On (Un)Congested Roads:

A Quantitative Analysis of Infrastructure Investment Efficiency using Truck GPS Data *

Simon Alder Zheng (Michael) Song Zhitao Zhu

April 28, 2023

Abstract

This study aims to quantify the gains from investments in a transportation network, where the elasticity of driving time to traffic ("congestion elasticity") may differ across roads. We use high-frequency GPS data from half a million Chinese trucks to measure traffic flow and uncover the congestion elasticity heterogeneity in China's city-to-city road links. We find that 34% of the links are uncongested and 28% are associated with a small elasticity. In contrast, using real-time traffic data for highways in England, we find that almost all are associated with a large congestion elasticity. We next incorporate congestion elasticity heterogeneity into a quantitative general equilibrium trade model with optimal route choices and structurally estimate the model. We find the returns to be highly unequal in China, and the heterogeneity in congestion elasticity can account for more than half of the dispersion. Our findings suggest a severe misallocation of road investments in China.

Keywords: Roads, Truck GPS Data, Heterogeneous Congestion Elasticities, Misallocation
JEL Codes: F11, H54, O18, R40

*Alder: Swiss National Bank and The University of North Carolina at Chapel Hill, simon.alder@snb.ch; Song: The Chinese University of Hong Kong, zheng.michael.song@gmail.com; Zhu: The Chinese University of Hong Kong, zhuzhitao@link.cuhk.edu.hk.

The paper was initially circulated under the title "Unequal Returns to China's Intercity Road Network". We thank Nathaniel Baum-Snow, Gilles Duranton, and seminar participants at the Chinese University of Hong Kong, Hong Kong University of Science and Technology, Fudan University, Peking University, Tinbergen Institute, Tsinghua University, China Economics Summer Institute, the NBER Chinese Economy Working Group Meeting, and the SED Annual Meeting for their helpful suggestions and comments. We also thank Lucas Risinger for his excellent research assistance. The views expressed in this paper are those of the authors and do not necessarily reflect those of the SNB.

1. Introduction

Large amounts of resources are devoted to transportation infrastructure across the world each year. In developing countries, the annual investment in road infrastructure is above 1% of GDP,[1] and this share has reached 2.4% in China in recent years.[2] The gain from investments depends crucially on their allocation. However, there are ample anecdotes about inefficient investments. In China, while some roads suffer from heavy congestion, many expressways remain underutilized or even empty most of the time, especially in the central and western regions. To gauge the potential misallocation and its welfare implications, we need systematic quantitative examinations of the benefits and costs of such investments.

There are two key challenges. On the empirical front, to accommodate the possibility of underutilized or uncongested roads, we must forgo the widely adopted approach that assumes homogeneous congestion elasticity, i.e., the elasticity of driving time to traffic, for all roads. Intuitively, the congestion elasticity becomes zero for an uncongested road because the driving time is independent of road traffic. In this study, we utilize high-frequency GPS data of half a million Chinese trucks to simultaneously measure the driving speed and traffic on each city-to-city road link in China. We then apply the classifier Lasso (C-Lasso) developed by Su, Shi and Phillips (2016) to uncover the heterogeneity in congestion elasticity. It turns out that more than one-third of the links are uncongested. Less than 40% of the links, referred to as congested, are associated with a congestion elasticity of 0.62, which is comparable to the estimates of homogeneous congestion elasticity in developed economies (see, e.g., Russo et al., 2021; Allen and Arkolakis, 2022; Herzog, 2022). The remaining links, referred to as semicongested, are associated with a congestion elasticity of 0.14.

On the theoretical front, the difficulty of evaluating road investment is rooted in the global nature of the returns. Investing in one road not only affects the driving speed and traffic on this road but also leads to a rerouting of traffic through other roads and to a reallocation of economic activities across locations. We apply a quantitative general equilibrium trade model with heterogeneous route selection and endogenous trade cost developed by Allen and Arkolakis (2019) to evaluate such welfare effects.[3] We extend the model to incorporate the observed heterogeneity in the congestion elasticity and structurally estimate the extended model to match the driving speed and traffic on each road link. We then compute the returns to road investment using the welfare elasticity implied by the structurally estimated model and the investment cost based on the construction, maintenance, and opportunity cost of land, which we obtain from land transaction and geographic data.

The average return from investing in Chinese roads is 8.9%, and there is a large dispersion in the

[1] According to Figure 16 in Fay et al. (2019), total capital allocations for roads are no smaller than 1 percent of GDP between 2009 and 2016 in 55 developing countries covered by the BOOST database.

[2] The official information from the website of the National Development and Reform Commission shows that the share of road construction investment in GDP has been around 2.4% since the 13th Five-Year Plan in China.

[3] Allen and Arkolakis (2019) is the working paper version of Allen and Arkolakis (2022). The former incorporates optimal route choice into an Armington model, while the latter incorporates that into a Ricardian model as in Eaton and Kortum (2002). We use an Armington model and thus cite the former. When we discuss aspects common to both papers or new in the latter, we cite Allen and Arkolakis (2022) accordingly.

returns. The average returns are 29.3% for congested links, 2.2% for semicongested links, and negative for uncongested links. The standard deviation across all links amounts to 29.7%. Alternative assumptions on labor mobility, agglomeration effects, and productivity growth have limited effects on the mean and standard deviation of these returns, which are, however, sensitive to the heterogeneity in congestion elasticity. We re-estimate a version of the model that restricts the congestion elasticity to be homogeneous. In that model, the mean of the returns drops from 8.9% to 3.3%, and the standard deviation is almost halved, from 29.7% to 12.4%.

The misallocation literature pioneered by Restuccia and Rogerson (2008) and Hsieh and Klenow (2009) provides a natural welfare interpretation of the highly unequal returns in China. However, the interpretation is complicated by the congestion elasticity heterogeneity and externalities that might violate the conditions for convexity of the optimal investment problem (Fajgelbaum and Schaal, 2020). To alleviate this concern, we numerically show that the dispersion of the returns can be mapped to the severity of misallocation or, equivalently, the potential for further welfare improvement. Using random groups of road links, the numerical exercises examine the cross-group relationship between the standard deviation of returns within a group and the welfare gain from optimally investing a fixed budget in the group. Consistent with the misallocation literature, we find the two indicators to be strongly positively correlated. A one-percentage-point increase in the standard deviation is associated with an average increase of 0.35 percentage points in returns on the optimal investment. We further examine whether China's road expansion between 2014 and 2018 was directed by the returns and find zero correlation. Moreover, the actual road investments in many regions turn out to be orthogonal to investments that have the same budget but are welfare-maximizing. These results strengthen the interpretation that the large dispersion in China's returns to road investments is a manifestation of misallocation.

Distinguishing between congested and uncongested road links is a cornerstone of the paper. The estimation is based on cross-link high-frequency panel data on driving time and traffic, which are constructed from the detailed spatiotemporal data on trucks' location and speed from their GPS records and matched to China's city-to-city road network. A homogeneous congestion elasticity, such as the one in the congestion equation proposed by Vickrey (1967), can be estimated using a standard fixed effects panel regression. However, the elasticity may be heterogeneous. We use C-Lasso to allow for cross-link heterogeneity in congestion elasticity. To address the concern of potential measurement errors in the high-frequency variables, we exploit supply-side shocks to construct an instrumental variable (IV) for the regressions. In the two-group specification, C-Lasso with an IV, implemented by penalized generalized method of moments (PGMM), distinguishes the congested road links with a positive congestion elasticity from the uncongested road links with a nonpositive congestion elasticity. In the three-group specification, C-Lasso further distinguishes between the congested road links with a large congestion elasticity and the semicongested road links with a small congestion elasticity. Importantly, congestion elasticity is a fundamental determinant of the degree to which traffic density affects speed, and congestion heterogeneity is not simply reflecting differences in traffic density or in observable link characteristics such as geography.

We conduct two main external validity checks. The first applies the same method to England, for which we can construct a comparable high-frequency panel dataset using detailed traffic data from more than 7,000 measurement sites along the highway network. We find that almost all the interregional highways in England are congested and associated with a large congestion elasticity, comparable to the estimate in Allen and Arkolakis (2022) for the US. The stark contrast suggests the importance of considering heterogeneous congestion elasticities for developing countries like China. We next examine the empirical relevance of the distinction between congested and uncongested road links through the lens of the fundamental law of road congestion (Duranton and Turner, 2011). Specifically, we estimate the elasticity of traffic with respect to cross-link variation in road capacity and expect the unit elasticity to hold only for congested links. Inspired by Duranton and Turner (2012) and Baum-Snow et al. (2020), we use China's road network in 1962 as an instrument with which to address the endogeneity of road capacity. The estimated elasticity is 1.1 and 0.7 for congested and uncongested links, respectively. Notably, the estimate for the uncongested links is significantly below one, implying that the uncongested links identified by C-Lasso are qualitatively different from the congested links as predicted by the fundamental law of road congestion.

Our model, structural estimation, and welfare calculations are extensions of Allen and Arkolakis (2019). The model is structurally estimated to minimize the distance between the predicted and observed traffic and driving time on each road. The match is reasonably good for traffic and nearly perfect for driving time, showing that the estimation strategy in Allen and Arkolakis (2019) can easily be extended to the model with heterogeneous congestion elasticities. The opportunity cost of land is a novel aspect in the calculation of total costs. We use the price of geocoded land transactions along road links to calculate the opportunity cost. Ignoring the opportunity cost would inflate the mean value of the returns on the congested links by a factor of two.

Our paper contributes to the following five broad areas in the literature. First, our analysis is closely related to the literature on quantitative general equilibrium trade models (see, e.g., Allen and Arkolakis, 2014; Redding, 2016; Redding and Rossi-Hansberg, 2017), especially the growing literature that quantifies the general equilibrium effects of transportation infrastructure (see, e.g., Redding and Turner, 2015; Donaldson and Hornbeck, 2016; Donaldson, 2018; Fajgelbaum and Schaal, 2020). In particular, our work is built on the framework of Allen and Arkolakis (2019). We deviate from these papers by introducing and quantifying heterogeneous congestion elasticities across groups of links. Although congestion has been shown to be prevalent in developed countries, like in the US according to Allen and Arkolakis (2022) and in England according to our analysis in Section 3.1.2, our GPS data on trucks in China reveal that the heterogeneity in congestion across links is a crucial feature. We show that characterizing the heterogeneity in the congestion elasticity is important for an accurate welfare evaluation and that ignoring this dimension would not only bias the magnitude of returns but also result in an underestimation of potential misallocation.

Our second contribution is the application of large, high-frequency GPS data of trucks and incorporating these data into a quantitative general equilibrium trade model. Such truck-level data are well-suited

for studying important questions in trade and transportation and have recently been used to explore various topics, including market power (Allen et al., 2020), the costs of delays and stops (Hernández, 2022), and internal trade barriers between regions (Krishna and van Leemput, 2018). In this paper, we combine the high-frequency GPS data and C-Lasso developed by Su, Shi and Phillips (2016) to estimate heterogeneous congestion elasticities across links, which are then used to discipline our theoretical model in the quantitative exercises. Thus, our exercises demonstrate the promising potential of big data and machine learning methods in combination with quantitative general equilibrium trade models.

The third contribution relates to resource misallocation, which has been studied extensively since Restuccia and Rogerson (2008) and Hsieh and Klenow (2009) and is reviewed in Jones (2013) and Restuccia and Rogerson (2017). Instead of focusing on resource allocation at the firm level, we explore the allocative efficiency of infrastructure investments across road links, which is more complicated due to the complementarities between investments across links. In the literature using quantitative general equilibrium trade models, Alder (2023) and Alder and Kondo (2019) examine the efficiency of road networks at the extensive margin, while Fajgelbaum and Schaal (2020) study the intensive margin of road network investments by focusing on the efficiency of allocating resources to expand a link. We focus on the intensive margin of road networks as in Fajgelbaum and Schaal (2020), but we differ from their study by introducing the heterogeneous congestion elasticities and allowing for externalities in productivities and amenities.

Our external validity check of the congestion measure also contributes to the literature on the fundamental law of road congestion, which is well documented by several studies in congested urbanized areas (see, e.g., Duranton and Turner, 2011; Hsu and Zhang, 2014; Chen and Klaiber, 2020; Ossokina, van Ommeren and van Mourik, 2021; Garcia-López, Pasidis and Viladecans-Marsal, 2022). Our work differs from the existing work by focusing more on intercity roads instead of within-city roads and by examining different groups of links based on their congestion elasticities. We find that the fundamental law of road congestion, as expected, holds for congested intercity road links.

Finally, we directly speak to the literature examining the impact of China's transportation infrastructure, including reduced-form studies (see, e.g., Faber, 2014; Baum-Snow et al., 2017, 2020; Banerjee, Duflo and Qian, 2020; He, Xie and Zhang, 2020), as well as quantitative studies (see, e.g., Roberts et al., 2012; Alder and Kondo, 2019; Fan, Lu and Luo, 2019; Xu and Yang, 2021; Ma and Yang, 2022; Egger, Loumeau and Loumeau, 2023). Some scholars have also evaluated investment efficiency at the aggregate level (see, e.g., Bai and Qian, 2010; Li, Wu and Chen, 2017; Wu, Feng and Wang, 2021). Our contribution is to introduce heterogeneity in road utilization into such an evaluation, which is usually overlooked in the previous literature, largely because of data unavailability. Our analysis reveals that taking this dimension into account allows for a more accurate evaluation, especially regarding the potential overestimation of the returns for uncongested links. Another difference between our work and the existing literature is that our approach allows us to extend the welfare evaluation from the aggregate to the link level. We are able to identify the return on investment in each link while taking into account the differences in

congestion elasticities, which is useful for deciding on the allocation of future investments. Furthermore, by comparing different investment scenarios (uniform vs. optimal allocation), we provide a gauge of the possible magnitude of welfare gains if future investments are efficiently allocated.

The remainder of the paper is organized as follows. Section 2 describes the data on cities, roads, and trucks. Section 3 discusses congestion and how it is measured. Section 4 introduces the quantitative general equilibrium trade model and the structural estimation. We then describe the welfare gains from infrastructure investments along different links of the network in Section 5 and analyze the misallocation of road investments in Section 6. Section 7 compares the results with those from the model with homogeneous congestion elasticity. Finally, Section 8 concludes the paper.

2. Data

We combine detailed data on Chinese cities and roads to construct a national road transportation network and then connect it to the spatiotemporal data based on GPS devices in half a million trucks in China. We also use several socioeconomic datasets to obtain regional statistics such as GDP, population, and nighttime luminosity. Furthermore, we use land transaction data to measure the opportunity cost of land and geographic characteristics to compute construction costs.

2.1. Cities and Roads

We treat each prefecture as a "city" or a node in the network. The geographic information is extracted from China's official map, "National Catalogue Service for Geographic Information".[4] The map provides each city's administrative boundary, which we use to compute its geometric center with ArcGIS. We exclude islands, where roads are not well connected to Mainland China. We also exclude northwestern areas (Gansu, Inner Mongolia, Ningxia, Qinghai, Tibet, and Xinjiang), where city size and population density are vastly different from those in other parts of China. The remaining 272 cities are used as nodes in our network.

The roads connecting cities are also obtained from the same map, which contains all county-level-and-above roads. As we focus on intercity transportation, we use national and provincial roads (highways and regular roads) to construct our road network. We obtain the lane information for each road from the navigation map at the Geographic Data Sharing Infrastructure, which is maintained by the College of Urban and Environmental Science, Peking University. We spatially match the navigation map with the official road map to obtain the lane information for each road segment. We then obtain the capacity of each road, which is the product of the average number of lanes of the road and its length.[5]

[4]The four provincial level cities, Beijing, Chongqing, Tianjin, and Shanghai, are treated equally as prefectures.
[5]We use the equidistant projection *WGS 1984 World Equidistant Cylindrical* throughout the paper.

2.2. Construction of Road Network

We construct a road network connecting neighboring cities through the "least-cost" routes from all possible routes connecting two neighboring cities. Following the conventional approach, we first construct the network based on the shortest route connecting the neighboring city-pair. The shortest routes are selected by distance.[6] In some special cases where the shortest route strays from the neighboring city-pair, we treat the city-pair as not directly connected.[7] This approach reduces the number of routes from 1,464 to 1,408.

Two aspects should be noted. First, the shortest routes are not selected by travel time, which is endogenously determined in our analysis, but rather by the road length. Second, shipments between the neighboring city-pair do not necessarily take the shortest routes. Drivers may choose to travel through another city to reach the destination because of their idiosyncratic preferences (Allen and Arkolakis, 2019, 2022), and our data analysis and modeling allow for this possibility.

The shortest route network covers 26.7% and 28.5% of national and provincial roads by length and capacity, respectively. We extend the road network by including the shortest s routes connecting two neighboring cities. The extension is implemented sequentially. The second shortest route is obtained by penalizing those segments that overlap with the shortest route.[8] By construction, the second shortest route differs from the shortest route, except for its essential segments, such as bridges and bottlenecks. The same criterion as that for the shortest route is adopted to drop the second shortest route straying from the neighboring city-pair. Among the neighboring city-pairs with both shortest and second-shortest routes, the total length of the second-shortest route is 81.7% longer than that of the shortest route, indicating a significant difference between the two routes.

We repeat the same procedure to select the sth shortest route by penalizing the segments that overlap with the $s-1$ shortest routes. The solid and dotted lines in Figure A.1 in Online Appendix A.1 plot the road coverage by length and capacity in the network based on the s shortest routes, respectively. We choose $s = 5$ as our benchmark. The corresponding road network covers 78.1% and 79.0% of all national and provincial roads in China in terms of length and capacity, respectively. Higher s have a small effect on the coverage of the corresponding road network.

For notational convenience, in the rest of the paper, we refer to the road network as the network based on the five shortest routes connecting neighboring cities. Figure A.2 in Online Appendix A.1 plots the road network. The routes between two neighboring cities are referred to as *a city-pair link* (or simply *link*), with $length_{kl}$, $dist_{kl}$, $lane_{kl}$ and cap_{kl} denoting the total length of the link for city-pair kl, average distance of this link, average number of lanes on the link, and capacity of the link, respectively. $length_{kl}$ and $dist_{kl}$ are

[6] The routes are obtained by "closest facility analysis" in the "Network Analyst" toolbox in ArcGIS.

[7] Fajgelbaum and Schaal (2020) define a route as straying from two neighboring cities if more than half of the route (by length) is outside the two cities, and we adopt the same criterion.

[8] We assign a penalized length of 1e+10 km (overwhelming the longest segment of 89 km) to the overlapped segments. Imposing a larger penalty has no effect on the sth shortest route.

directly measurable and correspond to the sum and mean of the five shortest paths, respectively.[9] $lane_{kl}$ is the average lane number across all segments along the link weighted by the length of each segment. Capacity, cap_{kl}, is the product of $lane_{kl}$ and $length_{kl}$.

2.3. Truck GPS Data

Trucking is the primary mode of domestic freight transport in China.[10] We obtain real-time truck GPS data from one of China's leading logistics services providers. The data cover a total of 562,980 trucks in 2018, accounting for 7.9% of all trucks in China in that year. Since trucks are primarily used for the shipment of goods across cities, we argue that the truck flows indicated by our GPS data capture mainly city-to-city goods flows.

There are obvious concerns regarding data representativeness, but these concerns are alleviated by the fact that the data are strongly correlated with other measures of economic activity. Using 2019 GPS records from the same source but aggregated to city-pairs, Chen et al. (2020) find truck flows to be strongly correlated with city-level GDP (correlation 0.90). Using our GPS records, we find a similar correlation (0.88) between truck flows and GDP across cities in 2018.

Another concern is related to the representativeness of truck speed for the speed of other vehicles. One may particularly worry about the difference in speed between cars and trucks under different congestion levels. We address this concern by using data from DiDi, China's equivalent of Uber. The DiDi data provide the average speed for each road segment in 5 cities (Chengdu, Jinan, Shenzhen, Suzhou, and Xi'an) in 2018.[11] We use our truck data to construct the average speed for each segment.[12] Figure A.3 in Online Appendix A.1 plots the segment-specific speed of cars (x-axis) and of trucks (y-axis). The speed of trucks is on average 32.4% lower than that of cars. However, the cross-segment correlation is high. We regress truck speed (ln) on car speed (ln) and find an estimated coefficient of 0.83 (with a standard error of 0.02). Moreover, the correlation is robust to grouping segments by speed.

We obtain the following real-time GPS records. For each month, our data cover two days (one weekday and one weekend in the middle of the month), and for each day, our data cover four hours (8 a.m. to 9 a.m., 12 a.m. to 1 p.m., 5 p.m. to 6 p.m. and 9 p.m. to 10 p.m.). These GPS records provide information on the latitude, longitude, and speed of each truck (identified by a unique ID), as well as the time and date of the observation. In theory, GPS records should be received every 10 seconds with a precision of 1e-5 degrees. In practice, there is some variation in the exact frequency of the signal, but this variation is small. In our

[9] The two measures are identical when the road network is based on the shortest path ($s = 1$). When $s > 1$, the overlapped segments are used only once for $length_{kl}$ to avoid double counting.

[10] Road freight, for example, accounted for 72.9% of total freight in China in 2019 according to China's Statistical Yearbook 2020.

[11] There are a total of 3,505 segments in the five cities, with an average length of 2.0 km. The data are publicly available in GAIA Open Dataset.

[12] The DiDi data provide geographic information for each segment. We assign each of the trucks in our sample to a segment when its distance to the segment is no greater than 30 meters.

sample, the frequency of GPS signals ranges from 2 seconds to one minute for 99.3% of the trucks.

We remove GPS records that are either irrelevant or likely to contain serious measurement errors. We first drop those records where a truck keeps zero speed for at least 15 minutes (39.4% of the GPS records) because such cases suggest parking or accidents rather than congestion. We also drop records where speed is either negative or greater than 120 km/h (0.005%), the highest speed limit in China. If the straight-line distance between any two locations of a truck is greater than 120 km per hour, then we drop all records of that truck in that hour (0.001%). We further drop records if they suggest that a truck's latitude or longitude is more than 1 degree away from its previous location (0.03%).

We then assign each observation to a location on a digitized map of China according to the longitude and latitude. We match these points to specific roads. Specifically, for each road, we code all trucks that are at most 30 meters away as driving on the road. In addition, by tracing each truck's exact location over time, we can identify the truck's driving direction, which enables us to select trucks and roads that share the same direction. The dashed line in Figure A.1 in Online Appendix A.1 plots the share of trucks covered by the road network based on the s shortest routes in all trucks on national and provincial roads in our GPS data. Our benchmark network (s=5) covers approximately 85% of trucks.

The above data preparation procedure gives us a comprehensive picture of China's trucks at any snapshot in our sample period. In practice, we treat each minute as a snapshot (recall that more than 99% of trucks have a frequency of GPS signals within a minute) and construct the following minute-by-minute variables from our truck data:

$speed_{klt}$: average speed among all trucks on link kl within minute t;

$traf_{klt}$: number of different truck IDs on link kl within minute t;

$density_{klt} \equiv traf_{klt}/cap_{kl}$: number of trucks per unit of area (1 km $\times$ 1 lane) on link kl in minute t.

We drop all observations in February and outliers with an absolute value of link-specific z scores above 3.[13] Figure 1 illustrates the GPS data as a heatmap, and Table A.1 in Online Appendix A.1 reports the summary statistics of the minute-by-minute density and speed of all the trucks used in the empirical analysis.

In addition to the congestion estimation in Section 3.1, where we exploit cross-minute variation within a year, the other analyses, including the empirical analysis for the fundamental law of road congestion, structural estimation, and quantitative analysis of returns and welfare, focus on the cross-sectional variation. Therefore, we aggregate over minutes within each year and obtain the following variables:

[13] The Chinese New Year for 2018 started on February 16. The number of trucks in February is only one-fourth of the average across other months. z_{klt} is constructed for each link-specific variable x_{klt}, with $z_{klt} \equiv |\frac{x_{klt}-mean_{kl}}{std_{kl}}|$, where $mean_{kl}$ and std_{kl} are the time-series mean and standard deviation of x_{klt}, respectively.

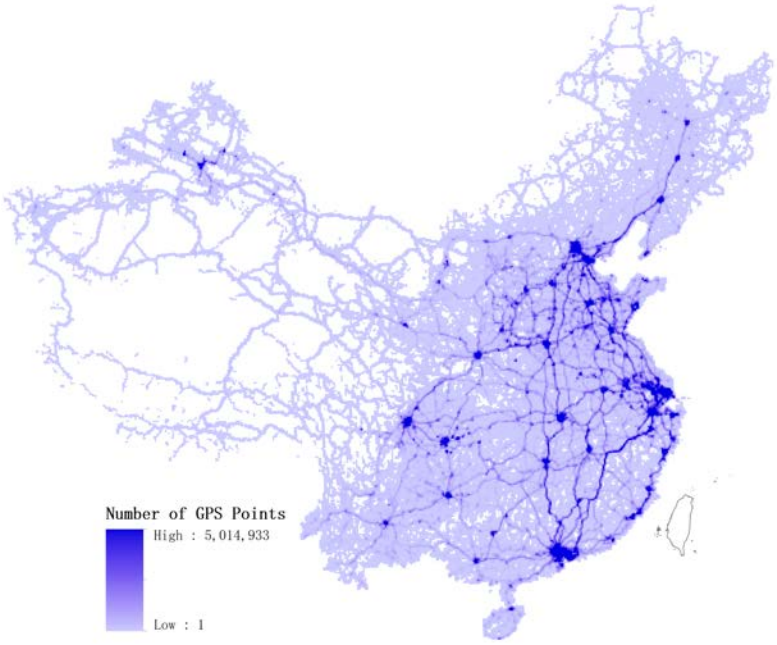

Figure 1. Heat Map of GPS Records of Heavy Trucks

Notes. This heatmap is based on the number of truck GPS signals in 2018 in cells with a size of 10 km by 10 km.

$speed_{kl}$: mean value of $speed_{klt}$ across all minutes within a year ;

$traf_{kl}$: mean value of $traf_{klt}$ across all minutes within a year.

The remaining three variables can be constructed correspondingly as

$density_{kl} \equiv traf_{kl}/cap_{kl}$: average number of trucks per unit of area (1 km $\times$ 1 lane) on link kl;

$hours_{kl} \equiv dist_{kl}/speed_{kl}$: average driving hours from cities k to l along link kl;

$flow_{kl} \equiv density_{kl} \times speed_{kl} \times lane_{kl}$: average traffic flow, or number of trucks, passing through link kl within an hour. The units of measure are consistent: $\frac{trucks}{hours} = \frac{trucks}{km \times lanes} \times \frac{km}{hours} \times lanes$.

2.4. Socioeconomic, Geographic and Land Transaction Data

For each city, we obtain GDP and population data for 2014 and 2018 from *China's Urban Statistical Yearbook* and each city's *Statistical Communiqué on Economic and Social Development*, which are used in the structural estimation and quantitative analyses. We also use population density and urban share, which is calculated as the share of nonagricultural population, for 1964 from *The Compilation of China's Population Statistic Materials (1949-1985)* and *Historical China County Population Census Data with GIS Maps (1953-2000)*. These variables serve as control variables in our analysis of the fundamental law of road congestion.

For the same purpose, we also exploit grid level nighttime luminosity from Flint by the China Remote Sensing Satellite Ground Station, Chinese Academy of Science, which is constructed based on the original data from Version 1 VIIRS Day/Night Band Nighttime Lights, with a grid size of 15 arc-seconds.

In addition, we also utilize geographic digital elevation model (DEM) data from Shuttle Radar Topography Mission (SRTM) digital elevation database with a spatial resolution of 90 meters. Based on the database, we calculate grid level geographic slope in percent using the "Slope" tool in ArcGIS, which is used to construct the geography-based shortest path and to scale road construction costs. We also use the DEM data to calculate grid level geographic ruggedness based on the tool of "Terrain Ruggedness Index (TRI)" from the toolbox of Arc Hydro developed by Riley, DeGloria and Elliot (1999). This variable is used to construct a control variable in our regressions of the fundamental law of road congestion.

Furthermore, we obtain land transaction data between 2013 and 2017 from `https://www.landchina.com`, which covers all land transactions in China's primary land market. The data of each land transaction include several useful variables, e.g., the address of the land parcel, the size, total payment, the date of transaction, land usage (industrial versus commercial), and so forth. We are able to locate each land parcel as a point in the map by geocoding the address via GaoDe Map API and to recover the unit price of each transaction based on the total payment and size. This information is used for our calculation of the opportunity cost of land.

3. Empirical Analysis of Congestion

Our GPS data include information on the location and speed of trucks, which we can use to measure the overall density of trucks on a given road segment and at a given time. The combination of speed and traffic density allows us to estimate a congestion equation and identify congested links in the network. Using variation in link capacity across roads, we also test the fundamental law of road congestion and compare congested and uncongested roads.

3.1. Congestion Measurement

We exploit the correlation between driving hours and vehicle density in *time series* data to measure congestion. The idea is illustrated in the two panels in Figure 2. The left and right panels plot the correlations for two links connected to Shanghai and Xi'an (the capital city of Shaanxi Province), respectively. As expected, the link connected to Shanghai is more crowded and slower than that connected to Xi'an. The key observation is that driving hours (inverse of speed) are positively correlated with truck density in the link connected to Shanghai. In contrast, no correlation between these two factors can be found in the link connected to Xi'an. Through the lens of a congestion equation, such as that proposed by Vickrey (1967), we interpret the links connected to Shanghai and Xi'an as congested and uncongested, respectively.

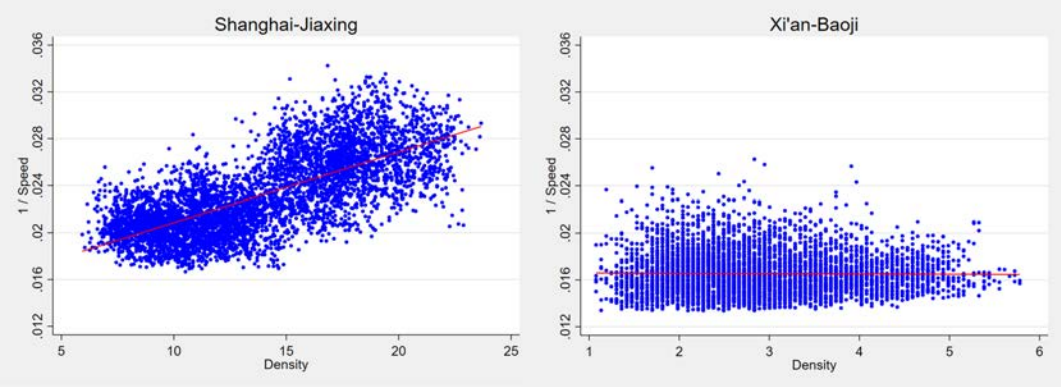

Figure 2. Examples of Links with and without Congestion

Notes. Each dot represents a link-minute observation from 2018 truck GPS data. The red line is the linear fit. The average driving speeds are 44.2 km/h and 61.1 km/h in the left and right panels, respectively.

To characterize the spatial distribution of congestion, we generalize the congestion equation in Vickrey (1967) and Allen and Arkolakis (2019) as follows:

$$\frac{1}{speed_{klt}} = \frac{1}{\delta_{0kl}} + \delta_{1kl}\left(\frac{traf_{klt}}{cap_{kl}}\right)^{\delta_{2kl}}. \tag{1}$$

Recall that $traf_{klt}$ denotes the number of trucks on city-pair link kl at time t and $speed_{klt}$ is the average speed of trucks on kl at time t.[14] δ_{0kl} can be interpreted as the free-flow speed. δ_{1kl} and δ_{2kl} govern the relationship between speed and density, which is allowed to differ across links. Note that, different from the standard congestion equation, δ_{ikl}, with $i \in \{0, 1, 2\}$, is link specific. δ_{2kl} is the parameter of interest and is referred to as the elasticity of congestion. Allowing the congestion equation to be link specific is the key deviation of our work from the specifications set forth in Vickrey (1967) and Allen and Arkolakis (2019).[15]

The transportation literature has documented that speed and traffic may be positively correlated on uncongested roads because of selection (see, e.g., Notley, Bourne and Taylor, 2009). We present a toy model in Online Appendix A.2 to illustrate the mechanism. Importantly, this selection biases downward δ_{2kl} only for uncongested roads and not for congested roads. Therefore, we expect δ_{2kl} to be positive only for congested links and to be zero or negative for uncongested roads.

[14] In Vickrey (1967), the left-hand variable is $hours_{klt}/dist_{kl}$, where $hours_{klt}$ denotes the travel duration from k to l at time t. Note that $1/speed_{klt}$ is equal to $hours_{klt}/dist_{kl}$. We use speed, as it is directly observed in the data.

[15] An alternative specification of congestion is used in Fajgelbaum and Schaal (2020), where the strengths of the effects of traffic flow and road capacity on shipping costs are allowed to be different from each other. We estimate the effect of traffic density on speed by exploiting time variation, and we allow the effect to be heterogeneous across links. The strength of road capacity could not be separately identified in our estimation because cap_{kl} is time-invariant and would be absorbed by link fixed effects FE_{kl}, as we illustrate in equation (2).

To estimate the congestion equation, we rewrite equation (1) as

$$\ln\left(\frac{1}{speed_{klt}}-\frac{1}{\delta_{0kl}}\right)=\delta_{2kl}\times\ln\left(traf_{klt}\right)+FE_{kl}+\epsilon_{klt}. \tag{2}$$

Here, δ_{1kl} and $\delta_{2kl}\times\ln\left(cap_{kl}\right)$ are absorbed by link fixed effects, FE_{kl}. When δ_{2kl} is restricted to be identical across links, equation (2) reduces to a fixed effects regression. Without loss of generality, we consider a group pattern of a general form for δ_{2kl}:

$$\delta_{2kl}=\sum_{h=1}^{H}\vartheta_h\mathbb{1}\left\{kl\in\mathcal{H}_h\right\}, \tag{3}$$

where H is the number of groups, $\mathcal{H}_h$ is the set of links in group h and $\mathbb{1}$ is the indicator function. To estimate equation (2), where δ_{2kl} follows equation (3), we employ the C-Lasso developed by Su, Shi and Phillips (2016). The estimation method identifies the latent group structure in panel data and we outline the basic idea and the econometric framework in Web Appendix B.1. In our context, this method can estimate the heterogeneity in the elasticity of congestion across links. The heterogeneity could be due to link characteristics such as geographical features or the position of the link in the network, and we experiment with various observable characteristics that might explain such heterogeneity, but the heterogeneity remains largely unexplained. Importantly, congestion elasticity is a fundamental determinant of the degree to which traffic density affects speed.

$speed_{klt}$ and $traf_{klt}$ are computed minute by minute in the four hours each day. Since our data cover one working day and one weekend day in each month of 2018 (except February), there are at most 5,280 observations of $speed_{klt}$ and $traf_{klt}$ for each link. Not all links have truck observations for every minute. Hence, our sample is an unbalanced panel, where the average number of observations for a link is 4,817. δ_{0kl}, the free-flow speed, is proxied by the maximum observed speed in the link.[16] Different from Allen and Arkolakis (2019, 2022), our estimation exploits time variation within a link rather than cross-link variation.

$traf_{klt}$ is measured by the number of trucks in our sample, which is only a small proportion of actual traffic. To correct the attenuation bias caused by measurement errors, we employ the following IV to estimate the congestion equation.

$$traf_{klt}^{IV}=\frac{1}{N_k-1}\sum_{m\in\mathcal{N}(k)\setminus\{l\}}\omega_{km}traf_{kmt}, \tag{4}$$

where $\mathcal{N}\left(k\right)$ indicates the set of $k's$ neighboring cities, N_k is the number of $k's$ neighboring cities and ω_{km} is the weight measured as the average number of trucks on link km across all t. We argue that $traf_{klt}^{IV}$ captures the "supply-side shocks" originating from the departure city of the pair, which can affect $speed_{klt}$

[16]For the links with the maximum observed speed exceeding the speed limit, we proxy δ_{0kl} by the link-specific speed limit, which is the average speed limit weighted by the length of each road in the link. The results are essentially the same when we use the speed limit to proxy δ_{0kl} for all links.

only through $traf_{klt}$. Based on this argument, we use $traf^{IV}_{klt}$ to instrument $traf_{klt}$.[17] We drop link kl in $traf^{IV}_{klt}$ because keeping it would violate the exclusion restriction. We cannot construct $traf^{IV}_{klt}$ for the links where the departure city has only one neighbor. As a result, the number of links in the estimation decreases from 1,408 to 1,392.

Our high-frequency data are at one-minute intervals. However, it is conceptually difficult to rationalize $traf_{klt}$ at the same frequency as that of supply-side shocks. Moreover, one-minute traffic averaged over all the links originating from the same city may still contain measurement errors. We, therefore, use the daily average value of $traf^{IV}_{klt}$ to instrument minute-by-minute $traf_{klt}$. Standard errors are clustered at the day level. Since our sample has one weekday and one weekend day for each month, changes in daily average of $traf^{IV}_{klt}$ may also reflect monthly changes in city k's supply-side factors.[18]

Our IV is used to correct for measurement errors in $traf_{klt}$ and does not address the selection bias driven by switchers between links, as presented in a toy model in Online Appendix A.2. It should be noted that the selection does not bias the estimate of the congestion elasticity for congested links. To the extent that our focus is on the classification between congested and uncongested links and the congestion elasticity for congested links, the estimation bias for uncongested links would be innocuous to our results. Moreover, we will provide evidence from cross-sectional variation that our results are in line with the fundamental law of road congestion.

The IV estimation is implemented by PGMM, as in Su, Shi and Phillips (2016) (see Online Appendix A.3 for the estimation procedure). The value of the information criteria (IC) decreases substantially when the number of groups, H, increases from 1 to 2, suggesting a significant improvement in the fit by allowing two different values of δ_{2kl}. The value of the IC decreases further when H increases to 3 and 4 and is minimized at $H = 5$; the marginal gains are, however, much less pronounced compared to those from increasing H from 1 to 2 and from 2 to 3.

The results are reported in Table 1. First, we find strong first-stage results (see Panel B). The daily average traffic departing from city k is indeed highly correlated with $traf_{klt}$, the minute-by-minute traffic on the links originating from k and arriving in l. The second-stage results differ substantially across different h values. When $H = 1$, i.e., δ_{2kl} is restricted to being identical across links, and the estimated elasticity of congestion is 0.15, barely a quarter of the estimate in Allen and Arkolakis (2022) and approximately 60% of that in Herzog (2022).

The estimates turn out to be substantially different once we allow for minimal heterogeneity with $H = 2$. The positive estimate of 0.45 is now much closer to the estimated elasticity of congestion in the literature. These links are regarded as congested. The negative estimate of δ_{2kl} for the other group is

[17] Specification (4) has been widely adopted in the literature. Autor, Dorn and Hanson (2013), for example, "instrument for growth in Chinese imports to the United States using the contemporaneous composition and growth of Chinese imports in eight other developed countries". See also Dube and Naidu (2015), who use changes in US funding to countries outside of Latin America as an instrument for changes in US funding to Colombia.

[18] Alternatively, we can also construct "demand-side shocks": $traf^{IV,demand}_{klt} = \frac{1}{N_l - 1} \sum_{m \in \mathcal{N}(l) \setminus \{k\}} \omega_m traf_{mlt}$. Using daily average $traf^{IV,demand}_{klt}$ as an instrument gives essentially the same results.

consistent with the positive correlations between speed and traffic on uncongested roads, which have been widely documented in the literature. We regard the links in the group with negative δ_{2kl} as uncongested links. With this classification, 58.5% of the links are congested.

Table 1. Estimation of Congestion Elasticity

	(1)	(2)	(3)	(4)	(5)	(6)
	One Group	Two Groups		Three Groups		
	All Links	Congested Links	Uncongested Links	Congested Links	Semi-congested Links	Uncongested Links
Share of Links	100%	58.5%	41.5%	34.4%	28.4%	37.2%
	Panel A: 2SLS					
Truck density (ln)	0.150***	0.453***	-0.399***	0.624***	0.138***	-0.424***
	(0.045)	(0.052)	(0.051)	(0.056)	(0.047)	(0.049)
	Panel B: First Stage					
IV (ln)	0.796***	0.829***	0.744***	0.813***	0.827***	0.755***
	(0.024)	(0.025)	(0.031)	(0.029)	(0.034)	(0.033)
Kleibergen-Paap rk Wald F Statistics	1,082.503	1,084.822	560.834	780.811	603.080	533.347
Link fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Observation	6,802,206	4,057,825	2,744,381	2,389,522	1,960,191	2,452,493

Notes. The table shows the estimates of the congestion elasticity δ_{2kl}. In column (1), the elasticity is constrained to be identical. In columns (2)-(3) we allow for 2 groups. In columns (4)-(6) we allow for 3 groups. Standard errors are reported in parentheses and are clustered at the level of link and day. We use the group classification implied by optimized tuning parameters (see Appendix A.3). * 10%, ** 5%, *** 1% significance level.

We then try $H = 3$. The PGMM separates a group with a small but positive estimate of δ_{2kl} from the congested links. The links in this group are referred to as semicongested links, while those links in the group with the much larger estimate of δ_{2kl} are referred to as congested links. The estimate of the homogeneous elasticity under $H = 1$ seems to pick up the elasticity for the middle group of semicongested links. Interestingly, the estimated elasticity for congested links, 0.62, is remarkably close to the estimate in Allen and Arkolakis (2022) and our estimate for England in Section 3.1.2. Based on this classification, 34.4% of the links are congested and 28.4% of the links are semicongested.

While the value of the IC continues to decrease slightly when $H = 4$ and is minimized at $H = 5$, we find that the gain stems mostly from exploiting heterogeneity among uncongested links, which is not our focus. The composition of congested and uncongested links is robust to larger H values. The marginal gain of IC is also diminishing. Therefore, we choose $H = 3$ as the benchmark case.

The corresponding spatial distribution of congestion is reported in Figure 3, where solid red, dashed orange and dotted green lines represent congested, semicongested, and uncongested links, respectively.

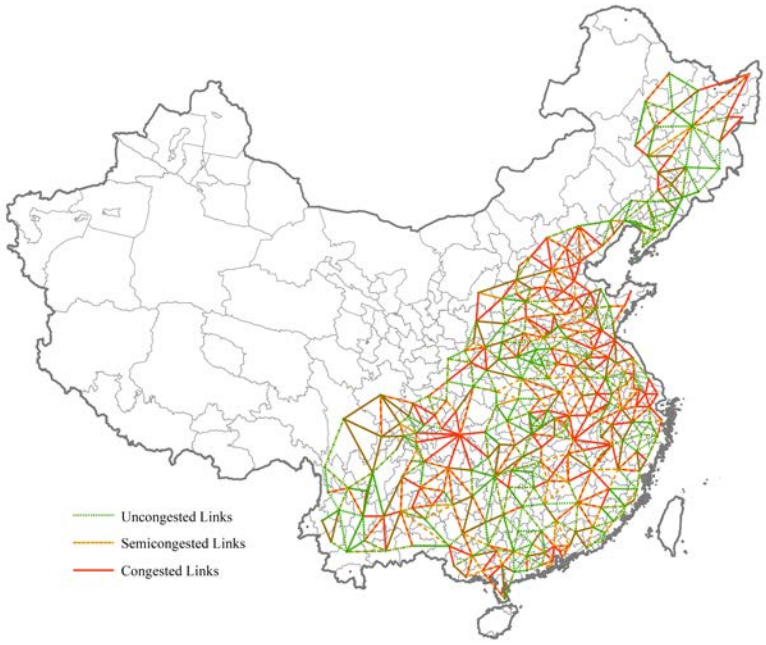

Figure 3. Spatial Distribution of Congestion Elasticities

Notes. Each straight line represents a link. Solid red, dashed orange, and dotted green lines represent congested, semicongested, and uncongested links, respectively. The classification is based on Table 1.

As a robustness check, we also consider road networks based on the shortest path in Table A.2 in Online Appendix A.3, the results of which are robust. For the network based on the shortest path, the PGMM identifies 54.6% congested links under two groups and 31.5% congested and 33.6% semicongested links under three groups.

3.1.1. Cross-Sectional Evidence

δ_{2kl} is estimated by using panel data. If $\delta_{2kl} > 0$ is a reasonable proxy for congestion, then the average speed in the links identified as congested should be low relative to their speed limit, and their average density is likely to be high. Note that the average speed and the average density of a link are absorbed by the fixed effects in equation (2) and have no effect on the estimation of δ_{2kl}. Therefore, the cross-link correlations can serve as an external validity check for our congestion measure.

We estimate an ordered probit model, where the dependent variable is 2 for congested links, 1 for semicongested links, and 0 for uncongested links. Column (1) of Table A.3 in Online Appendix A.4 shows that across links, congestion is negatively correlated with the ratio of average speed to the speed limit, which is consistent with Akbar et al. (2023) who use the difference between real-time and free-flow speeds as a measure of congestion delays. Column (3) shows that congestion is positively correlated with average truck density. Columns (2) and (4) show that the results are robust to adding control variables and their

quadratic terms. We include both the speed ratio and density in column (5). The estimated coefficients again have the expected signs and are stable. Column (6) shows that the result is also robust when aggregating the high-frequency data to 15-minute, instead of 1-minute, intervals.

In addition, the Pseudo R^2 is also informative. The small Pseudo R^2 in column (3) suggests that traffic volume alone cannot account for the heterogeneity in $\delta_{2,kl}$. Controlling for more geographic and network characteristics (e.g., latitude, longitude, height, length, ruggedness, and centrality) doubles the Pseudo R^2 in column (4), but the magnitude is still low, implying that the congestion heterogeneity remains largely unexplained. Hence, we regard the heterogeneity in the congestion elasticity as a basic feature of the current Chinese road network.

3.1.2. Congestion in England

We compare our findings with congestion patterns in England. Online Appendix A.5 provides further details, and we briefly discuss the main results here. England has traffic data from more than 7,000 measurement sites along its highway network. Importantly, English data include information on speed and traffic counts for every 15-minute interval, which allows us to study the relationship between traffic and speed similarly to our analysis for China, even though the origin of the data (from fixed measurement sites along roads) is different from that of the Chinese data (based on GPS devices in trucks). The results show that more than 60% of the links in England are congested and only 2.5% are uncongested (Table A.4 in Online Appendix A.5). In contrast, only one-third of the links in China are congested. Even if we focus on the economically most advanced areas in China, the share of congested links only increases slightly to 36%. However, the estimates of the elasticity of congestion are much closer: 0.51 and 0.53 for congested links in China and England, respectively (Table A.5 in Online Appendix A.5).

3.2. Fundamental Law of Road Congestion

In this section, we apply the idea of the fundamental law of road congestion (Duranton and Turner, 2011) as an external validity check and test whether the law holds for different groups of links, as classified by the estimated congestion equation (2). If the identification of congestion is valid, then we would expect the fundamental law of road congestion to hold for congested links but to fail for uncongested links. The test serves as a further external validity check for our congestion measure. In what follows, we use cross-sectional variations in city-pair link capacity. The potential endogeneity of road capacity across city-pair links could bias the OLS estimate, and we address this by constructing an instrument using historical roads.

We first run OLS regressions that exploit cross-sectional variation in link capacity:

$$\ln(VKT_{kl}) = \gamma_1 + \gamma_2 \times \ln(cap_{kl,-1}) + \rho \times Z'_{kl} + \epsilon_{kl}. \tag{5}$$

The dependent variable VKT_{kl} denotes vehicle-kilometers-travelled through link kl measured in 2018 and

is calculated as the product of the number of trucks passing through link kl, $flow_{kl}$, and link kl's length, $length_{kl}$.[19] The main independent variable $cap_{kl,-1}$ is the link capacity at the end of 2017. Z'_{kl} contains a series of control variables, including the average speed limit on a link, the average nighttime luminosity density and ruggedness along a link.[20] The average speed limit captures the road quality of the link, the average nightlight luminosity measures the density of economic activity, and the ruggedness proxies for geographic characteristics.

The OLS results are presented in Table A.6 in Online Appendix A.6. Column (1) reports the results from all links. The estimated elasticity of VKT with respect to link capacity is 1.1. The magnitudes of the estimated elasticities are similar among congested, semicongested, and uncongested links (columns (2) to (4)).

There are several sources of endogeneity that may bias the estimated elasticity. The main concern is that the correlation can be driven by reverse causality, i.e., link capacity endogenously responds to VKT and demand factors. To estimate the causal effect of road capacity, the literature on transportation infrastructure has developed several IVs. Following Duranton and Turner (2012) and Baum-Snow et al. (2020), we construct an IV based on the 1962 Chinese road network: the historical road density in the 30 km buffer zone of the geography-based least-cost path between two neighboring cities.[21] The relevance and exclusion restriction of the IV are discussed in detail in Online Appendix A.6.

The 2SLS estimation is implemented based on the following specification:

$$\ln(cap_{kl,-1}) = d_1 + d_2 \times \ln(hist_{kl}) + \varrho \times Z'_{kl} + \varepsilon_{kl}, \tag{6}$$

$$\ln(VKT_{kl}) = \gamma_1 + \gamma_2 \times \ln(\widehat{cap_{kl,-1}}) + \rho \times Z'_{kl} + \epsilon_{kl}. \tag{7}$$

Panel B of Table 2 reports the results for the first stage. Our IV is constructed to capture supply-side effects on the current link capacity. We find a highly significant positive correlation between the historical road density in a city pair and the current capacity of the city-pair link. This finding is consistent with those in

[19] $flow_{kl}$ in VKT_{kl} is comparable to "number of vehicles per lane per day passing any point", called "average annual daily traffic (AADT)" in Duranton and Turner (2011); "the average number of vehicles passing through the segment observation point between 7 am and 7 pm on a weekday", called "average weekday daytime traffic (AWDT)" in Hsu and Zhang (2014); and the number of cars passing any noise monitor on the urban road system in 20 minutes in Chen and Klaiber (2020).

[20] We use nighttime luminosity in 2016, i.e., before the year of our road capacity and truck flows observations, to avoid introducing bad controls. We average nighttime luminosity across grids in the 10 km buffer zone along a link. The 1 km buffer zone is dropped because road light emissions from the roads may intensify nighttime luminosity within this buffer. The ruggedness along a link is calculated as the average across all grids in the 1 km buffer zone along a link.

[21] Specifically, the instrument is the length of historical roads in the buffer zone divided by the area of the buffer zone times the length of the geography-based least-cost path. In practice, we follow a similar approach as that of Faber (2014) and discretize the map into a 1 km x 1 km grid, calculating the geography-based cost for crossing each cell as $cost = 1 + slope + 25 \times water$, where slope is measured in percent and water is a dummy variable equal to 1 when the cell belongs to water bodies. We then calculate the least-cost path between neighboring cities using "OptimalPathAsLine" in ArcGIS.

the literature (e.g., Duranton and Turner, 2012; Baum-Snow et al., 2020).[22]

Panel A of Table 2 reports the results for the second stage. We find a large difference in the estimated elasticity between congested and uncongested links, which is statistically significant. However, there is almost no difference in the point estimate between congested and semicongested links. Importantly, the estimated elasticity for uncongested links is significantly below one, violating the fundamental law of road congestion that "predicts that the elasticity of traffic to road capacity is at least 1" (Hsu and Zhang, 2014). Hence, as we would expect, our high-frequency traffic data confirm the fundamental law of road congestion for congested (and semicongested) but not for uncongested links.

Table 2. Fundamental Law of Road Congestion

	(1) All Links	(2) Congested Links	(3) Semicongested Links	(4) Uncongested Links
Dep: VKT (ln)		Panel A: 2SLS		
Capacity (ln, *lag*)	0.959***	1.061***	1.132***	0.699***
	(0.074)	(0.110)	(0.125)	(0.148)
Control variables	Yes	Yes	Yes	Yes
Observation	1,392	479	395	518
R^2	0.627	0.637	0.629	0.598
p value for H_0			0.035	
Dep: Capacity (ln, *lag*)		Panel B: First Stage		
Instrument variable	0.625***	0.697***	0.671***	0.527***
	(0.036)	(0.057)	(0.064)	(0.062)
Control variables	Yes	Yes	Yes	Yes
Observation	1,392	479	395	518
R^2	0.228	0.258	0.251	0.207
Kleibergen-Paap rk Wald F statistic	294.083	150.931	109.878	73.300

Notes. Robust standard errors are reported in parentheses. The sample is the cross-section of 2018. The congestion classification comes from Table 1. The IV is the length of historical roads in the buffer zone divided by the area of the 30 km buffer zone times the length of the geography-based least-cost path, where historical urban areas ("districts" in 1962) are excluded. The control variables include speed limit $(\ln, lag)$, night lights in 2016 $(\ln)$, population density in 1964 $(\ln)$, urban share in 1964 $(\ln)$ and ruggedness $(\ln)$. H_0: the coefficients of Capacity $(\ln, lag)$ in congested (column (2)) and uncongested (column (4)) groups are the same. * 10%, ** 5%, *** 1% significance level.

An important reason to identify congested links by δ_{2kl} is that the estimates of δ_{2kl} are crucial for the structural estimation and the simulation of the returns to road infrastructure, as is shown in Section 4.4.

[22] The literature sometimes uses historical road length as the IV (e.g., Duranton and Turner, 2012; Baum-Snow et al., 2020). Since the area of the buffer zone can be approximated by the product of the length of the geography-based least-cost path and the width of the buffer zone, our IV is essentially the length of the historical road normalized by the width of the buffer zone. Our results are robust to a wide range of buffer zones (see Figure A.8 in Online Appendix A.6). Using the historical road length as an IV generates very similar results. The estimated coefficient of ln *cap* is 0.9 (s.e. 0.2 and statistically indifferent from 1) for congested links and 0.7 (s.e. 0.2 and statistically significantly smaller than 1) for uncongested links.

Putting aside the role of δ_{2kl} in the model, a simple way of measuring congestion is to use the ratio of a link's average speed to its speed limit, as shown in Akbar et al. (2023), which is used in Section 3.1.1. We sort links into two groups in ascending order by the ratio of the average speed to the speed limit (top 35% and bottom 37%). The resulting traffic elasticity is reported in columns (1)-(2) of Table A.7 in Online Appendix A.6. The speed ratio might capture congestion more accurately when geographic obstacles are negligible. Hence, we estimate the traffic elasticity in the subsample of coastal provinces, which are less mountainous than inland provinces. The results are reported in columns (3)-(4) of Table A.7. The results are consistent with our baseline congestion estimation.

4. Model and Structural Estimation

4.1. Quantitative General Equilibrium Trade Model

We consider a discrete version of the quantitative general equilibrium trade model of Allen and Arkolakis (2014), where the economy consists of a finite number of locations indexed by $j \in \{1, \cdots, N\}$. The total number of workers is constant and denoted by $\bar{L}$. The benchmark model assumes perfect labor mobility across locations. Each worker supplies one unit of labor inelastically, which gives wage w_j as the only income, and perfectly free labor mobility implies welfare equalization across locations.

The preferences of a representative agent in location j are

$$U_j = \left(\sum_i q_{ij}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} u_j,$$

where σ is the elasticity of substitution, q_{ij} denotes the consumption of goods from location i, and u_j is the local amenity in location j.

Goods are produced in an Armington setting, where each location produces a distinct variety (with the same index as the producing location). Denote by q_j the quantity of goods produced in location j. The production technology is linear and uses labor as the only input: $q_j = A_j L_j$, where A_j and L_j are productivity and labor in location j, respectively. Assuming perfect competition, we have

$$w_j = p_j A_j, \tag{8}$$

where p_j is the factory-gate price of variety j in its home market.

The income of the representative agent stems from labor income, $Y_j = w_j L_j$, and trade is always balanced, $Y_j = \sum_i X_{ij}$, where X_{ij} denotes the value of location j's imports from location i. Utility maximization leads to a standard gravity equation: the value of location j's imports from location i follows

$$X_{ij} = \tau_{ij}^{1-\sigma} p_i^{1-\sigma} P_j^{\sigma-1} Y_j, \tag{9}$$

where τ_{ij}, p_i, P_j, and Y_j denote the iceberg trade cost, the factory-gate price, the Dixit-Stiglitz price index, and the total income in the location, respectively. The Dixit-Stiglitz price index follows

$$P_j \equiv \left(\sum_i \left(p_i \tau_{ij} \right)^{1-\sigma} \right)^{\frac{1}{1-\sigma}}, \tag{10}$$

where P_1 is normalized to one. The goods market clearing condition is

$$Y_j = \sum_i X_{ji}. \tag{11}$$

We allow for local productivity ($A_j = \bar{A}_j L_j^{\alpha}$) and amenity ($u_j = \bar{u}_j L_j^{\beta}$) externalities as in Allen and Arkolakis (2014), where $\alpha > 0$ and $\beta < 0$ govern the intensity of productivity and amenity spillovers, respectively. $\bar{A}_j$ and $\bar{u}_j$ are location-specific constant terms.

The representative agent's welfare, denoted by W_j, is

$$W_j = \frac{w_j}{P_j} u_j. \tag{12}$$

By embedding the gravity equation (9), welfare equalization, and externalities into the goods market clearing condition (11) and Dixit-Stiglitz price index (10), we can obtain an equation system summarizing the equilibrium of the model

$$w_j^{\sigma} L_j^{1-\alpha(\sigma-1)} = W^{1-\sigma} \sum_i \tau_{ji}^{1-\sigma} \bar{A}_j^{\sigma-1} \bar{u}_i^{\sigma-1} w_i^{\sigma} L_i^{1+\beta(\sigma-1)}, \tag{13}$$

$$w_j^{1-\sigma} L_j^{\beta(1-\sigma)} = W^{1-\sigma} \sum_i \tau_{ij}^{1-\sigma} \bar{A}_i^{\sigma-1} \bar{u}_j^{\sigma-1} w_i^{1-\sigma} L_i^{\alpha(\sigma-1)}. \tag{14}$$

Allen and Arkolakis (2014) prove that with free labor mobility, if $\alpha + \beta < 0$, $-1 < \beta < 0 < \alpha < 1$, then there exists a unique equilibrium with $L_j > 0 \; \forall j$.

4.2. Optimal Path Selection

Following Allen and Arkolakis (2019), we endogenize trade costs through traders' optimal path selection. The N locations in the economy are connected by a road network, where some location-pairs are directly connected while others are indirectly connected through other locations. We characterize the road connection through an infrastructure matrix, $T = [t_{ij}]$, where t_{ij} is the *direct* trade cost from location i to location j. "Direct" implies that $t_{ij} \in [1, \infty)$ if the two locations are directly connected by a link, and $t_{ij} = \infty$ otherwise. We use kl to denote a directly connected location-pair, as opposed to ij for a location-pair that can be directly or indirectly connected.

Consider a continuum of traders, $\nu \in [0, 1]$, who ship goods from i to j. Each trader selects one path

from i to j, which consists of a group of links. Denote by $\Re_{ij}$ the set of all possible paths starting from i and ending at j, and $r \in \Re_{ij}$ represents one possible path. In addition to instantaneous trade costs $\tilde{\tau}_{ij}(r)$,[23] each trader also faces a path-specific idiosyncratic trade cost shock, $\varepsilon_{ij}(r,\nu)$, which is drawn from a *Weibull* distribution:

$$Pr\left[\varepsilon_{ij}(r,\nu) \leq \varepsilon\right] = 1 - \exp\left(-\varepsilon^{\theta}\right), \tag{15}$$

where θ captures the heterogeneity of traders' route preferences. A higher θ value implies less dispersed preferences and, hence, more agreement concerning the optimal path. Therefore, each trader's total trade cost from i to j along path r is $\tilde{\tau}_{ij}(r)\,\varepsilon_{ij}(r,\nu)$. The optimal path for trader ν solves

$$\tau_{ij}(\nu) = \min_{r} \tilde{\tau}_{ij}(r)\,\varepsilon_{ij}(r,\nu). \tag{16}$$

The optimized $\tau_{ij}(\nu)$ is trader ν's total trade cost from i to j. Applying the above equation to all traders, we obtain the average trade cost from i to j, $\mathbb{E}_{\nu}\left[\tau_{ij}(\nu)\right]$, which is τ_{ij} in the general equilibrium model in Section 4.1,

$$\tau_{ij} = \left(c^{-\theta}\sum_{K=0}^{\infty}\sum_{r\in\Re_{ij,K}}\prod_{\iota=1}^{K}a_{r_{\iota-1}r_{\iota}}\right)^{-\frac{1}{\theta}} = \left(c^{-\theta}\sum_{K=0}^{\infty}\mathbf{A}_{ij}^{K}\right)^{-\frac{1}{\theta}} = \left(c^{-\theta}b_{ij}\right)^{-\frac{1}{\theta}}, \tag{17}$$

where $c \equiv \Gamma\left((\theta+1)/\theta\right)$, $\Re_{ijK}$ indicates the set of paths from i to j with length K, $a_{ij} \equiv t_{ij}^{-\theta}$, $\mathbf{A} \equiv [a_{ij}]$, $\mathbf{A}_{ij}^{K}$ is the ij-th element of $\mathbf{A}^{K}$, and $b_{ij} \equiv \sum_{K=0}^{\infty}\mathbf{A}_{ij}^{K}$.[24]

To illustrate the effect of infrastructure on trade costs, let us first take t_{kl} as exogenous parameters and examine their effects on trade costs. According to Allen and Arkolakis (2019), the elasticity of trade cost between any location-pair, τ_{ij}, with respect to t_{kl} is

$$\frac{d\ln\tau_{ij}}{d\ln t_{kl}} = \left(\frac{\tau_{ij}}{\tau_{ik}t_{kl}\tau_{lj}}c\right)^{\theta}. \tag{18}$$

Allen and Arkolakis (2019) also prove that the elasticity can be interpreted as the probability of a trader using link kl when considering moving goods from i to j. Let us denote by Ξ_{kl} the total trade flow through link kl, which is then given by

$$\Xi_{kl} = \sum_{i}\sum_{j}X_{ij}\left(\frac{\tau_{ij}}{\tau_{ik}t_{kl}\tau_{lj}}c\right)^{\theta}. \tag{19}$$

[23] This term is defined as the multiplication of direct trade cost associated with each link along path r. Allen and Arkolakis (2022) show that replacing multiplicative costs by additive costs can also generate approximately equal results.

[24] Here, $\mathbf{B} \equiv \sum_{K}\mathbf{A}^{K}$ if $\sum_{j}a_{ij} = \sum_{j}t_{ij}^{-\theta} < 1$ for $\forall i$.

4.3. Traffic Flow and Road Congestion

We assume that the traffic flow, or the number of trucks passing through a link kl, is determined by the trade flow:

$$\ln\left(flow_{kl}\right) = \zeta + \lambda \ln\left(\Xi_{kl}\right), \tag{20}$$

where $\lambda \in (0, 1]$. Following Allen and Arkolakis (2019), we let $\lambda = 1$ in the estimation. $flow_{kl}$ determines $hours_{kl}$ by the congestion equation (1). We close the model by assuming t_{kl}, the trade cost on direct link kl, to be determined by $hours_{kl}$:

$$t_{kl} = \exp\left(\kappa \times hours_{kl}\right), \tag{21}$$

where $\kappa > 0$, which ensures $t_{kl} \geq 1$, and $hours_{kl}$ denotes the travel duration from k to l. Hence, traffic flow affects trade costs through congestion, which in turn affects traffic flows through optimal path selection. The equilibrium is defined as follows:

Definition 1 (Equilibrium)**.** Consider an economy associated with exogenous location-specific amenities $\bar{u}_j$ and productivities $\bar{A}_j$, and link-specific road capacities cap_{kl}. An equilibrium of the economy consists of location-specific w_j, L_j, p_j, P_j, and Y_j, location-pair-specific τ_{ij} and X_{ij} and link-specific $hours_{kl}$, t_{kl}, Ξ_{kl}, $flow_{kl}$ and $traf_{kl}$ such that:

(i) Given $hours_{kl}$ and, hence, t_{kl} by (21), τ_{ij} is determined by the optimal path choice as in (17);

(ii) The equilibrium conditions, (13) and (14), pin down w_j and L_j, which determines p_j through (8), X_{ij} through (9), P_j through (10), and Y_j through (11);

(iii) τ_{ij} and X_{ij} determine link-specific trade flow, Ξ_{kl}, through (19), which implies link-specific traffic flow, $flow_{kl}$, by (20);

(iv) $hours_{kl}$ and $flow_{kl}$ pin down $traf_{kl}$ by the definition of $flow_{kl}$, which leads to $hours_{kl}^{new}$;

(v) $hours_{kl} = hours_{kl}^{new}$.

4.4. Structural Estimation

We now structurally estimate the model. Following Allen and Arkolakis (2019), we set two predetermined parameters: $\sigma = 9$ and $\lambda = 1$. The structural estimation is independent of α and β; for the welfare analysis, we use $\beta = -0.3$ and $\alpha = 0.1$, as in Allen and Arkolakis (2019). City-level GDP in the structural estimation is always set equal to the data. Therefore, the estimation does not depend on city-specific amenities $\bar{u}_i$ and productivity $\bar{A}_i$. This leaves three parameters to be structurally estimated: the shape parameter for traders' preference heterogeneity, θ; the cost parameter of $hours_{kl}$ in (21), κ; and the constant term in the traffic flow equation (20), ζ.

Let us denote $\phi \equiv \{\theta, \kappa, \zeta\}$. Given ϕ, we simulate the average number of trucks and driving hours for each direct link, denoted by $traf_{kl}^{sim}$ and $hours_{kl}^{sim}$, respectively. The algorithm for solving $traf_{kl}^{sim}$ and $hours_{kl}^{sim}$ is in Online Appendix A.7. We use nonlinear least squares (NLS) to estimate ϕ, i.e., minimizing the distance between $traf_{kl}^{sim}$, $hours_{kl}^{sim}$ and their observed counterparts in the data, denoted by $traf_{kl}^{data}$, $hours_{kl}^{data}$. Note that, by construction, $hours_{kl}^{sim}$ are always identical to $hours_{kl}^{data}$ for uncongested links. We estimate ϕ using

$$\hat{\phi} = \underset{\phi}{argmin} \sum_{kl} \left(\left(\ln\left(traf_{kl}^{sim}\right) - \ln\left(traf_{kl}^{data}\right) \right)^2 + \left(\ln\left(hours_{kl}^{sim}\right) - \ln\left(hours_{kl}^{data}\right) \right)^2 \right). \quad (22)$$

θ and κ affect both $traf_{kl}^{sim}$ and $hours_{kl}^{sim}$. The identification of ζ comes from $hours_{kl}^{sim}$, on which ζ has a first-order effect.[25]

4.5. Estimation Results

The estimation results are reported in Figure 4 and Table 3. Figure 4 shows that the estimated model fits the data well. The correlation coefficient between $\ln(traf_{kl}^{sim(\hat{\phi})})$ and $\ln(traf_{kl}^{data})$ is 0.67, which is higher than the correlation of 0.56 in Allen and Arkolakis (2019). $\ln(hours_{kl}^{sim(\hat{\phi})})$ fits $\ln(hours_{kl}^{data})$ almost perfectly, with a correlation coefficient of 0.97. The estimated parameter values, $\hat{\phi}$, are reported in Table 3. The estimated θ, κ, and ζ are 44.7, 0.02, and -5.7, respectively, and are precisely estimated.

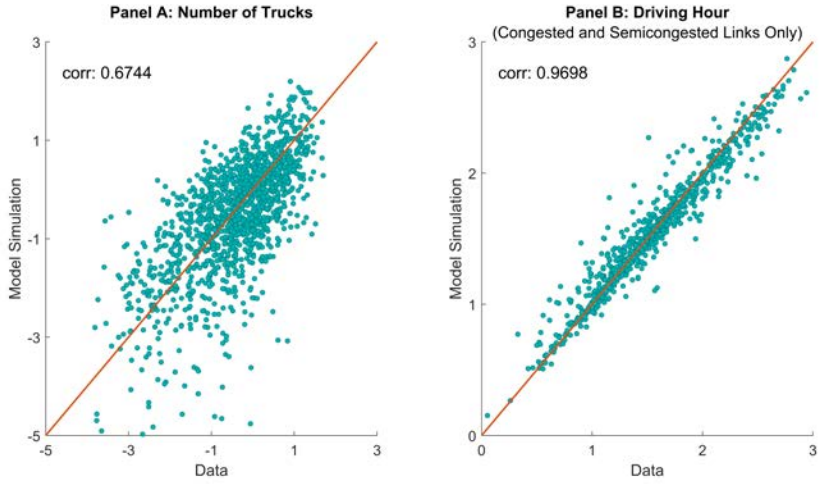

Figure 4. Model Fit

Notes. Panels A and B plot $\ln(traf_{kl})$ and $\ln(hours_{kl})$ in the data and simulated from the estimated model, respectively. Panel B shows only congested and semicongested links, as the simulated $\ln(hours_{kl})$ are, by construction, always identical to those in the data for uncongested links.

[25]Our estimation strategy is almost the same as that in Allen and Arkolakis (2019), but there is one important difference. Allen and Arkolakis (2019) do not include their data on hours from Google Maps in their NLS estimation. Our hours are from the real-time GPS data of actual trips, and we include this variable in the NLS.

For comparison, we also estimate the model using data from England. Figure A.9 in Online Appendix A.7 shows that the fit for $\ln(traf_{kl})$ is not bad for England, with a correlation coefficient of 0.49 for $\ln(traf_{kl})$ and an almost perfect match for $\ln(hours_{kl})$. [26]

Table 3. Results of the Structural Estimation of Model Parameters

	θ	κ	ζ
Estimated parameter	44.74	0.02	-5.68
(Standard error)	(0.45)	(0.00)	(0.00)

Notes. The table shows the results of the structural estimation of equation (22). Standard errors reported in parentheses are calculated through the numerical implementation of the covariance matrix of nonlinear least square estimation from Hansen (2022).

5. Returns to Investment

In this section, we first use the estimated model to compute the welfare gains from expanding the capacity of individual links. With the investment costs calculated from the data, we can compute the returns to capacity expansion on each link.

5.1. Welfare Gain and Elasticity

The welfare gain from expanding the capacity of individual links is calculated as

$$B\left(\Delta_{kl}\right) = \frac{W(\Delta_{kl}) - W}{W} \times GDP, \tag{23}$$

where W and $W(\Delta_{kl})$ represent the welfare in the benchmark model and the counterfactual with the capacity expansion for link kl, Δ_{kl}, respectively. GDP is set equal to its nominal value in 2018.

When we consider a small road expansion, $B\left(\Delta_{kl}\right)$ can be approximated by the "marginal gain":

$$B\left(d_{kl}\right) = \frac{d\ln W}{d\ln lane_{kl}} \times \frac{length_{kl} \times d_{kl}}{cap_{kl}} \times GDP, \tag{24}$$

where $cap_{kl} = length_{kl} \times lane_{kl}$, and d_{kl} indicates the infinitesimal expansion of the lane for link kl. $d\ln W / d\ln lane_{kl}$ is the welfare elasticity derived in Allen and Arkolakis (2019).

$$\frac{d\ln W}{d\ln lane_{mn}} = \sum_{ij}\sum_{kl} \frac{d\ln W}{d\ln \tau_{ij}} \frac{d\ln \tau_{ij}}{d\ln t_{kl}} \frac{d\ln t_{kl}}{d\ln lane_{mn}}, \tag{25}$$

[26] Table A.8 in Online Appendix A.7 shows the estimates for England. We find that the estimate of θ for England is slightly lower than in China, implying that English drivers behave slightly more heterogeneously than those in China. The estimate of κ, i.e., the sensitivity to driving hours, is higher for England. ζ captures the heterogeneity in the coverage of $flow_{kl}$ and in the unit of Ξ_{kl} between the two economies. Its estimates are, therefore, not directly comparable.

where $lane_{mn}$ denotes the average number of lanes on link mn, and we use variations in the number of lanes to represent variations in the capacity of the link.[27] Each of the three parts on the right-hand side of equation (25) can be written as a function of the parameters and observables. We present the formulas below but leave the derivations to Web Appendix B.2.

First, the elasticity of aggregate welfare with respect to trade cost is as follows:

$$\frac{d\ln W}{d\ln \tau_{ij}} = \begin{cases} -\frac{X_{ij}}{\sum_{i'} Y_i'} \times (1+\varsigma_i+\varpi_j) & if\ j \neq s \\ -\frac{X_{ij}}{\sum_{i'} Y_i'} \times (1+\varsigma_i) & if\ j = s \end{cases},$$

where s refers to the city for which nominal wage w_s is normalized to one. In the special case without externalities (i.e., $\alpha = 0$ and $\beta = 0$), $\varsigma_i = 0$ and $\varpi_j = 0$, and lowering trade costs on a link always yields aggregate welfare gains. The aggregate welfare elasticity is then exactly equal to the share of trade flows on the link in aggregate output, which coincides with the finding from Atkeson and Burstein (2010) and Lai, Fan and Qi (2020) that the global gain from the reduction in trade cost can be captured by the share of trade flow in global GDP. The general forms of ς_i and ϖ_j are provided in Web Appendix B.2.

The second term in equation (25) can be written as

$$\frac{d\ln \tau_{ij}}{d\ln t_{kl}} = \left(\frac{\tau_{ij}}{\tau_{ik}t_{kl}\tau_{lj}}c\right)^{\theta}, \tag{26}$$

where $c \equiv \Gamma\left((\theta+1)/\theta\right)$. The elasticity of τ_{ij} with respect to τ_{kl} is proportional to the ratio of the trade cost between cities i and j to that on the path connecting cities i and j through link kl. The elasticity is nonnegative and smaller than 1. As θ increases (traders' heterogeneity decreases), traders' selections become closer to the shortest path between i and j.

Third, by substituting equations (20) and (21) into the congestion equation (1) and differentiating both sides with respect to $\ln lane_{mn}$, we obtain

$$\frac{d\ln t_{kl}}{d\ln lane_{mn}} = \frac{J_{kl}}{1-J_{kl}} \times \kappa \times hours_{kl} \times \left(\lambda \sum_{k'l'} \frac{d\ln \Xi_{kl}}{d\ln t_{k'l'}} \frac{d\ln t_{k'l'}}{d\ln lane_{mn}} - \mathbf{1}\{kl=mn\}\right), \tag{27}$$

where $J_{kl} \equiv \delta_{1kl} \times \delta_{2kl} \times \left(\left(\exp(\zeta) \times \Xi_{kl}^{\lambda} \times hours_{kl}\right)/\left(dist_{kl} \times lane_{kl}\right)\right)^{\delta_{2kl}} \times (dist_{kl}/hours_{kl})$, $J_{kl}/(1-J_{kl})$ is nonnegative, and the formula for $d\ln \Xi_{kl}/d\ln t_{k'l'}$ is presented in Web Appendix B.2. For congested and semicongested links, an increase in lane capacity in a link mn directly affects $hours_{mn}$ and t_{mn}. Moreover, it also indirectly affects t_{kl} on the other links. Trade costs and flows are adjusted accordingly until the system reaches a fixed point. We can obtain the elasticity of t_{kl} with respect to $lane_{mn}$ by solving the equation system (27). For uncongested links, the elasticity is always zero since $hours_{mn}$ or t_{mn} is constant.[28]

[27] Similar to kl, mn also represents a directly connected city-pair.

[28] We compare the welfare elasticities between China and England in Table A.9 in Online Appendix A.8 and find

With the welfare elasticity from equation (25), $B\left(d_{kl}\right)$ can easily be calculated by equation (24). Figure A.10 in Online Appendix A.8 visualizes $B\left(d_{kl}\right)$ for expanding each link at a time in the structurally estimated model, where the parameter values are from Table 3. We will assess the accuracy of approximating the welfare gain by $B\left(d_{kl}\right)$ in Section 5.3.

5.2. Cost Calculation Procedure

We assume the unit construction cost for a link to be proportional to its slope:[29]

$$UC_{kl}^{S} \propto 1 + slope_{kl}. \tag{28}$$

We measure $slope_{kl}$ by the average slope across all grids in the 1 km buffer zone of link kl. We then calibrate the unit construction cost for one specific link so that the average of unit construction costs across all links, weighted by each link's area, matches the national average construction cost of 8.7 million RMB per lane-km in 2013-17.[30] The more detailed procedures for calculating the unit cost are provided in Online Appendix A.8.

Our approach regarding the opportunity cost of land is based on the land price information for each year between 2013 and 2017. After geocoding each land parcel's address as a point using the GaoDe Map API, we assign these parcels to links according to the criterion that the point is within a 1 km buffer zone of the road. Then, for each segment, we calculate the mean value of land prices by taking the average across all land parcels that are matched. By taking the average across all segments of a link, we obtain the average unit land price for each link, denoted by UC_{kl}^{L}.[31]

The unit costs for improving each link are presented in Figures A.11 and A.12 in Online Appendix A.8. As expected, the construction costs are much higher in mountainous southeastern and southwestern areas, and the land prices are much higher in more developed areas like Yangtze and Pearl River Deltas and Beijing-Tianjin-Hebei.

By putting the above two unit costs together, we obtain the total cost of road expansion on link kl:

$$C\left(\Delta_{kl}\right) = \left(UC_{kl}^{S} + UC_{kl}^{L}\right) \times length_{kl} \times \Delta_{kl}, \tag{29}$$

that the elasticity for England is much larger than that for China.

[29]The relationship between construction costs and slope is similar to that in Faber (2014), but his equation also includes other determinants such as urban lands, which are already at least partially captured through our data on land prices.

[30]The data source of the calculation is *Statistical Report of the Development of Transportation Industry*, and the annual construction costs are adjusted to the 2018 price level. This number for the national average construction cost is very close to that obtained from the Roads Cost Knowledge System (ROCKS) developed by the World Bank's Transport Unit according to Collier, Kirchberger and Söderbom (2016), which reports USD 2,838,562 per km for 4-lane expressways measured in 2000 price level, or equivalently 8.9 million RMB per lane-km measured in 2018 price level.

[31]The annual land price for each link is also adjusted to the 2018 price level. We conduct robustness checks regarding different land types in Online Appendix A.8.

The "marginal cost", $C(d_{kl})$, can be defined analogously, with Δ_{kl} replaced by d_{kl}.

5.3. Returns

The return from expanding one link is given by

$$R(\Delta_{kl}) = \frac{B(\Delta_{kl})}{C(\Delta_{kl})} - M, \tag{30}$$

where $B(\Delta_{kl})$ and $C(\Delta_{kl})$ follow equations (23) and (29), respectively, and M denotes the annual maintenance cost. Following Office of the State Auditor (2002) and Allen and Arkolakis (2019, 2022), we assume $M = 5\%$ for both types of costs under the assumption of a 20-year depreciation schedule. The return can also be approximated by $R(d_{kl}) = B(d_{kl})/C(d_{kl}) - M$. We first use $R(d_{kl})$ to report the statistics of the return to investment on each link and then discuss the difference between $R(d_{kl})$ and $R(\Delta_{kl})$ at the end of this section.

The upper-left panel of Table 4 reports the mean and standard deviation of $R(d_{kl})$. The returns to investments in the Chinese city-to-city links are low on average. The mean value is merely 8.9%, which is small compared to the mean of 108% for the segments in the US highway system in Allen and Arkolakis (2022). However, our returns are not directly comparable with those in Allen and Arkolakis (2022) for three reasons. First, the estimated congestion elasticity is 0.74 for all the road segments in their model, while only the congested links are associated with a similar congestion elasticity of 0.62 in our model. If we focus on the congested links, then the mean of their returns increases to 29.3%. Second, the opportunity costs of land account for more than one-third of the total costs of road investments in our model. However, they are either ignored in the literature or approximated by urbanization (e.g., Faber, 2014; Allen and Arkolakis, 2022; Alder, 2023). We show, in the second column of Table A.11 in Online Appendix A.8, that excluding land costs further increases the mean of the returns to 65.8% among congested links. Moreover, despite the low returns on average, investing in some links can generate very high returns. We list the top 20 links according to returns in Table A.12 in Online Appendix A.8. There are 6 links with returns above 200%. When we exclude the opportunity costs of land, 10 links have returns above 400%, comparable to the returns for the top ten links in Allen and Arkolakis (2022). Lastly, the returns reported above are before borrowing costs, while the returns in Allen and Arkolakis (2022) already include borrowing costs. If we assume a borrowing cost of 3% as in Allen and Arkolakis (2022), then the average return would be even lower.

The presence of semicongested and uncongested links is the main reason for the average of the returns being so low in China. Among the semicongested links, the returns are very close to zero. Expanding the uncongested links brings about zero benefits and, therefore, delivers returns of -5%, which capture the costs of depreciation and maintenance. The large variation in returns among the three groups of links also leads to a large dispersion in returns across all links. The standard deviation of the returns is 29.7%, more

than three times as large as the mean. As shown in Section 7, the standard deviation is halved in a model without congestion elasticity heterogeneity.

In addition to the unweighted average of $R(d_{kl})$, it is useful to construct the average of $R(d_{kl})$ weighted by costs: $\sum_{kl} \frac{C(d_{kl})}{\sum_{kl} C(d_{kl})} \times R(d_{kl})$. Links differ in length and construction costs, and these features might be correlated with the returns. Hence, the unweighted average may be affected by how and where the road network is split into individual links. Thus, we prefer the weighted average to the unweighted one, and Table 4 shows that the weighted average of $R(d_{kl})$ is substantially lower than the unweighted average. If we assume a borrowing cost of 3% as in Allen and Arkolakis (2022), then the weighted average return would be zero.

Table 4. Returns to Road Investment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Heterogenous δ_2				Homogeneous δ_2 ($\delta_2 = 0.15$)			
	All links	Congested links	Semi-congested links	Un-congested links	All links	Congested links	Semi-congested links	Un-congested links
Mean	0.09	0.29	0.02	-0.05	0.03	0.04	0.04	0.03
	(0.30)	(0.42)	(0.11)	(0.00)	(0.12)	(0.09)	(0.13)	(0.14)
Weighted mean	0.03	0.14	-0.01	-0.05	-0.01	-0.00	-0.01	-0.02
	(0.20)	(0.29)	(0.06)	(0.00)	(0.06)	(0.06)	(0.07)	(0.06)

Notes. The first two rows of this table report the mean and standard deviation (in parenthesis) of $R(d_{kl})$. The last two rows report the statistics weighted by cost $C(d_{kl})$. Congested, semicongested, and uncongested links are classified in Table 1.

Although labor mobility and externalities do not affect our structural estimation, they do affect the welfare implications.[32] We check the robustness of the above results by conducting the same counterfactual exercise in the model without labor mobility and externalities, the results of which are reported in the third and fourth columns of Table A.11 in Online Appendix A.8. The differences are quantitatively small in the case of immobile workers. The returns change more significantly when α and β are set to zero. For congested and semicongested links, the weighted mean of their returns drop by 2 and 0.3 percentage points to 12.2% and -1.7%, respectively.

Despite the tractability of $R(d_{kl})$, it should be noted that $R(\Delta_{kl}) \approx R(d_{kl})$ only for sufficiently small expansion. To evaluate the accuracy of the approximation, we obtain $R(\Delta_{kl})$ in equation (30) by numerically solving the model for different Δ_{kl} on each link kl. As expected, $R(\Delta_{kl})$ and $R(d_{kl})$ are almost identical for small Δ_{kl} (see Figure A.13 in Online Appendix A.8). However, $R(\Delta_{kl})$ becomes lower as road expansion increases, showing a strong pattern of diminishing returns.

[32] See Donaldson and Hornbeck (2016), who assume perfect mobility, and Tombe and Zhu (2019), who assume limited mobility.

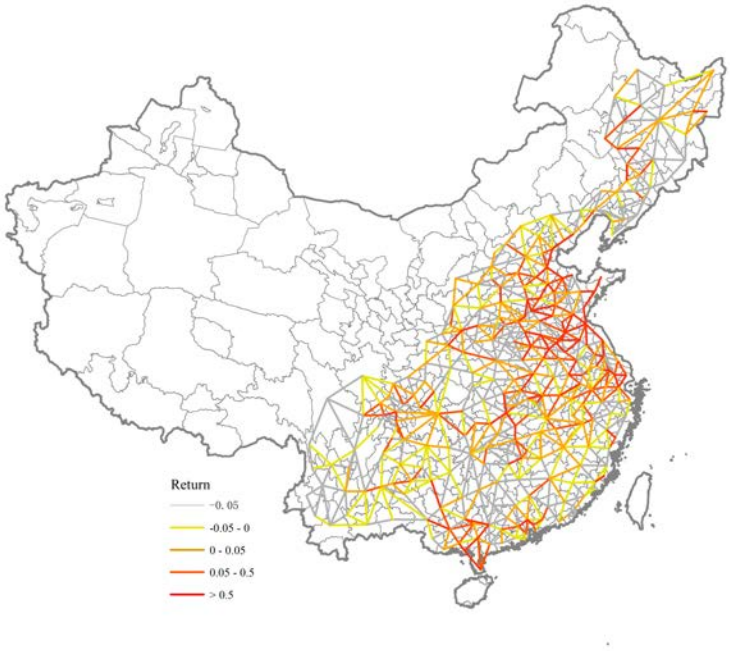

Figure 5. Spatial Distribution of Returns

Notes. Each straight line represents a link. The darkness of the color represents the magnitude of the return $R(d_{kl})$, darker colors implying higher returns.

A more policy-relevant approach is to compute $R(\Delta_{kl})$ associated with adding one lane to each link. For a two-lane link, this means a capacity expansion of 50%. The results are reported in the last four columns of Table A.11 in Online Appendix A.8. We assume that $\Delta_{kl} = 1$ increases the width of the link by 3.5 meters. The mean of these returns for congested links is 20.9%, about one-third lower than the mean of $R(d_{kl})$. However, the two measures are highly correlated (see Figure A.14 in Online Appendix A.8). $R(\Delta_{kl})$ with $\Delta_{kl} = 1$ are also highly unequal in China's road network. The ranks of links based on these two measures are almost identical.

5.4. Productivity Growth and Returns

Our estimates reveal that returns to road investment are, on average, low but highly unequal. An obvious caveat is that the static model tends to underestimate the returns in a fast-growing economy like China. Expanding a road link, albeit having low returns under the current economic condition, could deliver much higher future returns because of strong productivity growth. While we are unable to forecast productivity growth, it is useful to redo all the welfare calculations in an earlier year by assuming perfect foresight of economic fundamentals in 2018. Specifically, let us consider an economy in which the city-to-city links follow China's road network in 2014 and in which the city productivity and amenities are set equal to those calibrated to city GDP and population in that year. In other words, the economy is calibrated

to fit China's economic geography in 2014. The parameters, θ, κ, and ζ, are the same as the structurally estimated values. We then conduct two exercises. In the first exercise, we perform the same welfare calculations for returns to road investment in the 2014 economy. In the second exercise, we evaluate the returns in the economy with the road network in 2014 but with productivities and amenities in 2018.

In the first counterfactual, we calculate $R(d_{kl})$ in the estimated model with the actual number of lanes for each link at the beginning of 2014, the earliest year when the lane information is available, and GDP and population in each city in 2014. In the second exercise, we back out productivity and amenity in each city using city GDP and population in 2018.[33] The same method can also be used to back out city productivity in 2014. Figure A.15 in Online Appendix A.8 shows significant productivity growth in most cities. The annualized productivity growth has an average of 5.1% across the cities. We then impose the 2018 productivities and amenities into a model with the 2014 link capacity and simulate GDP, population, and traffic flows. The simulated data are used to calculate $R(d_{kl})$ in the second counterfactual.

The level of returns differs significantly between counterfactuals with and without productivity growth. The mean of returns for congested links is 35.9% in the counterfactual with productivity growth, much larger than the 25.4% in the first counterfactual without productivity growth. However, replacing productivities and amenities in 2018 with those in 2014 does not change our main finding of highly unequal returns in China's road network. Figure A.16 in Online Appendix A.8 shows that $R(d_{kl})$ in the two counterfactuals are highly correlated (correlation 0.96). The large dispersion of the returns does not narrow unless productivity converges rapidly across cities, a pattern that has not yet appeared in the data (see Figure A.15 in Online Appendix A.8 for the highly persistent productivity disparity).

6. Allocative Efficiency

Dispersion in returns measures misallocation in a convex optimization problem. A well-known example is the variance in the marginal product of capital as summary statistics of capital misallocation in Hsieh and Klenow (2009). However, the mapping between misallocation and the dispersion of returns can be blurred by the presence of investment frictions (see, e.g., Asker, Collard-Wexler and Loecker, 2014; Song and Wu, 2015; Bils, Klenow and Ruane, 2021). More fundamentally, we cannot prove the convexity of the optimal allocation of road capacity in our model.[34] A local optimum with equalized returns may be suboptimal compared to an allocation associated with unequal returns but sufficiently close to the global optimum.

[33] $\bar{A}$ and $\bar{u}$ are calibrated to match city GDP and population through the production function and welfare equalization.

[34] Fajgelbaum and Schaal (2020) prove the convexity of an optimal infrastructure allocation problem in a neoclassical trade model with congestion if the elasticity of trade costs with respect to traffic is large relative to the elasticity of trade costs with respect to infrastructure. A key difference between our framework and theirs is that we take into account congestion heterogeneity. Furthermore, we allow for productivity and amenity externalities, as in Allen and Arkolakis (2022).

To address the allocation efficiency of road investments more rigorously, we consider two further exercises: one that numerically establishes a positive mapping between dispersion in returns and the welfare gain through the optimization of new investments in link capacity and another that shows that the actual road expansion in previous years is orthogonal to the optimal expansion predicted by the model.

6.1. Numerical Analysis

In this subsection, we numerically explore the welfare implications of the dispersion in $R(d_{kl})$. Specifically, we first calculate the returns to optimized investments in a group of links and then compare the returns across groups. The idea is that if the dispersion in $R(d_{kl})$ captures misallocation, the returns to optimized investments should be higher in the groups associated with more dispersed $R(d_{kl})$. To make the analysis more policy-relevant, we assume the new investment to be constrained by a limited budget and the existing capacity to be irreversible. Such irreversibility rules out the possibility of expanding the capacity of a link by reducing existing capacity elsewhere.

To evaluate returns to optimized investments, we need to generalize the returns to expansion of one link in equation (30) to allow for expanding multiple links simultaneously. Denote by $G(g)$ a group of links, indexed by g, and by $\Delta_{G(g)}$ the set of Δ_{kl} $\forall kl \in G(g)$. The return of a joint road expansion is defined as

$$R\left(\Delta_{G(g)}\right) = \frac{B\left(\Delta_{G(g)}\right)}{C\left(\Delta_{G(g)}\right)} - M, \tag{31}$$

where $B\left(\Delta_{G(g)}\right)$ and $C\left(\Delta_{G(g)}\right)$ also generalize the link-specific welfare gain and total cost in equation (23) and (29):

$$B\left(\Delta_{G(g)}\right) = \frac{W\left(\Delta_{G(g)}\right) - W}{W} \times GDP, \tag{32}$$

$$C\left(\Delta_{G(g)}\right) = \sum_{kl \in G(g)} \left(UC_{kl}^{S} + UC_{kl}^{L}\right) \times length_{kl} \times \Delta_{kl}. \tag{33}$$

Here, $W\left(\Delta_{G(g)}\right)$ stands for the counterfactual welfare associated with $\{\Delta_{kl}\}_{kl \in G(g)}$.[35]

We now optimize road investments. Denote by I_g the investment budget for the links in $G(g)$. The optimization solves $\max_{\Delta_{G(g)}} W\left(\Delta_{G(g)}\right)$, subject to $\sum_{kl}\left(UC_{kl}^{S} + UC_{kl}^{L}\right) \times length_{kl} \times \Delta_{kl} = I_g$ and $\Delta_{kl} \geq 0$ $\forall kl \in G(g)$.[36] The restriction $\Delta_{kl} \geq 0$ captures the irreversibility. We consider a budget that amounts to 0.2% of GDP in our model, or one-tenth of China's average annual road infrastructure investment in 2014-18.

[35] Since these links can be substitutes or complements to each other, the returns of a joint expansion are not an aggregation over the returns to single expansion. Interestingly, Figure A.13 in Online Appendix A.8 shows that the weighted average of $R(d_{kl})$ over all congested or semicongested links is very close to $R\left(\Delta_{G(g)}\right)$.

[36] Different from the optimization problem via the intensive margin, Alder and Kondo (2019) and Alder (2023) consider the extensive margin of the road network, i.e., which cities should be directly connected by a national highway network.

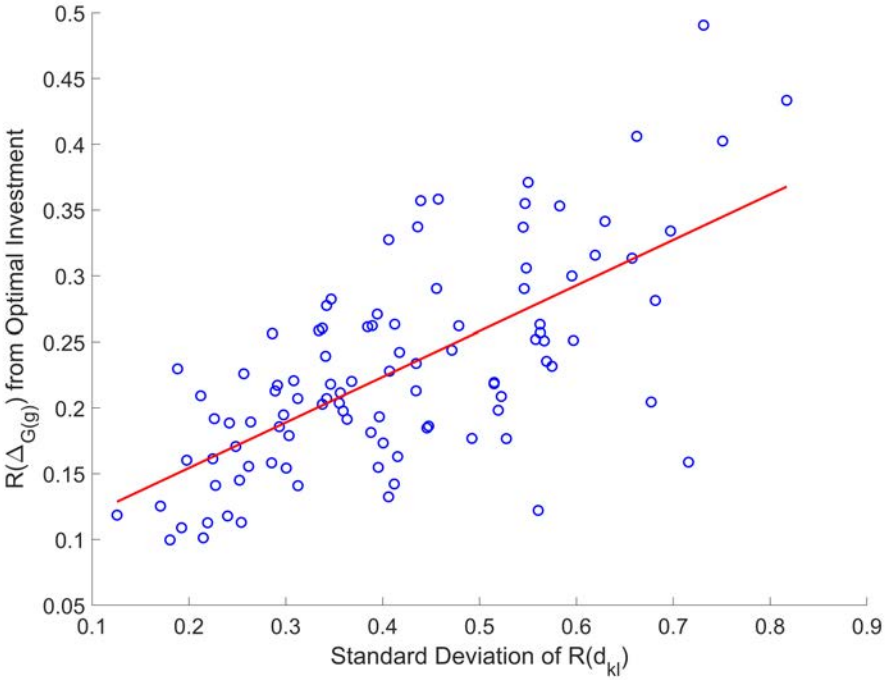

Figure 6. Dispersion of $R\left(d_{kl}\right)$ and Welfare Gains

Notes. Each circle represents a group consisting of 20 congested links. The red line denotes the linear fit. The horizontal axis shows the standard deviation of $R\left(d_{kl}\right)$ across 20 links within each group, and the vertical axis shows the $R\left(\Delta_{G(g)}\right)$ from optimal investments based on equation (31).

The optimization of investment in China's road network with 1,392 links is challenging. To reduce the computational cost, we narrow the problem to the optimization of investment in subsets of links. Specifically, we generate 100 subsets, each with 20 randomly selected congested links. Each subset is referred to as a random group. As the number of random groups approaches infinity, the cross-group mean of the within-group mean and standard deviation of returns should converge to the mean and standard deviation of returns for all congested links. Table A.13 in Online Appendix A.9 shows that these values are indeed very close to each other.

We use "MultiStart" with "fmincon" in MATLAB to search for the globally optimal budget allocation.[37] Figure 6 plots the relationship between the standard deviation of $R\left(d_{kl}\right)$ for $kl \in G\left(g\right)$ before the investments and the return to the optimally allocated investment budget for $G\left(g\right)$. They are positively correlated across random groups (correlation 0.68). The simple linear function has a decent fit (R^2 of 0.45). The slope of the fitted line is 0.35 (standard error of 0.04), implying that increasing the standard deviation of the return to single expansion from 0.30 (25th percentile) to 0.55 (75th percentile) increases the returns

[37] 500 trial points are called to conduct a global search. Increasing the number of trial points gives the same results. In Figure A.17 in Online Appendix A.9, we select one group and report the distribution of the optimized welfare across the 500 trial points, and they are so close to each other that the difference can be negligible, e.g., the smallest welfare is only around 0.00016% smaller than the largest welfare. In other words, it is unlikely that we are stuck in a local optimum that is drastically different from the global optimum for this particular group.

to an optimized joint expansion from 18.8% to 27.4%.

Table 5. $R\left(\Delta_{G(g)}\right)$ and Postallocation Dispersion from a Given Budget

	(1)	(2)	(3)
	Heterogeneous δ_2		Homogeneous δ_2 ($\delta_2 = 0.15$)
	Random Group (Congested Links)	Random Group (All Links)	Random Group (All Links)
	Panel A: Benchmark Model		
Std of $R\left(d_{kl}\right)$	0.42	0.25	0.11
	Panel B: Optimal Investment		
Std of $R\left(d_{kl}\right)$	0.04	0.03	0.01
$R\left(\Delta_{G(g)}\right)$	22.9%	9.3%	6.3%
	Panel C: Uniform Allocation		
Std of $R\left(d_{kl}\right)$	0.24	0.16	0.07
$R\left(\Delta_{G(g)}\right)$	10.3%	0.2%	0.2%

Notes. The budget amounts to 0.2% of GDP in the benchmark model. The standard deviations are the cross-group average of the within-group standard deviation of $R\left(d_{kl}\right)$ across the 20 links. The returns from joint expansions are also the average of $R\left(\Delta_{G(g)}\right)$ across all 100 groups g. In Panel A, the standard deviation is based on the $R\left(d_{kl}\right)$ before the investment; in Panel B and C, the standard deviations are based on $R\left(d_{kl}\right)$ after the investments. In column (3), all standard deviations are calculated based on $R\left(d_{kl}\right)$ with homogeneous δ_2, while the return from joint expansions is calculated by combining heterogeneous δ_2 with investments based on homogeneous δ_2 in order to align with the exercises in Section 7.

We now analyze the dispersion in the postinvestment returns. Investment optimization with a reasonable budget almost completely eliminates such dispersion. The average standard deviation drops from 0.42 to 0.04 (column (1) of Table 5). We also zoom into a random group to visualize such dispersion elimination in Figure A.18 in Online Appendix A.9: the optimal allocation leads to equalized returns across links with sufficiently high preinvestment returns; due to the limited budget, the other links do not receive investments (the upward-sloping part of the triangles). The returns to optimized investment, $R\left(\Delta_{G(g)}\right)$, have a mean of 22.9% across random groups. To gauge the welfare gain from investment optimization, we consider a uniform allocation of the investment budget that makes the expansion proportional for each link: $\Delta_{kl}/lane_{kl} = I_g/\left(\sum_{kl} C\left(\Delta_{kl}\right)\right)$ is the same for all $kl \in G\left(g\right)$. $R\left(\Delta_{kl}\right)$ under the uniform allocation is 10.3% (Panel C of Table 5), only half of $R\left(\Delta_{kl}\right)$ under optimal investments. The dispersion of $R\left(d_{kl}\right)$ under the uniform allocation is also much larger than that under optimal investment, by a factor of 6.

6.2. Misallocation of Actual Road Expansion

In addition to the numerical exercises based on hypothetical investments, we further support the misallocation interpretation by examining actual road expansion in this section. If capacity expansion

is more likely to take place on roads with higher returns, then unequal returns might be a transitory phenomenon, and previous investments may correct, rather than worsen, the misallocation. However, we find no such patterns. We compute the returns in the model using the initial road network in 2014, $\bar{A}_j$ and u_j re-calibrated to China's economic geography in the same year. Figure 7 shows that the expansions of link capacity in 2014-18 (see Figure A.19 in Online Appendix A.9) are not correlated with the initial returns in 2014 (correlation of 0.04). Moreover, $R\left(d_{kl}\right)$ are highly correlated between 2014 and 2018 (correlation of 0.96), suggesting that road investments in 2014-18 have little effect on road allocation efficiency.

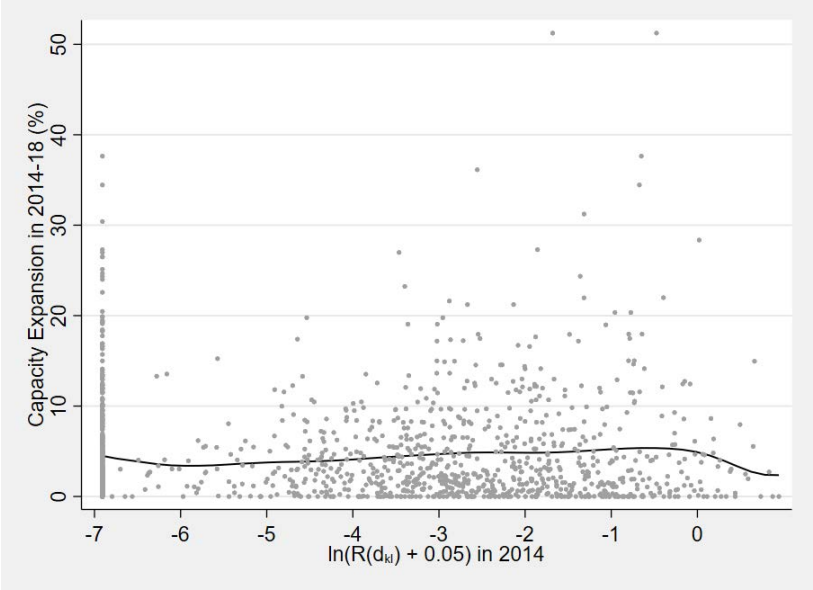

Figure 7. Returns and Link Capacity Expansion

Notes. Each dot represents a link. The black line denotes the polynomial fit. $R\left(d_{kl}\right)$ are calculated based on the initial road network in 2014.

An interesting question is how the actual expansion differs from the optimal expansion of link capacity. For ease of computation, we confine the optimization problem to congested links in a specific province or region. We first focus on the 21 congested links in Guangdong, China's largest province, by GDP. We impute the budget for the actual expansion of congested links in Guangdong between 2014 and 2018, which is approximately 0.1% of Guangdong's annual GDP. We next optimize the budget across the 21 links in the model with $\bar{A}_j$ and u_j re-calibrated to China's economic geography in 2014. The y-axis for each circle and square in the left-top panel of Figure 8 represents the actual and optimal expansion of a link, respectively. The x-axis is $R\left(d_{kl}\right)$. The weak correlation between the actual expansions of links and their returns in Guangdong is no different from that pattern found for the whole country in Figure 7. Optimization directs the limited budget only to those four links with very high returns. The capacity of the link with the highest return, Foshan-Guangzhou, is expanded by 16.3% under the optimal allocation. However, the actual and optimal expansions are only weakly correlated (correlation -0.15), which is additional evidence of road infrastructure investment misallocation in Guangdong. Correcting the misallocation delivers sizable

welfare gains. The actual budget allocation yields $R\left(\Delta_{G(g)}\right)$ of 15.7%. Optimizing this allocation can increase the returns fourfold to 67.6%. Similar patterns can be found when we use other provinces or regions, like Shangdong, Chongqing-Sichuan, and Beijing-Tianjin-Hebei in the remaining panels of Figure 8.[38] We again find that postinvestment returns, $R\left(d_{kl}\right)$, are almost equalized among the links expanded by the optimal investments (see Figure A.20 in Online Appendix A.9).

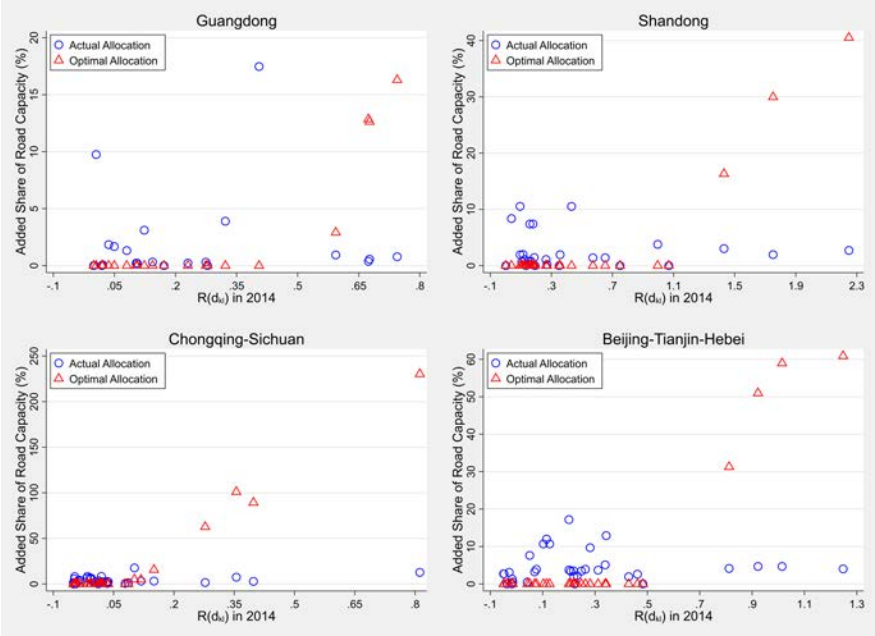

Figure 8. Optimal Capacity Reallocation

Notes. Each circle or triangle represents a link. The vertical axis shows the percentage increment based on the capacity of the initial road network in 2014. The horizontal axis shows the $R\left(d_{kl}\right)$ in 2014.

7. Returns and Welfare under Homogeneous Congestion Elasticity

A key contribution of this paper is that it allows for heterogeneity in the elasticity of congestion in the road network. In this section, we conduct a comparison to illustrate the extent to which assuming homogeneous congestion elasticity affects the estimated returns and corresponding welfare implications.

Denote by δ_2 the homogeneous elasticity of congestion. We set δ_2 to the point estimate of 0.15 from the first column of Table 1, where δ_{2kl} is restricted to be identical across all links. We structurally re-estimate the four parameters θ, κ, and ζ. The estimates are reported in Table A.14 in Online Appendix A.10. Interestingly, the resulting model parameter estimates are similar to those in the benchmark model,

[38] $R\left(\Delta_{G(g)}\right)$ would increase from 41.6% to 177.5% for Shandong, from 1.2% to 21.0% for Chongqing-Sichuan, and from 24.6% to 86.4% for Beijing-Tianjin-Hebei, when we move from actual to optimal road expansions.

implying that the structural estimation is insensitive to heterogeneity in the elasticity of congestion. We refer to the re-estimated model under $\delta_2 = 0.15$ as the model with homogeneous elasticity of congestion.

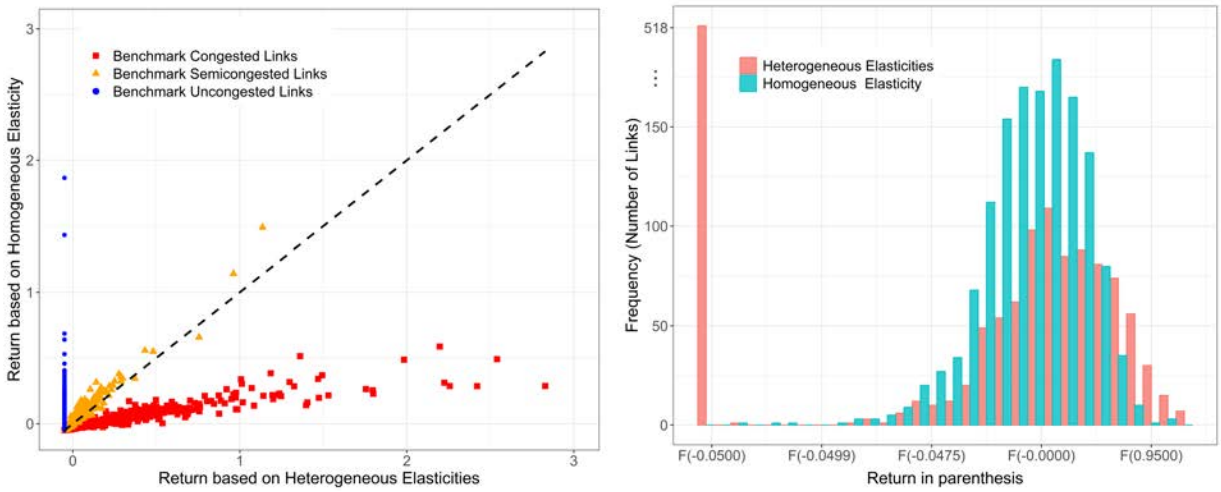

Figure 9. Comparison of Marginal Returns with and without Congestion Heterogeneity

Notes. The left panel shows a scatterplot of $R(d_{kl})$ and $\bar{R}(d_{kl})$ in each link with heterogeneous and homogeneous congestion elasticities. The dashed line represents the 45-degree line. The right panel shows the distributions of $R(d_{kl})$ and $\bar{R}(d_{kl})$ with heterogeneous and homogeneous congestion elasticities. $F(x)$ is a transformation function, defined as $F(x) = (x + 0.05 + 1e-5)$, where x represents $R(d_{kl})$ (or $\bar{R}(d_{kl})$). The horizontal axis in the left panel and the green bar in the right panel correspond to $R(d_{kl})$, and the vertical axis in the left panel and the red bar in the right panel correspond to $\bar{R}(d_{kl})$.

Denote by $\bar{R}(d_{kl})$ and $\bar{R}(\Delta_G(g))$ the return to expanding only link kl and expanding a group of links $kl \in G(g)$, respectively, in the model with homogeneous elasticity.[39] The weighted mean of $\bar{R}(d_{kl})$ becomes negative (see column (5) of Table 4). The main reason is that the returns for congested links drop from 14.2% to zero (see column (6) of Table 4) as well as the left panel of Figure 9. In addition, for uncongested links $R(d_{kl})$ is –5% but the weighted mean of $\bar{R}(d_{kl})$ becomes -1.5%. $R(d_{kl})$ and $\bar{R}(d_{kl})$ are much more similar among semicongested links than among other groups of links. The right panel of Figure 9 compares the distribution of $\bar{R}(d_{kl})$ with that of $R(d_{kl})$, and we observe that the model with homogeneous elasticities underestimates the dispersion.

The above results suggest that model misspecification regarding the elasticity of congestion implies very different policy implications for road investments. We are particularly interested in how much the returns from optimal expansions might be affected if the optimization conducted by the social planner is subject to the homogeneous elasticity. To quantify the welfare effect, we conduct the optimal investment exercises based on random groups, similar to the approach in Section 6.1. Different from the random groups that are drawn only from congested links in Section 6.1, we now work with 100 groups with 20 links randomly drawn from all links, regardless of their congestion elasticity. This approach allows for a

[39] Given these two notations, it should be pointed out that the returns in columns (5)-(8) of Table 4 correspond to $\bar{R}(d_{kl})$ and the returns in column (3) of Table 5 correspond to $\bar{R}(d_{kl})$ and $\bar{R}(\Delta_G(g))$.

less biased welfare comparison between homogeneous and heterogeneous elasticities. As shown in Table A.13 in Online Appendix A.10, the cross-group means of the within-group means and standard deviations of $\bar{R}(d_{kl})$ are again close to those of the full sample of all links. We then optimally allocate the investment budget across links for each group to maximize welfare in the model with homogeneous and heterogeneous elasticities, subject to the same budget and irreversibility constraints. Given these optimal investments, we evaluate the corresponding welfare gain in the (factual) case with heterogeneous elasticities. As shown in columns (2) and (3) of Table 5, $\bar{R}(\Delta_G(g))$ is 6.3% using the model with homogeneous congestion elasticity, which is only around two-thirds of the mean in the benchmark model with heterogeneous congestion elasticities. We conclude that model misspecification can lead to significantly different welfare effects.

8. Conclusion

We quantify the returns to road infrastructure investments in China using a quantitative general equilibrium trade model with optimal path choice and road congestion. Combined with detailed data on traffic flows from GPS devices on half a million trucks, the model captures the key features of the observed traffic patterns, such as heterogeneous congestion elasticities across links. We then structurally estimate the model and use it to predict the returns to expanding road capacity in each link. The predicted returns are low, on average, and highly dispersed across the network. We show numerically that optimized investments with a reasonable budget can reduce this dispersion substantially and generate sizable welfare gains. However, we do not find evidence that the actual link capacity expansions in China during 2014-18 reduce the dispersion and improve the allocation efficiency of link capacity.

Our analysis is conducted in a static environment. It is possible that the returns to infrastructure investments are underestimated by ignoring their contributions to long-run growth (see, e.g., Xiong, 2018; Ma and Yang, 2022), and the bias could be larger for those returns in less developed regions. We show through robustness checks that the average of the returns is affected by productivity growth. However, since productivity differences are highly persistent across Chinese cities, feeding the model with actual productivity growth does not change the dispersion in the returns. It is, therefore, hard to argue that the misallocation that we find based on the static model is due to the exclusion of dynamics. That being said, it will be interesting to explore the long-term effects of misallocation we identify in 2018 due to the persistence of the distribution of economic activity arising from migration frictions, capital adjustment costs, and dynamic agglomeration spillovers (e.g., Caliendo, Dvorkin and Parro, 2019; Allen and Donaldson, 2022; Kleinman, Liu and Redding, 2023). We leave it to future work.

While our findings suggest a severe misallocation of road investments in China, we are aware of two caveats. First, the budget allocation in our numerical analysis of misallocation is optimized only among a subsample of links. A second issue is that the congestion status of a link, as we measure it, may itself be endogenous. We think of at least part of congestion elasticity δ_{2kl} as a link characteristic that does not itself

depend on traffic volume but determines how much speed falls as traffic increases. While one might expect that geography or other observable factors determine how quickly a link becomes congested when traffic increases, we experiment with various geographic and network characteristics and find that congestion heterogeneity remains largely unexplained. Hence, the heterogeneity in the congestion elasticity appears to be a feature of the current Chinese road network. However, it is also reasonable to expect this elasticity to change if there are very large changes in traffic flows or a considerable reshaping of the network. We plan to study this issue in future work using GPS data over a longer period.

An obvious reason for the highly unequal returns is that the objective of road investments is broader than that of maximizing welfare, as defined in the model. For instance, the expansion of a low-return link can serve as a policy instrument for reducing income inequality across cities. Financial constraints might also be an important factor. There is a positive correlation between road investments and the ratio of land sales to GDP across cities, suggesting that local governments rely on land sales to finance infrastructure investments (Jiang, Miao and Zhang, 2022). Possible explanations can be explored in future research.

References

Akbar, P. A., V. Couture, G. Duranton, and A. Storeygard (2023): "Mobility and Congestion in Urban India," *American Economic Review*, 113, 1083–1111.

Alder, S. (2023): "Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development," Available at SSRN 2856050.

Alder, S., and I. Kondo (2019): "Political Distortions and Infrastructure Networks in China: A Quantitative Spatial Equilibrium Analysis," Working Paper.

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

——— (2019): "The Welfare Effects of Transportation Infrastructure Improvements," Working Paper.

——— (2022): "The Welfare Effects of Transportation Infrastructure Improvements," *The Review of Economic Studies*, 89, 2911–2957.

Allen, T., D. Atkin, S. Cantillo, and C. Hernandez (2020): "Trucks," Working Paper.

Allen, T., and D. Donaldson (2022): "Persistence and Path Dependence in the Spatial Economy," National Bureau of Economic Research Working Paper No. w28059.

Asker, J., A. Collard-Wexler, and J. D. Loecker (2014): "Dynamic Inputs and Resource (Mis)Allocation," *Journal of Political Economy*, 122, 1013–1063.

Atkeson, A., and A. T. Burstein (2010): "Innovation, Firm Dynamics, and International Trade," *Journal of Political Economy*, 118, 433–484.

Autor, D. H., D. Dorn, and G. H. Hanson (2013): "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 103, 2121–2168.

Bai, C.-E., and Y. Qian (2010): "Infrastructure Development in China: The Cases of Electricity, Highways, and Railways," *Journal of Comparative Economics*, 38, 34–51.

Banerjee, A., E. Duflo, and N. Qian (2020): "On the Road: Access to Transportation Infrastructure and Economic Growth in China," *Journal of Development Economics*, 145, 102442.

Baum-Snow, N., J. V. Henderson, M. A. Turner, and Q. Zhang (2020): "Does Investment in National Highways Help or Hurt Hinterland City Growth?" *Journal of Urban Economics*, 115, 103124.

Baum-Snow, N., L. B. J. V. Henderson, M. A. Turner, and Q. Zhang (2017): "Roads, Railroads, and Decentralization of Chinese Cities," *Review of Economics and Statistics*, 99, 435–448.

Bils, M., P. J. Klenow, and C. Ruane (2021): "Misallocation or Mismeasurement?" *Journal of Monetary Economics*, 124, S39–S56.

Caliendo, L., M. Dvorkin, and F. Parro (2019): "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock," *Econometrica*, 87, 741–835.

Chen, J., W. Chen, E. Liu, J. Luo, and Z. M. Song (2020): "The Economic Impact of COVID-19 in China: Evidence from City-to-City Truck Flows," Working Paper.

Chen, T., and J. K.-s. Kung (2019): "Busting the "Princelings": The Campaign against Corruption in China's Primary Land Market," *The Quarterly Journal of Economics*, 134, 185–226.

Chen, W., and H. A. Klaiber (2020): "Does Road Expansion Induce Traffic? An Evaluation of Vehicle-Kilometers Traveled in China," *Journal of Environmental Economics and Management*, 104, 102387.

Collier, P., M. Kirchberger, and M. Söderbom (2016): "The Cost of Road Infrastructure in Low-and Middle-Income Countries," *The World Bank Economic Review*, 30, 522–548.

Donaldson, D. (2018): "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, 108, 899–934.

Donaldson, D., and R. Hornbeck (2016): "Railroads and American Economic Growth: A "Market Access" Approach," *The Quarterly Journal of Economics*, 131, 799–858.

Dube, O., and S. Naidu (2015): "Bases, Bullets, and Ballots: The Effect of US Military Aid on Political Conflict in Colombia," *The Journal of Politics*, 77, 249–267.

Duranton, G., and M. A. Turner (2011): "The Fundamental Law of Road Congestion: Evidence from US Cities," *American Economic Review*, 101, 2616–2652.

——— (2012): "Urban Growth and Transportation," *Review of Economic Studies*, 79, 1407–1440.

Eaton, J., and S. Kortum (2002): "Technology, Geography, and Trade," *Econometrica*, 70, 1741–1779.

Egger, P. H., G. Loumeau, and N. Loumeau (2023): "China's Dazzling Transport-Infrastructure Growth: Measurement and Effects," *Journal of International Economics*, 142, 103734.

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

Fajgelbaum, P. D., and E. Schaal (2020): "Optimal Transport Networks in Spatial Equilibrium," *Econometrica*, 88, 1411–1452.

Fan, J., Y. Lu, and W. Luo (2019): "Valuing Domestic Transport Infrastructure: A View from the Route Choice of Exporters," *Review of Economics and Statistics*, 1–46.

Fay, M., S. Han, H. I. Lee, M. Mastruzzi, and M. Cho (2019): "Hitting the Trillion Mark: A Look at How Much Countries Are Spending on Infrastructure," World Bank Policy Research Working Paper 8730.

Garcia-López, M.-À., I. Pasidis, and E. Viladecans-Marsal (2022): "Congestion in Highways When Tolls and Railroads Matter: Evidence from European Cities," *Journal of Economic Geography*, 22, 931–960.

Hansen, B. (2022): *Econometrics*, Princeton: Princeton University Press.

He, G., Y. Xie, and B. Zhang (2020): "Expressways, GDP, and the Environment: The Case of China," *Journal of Development Economics*, 145, 102485.

Hernández, C. E. (2022): "Stops and Delays in Road Freight Transport: Variation across Trips, Trucks, Drivers, and Carriers," Working Paper.

Herzog, I. (2022): "The Marginal External Cost of Traffic in Greater London," Working Paper.

Hsieh, C.-T., and P. J. Klenow (2009): "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 124, 1403–1448.

Hsu, W.-T., and H. Zhang (2014): "The Fundamental Law of Highway Congestion Revisited: Evidence from National Expressways in Japan," *Journal of Urban Economics*, 81, 65–76.

Jiang, S., J. Miao, and Y. Zhang (2022): "China's Housing Bubble, Infrastructure Investment, and Economic Growth," *International Economic Review*, 63, 1189–1237.

Jones, C. I. (2013): "Misallocation, Economic Growth, and Input-Output Economics," in *Advances in Economics and Econometrics, Tenth World Congress, Volume II* ed. by Acemoglu, D., M. Arellano, and E. Dekel: Cambridge University Press, , 419–455.

Kleinman, B., E. Liu, and S. J. Redding (2023): "Dynamic Spatial General Equilibrium," *Econometrica*, 91, 385–424.

Krishna, P., and E. van Leemput (2018): "Quantifying Sources of Internal Trade Barriers: Evidence from Indian Truck GPS Data," Working Paper.

Lai, E. L.-C., H. Fan, and H. S. Qi (2020): "Global Gains from Reduction in Trade Costs," *Economic Theory*, 70, 313–345.

Li, Z., M. Wu, and B. R. Chen (2017): "Is Road Infrastructure Investment in China Excessive? Evidence from Productivity of Firms," *Regional Science and Urban Economics*, 65, 116–126.

Lin, Y., Y. Qin, Y. Yang, and H. Zhu (2020): "Can Price Regulation Increase Land-Use Intensity? Evidence from China's Industrial Land Market," *Regional Science and Urban Economics*, 81, 103501.

Ma, L., and T. Yang (2022): "The Distributional Impacts of Transportation Networks in China," Working Paper.

Notley, S., N. Bourne, and N. Taylor (2009): "Speed, Flow and Density of Motorway Traffic," TRL Insight Report S003.

Office of the State Auditor (2002): "Capitalization and Depreciation of Infrastructure," Presentation to: Mississippi Association of Governmental Purchasing and Property Agents.

Ossokina, I. V., J. van Ommeren, and H. van Mourik (2021): "Do Highway Widenings Reduce Congestion?", Working Paper.

Redding, S. J. (2016): "Goods Trade, Factor Mobility and Welfare," *Journal of International Economics*, 101, 148–167.

Redding, S. J., and E. Rossi-Hansberg (2017): "Quantitative Spatial Economics," *Annual Review of Economics*, 9, 21–58.

Redding, S. J., and M. A. Turner (2015): "Transportation Costs and The Spatial Organization of Economic Activity," *Handbook of Regional and Urban Economics*, 5, 1339–1398.

Restuccia, D., and R. Rogerson (2008): "Policy Distortions and Aggregate Productivity with Heterogeneous Establishments," *Review of Economic Dynamics*, 11, 707–720.

——— (2017): "The Causes and Costs of Misallocation," *Journal of Economic Perspectives*, 31, 151–74.

Riley, S. J., S. D. DeGloria, and R. Elliot (1999): "Index that Quantifies Topographic Heterogeneity," *Intermountain Journal of Sciences*, 5, 23–27.

Roberts, M., U. Deichmann, B. Fingleton, and T. Shi (2012): "Evaluating China's Road to Prosperity: A New Economic Geography Approach," *Regional Science and Urban Economics*, 42, 580–594.

Russo, A., M. W. Adler, F. Liberini, and J. N. van Ommeren (2021): "Welfare Losses of Road Congestion: Evidence from Rome," *Regional Science and Urban Economics*, 89, 103692.

Song, Z., and G. L. Wu (2015): "Identifying Capital Misallocation," Working Paper.

Su, L., Z. Shi, and P. C. Phillips (2016): "Identifying Latent Structures in Panel Data," *Econometrica*, 84, 2215–2264.

Tombe, T., and X. Zhu (2019): "Trade, Migration, and Productivity: A Quantitative Analysis of China," *American Economic Review*, 109, 1843–1872.

Vickrey, W. (1967): "Optimization of Traffic and Facilities," *Journal of Transport Economics and Policy*, 1, 123–136.

Wu, G. L., Q. Feng, and Z. Wang (2021): "A Structural Estimation of the Return to Infrastructure Investment in China," *Journal of Development Economics*, 152, 102672.

Xiong, W. (2018): "The Mandarin Model of Growth," Working Paper.

Xu, Y., and X. Yang (2021): "Access to Ports and the Welfare Gains from Domestic Transportation Infrastructure," *Journal of Urban Economics*, 126, 103392.

Xu, Z., J. Huang, and F. Jiang (2017): "Subsidy Competition, Industrial Land Price Distortions and Overinvestment: Empirical Evidence from China's Manufacturing Enterprises," *Applied Economics*, 49, 4851–4870.

Online Appendix for

On (Un)Congested Roads: A Quantitative Analysis of Infrastructure Investment Efficiency using Truck GPS Data

Simon Alder[1] Zheng (Michael) Song[2] Zhitao Zhu[3]

April 28, 2023

A.1. Data

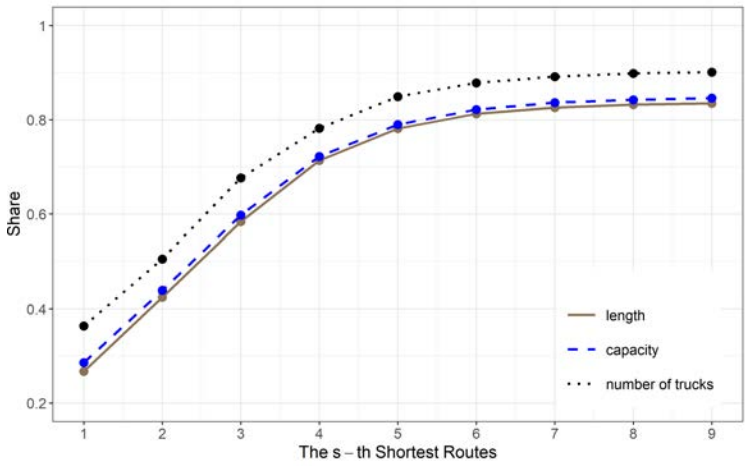

Figure A.1. Road Coverage of the s-Shortest Routes

Notes. The solid brown line, dashed blue line, and dotted black line represent the coverage of road length, road capacity, and number of trucks, respectively, for different numbers of shortest paths.

[1]Swiss National Bank and The University of North Carolina at Chapel Hill, simon.alder@snb.ch
[2]The Chinese University of Hong Kong, zheng.michael.song@gmail.com
[3]The Chinese University of Hong Kong, zhuzhitao@link.cuhk.edu.hk

Figure A.2. Road Network

Notes. The orange lines show the road network based on five shortest paths, and the red points denote cities.

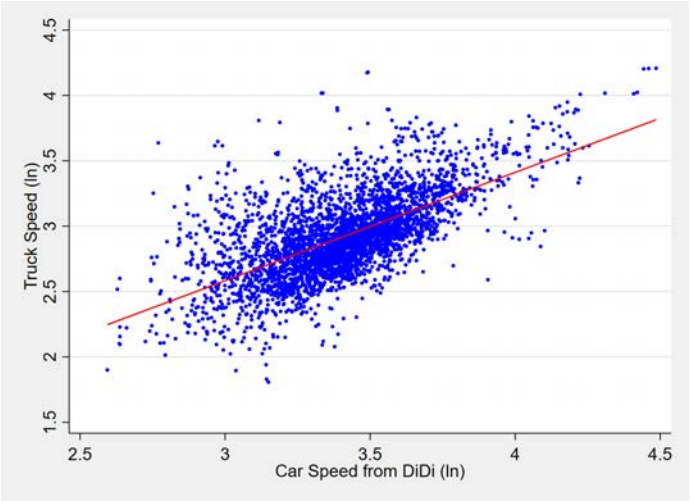

Figure A.3. Car and Truck Speed

Notes. The figure shows the correlation between the truck speed in the GPS data and the car speed data from DiDi, both on logarithmic scale. The correlation coefficient between truck speed and car speed (both in log) is 0.64.

Table A.1. Summary Statistics for Trucks on the Road Network

	Mean	Standard Deviation	The 25th percentile	The 50th percentile	The 75th percentile
Truck Density	4.42	3.83	1.82	3.45	5.91
Truck Speed	56.99	11.86	49.76	58.36	65.64

Notes. The summary statistics are calculated based on the data in which each link-minute is one observation. The units for truck density and speed are number of trucks per lane-kilometer (multiplied by 100) and kilometer per hour, respectively.

A.2. Implication of Endogenous Vehicle Composition for the Correlation between Speed and Density

In this section, we provide a toy model to show that the endogenous variation in the composition of vehicles may lead to negative δ_2, i.e., a positive correlation between speed and density. Notley, Bourne and Taylor (2009) argue that "free-flow" speed is a linear function of the fraction of heavy goods vehicles (HGV), which drive at a lower speed and are also more likely to select low-density roads. Importantly, the effect of the composition of vehicles is different for congested and uncongested links. On uncongested links, vehicles do not impede each others' ability to drive at their individually desired speed, and the observed average speed depends on the composition of vehicles, e.g., on the fraction of HGV (which want to drive more slowly). As low-speed vehicles actively choose to drive on low-density roads, the estimation of (2) from time variation can yield a negative and significant value of δ_2 for uncongested roads. In contrast, this selection matters relatively less, or is even irrelevant, for the estimation of (2) on congested roads.

The mechanism can be explicitly demonstrated in a toy model. Consider, without loss of generality, an economy with a congested link and an uncongested link. Denote by $traf_{Ct}$ and $traf_{Ut}$ the number of trucks on the congested and uncongested links, respectively. Assume $traf_{Ct} = C - V_t$ and $traf_{Ut} = U_t + V_t$, where C is the constant truck flow on the congested link, U_t is the time-varying truck flow on the uncongested link, with $U_t < C$, and V_t denotes the switchers (HGV) from the congested to uncongested link. We further assume that the speed on the congested link follows (2); the speed of U_t is s^U and the speed of the switchers (HGV) is $s^V < s^U$ on the uncongested link, which implies that trucks switching to the uncongested link with lower density, V_t, are of low-speed. We impose a linear relationship to govern the size of switchers in response to U_t: $V_t = \eta\,(C - U_t)$, where $\eta < 1$. $\eta < 1$ ensures that the variations in density on uncongested links will not be dominated by the switchers. The only source of exogenous variation is U_t.

The presence of switchers causes speed to correlate to density on the uncongested link by changing the composition of low- and high-speed trucks over time. The average speed on the uncongested link is $\frac{U_t}{\eta C+(1-\eta)U_t}s^U + \frac{\eta(C-U_t)}{\eta C+(1-\eta)U_t}s^V$. $\eta > 0$ is sufficient for the average speed to correlate positively to the density, $\eta C + (1-\eta)\,U_t$. The correlation leads to a negative δ_2 for the uncongested link. Nevertheless, the OLS estimate of δ_2 will not be biased for the congested link. In other words, our criterion for congestion will not mistake congested links for uncongested, nor the other way around.

If we further assume V_t to respond to shocks to speed on the congested link, ϵ_t in the congestion equation (2), the OLS estimate of δ_2 for the congested links will also be biased downwards. For instance, traffic accidents may slow down speed and cause detours at the same time. When V_t is correlated with both U_t and s^U, our criterion for congestion is likely to mistake some congested links for uncongested. We do not have a good time-varying IV to address the "double" endogeneity issue. However, we will present indirect evidence from the tests of the fundamental law of road congestion, which is in line with our congestion measure and suggests a relatively small downward bias of δ_2 for congested links.

A.3. C-Lasso and Penalized GMM

Estimation Procedure

PGMM is implemented as follows. The tuning parameters in Su, Shi and Phillips (2016) are specified as $\bar{\lambda} = \iota \times var(y) \times T^{-1/3}$, where $var(y)$ indicates the variance of the dependent variable, and T is the number of periods. Because our panel data is unbalanced, we use the average number of periods across all links for T. The remaining part of the tuning parameter, ι, and the number of groups, H, is obtained by minimizing the IC value.

The solid and dotted lines in Figure A.4 plot the IC values against ι under $H = 2$ and $H = 3$, respectively. For $H = 2$, the IC value is minimized at $\iota = 0.2924$, which classifies 578 links into the congested groups and 814 links into the uncongested group. For $H = 3$, the optimal value is $\iota = 0.7533$. 479, 395, and 518 links are classified into the congested, semicongested, and uncongested groups, respectively. The figure also indicates that $H = 3$ is slightly better than $H = 2$. But both are far better than $H = 1$, for which the IC value is -1.2e-3. [4]

Estimation Results

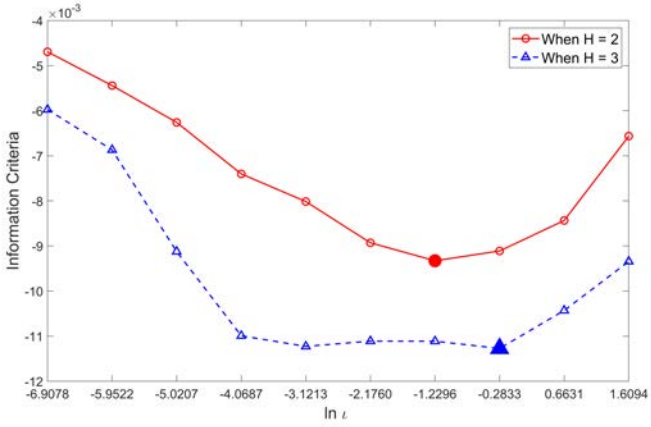

Figure A.4. Information Criteria under Various Groups and Tuning Parameters

Notes. The horizontal axis reports $\ln \iota$, where ι represents the constant part of tuning parameter. We report cases where the constant part of the tuning parameter $\iota \in [0.0010, 5]$, and select 10 values in a geometrically increasing sequence by following Su, Shi and Phillips (2016). The filled point and triangle indicate the optimal ι with the smallest information criteria. Cases beyond this range have also been checked, and they give larger values in terms of the information criteria.

[4] A naive way of showing the heterogeneity in the congestion elasticity is to conduct the link-by-link IV estimation. Such an exercise shows that the share of the congested links, selected by a positive estimate of the elasticity, is 56%. We can consider this exercise as an extreme case of C-Lasso, in which the tuning parameter is set such that C-Lasso classifies each link into a group.

Table A.2. Congestion Estimation based on the Shortest Path Network

	(1)	(2)	(3)	(4)	(5)	(6)
	One Group	Two Groups			Three Groups	
	All Links	Congested Links	Uncongested Links	Congested Links	Semi-congested Links	Uncongested Links
Share of Links	100%	54.60%	45.40%	31.54%	33.62%	34.84%
				2SLS		
Truck density (ln)	0.079**	0.369***	-0.388***	0.490***	0.061	-0.386***
	(0.040)	(0.042)	(0.040)	(0.052)	(0.049)	(0.038)
				First Stage		
IV (ln)	0.591***	0.644***	0.522***	0.681***	0.462***	0.613***
	(0.034)	(0.037)	(0.040)	(0.045)	(0.038)	(0.044)
Kleibergen-Paap rk Wald F Statistics	300.186	308.274	172.095	224.920	146.254	191.799
Link fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Observation	6,340,396	3,548,245	2,792,151	2,093,134	2,098,049	2,149,213

Notes. Standard errors reported in parentheses are clustered at the level of link and day. * 10%, ** 5%, *** 1% significance level.

A.4. Cross-Sectional Evidence on Congestion Measurement

Table A.3. Cross-link Correlations, Ordered Probit Model

	(1)	(2)	(3)	(4)	(5)	(6)
			Full Sample			15-min Interval
Speed / Speed limit	-1.634***	-1.592***			-1.575***	-1.279***
	(0.360)	(0.418)			(0.422)	(0.407)
Density (ln)			0.282***	0.396***	0.394***	0.329***
			(0.045)	(0.063)	(0.063)	(0.061)
Control variables		Yes		Yes	Yes	Yes
Observation	1,392	1,392	1,392	1,392	1,392	1,392
Pseudo R^2	0.011	0.032	0.026	0.046	0.054	0.039

Notes. The dependent variable is 0, 1, and 2 for uncongested, semicongested, and congested links, respectively. Robust standard errors are reported in parentheses. "Speed / Speed limit" denotes the ratio of average speed to the speed limit. The controls include longitude, latitude, altitude, bilateral distance, ruggedness, and the centrality of each link. We include both linear and quadratic terms of the control variables. The centrality is measured as how many times the link would be passed through by the shortest route when we consider all city pairs. * 10%, ** 5%, *** 1% significance level.

A.5. Congestion in England

England has detailed data on traffic volumes and speed at thousands of measurement sites installed along its highway network, which we use to measure congestion in England, serving as a comparison with our results of congestion estimation from the Chinese traffic data based on GPS devices on trucks.

City and Road Network

We define the national highway network in England as all roads that are maintained by *Highways England*, which is a national government company that is responsible for the motorways and major (trunk) roads in England. Figure A.5 shows the extent of the network.

We define the administrative units in England following the Nomenclature of Territorial Units for Statistics (NUTS 2), which divides England into 33 regions. We restrict the sample to the 28 regions that are connected by the highways maintained by *Highways England* and have measurement sites with non-missing observations.[5] The nodes of the infrastructure network correspond to the points on the network that are closest to the centroids of each region.

Traffic Data from Measurement Sites

We use traffic data from *Highway England*'s Motorway Incident Detection and Automatic Signalling (MIDAS), which monitors traffic using sensors installed in the road. The right panel of Figure A.5 shows the distribution of the measurement sites. The data contain information on traffic flow and speed for more than 7,000 measurement sites during each 15-minute interval.[6]

A highway link between two nodes can have many measurement sites, and we average speed and density across all measurement sites on the same link in order to have a comparable data set as for China. Specifically, we use the average speed over all measurement sites on link kl within each 15-minute interval to measure $speed_{kl}$. We calculate the density at each measurement site by dividing traffic flow by speed, and adjust it using lane numbers.[7] Then, we use the average density across all measurement sites on link kl as $density_{kl}$.[8] Finally, to be consistent with our data management for China's data, we also calculate $flow_{kl}$ by using $density_{kl}$.

[5] The gross domestic product of each region is available from the Office for National Statistics.

[6] For a sub-sample of the measurement sites, we also collected data on travel times from Google Maps. The travel speed implied by the Google data is consistent with the speed in the MIDAS data. The advantage of the MIDAS data is that we also have the traffic flow, which allows us to analyze congestion.

[7] The traffic flow is measured as the number of vehicles per 15-minute interval. Hence, the unit of traffic density is $\frac{\text{vehicles/15-min}}{\text{km/15-min}} = \frac{\text{vehicles}}{\text{km}}$. As the unit of density is $\frac{\text{vehicles}}{\text{km} \times \text{lanes}}$, we need to divide the previous term by the number of lanes to get density.

[8] The road network is defined at the level of links and we abstract from heterogeneity across measurement sites on the same link by using the average for each link.

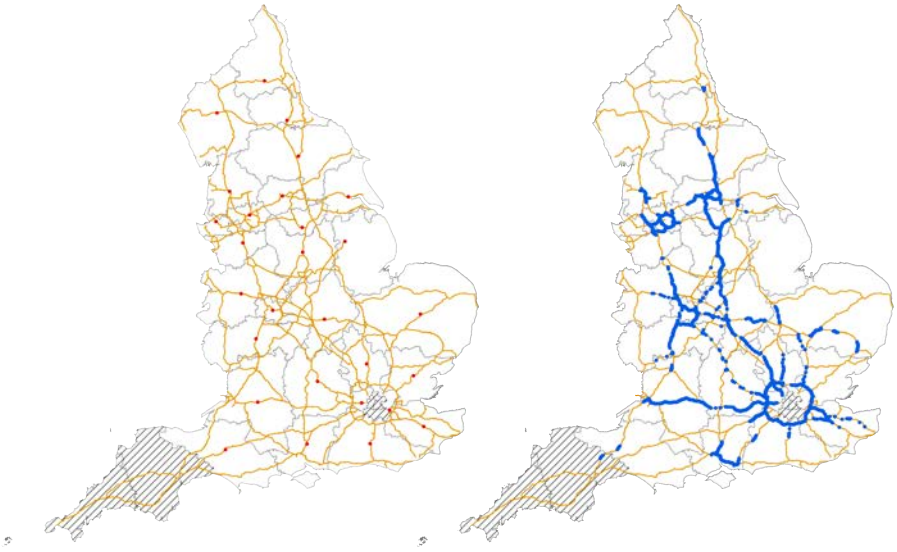

Figure A.5. England's Roads, Regions, and Measurement Sites

Notes. The non-shaded areas are the 28 regions in our sample, and the orange lines represent the roads managed by *Highway England.* The red points on the left map represent the points on the roads that are nearest to the centroids of each region, and the blue points represent the measurement sites.

Congestion Measurement and Comparison with China

We now apply the same method to estimate congestion in England. The goal is to compare China with a mature economy and to make sense of its small share of congested links. Before preceding, let us first compare highways and their related economic conditions between the two economies. The reason for focusing on highways is that the MIDAS data we will use only covers highways in England.

Table A.5 reports the summary statistics.[9] The first and third columns are for China and England. The middle column reports the statistics for three economically advanced regions in China that are more comparable to England. The first seven rows compare China and England on the capacity of highways (the first row) and their density (from the second to seventh rows). Not surprisingly, the statistics suggest a huge difference between China and England. In particular, despite the rapid expansion of China's highway system over the past two decades, the density of highways in China is still about half of that in England. Interestingly, the density of highway in the three economically advanced regions in China is very similar to that in England. Rows (8)-(10) compare the economic conditions between the two economies. Again, the three economically advanced regions are more comparable to England, though the income gap is still sizable.

[9]The gridded GDP data are obtained from the *Global Dataset of Gridded Population and GDP (1980-2010 Estimations and 2020-2100 Scenarios)* for 2010, where the grid size is 0.5×0.5 degree and the unit is Billion $US\$2005/yr$ in PPP. For population, we do not refer to the same source because it is constructed for every ten years and thus there is no information for 2018. The population data for 2018 are obtained from *LandScan Datasets*, where the resolution is 1 km.

The MIDAS data we use for England report the average traffic speed and flow every 15 minutes at 7,426 measurement sites, which are all on highways. In contrast, our Chinese GPS data covers the company's trucks on all types of Chinese roads. We add several restrictions to the two samples to facilitate a direct comparison between China and England. We first restrict our attention to *trucks* or truck-like long vehicles on highways. The MIDAS data do not explicitly distinguish trucks from other vehicles. However, it reports traffic flows conditional on the length of the vehicle. For traffic flows in the MIDAS data, we will only use long vehicles with a length above 6.6 meters, which are arguably the closest counterpart of trucks in our GPS records.[10]

Another important difference is that our GPS records cover four hours a day and two days each month, while MIDAS has data for every 15-minute interval.[11] We extract a sample from the MIDAS data with the same time structure as the GPS data for China, and extract a sample from our GPS records that consist of four minutes (15, 30, 45, 60) each hour for each link in China.

We then use the two restricted samples to estimate the same congestion equation. We find nearly all links in England (119 out of 122) are congested. The share of congested and semicongested ($\delta_2 > 0$) links in China's highway network is 73.3%. While the share is higher than 62.9% in our benchmark road network, it is still much lower than the congestion share in England (see Table A.4 for the congestion estimation results for England and Figure A.5 for the comparison of the spatial distribution of δ_2).

The remarkable differences may be attributed to the fact that our road network for England only consists of highways. To make the samples more comparable, we construct China's city-to-city highway network and re-estimate (2) using C-Lasso and observations with a 15-minute interval.[12] The share of congested and semicongested links changes modestly to 67.3% and 73.2% under $H = 2$ and 3, respectively, in China's highway system.

To conclude, congestion heterogeneity is a salient feature of China's road network. The heterogeneity is much less pronounced in England. While the English traffic data is detailed and allows for a similar empirical analysis as in China, an important difference between England and China (and many European countries) is that England spent less on road infrastructure over the past decades, and its traffic density is relatively high.[13] These differences are, on the one hand, a challenge for the comparison between England and China, however, they are, on the other hand, also an opportunity to consider potential long-term dynamics when China moves towards the steady state with higher traffic density and lower investments.

[10] The MIDAS data group vehicles into four types by length: (1) less than 5.2 meters, (2) between 5.2 and 6.6 meters, (3) between 6.6 and 11.6 meters, and (4) above 11.6 meters.

[11] Recall that the MIDAS data only report the average speed of all vehicles passing through a measurement site over every 15-minute interval and does not condition the average speed on vehicle type.

[12] This new network is constructed following the same approach as that in Section 2.2, but we replace the provincial and national roads with highways. The road network for England is based on the shortest route, which covers 63.0% of the highways. To facilitate comparison, the highway network for China is also based on the shortest route.

[13] For a comparison of England's network to European networks, see e.g. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/212590/action-for-roads.pdf.

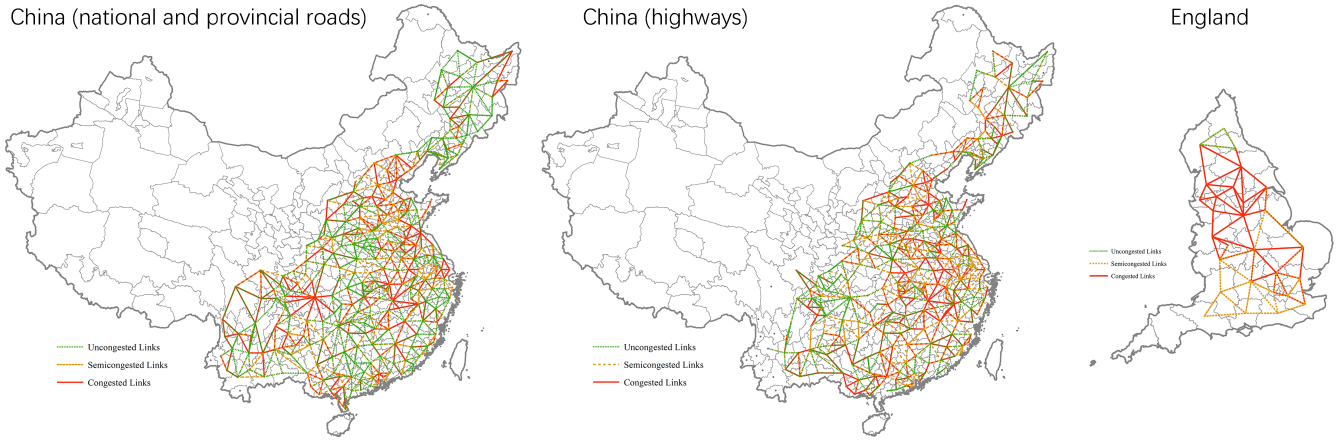

Figure A.6. Comparison of δ_{2kl}s' Spatial Distribution

Notes. The solid red, dashed orange, and dotted green lines indicate congested, semicongested, and uncongested links, respectively. The δ_2 of congested links is 0.753, 0.488, and 0.533 in the left, middle, and right sub-figure, respectively; The δ_2 of semicongested links are 0.216, 0.122, and 0.392 in the left, middle, and right sub-figure, respectively. This figure is based on the 4-min-per-hour sample.

Table A.4. Estimation of Congestion Elasticity for England

	(1)	(2)	(3)	(4)	(5)	(6)
	One Group	Two Groups		Three Groups		
	All Links	Congested Links	Uncongested Links	Congested Links	Semi-congested Links	Uncongested Links
Share of Links	100%	73.33%	26.23%	62.30%	35.25%	2.45%
	2SLS					
Truck density (ln)	0.478***	0.514***	0.356***	0.533***	0.392***	0.035
	(0.052)	(0.050)	(0.065)	(0.052)	(0.054)	(0.092)
	First Stage					
IV (ln)	0.492***	0.516***	0.414***	0.530***	0.442***	0.103
	(0.053)	(0.052)	(0.064)	(0.054)	(0.054)	(0.099)
Kleibergen-Paap rk Wald F Statistics	9,913.993	7,280.579	2,894.091	6,523.472	3,601.946	266.808
Link fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Observation	26,075	19,440	6,635	16,416	9,070	589

Notes. The table shows the estimates of the congestion elasticity δ_{2kl}. In column (1), the elasticity is constrained to be identical. In columns (2)-(3) we allow for 2 groups. In columns (4)-(6) we allow for 3 groups. Standard errors are reported in parentheses and are clustered at the level of link and day. We use the group classification implied by optimized tuning parameters. * 10%, ** 5%, *** 1% significance level.

Table A.5. Comparison of China and England's Highways

	(1) China	(2) China Yangtze River Delta, Pearl River Delta and Beijing-Tianjin-Hebei	(3) England
Lane number	4.27	4.89	4.56
Highways' length density	2.70	4.01	5.09
Highways' area density	11.53	19.62	23.22
Average distance from grids (1km) to highways	24.54	11.84	10.46
Average distance from grids (10km) to highways	24.56	11.89	10.46
Percent of grids (1km) where highways go through	3.0%	4.9%	5.1%
Percent of grids (10km) where highways go through	25.1%	36.8%	39.2%
Population density (person / squared km)	265.88	550.57	393.06
GDP per capita (thousand US dollar)	7.05	13.20	29.00
Average night light	44.74	63.8	74.82
Gini coefficint of GDP	0.53	0.52	0.34
Gini coefficint of GDP per capita	0.29	0.30	0.22
Share of congested links ($\delta_{2,kl}$)	33.3% (0.49)	36.1% (0.51)	62.3% (0.53)
Share of semicongested links ($\delta_{2,kl}$)	39.9% (0.12)	47.1% (0.13)	35.2% (0.39)

Notes. Highways' length density is calculated as the length of highways divided by areas of cities, with unit $km/100km^2$. Highways' area density is calculated as the area of highways (which is the multiplication of length and lane number) divided by areas of cities, with unit $km \times lane/100km^2$. We construct two types of grids covering the cities, one with size $1km \times 1km$, and the other with $10km \times 10km$. Average distance from grids (1km or 10km) to highways is the average value of the distances from each grid's center point to the nearest highways, and percent of grids (1km or 10km) where highways go through is the percentage of grids which intersect with highways. The population, GDP and night light are all gridded dataset: the population comes from LandScan Datasets for 2018, where the resolution is 1 km; the GDP comes from Global Dataset of Gridded Population and GDP (1980-2010 Estimations and 2020-2100 Scenarios) for 2010, where the grid size is 0.5×0.5 degree and the unit is Billion $US\$2005/yr$ under PPP; the night light comes from Flint from China Remote Sensing Satellite Ground Station, Chinese Academy of Science, which is originally from Version 1 VIIRS Day/Night Band Nighttime Lights for 2018 with grid size as 15 arc-second.

A.6. Fundamental Law of Road Congestion

We discuss the relevance and exclusion restriction of our IV based on historical roads in 1962. We first discuss the relevance. The relevance of the instrument relies on the argument that constructing new roads alongside or even on top of existing roads is less expensive. The strong first stage confirms that historical roads are a good predictor of today's roads.

The exogeneity of the instrument requires that historical roads cannot be correlated with today's traffic through channels other than today's road capacity. We address this from three perspectives. First, from the perspective of an institutional background, Baum-Snow et al. (2020) discuss in detail that historical roads in 1962 were built under an agricultural and planned economy and for the purpose of connecting rural areas to cities rather than of constructing intercity links. Second, we use the hypothetical geography-based least-cost path, rather than the actual link, to alleviate the endogeneity of the IV. As discussed in Baum-Snow et al. (2020), there may still be serially correlated unobservable factors (like local exogenous productivities or demand for road expansion from urban areas), which can predict roads in 1962 as well as today's flows within the region. For related reasons, we exclude urban areas defined by the 1962 administrative boundary map from the buffer zone.[14] Specifically, when constructing our instrument, all road lines and buffer zones in historical urban areas are excluded. We illustrate this instrument in an example in Figure A.7. Third, some control variables are included in the regression to ensure the conditional exogeneity. We directly control for population density in 1964 to capture the historical economic conditions and urban share in 1964 to proxy for the historical industrial structure.[15]

[14]Using the 1962 map alleviates the concern that today's urbanization is a result of road expansion and is, thus, endogenous. The 1962 county-level boundary is from "Administrative Boundary Maps of China, 1949-2014", published by China Data Center. All "districts (*qu*)" are considered urban areas.

[15]1964 is the year closest to 1962 for which we can access such variables.

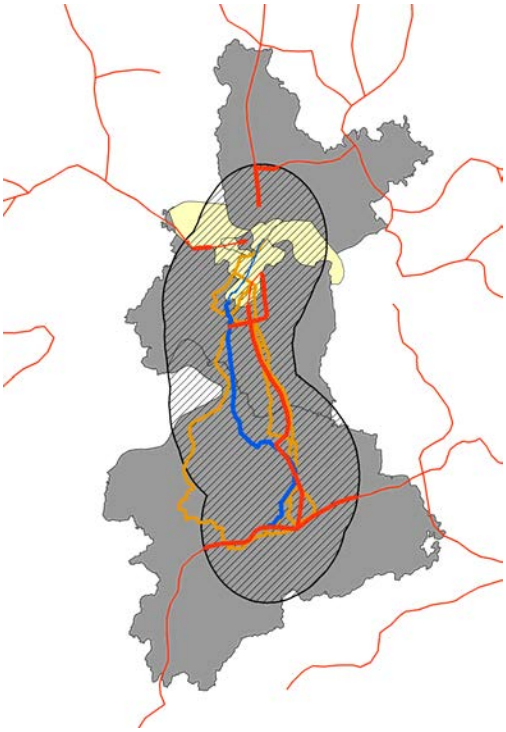

Figure A.7. Illustration of the Instrument for Road Capacity

Notes. The figure shows Wuhan and Xianning as an example. The orange line indicates the modern link consisting of the five least-cost routes, the blue line indicates the geography-based least-cost route, and the red lines indicate historical roads in 1962. The yellow areas are districts in 1972 and the shaded area is the 30 km buffer zone along the blue line.

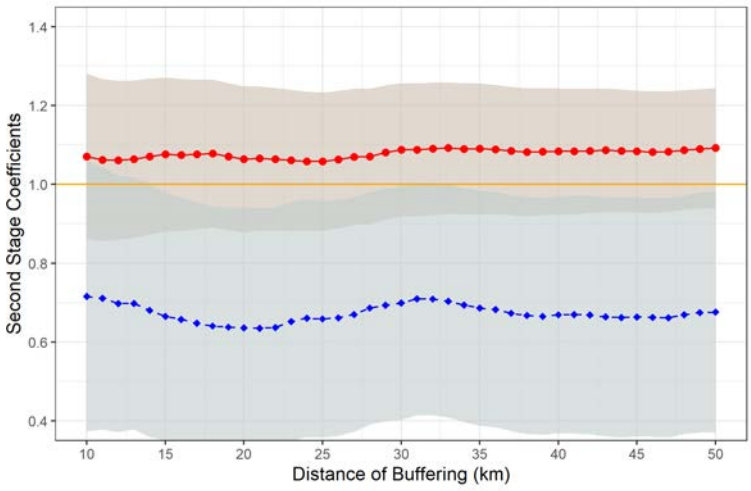

Figure A.8. Robustness Checks based on Different Buffering Distances

Notes. The red solid line represents the congested and semicongested links and the blue dashed line represents the uncongested links. The shade areas indicate the 95% confidence intervals.

Table A.6. Fundamental Law of Road Congestion, OLS Regression Results

Dep: VKT (ln)	(1) All Links	(2) Congested Links	(3) Semicongested Links	(4) Uncongested Links
Capacity (ln, *lag*)	1.093***	1.156***	1.192***	0.938***
	(0.035)	(0.057)	(0.067)	(0.059)
Control variables	Yes	Yes	Yes	Yes
Observation	1,392	479	395	518
R^2	0.631	0.640	0.630	0.611
p value for H_0			0.000	

Notes. Robust standard errors are reported in parentheses. The sample is the cross-section of 2018. The congestion classification comes from Table 1. The control variables include speed limit (ln, *lag*), night light in 2016 (ln), population density in 1964 (ln), urban share in 1964 (ln) and ruggedness (ln). H_0: the coefficients of Capacity $(\ln, lag)$ in congested (column (2)) and uncongested (column (4)) groups are the same. * 10%, ** 5%, *** 1% significance level.

Table A.7. Fundamental Law of Road Congestion, Robustness Checks based on Speed Ratio

	(1)	(2)	(3)	(4)
	Full Sample		Costal Provinces	
	Top 35%	Bottom 37%	Top 35%	Bottom 37%
Dep: VKT (ln)	Panel A: 2SLS			
Capacity (ln, *lag*)	1.346***	0.876***	1.503***	0.411
	(0.115)	(0.134)	(0.238)	(0.500)
Control variables	Yes	Yes	Yes	Yes
Observation	488	516	132	139
R^2	0.724	0.662	0.705	0.559
p value for H_0	0.005		0.000	
Dep: Capacity (ln, *lag*)	Panel B: First Stage			
Instrument variable	0.618***	0.590***	0.554***	0.355***
	(0.065)	(0.059)	(0.093)	(0.094)
Control variables	Yes	Yes	Yes	Yes
Observation	488	516	132	139
R^2	0.260	0.274	0.308	0.263
Kleibergen-Paap rk Wald F statistic	89.272	100.895	35.148	14.123

Notes. Robust standard errors in parentheses. The sample is the cross-section of 2018, and covers coastal provinces (Liaoning, Tianjin, Hebei, Shandong, Jiangsu, Shanghai, Zhejiang, Fujian, Guangdong, and Guangxi). The IV is the multiplication of historical road density in the 30 km buffering zone of the geography-based least cost path, and the length of the geography-based least cost path, where historical urban areas ("districts" in 1962) are excluded. The control variables include speed limit $(\ln, lag)$, night light in 2016 $(\ln)$, population density in 1964 $(\ln)$, urban share in 1964 $(\ln)$, and ruggedness $(\ln)$. H_0: the coefficients of Capacity $(\ln, lag)$ are the same in columns (1) and (2) (or in columns (3) and (4)). * 10%, ** 5%, *** 1% significance level.

A.7. Structural Estimation

Algorithm

Our estimation of θ, κ, and ζ is based on the following algorithm.

- Step 1: Start with a guess for $hours_{kl}$, denoted by $hours_{kl}^{guess}$. Then, we obtain t_{kl} and τ_{ij} from (21) and (17), respectively.
- Step 2: We use city-level GDP, denoted by Y_i^{data}, for Y_i. The goods market clearing and Dixit-Stiglitz price index, (11) and (10), solve p_i and P_i. Then, (9) gives $X_{ij} = \tau_{ij}^{1-\sigma} p_i^{1-\sigma} P_j^{\sigma-1} Y_j$. Substituting X_{ij} into (19), we obtain Ξ_{kl}.
- Step 3: Substituting Ξ_{kl} into (20), we obtain $flow_{kl}$, which leads to $traf_{kl}$ by the definition of traffic flow and, then, $hours_{kl}^{updated}$ by the congestion equation (1).
- Step 4: Iterate on $hours_{kl}^{updated}$ until convergence. $traf_{kl}^{sim}$ and $hours_{kl}^{sim}$ are $traf_{kl}$ and $hours_{kl}$ associated with the converged $hours_{kl}$.

Estimation Results for England

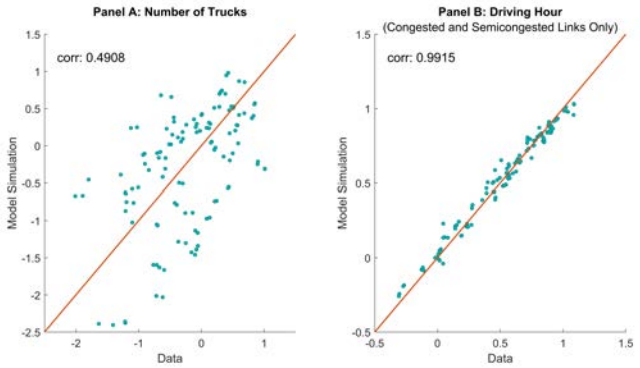

Figure A.9. Model Fit for England

Notes. Panels A and B plot $\ln(traf_{kl})$ and $\ln(hours_{kl})$ in the data and simulated from the estimated model, respectively. Panel B shows only congested and semicongested links, as the simulated $\ln(hours_{kl})$ are, by construction, always identical to those in the data for uncongested links.

Table A.8. Results of the Structural Estimation of Model Parameters for England

	θ	κ	ζ
Estimated parameter	38.10	0.07	0.61
(Standard error)	(3.71)	(0.01)	(0.06)

Notes. The table shows the results of the structural estimation of equation (22). Standard errors reported in parentheses are calculated through the numerical implementation of the covariance matrix of nonlinear least square estimation from Hansen (2022).

A.8. Returns to Investment

Comparison of Welfare Elasticities for China and England

We compare the welfare elasticity between China and England in Table A.9. The elasticity for England turns out to be much larger than that for China: The mean value of the welfare elasticity for China is almost one-fifth as large as that for England. This comparison not only indicates the huge contrast in the current economic stage between China and England but also implies the potential level of welfare gain when China moves to the steady state like England (with very high road density and low road investment) in the future.

Table A.9. Comparison of Welfare Elasticity between China and England

	(1) China	(2) England
mean	0.22	0.98
std	0.34	0.71
25%	0.05	0.37
50%	0.11	0.88
75%	0.29	1.35

Notes. The table reports welfare elasticity multiplied by 10,000.

Welfare Gain

Figure A.10. Gains from Road Expansion

Notes. Each straight line represents a link. The darkness of the color represents the magnitude of the welfare gain $B\left(d_{kl}\right)$ in equation (24). Darker colors imply larger welfare gains. For visualization purposes, d_{kl} is set to 1 in this figure.

Cost Calculation Procedure

In this subsection, we discuss the two steps to calculate the topography-related road construction cost in more detail, and also discuss the robustness check for the calculation of the opportunity cost of land.

First, we discuss how we obtain the national average value of 8.7 million RMB per lane-km. According to the *Statistical Report of the Development of Transportation Industry* from 2013 to 2017, the total investment on road construction during this period is 7,388,489 million RMB, and the increments of roads are 350,300 kilometers. We assume on average the regular national and provincial roads have 4 lanes (2 lanes for each direction), and the statistical data tell us that the average lane number for national and provincial highways is around 4.5. Then, the total increment of national and provincial roads (including regular roads and highways) is 849,500 lane-km. This leads to an average unit cost as 8.7 million RMB per lane-km.

In the second step, we illustrate how to combine the average cost with the geographic slope to calculate the unit construction cost of each link. The slope of each link proxies the relative values of unit construction costs, and we transform the relative values proxied by the slope to the unit construction costs in terms of RMB according to the following specification:

$$UC_{kl}^{S} \propto 1 + slope_{kl},$$

by requiring the weighted average of UC_{kl}^{S} equal to the average cost from the first step. In practice, suppose we know the unit construction cost for one specific link. Since the slope determines the relative level of unit construction costs, all other unit construction costs can be computed subsequently. Then, by taking the weighted average of these unit construction costs and requiring the weighted average to equal 8.7 million RMB per lane-km, we can easily solve for the appropriate factor and then get the exact values for the construction costs of all links. This allows us to finally recover the unit costs for each link. We visualize the spatial distribution of the unit construction cost in Figure A.11, and report the summary statistics in columns (1)-(2) of Table A.10.

Third, we discuss the robustness of our calculation of the opportunity cost of land. Based on our approach as described in the main text, there are around 15.5% of road segments (in terms of capacity) that cannot be matched with any land parcel, and these segments are excluded when we calculate the land cost. Although these segments are located in less urbanized areas with potentially different prices, we do not expect that these missing values would affect the validity of our calculation. The reason for this expectation is that we also find that along roads, the share of industrial land is higher than the share of commercial and residential land, and the price of the former is generally lower everywhere due to local governments' subsidies (Xu, Huang and Jiang, 2017; Chen and Kung, 2019; Lin et al., 2020). As a robustness check, for each such segment, we select a similar segment with land price information by using propensity score matching based on the night light, population density, ruggedness, urban share, and distance to the closest city center of each segment. We assign the land prices of the matched segments to the segments with unobserved prices and recalculate the land cost, which yields very similar results.

Figure A.11. Unit Construction Costs

Notes. Each straight line represents a link. Darker colors represent higher unit road construction costs, measured in million Yuan (2018) per lane-km.

Figure A.12. Unit Opportunity Costs of Land

Notes. Each straight line represents a link. Darker colors represent higher opportunity costs of land, measured in million Yuan (2018) per square km.

Table A.10. Summary Statistics of Costs from One Additional Lane

	(1)	(2)	(3)	(4)
	Total Construction Cost		Total Opportunity Cost of Land	
	Congested and Semicongested Links	All Links	Congested and Semicongested Links	All Links
mean	7241.62	7461.42	2734.28	2536.05
std	8294.15	8386.45	2334.01	2093.26
25%	2113.84	2223.43	1265.54	1202.36
50%	4152.30	4567.14	2050.59	1941.94
75%	9269.46	9745.89	3476.31	3276.10

Notes. The unit is million RMB.

Returns

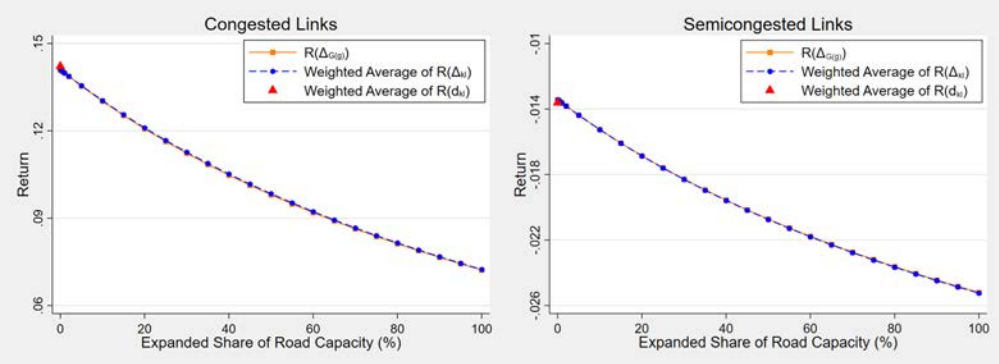

Figure A.13. Diminishing Returns of Capacity Expansion

Notes. The red triangle indicates the weighted (by cost) average of $R\left(d_{kl}\right)$, the blue dots indicate the weighted (by cost) average of $R\left(\Delta_{kl}\right)$ with the same percentage of lane expansion for all links, and the orange squares indicate $R\left(\Delta_{G}\left(g\right)\right)$ from the joint road expansion with the same percentage. The percentages of lane expansion are reported in the horizontal axis.

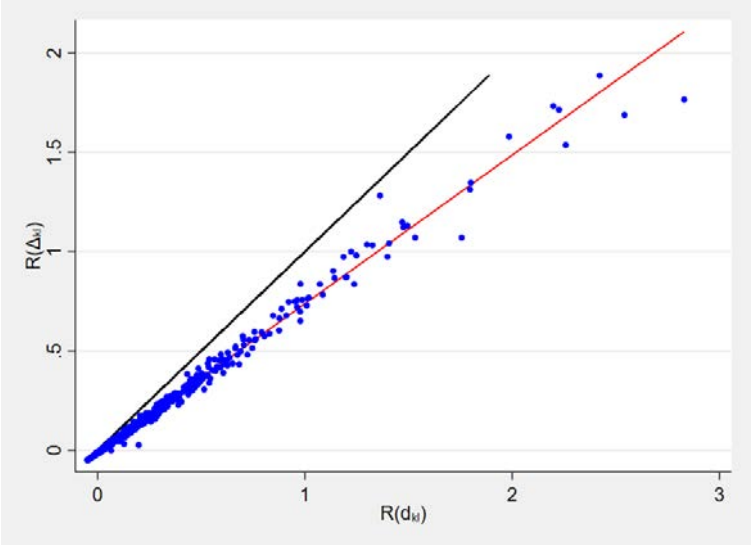

Figure A.14. Extensions of Return Calculation

Notes. The black line is the 45-degree line, and the red line is the linear fit. $R\left(\Delta_{kl}\right)$ at the vertical axis is calculated based on $\Delta_{kl}=1$ for each link kl.

Table A.11. Returns to Expanding One Link's Capacity, Robustness Checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$R(d_{kl})$				$R(\Delta_{kl})$			
	Benchmark	Excluding Land Cost	No Labor Mobility	No Externality	Benchmark	Excluding Land Cost	No Labor Mobility	No Externality
	Congested Links							
mean	0.29	0.66	0.30	0.25	0.21	0.48	0.21	0.18
std	0.42	1.19	0.51	0.37	0.32	0.89	0.38	0.28
wgt mean	0.14	0.44	0.16	0.12	0.10	0.31	0.11	0.08
wgt std	0.29	1.23	0.36	0.25	0.22	0.90	0.28	0.19
25%	0.02	0.04	0.00	0.01	-0.00	0.02	-0.01	-0.01
50%	0.15	0.30	0.14	0.13	0.11	0.20	0.09	0.09
75%	0.40	0.78	0.38	0.36	0.29	0.59	0.29	0.27
	Semicongested Links							
mean	0.02	0.08	0.01	0.01	0.01	0.05	-0.00	0.00
std	0.11	0.21	0.10	0.10	0.09	0.17	0.08	0.08
wgt mean	-0.01	0.02	-0.02	-0.02	-0.02	0.00	-0.02	-0.02
wgt std	0.06	0.11	0.05	0.05	0.04	0.09	0.04	0.04
25%	-0.04	-0.03	-0.04	-0.04	-0.04	-0.03	-0.04	-0.04
50%	-0.01	0.01	-0.02	-0.01	-0.02	-0.00	-0.02	-0.02
75%	0.03	0.09	0.03	0.03	0.02	0.07	0.01	0.01

Notes. Only congested and semicongested links are reported in this table, and the returns for uncongested links are -0.05. The benchmark case includes the land cost, and the labor mobility and externality ($\alpha = 0.1, \beta = -0.3$) are allowed. $R(\Delta_{kl})$ is calculated based on $\Delta_{kl} = 1$ for each link kl. "wgt mean" and "wgt std" indicate cost-weighted mean and standard deviation.

Table A.12. Top 20 Links in Terms of Return

(1)	(2)	(3)	(4)	(5)
Rank	Including Land Cost		Excluding Land Cost	
	Link	Return	Link	Return
1	Qingdao-Yantai	2.83	Shanghai-Suzhou	16.20
2	Yueyang-Jingzhou	2.54	Taizhou-Yangzhou	6.30
3	Qingdao-Weifang	2.42	Shanghai-Jiaxing	6.23
4	Bozhou-Huaibei	2.26	Yangzhou-Taizhou	5.63
5	Taizhou-Yangzhou	2.23	Bozhou-Huaibei	4.96
6	Shangqiu-Bozhou	2.20	Changzhou-Wuxi	4.81
7	Yangzhou-Taizhou	1.98	Shangqiu-Bozhou	4.72
8	Neijiang-Ziyang	1.80	Suzhou-Wuxi	4.63
9	Tianjin-Tangshan	1.80	Bozhou-Fuyang	4.29
10	Yantai-Qingdao	1.76	Yueyang-Jingzhou	4.08
11	Binzhou-Dongying	1.53	Qingdao-Weifang	3.94
12	Lianyungang-Linyi	1.49	Wuxi-Changzhou	3.93
13	Heze-Jining	1.47	Qingdao-Yantai	3.70
14	Bozhou-Fuyang	1.47	Suzhou-Jiaxing	3.36
15	Xuzhou-Zaoyang	1.41	Guangzhou-Foshan	3.35
16	Shanghai-Suzhou	1.40	Tianjin-Tangshang	3.34
17	Lishui-Wenzhou	1.36	Heze-Jining	3.20
18	Taizhou-Yancheng	1.33	Foshan-Guangzhou	3.18
19	Zhoukou-Kaifeng	1.30	Taizhou-Yancheng	3.17
20	Huaian-Lianyungang	1.25	Cangzhou-Tianjin	2.90

Notes. The table reports $R(d_{kl})$.

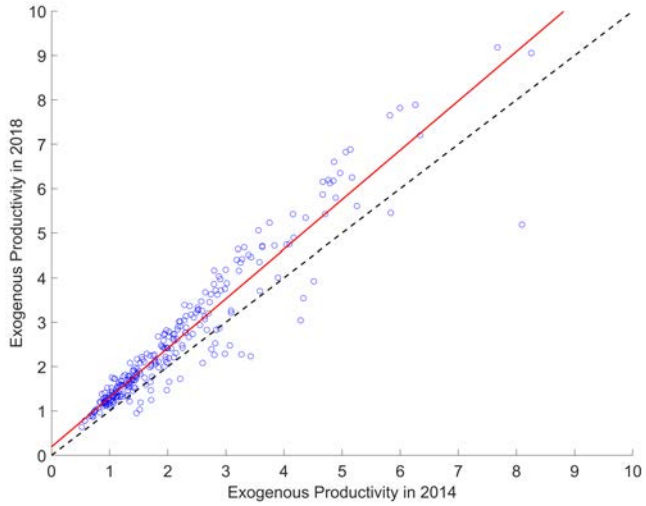

Figure A.15. Exogeneous Productivity

Notes. The dashed black line is the 45-degree line, and the red-solid line is the linear fit.

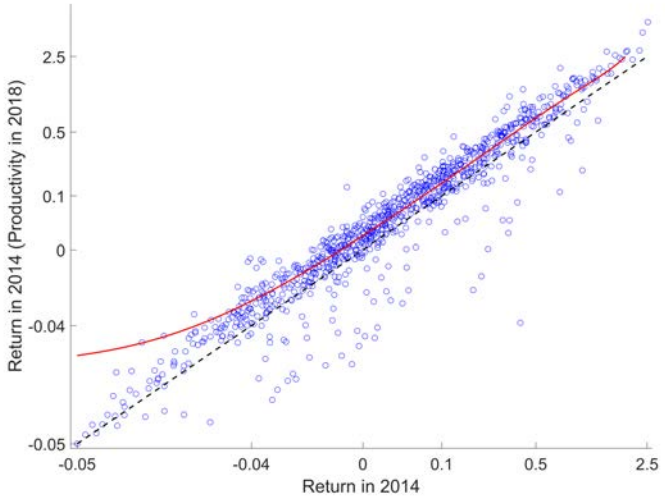

Figure A.16. Comparison of Returns $R\left(d_{kl}\right)$ between Scenarios

Notes. The dashed black line is the 45-degree line, and the solid red line is the polynomial fit of degree 4. Other degree levels only slightly change the fitted line between -0.05 and -0.04, and do not affect the shape of the fitted line in other ranges.

A.9. Allocation Efficiency

Numerical Exercises

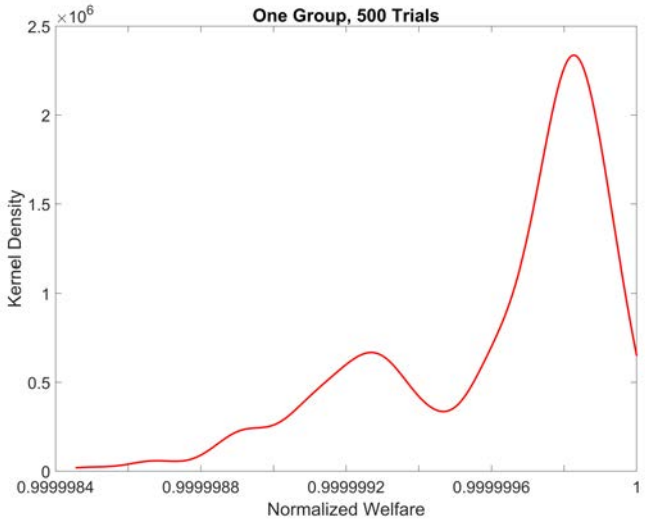

Figure A.17. Welfare Distribution across 500 Trials

Notes. For comparison purposes, we normalize the largest welfare as 1 on the horizontal axis.

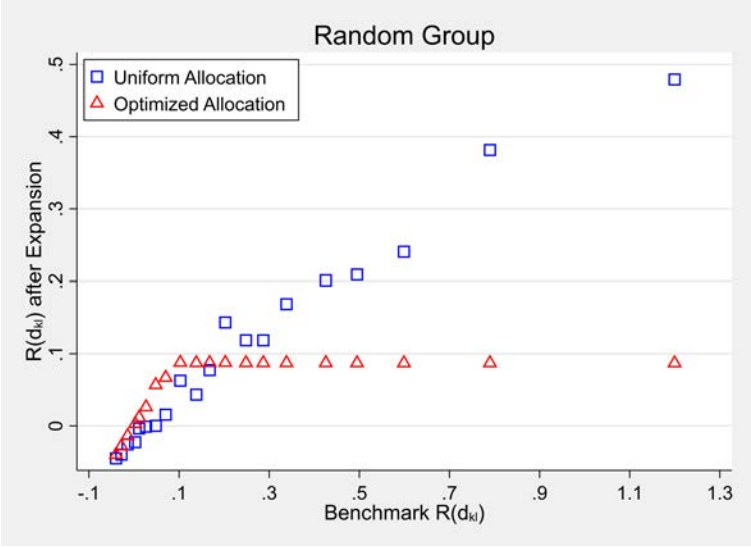

Figure A.18. Optimal Capacity Expansion and Post-Optimization Return Equalization

Notes. Each triangle represents a link in the random group. The x- and y-axis plot $R\left(d_{kl}\right)$ before and after the optimized investment allocation, respectively.

Table A.13. Returns to Road Investment (Random Groups)

	(1)	(2)	(3)	(4)
	Heterogeneous δ_2		Homogeneous δ_2 ($\delta_2 = 0.15$)	
	subsample of congested links	full sample	subsample of congested links	full sample
	Panel A: Average of weighted mean of $R(d_{kl})$ or $\bar{R}(d_{kl})$ across 100 groups			
Wgt mean	0.16	0.04	-0.00	-0.01
	(0.29)	(0.17)	(0.06)	(0.06)
	Panel B: Weighted mean of $R(d_{kl})$ or $\bar{R}(d_{kl})$			
Wgt mean	0.14	0.03	-0.00	-0.01
	(0.29)	(0.20)	(0.06)	(0.06)

Notes. This table reports the weighted statistics of $R(d_{kl})$, where "wgt" indicates "weighted". In panel A, We first calculate the weighted mean and standard deviation of $R(d_{kl})$ across the 20 links within each group and then take the average across the 100 groups. Thus, Panel A reports the across-group average of within-group means and within-group standard deviations in parentheses. In Panel B, as a reference, we repeat the weighted statistics of the returns from Table 4. In columns (1) and (3), we focus on links selected from the subsample of congested links, while in columns (2) and (4), we focus on the full sample of all links. Congested links are selected based on heterogeneous δ_2.

Misallocation of Actual Road Expansion

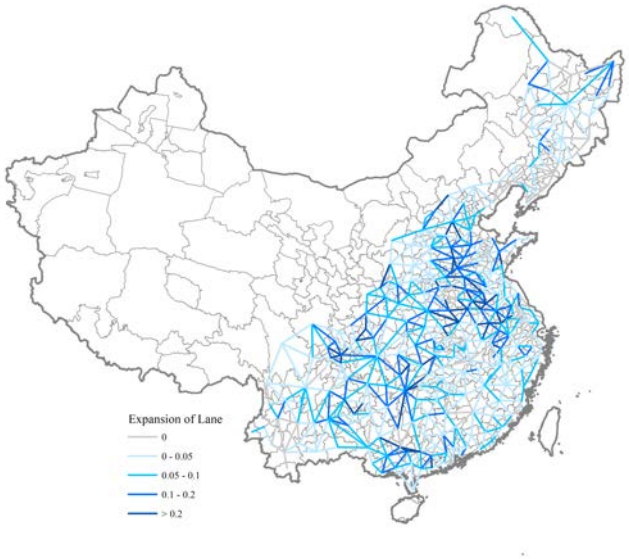

Figure A.19. The Expansion of Lanes in 2014-18

Notes. Each straight line represents a link. The darkness of the color represents the magnitude of lane expansion between 2014 and 2018. Darker colors imply larger lane expansion.

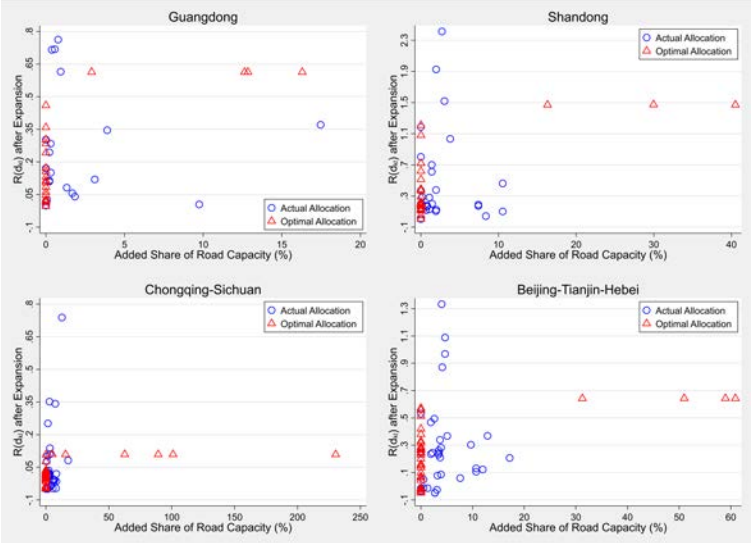

Figure A.20. Return Equalization after Optimal Expansions

Notes. Each circle or triangle represents a link. The horizontal axis shows the percentage increment based on the capacity of the initial road network in 2014. The vertical axis shows the $R(d_{kl})$ based on the expanded road network.

A.10. Returns and Welfare under Homogeneous Congestion Elasticity

Table A.14. Comparison of Estimation Results

	(1) Homogeneous δ_2 ($\delta_2 = 0.15$)	(2) Comparison: Heterogenous δ_2
θ	45.96	44.74
	(0.13)	(0.45)
κ	0.02	0.02
	(0.00)	(0.00)
ζ	-4.41	-5.68
	(0.11)	(0.00)

Notes. Standard errors reported in parentheses are calculated through the numerical implementation of the covariance matrix of nonlinear least square estimation from Hansen (2022).