector i of region d, and is linked to E_d^i defined before via $M_o^i = \sum_d E_d^i \pi_{od}^i$.

[25] Notice the system of Equation (C.5) is homogeneous of degree one in T_d^i for any i. That is, fixing i, scaling up T_d^i by the same factor scales nominal price and wage proportionally but does not affect real allocations. Therefore, we normalize $T_d^i = 1$ for a chosen region d for all i.

cost $\bar{\kappa}$. Since the number of parameters is equal to the number of moments, calibrating these parameters is to solve the system of equations such that the model moments are equal to their data counterparts listed in Table 4. We solve the system of equations using an iterative procedure based on a line search method. The equations are solved such that the maximum distance between the data moments and the model moments is less than 1%, and the maximum difference in the inner loop is smaller than $1e-5$.

References

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

_ and _ , "The Welfare Effects of Transportation Infrastructure Improvements," *NBER Working Paper No. 25487*, 2019.

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.

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.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian, "On the Road: Access to Transportation Infrastructure and Economic Growth in China," *Journal of Development Economics*, 2020, p. 102442.

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

Baum-Snow, Nathaniel, J Vernon Henderson, Matthew A Turner, Qinghua Zhang, and Loren Brandt, "Does Investment in National Highways Help or Hurt Hinterland City Growth?," *Journal of Urban Economics*, 2020, *115*, 103124.

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

Combes, Pierre-Philippe and Laurent Gobillon, "The Empirics of Agglomeration Economies," in "Handbook of Regional and Urban Economics," Vol. 5, Elsevier, 2015, pp. 247–348.

Coşar, A Kerem and Banu Demir, "Domestic Road Infrastructure and International Trade: Evidence from Turkey," *Journal of Development Economics*, 2016, *118*, 232–244.

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.

Eaton, Jonathan and Samuel Kortum, "Technology, Geography, and Trade," *Econometrica*, September 2002, *70* (5), 1741–1779.

Faber, Benjamin, "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System," *Review of Economic Studies*, 2014, *81* (3), 1046–1070.

Facchini, Giovanni, Maggie Y Liu, Anna Maria Mayda, and Minghai Zhou, "China's "Great Migration": The Impact of the Reduction in Trade Policy Uncertainty," *Journal of International Economics*, 2019.

Fajgelbaum, Pablo D and Edouard Schaal, "Optimal Transport Networks in Spatial Equilibrium," *Econometrica*, 2020, *88* (4), 1411–1452.

Fan, Jingting, "Internal geography, labor mobility, and the distributional impacts of trade," *American*

Economic Journal: Macroeconomics, 2019, *11* (3), 252–88.

Fan, Shenggen and Connie Chan-Kang, *Road Development, Economic Growth, and Poverty Reduction in China*, Vol. 12, Intl Food Policy Res Inst, 2005.

Hillberry, Russell and David Hummels, "Trade Responses to Geographic Frictions: A Decomposition Using Micro-Data," *European Economic Review*, 2008, *52* (3), 527–550.

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

Rauch, James E, "Networks Versus Markets in International Trade," *Journal of International Economics*, 1999, *48* (1), 7–35.

Roberts, Mark, Uwe Deichmann, Bernard Fingleton, and Tuo Shi, "Evaluating China's Road to Prosperity: A New Economic Geography Aapproach," *Regional Science and Urban Economics*, 2012, *42* (4), 580–594.

Shi, Hao and Shaoqing Huang, "How Much Infrastructure is Too Much? A New Approach and Evidence from China," *World Development*, 2014, *56*, 272–286.

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

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

Tombe, Trevor and Xiaodong Zhu, "Trade, Migration, and Productivity: A Quantitative Analysis of China," *American Economic Review*, 2019, *109* (5), 1843–72.