Journal of International Economics 157 (2025) 104140

Contents lists available at ScienceDirect

Journal of International Economics

journal homepage: www.elsevier.com/locate/jie

Full length articles

Chinese roads in India: The effect of transport infrastructure on economic development☆

Simon Alder *

Swiss National Bank, Switzerland
University of North Carolina at Chapel Hill, United States of America

ARTICLE INFO

Dataset link: Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development (Original data)

JEL classification:
F11
F14
F15
O11
O18
R12
R13

Keywords:
Transport infrastructure
Highways
Network design
Trade
Economic growth
Regional development
India
China
Geographic information system

ABSTRACT

India and China followed different strategies in the design of their recent highway networks. India first focused on connecting the four largest economic centers of the country, the Golden Quadrilateral, while China had the explicit strategy of connecting intermediate-sized cities. This paper analyzes the aggregate and distributional effects of transport infrastructure in India based on a general equilibrium trade framework. I compare the effect of the Golden Quadrilateral to a counterfactual network that connects India's intermediate-sized cities. To construct the counterfactual network, I propose a heuristic network design algorithm to maximize aggregate real income net of road construction costs in the general equilibrium model, and I show that the heuristic algorithm provides a good approximation of the optimal network. The results suggest that the actual network led to sizable aggregate gains but unequal effects across regions. The income-maximizing counterfactual network is substantially larger than the actual Indian network, would imply further aggregate gains, and would benefit the lagging regions of India.

☆ I am grateful to the editor, Treb Allen, and the anonymous referees for their comments, which significantly improved the paper. I would also like to thank Fabrizio Zilibotti for his support and guidance. I also thank Simeon D. Alder, George-Marios Angeletos, Costas Arkolakis, Marco Bassetto, Timo Boppart, Filippo Brutti, Marius Brülhart, Adrien Bussy, Jonathan Colmer, Kerem Coşar, Guido Cozzi, Gregory Crawford, Andreas Moxnes, Mariacristina De Nardi, Dave Donaldson, David Dorn, Jonathan Eaton, Gino Gancia, Ed Glaeser, Vernon Henderson, Lutz Hendricks, Roland Hodler, Illenin Kondo, Peter Kondor, Michael König, Rafael Lalive, Sergi Jiménez-Martín, Omar Licandro, Thierry Mayer, Andreas Müller, Alessandro Pavan, Michelle Rendall, José-Víctor Ríos-Rull, Dominic Rohner, Esteban Rossi-Hansberg, Lin Shao, Kjetil Storesletten, Viktor Tsyrennikov, Ashish Vachhani, Valentin Verdier, Rainer Winkelmann, Christoph Winter, Gabriel Zucman, Josef Zweimüller, Robert Zymek, and participants in presentations for helpful comments. Sebastian Ottinger provided outstanding research assistance. Ronald Schmidt and Larry Crissman provided valuable support with GIS software and data. Financial support from the European Research Council (ERC Grant IPCDP-229883) and from the University of North Carolina at Chapel Hill is gratefully acknowledged. The views expressed in this paper are those of the author and do not necessarily reflect those of the SNB.

* Correspondence to: Swiss National Bank, Börsenstrasse 15, 8001 Zürich, Switzerland
E-mail address: simon.alder@snb.ch.

https://doi.org/10.1016/j.jinteco.2025.104140
Received 27 April 2024; Received in revised form 18 July 2025; Accepted 19 July 2025

Available online 5 August 2025

1. Introduction

India and China followed different strategies in the design of their recent highway networks. The Indian government launched a national highway project in 2001 that improved connections between the four largest economic centers Delhi, Mumbai, Chennai, and Kolkata with a network known as the "Golden Quadrilateral" (GQ). In contrast, China built a National Expressway Network (NEN) that had the explicit goal of connecting all intermediate-sized cities with a population above 500,000 and all provincial capitals with modern highways. This led to stark differences in the modern highway networks of the two countries, as shown in Fig. 1. Overall, China invested about ten times more in its highway network than India, which is seen as being severely constrained by its insufficient infrastructure (Harral et al., 2006). Furthermore, China and India had different spatial development patterns over the past decades (Desmet et al., 2013; Chaudhuri and Ravallion, 2006).

Transport infrastructure can be an important determinant of development (Donaldson, 2015), and it is crucial to understand how such differences in infrastructure policies affect economic development. In this paper, I compare the aggregate and distributional effects of the GQ to a counterfactual Indian network that maximizes aggregate real income in a general equilibrium trade model based on Donaldson and Hornbeck (2016). The nodes of the network are all Indian cities with a population above 500,000 and all state capitals, thereby implementing the Chinese policy of connecting intermediate-sized cities. I propose a heuristic network design algorithm that maximizes aggregate real income net of road construction costs. While the solution of the algorithm is not guaranteed to be the global optimum, I obtain an estimate of the maximum income that can be reached with the globally optimal network, together with confidence bounds around the estimated maximum. I show that the solution of the algorithm is close to the upper bound of the confidence interval, and the solution is, thus, unlikely to differ significantly from the global optimum.

The actual and counterfactual networks are evaluated in a general equilibrium trade model based on Donaldson and Hornbeck (2016) who estimate the effect of railways on agricultural land values in the US. The model allows for trade among locations that are assumed to differ in productivity as in Eaton and Kortum (2002). Trade flows are subject to trade costs that depend on the transport infrastructure. Donaldson and Hornbeck (2016) show that in this framework, the general equilibrium effects of changes in the transport network are captured by a measure of market access. A location's market access is the sum over the incomes of all other locations, discounted by the bilateral trade costs and by the other locations' market access. The model yields a gravity equation for bilateral trade that can be aggregated over destinations to obtain a log-linear relationship between income and market access that can be estimated using panel data for Indian districts' nominal income and market access. The market access measures are general equilibrium outcomes that I obtain from the model for each set of trade costs. For a given transport network, these trade costs can be derived from the computed shortest path between all district centroids. Hence, the bilateral trade costs can be calculated for the transport network in 2000 (before the construction of the GQ), in 2012 (after the completion of the GQ), and for the counterfactuals.

The observed changes in market access from 2002 to 2012 due to the construction of the GQ allow me to estimate the elasticity of income with respect to market access while controlling for unobserved heterogeneity with district fixed effects. The estimated elasticity is consistent with the model for reasonable parameter values. I then use the full structure of the general equilibrium model to quantify the aggregate and distributional effects of the networks and design the counterfactual network that maximizes the general equilibrium income net of road construction costs.

The paper makes three contributions. First, I quantify the aggregate effect of the GQ that connected India's four largest economic centers. The result suggests that aggregate real GDP (net of construction and maintenance costs) would have been 0.6% lower in 2012 if the GQ had not been built.

Second, I propose a heuristic network design algorithm to construct the income-maximizing highway network among India's intermediate-sized cities. The algorithm solves the discrete problem of which cities should be directly connected, which is an important margin in both India's and China's national infrastructure plans because both countries had the goal of developing a new national network of a given road standard. The network design algorithm starts from the full network and iteratively removes and adds links based on their effect on general equilibrium real income net of road construction and maintenance costs until no further increases in net income are possible.

The resulting network is substantially larger than the GQ and would connect most intermediate-sized cities. It would cost more than three times as much as the GQ, but it would lead to a net increase in aggregate income (relative to the GQ) of 4.3%. I also compare the effect to an alternative counterfactual network that imposes a budget constraint equal to the budget of the first two phases of the Indian Highway Development Project (which includes additional highway connections besides the GQ). The resulting network would imply an increase in net income of 3.2%. While these counterfactual network designs are computed with a heuristic algorithm and are not guaranteed to be the global optimum, I show statistically that they are likely within narrow bounds of the optimal network.[1] To obtain these bounds, I estimate a Generalized Extreme Value distribution using net incomes implied by the algorithm's solution for a large number of random starting networks. The upper bound of the 95% confidence interval of the estimated maximum is less than 0.01% above the solution of the algorithm starting from the full network. Furthermore, the results provide a lower bound for the net gains that could be achieved with the optimal network. The analysis yields new and important findings in the Indian context, suggesting that there are large additional income gains from building the optimal transport network. Furthermore, the income-maximizing network has more star-shaped links to the center of the country and thus differs from the actual and the planned transport networks not only in overall length but also in its structure.

[1] Finding the globally optimal network in this framework is challenging because there can be positive or negative complementarities between the links and the discrete problem of choosing the optimal combination of links is extremely large.

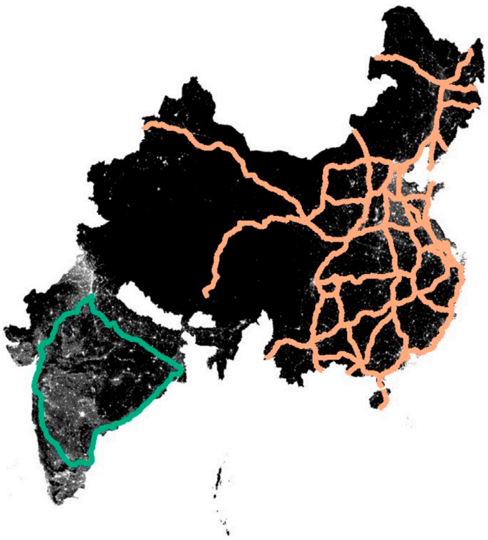

Fig. 1. Indian and Chinese highway projects.
The figure shows two major highway investment projects in India (Golden Quadrilateral, in green) and China (National Expressway Network, in red). The image in the background shows the night-time light intensity.

The third contribution is to evaluate the distributional consequences of the actual and counterfactual networks. This is particularly important in the Indian context due to its low rate of convergence among districts. I show that the network that maximizes aggregate net income may benefit initially less developed districts that, in comparison to the GQ, gain from the denser optimal network that reaches lagging regions.

The remainder of the paper is structured as follows. Section 2 reviews the related literature. Section 3 discusses the transport infrastructure in India and China. Section 4 presents the conceptual framework. Section 5 describes the heuristic network design algorithm and how the model is used to compute counterfactuals. Section 6 discusses the data and Section 7 the estimation strategy and results. Section 8 presents the effects of the actual and counterfactual networks. Section 9 discusses the robustness of the results and Section 10 concludes the paper.

2. Related literature

The role of transport infrastructure for development has recently been the subject of a growing literature that often uses detailed geographic information such as the location of transport infrastructure and data on outcomes at the sub-national level.[2] My methodology for evaluating the impact of infrastructure builds on Donaldson and Hornbeck (2016). They use a general equilibrium trade model to estimate the effect of an expansion of the American railway network on agricultural land values in the 19th century. Donaldson and Hornbeck (2016) also compare the effect of the actually built railway network to counterfactual scenarios in which railways are replaced by an extension of the canal network or a reduction in the cost of wagon transport on country roads. My counterfactual analysis differs from theirs by using the general equilibrium model to design a new network that approximately maximizes aggregate income net of road construction costs. I then compare the aggregate and distributional consequences of various alternative network designs.

To construct the counterfactual networks, I propose a heuristic network design algorithm that adds and removes links based on construction costs and aggregate income implied by trade costs in the general equilibrium framework.[3] Iterative procedures to solve related problems with complementarities are applied in Jia (2008), Antras et al. (2017), and Arkolakis et al. (2025), but not in the context of transport infrastructure networks.[4] Allen and Arkolakis (2022) propose a general equilibrium gravity model that allows for a characterization of the welfare effects from investments in each segment of a transport network. They apply the framework to the U.S. interstate highway network and find that the effects differ substantially across segments.[5] Fajgelbaum and Schaal (2020)

[2] See for example recent surveys by Breinlich et al. (2013), Redding and Turner (2015), and Donaldson (2015). The general decline of transport costs for goods and its implication for urban and regional development is discussed in Glaeser and Kohlhase (2004). See also Ahlfeldt et al. (2015) for a quantitative model that can be used to study the effect of transport infrastructure within cities.

[3] Heuristic algorithms have been applied in the network design literature, such as Gastner (2005) and Gastner and Newman (2006). In the economics literature, Burgess et al. (2015) and Balboni (2025) construct counterfactual networks by ranking pairs of cities by initial market potential based on Euclidean distance and connecting those with the highest rank, but they do not select links in order to maximize aggregate net income based on the general equilibrium model.

[4] Furthermore, the conditions in these papers to find global optima are not satisfied in my case of a transport network.

study the optimal infrastructure network design and propose a general equilibrium trade model with congestion in transport and a continuous infrastructure investment choice. When congestion is sufficiently strong, their framework leads to a convex optimization problem and allows them to compute the globally optimal transport network in Europe. I focus on the discrete choice of which links between any nodes should be part of the new highway network, instead of a continuous choice of the level of infrastructure investment across links. In the context of an emerging economy like India that plans a new national infrastructure network, this discrete choice of where the new infrastructure should be built is an important margin, since links may have a given standard, but the planners need to decide where to place them.[6]

While the empirical analysis of this paper builds on general equilibrium trade theory, it is also related to studies on the local effects of transport infrastructure such as the GQ. For example, Ghani et al. (2016) study the effects of the GQ on firms located in the proximity of the new highways and find positive effects on manufacturing activity. I rely on the identification strategy based on non-nodal districts (i.e., an inconsequential places approach) as in Chandra and Thompson (2000) but use it in the general equilibrium framework to estimate the effect of market access on income.[7] Furthermore, I apply an IV approach to account for the spatial dependence in incomes that enter the market access measures.

Some recent studies analyze transport infrastructure in India based on general equilibrium models. Asturias et al. (2019) quantify the effect of the GQ based on a model of oligopolistic competition applied to Indian states. Van Leemput (2021) analyzes internal and external trade barriers in India using state-level trade data. Allen and Atkin (2022) consider the effect of changes in the Indian highway network on the agricultural sector. Donaldson (2018) estimates the effect of railways in colonial India. My analysis differs from the above studies by estimating the effect of transport infrastructure in India through market access as in Donaldson and Hornbeck (2016), designing a counterfactual network that maximizes aggregate net income, and quantifying the aggregate and distributional implications from various alternative transport networks.

Several studies focus on the effect of transportation infrastructure in other countries.[8] Faber (2014) studies the NEN in China and uses the minimum spanning tree among the targeted cities as an instrument for the actual network. I follow his approach of modeling road construction costs based on topographical features and identifying cities that fulfill the Chinese criteria, but I apply the Chinese strategy to India and construct a counterfactual network that maximizes net income in a general equilibrium framework.

3. Transport infrastructure in India and China

Infrastructure is a key determinant of transport costs and trade (Limao and Venables, 2001) and investments in transport infrastructure have been used extensively to promote development (World Bank, 2007a). India and China have both invested in their transport infrastructure during the past decades, but with different intensities and strategies (Harral et al., 2006). Fig. 1 shows the two major highway projects in the two countries, India's GQ and China's NEN, and I discuss their characteristics and implications for transport costs below.

3.1. Past investments in transport infrastructure

In the early 1990s, the Indian road infrastructure was superior to the Chinese in terms of total kilometer length and kilometer per person, but both countries had about the same low quality of roads. Travel speeds on roads were further reduced by the simultaneous use by pedestrians and slow vehicles.[9] Over the 1990s, China's highway and railway network developed significantly faster than the Indian counterpart. In particular, China built the NEN (shown in red in Fig. 1) with the explicit objective of connecting all cities with more than 500,000 people and all provincial capitals in a modern highway system.[10] At that time, China's transport infrastructure was at risk of becoming a constraint for economic development, which was gaining speed since the reforms started in the late 1970s (Asian Development Bank, 2007). The new network had reached a length of 40,000 km by 2007 and it continued to be expanded. It consists of four-lane limited access highways that allowed significantly higher driving speeds than the existing roads.[11]

India also invested in its road infrastructure, but about ten times less than China and with a focus on the main economic centers. In particular, it launched a National Highways Development Project (NHDP) in 2001 and the first achievement of that project was the GQ, which connects the four major economic centers with four-lane highways (shown in green in Fig. 1). Construction, mostly

[5] Felbermayr and Tarasov (2022) model the endogenous distribution of transport infrastructure on a line and consider the density of transport infrastructure when approaching national borders. Blouri and von Ehrlich (2020) compute the optimal allocation of different transfers across European regions in a spatial equilibrium model and also consider transport infrastructure investments as a transfer type.

[6] The discrete nature of the decision is particularly relevant in the context of emerging and developing countries where entirely new networks might be planned, or the initial network is sparse. The Indian, as well as the Chinese government, sought to build new modern networks with a certain capacity of each link, more specifically highways of international standards with four to six lanes (see Faber, 2014; Ghani et al., 2016).

[7] See also Michaels (2008), Banerjee et al. (2020), Datta (2012), and Ghani et al. (2016) for examples of similar identification strategies.

[8] See, for example, Allen and Arkolakis (2014), Jaworski and Kitchens (2019), Baum-Snow et al. (2020), Roberts et al. (2012), Jedwab and Storeygard (2022), Bird and Straub (2020), Morten and Oliveira (2024), and Coşar and Demir (2016), Storeygard (2016), among others.

[9] The railway infrastructure in the two countries was similar in terms of passengers but the Chinese railways transported four times more freight than the Indian railways. The numbers in this section are taken (if not otherwise stated) from Harral et al. (2006).

[10] This is also referred to as the National Trunk Highway System. The program was later expanded to include all cities with more than 200,000 people. See World Bank (2007b), Roberts et al. (2012), and Faber (2014) for a discussion.

[11] A description of the history of the Chinese highway network and its different components was provided by the Australian Consortium for the Asian Spatial Information and Analysis Network (ACASIAN), People's Republic of China spatio-temporal expressway database.

upgrades of existing highways to higher quality, began in 2001 and was completed by 2012 with a total network length of 5846 km and at approximately a cost of USD 5.4 billion (1999 prices).[12] The NHDP in India was not restricted to the GQ and also included the so-called North-South and East-West (NS-EW) Corridors. However, the NS-EW were delayed such that by 2006 only 10% were built (the GQ was by then 95% complete) and the NS-EW were still not finished by 2012 (Ghani et al., 2016).

3.2. Implications for transport costs

The GQ in India, like the NEN in China, has significantly reduced the transport times between places that were connected by these new highways. The average driving speed on a conventional national highway (i.e. a highway that was not upgraded or built as part of the NHDP) was below 40 km/h (World Bank, 2002, 2005), while the driving speed on the GQ is around 75 km/h.[13] However, there is ample evidence that, even today, insufficient transport infrastructure is a severe constraint for the Indian economy. Raghuram Rajan, former Governor of the Reserve Bank of India, stated that India needs to improve its infrastructure with the same intensity in order to catch up with China (FAZ, 2013). The same view is held by the World Bank and several consultancies and logistic firms, stating that a lack of adequate infrastructure hampers the regional development in India (World Bank, 2008; Urban Land Institute and Ernst & Young, 2013).

3.3. Roads and other transport infrastructure

The road investment projects described above were among the largest inter-city transport infrastructure investments in the two countries and dominated investments in other means of transportation. The spending on the NEN in China was around USD 30 billion per year, roughly three times as much as its investments in the national railway system during the period 1992–2002. The importance of highways relative to railways also increased in India and the share of expenditures on railways in total transport infrastructure declined from 50% in the 1990s to 30% by the end of the 2000s (Indian Ministry of Railways). Today, roads are the most important transport mode in India, carrying 60% of the freight turnover compared to 31% for railways.[14] The highway projects undertaken in the two countries are therefore crucial parts of their transport strategies and of high importance for the development of the two countries. More recently there are also efforts to improve rail connections and environmental outcomes play an important role in these considerations as well. Interestingly, the newly proposed Indian dedicated freight corridors focus on improving the railway connections among the four economic centers that were also targeted by the GQ (World Bank, 2015).

3.4. Chinese roads in India

India currently faces substantial constraints due to insufficient transport infrastructure, which is less the case for China. Furthermore, China has experienced stronger spatial convergence. A natural question therefore is how India would develop if it invested in transport infrastructure like China. To answer this question, I propose a counterfactual road network for India that connects intermediate-sized cities in a way that approximately maximizes aggregate net income. I use a heuristic algorithm that balances the income gains from reduced trade costs in the general equilibrium framework against the road construction costs predicted by the topography. The next section presents the general equilibrium framework that is used to design the counterfactual network and to quantify and compare the effects of actual and counterfactual networks.

4. Conceptual framework

The setup is a general equilibrium trade model based on Donaldson and Hornbeck (2016). They derive from a version of the Eaton and Kortum (2002) model an expression for the impact of transport infrastructure on income.[15] That expression captures the “market access” of a location, which is the sum over trading partners’ income, discounted by the bilateral trade costs and by the market access of the trading partners. They use this framework to estimate the effect of the expansion of the American railway network on land prices. I estimate the effect of the Indian transport network on income and use the model to design the income-maximizing network (see Section 5).

[12] See the webpage of the National Highway Authority of India (https://nhai.gov.in/) for details on individual segments. The cost estimates are based on Ghani et al. (2016).

[13] The official speed limit was increased to 100 km/h in 2007, but the actual driving speed is significantly lower. This was derived by selecting a random sample of locations and exporting bilateral transport times with a routine from Google Maps.

[14] The share of highways in the total freight turnover is even higher in India than in China (as described in KPMG, 2013, Logistics games changers: Transforming India’s logistic industry).

[15] The presentation in this section focuses on the key aspects of the model. The details are discussed in Appendix A.

4.1. Trade between Indian districts

The basic setup is a trade model with the production factors land, labor, and capital. In the benchmark model, I take a long-term perspective and assume labor is mobile and real incomes per capita are equalized across locations. For the analysis of convergence during the period 2001–2012, I use a variation of the model with immobile labor.

The economy consists of many trading regions (i.e., Indian districts), where the origin of a trade flow is denoted by o and the destination by d. Each district produces varieties indexed by j with a Cobb–Douglas technology using land (L), labor (H), and capital (K), and an exogenous productivity shifter ($z_o(j)$) drawn from a Fréchet distribution as in Eaton and Kortum (2002).

Trade costs between locations o and d are modeled according to an "iceberg" assumption: for one unit of a good to arrive at its destination d, $\tau_{od} \geq 1$ units must be shipped from origin o. This implies that if a good is produced in location o and sold there at the price $p_{oo}(j)$, then it is sold in location d at the price $p_{od}(j) = \tau_{od} p_{oo}(j)$. With perfect competition, prices equal the marginal costs of producing each variety, $p_{oo}(j) = MC_o(j) = \frac{q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma}}{z_o(j)}$, where q_o is the land rental rate, w_o is the wage, r_o is the interest rate, and α and γ are the factor shares. Consumers have CES preferences and search for the cheapest price of each variety (including trade costs), such that prices in each district are governed by the productivity distribution across districts. Eaton and Kortum (2002) show that this implies a CES price index

$$P_d = \mu\Big(\sum_o \big[T_o\big(\tau_{od} q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma}\big)^{-\theta}\big]\Big)^{-\frac{1}{\theta}} \equiv CMA_d^{-\frac{1}{\theta}}, \tag{1}$$

where I follow Donaldson and Hornbeck (2016) and define the sum over origins' factor costs as "consumer market access" (CMA), because it measures district d's access to goods at low prices.[16]

4.1.1. Trade flows and gravity

With expenditure shares as in Eaton and Kortum (2002) and assuming that a district's total expenditure equals income ($X_d = Y_d$), we obtain the gravity equation

$$X_{od} = \underbrace{T_o\big(q_o^\alpha w_o^\gamma\big)^{-\theta}}_{\text{Origin's productivity and factor costs}} \times \underbrace{\tau_{od}^{-\theta}}_{\text{Trade costs}} \times \underbrace{Y_d}_{\text{Destination's income}} \times \underbrace{\kappa_1 CMA_d^{-1}}_{\text{Destination's CMA}}. \tag{2}$$

Trade from o to d depends positively on the origin's competitiveness (productivity) and the destination's income but negatively on the consumer market access of the destination and on the bilateral trade costs. This feature of a gravity equation is shared by a large class of models and has found strong support in the data.

4.1.2. Market access and income

Summing the gravity equation over destinations d yields total income of origin o,

$$Y_o = \sum_d X_{od} = \kappa_1 T_o(q_o^\alpha w_o^\gamma)^{-\theta} \sum_d \big[\tau_{od}^{-\theta} CMA_d^{-1} Y_d\big], \tag{3}$$

where Donaldson and Hornbeck (2016) define "firm market access" of district o as $FMA_o \equiv \sum_d \tau_{od}^{-\theta} CMA_d^{-1} Y_d$. I assume that trade costs are symmetric,[17] in which case a solution must satisfy $FMA_o = \rho CMA_o$ for $\rho > 0$. Donaldson and Hornbeck (2016) refer to this as "market access" (MA). In this setting, we then get

$$MA_o = \rho \sum_d \tau_{od}^{-\theta} MA_d^{-1} Y_d, \tag{4}$$

and Eq. (3) for income becomes

$$Y_o = \kappa_1 T_o(q_o^\alpha w_o^\gamma)^{-\theta} MA_o. \tag{5}$$

Eqs. (4) and (5) are two key model equations and they summarize how trade costs affect income. While Eq. (5) implies a relationship between income and market access, Eq. (4) shows that this market access measure is the channel through which transport costs affect income. An appealing property of the model is that it is a general equilibrium framework and thus allows quantifying aggregate effects.

[16] Using the fact that the rental rate for capital is equalized everywhere to $r_o = r$, we can define the constant $\kappa_1 \equiv \mu^{-\theta} r^{-(1-\alpha-\gamma)\theta}$ as described in the appendix.

[17] The travel times through the road network that I obtain from the shortest path algorithm are also symmetric because the driving speed on roads is assumed to be the same in both directions.

4.1.3. Labor mobility

In the benchmark, I assume that labor is fully mobile across districts, and thus real incomes per capita equalize. This provides a long-term perspective for the distribution of economic activity in the counterfactuals and network designs described below. The income of each district can then be written as

$$Y_o = \left(\kappa_1 T_o\right)^{\frac{1}{1+\theta\alpha}} \left(\frac{\alpha}{L_o}\right)^{\frac{-\theta\alpha}{1+\theta\alpha}} \left(\bar{U}\right)^{\frac{-\theta\gamma}{1+\theta\alpha}} \rho^{\frac{-\gamma}{1+\theta\alpha}} \left(MA_o\right)^{\frac{1+\gamma}{1+\theta\alpha}}, \quad (6)$$

where I used the Cobb–Douglas production function and $w_o = \bar{U} P_o$. An analogous expression can be derived for labor as a function of market access, which I discuss in the appendix. An alternative version of the model with immobile labor is discussed in the appendix.

5. Counterfactuals and network design

5.1. Counterfactuals

In the baseline model, I assume that labor is mobile. We can solve the model for labor as shown in Appendix A.6 and back out the unobserved constant terms for productivity and land from observed population and trade costs in the actual economy in 2012.

For given productivity and land, one can then solve the model using counterfactual trade costs. As described in Donaldson and Hornbeck (2016), the counterfactuals can be computed either assuming that (i) the total population adjusts and $\bar{U}$ remains constant, or (ii) that the total population is fixed (but population shares can change due to labor mobility) and $\bar{U}$ adjusts. I focus on the latter case, i.e., the counterfactuals are computed under the constraint that the total population implied by the model counterfactual equals the one in 2012. Consequently, the utility level (which equals real income) adjusts with trade costs.[18]

5.2. Network design

Using the framework above, we can compute counterfactual incomes for different trade costs (i.e., different highway networks). We can then optimize over highway networks to maximize income net of road construction costs.

I model this as a discrete choice of which links should get a modern highway of the same quality as the GQ. This fits the Indian and Chinese context well because both countries planned to build a new highway network with links of a given capacity of four to six-laned dual-carriageway highways and had to decide where these links should be placed (see Faber, 2014; Ghani et al., 2016). Finding the optimal network in this framework is challenging because there can be positive or negative complementarities between the highway links, and the number of potential combinations of links that can form a network is extremely large. While the problem of finding the globally optimal network in this discrete context has, to the best of my knowledge, not been solved, I propose a heuristic algorithm to find a (local) optimum, and I show statistically through simulations that the local solution is within narrow bounds of an estimate of the global maximum income.[19] The heuristic algorithm is an iterative procedure that starts from the full network and then removes and adds links sequentially based on their effects on net income until no further improvements are possible.[20]

5.2.1. Heuristic network design algorithm with single deviations

I assume that there is a national planner who has the objective of maximizing the national real income net of road construction costs. Given the assumption of the model, real income equals $\bar{U}$ and is constant across locations but varies over counterfactuals. The planner takes as given the productivity terms B_o, the geographic characteristics determining road construction costs, and pre-existing lower-quality roads, and then designs the network of highways of the same quality as the GQ.

We denote the net real income implied by a network G as $W(G) = Y^r(G) - b(G)$. $Y^r(G) = \bar{U}(G) \times Y_{Nat}$ is national real income, which is computed as the utility level scaled by the national GDP in the actual economy (2012 prices). Recall that the utility level corresponds to real income and is equalized across locations due to the assumption of mobile labor and that the utility level is set to 1 in the actual economy. $b(G)$ is the cost of the network construction.

The net effect of adding a link n given a starting network G is evaluated as

$$\Delta W(G, G_{-n}) = Y^r(G) - Y^r(G_{-n}) - b_n, \quad (7)$$

[18] Furthermore, the price of capital is computed as the income-weighted national average price index. Capital is assumed to be freely mobile across Indian districts and the international capital market. This requires choosing an appropriate price index for capital, for which I use a national weighted average. Since this depends on market access, the price of capital is endogenous and needs to be computed as part of the equilibrium (see also Appendix A.6.1). The price of capital could also be weighted differently, e.g., taking the price of one location, see Donaldson and Hornbeck (2016), for a discussion.

[19] Gastner (2005) and Gastner and Newman (2006) consider the problem of forming a network that connects facilities in an optimal way. In their case, the objective function is a weighted sum of the road construction costs and of total travel costs through the network, and they use a similar heuristic algorithm to search for the optimal network. I rely on a similar heuristic algorithm but with an objective function based on net income in the general equilibrium framework of Section 4. Note that Gastner and Newman (2006) do not have general equilibrium income effects because they compute the total travel costs rather than the implied aggregate income.

[20] Fajgelbaum and Schaal (2020) solve for the optimal network in a different general equilibrium trade model and show under which assumptions the planner problem is globally convex. They focus on the intensive margin and model road congestion, while I solve the discrete problem of which nodes should be directly connected by roads of a given quality, which was a key margin for the Indian government when planning its new national highway network.

where b_n is the construction cost of link n (which does not depend on the rest of the network). Starting from the full network, the algorithm first evaluates all links and then removes the link with the lowest net gain. In the second step, the algorithm evaluates all missing links that could be added, and adds the best link in case it leads to a positive net gain. The algorithm continues iterating in this way until no further local improvements are possible. Hence, this version of the algorithm is based on single deviations.

5.2.2. Heuristic network design algorithm with multiple simultaneous deviations

Given the discrete nature of the problem and the complementarity between highway links, there may be multiple local optima. The algorithm may get trapped in a local optimum if, for example, the solution contains a link n that delivers a positive net gain, but there is a potential link m (not currently in the network) that could replace n and deliver an even higher net gain. An algorithm that only evaluates single deviations will not find such improvements. Similarly, a current network might contain two links that are highly complementary to each other, and removing one individually reduces net income, but they could jointly be removed and replaced by a different link that increases net income more. Such dependencies are a key reason why this problem is difficult to solve. While considering all possible combinations is infeasible, it is possible to refine the algorithm to consider besides single deviations also multiple simultaneous deviations (e.g., checking if there are two links that could be replaced by one single link). However, this is computationally costly. I limit this multi-deviation search to after the single-deviation algorithm has converged to a local optimum. Furthermore, I only allow simultaneous addition and removal of up to two links each. When adding multiple links simultaneously, I only consider the k links that had the highest net return in the previous iteration; when simultaneously removing links, then I only consider the k links that had the lowest (negative) return in the previous iteration.[21] Allowing for multiple simultaneous deviations can improve the solution, but it still does not guarantee that the algorithm will converge to a global optimum because it is infeasible to check for all possible simultaneous deviations.

5.2.3. Convergence of heuristic algorithm

While the heuristic algorithm is not guaranteed to converge to the global optimum from an arbitrary starting network, there is a neighborhood around the global optimum from which it is guaranteed to converge. The reason is that for some given network G, the algorithm conducts an exhaustive search over all possible single deviations (i.e., adding or removing one link) and computes real net income in each case. If there are M nodes in the network, then there are $(M^2 - M)/2$ symmetric off-diagonal links to be added or dropped. This corresponds to the number of single deviations that are evaluated. Given that the algorithm conducts an exhaustive search over all possible single deviations, the minimum number of networks from which the algorithm is guaranteed to converge to the global optimum is also equal to $(M^2 - M)/2$. With multiple simultaneous deviations, this number is higher. When one draws random starting networks, and the number of draws approaches infinity, then the heuristic algorithm will eventually reach a network from which it converges to the global optimum. For a network with 68 nodes, as in the present case, it is not feasible to evaluate all possible network configurations. However, as I will show next, one can use statistical methods to obtain an estimate of the maximum income as well as statistical bounds around it.

5.2.4. Generalized extreme value distribution: Theory

When there are potentially many local optima, an obvious question is how we can find the global optimum. I will show below that one can use statistical methods from extreme value theory to obtain a point estimate of the maximum possible income as well as statistical bounds around it.

Let us denote the heuristic algorithm as F, starting networks as G_i, the locally optimal network that is found by F when starting from G_i as G_i^*, and the net income associated with G_i^* as W_i^*. Hence, $G_i^* = F(G_i)$ and $W_i^* = W(G_i^*)$. G_i are random symmetric adjacency matrices (without self-loops) of random size between the empty and the full network. Suppose we draw a large number of G_i and compute W_i^*. We now want to obtain an estimate of the population maximum from the observed sample maxima. Based on the sample of maxima, we can apply statistical methods to estimate the maximum. The extremal type theorem (Coles, 2001) states that if the distribution of maxima converges to a non-degenerate distribution, then it converges to one of the three distributions in the family of the Generalized Extreme Value (GEV) distributions.[22] I briefly discuss the conditions under which this applies. First, the distribution must be continuous. Although this is not exactly the case with this discrete problem, a continuous distribution is a good approximation in this case because there are a large number of local optima with distinct income levels that mimic a continuous distribution, as we will see below. Second, the draws must be independent, which would be violated, for example, if many draws converge to the same value such that there is clustering. This is not the case here either because almost every draw of a random starting network yields a slightly different network and, thus, a slightly different income. Third, the draws must be from the same population, which is the case here as well. Fourth, it must be possible to rescale the sequence of maxima such that the distribution converges to a non-degenerate distribution function, which will be the case in the estimation shown below. Finally, the estimation using maximum likelihood imposes certain regularity conditions on the parameter estimates, which are satisfied here.

[21] I set $k = 70$ in the multiple deviations case. In the single deviation case, I set $k = 1$. It could also be set to $k = 0$, but since the ranking of links is already available from the previous iteration, the computational burden with $k = 1$ is negligible. However, with $k = 70$, the computational burden is large, and the algorithm spends most of the time checking whether there are multiple simultaneous deviations that increase net income.

[22] Note the similarity with the central limit theorem. The central limit theorem essentially states that when we draw a large enough sample from a population, the distribution of sample means converges to the normal distribution, regardless of the population distribution. Similarly, the extremal types theorem states that when we draw a large enough sample of maxima from a population, then, if the distribution of the sample maxima converges, then it converges to one of the three extremal types distributions (Gumbel, Fréchet and Weibull), regardless of the distribution of the population.

5.2.5. Generalized extreme value distribution: Applications in the literature on combinatorial optimization problems

The literature on combinatorial optimization problems contains several applications of extreme value estimation, but to the best of my knowledge, not in the context of a general equilibrium trade model. Los and Lardinois (1982) use a heuristic algorithm to solve a transportation network problem. They also draw random starting points to generate a distribution of local optima and apply extreme value theory to obtain an estimate of the global maximum and determine when to stop sampling. They find that this method works well for heuristic algorithms that are of intermediate sophistication because they generate a sufficiently large number of local optima to be able to estimate an extreme value distribution but find good enough local optima such that the global maximum can be estimated precisely. Among the two versions of the heuristic algorithms described above, the one with single deviations is better suited for this purpose because, with multiple simultaneous deviations, it often converges to the same optimum. Loosely speaking, the heuristic algorithm must allow us to sample enough observations in the right tail of the distribution in order to estimate the maximum, but it does not have to be the best possible algorithm.

Giddings et al. (2014) provide a review and critique of the literature using statistical methods to estimate the maximum in optimization problems. They point out that, for example, when a heuristic algorithm delivers the same solution for multiple starting points, then the assumption of independent draws is not satisfied. Similarly, the assumption of a continuous distribution is not satisfied in combinatorial problems when the number of draws relative to the number of unique local optima becomes large. Carling and Meng (2015) empirically assess the concerns raised by Giddings et al. (2014) by considering a large set of optimization problems for which the optimum is known. They find that the statistical bounds are reliable if the heuristic solutions are good enough. Cicirello and Smith (2004) estimate extreme value distributions to choose among different search heuristics and find that it improves performance in different benchmarks.

As the simulation below will show, the optimization problem and the heuristic algorithm are suitable for using statistical techniques to estimate the maximum.

5.2.6. Estimating the maximum income using the generalized extreme value distribution

To estimate the maximum income and obtain the statistical bounds, I generate more than 13,000 random networks by creating random symmetric adjacency matrices without self-loops. The networks can be of arbitrary size between the empty networks and the full networks. I then use these random graphs as initial networks in the algorithm, compute the optimal network implied by each of them, and compute the net income for each network solution.[23] 99% of the simulations are based on the algorithm with single deviations, which is much faster and yields a large number of distinct local optima (in 94% of the cases), which has the advantages described above. The estimation also includes 100 simulations with the multi-deviation algorithm, which will be discussed in more detail below.[24]

Using this sample of net incomes, I then fit a Generalized Extreme Value Distribution (GEV) to obtain an estimate of the maximum net income that is generated by such a network, as well as confidence bounds. The fit of the distribution is good and is consistent with a Weibull distribution (see also the diagnostic plots in Figure A2). The regularity conditions on the parameter estimates are satisfied.[25]

The results of the estimation based on the simulation are shown in Fig. 2. The key insight is that the solution from the full-start network is close to the upper bound of the confidence interval around the estimated maximum. More precisely, the full-start solution when using multiple simultaneous deviations is less than 0.01% below the upper bound of the confidence interval. This difference is not economically meaningful compared to the size of the income gains and to other sources of uncertainty such as parameter calibration. Since the confidence interval is narrow and the full-start solution is close to the upper bound, it is statistically unlikely that the global optimum is significantly higher than the full-start solution.

There are a number of other interesting observations from the figure. First, the differences overall are not large. The heuristic algorithm that starts from the empty network and only evaluates single deviations is 0.3% below the benchmark (the algorithm starting from the full network and allowing multiple simultaneous deviations). The solution can be improved substantially when allowing multiple simultaneous deviations. In that case, income is 0.1% lower than the benchmark, which is already a fairly small difference considering other sources of uncertainty. Second, the solution of the algorithm, when starting from the full network is already relatively good even with single deviations, and the difference to the maximum is economically small (0.02% below the benchmark). Third, the estimated maximum is practically identical to the benchmark (difference less than 0.01%). Fourth, the 99% confidence bounds are very narrow, and the upper bound is close to the benchmark (difference less than 0.01%). Fifth, the observed maximum from the 13,000 simulations is practically identical to the benchmark (difference less than 0.01%). There are six random-start solutions that are better than the full-start solution, but only marginally (see also Fig. 4). Overall, the simulation shows that a GEV can fit the distribution of maxima well, and the full-start solution is close to the upper bound of the confidence interval around the estimated maximum, showing that it is unlikely that the global maximum is significantly above the full-start solution.

[23] For the simulation, I use the observed data for India and parameters discussed in Sections 6 and 8.

[24] The estimated maximum is almost identical when these 100 simulations with multiple simultaneous deviations are included (difference of less than 0.01%).

[25] 6% of the solutions are not unique and are excluded from the estimation. The estimation is based on Jalbert et al. (2024) and uses the block-maxima method. I set the length of the return period equal to the size of the sample. However, the estimated maximum is similar when using the exceedance over threshold method with a threshold equal to the 95th percentile. The results are also similar when using larger return periods.

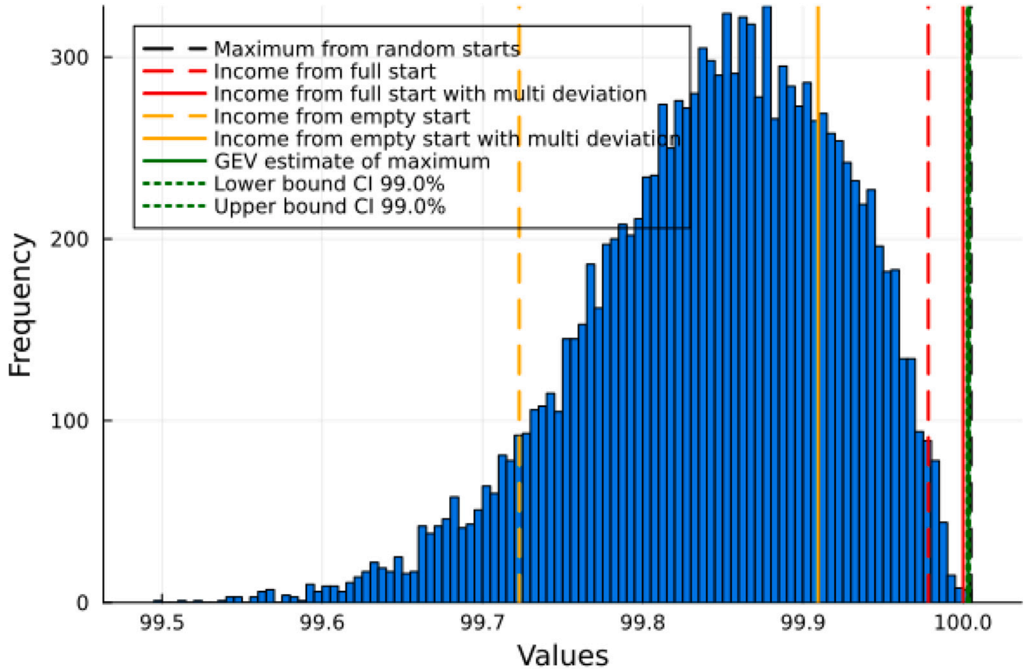

Fig. 2. Distribution of returns and estimated maximum.
The blue bars show the histogram of net real incomes from 13 100 simulations with random start networks. The vertical lines represent the solution when starting from the full or the empty network with single- or multi-deviations, the estimated maximum based on a fitted Generalized Extreme Value Distribution, and the observed maximum in the simulation. All values are scaled relative to the net income of the solution when starting from the full network and allowing simultaneous multi-deviations of four links, which is normalized to 100.

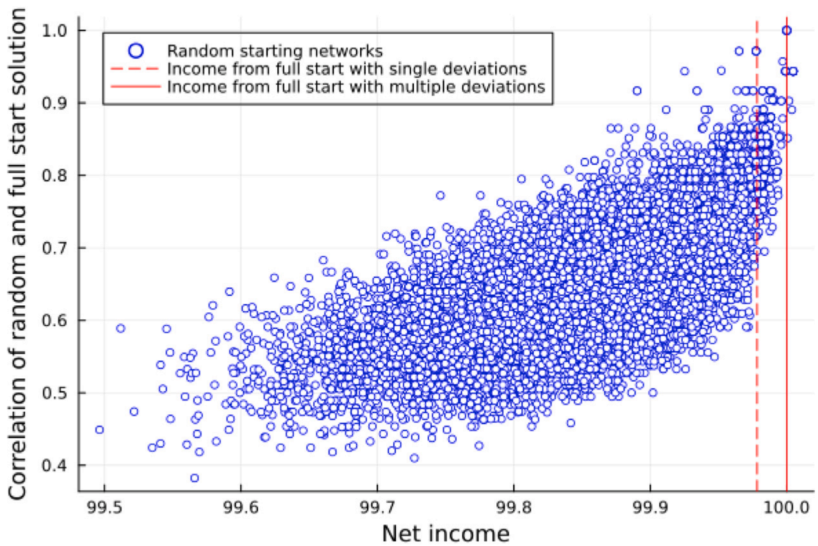

Fig. 3. Correlation of each random-start network with the full-start solution (Jaccard distance).
Each circle corresponds to a solution of the network design algorithm. The x-axis is the net income of this solution. The y-axis is the correlation (1-Jaccard distance) between the solution from the random start and the full-start solution. The horizontal line shows the income of the network when starting from the full network and using single- or multi-deviations. All incomes are relative to the full-start solution, which is normalized to 100.

5.2.7. Comparison of networks from different local optima

As shown above, the solutions when starting from random networks are within a relatively narrow range, and the solution when starting from the full network is practically identical to the estimated maximum and to the observed maximum across all simulations. We now assess how similar the networks that yield these incomes are. Figure A1 shows four different networks that correspond to

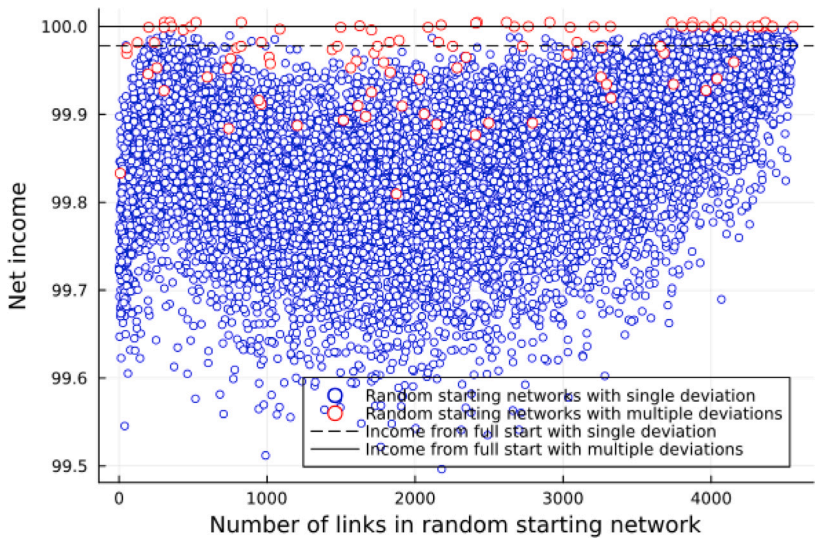

Fig. 4. Size of random start network and income of the resulting locally optimal network.
Each circle corresponds to a solution of the network design algorithm. The *y*-axis is the net income of this solution. The *x*-axis is the size of the corresponding random start network. The horizontal line shows the income of the network when starting from the full network with multiple deviations or single deviations.

the full-start solution, the best two random-start solution, and the worst random-start solution, respectively. The first three networks show only very minor differences. The worst network differs a bit more but still has a similar structure. The difference in net income between the best and the worst networks is 0.5%.

To assess the similarities across all networks, Fig. 3 shows the correlations between the links in each random-start network with the full-start network (vertical axis) and associates this with the income of each network (horizontal axis). The figure shows a clear positive correlation between income implied by a network and its similarity with the full-start solution.

5.2.8. Size of the random start network

Fig. 4 shows the income implied by each random-start solution (vertical axis) against the size of the start network (horizontal axis, count of the number of links). The red circles are those random-start networks for which multiple simultaneous deviations are used (Section 5.2.2). The figure conveys two important messages. First, the red circles tend to be higher up, which shows that allowing for multiple simultaneous deviations leads to higher incomes on average and avoids being trapped in a relatively low-income local optimum. Second, and more surprisingly, random-start solutions tend to have higher incomes if the random network that was used as a start was larger. In particular, when the start network is very large, then it is rare to converge to a network that is more than 0.1% below the full-start solution with multiple deviations. This result shows that it is not a coincidence that the full-start solution yields a higher income than most other starting networks because larger start networks systematically lead to a higher income in the resulting local optimum. A possible explanation is that when networks are sparse, small deviations are more likely to eventually lead to larger differences in incomes in the local optimum that is reached. Intuitively, path dependence might be stronger in sparser networks. Overall, the results suggest that starting the heuristic network design algorithm from the full network is a safe option in the sense that the solution is close to the estimated maximum. Furthermore, it is also computationally more efficient than evaluating a large number of random start networks, which will be discussed next.

5.2.9. Algorithm speed

The network design algorithm is sufficiently fast to start from the full network and to simulate many different random networks. Starting from the full network and only allowing single deviations, it takes about 15 minutes to compute the optimal network on one computing node with 48 cores using multi-threading in Julia. When allowing up to 4 simultaneous deviations and, in each case, considering 70 links, then it takes about 70 minutes.[26] Starting from a random network takes, on average, about 4 minutes with single deviations. Given that for each random starting network, the chance of obtaining a solution that is better than the full-start solution is small, one would need to draw very many random start networks. Alternatively, one may draw a smaller number of random starting networks but allow for multiple simultaneous deviations. However, both alternatives are computationally more costly than starting once from the full-start network, even when allowing for multiple simultaneous deviations.

Recall that in each iteration of the algorithm, it checks whether it can add or remove a link to increase net income, such that it needs to solve the model $(N^2 - N)/2$ times per iteration (which is the number of symmetric links excluding self-loops). The

[26] These times should be considered estimates and not exact benchmarking because the algorithm was run on a computing cluster with a batch processing scheduler that may assign jobs to different nodes that do not necessarily have the exact same hardware specification.

algorithm makes extensive use of multi-threading since, for a given network, every single deviation can be evaluated in parallel, but the time it takes to compute the optimal network increases rapidly in the size of the network. However, it is feasible to compute the optimum for larger networks. For example, when doubling the number of nodes from 68 to 136, then it takes about 2 hours when using single deviations.[27]

6. Data and estimation

The network design approach discussed in Section 5 can now be applied to study how India would have developed if it had optimally implemented the Chinese strategy of connecting intermediate-sized cities. To this aim, we first take the model from Section 4 to the data by estimating one of the key relationships in the model, namely the effect of market access on income in Eq. (6).

The data required for the estimation of Eq. (6) and for the counterfactual analyses are the income and population of each location and bilateral trade costs. I focus on the period from 2002 to 2012 and estimate the equation in differences. I have data on 636 mainland Indian districts, but there are a few missing income observations such that the estimation is based on 631 districts. The data and estimation are discussed below, and further details on the data and descriptive statistics can be found in Appendix B.

6.1. Data

I rely on geo-coded data on income and road infrastructure over time. For the construction of the counterfactual highway networks, I additionally need data on the topography in order to predict road construction costs.

6.1.1. Income, population, and night-time lights

I use Indian districts with administrative boundaries as of 2011. Estimates for GDP for the years 2001–2013 are available from Nielsen, a commercial data provider. Nielsen constructs the district-level GDP estimates by combining various sources such as the Annual Survey of Industries, the National Sample Survey, and the Economic Census. I use these data as the baseline in the empirical analysis and the quantitative analysis based on the model. The first and last three years are averaged before computing the long differences, and I refer to the start and end years as 2002 to 2012. For the population data, I use the 2001 and 2011 censuses. I also use light at night for the comparison to China at the cell level. Growth in light at night measured by weather satellites has been shown to be a good proxy for income growth in certain situations (Henderson et al., 2012; Chen and Nordhaus, 2011). Two important advantages of the light data are that it has a high spatial resolution and is independent of countries' statistical procedures, making it well-suited for the analysis of spatial convergence.

6.1.2. Trade costs based on road network

Transport infrastructure affects economic activity in several dimensions, such as the time it takes to move goods and people, pecuniary costs from tolls, or risks associated with the use of inadequate or overused infrastructure. I will focus on transport times as a determinant of transport costs. Higher road quality, limited access, and more capacity are all reflected in the time it takes to move goods between two locations on a new or upgraded road.

The counterfactual analysis requires information on the transport times between all pairs of Indian districts for different versions of (actual and counterfactual) transport networks. While the transport times on the current network could be derived from automated searches on applications like Google Maps, this is not the case for past or counterfactual networks. The approach used here is to model the network using GIS and then apply an algorithm that finds the shortest path (in terms of transport time) between any two locations on the digitized road network. The advantage of this approach is that the same algorithm can compute all bilateral transport times for different road networks. The required inputs are the geographically referenced roads and the driving speed on different types of roads. I take the driving speeds on different types of existing roads from surveys conducted by the World Bank. These surveys suggest that the average driving speed on a conventional highway is about 35 km/h (see also Section 3). For the driving speed on the GQ, I use 75 km/h, which is an average that I obtain from automated searches on routes along the GQ using Google Maps. The travel map consists of pre-existing lower-quality roads and conventional highways as well as the GQ. The data sources and a more detailed description of the data preparation are provided in Appendix B.4.

With these inputs, it is possible to construct a grid of India where the value of each 1×1 km cell represents the speed of traveling through this cell. Such a grid of transport costs is shown in Figure A3 in the appendix. I use the fast marching algorithm as in Allen and Arkolakis (2014) and Allen and Atkin (2022) to compute bilateral travel times through the given road map.[28] The algorithm calculates the cheapest way to travel from one location (district centroids, represented by dots in Figure A3) to another location. Depending on the road infrastructure and thus on the transport costs in each cell, the cheapest path may not be the shortest in terms of distance. More importantly, the transport times associated with the cheapest path change when the infrastructure is improved, thus generating time variation in transport costs.

[27] For these results, I generated random economies with different numbers of nodes. I also repeated the analysis from Figs. 3 and 4, although with much smaller sample sizes, given that it takes longer to compute one simulation. The results are similar and show that the solution quality increases with the size of the starting network and that solutions with higher incomes have a higher correlation with the network from the full-start solution. Hence, these simulations support the finding that starting from the full network is a suitable way to find the optimal network.

[28] The code for the fast marching algorithm is from the "accurate fast marching" Matlab toolbox by Dirk-Jan Kroon. Note that in the network design algorithm, I model the transport network as a graph and use the Floyd-Warshall algorithm to compute the shortest paths using the Graphs.jl package.

Following Roberts et al. (2012), I assume that there are economies of scale in transport, such that transport costs increase less than proportionally in transport times.[29] More precisely, I calculate iceberg trade costs between an origin o and a destination d as

$$TradeCosts_{od} = 1 + (\gamma_T \times TransportTime_{od})^{0.8}. \tag{8}$$

γ_T is chosen such that the median iceberg trade cost is 1.25 for the network without the GQ.[30] The transport costs within districts are set to 1.

Although the analysis undertaken here captures a key aspect of the modern transport infrastructure in India and China, some caveats must be pointed out. The first concern is the omission of other types of domestic transport infrastructure such as railways or urban transport systems such as subways. Second, access to international markets via seaports or airports is not modeled as part of the general equilibrium framework, but one can add international trade as additional income in districts with major ports as in Donaldson and Hornbeck (2016). Third, villages' access to the transport infrastructure via rural roads is not considered since I focus on major roads. Finally, non-transport infrastructure such as electricity and water also affect economic development. These caveats would limit the validity of the exercise here only if the omitted factors were time-varying at the district level and correlated with the explanatory variable market access. Section 7.2 discusses how I address this with my empirical strategy.

6.1.3. *Road construction costs based on topography*

In order to construct the counterfactual networks that connect the cities that would be targeted by the Chinese policy, one first needs to obtain a measure for road construction costs on the Indian terrain. I follow Faber (2014) and assume that the construction costs on a given 1×1 km cell of land depends on the slope and the share of water and built-up area in the following way:

$$ConstructionCosts_c = 1 + Slope + 25 \times Builtup + 25 \times Water. \tag{9}$$

$Slope$ is measured in percent and $Builtup$ and $Water$ are binary indicators that take the unit value if the majority of the cell is built up or water, respectively.[31] Applying this formula using detailed terrain data produces a 1×1 km grid of construction costs for the entire Indian landscape. Given this grid of construction costs, one can in a second step apply the Dijkstra (shortest path) algorithm to find the cheapest connection between any two given points through the cost grid and back out the corresponding path.[32] The procedure is illustrated in Figure A4 in the appendix, where the cells represent different construction costs (based on Eq. (9)) and the lines are the least-cost paths to connect the locations (shown as circles). For each resulting counterfactual network, one can calculate the total road construction costs based on the topography. An important feature of this setting is that the total road construction costs based on the topography can be calibrated to the Indian context using the actual cost of the GQ in USD.[33] Hence, only the *relative* weights of the different components in Eq. (9) matter.

7. Estimation

7.1. *Empirical strategy*

Estimating Eq. (6) in a cross-section would require to control for relevant district characteristics, which are difficult to obtain. Therefore, Eq. (6) will be estimated with a fixed effect panel regression, or equivalently in differences between the two periods, that relies on the time variation within districts. This allows accounting for the unobserved heterogeneity across districts. Eq. (10) shows the different components of Eq. (6) over time:

$$\ln(Y_{o,t}) = \underbrace{-\frac{\theta\alpha}{1+\theta\alpha}\ln\left(\frac{\alpha}{L_o}\right)}_{\text{Constant over time}} + \underbrace{\frac{1}{1+\theta\alpha}\ln(\kappa_{1,t}) - \frac{\gamma\theta}{1+\theta\alpha}\bar{U}}_{\text{Constant over locations}} + \underbrace{\frac{1}{1+\theta\alpha}\ln(T_{o,t})}_{\text{Productivity}} + \underbrace{\frac{1+\gamma}{1+\theta\alpha}\ln(MA_{o,t})}_{\text{Market access}}. \tag{10}$$

The first term on the right-hand side in Eq. (10) collects parameters and land endowments, which are assumed to be constant over time and thus absorbed by the district fixed effects. The second term includes country characteristics (the interest rate inside of κ and the common utility level) that are absorbed by state-year fixed effects or a constant. The next term is the productivity of each district, $T_{o,t}$, which can potentially vary over time and districts. Since productivity is unobserved, there could be endogeneity concerns, but I will argue below that my identification strategy uses exogenous variation in transport infrastructure such that there

---

[29] This is a common assumption, see for example also Au and Henderson (2006) who assume that transport costs increase less than proportionally in *distance*. The value of 0.8 is close to what Roberts et al. (2012) obtain for the rural sector in their analysis of the Chinese NEN. For the city-level data, see Section C.3.

[30] This is calculated based on the median distance to be traveled through the Indian road network and the average cost per kilometer based on evidence in Limao and Venables (2001). See also Baum-Snow et al. (2020) for a similar calculation for China.

[31] The implication of this formulation is that a 25 percentage points increase in slope raises the road construction costs in the same way as when the road has to be built through an area with existing houses, other infrastructure, or water. Different from Faber (2014), my formulation does not include wetlands.

[32] This algorithm is implemented in the ArcGIS Network Analyst extension. The algorithm has already been widely used in the economics literature, for example in Dell (2015), Faber (2014), Donaldson and Hornbeck (2016), and Donaldson (2018).

[33] I build the GQ on the road construction cost surface based on Eq. (9) to obtain the total costs based on the topography. The ratio to the cost of the GQ in USD (from Collier et al., 2016, or Ghani et al., 2016) then allows me to scale the road construction costs of all counterfactual networks accordingly. The advantage of this approach is that it reflects the actual cost of road construction in India, but it relies on the assumption that (conditional on the topography) the cost of the GQ can be applied to other areas within the country. In my baseline, I use the per-kilometer cost in Collier et al. (2016) and the length of the planned Indian highway network to compute the total cost.

is no effect of unobserved productivity changes on market access. The last term in Eq. (10) shows the effect of market access, which is computed in Eq. (4) for different transport infrastructures.

The model predicts a constant elasticity of income with respect to market access,

$$\beta = \frac{1+\gamma}{1+\theta\alpha}. \tag{11}$$

The elasticity predicted by the model for assumed parameter values, $\theta = 8$ (trade elasticity), $\gamma = 0.5$ (labor share), and $\alpha = 0.18$ (land share), is around 0.6. As we will see below, this value is close to the estimate obtained in the empirical analysis.

7.2. Identification

Identifying the causal effect of infrastructure on income is challenging for several reasons. First, the choice of where to build infrastructure is not exogenous. In particular, the GQ had the explicit goal of connecting the four largest economic centers. This raises the concern that infrastructure may have been built where high growth was expected. But the clear objective of the GQ also poses an advantage for identification. By connecting the four largest centers, it affected districts that happened to be in between two important cities. By excluding the nodes of the network, it is, therefore, possible to exploit plausibly exogenous variation in transport infrastructure in districts that were accidentally affected by the GQ. This identification strategy was proposed by Chandra and Thompson (2000) for the U.S. interstate highway system. I follow this strategy and exclude the nodal cities and the corresponding districts.[34] One caveat with this approach is that the planners may have designed the network such that it also connects other cities on the way. For example, the GQ makes a detour to connect Bangalore, even though it was not explicitly among the four targeted cities. I show that excluding Bangalore does not substantially change the results. This does not rule out that other cities on the way were targeted, e.g., because of their future growth prospects, but I also show in the robustness section that there is no evidence that the GQ targeted locations that were already growing more before the construction.

A second challenge to identification is that shocks to income may be spatially correlated. Since the market access of o sums over incomes of trading partners d and a spatially correlated income shock may affect both o and d, changes in market access over time are likely to be correlated with o's own income. Therefore, an observed correlation between income and market access can arise even if there was no change in trade costs. To address this, I instrument the market access measure from Eq. (4) with a measure where I hold income fixed in the initial year, hence only exploiting the variation due to changes in transport infrastructure (and thus bilateral trade costs). Eq. (12) shows this version of the market access equation,[35]

$$MA_{o,t} = \rho \sum_{d} \tau_{od,t}^{-\theta} MA_{d,t}^{-1} Y_{d,2000}. \tag{12}$$

A third challenge is that there may have been other policies that affected the return to the GQ, such as complementary infrastructure investments. For example, an area that was newly connected by the GQ might find it worthwhile to then build additional roads connecting to the GQ, but they might not build these roads without the GQ. The same might apply to ports, railways, or airports. This is a potential threat to identification if these policies are targeted toward areas that received better connectivity due to the GQ during the same time period.

7.3. Estimate the elasticity of income with respect to market access

The estimation of β is based on Eq. (10), suggesting a log-linear relationship between real income and market access. I estimate the equation in differences to account for the unobserved time-invariant district characteristics. Furthermore, I add geographic controls and state fixed effects as shown below.

The first column of Table 1 shows OLS results. The t values with the wild bootstrap method are in parentheses and allow for clustering at the state level because the errors could be correlated within this administrative unit, for example, due to state-wide policies or other shocks. The OLS estimate has approximately the expected magnitude and is estimated relatively precisely. The second column then applies the two key aspects of the identification strategy. First, it excludes the 4 targeted districts and only exploits the variation in other districts that were not directly targeted by the roads connecting the four largest centers. Second, it instruments growth in market access (which can be due to changes in trade costs and income) with growth in market access with fixed income. This accounts for the fact that districts' incomes depend on each other through the general equilibrium relationships. The estimate is somewhat lower but roughly of the same magnitude, which suggests that the observed correlation between market access and income is not driven by the endogenous location of infrastructure in the four nodes.[36] In column 3, I exclude a fifth node,

[34] Redding and Turner (2015) provide a discussion of the identification issues related to transport infrastructure and they refer to this strategy as the 'inconsequential places approach'. There are several variations of this approach. Michaels (2008) studies the U.S. highway network and uses the orientation to the next large city as an instrument. Banerjee et al. (2020), Ghani et al. (2016), Asturias et al. (2019), and Khanna (2016) study the effects of transport infrastructure in China and India and they use the straight line between the nodes as an instrument. I use the actual path of the GQ and exclude the targeted nodes.

[35] Donaldson and Hornbeck (2016) also consider a market access measure where they hold population constant in the initial year to estimate the effect of market access on land values.

[36] Further evidence against the concern that the location of transport infrastructure is driven by economic performance is provided in the robustness section. There, I show that changes in market access due to the construction of the GQ are not significantly correlated with districts' growth trends prior to the start of the NHDP.

Table 1
Elasticity of income with respect to market access.

	(1)	(2)	(3)	(4)	(5)
Log market access	0.760	0.600	0.585	0.584	0.501
	(3.16)	(2.57)	(2.43)	(1.81)	(1.60)
Estimation	OLS	2SLS	2SLS	2SLS	2SLS
Excluded target nodes	0	4	5	5	5
Geog. controls	No	No	No	Yes	Yes
State FE	No	No	No	No	Yes
N	631	620	619	619	619
Rsq.	0.054	0.049	0.047	0.085	0.339

The table shows estimates of the elasticity of nominal GDP with respect to market access. The dependent variable is the log difference in GDP in each district between the years 2002 and 2012, where the start and end periods are averaged over three years. The explanatory variable is the log difference in market access between 2002 and 2012 computed based on Eq. (4) and instrumented with the market access with constant GDP in Eq. (12). Observations are weighted by the log of 2002 income. Standard errors are clustered at the state level and estimated with the wild bootstrap method (Roodman et al., 2019), the corresponding t values are in parentheses.

Bangalore, that also appears to have been targeted — although it was not explicitly stated in the objective. Columns 4 and 5 then include geographic controls (log distance from the coast, longitude, latitudes, and their squares and interactions) and state fixed effects. The elasticity is estimated less precisely, but the magnitude remains similar. Furthermore, the magnitude of the elasticity is in line with the theoretical prediction of the model when assuming reasonable parameter values.

In the next section, I use the full structure of the model and parameter values that are consistent with the above estimation to quantify the aggregate and distributional effects of the GQ and counterfactual networks. Section 9 discusses the robustness to growth prior to the GQ, different values of the trade elasticity, and population growth.

8. Aggregate and distributional effects of transport infrastructure

In the previous section, I estimated the effect of transport infrastructure on economic activity through the channel of market access and found that the estimate is in line with the model's prediction for reasonable parameter values. I now use the full structure of the model to predict how much each location's income changes for factual and counterfactual changes to the network. Because the framework captures general equilibrium effects of transport infrastructure, this allows for analyzing the aggregate and distributional consequences. The results are derived from solving the model for each version of the trade costs implied by the actual and counterfactual transport networks and comparing the resulting income to the income levels implied by the actual GQ network (see Section 5). Given the assumption of freely mobile labor, utility is equalized across locations in each equilibrium, but the total real income of each location varies due to changes in trade costs and adjustments in population and capital.

8.1. Actual network

I evaluate the effect of the GQ by constructing a transport network in 2012 without the GQ (i.e., with only conventional highways but with productivities implied by 2012 data) and then comparing the predicted income to the actual income. I first discuss the aggregate effects and then consider the distributional implications.

8.1.1. Aggregate effect of GQ

With trade costs implied by a network without the GQ, the model suggests that aggregate real income in 2012 would be 1.1% lower if the GQ had not been built (see Table 2).[37]

This loss in income needs to be compared to the savings in construction and maintenance costs when the GQ is not built. The total budget of the GQ and NS-EW (phase 1) amounted to USD 7 billion (Ghani et al., 2016). The GQ accounts for 78% of the total length of phase 1, which would correspond to USD 5.46 billion (in 1999 prices, without interests). I do not have data on actual expenditures, which may be higher. An alternative way to estimate the costs is to use the average cost per km from other sources. Collier et al. (2016) provide such estimates for new 4-lane expressways (see their Table S.2), and using these estimates implies a cost of USD 16.6 billion (2000 prices). I use this more conservative estimate to compute net income.

In order to quantify the annual cost, I need to assume a cost of capital and maintenance. I use a cost of capital of 5% for all counterfactual comparisons, which is a conservative assumption since during the past decade the government's cost of capital was lower. Furthermore, I assume maintenance costs of 12% of the construction costs, which is approximately what Allen and Arkolakis (2014) report for the U.S. interstate highway network. Based on these assumptions, the annual net effect on real income when removing the GQ implies a loss of around 0.6% per year (see the last column in Table 2). These effects are large but reasonable given that the median trade cost between the 68 nodes fell by 1.1% due to the construction of the GQ.

[37] For comparability with the network designs that connect the 68 intermediate-sized cities as shown in Section 8.2, this income loss due to the removal of the GQ is computed based on an economy consisting of the 68 nodes.

Table 2
Aggregate effects of actual and counterfactual transport networks.

	Cost (%)	Income (%)	Net income (%)
Removing GQ	−0.59	−1.14	−0.55
Net income-maximizing	1.56	5.83	4.27
Net income-maximizing, NHDP budget	0.51	3.68	3.17

The table summarizes the aggregate effects of the actual and counterfactual networks in percent changes. The changes in construction costs, income, and net income due to each network are shown in percentages of GDP with the actual network (i.e., with the GQ). Annual costs are based on 5% cost of capital and 12% maintenance costs. The first row shows the effect of removing the actual network (GQ). The counterfactual networks in the second and third rows are assumed to replace the GQ.

8.1.2. Distributional effects of GQ

China and India have experienced different regional development patterns. In particular, India had less spatial convergence and some "lagging regions" (Chaudhuri and Ravallion, 2006). From this perspective, an important question is how transport infrastructure may contribute to these differences in regional development. One advantage of the general equilibrium approach used here is that the effects of different transport networks can be assessed both at the aggregate and at the local level, allowing to analyze the distributional consequences and the regional development patterns with each transport network. As will be shown below, the effects on the local development of Indian districts differ over the various versions of the transport networks.

Figure A6 in the appendix shows the effects of the GQ at the level of Indian districts.[38] The numbers represent the percentage change in population from building the GQ relative to the network without the GQ. As expected, the effects are strongest along the path of the newly built or upgraded highways, and there is considerable localized variation in the changes across districts.[39]

8.2. Income-maximizing counterfactual network

The counterfactual exercise asks how India would develop if it had built a highway network among its intermediate-sized cities and connected them in an approximately optimal way. I first identify the Indian cities that would have been chosen by the Chinese policy. 68 Indian cities fulfill one of the two criteria, i.e., having a population above 500,000 or being a state capital. The locations of these cities are shown in Figure A5 in the appendix. I then design counterfactual networks among the targeted cities based on the general equilibrium model and the network design algorithm described in Section 5.2.

Benchmark. The solution of the network design algorithm in Section 5.2 is shown in Fig. 5(a). As we can see, this is a substantially larger network than the GQ, and it would cost more than four times as much. The network reaches parts of the countries that the GQ, which connected the four largest economic centers, did not reach. Furthermore, the network has a star-shape in the center, and its structure is very different from the GQ. Despite the higher costs, the counterfactual network delivers sizable aggregate welfare gains compared to the actual network, i.e., compared to the GQ, and would increase real income net of construction and maintenance cost by 4.3% (see Table 2). Furthermore, it has different distributional implications, which will be discussed below.

The effect of the income-maximizing network on aggregate real income net of road construction costs is, like the effect of the GQ, large but again reasonable. The median trade cost reduction between the 68 nodes due to the counterfactual network is 4.4% in comparison to the GQ. The aggregate effect is also in line with evidence from the Chinese NEN, for which Roberts et al. (2012) find an aggregate real effect (before costs) of 6%.

Alternative network designs. The structure of the network is similar when using the heuristic network design algorithm with only single deviations, which is shown in Fig. 5(b). However, real net income would be slightly lower (by 0.02%, see Fig. 2).

The baseline network design assumes that there is no budget constraint, and the government can borrow at a constant world interest rate. The optimal network is indeed substantially larger and costs more than four times as much as the GQ. Furthermore, it also costs more than the planned budget for phases 1 and 2 of the NHDP, which include additional highways besides the GQ but were not completed by the end of my sample period. The total length of planned highways in phases 1 and 2 is 2.1 times the length of the GQ. To assess the implications of setting a budget constraint, I take the total cost of phases 1 and 2 as a budget constraint and use the network design algorithm to find the best solution given this constraint. The result is shown in Fig. 5(c). As can be seen in the figure, the budget-constrained optimal network consists of a subset of the unconstrained optimal network. It includes three of the four edges of the GQ (it is missing the link in the South-East from Kolkata to Chennai) and a cross through the center. There are two important differences between this counterfactual and the actually planned highways during phases 1 and 2 (see Ghani et al., 2016, for a map). The first difference is the missing link from Kolkata to Chennai. The second difference is that the budget-constrained counterfactual network has, like the unconstrained income-maximizing network, a central node in Nagpur, while the East-West corridor in the actual plans crosses further to the north. In the aggregate, this budget-constrained network would still increase national income by 3.2% (see Table 2).

38 The counterfactual district-level impacts are computed based on the model with mobile labor. Since real income per capita is equalized across locations, I show the effect on population. This illustrates the differences across districts, but the aggregate effect varies across counterfactuals.

39 Table A6 in the Appendix estimates convergence regressions for different network designs under the assumption of immobile labor and compares the convergence between India and China. The findings suggest that the GQ reduced convergence and that convergence is generally lower in India than in China.

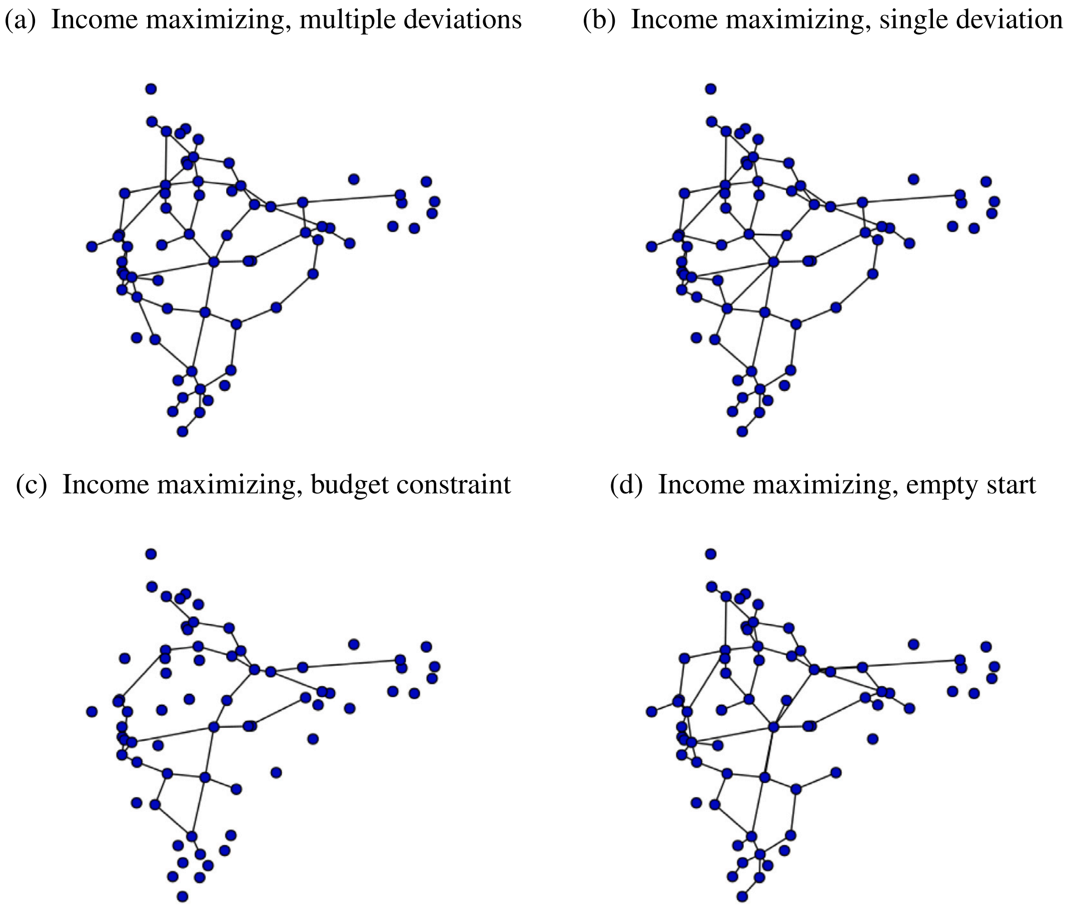

Fig. 5. Counterfactual networks.
All figures show networks designed with the algorithm that approximately maximizes net income among the 68 targeted cities, but with different constraints and initial conditions. Fig. 5(a) has no constraint, and the algorithm starts from the full network and uses multiple simultaneous deviations. Fig. 5(b) also shows the solution when the algorithm starts from the full network but now only uses single deviations. Fig. 5(c) imposes a budget equal to the estimated costs of phases 1 and 2, and the algorithm starts from the full network. Fig. 5(d) starts from the empty network and only uses single deviations.

A third alternative network design is to start from the empty network instead of from the full network. The result is shown in Fig. 5(d). While the structure is similar to Fig. 5(a), there are some differences, and the network starting from the empty network implies a 0.3% lower aggregate real income net of costs (see Fig. 2). This result is based on the version with single deviations, and the gap shrinks to 0.1% when using multiple simultaneous deviations. In the appendix, I show additional versions of the network, including the best and the worst local optima that the algorithm converges to.

8.2.1. *Aggregate effects of the counterfactual infrastructure*

Overall, the income-maximizing counterfactual network delivers sizable aggregate welfare gains compared to the actual network, i.e., compared to the GQ, and would increase real income net of construction and maintenance cost by 4.3%. The total construction costs are estimated based on the length of the network and costs per kilometer in Collier et al. (2016).[40]

There are some important assumptions that I made for the above calculations. First, I assume that India can raise the necessary capital to finance this substantially larger investment. I use a cost of capital of 5% for all calculations, which is above the past cost of capital of the GQ. This partly accounts for the possibility that the larger networks may be more costly to finance than the GQ. Second, the predicted cost of the counterfactual network is based on extrapolating from the cost of the GQ to other locations. Although the topography is taken into account, there may be unobserved factors (such as local regulation and land acquisition) that make road construction more expensive when implementing the counterfactual network in different parts of the country. However,

[40] Note that the estimate of the total cost is used to map the topography costs to USD, i.e., the cost of a particular segment depends on the local topography, but the costs are scaled such that they sum to the total cost based on Collier et al. (2016). The costs would be about half based on the budget (see Ghani et al., 2016).

given the magnitudes, it seems unlikely that this would overturn the overall result. Furthermore, I have chosen the higher estimate of construction costs based on Collier et al. (2016) instead of the budget. Third, I have abstracted from political economy considerations that affect the implementation of national infrastructure projects, and the benefits from the counterfactual network should be viewed as the gains that could be obtained without the potential political frictions. Finally, the comparison of the actual network and the income-maximizing network depends on the model. If the model is misspecified, then the implied optimal network may differ from the actual network, even if the latter is really optimal. Since the same model is used to assess the welfare gains from the implied optimal network compared to the actual network, these gains could simply be due to model misspecification. However, the general equilibrium gravity model used here has been widely and successfully applied in the quantitative analysis of spatial data (see Redding and Rossi-Hansberg (2017) for an overview). Furthermore, the key elasticity of income with respect to market access is consistent with the estimates based on observed time variation in the transport network in India, and I design and evaluate the counterfactual network in the same context and with the same model.

8.2.2. Distributional effects of the counterfactual infrastructure

Replacing the GQ with the counterfactual network that connects intermediate-sized cities in an approximately income-maximizing way would increase the market access of regions in the center and in the east, which have been neglected by the GQ. Particularly the central region had districts with low initial density and also low subsequent growth (see Figures A8 and A9 in the appendix). This area would become better connected by the counterfactual network and experience increases in market access and higher district-level income. Figure A7 in the appendix shows these distributional effects of the counterfactual relative to the GQ.

9. Robustness

This section discusses the robustness of the results to growth prior to the GQ, different values of the trade elasticity, and population growth.

9.1. Trends in district growth prior to road investment

The identification strategy used in Section 7.3 relies on the assumption that non-nodal districts were quasi-randomly affected by the GQ that connected the four largest economic centers. One may have the concern that the structure of the GQ was chosen precisely because it goes through certain non-nodal regions. One possibility could be that the GQ was planned such that it goes through regions that were already growing fast. Alternatively, the highways could also have been constructed to trigger growth where it has been particularly low. While I do not have GDP data prior to 2001, one can use light data to analyze this. Since the light data goes back to 1992 and the NHDP started after 2000, it is possible to test whether districts' growth rates prior to the NHDP are related to the subsequent reduction in travel costs due to new roads. To this aim, I estimate the specifications of Table 1 again but use as the dependent variable growth in light between 1992 and 2000, i.e., prior to investment. If it were the case that transport infrastructure was improved precisely in those districts that were already growing fast, then we should observe a positive correlation between increases in market access due to the GQ and the growth rate prior to its construction. The results are shown in Appendix Table A2. All point estimates are negative or near zero, suggesting that, if anything, locations with weaker growth were more likely to gain better market access. The estimates are insignificant except for the specification in column 5 with state fixed effects.

9.2. Alternative values for the trade elasticity

When solving the system of equations in (4) numerically to obtain the general equilibrium market access measures, it is necessary to choose a value for the trade elasticity parameter θ. Donaldson and Hornbeck (2016) obtain an elasticity of 8.2 and I use a value of 8 for the baseline. However, alternative values for the trade elasticity have been found in the literature as well (see e.g. Simonovska and Waugh, 2014). Appendix Tables A3 and A4 report the estimated effect of market access on GDP with values for the trade elasticity θ of 4 and 12. The point estimates change with θ as predicted by the model. In particular, the model-implied elasticity with $\theta = 4$ is 0.87, which is within the range of estimates in Table A3. Analogously, the model-implied elasticity with $\theta = 12$ is 0.47, which is within the range of estimates in Table A4.

9.3. Effects of market access on population

In the conceptual framework used for the counterfactual analysis, I take a long-term perspective and assume labor is mobile. In the medium term, labor mobility is likely to be limited. According to Census of India in 2001, about 30% of the Indian population live in a different place than at birth. But out of the total number of migrants, 60% migrated within the same district and therefore not across the units of my empirical analysis. Although it does not appear that labor mobility was large on average during the sample period used for the estimation, this does not necessarily imply that the population did not move in response to changes in transport infrastructure.[41]

[41] An alternative approach would be to consider a model with idiosyncratic location preferences (see e.g. Redding, 2016), but this would require estimating additional parameters.

In order to estimate the response of the population to road investments, I regress the decennial log change in each district's population between the 2001 and 2011 censuses on the change in market access due to transport investments. The results in Appendix Table A5 show that there is no significant effect of market access on population. Although all point estimates are positive, they are relatively small and well below the elasticity predicted by a model with mobile labor and reasonable parameter values. However, recall from Table 1 that using income data from 2002–2012, the estimated elasticity of income with respect to market access is consistent with a model with mobile labor.[42]

10. Conclusion

Investments in transport infrastructure are often used to foster economic development, as it is generally believed that insufficient transport infrastructure is an important constraint in many countries. However, the impact of these investments is difficult to identify due to the general equilibrium consequences of transport networks. Furthermore, it is challenging to determine the optimal way to design transport networks, although large sums are invested, and the gains from an optimal design could be large.

This paper contributes to our understanding of the effects of transport infrastructure on development by analyzing a major Indian highway project in a general equilibrium trade framework and comparing the effects to a counterfactual that approximately maximizes income. I combine the theoretical framework with subnational income data and geographically referenced information to measure income, terrain features, and road infrastructure at a high spatial resolution. I propose a heuristic network design algorithm that maximizes the real income net of road construction costs and show that the solution of the heuristic algorithm is statistically unlikely to deviate significantly from the global optimum.

The findings suggest that the actual network, the GQ, led to sizable aggregate net gains but unequal effects across regions because it targeted the four largest economic centers. The counterfactual that approximately maximizes income is substantially larger than the existing network, and it would lead to large additional gains net of construction costs. Furthermore, the actual and counterfactual networks have different distributional implications. The previously less developed regions that were neglected by the GQ would benefit from the counterfactual network that integrates regions through their intermediate-sized cities.

The implications of the findings above may extend to other countries. The theoretical framework allows quantifying the aggregate and distributional effects, and I find that both are important. This suggests that the debates about infrastructure investments in other countries should give careful consideration to the aggregate as well as distributional consequences of alternative networks. Furthermore, I show that the gains from the optimal design of transport networks are large. The heuristic network design algorithm could potentially be applied to other models, data, and contexts where the extensive margin of road investments is important.

The above results from the counterfactual network are based on an algorithm that has the objective of maximizing aggregate real income net of construction costs. In future work, it would be interesting to consider how other objectives such as distributional or political considerations can be taken into account directly in the design of the network.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jinteco.2025.104140.

Data availability

Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development (Original data) (Mendeley Data)

[42] Donaldson and Hornbeck (2016) also find that the elasticity of population with respect to market access is lower than predicted by the model, see their Table V.

References

Ahlfeldt, Gabriel M., Redding, Stephen J., Sturm, Daniel M., Wolf, Nikolaus, 2015. The economics of density: Evidence from the Berlin wall. Econometrica 83 (6), 2127–2189.

Allen, Treb, Arkolakis, Costas, 2014. Trade and the topography of the spatial economy. Q. J. Econ. 129 (3), 1085–1140.

Allen, Treb, Arkolakis, Costas, 2022. The welfare effects of transportation infrastructure improvements. Rev. Econ. Stud. 89 (6), 2911–2957.

Allen, Treb, Atkin, David, 2022. Volatility and the gains from trade. Econometrica 90 (5), 2053–2092.

Antras, Pol, Fort, Teresa C., Tintelnot, Felix, 2017. The margins of global sourcing: Theory and evidence from US firms. Am. Econ. Rev. 107 (9), 2514–2564.

Arkolakis, Costas, Eckert, Fabian, Shi, Rowan, 2025. Combinatorial Discrete Choice: Theory and Application to Multinational Production. NBER Working Paper No. 31877, https://www.nber.org/papers/w31877.

Asian Development Bank, 2007. Retrospective Analysis of the Road Sector, 1997-2005. Asian Development Bank.

Asturias, Jose, García-Santana, Manuel, Ramos, Roberto, 2019. Competition and the welfare gains from transportation infrastructure: Evidence from the Golden Quadrilateral of India. J. Eur. Econ. Assoc. 17 (6), 1881–1940.

Au, Chun-Chung, Henderson, J. Vernon, 2006. Are Chinese cities too small? Rev. Econ. Stud. 73, 549–576.

Balboni, Clare, 2025. In harm's way? infrastructure investments and the persistence of coastal cities. Amer. Econ. Rev. 115 (1), 77–116, https://doi.org/10.1257/aer.20191943.

Banerjee, Abhijit, Duflo, Esther, Qian, Nancy, 2020. On the road: Access to transport infrastructure and economic growth in China. J. Dev. Econ. 145, 102442.

Baum-Snow, Nathaniel, Brandt, Loren, Henderson, J. Vernon, Turner, Matthew A., Zhang, Qinghua, 2020. Does investment in national highways help or hurt hinterland city growth?. J. Urban Econ. 115, 103124, https://doi.org/10.1016/j.jue.2018.05.001.

Bird, Julia, Straub, Stéphane, 2020. The Brasilia Experiment: The Heterogeneous Impact of Road Access on Spatial Development in Brazil. World Development 127, https://doi.org/10.1016/j.worlddev.2019.104739.

Blouri, Yashar, von Ehrlich, Maximilian, 2020. On the optimal design of place-based policies: A structural evaluation of EU regional transfers. J. Int. Econ. 125, 103319.

Breinlich, Holger, Ottaviano, Gianmarco I.P., Temple, Jonathan R.W., 2013. Regional growth and regional decline. Handbook of Economic Growth.

Burgess, Robin, Jedwab, Remi, Miguel, Edward, Morjaria, Ameet, 2015. The value of democracy: evidence from road building in Kenya. Am. Econ. Rev. 105 (6), 1817–1851.

Carling, Kenneth, Meng, Xiangli, 2015. Confidence in heuristic solutions? J. Global Optim. 63 (2), 381–399.

Chandra, Amitabh, Thompson, Eric, 2000. Does public infrastructure affect economic activity? Evidence from the rural interstate highway system. Reg. Sci. Urban Econ. 30 (4), 457–490.

Chaudhuri, Shubham, Ravallion, Martin, 2006. Partially Awakened Giants: Uneven Growth in China and India. World Bank Policy Research Working Paper 4069, http://hdl.handle.net/10986/8877.

Chen, Xi, Nordhaus, William D., 2011. Using luminosity data as a proxy for economic statistics. Proc. Natl. Acad. Sci. 108 (21), 8589–8594.

Cicirello, Vincent A., Smith, Stephen F., 2004. Heuristic selection for stochastic search optimization: Modeling solution quality by extreme value theory. In: Principles and Practice of Constraint Programming-CP 2004: 10th International Conference, CP 2004, Toronto, Canada, September 27-October 1, 2004. Proceedings 10. Springer Berlin Heidelberg.

Coles, Stuart, 2001. An Introduction to Statistical Modeling of Extreme Values. Springer, London.

Collier, Paul, Kirchberger, Martina, Söderbom, Mans, 2016. The cost of road infrastructure in low-and middle-income countries. World Bank Econ. Rev. 30 (3), 522–548.

Coşar, A. Kerem, Demir, Banu, 2016. Domestic road infrastructure and international trade: Evidence from Turkey. J. Dev. Econ. 118.

Datta, Saugato, 2012. The impact of improved highways on Indian firms. J. Dev. Econ. 99, 46–57.

Dell, Melissa, 2015. Trafficking networks and the Mexican drug war. Am. Econ. Rev. 105 (6), 1738–1779.

Desmet, Klaus, Ghani, Ejaz, O'Connell, Stephen, Rossi-Hansberg, Esteban, 2013. The spatial development of India. J. Reg. Sci. 55, 10–30.

Donaldson, Dave, 2015. The gains from market integration. Annu. Rev. Econ. 7, 619–647.

Donaldson, Dave, 2018. Railroads and the Raj: Estimating the impact of transport infrastructure. Am. Econ. Rev. 108 (4–5), 899–934.

Donaldson, Dave, Hornbeck, Richard, 2016. Railroads and American economic growth: A market access approach. Q. J. Econ. 131 (2), 799–858.

Eaton, Jonathan, Kortum, Samuel, 2002. Technology, geography, and trade. Econometrica 70 (5), 1741–1779, https://doi.org/10.1111/1468-0262.00352.

Faber, Benjamin, 2014. Trade integration, market size, and industrialization: Evidence from China's national trunk highway system. Rev. Econ. Stud. https://doi.org/10.1093/restud/rdu010.

Fajgelbaum, Pablo D., Schaal, Edouard, 2020. Optimal transport networks in spatial equilibrium. Econometrica 88 (4), 1411–1452, https://doi.org/10.3982/ECTA15213.

Felbermayr, Gabriel J., Tarasov, Alexander, 2022. Trade and the spatial distribution of transport infrastructure. J. Urban Econ. 130, 103473, https://doi.org/10.1016/j.jue.2022.103473.

Frankfurter Allgemeine Zeitung (FAZ), 2013. Indien wird China überholen. www.faz.net (in German), https://www.faz.net/aktuell/finanzen/devisen-rohstoffe/indiens-zentralbank-gouverneur-rajan-indien-wird-china-ueberholen-12592912.html.

Gastner, Michael T., 2005. Spatial Distributions: Density-Equalizing Map Projections, Facility Location, and Two-Dimensional Networks (Thesis). The University of Michigan, https://hdl.handle.net/2027.42/125368.

Gastner, Michael T., Newman, M.E.J., 2006. Optimal design of spatial distribution networks. Phys. Rev. E 74 (1), https://doi.org/10.1103/PhysRevE.74.016117.

Ghani, Ejaz, Goswami, Arti Grover, Kerr, William R., 2016. Highway to success: The impact of the Golden Quadrilateral project for the location and performance of Indian manufacturing. Econ. J. 126 (591), 317–357, https://doi.org/10.1111/ecoj.12207.

Giddings, Angela P., Rardin, Ronald L., Uzsoy, Reha, 2014. Statistical optimum estimation techniques for combinatorial optimization problems: A review and critique. J. Heuristics 20, 329–358, https://doi.org/10.1007/s10732-014-9243-4.

Glaeser, Edward L., Kohlhase, Janet E., 2004. Cities, regions and the decline of transport costs. Pap. Reg. Sci. 83 (1), 197–228, https://doi.org/10.1007/s10110-003-0183-x.

Harral, Clel, Sondhi, Jit, Chen, Guang Zhe, 2006. Highway and Railway Development in India and China, 1992-2002. Transport Note TRN-32, The World Bank, Washington, D.C, https://hdl.handle.net/10986/11774.

Henderson, J. Vernon, Storeygard, Adam, Weil, David N, 2012. Measuring economic growth from outer space. Amer. Econ. Rev. 102 (2), 994–1028, https://doi.org/10.1257/aer.102.2.994.

Jalbert, J., Farmer, M., Gobeil, G., Roy, P., 2024. Extremes.jl: Extreme value analysis in julia. J. Stat. Softw. 109 (6), 1–35, https://doi.org/10.18637/jss.v109.i06.

Jaworski, Taylor, Kitchens, Carl T., 2019. National policy for regional development: Evidence from appalachian highways. Rev. Econ. Stat. https://doi.org/10.1162/rest_a_00808.

Jedwab, Remi, Storeygard, Adam, 2022. The average and heterogeneous effects of transportation investments: Evidence from sub-Saharan Africa 1960–2010. J. Eur. Econ. Assoc. 20 (1), 1–38, https://doi.org/10.1093/jeea/jvab027.

Jia, Panle, 2008. What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry. Econometrica 76 (6), 1263–1316, https://doi.org/10.3982/ECTA6649.

Khanna, Gaurav, 2016. The road oft taken: The route to spatial development. SSRN: https://ssrn.com/abstract=2426835.

Limao, Nuno, Venables, Anthony J., 2001. Infrastructure, geographical disadvantage, transport costs, and trade. World Bank Econ. Rev. 15 (3), 451–479, https://doi.org/10.1093/wber/15.3.451.

Los, Marc, Lardinois, Christian, 1982. Combinatorial programming, statistical optimization and the optimal transportation network problem. Transp. Res. Part B: Methodol. 16 (2), 89–124, https://doi.org/10.1016/0191-2615(82)90030-3.

Michaels, Guy, 2008. The effect of trade on the demand for skill: Evidence from the interstate highway system. Rev. Econ. Stat. 90 (4), 683–701, https://doi.org/10.1162/rest.90.4.683.

Morten, Melanie, Oliveira, Jaqueline, 2024. The effects of roads on trade and migration: Evidence from a planned capital city. Amer. Econ. J. Appl. Econ. 16 (2), 389–421, https://doi.org/10.1257/app.20180487.

Redding, Stephen J., 2016. Goods trade, factor mobility and welfare. J. Int. Econ. 101, 148–167, https://doi.org/10.1016/j.jinteco.2016.04.003.

Redding, Stephen J., Rossi-Hansberg, Esteban, 2017. Quantitative spatial economics. Annu. Rev. Econ. https://doi.org/10.1146/annurev-economics-063016-103713.

Redding, Stephen J, Turner, Matthew A, 2015. Transportation costs and the spatial organization of economic activity. Handbook Regional Urban Econ. 5, 1339–1398, https://doi.org/10.1016/B978-0-444-59531-7.00020-X.

Roberts, Mark, Deichmann, Uwe, Fingleton, Bernard, Shi, Tuo, 2012. Evaluating China's road to prosperity: A new economic geography approach. Reg. Sci. Urban Econ. 42, 580–594, https://doi.org/10.1016/j.regsciurbeco.2012.01.003.

Roodman, David, Nielsen, Morten Orregaard, MacKinnon, James G., Webb, Matthew D., 2019. Fast and wild: Bootstrap inference in Stata using boottest. Stata J. 19 (1), 4–60, (2019), https://doi.org/10.1177/1536867X19830877.

Simonovska, Ina, Waugh, Michael E., 2014. The elasticity of trade: Estimates and evidence. J. Int. Econ. 94, 34–50, https://doi.org/10.1016/j.jinteco.2013.10.001.

Storeygard, Adam, 2016. Farther on down the road: Transport costs, trade and urban growth in sub-Saharan Africa. Rev. Econ. Stud. 83 (3), 1263–1295, https://doi.org/10.1093/restud/rdw020.

Urban Land Institute, Ernst & Young, 2013. Infrastructure 2013: Global Priorities, Global Insights. Washington, D.C.: Urban Land Institute, https://uli.org/wp-content/uploads/ULI-Documents/Infrastructure-2013.pdf.

Van Leemput, Eva, 2021. A passage to India: Quantifying internal and external barriers to trade. J. Int. Econ. 131, 103473, https://doi.org/10.1016/j.jinteco.2021.103473.

World Bank, 2002. India's transport sector: The challenges ahead. https://hdl.handle.net/10986/15407.

World Bank, 2005. India: Road transport service efficiency study. Energy Infrastruct. Oper. Div. https://hdl.handle.net/10986/8356.

World Bank, 2007a. A Decade of Action in Transport. an Evaluation of World Bank Assistance to the Transport Sector, 1995-2005. World Bank, Washington, D.C, https://hdl.handle.net/10986/6695.

World Bank, 2007b. An Overview of China's Transport Sector - 2007. EASTE Working Paper 15, http://documents.worldbank.org/curated/en/337221468024252254.

World Bank, 2008. India: Accelerating Growth and Development in the Lagging Regions of India. Poverty Reduction and Economic Management Report No. 41101-IN, http://documents.worldbank.org/curated/en/587041468040572373.

World Bank, 2015. New rail corridors in India cut harmful gasses while boosting speed. http://documents.worldbank.org/curated/en/536381639368749847.