Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development*

Simon Alder
University of North Carolina at Chapel Hill
January 2019

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, while China had the explicit strategy of connecting intermediate-sized cities. The two countries also experienced different regional development patterns, with stronger convergence in China. This paper analyzes the aggregate and distributional effects of transport infrastructure in India based on a general equilibrium trade framework. I compare the effects of a recent highway project that improved the connections between Delhi, Kolkata, Chennai, and Mumbai to a counterfactual Indian highway network that mimics the Chinese strategy of connecting intermediate-sized cities. The counterfactual network among these cities is designed to approximately maximize net income based on the general equilibrium framework and road construction costs. I use satellite data on night lights to estimate the model at the level of Indian districts. The results suggest that the actual network led to large 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.

JEL Codes: F11, F14, F15, O11, O18, R12, R13

Keywords: Transport Infrastructure, Highways, Network Design, Trade, Economic Growth, Regional Development, India, China, Geographic Information System, Satellite Data, Night Lights.

*I would like to thank Fabrizio Zilibotti for his support during this project. I also thank Simeon D. Alder, Treb Allen, George-Marios Angeletos, Costas Arkolakis, Marco Bassetto, Timo Boppart, Filippo Bratti, Marius Brühlhart, Adrien Bussy, Jonathan Colmer, Kerem Çağrı, 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, Illelin Kondor, Peter Kondor, Michael König, Rafael Lalinde, Sergi Jiménez-Martín, Omar Licandro, Thierry Mayer, Andreas Müller, Alessandro Pavan, Michelle Rendall, José-Victor Rios-Rull, Dominic Rohner, Esteban Rossi-Hansberg, Lin Shao, Kjetil Storesletten, Viktor Tsyrennikov, Ashish Vachhan, Valentin Verdier, Rainer Winkelmann, Christoph Winter, Gabriel Zucman, Josef Zweimüller, Robert Zymek, and participants in presentations at the University of Zurich, University of Lausanne, University of Oslo, Arizona State University, University of Virginia, Bank of Canada, World Bank, NBER Summer Institute Workshop on Urban Economics, Barcelona Summer Forum, SNF-CEPR Conference, Midwest Macro Meeting, World Congress and Asia Meetings of the Econometric Society, Indian Statistical Institute, and Indian School of Business 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.
1 Introduction

China and India, the two most populous countries in the world, are developing at unprecedented rates. Yet, their spatial, or regional, development patterns are surprisingly different. Throughout China, new clusters of economic activity are emerging and there is a stronger pattern of convergence within the country. In contrast, a substantial number of Indian districts of intermediate density experience low growth and there is generally less convergence. While such differences in the spatial development of China and India have been documented in the literature (Desmet et al., 2013; Chaudhuri and Ravallion, 2006), we still lack precise explanations and possible policy measures.

This paper links the differences in the spatial development of the two countries to their major transport 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 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 Figure 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).

If transport infrastructure is a determinant of development, then one may ask how a network should be designed in order to foster growth and regional development. In this paper, I compare the aggregate and distributional effects of the GQ to a counterfactual Indian network that approximately maximizes aggregate income net of road construction costs in a general equilibrium framework. 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.

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 productivities 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
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.

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.

Since official district-level GDP data are not available for the entire period, I use night light data to measure real income and I compare the results with alternative data sources. Night time panel data is readily available at a high spatial resolution and has been shown to correlate strongly with real GDP (see e.g. Henderson et al., 2012). 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 1999 (before the construction of the GQ), in 2012 (after completion of the GQ), and for the counterfactuals.

The observed changes in market access from 1999 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. I then use this estimated elasticity in the model to predict districts’ incomes for each transport network. The general equilibrium model allows me to compare the aggregate and distributional implications from various actual and counterfactual networks.

The analysis 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 2.44 percent lower in 2012 if the GQ had not been built.

Second, I use the estimated model and a heuristic network design algorithm to construct the income-maximizing highway network among India’s intermediate-sized cities. The algorithm starts from the full network and iteratively removes or 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 it would cost more than eight times as much, but it would lead to a net increase in aggregate income (relative to the GQ) equivalent to 2.77 percent of GDP in 2012. I also compare the effect to an alternative counterfactual network that is designed to approximately equalize marginal costs and benefits without the constraint that all cities that would be targeted by the Chinese policy are connected. The resulting network is still more than seven times larger than the GQ and connects most intermediate-sized cities. It would imply an increase in net income of 2.79. While these counterfactual network designs are computed with a heuristic algorithm and do not necessarily represent the global optimum, 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.

---

1. Finding the optimal network in this framework is challenging, because there can be positive or negative complementarities between the links. The number of potential combinations of links that can form a network is extremely large and I use a heuristic algorithm to approximate the optimal network. In each iteration, the algorithm removes or adds links and recomputes the trade costs from the bilateral shortest paths and the implied general equilibrium income. The iterative procedure is also described in Gastner (2005), but not using a general equilibrium model.

2. I also compare the results to several alternative approaches to design the counterfactual network, including the least-cost network and ad hoc ways of implementing the Chinese strategy with a certain number of corridors. The results show that the optimal network based on the heuristic algorithm has a different structure and is substantially better in terms of net income.
In addition, the approximately 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.

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 how the transport network contributes to this lack of spatial convergence. To this aim, I compute the predicted income distribution across districts in 2012 for each version of the network and obtain the implied convergence rates from growth regressions starting in 1999. The results show that the network that maximizes aggregate net income would increase convergence compared to the GQ, because initially less developed districts benefit the most from the optimal network that is denser and reaches into lagging regions. In contrast, the GQ tends to reduce spatial convergence compared to the initial network, because the GQ targets the four already densest economic centers and many of the poorest regions do not benefit from it. This convergence pattern in India differs from the one in China, where poorer regions grew more rapidly. The light data is well suited for a comparison of spatial convergence in the two countries, because it measures income at a high spatial resolution and in a consistent way across space. I exploit this feature by constructing equally sized cells in both countries and then compare the convergence rates across cells in the two countries. The results show that China indeed had higher rates of spatial convergence than India, consistent with China having a highway network that benefits poorer regions more than the Indian network that targeted the densest economic centers.

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 and the empirical strategy. Section 5 discusses the data and the estimation of the model. Section 6 presents the effects of the actual and counterfactual networks. Section 7 discusses the robustness of the results and Section 8 concludes.
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. 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. I adapt the framework to consider the effect on real income across Indian districts as measured by night lights. 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 rely on a network design algorithm that adds and removes links based on construction costs and aggregate income implied by trade costs in the general equilibrium framework. Iterative procedures to solve related problems with complementarities are applied in Jia (2008), Antras et al. (2016), and Arkolakis and Eckert (2017), but not in the context of transport infrastructure. Allen and Arkolakis (2016) 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. Fajgelbaum and Schaal (2017)}
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 in each link of an underlying graph. 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. Another feature of this study is that I can first estimate the effect of transport infrastructure investments in India using variation in trade costs due to the construction of the GQ, and then search for the optimal network in the same context.

While the empirical analysis of this paper builds on general equilibrium trade theory, it is also related to recent studies on the local effects of transport infrastructure such as the GQ. For example, Datta (2012) and 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. An important aspect of these studies is the identification of exogenous sources of variation in transport infrastructure. They rely on an identification strategy similar to the one proposed by Chandra and Thompson (2000) who estimate the effect of U.S. highways on counties that lie between large nodal cities. This is based on the observation that the highways are built to connect larger cities and thereby pass through other counties, which consequently obtain access to the new transport infrastructure without being targeted themselves. I rely on the identification strategies based on non-nodal cities and straight lines to estimate the effect of the GQ on Indian sub-districts using light data. Aggarwal (2016) and Asher and Novosad (2016) use regression discontinuity designs to estimate the effect of rural roads in India. See Redding and Turner (2015) for a detailed
strategy based on non-nodal districts as in Chandra and Thompson (2000), but use it in the general equilibrium framework of Donaldson and Hornbeck (2016) in order to estimate the effect of market access on income.

Some recent studies analyze transport infrastructure in India based on general equilibrium models. Asturias et al. (2016) quantify the effect of the GQ based on a model of oligopolistic competition applied to Indian states. Van Leemput (2015) analyzes internal and external trade barriers in India using state-level trade data. Allen and Atkin (2016) 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) and using this estimate to predict aggregate and distributional implications from various counterfactual transport network designs.

Several studies focus on the effect of transportation infrastructure in other countries. 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 connect cities in an approximately income-maximizing way in a general equilibrium framework. Banerjee et al. (2012) estimate the effect of transport infrastructure in China and use straight lines between important cities as instruments for the potentially endogenous location of infrastructure. Baum-Snow et al. (forthcoming) analyze the effect of transport infrastructure on the decentralization of Chinese cities. They use exogenous variation based on historical maps of transport infrastructure and also use light data as a measure for economic activity. Baum-Snow et al. (2017) estimate the effect of Chinese highways on growth in regional primates and hinterland prefectures and distinguish the effect of domestic and international market access. Roberts et al. (2012) also estimate the effect of Chinese highways but use a structural new economic geography model. Allen and Arkolakis (2014) develop a spatial equilibrium framework with mobile labor and apply it to quantify the general equilibrium effects of the U.S. interstate highway network.9 Ahlfeldt et al. (2015) use a quantitative spatial
discussion of different identification strategies to estimate the effect of transport infrastructure.

9See also Atack et al. (2008), Herrendorf et al. (2012), and Jaworski and Kitchens (2016) for studies
model with mobile workers in order to analyze the effect of the Berlin Wall. I differ from these studies by comparing the aggregate and distributional effects of the actual network to a counterfactual network that connects targeted cities in an approximately income-maximizing way.

The market access approach used in this paper is closely related to models in the new economic geography literature (see Fujita et al., 1999). Several authors analyze the role of market access (or market potential), which can be affected by transport costs (Puga, 2002; Redding and Venables, 2004; Hanson, 2005; Redding and Sturm, 2008; Head and Mayer, 2011, 2013). They find that market access is associated with trade, income, and population within and between countries. This paper also relates more broadly to a large literature on trade, in particular on the gravity structure (see for example Anderson and van Wincoop, 2003; Allen and Arkolakis, 2014; Redding, 2016; Redding and Rossi-Hansberg, 2017). Head and Mayer (2011) point out that the gravity structure and market access can be derived from various trade models with different market structures and sources of gains from trade. The geographically coded digital transport network that is used here can model explicitly how trade costs and thus proximity change due to transport infrastructure. Thus, changes in transport costs generate variation in market access which allows to study the relationship between income and market access over time.\(^\text{10}\) While these models are static, Desmet and Rossi-Hansberg (2014) propose a model of spatial development based on technology spillovers where growth depends on the density of economic activity.\(^\text{11}\)

\(^{10}\) The present analysis is also related to a literature studying the role of transport infrastructure within cities. Baum-Snow (2007) uses planned routes as instruments for the actual transport network that was developed in the U.S. and finds that highways led to a decentralization of population. Duranton and Turner (2012) estimate the effect of highways within U.S. cities’ boundaries on their employment growth and conduct policy experiments that extend the highway network. Duranton et al. (2014) estimate the effect of highways on intercity trade flows.

\(^{11}\) The assessment of the development effects of transport infrastructure naturally relates to cost-benefit analyses of individual infrastructure projects. For example, as a major investor in transport infrastructure in developing countries, the World Bank has developed procedures to evaluate the effec-
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). Figure 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. 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 Figure 1) with the explicit objective of connecting all cities with more than 500,000 people and all provincial capitals in a modern highway system. 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 speed than the existing roads.

---

12 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).

13 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 Chinese Ministry of Transportation (2004), World Bank (2007b), Roberts et al. (2012), and Faber (2014) for a discussion.

14 A description of the history of the Chinese highway network and its different components is provided by ACASIAN. See www.acasian.com for further details.
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 Figure 1). Construction, mostly upgrades of existing highways to higher quality, began in 2001 and was completed by 2012 with a total network length of 5,846 km and at approximately a cost of USD 5.4 billion (1999 prices).\textsuperscript{15} 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. 2015).

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 which 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.\textsuperscript{16} 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; DHL, 2007; Ernst and Young, 2013; KPMG, 2013).

\textsuperscript{15}See the webpage of the National Highway Authority of India (http://www.nhai.org/index.asp) for details on individual segments. The cost estimates are based on Ghani et al. (2016).

\textsuperscript{16}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.
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, 2012). Today, roads are the most important transport mode in India, carrying 60% of the freight turnover compared to 31% for railways. 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. Furthermore, I compare the results to alternative counterfactual networks that minimize construction costs or implement further policy objectives. The next section presents the general equilibrium framework that is used to design the counterfactual network and to quantify and com-

---

17 The share of highways in the total freight turnover is even higher in India than in China (KPMG, 2013).
pare 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. 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 by adapting the framework to a version which can be estimated with light data as a measure of real income.

4.1 Trade between Indian Districts

The basic setup is a trade model with the immobile production factors land and labor and the mobile factor capital. The economy consists of many trading regions (i.e. Indian districts), where the origin of a trade 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,

\[
p_{oo}(j) = MC_o(j) = \frac{q_o w_o r_o^{1-\alpha-\gamma}}{z_o(j)},
\]

where \( q_o \) is the land rental rate, \( w_o \) is the wage,

---

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

19 Over the sample period of 12 years that I have available to estimate the model based on the Indian GQ, the assumption of immobile labor is more appropriate because labor mobility is relatively low across Indian districts. However, the analysis could also be carried out when assuming labor is mobile. Section 7.7 discusses labor mobility in more detail.
$r_o$ is the interest rate, and $\alpha$ and $\gamma$ 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

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

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.\textsuperscript{20} This equation provides a relationship between prices and consumer market access, which will be exploited below to derive real income.

### 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 ($X_d = Y_d$),\textsuperscript{21} we obtain the gravity equation

$$X_{od} = \frac{T_o (q_o^\alpha w_o^\gamma )^{-\theta}}{\tau_{od}^{-\theta}} \times \frac{T_{od}^{-\theta}}{CMA_d^{-\frac{1}{\theta}}} \times Y_d,$$

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

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

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

\textsuperscript{20}Using the fact that the rental rate for capital is equalized everywhere to $r_o = r$, we can define the constant $\kappa_1 \equiv \mu r^{-\theta - (1-\alpha-\gamma)\theta}$ as described in the appendix.

\textsuperscript{21}Note that capital is mobile but labor is not. The assumption that income equals expenditures therefore implies that capital rents are consumed where the capital is used for production.
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\(^{22}\), 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

\[
MA_o = \rho \sum_d \tau_{od}^{-\theta} MA_d^{-1} Y_d, \quad (4)
\]

and Equation (3) for income becomes

\[
Y_o = \kappa_1 T_o (q_o^\alpha w_o^\gamma) - \theta MA_o, \quad (5)
\]

Equations (4) and (5) are two key model equations and they summarize how trade costs affect income. While Equation (5) implies a relationship between income and market access, Equation (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 to quantify aggregate effects.

### 4.1.3 Real Income

The framework summarized by Equations (4) and (5) suggests a relationship between transport costs and income that can be estimated with district-level data.\(^{23}\) I will use night lights as a measure of real income and therefore rewrite the income equation accordingly. To this aim, I use the property that the price index is related to market access,

\[
P_d = (\rho^{-1} MA_d)^{-\frac{1}{\theta}}. \quad (6)
\]

This allows rewriting Equation (4) as

\[
MA_o = \rho \frac{1+\theta}{\sigma} \sum_d \tau_{od}^{-\theta} MA_d^{-\frac{(1+\theta)}{\sigma}} Y_d, \quad (7)
\]

\(^{22}\)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.

\(^{23}\)Donaldson and Hornbeck (2016) solve for land rental rates to quantify how the effect of railways is capitalized into the fixed factor land.
where \( Y_{d}^{r} = \frac{Y_{d}}{P_{o}} \) denotes real income. Equation (5) can also be written in terms of real income by using the price index in Equation (6). Furthermore, the wage and land rental rates can be substituted using the factor income shares to obtain\(^{24}\)

\[
Y_{o}^{r} = \left( \kappa_{2}T_{o} \right)^{\frac{1}{1+\theta(\alpha+\gamma)}} \left( \frac{\alpha}{L_{o}} \right)^{-\theta\alpha} \left( \frac{\gamma}{H_{o}} \right)^{-\theta\gamma} \left( MA_{o} \right)^{1+\theta(1+\alpha+\gamma)} \left( 1+\theta(1+\alpha+\gamma) \right)^{\theta},
\]

where \( \kappa_{2} = \kappa_{1} \rho^{-\frac{1+\theta(\alpha+\gamma)}{\theta}}. \) The general equilibrium effect of transport infrastructure on income is captured by the market access measures. The elasticity of income with respect to market access in Equation (8) will be estimated using panel data on real income (proxied by luminosity) and market access measures that solve Equation (7). This estimate will then be used for the counterfactual analysis.

4.2 Empirical Strategy

Estimating Equation (8) in a cross-section would require to control for relevant district characteristics, which are difficult to obtain. Therefore, the above equation will be estimated with a fixed effect panel regression that relies on the time variation within districts. This allows accounting for the unobserved heterogeneity across districts. Equation (9) shows the different components of Equation (8) over time:

\[
\ln \left( Y_{o,t}^{r} \right) = -\frac{\theta\alpha}{1+\theta(\alpha+\gamma)} \ln \left( \frac{\alpha}{L_{o}} \right) - \frac{\theta\gamma}{1+\theta(\alpha+\gamma)} \ln \left( \frac{\gamma}{H_{o}} \right) + \frac{1}{1+\theta\alpha+\theta\gamma} \ln \left( \kappa_{2,t} \right) + \frac{1}{1+\theta(\alpha+\gamma)} \ln \left( T_{o,t} \right) + \frac{1+\theta(1+\alpha+\gamma)}{(1+\theta(\alpha+\gamma))\theta} \ln \left( MA_{o,t} \right). \tag{9}
\]

The right-hand side of the first line in Equation (9) collects parameters and factor endowments, which are assumed to be constant over time and thus absorbed by the district fixed effects. The second line includes country characteristics (the interest

\(^{24}\)With the Cobb-Douglas production function and competitive markets, we have \( q_{o} = \frac{Q_{o}}{L_{o}} \) and \( w_{o} = \frac{\gamma Y_{o}}{H_{o}}. \)
rate inside of $\kappa$) that are absorbed by state-year fixed effects. 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 is no effect of unobserved productivity changes on market access. Furthermore, part of the unobserved heterogeneity is absorbed by the state-year fixed effects. The last line in Equation (9) shows the effect of market access.

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

$$\beta = \frac{1 + \theta(1 + \alpha + \gamma)}{(1 + \theta(\alpha + \gamma))\theta}. \quad (10)$$

The panel fixed effects specification corresponding to Equation (9) to estimate $\beta$ is

$$\ln(Y_{o,s,t}^r) = \phi_o + \delta_{s,t} + \beta \ln(MA_{o,t}) + X_o\pi_t + \varepsilon_{o,s,t}, \quad (11)$$

where $\phi_o$ is a location fixed effect, $\delta_{s,t}$ is a state-year fixed effect, and $X_{s,t}$ is a vector of district characteristics such as distance from the coast and share of households with electricity in the initial year, interacted with a year fixed effect. Changes in the transport network imply different trade costs. Solving Equation (7) for each set of trade costs implies different market access measures, which generates the time variation in MA in Equation (11).

### 4.3 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 which 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. 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 Equation (7) 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). Equation (12) shows this version of the market access equation,

\[ MA_{o,t} = \rho \sum_d \tau_{od,t}^{1+\theta} MA_{d,t}^{-\theta} Y_d,1999. \]  

4.4 Counterfactual Predictions

\( \beta \) represents the elasticity of real income with respect to market access, as shown in Equation (8). The general equilibrium market access measures implied by Equation (7) themselves also depend on income. Furthermore, capital mobility implies that the

25 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. (2012), Ghani et al. (2016), Asturias et al. (2016), 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.

26 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.
real interest rate is constant.\textsuperscript{27} I jointly solve this system of equations for each version of the transport network, holding constant the immobile production factors, districts’ technology levels, and the real interest rate. Another approach in the literature is to use the estimate of $\beta$ in Equation (11) to predict changes in income based on counterfactual market access measures, but holding income in the market access equation fixed instead of jointly solving Equations (7) and (8). Hence, market access changes over the counterfactuals only due to the trade costs and does not account for the reallocation of income.\textsuperscript{28} I will focus on the first approach and solve Equations (7) and (8) jointly to predict income.

For the computation of the market access measures that solve Equation (7), a value for the trade elasticity $\theta$ is required. As a benchmark, I use a value of 8.2 as in Donaldson and Hornbeck (2016) and I report results with alternative values for $\theta$ in the robustness section.\textsuperscript{29}

5 Data and Estimation

The data required for the estimation of Equation (11) are income of each location and bilateral trade costs. I focus on the period from 1999 to 2012 and estimate the equation in differences. I have data on 636 mainland Indian districts, but there are a few missing observations such that the estimation is based on 626 districts. The data and estimation are discussed below and further details on the data and descriptive statistics can be found in Appendix B.

\textsuperscript{27}Capital is assumed to be freely mobile across Indian districts and the international capital market. This requires choosing an appropriate price for capital. In the baseline, I will choose the price index of Mumbai, but I also used a national weighted average (see also Appendix A.8).

\textsuperscript{28}See Head and Mayer (2013) and Donaldson and Hornbeck (2016) for a comparison of different approaches to quantify the effect of trade costs.

\textsuperscript{29}Donaldson and Hornbeck (2016) also estimate the effect of transport infrastructure through market access and this value is a natural benchmark. Donaldson (2018) obtained a value of 3.8, but in the context of the Indian colonial railway network. Using international trade data, Simonovska and Waugh (2014) find a value of 4.10. As I show in the robustness section, using values for $\theta$ below 8.2 tends to make the aggregate effects somewhat larger, but less than proportional because $\theta$ affects both the elasticity $\beta$ and the market access measures. See also Donaldson and Hornbeck (2016) for a comparison of aggregate effects with different $\theta$. 
5.1 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.

5.1.1 Real Income

I use Indian districts with administrative boundaries as of 2011. Official GDP data is not available for the full period and I therefore use night lights as a measure for real income in my baseline. The Indian Planning Commission publishes district level GDP data, but only for the years 1999 - 2004. This is not a long enough period to evaluate the effect of the GQ, but I use this data in robustness section 7.3 in order to estimate the relationship between GDP and light and then predict GDP for the years after 2004. Estimates for real GDP in the later years are available from Nielsen, a commercial data provider, and I report the results using this data as a robustness check in Section 7.4.

Growth in light at night measured by weather satellites has been shown to be a good proxy for income growth (Henderson et al., 2012). Two important advantages of the light data are that it has a high spatial resolution and is independent of countries’ statistical procedures. It is particularly useful when official GDP figures are not available, for example for sub-national administrative units such as Indian districts.\footnote{See also Ma et al. (2012) and Hodler and Raschky (2014) for applications of light data at the sub-national level. A number of other studies have also shown that light correlates with economic activity, for example Elvidge et al. (1997), Ghosh et al. (2010), and Chen and Nordhaus (2011).} In the DMSP-OLS version of the light data that is used here, the intensity of light of each pixel is measured on a scale from 0 to 63. The resolution is 30 arc-seconds, which is less than one kilometer at the equator. I calculate the sum of light of all pixels within a district in each year, holding the administrative boundaries fixed in 2011. This allows me to construct a panel of districts with light data from 1992 to 2013. In order to reduce measurement error, I average the three pre-construction years and the three final years. The data source and details on the light data preparation are provided in Appendix B.1.

While the estimation is based on light data, the results of the counterfactual analysis are in terms of GDP. In order to translate growth in light to growth in GDP, I use the
estimated elasticity of GDP with respect to light obtained by Henderson et al. (2012) from a large panel of countries. They find a log-linear relationship between GDP and light with an elasticity of about 0.3 when estimated in long differences.

There are some caveats when using light data. First, the light intensity is top-coded at 63 for extremely bright pixels, such that growth may be underestimated for dense city centers. Less than 10% of the districts have at least one pixel that is top coded in 1999, but this fraction increases to more than 20% by 2013. This could imply that growth is underestimated in the brightest cities. Interestingly, the evidence cited in the introduction suggests that the locations of intermediate density are the ones that have surprisingly low growth, which cannot be explained by the top coding. A second caveat is that light emissions depend on other infrastructure, in particular the electricity network. To address this, I control for the share of households with access to electricity in 2001. Furthermore, the transport infrastructure itself leads to light emissions (e.g. from traffic and street lights) that could potentially overstate income growth generated by the road. To address this, I show in Section 7.2 that the estimation results are robust to removing the light pixels that overlap with the GQ.

5.1.2 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 the 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

\[3^{1/2}\text{The alternative would be to use a radiance-calibrated version of the data, but this is unfortunately not available in all years.}\]
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 data sources and a more detailed description of the data preparation are provided in Appendix B.3.

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 A1 in the appendix. I use the fast marching algorithm as in Allen and Arkolakis (2014) and Allen and Atkin (2016) to compute bilateral travel times. The algorithm calculates the cheapest way to travel from one location (district centroids, represented by dots in Figure A1) 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 the transport costs.

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. More precisely, I calculate iceberg trade costs between an origin \( o \) and a destination \( d \) as

\[
\text{TradeCosts}_{od} = 1 + \gamma \text{TransportTime}_{od}^{0.8}.
\]  

\( \gamma \) is chosen such that the median iceberg trade cost is 1.25 for the network without the GQ. The transport costs within districts is set to 1.

Although the analysis undertaken here captures a key aspect of the modern trans-

---

32 The code for the fast marching algorithm is from the "accurate fast marching" Matlab toolbox by Dirk-Jan Kroon.

33 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.

34 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. (2017) for a similar calculation for China.
port 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 sea ports or airports is not modeled as part of the general equilibrium framework, although I show that the results are robust to adding international trade as additional income in districts with major ports. 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. The concern regarding electricity can be addressed by controlling for the share of households with access to electricity. Furthermore, it should be noted that I control for location and state-year fixed effects that partly absorb that coastal regions may have developed differently because of international trade. More generally, the above 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 4 discusses how I address this with my empirical strategy.

5.1.3 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 1x1 km cell of land depends on the slope and the share of water and built up area in the following way:

\[
\text{ConstructionCosts}_c = 1 + \text{Slope} + 25 \times \text{Builtup} + 25 \times \text{Water}.
\] (14)

Slope is measured in percent and Builtup and Water are binary indicators which take the unit value if the majority of the cell is built up or water, respectively.\(^{35}\) Applying this formula using detailed terrain data produces a 1x1 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.

\(^{35}\)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.
between any two given points through the cost grid.\textsuperscript{36} The procedure is illustrated in Figure A2 in the appendix, where the cells represent different construction costs (based on Equation 14) 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.\textsuperscript{37} Hence, only the relative weights of the different components in Equation (14) matter.

5.2 Estimation

The estimation of $\beta$ is based on Equation (9) of the model, suggesting a log-linear relationship between real income and market access. This equation is estimated using variation in real income and market access over time, where real income is approximated by night lights. The corresponding empirical specification is shown in Equation (11). As discussed in Section 4.3, the time variation in market access can be due to changes in trade costs or neighbors’ incomes, but I instrument market access with a measure that only varies due to changes in trade costs (Equation 12).

Table 1 shows the IV estimates for Equation (11). As a benchmark, column 1 of Table 1 uses the full sample of mainland Indian districts and regresses the logarithm of light on the logarithm of market access, controlling for district fixed effects, state-year fixed effects, distance from the coast, and initial electrification (both interacted with a year indicator). Observations are weighted by the logarithm of districts’ initial light in 1999 and standard errors are clustered at the state-level.\textsuperscript{38}

\textsuperscript{36}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).

\textsuperscript{37}I build the GQ on the road construction cost surface based on Equation (14) to obtain the total costs based on the topography. The ratio to the cost of the GQ in USD (from 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.

\textsuperscript{38}In all regressions, the light data is averaged over three years to reduce measurement error. The first stage, shown in Appendix Table A13, is strong and the elasticity between the two market access measures is 1.06. Table A14 shows the OLS results where the market access measure varies over time because of changes in the trade costs and neighbors’ incomes. The OLS estimates are larger because of the spatial dependence in income.
Table 1: Elasticity of light with respect to market access

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Market Access</td>
<td>0.614</td>
<td>0.643</td>
<td>0.649</td>
<td>0.657</td>
<td>0.657</td>
</tr>
<tr>
<td></td>
<td>(0.334)</td>
<td>(0.356)</td>
<td>(0.358)</td>
<td>(0.381)</td>
<td>(0.294)</td>
</tr>
<tr>
<td>Excluded nodal districts</td>
<td>None</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Weighting</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Standard errors</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Robust</td>
</tr>
<tr>
<td>N</td>
<td>626</td>
<td>613</td>
<td>612</td>
<td>612</td>
<td>612</td>
</tr>
<tr>
<td>Rsq.</td>
<td>0.508</td>
<td>0.502</td>
<td>0.502</td>
<td>0.495</td>
<td>0.495</td>
</tr>
</tbody>
</table>

The table shows 2SLS estimates of the elasticity of light with respect to market access. The dependent variable is the logarithm of the sum of light in each district in the years 1999 and 2012, where the start and end periods are averaged over three years. The explanatory variable is market access computed based on Equation (7) and instrumented with the market access with constant light in Equation (12). All regressions include district fixed effects, state-year fixed effects, and controls for distance to the coast and the level of electrification in 2001 interacted with a year fixed effect. Column 1 shows the effect in the full sample and columns 2 - 5 exclude four or five nodal districts as stated in the table. Columns 1 - 3 weigh by the logarithm of initial sum of light. Standard errors (in parentheses) are clustered at the state-level except in column 5.

The estimated IV coefficient for the effect of market access on light in Table 1 implies that a one percent increase in market access is associated with a 0.614 percent increase in light. The estimate is significant at 10% when clustering at the state-level and at 5% when using robust standard errors (see column 5). As emphasized above, estimating the causal effect of transport infrastructure is challenging because the location of the infrastructure may be endogenous and there is spatial dependence in income. The latter is addressed by instrumenting market access with a measure that varies only due to trade costs. To address the potentially endogenous location of infrastructure, I apply the identification strategy outlined in Section 4.3 and discussed below.

First, it should be noted that the panel structure of the data helps addressing concerns due to time-invariant unobserved location characteristics, which here are absorbed by the district fixed effects. It is in principal possible that there are relevant time-varying district characteristics, but it is less likely – given the national scope of

---

39 It is important to note that this elasticity of light with respect to market access cannot directly be compared to the elasticity predicted by the model, because light is used as a proxy for income. The aggregate results will be translated to GDP using the elasticity of GDP with respect to light estimated in Henderson et al. (2012). See also Section 7.3 for a discussion.
the NHDP – that such shocks would also affect the transport network. Furthermore, time-varying heterogeneity at a higher level of aggregation is absorbed by state-year fixed effects, which allow for differences in states’ growth trends.

In the second column, following an identification strategy similar to Chandra and Thompson (2000), I exclude the nodal districts and only exploit the variation in other districts that were not directly targeted by the roads connecting the four largest centers. The point estimate changes only marginally to 0.643. This suggests that the observed correlation between market access and light is not driven by the endogenous location of infrastructure in the four nodes.\(^40\) In column 3, I exclude a fifth node, Bangalore, that also appears to have been targeted – although it was not explicitly stated in the objective. The estimate is 0.649 and similar when excluding four nodes. The regressions shown in column 1 - 3 weigh observations by the logarithm of sum of light in 1999. Columns 4 and 5 are based on unweighted regressions, which does not alter the point estimates substantially. Finally, column 5 shows the result when using robust standard errors, which leads to more precise estimates.\(^41\) For the counterfactual exercise below, I rely on the estimate from column 3 in Table 1, which weighs observations by the log of the initial sum of light and excludes the five nodes. Section 7 discusses the robustness of the results to growth prior to the GQ, light emissions from roads, alternative GDP measures, international trade, different values of the trade elasticity, and population growth.

### 6 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. This estimate can be used in the model to predict how much each district’s income changes for various counterfactual networks. Because the framework captures general equilibrium effects of transport infrastructure,

\(^{40}\)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.

\(^{41}\)Columns 1 - 4 allow errors to be correlated within states, but the number of clusters is relatively small with 32 states in the sample.
this allows 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 observed levels in 2012 (see Section 4.4). The aggregate effects of each network are summarized in Table 2 and discussed in detail below.

6.1 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) and then comparing the predicted income to the actual income. I first discuss the aggregate effects and then consider the distributional implications.

6.1.1 Aggregate Effect of GQ

Based on the estimated elasticity $\beta$ and trade costs implied by a network without the GQ, the model suggests that aggregate light in 2012 would be 8.44 percent lower if the GQ had not been built. To predict the change in GDP from the estimated effect on light, I use an elasticity of income with respect to light of 0.3, which is approximately what Henderson et al. (2012) find in long difference estimations across many countries. A 8.44 percent change in light corresponds to a 2.53 percent change in GDP, as shown in Table 2. Aggregate GDP in India in 2012 was USD 1,099 billion (in 1999 prices) such that a 2.53 percent difference would correspond to an annual change in GDP of roughly USD 28 billion.

The total costs of the GQ (including interest until 2012) amounted to about USD 6.2 billion (1999 prices).\(^\text{42}\) 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

\(^{42}\)The total budgeted construction costs of phase 1 of the highway development program (that included the GQ and NS-EW) are reported in Ghani et al. (2016) as USD 7 billion. I approximate the share of the costs due to the GQ based on its length, which is about 78% of the total of phase 1. Furthermore, I assume that the costs accrued equally over this period. In what follows, the base year is omitted, but all values are in USD at 1999 prices.
Table 2: Aggregate effects of actual and counterfactual transport networks

<table>
<thead>
<tr>
<th></th>
<th>Costs</th>
<th>Income</th>
<th>Net income</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% USD</td>
<td>% USD</td>
<td>% USD</td>
</tr>
<tr>
<td>Removing GQ</td>
<td>-0.10</td>
<td>-1.05</td>
<td>-2.53</td>
</tr>
<tr>
<td>Income-maximizing network, all 68 cities</td>
<td>0.69</td>
<td>7.63</td>
<td>3.46</td>
</tr>
<tr>
<td>Income-maximizing network, unconstrained</td>
<td>0.65</td>
<td>7.16</td>
<td>3.44</td>
</tr>
<tr>
<td>Least-cost network, all 68 cities</td>
<td>0.21</td>
<td>2.31</td>
<td>0.24</td>
</tr>
<tr>
<td>Rays and corridors, all 68 cities</td>
<td>0.39</td>
<td>4.31</td>
<td>2.10</td>
</tr>
<tr>
<td>Income-maximizing network, least-cost budget</td>
<td>0.20</td>
<td>2.24</td>
<td>1.93</td>
</tr>
<tr>
<td>Income-maximizing network, NHDP budget</td>
<td>0.16</td>
<td>1.74</td>
<td>1.61</td>
</tr>
</tbody>
</table>

The table summarizes the aggregate effects of the actual and counterfactual networks. The changes in construction costs, income, and net income due to each network are shown in percentages of actual GDP in 2012 and in billion USD (1999 prices). 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 row and below are assumed to replace the GQ and the construction costs of the GQ are subtracted from the construction costs of the counterfactual.

(2014) report for the U.S. interstate highway network. Based on these assumptions, the annual net effect on income when removing the GQ implies a decline of around 2.44% of GDP, or USD 27 billion.

The result above depends on the two elasticities (GDP with respect to light and light with respect to market access). In robustness section 7.3, I show the results when first predicting district-level GDP based on the light data. This allows me to directly estimate the effect of market access on real GDP and then perform the counterfactual analysis with that estimate and the predicted GDP measures. The results suggest that the effect of the GQ is somewhat larger. Another assumption when computing net effects concerns the costs of the infrastructure construction. As pointed out by Ghani et al. (2016), there is some uncertainty regarding the exact costs of the GQ. However, the assumption on cost of capital is conservative and, given the magnitudes, it is unlikely that the true costs are so different that they would overturn the overall result.

6.1.2 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 ana-
lyze 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 substantially over the various versions of the transport networks.

Figure 2 shows the effects of the GQ at the level of Indian districts. The numbers represent the percentage gains 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 variation in the gains across districts. The largest beneficiaries had a more than 7 percent higher GDP level in 2012 than they would have had in the absence of the GQ.\footnote{Although transport costs did not increase anywhere (it is assumed that the GQ was added to an otherwise unchanged transport network), there are also some districts with losses from the infrastructure investments due to general equilibrium effects. However, there are only seven such districts and their losses are relatively small. Note that the distributional effects here are driven by differences in market access across districts, and not by differential effects of market access across districts. Alder et al. (2017) discuss potential heterogeneity in the effect of transport infrastructure in India using a two-sector model.}

6.1.3 Spatial Convergence

Many of the largest beneficiaries of the GQ tended to also have high initial density. This has consequences for the spatial convergence patterns and I show in Table 3 that convergence across Indian districts may have been reduced by the GQ. As shown in Table 3 columns 1 and 2, the coefficient is -0.170 with the GQ, and -0.176 without the GQ, suggesting that spatial convergence could have been larger without the GQ. Column 3 shows the result for the counterfactual network that will be discussed below (see also Section 6.2.4).

Columns 4 and 5 compare the spatial convergence in India and China. Since the two countries have different administrative divisions, I define squares of 20 km in both countries in order to make the spatial units comparable.\footnote{The qualitative results are the same when using larger cells. With 100 km squares, the coefficients are -0.14 for India and -0.25 for China. Note that cells with no light in any year are excluded.} I then measure the intensity of light in each square and year for both countries and regress the log difference in light density between 1999 and 2012 on the initial log light density in 1999 (again averaging three years). The results from these unconditional convergence regressions show that there is stronger convergence in China than in India, which is in line with the literature cited in the introduction. Desmet et al. (2013) use employment data and they also find that there are important differences in the spatial development of India.
and China, especially a lack of growth in India’s intermediate-sized cities. In the next section, I design a counterfactual Indian road network that connects intermediate-sized cities like the Chinese NEN.\textsuperscript{45} I use a heuristic algorithm to search for the network that maximizes aggregate net income and I then investigate the distributional implications and spatial convergence.

### Table 3: Spatial convergence

<table>
<thead>
<tr>
<th></th>
<th>(1) GQ</th>
<th>(2) No GQ</th>
<th>(3) Max. Income</th>
<th>(4) India (squares)</th>
<th>(5) China (squares)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log mean light in 1999</td>
<td>-0.170</td>
<td>-0.176</td>
<td>-0.187</td>
<td>-0.175</td>
<td>-0.288</td>
</tr>
<tr>
<td></td>
<td>(0.0241)</td>
<td>(0.0239)</td>
<td>(0.0239)</td>
<td>(0.00877)</td>
<td>(0.00574)</td>
</tr>
<tr>
<td>N</td>
<td>630</td>
<td>630</td>
<td>630</td>
<td>7283</td>
<td>11317</td>
</tr>
<tr>
<td>Rsq.</td>
<td>0.248</td>
<td>0.265</td>
<td>0.290</td>
<td>0.138</td>
<td>0.276</td>
</tr>
</tbody>
</table>

The dependent variable in columns 1 to 3 is the log difference in light in India between 1999 and 2012, where income in 2012 is obtained from the model for various networks. The explanatory variable is initial log light density. Columns 1 to 3 show convergence with the actual network in 2012 (including the GQ), without the GQ, and with the income-maximizing network, respectively. The dependent variable in columns 4 and 5 is the log difference in actual light between 1999 and 2012 for 20km cells in India and China, respectively, and the explanatory variable is initial log light density in each cell. Robust standard errors are in parentheses.

### 6.2 Counterfactual Networks

The counterfactual exercise asks how India would develop if it had built a highway network that targets intermediate-sized cities like the Chinese NEN and connects these cities 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 A3 in the appendix. I then design counterfactual networks among the targeted cities based on the general equilibrium model. The approach to design the approximately income-maximizing network as well as other counterfactual networks is discussed below and the details can be found in Appendix C.

\textsuperscript{45}It should be emphasized here that the evidence described above does not imply that the NEN led to stronger convergence in China. The goal of this paper is to quantify the effect of transport infrastructure in India, but it would be possible to apply this to China and analyze whether and how the NEN affected spatial convergence.
Figure 2: Percent increase in GDP generated by the GQ

The map shows the boundaries of Indian districts. Darker areas represent higher percentage difference in GDP generated by the GQ relative to the network without the GQ.

6.2.1 Algorithm to Design Income-Maximizing Network

Based on the locations of the targeted 68 Indian cities and the road construction costs between them (equation 14), I build a counterfactual network that connects these cities in a way that approximately maximizes aggregate net income in the general equilibrium gravity model of Section 4. 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, and 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 context has to the best of my knowledge not been solved, heuristic algorithms can be used to balance the gains in income against construction costs to approximate the optimal network.

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 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.\footnote{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. Gastner (2005) discusses various heuristic algorithms to design transport networks, including the iterative procedure that is used here.}

The algorithm used here is an iterative procedure that starts from the full network and then removes and adds links based on their effects on net income until no further improvements are possible. I show below that the effects are robust to using the empty network, the GQ, or random networks as starting points. While these various starting points make it less likely that I find only a local optimum, there is no guarantee that the solution is the global optimum. However, the gains from the resulting network can be viewed as a lower bound for the welfare gains that are possible from the optimal transport network design in a general equilibrium gravity model. The individual steps are discussed below and in Appendix C.

**Baseline Network Design** Starting from the full network, the algorithm removes each link individually from the network and recomputes the resulting shortest paths and trade costs among all 68 cities. It then uses the general equilibrium model to compute the effect on aggregate income (as described in Sections 4.4 and A.8) for each link. Using data on the topography, I can approximate the construction costs and thus calculate the change in aggregate real income net of construction costs for each link. This allows me to identify the links with the smallest (most negative) net effects.

The algorithm first reduces the network by removing the 5\% links with the most negative effect on net income.\footnote{\label{fn:5}I remove the 5\% least beneficial links together rather than only one link in each iteration in order to reduce computation time. Section C.2 in the appendix discusses a case where only one link is removed in each iteration.} After each round of removing links, the algorithm then searches for links that could be added to increase net income, or are necessary to ensure that all 68 cities are connected (this is relaxed below). The procedure then starts again from the beginning and removes and adds links gradually until no gains in net income are possible. The resulting network is shown in Figure 3a. As we can see, this is a substantially larger network than the GQ and it would cost more than eight times...
The top row shows networks designed with the algorithm that approximately maximizes net income. Figure 3a has the constraint that all 68 cities are connected, while Figure 3b does not impose this. The figures in the bottom row show the minimum spanning tree (Figure 3c) and the network based on the additional Chinese strategy of building rays from the capital city and corridors (Figure 3d). Also see Appendix Section C.3.

as much. Before discussing the effects of this network, I compare it to two alternative designs.

**Alternative Network Designs in the General Equilibrium Model** The network described above strictly implements the Chinese policy of connecting all targeted cities and the network may be too large in terms of net aggregate income. I therefore also find the approximation of the income-maximizing network without the constraint that
all 68 cities have to be connected. Using the above procedure and starting from the full network, the algorithm adds and removes links until the marginal costs and benefits of roads are approximately equalized. The resulting network, shown in Figure 3b, is still more than seven times larger than the GQ. Five of the 68 targeted cities are not connected, but the overall structure of the network is similar.48

I compare the above approximation of the income-maximizing network (without the constraint that all 68 cities are connected) to the solution when starting from the empty network and sequentially increasing it. The result is shown in Appendix Figure A4a and the structure is similar to Figure 3b. While the overlap is not perfect and there is no guarantee that the globally optimal network shares the same links, the similarity of the networks resulting from the two opposite starting points is reassuring. This approach of using an iterative procedure that starts from the full and the empty network is related to Jia (2008) and Antras et al. (2016). They show that in their case the global optimum is within the bounds obtained from sequential reduction and sequential increase of the network, but the conditions for these bounds do not hold here.49 To further test the robustness of the algorithm, I also use 100 random networks as starting points and Figure A5 shows how the number of links, construction costs, income, and net income converge. While I cannot formally put a bound on the difference between the solution of the heuristic algorithm and the global optimum, the random starting points converge to similar net income levels.50

An alternative approach to compute the optimal transport network in a spatial equilibrium model is presented by Fajgelbaum and Schaal (2017) and they apply their framework to the European road network. They consider a case where road investments on a given underlying graph are continuous and transport is subject to congestion, which allows them to obtain the globally optimal transport network. In my case, the decision where modern highways are built is discrete and this leads to a complex combinatorial problem that I solve with the heuristic algorithm. This approach can be applied to a standard spatial equilibrium model and in a context where planners face a discrete decision regarding where to build a new transport infrastructure network,

---

48 In Section 6.2.5, I also consider alternative ways of connecting intermediate-sized cities in India that do not rely on the general equilibrium model. These networks are shown in Figures 3c and 3d.
49 This depends on the complementarities between different links. In my context, a highway can be a complement or substitute to others.
50 I also use the GQ as the starting point and then sequentially increase the network.
instead of the continuous choice of how much to invest on each link of a given underlying graph. The extensive margin may be particularly relevant in applications in emerging and developing countries that plan new types of infrastructure networks or where the initial network is sparse.

6.2.2 Aggregate Effects of the Counterfactual Infrastructure

The second row in Table 2 shows the aggregate effects of the counterfactual network that connects all 68 cities in a way that approximately maximizes net aggregate income. GDP in 2012 would have been 3.46% higher with this counterfactual network than with the GQ. Based on the topography, the counterfactual network is predicted to be more than eight times as expensive to build as the GQ, but it would imply annual net gains (relative to the GQ) of 2.77% of GDP, roughly USD 30 billion.\footnote{Including interests until 2012, the GQ cost approximately USD 6.2 billion (1999 prices) and the annual cost of capital and maintenance is therefore USD 1.05 billion. The counterfactual network is 8.24 times as costly as the GQ, i.e. USD 51.09 billion. The counterfactual is assumed to replace the GQ, hence the cost net of the actual network is USD 44.89. This implies an annual cost of capital and maintenance of USD 7.63 billion.}

I then consider the counterfactual network without the constraint that all 68 cities must be connected, i.e. the marginal costs and benefits are approximately equalized. This counterfactual network, shown in Figure 3b, is similar but somewhat smaller than the previous network and it does not strictly implement the Chinese policy because five intermediate-sized cities are not connected. As shown in the third row of Table 2, the aggregate effect on income (before costs) is 3.44 percent of GDP. The net effect is 2.79 percent of GDP and therefore, as expected, larger than the constrained network that connects all 68 cities.

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 that make road construction more expensive when implementing the counterfactual network in different parts of the country. However, given the magnitudes, it seems...
unlikely that this would overturn the overall result. 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, I can estimate the model based on observed time variation in the transport network in India and then design and evaluate the counterfactual network in the same context.

6.2.3 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 Figure A6 in the appendix). This area would become better connected by the counterfactual network and experience increases in market access and higher income. Figure 4 shows these distributional effects of the counterfactual relative to the GQ.

6.2.4 Spatial Convergence with Counterfactual Infrastructure

As observed above, the counterfactual network that is designed to maximize aggregate net income benefits districts with low and intermediate light density. This is relevant for the spatial convergence patterns and it is particularly important in India because

Furthermore, I consider in Appendix C.4 a counterfactual network that approximately maximizes aggregate net income with the constraint that the construction costs do not exceed the budget that the Indian government planned for the first two phases of the National Highway Development Project.
Figure 4: Percent increase in GDP from replacing GQ with counterfactual network that connects 68 cities

The map shows the boundaries of Indian districts. Darker areas have higher percentage difference in GDP generated by replacing the GQ with the counterfactual network that connects all 68 cities in a way that approximately maximizes income.

It appears to have relatively low spatial convergence. Column 3 in Table 3 shows the results from an unconditional convergence regression where the dependent variable is the log difference in light density predicted by the counterfactual network. The point estimate suggests that the counterfactual network leads to stronger convergence than the GQ (which is shown in column 1).

6.2.5 Counterfactual Networks Without General Equilibrium Model

I also consider alternative ways of implementing the Chinese policy of connecting intermediate-sized cities in India that do not rely on the general equilibrium model. This includes the least-cost network (Figure 3c) and an ad hoc design based on additional policies by the Chinese government to construct a certain number of rays and corridors (Figure 3d). The results show that the optimal network can achieve substantially higher aggregate net income than these alternative networks.

The design of these networks are discussed in more detail in Appendix C. There, I also analyze the effects of other planned parts of the NHDP and of the income-maximizing network with the budget of the planned highways. Furthermore, I calculate the income-maximizing network with the constraint that the budget does not exceed the least-cost network.
The least-cost network relies only on the location of the cities and the topography. The Kruskal algorithm (Kruskal, 1956) finds the minimum spanning tree that connects all targeted cities at the lowest cost, without taking into account the gains from the better infrastructure.\textsuperscript{54} As shown in the third column of Table 2, the least-cost network in the aggregate implies about a 0.24% higher income than the GQ, or USD 2.63 billion per year. It would cost USD 2.31 billion per year and imply an annual net gain of about 0.03% of GDP, or USD 0.31 billion, when replacing the GQ. The least-cost network therefore is only slightly better than the GQ in terms of aggregate net income and the other counterfactuals yield substantially larger gains. However, the distributional consequences, shown in Figure A7, are qualitatively similar to the previous counterfactual networks. The least-cost network connects all intermediate-sized cities and thus reaches some of the previously lagging regions that did not benefit from the GQ.

The network with a given number of rays and corridors as shown in Figure 3d represents an example of a network that planners could have designed without using a general equilibrium model or local income data, instead only relying on the location of the intermediate-sized cities and connecting them with rays and corridors. Table 2 shows that the ad hoc network based on rays and corridors would generate 2.1% higher income in 2012 relative to the GQ and yield annual net gains of 1.71% of GDP, or USD 19 billion. Comparing this ad hoc network to the income-maximizing network design illustrates that the latter yields substantially larger gains. Figure A8 shows that the distributional implications of the ad hoc network are qualitatively similar to the approximation of the income-maximizing network. In particular, the lagging regions that did not benefit from the GQ would gain from the network that connects intermediate-sized cities with rays and corridors.

7 Robustness

This section discusses the robustness of the results to growth prior to the GQ, light emissions from roads, alternative GDP measures, international trade, different values of the trade elasticity, and population growth.

\textsuperscript{54}See Faber (2014) for an application of this network design as an instrument for the Chinese highway network.
7.1 Trends in District Growth Prior to Road Investment

The identification strategy used in Section 5.2 relies on the assumption that non-nodal districts were 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. 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 1999, 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. In none of the specifications is the estimate significant. All point estimates are negative, suggesting that, if anything, locations with weaker growth were more likely to gain better market access.

7.2 Excluding Light Pixels that Overlap with the GQ

The satellite images of night lights show clearly that the areas along roads are substantially brighter than other parts of the country. One concern when using the night lights data to estimate the effects of transport infrastructure on income may therefore be that light emissions of the road overstate its effect. As a further robustness check, I exclude pixels from the light image that overlap with the GQ. This addresses the concern that light emissions from the road may overstate income, but it has the downside that a part of the economic activity is taking place along the road and is thus excluded. As shown in Table A3 in the appendix, the estimates are similar to the baseline.
7.3 Predicting GDP Based on Lights

There is no official GDP data available at the district level for my sample period, but the Indian Planning Commission published the data for the years 1999 to 2004. This data can be used to estimate the relationship between GDP and light and then predict GDP for the later years based on the light data.\(^{55}\) The advantage of this approach (instead of directly using the light data) is that the estimate for \(\beta\) can be compared to the prediction from the model.\(^{56}\) As shown in Appendix Table A4, using the prediction as the dependent variable and in the market access measures in Equation (11) yields a point estimate for the elasticity \(\beta\) of 0.245, although it is imprecisely estimated with a standard error of 0.171. This point estimate, which is for the elasticity of real income (not light as in the baseline) with respect to market access, can now be compared to the elasticity predicted by the model for reasonable parameter values. In particular, using a capital share of 0.3 and a trade elasticity of 8.2, the model predicts a \(\beta\) around 0.27. The estimate of 0.245 is therefore comparable but somewhat lower than the one predicted by the model for these parameter values. Appendix Table A5 shows the aggregate effects from the counterfactuals when the estimate for \(\beta\) of 0.245 is used in the model. The magnitude of the effects is slightly larger but comparable to the baseline results.\(^{57}\) While this is reassuring, the caveat with this approach is that it uses a predicted value for income based on the cross-sectional relationship between district-level GDP and light.

7.4 District-Level GDP Data

While official district-level GDP is not available for the relevant sample period, a commercial data provider, Nielsen, constructs estimates for district-level GDP by combining

\(^{55}\)There are some missing values and the sample appears to be too short to estimate the relationship between GDP and light in a long difference specification. I therefore estimate the elasticity of income with respect to GDP based on the cross-section in 1999. The estimated elasticity in the cross section is 0.45 (with a robust standard error of 0.0325) and thus somewhat higher than the estimate of 0.3 from the long differences regression in Henderson et al. (2012).

\(^{56}\)In the model, \(\beta\) represents the elasticity of real income with respect to market access, but light is used as a proxy for income both in the estimation and in the counterfactuals. Since the relationship between income and light is not linear and income also enters the market access measures, the interpretation of \(\beta\) is different when light is used as a proxy for income.

\(^{57}\)Note that using an elasticity of \(\beta = 0.27\) as predicted by the model instead of 0.245 would lead to somewhat larger aggregate effects as well.
ing various sources such as the Annual Survey of Industries, the National Sample Survey, and the Economic Census. The district-level GDP data is available on an annual basis since 2001. While Nielsen provides real GDP data, it states that it is based on a state-level (and in some cases national-level) deflator, implying that district-level differences in inflation might not be accounted for appropriately. Furthermore, the estimates depend on various assumptions on how to combine the different data sets and the quality of the resulting GDP estimates has not yet been extensively tested. These are important disadvantages compared to the light data, which is a measure of real income that does not depend on the statistical procedure. But the advantage of the Nielsen data is that it attempts to directly measure district-level GDP and is not subject to some of concerns with the light data such as top coding or light emissions from roads as discussed in Section 7.2. Furthermore, there is no reason to expect that measurement error in the GDP data and in the light data are highly correlated and it is therefore useful to consider both measures.

Table A6 shows the elasticity of income with respect to market access when using the district-level GDP data from Nielsen. The results suggest an elasticity of 0.566 for the preferred specification in column 3. It is reassuring that there is a statistically and economically significant and positive effect when using this alternative data. The estimate is larger than the one obtained in Section 7.3 based on the predicted GDP data, but the elasticity predicted by the model for reasonable parameter values, 0.275, is not rejected.

Table A7 shows the aggregate effects of each network when using the elasticity of income with respect to market access from column 3 of Table A6. The results suggest that GDP in 2012 would be 8.1% lower if the GQ had not been built, which is a substantially stronger effect than the ones based on light in Tables 2 and A5. The aggregate effect of the counterfactual income-maximizing network is 7.4% and the magnitudes of the effects with the Nielsen GDP data overall seem surprisingly large. However, the fact that the results are stronger than with the light data provides evidence against the concern that the light data might overstate the effect of transport infrastructure due to direct light emissions of roads or other factors specific to the light data.
7.5 International Trade

The model and the empirical analysis so far have focused on the domestic Indian economy and did not consider the role of international markets. In this section, I check whether the results are robust to including international trade flowing through major Indian ports. The Indian government publishes data on traffic handled at the 12 major ports in India in each year, reporting tons of loaded and unloaded traffic, separately for coastal traffic and overseas traffic (see Appendix B.4). I use the total of exports and imports of overseas traffic in 1999 and 2012 (averaging over three years). Similar to the approach in Donaldson and Hornbeck (2016), I then inflate the income of the districts that host the ports by the international trade flows. The results are shown in Appendix Table A8. The estimates don’t change substantially when this alternative income measure is used.

7.6 Alternative Values for the Trade Elasticity

When solving the system of equations in (7) numerically to obtain the general equilibrium market access measures, it is necessary to choose a value for the trade elasticity parameter $\theta$. Donaldson and Hornbeck (2016) obtain an elasticity of 8.2 and I used this value for the baseline. However, alternative values for the trade elasticity have been found in the literature as well (see footnote 29). Appendix Table A9 reports the estimated effect of market access on light with values for the trade elasticity $\theta$ of 4, 6, 10, and 12. The point estimates for the preferred specifications in columns 3 range from 0.481 to 1.181 and are generally significant at the 10 percent level. That the estimated elasticity varies in $\theta$ is not surprising given that the elasticity predicted by the model also depends on $\theta$. The aggregate effects also vary somewhat with the trade elasticity, but not as much because $\theta$ affects both $\beta$ and the market access measures.\(^{58}\) These results for the aggregate effects with $\theta$ equal to 4 and 12 are shown in Tables A10 and A11 in the appendix.

\(^{58}\)Donaldson and Hornbeck (2016) also find that the aggregate effects are not sensitive to the choice of $\theta$ over a relatively wide range.
7.7 Effects of Market Access on Population

In the conceptual framework and in the empirical analysis so far, I have abstracted from changes in population across districts. The reason for this approach is that population is unlikely to be completely mobile over the period during which the transport networks are assessed here. According to Census of India (2001), about 30 percent of the Indian population live in a different place than at birth. But out of the total number of migrants, 60 percent 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, this does not establish that population did not move in response to changes in transport infrastructure. In order to test this directly, I regress the decennial log change in each district’s population between the 2001 and 2011 census on the change in market access due to transport investments. The results in Appendix Table A12 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.

8 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 and we often lack sources of exogenous variation in infrastructure. 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 approxi-

\footnote{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. It is also unclear whether the post-treatment period until the 2011 census is long enough to estimate labor mobility.}

\footnote{My preferred specification in column 3 where I exclude five nodes yields an estimate of 0.0619. Donaldson and Hornbeck (2016) find estimates between 0.197 and 0.314 for the U.S. for a longer time period and they also report that these estimates are below the prediction for their preferred parameters.}
mately maximizes income. I combine the theoretical framework with satellite data and geographically referenced information to measure income, terrain features, and road infrastructure at a high spatial resolution.

The findings suggest that the actual network, the GQ, led to large positive 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 and failed to converge would benefit from the counterfactual network that integrates regions that have intermediate-sized cities. The results suggest that the counterfactual implies stronger convergence across districts.

The implications of the findings above may extend to other countries. The theoretical framework allows to quantify 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. Data on geographic and economic characteristics based on high-resolution satellite images is commonly available. Therefore, with the data and methods applied in this paper, the effects of infrastructure projects and the predictions of the benefits from new and optimally designed networks could be considered in many other settings. Furthermore, the above results from the counterfactual network are based on an algorithm that has as the objective to maximize aggregate real income net of construction costs. In future work, it would be interesting to consider how distributional objectives can be taken into account directly in the design of the network.

References


Ernst and Young 2013. “Infrastructure 2013: Global Priorities, Global Insights”.


Harral, Clel, Jit Sondhi, and Guang Zhe Chen 2006. “Highway and Railway De-


KPMG 2013. “Logistics Games Changers: Transforming India’s Logistic Industry.”

Limao, Nuno and Anthony J. Venables 2001. “Infrastructure, Geographical Disad-


Roberts, Mark, Uwe Deichmann, and Bernard Fingleton, Tuo Shi 2012. “Evaluating China’s road to prosperity: A new economic geography approach.” Regional
World Bank 2015. “New Rail Corridors in India Cut Harmful Gasses While Boosting Speed.”
ONLINE APPENDIX

Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development

Simon Alder

Section A of this appendix presents the model details. Section B discusses the data sources and the data preparation. Section C explains the algorithm to approximate the income-maximizing network. Section D includes additional tables and figures.

A Model Details

This section provides a detailed discussion of the model presented in Section 4. The framework is based on Donaldson and Hornbeck (2016) and Eaton and Kortum (2002).

Donaldson and Hornbeck (2016) derive a reduced form expression for the impact of railroads on land values from general equilibrium trade theory. I adapt their framework to a version which can be estimated with satellite data on night lights, thus making it suitable for my estimation and counterfactual analysis across 636 Indian districts. Furthermore, I focus on the case with immobile labor as this is the more realistic assumption during the 12-year period which I consider.

The basic setup is a trade model as in Eaton and Kortum (2002) with the immobile production factors land and labor and the mobile factor capital. The economy consists of many trading regions (Indian districts), where the origin of a trade is indexed by $o$ and the destination by $d$.

A.1 Preferences

Consumers have CES preferences over a continuum of differentiated goods indexed by $j$,

$$U_o = \left( \int_j x_o(j) \frac{\sigma - 1}{\sigma} d j \right)^{\frac{\sigma}{\sigma - 1}},$$

where $x_o(j)$ is the quantity consumed of variety $j$ by a consumer in district $o$ and $\sigma > 0$ is the elasticity of substitution between goods. Consumers in location $o$ maximize $U_o$. 
subject to
\[ \int_j p_o(j)x_o(j)\,dj = y_o \]
where \( y_o \) is income per capita in district \( o \). This yields a demand function for variety \( j \) equal to
\[ x_o(j) = \frac{y_o}{P_o} \left( \frac{p(j)}{P_o} \right)^{-\sigma}, \]
where \( P_o \) is the a CES price index of the form
\[ P_o = \left( \int_j p_o(j)^{1-\sigma} \,dj \right)^{\frac{1}{1-\sigma}}. \]

Indirect utility of a consumer who has income \( y_o \) and faces prices \( P_o \) then is
\[ V(P_o, y_o) = \frac{y_o}{P_o}. \]

A.2 Production Technology

Each district produces varieties with a Cobb-Douglas technology using land (L), labor (H), and capital (K),
\[ x_o(j) = z_o(j) \left( L_o(j) \right)^\alpha \left( H_o(j) \right)^\gamma \left( K_o(j) \right)^{1-\alpha-\gamma}, \]
where the amounts of land and labor in \( o \) are fixed but capital is mobile across districts. \( z_o(j) \) is an exogenous productivity shifter as explained below. The production function implies marginal costs
\[ MC_o(j) = \frac{q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma}}{z_o(j)}, \]
where \( q_o \) is the land rental rate, \( w_o \) is the wage, and \( r_o \) is the interest rate. Following Eaton and Kortum (2002), each district draws its productivity \( z_o(j) \) from a Fréchet distribution with CDF
\[ F_o(z) = Pr[Z_o \leq z] = \exp(-T_o z^{-\theta}), \]
where $\theta > 1$ governs the variation of productivity (comparative advantage) and $T_o$ is a district’s state of technology (absolute advantage).

### A.3 Transport Costs and Prices

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 $p_{od}(j) = \tau_{od} p_{oo}(j)$.

We assume perfect competition such that prices equal the marginal costs of producing each variety:

\[
\begin{align*}
p_{oo}(j) & = MC_o(j) = \frac{q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma}}{z_o(j)} \\
p_{od}(j) & = \tau_{od}MC_o(j) = \tau_{od} \frac{q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma}}{z_o(j)} \\
z_o(j) & = \tau_{od} \frac{q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma}}{p_{od}(j)}
\end{align*}
\]

(A1)

Consumers search for the cheapest price of each variety, such that the distribution of prices is governed by the productivity distribution. Eaton and Kortum (2002) show, by substituting Equation (A1) into the distribution of productivity, that district $o$ offers district $d$ a distribution of prices

\[
G_{od}(p) = Pr[P_{od} \leq p] = 1 - F_o \left[ \frac{\tau_{od} q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma}}{p} \right] = 1 - \exp \left[ - T_o \left( \tau_{od} q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma} \right)^{-\theta} p^\theta \right].
\]

District $d$ buys variety $j$ from another district if at least one district offers a lower price than itself. The distribution of prices for what district $d$ purchases then is

\[
G_d(p) = Pr[P_d \leq p] = 1 - \prod_o \{1 - G_{od}(p)\}.
\]
Inserting for $G_{od}(p)$ yields

$$G_{d}(p) = 1 - \prod_{o} \left\{ \exp \left[ - T_{o} (\tau_{od} q_{o}^{\alpha} w_{o}^{\gamma} r_{o}^{1-\alpha-\gamma} )^{-\theta} p^{\theta} \right] \right\}$$

$$= 1 - \exp \left[ - \sum_{o} \left\{ T_{o} (\tau_{od} q_{o}^{\alpha} w_{o}^{\gamma} r_{o}^{1-\alpha-\gamma} )^{-\theta} \right\} p^{\theta} \right]$$

$$= 1 - \exp \left[ - \Phi_{d} p^{\theta} \right]$$

where the destination-specific parameter $\Phi_{d} = \sum_{o} \left\{ T_{o} (\tau_{od} q_{o}^{\alpha} w_{o}^{\gamma} r_{o}^{1-\alpha-\gamma} )^{-\theta} \right\}$ summarizes the exposure of destination $d$ to technology in other districts, factor costs, and trade costs.

Eaton and Kortum (2002) show that the price index takes the form

$$P_{d} = \mu \Phi_{d}^{-\frac{1}{2}}$$

with

$$\mu = \left[ \Gamma \left( \frac{\theta + 1 - \sigma}{\theta} \right) \right]^{\frac{1}{1-\sigma}}$$

where $\Gamma$ is the Gamma function. The rental rate for capital is equalized everywhere to $r_{o} = r$ because capital is perfectly mobile. Donaldson and Hornbeck (2016) then define

$$\kappa_{1} = \mu^{\theta} r^{-(1-\alpha-\gamma)\theta}$$

and rearrange Equation (A2) to

$$P_{d}^{-\theta} = \kappa_{1} \sum_{o} \left\{ T_{o} (\tau_{od} q_{o}^{\alpha} w_{o}^{\gamma} r_{o}^{1-\alpha-\gamma} )^{-\theta} \right\}$$

$$= \kappa_{1} \sum_{o} \left\{ T_{o} (q_{o}^{\alpha} w_{o}^{\gamma})^{-\theta} \tau_{od}^{-\theta} \right\} \equiv CMA_{d}.$$  

They refer to $CMA_{d}$ as “consumer market access” because it measures district $d$’s access to cheap goods (i.e. low production costs in supplying district and low trade costs).
A.4 Trade Flows and Gravity

Eaton and Kortum (2002) show that the fraction of expenditure of district $d$ on goods from district $o$ is

$$X_{od} = \frac{T_o(q_o^\alpha w_o^\gamma)^{r_{1-\alpha-\gamma}}}{\Phi_d}$$

$$= \frac{T_o(q_o^\alpha w_o^\gamma)^{r_{1-\alpha-\gamma}}}{\sum_o [T_o(q_o^\alpha w_o^\gamma)^{r_{1-\alpha-\gamma}}]^\theta \tau_{od}^{-\theta}}.$$

(A4)

Assuming that aggregate expenditures equal aggregate income ($X_d = Y_d$) and canceling out the interest rate, this can be rearranged to

$$X_{od} = \frac{T_o(q_o^\alpha w_o^\gamma)^{\theta}}{\sum_o [T_o(q_o^\alpha w_o^\gamma)^{\theta}]^{-1} \tau_{od}^{-\theta}}.$$

Using Equation (A3), the competitiveness of the destination’s market can be written as

$$\sum_o [T_o(q_o^\alpha w_o^\gamma)^{\theta}] = \frac{CMA_d}{\kappa_1},$$

which yields

$$X_{od} = \frac{T_o(q_o^\alpha w_o^\gamma)^{\theta}}{\sum_o [T_o(q_o^\alpha w_o^\gamma)^{\theta}]^{-1} \tau_{od}^{-\theta}}.$$

(A5)
A.5 Consumer market access and firm market access

Equation (A5) is a gravity equation with the standard features that trade increases in income of the destination and in productivity of the origin, while trade decreases in production costs, trade costs, and in consumer market access of the destination. Summing the gravity equation over destinations \( d \) and assuming that goods markets clear yields total income of origin \( o \),

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

(A6)

Donaldson and Hornbeck define “firm market access” of district \( o \) as

\[
FMA_o \equiv \sum_d \tau_{od} CMA_d^{-1} Y_d,
\]

(A7)

such that

\[
Y_o = \kappa_1 T_o (q_o^\alpha w_o^\gamma)^{-\theta} FMA_o.
\]

(A8)

\( FMA_o \) depends positively on all other destination’s income \( Y_d \) and negatively on their \( CMA_d \) (since a higher consumer market access in \( d \) implies that district \( o \) faces more competition when exporting to \( d \)). Using Equation (A8), we have

\[
\frac{Y_o}{\kappa_1 FMA_o} = T_o (q_o^\alpha w_o^\gamma)^{-\theta}
\]

which can be substituted into the definition of \( CMA_d \) to obtain

\[
CMA_d = \kappa_1 \sum_o T_o (q_o^\alpha w_o^\gamma)^{-\theta} \tau_{od}^{-\theta}
\]

\[
= \sum_o \tau_{od}^{-\theta} FMA_o^{-1} Y_o
\]

\[
CMA_o = \sum_d \tau_{od}^{-\theta} FMA_d^{-1} Y_d.
\]

(A9)

Following Donaldson and Hornbeck (2016), if trade costs are symmetric, then a solution to the Equations (A7) and (A9) must satisfy \( FMA_o = \rho CMA_o = MA_o \) for
\( \rho > 0 \) and they refer to this term as “market access”. In this setup, we then get

\[
MA_o = \rho \sum_d \tau_{od}^\theta MA_d^{-1} Y_d. \tag{A10}
\]

This system of non-linear equations captures the general equilibrium effects of the bilateral trade costs \( \tau_{od} \), because a decline in the trade costs of \( d \) enters in \( MA_d \) and will have an effect on the market access measure of \( o \).

### A.6 Measuring real market access with light

I adapt the approach of Donaldson and Hornbeck (2016) to incorporate light as a measure for real income. The starting point is Equation (A10). I then use the fact that the sum of light in a district \( o \) measures aggregate real economic activity

\[
Y_d = Y_d^r \times P_d
\]

such that

\[
MA_o = \rho \sum_d \tau_{od}^\theta MA_d^{-1} P_d Y_d^r
\]

Using the equation for the price index,

\[
P_d = (\rho^{-1} MA_d)^{-\frac{1}{\theta}},
\]

we obtain

\[
MA_o = \rho \frac{(1+\theta)}{\theta} \sum_d \tau_{od}^\theta MA_d^{-(1+\theta)} Y_d^r. \tag{A11}
\]

### A.7 Income and Market Access with Immobile Labor

Donaldson and Hornbeck (2016) proceed to solve Equation (A8) for land prices. I instead solve for real income, which in the empirical analysis I can approximate with luminosity. Using the result that firm market access equals consumer market access (up to a scale), this yields

\[
Y_o = \kappa_1 T_0(q_o^\alpha w_o^\gamma)^{-\theta} MA_o \tag{A12}
\]
where income is a function of productivity, factor prices, and market access. The constant \( \kappa_1 \) includes the interest rate, which is equalized across districts because of full capital mobility. The rental rates for the immobile factors land and labor are related to their income share according to the Cobb-Douglas production function, such that

\[
q_o L_o = \alpha Y_o
\]

\[
w_o H_o = \gamma Y_o.
\]

Using this in Equation (A12) and solving for income yields

\[
Y_o = \left( \kappa_1 T_o \right)^{\frac{1}{1+\theta(\alpha+\gamma)}} \left( \frac{\alpha}{L_o} \right)^{\frac{-\theta}{1+\theta(\alpha+\gamma)}} \left( \frac{\gamma}{H_o} \right)^{\frac{-\theta}{1+\theta(\alpha+\gamma)}} \left( MA_o \right)^{\frac{1}{1+\theta(\alpha+\gamma)}}
\]

(A13)

As mentioned above, luminosity measures real economic activity. I therefore use the relationship between the price index and market access,

\[
P_o = (\rho^{-1} MA_o)^{-\frac{1}{\theta}},
\]

(A14)

to obtain

\[
Y_o = \frac{Y_o}{P_o} = \left( \kappa_2 T_o \right)^{\frac{1}{1+\theta(\alpha+\gamma)}} \left( \frac{\alpha}{L_o} \right)^{\frac{-\theta}{1+\theta(\alpha+\gamma)}} \left( \frac{\gamma}{H_o} \right)^{\frac{-\theta}{1+\theta(\alpha+\gamma)}} \left( MA_o \right)^{\frac{1+\theta(1+\alpha+\gamma)}{1+\theta(\alpha+\gamma)}}
\]

(A15)

where \( \kappa_2 = \kappa_1 \rho^{-\frac{1+\theta(\alpha+\gamma)}{\theta}} \). After taking logs, the determinants of real income can be grouped as follows

\[
\ln \left( Y_o \right) = \frac{1}{1+\theta(\alpha+\gamma)} \ln \left( \kappa_2 \right) - \frac{\theta\alpha}{1+\theta(\alpha+\gamma)} \ln \left( \frac{\alpha}{L_o} \right) - \frac{\theta\gamma}{1+\theta(\alpha+\gamma)} \ln \left( \frac{\gamma}{H_o} \right)
\]

\[
+ \frac{1}{1+\theta(\alpha+\gamma)} \ln \left( T_o \right) + \frac{1+\theta(1+\alpha+\gamma)}{(1+\theta(\alpha+\gamma))\theta} \ln \left( MA_o \right).
\]

(A16)

This equation suggests a log-linear relationship between real income and market access, where the effect of transport infrastructure goes through this measure of market access.
access. The elasticity of income with respect to market access,
\[ \beta = \frac{1 + \theta(1 + \alpha + \gamma)}{(1 + \theta(\alpha + \gamma))\theta}, \]
can be estimated using variation in income and market access over time.

### A.8 Counterfactual Predictions

In order to predict income for counterfactual transport networks, I first compute the bilateral shortest paths and the implied iceberg trade costs for each network and then solve for the new equilibrium. This requires jointly solving Equations (A11) and (A15). However, productivity, immobile production factors, and \( \kappa_2 \) are unobserved in Equation (A15). In order to solve the model for all counterfactual networks, I first back out the unobserved terms using observed data in 2012 and then solve for equilibrium outcomes of the counterfactuals while holding this term constant.\(^{61}\) These steps are explained in more detail below.\(^{62}\)

I first solve the market access Equation (A11) with data in the year 2012, i.e. with the values of trade costs implied by the actual transport network (including the GQ) and with 2012 lights data.

Second, we observe in Equation (A15) that the rental rate of capital is included in \( \kappa_2 \). This rental rate of capital relative to the price index has to be constant because of international capital mobility, but we need to make an appropriate choice for the price index. Donaldson and Hornbeck (2016) use the price index of the main financial center and a national weighted average. Using Equation (A14), the price index can be written as a function of market access,
\[ P = \sum_d \lambda_d P_d = \sum_d \lambda_d (\rho^{-1} MA_d)^{-\theta}, \]  
(A17)

where \( \lambda_d \) are the weights. In the baseline I use the price index of Mumbai (i.e. the weights for all other districts are 0) but I have also used a national average weighted

\(^{61}\)Donaldson and Hornbeck (2016) use this approach in a setting with mobile labor where utility is equalized across locations.

\(^{62}\)The system can be solved with a standard solver in Matlab.
by districts’ real income. The rental rate of capital can then be written as

\[ r = \bar{r} \times P = \bar{r} \times \sum_d \lambda_d (\rho^{-1} MA_d)^{-\frac{1}{\theta}}, \]  

(A18)

where \( \bar{r} \) is constant and determined by world markets.

Third, using the price index and defining \( \kappa_3 \equiv \mu^{-\theta} \bar{r}-(1-\alpha-\gamma)\theta \rho^{-\frac{1+\theta(\alpha+\gamma)}{\theta}}, \) Equation (A15) can then be written as

\[ Y_o^{r} = \left( \kappa_3 T_o \right) \left[ \frac{\alpha}{E_o} \left( \frac{\gamma}{H_o} \right)^{-\theta_{\alpha}} \right] \left( \frac{P}{B_i} \right)^{-\frac{-1}{\theta}} \left( MA_o \right)^{\frac{1+\theta(1+\alpha+\gamma)}{(1+\theta(\alpha+\gamma))\theta}}. \]  

(A19)

The unobserved term \( B_i \) is assumed to be constant over the counterfactuals and I can therefore jointly solve Equations (A11), (A17), and (A19) for all counterfactuals. The effect of each counterfactual on income is then expressed relative to actual income in 2012.

Holding \( B_i \) constant at its 2012 value also assumes that productivities are constant. This implies that overall productivity levels are not directly affected by improvements in transport infrastructure. Although this is a strong assumption, we would expect that such effects are positive and reflected in the estimation of \( \beta \).

B Data Sources

B.1 Administrative Boundaries and Light Data

The district boundaries are from 2011 and the light data is provided by the Earth Observation Group (2015) of the National Geophysical Data Center of the United States. The satellite images originate from the Defense Meteorological Satellite Program (DMSP) Operational Linescan System (OLS) to detect cloud cover. The data is available from 1992 to 2012 as composites over cloud-free evenings. The raster are 30 arc second grids, spanning -180 to 180 degrees longitude and -65 to 75 degrees latitude. To derive a measure of economic activity for each district, I aggregate light within Indian district boundaries using an equal area projection. The light summary statistics of the sample
of mainland Indian districts are presented in Table A1.

B.2 District-Level GDP

The district-level GDP data for the period 1999-2004 is obtained from the Indian Planning Commission.\(^{63}\) I use the GDP series at constant prices. The district-level GDP for the period 2001 to 2013 is obtained from Nielsen.\(^{64}\) I again use the GDP series at constant prices, but it should be noted that Nielsen uses state-and national-level deflators. 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.

B.3 Transport Infrastructure and Terrain

I use geographic information system (GIS) methods to process the spatial data. Digital maps with the location of the actual Indian transport infrastructure are taken from three sources: CIESIN (2013) provides a digitized road network that includes both highways and local roads. Esri (2013) also has digitized roads but is limited to the national highway networks. These first two sources allow me to localize the current transport infrastructure in space, but they do not allow to accurately track changes over time and cannot distinguish the higher quality of today’s GQ. Therefore, I use as a third source PDF maps of the NHDP issued by the National Highway Authority of India (NHAI, 2010 and NHAI, 2013). These maps, which were digitized manually, show the location of several new highways, including the GQ and the completed parts of the North-South and East-West Corridors. The average driving speed on existing roads are taken from several transport efficiency studies. World Bank (2005) reports that the typical driving speed on the existing Indian national and state highways is between 30 and 40 km/h and I therefore assume a speed of 35 km/h for all highways built before the start of the NHDP.\(^{65}\) For areas where there are no roads reported in the digitized maps, I assume a travel speed of 10 km/h, which corresponds to the speed on unpaved roads (Roberts et al., 2012). The travel speed on the counterfactual network is

\(^{65}\)These estimates are in line with more recent numbers by KPMG (2013).
taken to be the same as for the Chinese expressways and the GQ, which according to google maps is 75 km/h. To take into account the average waiting time of trucks at state borders, I include in the digital network a cost when crossing state borders equivalent to three hours.\textsuperscript{66} For a comparison of the highway networks, the digital maps of the Chinese expressway network were obtained from ACASIAN (2013).

In order to determine the construction costs for the counterfactual roads, I need digitized information on the terrain. I use digital elevation data produced by Jarvis et al. (2008) for a measure of slope. For land cover, I use the classification by the Global Land Cover Facility (2013) at the University of Maryland Department of Geography.

\subsection*{B.4 Trade Through Major Indian Ports}

The Indian government publishes data on traffic handled at the following 12 major ports in India in each year, reporting tons of loaded and unloaded traffic, separately for coastal traffic and overseas traffic: Kamarajar, Chennai, New Mangalore, Chidambaranar, Kandla, Mormugao, Vishakhapatnam, Jawahar Lal Nehru, Paradip, Mumbai, Cochin, Kolkata and Haldia.\textsuperscript{67} For the average value per ton, I use $668, which was the average value per ton of imports to the US in the year 2000 (United States Department of Transportation, 2002). In order to calculate the ratio of trade to income, I obtain the Gross District Product in 1999 from the Indian Planning commission. This data is only available for the years 1999-2004 and it has some missing values. I impute the missing values based on the light data and then use the data in 1999 to obtain the ratio between trade and income. Based on this ratio, the light measures in districts with a major port are then scaled up accordingly.

\subsection*{B.5 Descriptive Statistics}

The descriptive statistics for the main variables market access and light growth are shown in Table \ref{tab:stats}. The sample consists of 636 districts in mainland India. Market access is shown without the GQ (the network before 2000) and with the GQ (the network in 2012). These first two market access measures are computed with light data.

\textsuperscript{66} About one fourth of the transport times of trucks between Delhi and Kolkata was spent at state border checkpoints (World Bank, 2002).

\textsuperscript{67} The data is obtained from https://data.gov.in/catalog/traffic-handled-major-ports-india.
from 1999. The third variable shows the log change in market access implied by the GQ when holding light fixed in 1999. The next market access measure is based on the network with the GQ in 2012, using the light data from the same year. The log change in market access with variable light is 33% and with constant light it is 8%. The log change in light was 56% over the entire period. It should be noted that the calibration of the satellites differ such that it is difficult to directly compare light values over time. However, this does not affect the empirical analysis because potential differences in the calibration are absorbed by the year fixed effects.

C Construction of Counterfactual Networks

This appendix section discusses how the various counterfactual networks are constructed.

C.1 Approximation of Income-Maximizing Network

This section presents the heuristic algorithm to construct the counterfactual network among the targeted 68 cities. The objective is to maximize aggregate real income net of road construction and maintenance costs.

Step 1: Initial Network I start with a fully connected network, i.e. all 68 cities that are targeted by the policy are assumed to be connected with a direct modern highway link of the quality of the GQ. I assume that travel times are symmetric, such that I need to consider 2244 links.

Step 2: Removal of Each Link I then iterate over each of the 2244 links and turn the modern highway link (driving speed 75 km/h) into a conventional highway (driving speed 35 km/h). For each resulting network, I recompute the bilateral shortest paths as well as the resulting equilibrium income. These steps are explained in detail below.

Step 2a: Shortest Paths I recompute all bilateral driving times based on a shortest path algorithm. I do this with a Dijkstra algorithm that computes the shortest path among nodes in the network after turning one link into a regular highway. For each
version of the network (i.e. for each removal of a link), I obtain a matrix of the bilateral driving times among all nodes.\footnote{Different from Section 5.1.2, driving times here are computed using a version of Dijkstra’s algorithm based on a graph instead of the Fast Marching Algorithm on a cost surface. The implementation of Dijkstra’s algorithm is from the Octave Network Toolbox, which can be obtained at http://aeolianine.github.io/octave-networks-toolbox.}

**Step 2b: General Equilibrium Income** For each network and implied iceberg trade costs, I recompute the general equilibrium income by jointly solving market access, income, and the price of capital as discussed in Sections 4.4 and A.8. This yields a measure of aggregate real income, \( Y^r_n = \sum_o Y^r_{o,n} \), where \( n \) denotes the version of the network.

**Step 2c: Net Change in Aggregate Income** The net effect of adding or removing a highway link relative to an initial network is the change in aggregate real income per year, net of annual capital and maintenance costs,

\[
\Delta W_{n,n-1} = Y^r_n - Y^r_{n-1} - b_{n,n-1},
\]

where I use \( n \) to denote the new network and \( n - 1 \) the previous network. \( b_{n,n-1} \) is the annual capital and maintenance cost of the link that is added or removed (\( b_{n,n-1} \) is negative when a link is removed).

One challenge is that the construction costs in the data are in units based on the topographical features as in Equation (14). In order to compute the net gain in the objective function, I calculate the ratio of the GQ’s construction costs in USD (from Ghani et al., 2016) to its construction costs based on the topography (by drawing the GQ on the construction cost surface). This ratio allows me to express any road segment’s construction costs in USD. Then, assuming that the cost of capital is 5\% and the maintenance cost is 12\% of the construction costs, I calculate the annual cost of each road segment. The resulting change in net income from a link then is

\[
\Delta W_{n,n-1} = Y^r_n - Y^r_{n-1} - 0.17 \omega c_{n,n-1},
\]

where \( \omega \) is the ratio of construction costs of the GQ in USD and to the construction costs based on the topography, and \( c_{n,n-1} \) is the change in the topography costs.
Step 3: Removal of Least Beneficial Links  In the above step 2, I calculated the net change in income from removing each modern highway link. In the first iteration, starting from the fully connected network, there are 2244 links to be calculated. I then select those links that lead to the largest increase in net income when being removed. For computational reasons, I delete the 5% least beneficial links at once.

Step 4: Add Most Beneficial Link  It is possible that the sequential elimination of links changes the network such that there are potentially new links with a positive net benefit. I therefore include a step that is essentially the reverse of steps 2 and 3, i.e. it iterates over all possible new links and then adds the most useful one if it implies an increase in net income.\(^69\)

Step 5: Connecting all 68 Cities  The above steps may imply that some of the 68 cities are not connected because the marginal costs of connecting them exceeds the marginal benefit. For my baseline counterfactual that connects all 68 cities, I therefore add a step that connects the isolated cities by using the most beneficial link.

I then go back to step 2 and iterate through the algorithm until neither removing nor adding a link (while ensuring that all 68 cities are connected) leads to an increase in net aggregate income. The resulting network, shown in Figure 3a, is then used as an approximation of the income-maximizing network. I combine this network with the existing network of conventional highways to obtain a national transport network. I then use the fast marching algorithm in order to compute the driving times among all 636 Indian districts, as described in Section 5.1.2.

C.2 Alternative Versions of the Network Design Algorithm

I consider a number of alternative versions of the network design algorithm.

Equalizing Marginal Costs and Benefits  The links added in step 4 to ensure that all cities are connected to the network may imply a negative net income change. I therefore also compute the network without this step, i.e. the algorithm stops only

\(^69\) Gastner (2005) also includes this step to add links.
when no links can be added or removed to increase net income, i.e. marginal costs and benefits are approximately equalized. Five cities would not be connected, mostly in periphery regions in the north and east. This network is shown in Figure 3b and in the counterfactual analysis this network is used as an approximation of the unconstrained maximization of net aggregate income.

**Starting from Empty Network or from the Actual Network** As discussed in the main text, one obvious caveat of this heuristic approach is that there is no guarantee that it finds the globally optimal network. To partly address this concern, I also compare the resulting network when starting from the empty network or from the GQ and sequentially adding and removing links. The result for the empty network as starting point is shown in Figure A4a and the result for the GQ as starting point is similar.

**Starting from Random Networks** I also start the algorithm from random networks instead of from the full or empty network. I generate 100 random networks and use them as starting point in the algorithm to find the income-maximizing network. Figure A5 shows how the number of links, construction costs, income, and income net of construction costs converge in each iteration.

**Deleting Only One Link in Each Iteration** In the baseline version of the network design algorithm presented above, I delete 5% of the least-beneficial links in each iteration. I then check whether there are links that can be added in order to increase aggregate net income and, if applicable, then I add the most beneficial one. The reason for deleting 5% at once is the computational cost that makes it difficult to compute and compare several different networks as I do here. However, it is feasible to delete only one link in each iteration when changing the way in which the shortest paths are computed. In particular, I restrict the set of paths that are recomputed after a link is removed by only considering those paths that in the previous iteration actually used the link that is being removed. Since only a small fraction of paths use a given link,

---

70 The comparison of the solution when starting from the full network and from the empty network is related to Jia (2008) and Antras et al. (2016).

71 The computation time is to a large extent due to the many links that need to be considered, which includes solving for the optimal paths and general equilibrium incomes. The system of equations to obtain income converges and I use the solution of the previous network as a starting point.
this reduces the time it takes to compute the shortest path. This is evidently only possible in the step that reduces the network. I therefore use a version of the algorithm that removes links until no further improvements are possible, but without step 4 that checks whether links can be added. This is computationally feasible and yields as a solution a network with 124 (symmetric) links. When the same algorithm (i.e. only removing links) is used when 5% of the least-beneficial links are removed at once, then the resulting network has 118 links. Hence, in this example the algorithm that deletes the 5% of least-beneficial links at once in each iteration leads, not surprisingly, to a somewhat smaller network than if only one link is removed in each iteration. The observation that the difference is rather small is reassuring, although there is no guarantee that this is also the case in the baseline where links are added as well.

C.3 Alternative Network Designs

Least-Cost Network  The minimum spanning tree is the cheapest way to connect a given set of nodes in a network. Faber (2014) uses this network as an instrument to predict the actual highway network among the cities targeted by the Chinese policy. This least-cost network can be computed using the Kruskal algorithm (Kruskal, 1956), which uses as inputs all bilateral construction costs and finds the minimal links needed to connect all cities at least once to the common network. The resulting counterfactual network is shown in Figure 3c. It represents the cheapest way to formally fulfill the Chinese policy objective (connecting cities which have a population above 500’000 or are provincial capitals) in India. However, the resulting network does not take into account the benefit from building a road and only attempts to minimize the construction costs.

Counterfactual Network Based on Additional Policy to Construct ‘Rays and Corridors’  An alternative way to replicate the NEN in India is to use additional policies that the Chinese government specified. In particular, it stated that the targeted cities should be connected with rays out of the capital city and with horizontal and vertical corridors. Figure 3d shows an example where these additional criteria are implemented. This is a network that planners would have been able to design without a general equilibrium model and only using the location of the intermediate-sized cities.
The resulting network resembles the structure of the Chinese network, but the disadvantage is that there are various ways of connecting the targeted cities with rays and corridors and the example presented here is constructed in an ad hoc way.

**Other Planned Highway Projects** Policy makers seem to have been aware of the need to also connect other parts of the country and the NHDP did include plans for other highway connections besides the GQ. In particular, the government planned as part of the National Highway Development Project also the NS-EW corridors, which cross through regions that were not reached by the GQ. However, these other projects were delayed and only partially completed (see also Ghani et al, 2015). A complementary exercise is therefore to compare the effects from the completed parts of the NS-EW to the effects from the counterfactual network that follows the Chinese strategy. Figure A9 shows that adding the completed parts of the NS-EW to the GQ benefits additional regions, but the gains are not as widely distributed as with the counterfactual that connects intermediate-sized cities. I consider below what the effects would be of an approximately optimal network that has the same budget as the actually planned highway networks, i.e. the GQ and NS-EW corridors.

**C.4 Approximation of Income-Maximizing Networks for a Given Total Costs**

As an alternative way of designing the network, I can also compute a network that has a given total cost and choose the links in an optimal way given this budget constraint.

**Same Cost as Minimum Spanning Tree** The least-cost network leads to a lower increase in net income than the approximation of the income-maximizing network. In order to investigate whether this is mostly due to the smaller size of the network or its structure, I use the iterative procedure to compute a network that has the same costs as the least-cost network and maximizes the net income gain under this constraint.

Figure A4b shows the resulting network. Since the minimum spanning tree is the cheapest way of connecting the 68 cities in one common network, a change in the network (while keeping costs constant) must imply that some cities are not connected anymore. Hence, by weighting income gains against road construction costs, the algo-
algorithm tends to provide better connections to locations with larger effects on aggregate income, at the cost of not including certain cities in the network, as can be seen in Figure A4b. However, the resulting network more closely resembles a transport network than the least-cost network.

**Same Budget as Phase 1 and 2 of the National Highway Development Project**

The total budget for the GQ and the NS-EW corridors was almost USD 14 billion (Ghani et al., 2016). One interesting counterfactual exercise is therefore to ask what would be the income-maximizing network that could be built at that cost. This also addresses the question of whether the Indian government could expect being able to raise the funding that would be necessary to finance the counterfactual network, as discussed in Section 6.2.2. The counterfactual network that approximately maximizes net aggregate income under the constraint that the total costs do not exceed the planned budget is shown in Figure A4c. Table 2 shows that there would be an annual net gain of 1.45% of GDP from this network.

**Same Cost as Network with Rays and Corridors**

Figure A4d shows a network that has about the same overall length as the network in Figure 3d (based on an ad-hoc way of constructing rays and corridors as suggested by the additional Chinese policy), but it approximates the constrained income-maximizing network using the algorithm described above. As is clear from the comparison, the structure of the two networks is different despite the fact that they are constructed to connect the same cities. When balancing aggregate income and road construction costs, the algorithm tends to build more connections to locations with larger effects on aggregate net income.

**C.5 Mapping the Aggregate Economy to the Nodes of the Network**

A further challenge is how to map the aggregate Indian economy to the nodes of the network. The changes in income are computed only for the 68 cities that are targeted by the Chinese policy. By using the Indian aggregate GDP in 1999 for the total income of the 68 cities, I assume that they are a good representation of the overall Indian economy. An alternative approach would be to approximate total income by the share of national GDP that is due to the 68 cities. The local GDP data is not available to
calculate GDP of the 68 cities or their total, but one could approximate the share with the light data (i.e. comparing the share of light from the 68 cities to national light in 1999). However, this is likely to severely underestimate the benefits of the national highway network, since the other areas besides the 68 cities would be assumed to not have any gains from the network, while the network actually does reach a large part of the country.
### D Additional Tables and Figures

Table A1: Descriptive Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market access without GQ, light in 1999</td>
<td>636</td>
<td>863.92</td>
<td>218.07</td>
</tr>
<tr>
<td>Market access with GQ, light in 1999</td>
<td>636</td>
<td>933.87</td>
<td>233.61</td>
</tr>
<tr>
<td>Log difference in market access from GQ, constant light</td>
<td>636</td>
<td>.08</td>
<td>.06</td>
</tr>
<tr>
<td>Market access with GQ, light in 2012</td>
<td>636</td>
<td>1203.12</td>
<td>290.83</td>
</tr>
<tr>
<td>Log difference in market access from GQ, variable light</td>
<td>636</td>
<td>.33</td>
<td>.07</td>
</tr>
<tr>
<td>Log difference in light</td>
<td>630</td>
<td>.56</td>
<td>.54</td>
</tr>
</tbody>
</table>

The table shows the summary statistics for the general equilibrium market access measures and light data. The market access measures are derived from equations (7) and (12) where real income is proxied by light in 1999 or 2012. The light observations for 1999 are averages over 1998, 1999, and 2000. The light observations for 2012 are averages over 2011, 2012, and 2013. The trade elasticity is set to $\theta = 8.2$. The sample consists of 636 Indian mainland districts in 2011.
Table A2: Light growth prior to road investment

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Market Access</td>
<td>-0.446</td>
<td>-0.519</td>
<td>-0.503</td>
<td>-0.434</td>
</tr>
<tr>
<td></td>
<td>(0.365)</td>
<td>(0.342)</td>
<td>(0.343)</td>
<td>(0.355)</td>
</tr>
<tr>
<td>Excluded nodal districts</td>
<td>None</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Weighting</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Standard errors</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Cluster</td>
</tr>
<tr>
<td>N</td>
<td>620</td>
<td>607</td>
<td>606</td>
<td>606</td>
</tr>
<tr>
<td>Rsq.</td>
<td>0.388</td>
<td>0.385</td>
<td>0.386</td>
<td>0.354</td>
</tr>
</tbody>
</table>

The table shows 2SLS estimates of the elasticity of pre-investment light with respect to market access. The dependent variable is the logarithm of the sum of light in each district in the years 1993 and 1998 (averaging over three years). The explanatory variable is market access computed based on Equation (7) and instrumented with the market access with constant light in Equation (12). All regressions include district fixed effects, state-year fixed effects, and controls for distance to the coast and the level of electrification in 2001 interacted with a year fixed effect. Column 1 shows the effect in the full sample and columns 2 - 5 exclude four or five nodal districts as stated in the table. Columns 1 - 3 weigh by the logarithm of initial sum of light. Standard errors (in parentheses) are clustered at the state-level except in column 5.

Table A3: Excluding from light data the pixels that overlap with GQ

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Market Access</td>
<td>0.602</td>
<td>0.639</td>
<td>0.642</td>
<td>0.652</td>
</tr>
<tr>
<td></td>
<td>(0.336)</td>
<td>(0.360)</td>
<td>(0.363)</td>
<td>(0.386)</td>
</tr>
<tr>
<td>Excluded nodal districts</td>
<td>None</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Weighting</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Standard errors</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Cluster</td>
</tr>
<tr>
<td>N</td>
<td>626</td>
<td>613</td>
<td>612</td>
<td>612</td>
</tr>
<tr>
<td>Rsq.</td>
<td>0.505</td>
<td>0.500</td>
<td>0.500</td>
<td>0.494</td>
</tr>
</tbody>
</table>

The table shows the 2SLS estimates of the elasticity of light with respect to market access. The dependent variable is the logarithm of the sum of light in each district in the years 1999 and 2012 (averaging over three years), but excluding pixels (1x1 km) that overlap with the GQ. The explanatory variable is market access computed based on Equation (7) and instrumented with the market access with constant light in Equation (12). All regressions include district fixed effects, state-year fixed effects, and controls for distance to the coast and the level of electrification in 2001 interacted with a year fixed effect. Column 1 shows the effect in the full sample and columns 2 - 5 exclude four or five nodal districts as stated in the table. Columns 1 - 3 weigh by the logarithm of initial sum of light. Standard errors (in parentheses) are clustered at the state-level except in column 5.
Table A4: Elasticity of predicted GDP with respect to market access

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Market Access</td>
<td>0.232</td>
<td>0.241</td>
<td>0.245</td>
<td>0.247</td>
<td>0.247</td>
</tr>
<tr>
<td></td>
<td>(0.163)</td>
<td>(0.169)</td>
<td>(0.171)</td>
<td>(0.183)</td>
<td>(0.151)</td>
</tr>
<tr>
<td>Excluded nodal districts</td>
<td>None</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Weighting</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Standard errors</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Robust</td>
</tr>
<tr>
<td>N</td>
<td>627</td>
<td>614</td>
<td>613</td>
<td>614</td>
<td>614</td>
</tr>
<tr>
<td>Rsq.</td>
<td>0.518</td>
<td>0.510</td>
<td>0.510</td>
<td>0.500</td>
<td>0.500</td>
</tr>
</tbody>
</table>

The table shows the 2SLS estimates of the elasticity of GDP with respect to market access. The dependent variable is the logarithm of GDP predicted by light in each district in the years 1999 and 2012 (averaged over three years). The explanatory variable is market access computed based on Equation (7) and instrumented with the market access with constant GDP in Equation (12). The GDP data in the market access measures is also predicted based on lights. All regressions include district fixed effects, state-year fixed effects, and controls for distance to the coast and the level of electrification in 2001 interacted with a year fixed effect. Column 1 shows the effect in the full sample and columns 2 - 5 exclude four or five nodal districts as stated in the table. Columns 1 - 3 weigh by the logarithm of initial sum of light. Standard errors (in parentheses) are clustered at the state-level except in column 5.

Table A5: Aggregate effects of transport networks in percent of 2012 GDP, GDP predicted based on light, $\beta=0.245$

<table>
<thead>
<tr>
<th></th>
<th>Costs</th>
<th>Income</th>
<th>Net income</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% USD</td>
<td>% USD</td>
<td>% USD</td>
</tr>
<tr>
<td>Removing GQ</td>
<td>-0.10</td>
<td>-1.05</td>
<td>-3.04</td>
</tr>
<tr>
<td>Income-maximizing network, all 68 cities</td>
<td>0.69</td>
<td>7.63</td>
<td>3.72</td>
</tr>
<tr>
<td>Income-maximizing network, unconstrained</td>
<td>0.65</td>
<td>7.16</td>
<td>3.67</td>
</tr>
<tr>
<td>Least-cost network, all 68 cities</td>
<td>0.21</td>
<td>2.31</td>
<td>0.27</td>
</tr>
<tr>
<td>Rays and corridors, all 68 cities</td>
<td>0.39</td>
<td>4.31</td>
<td>2.31</td>
</tr>
<tr>
<td>Income-maximizing network, least-cost budget</td>
<td>0.20</td>
<td>2.24</td>
<td>2.10</td>
</tr>
<tr>
<td>Income-maximizing network, NHDP budget</td>
<td>0.16</td>
<td>1.74</td>
<td>1.85</td>
</tr>
</tbody>
</table>

The table summarizes the aggregate effects of the actual and counterfactual networks when GDP is predicted based on lights and $\beta$ is set to 0.245. The changes in construction costs, income, and net income due to each network are shown in percentages of GDP in 2012. 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 row and below are assumed to replace the GQ and the construction costs of the GQ are subtracted from the construction costs of the counterfactual.
Table A6: Elasticity of real GDP (Nielsen) with respect to market access

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Market Access</td>
<td>0.605</td>
<td>0.635</td>
<td>0.566</td>
<td>0.534</td>
<td>0.534</td>
</tr>
<tr>
<td></td>
<td>(0.296)</td>
<td>(0.300)</td>
<td>(0.332)</td>
<td>(0.342)</td>
<td>(0.213)</td>
</tr>
<tr>
<td>Excluded nodal districts</td>
<td>None</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Weighting</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Standard errors</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Robust</td>
</tr>
<tr>
<td>N</td>
<td>633</td>
<td>620</td>
<td>619</td>
<td>619</td>
<td>619</td>
</tr>
<tr>
<td>Rsq.</td>
<td>0.356</td>
<td>0.345</td>
<td>0.353</td>
<td>0.359</td>
<td>0.359</td>
</tr>
</tbody>
</table>

The table shows the 2SLS estimates of the elasticity of real GDP with respect to market access. The dependent variable is the logarithm of real GDP from Nielsen for the years 2002 and 2012 (averaging over three years). The explanatory variable is market access computed based on Equation (7) and instrumented with the market access with constant GDP in Equation (12). All regressions include district fixed effects, state-year fixed effects, and controls for distance to the coast interacted with a year fixed effect. Column 1 shows the effect in the full sample and columns 2 - 5 exclude four or five nodal districts as stated in the table. Columns 1 - 3 weigh by the logarithm of initial sum of light. Standard errors (in parentheses) are clustered at the state-level except in column 5.
Table A7: Aggregate effects of transport networks based on real GDP (Nielsen)

<table>
<thead>
<tr>
<th>Cost</th>
<th>Income</th>
<th>Net income</th>
</tr>
</thead>
<tbody>
<tr>
<td>% USD</td>
<td>% USD</td>
<td>% USD</td>
</tr>
<tr>
<td>Removing GQ</td>
<td>-0.10</td>
<td>-1.05</td>
</tr>
<tr>
<td>Income-maximizing network, all 68 cities</td>
<td>0.69</td>
<td>7.63</td>
</tr>
<tr>
<td>Income-maximizing network, unconstrained</td>
<td>0.65</td>
<td>7.16</td>
</tr>
<tr>
<td>Least-cost network, all 68 cities</td>
<td>0.21</td>
<td>2.31</td>
</tr>
<tr>
<td>Rays and corridors, all 68 cities</td>
<td>0.39</td>
<td>4.31</td>
</tr>
<tr>
<td>Income-maximizing network, least-cost budget</td>
<td>0.20</td>
<td>2.24</td>
</tr>
<tr>
<td>Income-maximizing network, NHDP budget</td>
<td>0.16</td>
<td>1.74</td>
</tr>
</tbody>
</table>

The table summarizes the aggregate effects of the actual and counterfactual networks. The estimation and the counterfactuals are based on the GDP data from Nielsen. The changes in construction costs, income, and net income due to each network are shown in percentages of actual national GDP in 2012 and in billion USD (1999 prices). 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 row and below are assumed to replace the GQ and the construction costs of the GQ are subtracted from the construction costs of the counterfactual.

Table A8: Including trade in 12 major ports

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Market Access</td>
<td>0.574</td>
<td>0.595</td>
<td>0.601</td>
<td>0.583</td>
</tr>
<tr>
<td></td>
<td>(0.303)</td>
<td>(0.324)</td>
<td>(0.326)</td>
<td>(0.349)</td>
</tr>
<tr>
<td>Excluded nodal districts</td>
<td>None</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Weighting</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Standard errors</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Cluster</td>
</tr>
<tr>
<td>N</td>
<td>626</td>
<td>613</td>
<td>612</td>
<td>612</td>
</tr>
<tr>
<td>Rsq.</td>
<td>0.507</td>
<td>0.500</td>
<td>0.500</td>
<td>0.494</td>
</tr>
</tbody>
</table>

The table shows 2SLS estimates of the elasticity of light with respect to market access. The dependent variable is the logarithm of the sum of light in each district in the years 1999 and 2012 (averaging three years). The explanatory variable is market access computed based on Equation (7) and instrumented with the market access with constant light in Equation (12). Exports and imports from major ports are added to the income in the districts where the ports are located to compute market access. All regressions include district fixed effects, state-year fixed effects, and controls for distance to the coast and the level of electrification in 2001 interacted with a year fixed effect. Column 1 shows the effect in the full sample and columns 2 - 5 exclude four or five nodal districts as stated in the table. Columns 1 - 3 weigh by the logarithm of initial sum of light. Standard errors (in parentheses) are clustered at the state-level except in column 5.
Table A9: Effect of Transport Infrastructure Investments with Alternative $\theta$

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Market Access</td>
<td>1.181</td>
<td>0.835</td>
<td>0.556</td>
<td>0.481</td>
</tr>
<tr>
<td></td>
<td>(0.636)</td>
<td>(0.455)</td>
<td>(0.310)</td>
<td>(0.271)</td>
</tr>
<tr>
<td>Assumed value for $\theta$</td>
<td>4</td>
<td>6</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Excluded nodal districts</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Weighting</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Standard errors</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Cluster</td>
</tr>
<tr>
<td>N</td>
<td>612</td>
<td>612</td>
<td>612</td>
<td>612</td>
</tr>
<tr>
<td>Rsq.</td>
<td>0.501</td>
<td>0.501</td>
<td>0.502</td>
<td>0.503</td>
</tr>
</tbody>
</table>

The table shows the 2SLS estimates of the elasticity of light with respect to market access. The dependent variable is the logarithm of the sum of light in each district in the years 1999 and 2012 (averaged over three years). The explanatory variable is market access computed based on Equation (7) and instrumented with the market access with constant light in Equation (12) with $\theta$ as shown in the table. All regressions include district fixed effects, state-year fixed effects, and controls for distance to the coast and the level of electrification in 2001 interacted with a year fixed effect. Standard errors (in parentheses) are clustered at the state-level.

Table A10: Aggregate effects of actual and counterfactual transport networks with $\theta = 4$

<table>
<thead>
<tr>
<th></th>
<th>Costs</th>
<th>Income</th>
<th>Net income</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% USD</td>
<td>% USD</td>
<td>% USD</td>
</tr>
<tr>
<td>Removing GQ</td>
<td>-0.10</td>
<td>-1.05</td>
<td>-3.72</td>
</tr>
<tr>
<td></td>
<td>-40.83</td>
<td>4.25</td>
<td>46.67</td>
</tr>
<tr>
<td></td>
<td>-3.62</td>
<td>4.26</td>
<td>46.78</td>
</tr>
<tr>
<td></td>
<td>-39.77</td>
<td>2.61</td>
<td>28.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income-maximizing network, all 68 cities</td>
<td>0.69</td>
<td>7.63</td>
<td>4.94</td>
</tr>
<tr>
<td></td>
<td>54.30</td>
<td>4.25</td>
<td>46.67</td>
</tr>
<tr>
<td>Income-maximizing network, unconstrained</td>
<td>0.65</td>
<td>7.16</td>
<td>4.91</td>
</tr>
<tr>
<td></td>
<td>53.93</td>
<td>4.26</td>
<td>46.78</td>
</tr>
<tr>
<td>Least-cost network, all 68 cities</td>
<td>0.21</td>
<td>2.31</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>2.91</td>
<td>0.05</td>
<td>0.60</td>
</tr>
<tr>
<td>Rays and corridors, all 68 cities</td>
<td>0.39</td>
<td>4.31</td>
<td>3.00</td>
</tr>
<tr>
<td></td>
<td>33.00</td>
<td>2.61</td>
<td>28.69</td>
</tr>
<tr>
<td>Income-maximizing network, least-cost budget</td>
<td>0.20</td>
<td>2.24</td>
<td>2.80</td>
</tr>
<tr>
<td></td>
<td>30.81</td>
<td>2.60</td>
<td>28.57</td>
</tr>
<tr>
<td>Income-maximizing network, NHDP budget</td>
<td>0.16</td>
<td>1.74</td>
<td>2.37</td>
</tr>
<tr>
<td></td>
<td>26.07</td>
<td>2.21</td>
<td>24.33</td>
</tr>
</tbody>
</table>

The table summarizes the aggregate effects of the actual and counterfactual networks. The changes in construction costs, income, and net income due to each network are shown in percentages of actual GDP in 2012 and in billion USD (1999 prices). 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 row and below are assumed to replace the GQ and the construction costs of the GQ are subtracted from the construction costs of the counterfactual.
Table A11: Aggregate effects of actual and counterfactual transport networks with $\theta = 12$

<table>
<thead>
<tr>
<th></th>
<th>Costs</th>
<th>Income</th>
<th>Net income</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% USD</td>
<td>% USD</td>
<td>% USD</td>
</tr>
<tr>
<td>Removing GQ</td>
<td>-0.10</td>
<td>-2.11</td>
<td>-2.02</td>
</tr>
<tr>
<td>Income-maximizing network, all 68 cities</td>
<td>0.69</td>
<td>3.09</td>
<td>0.26</td>
</tr>
<tr>
<td>Income-maximizing network, unconstrained</td>
<td>0.65</td>
<td>3.07</td>
<td>2.82</td>
</tr>
<tr>
<td>Least-cost network, all 68 cities</td>
<td>0.21</td>
<td>2.31</td>
<td>0.05</td>
</tr>
<tr>
<td>Rays and corridors, all 68 cities</td>
<td>0.39</td>
<td>1.87</td>
<td>1.47</td>
</tr>
<tr>
<td>Income-maximizing network, least-cost budget</td>
<td>0.20</td>
<td>1.68</td>
<td>1.47</td>
</tr>
<tr>
<td>Income-maximizing network, NHDP budget</td>
<td>0.16</td>
<td>1.38</td>
<td>1.22</td>
</tr>
</tbody>
</table>

The table summarizes the aggregate effects of the actual and counterfactual networks. The changes in construction costs, income, and net income due to each network are shown in percentages of actual GDP in 2012 and in billion USD (1999 prices). 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 row and below are assumed to replace the GQ and the construction costs of the GQ are subtracted from the construction costs of the counterfactual.

Table A12: Effect of transport infrastructure investments on population

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Market Access</td>
<td>0.121</td>
<td>0.0975</td>
<td>0.0629</td>
<td>0.0536</td>
<td>0.0536</td>
</tr>
<tr>
<td></td>
<td>(0.111)</td>
<td>(0.115)</td>
<td>(0.124)</td>
<td>(0.124)</td>
<td>(0.0969)</td>
</tr>
<tr>
<td>Excluded nodal districts</td>
<td>None</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Weighting</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Standard errors</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Robust</td>
</tr>
<tr>
<td>N</td>
<td>627</td>
<td>614</td>
<td>613</td>
<td>614</td>
<td>614</td>
</tr>
<tr>
<td>Rsq.</td>
<td>0.198</td>
<td>0.200</td>
<td>0.206</td>
<td>0.246</td>
<td>0.246</td>
</tr>
</tbody>
</table>

The table shows 2SLS estimates of the elasticity of population with respect to market access. The dependent variable is the logarithm of population in the years 2001 and 2011. The explanatory variable is market access computed based on Equation (7) and instrumented with the market access with constant light in Equation (12). All regressions include district fixed effects, state-year fixed effects, and controls for distance to the coast and the level of electrification in 2001 interacted with a year fixed effect. Column 1 shows the effect in the full sample and columns 2 - 5 exclude four or five nodal districts as stated in the table. Columns 1 - 3 weigh by the logarithm of initial sum of light. Standard errors (in parentheses) are clustered at the state-level except in column 5.
Table A13: First stage regression of market access

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Market Access</td>
<td>1.064</td>
<td>1.066</td>
<td>1.066</td>
<td>1.064</td>
<td>1.064</td>
</tr>
<tr>
<td></td>
<td>(0.0281)</td>
<td>(0.0288)</td>
<td>(0.0289)</td>
<td>(0.0287)</td>
<td>(0.00924)</td>
</tr>
<tr>
<td>Excluded nodal districts</td>
<td>None</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Weighting</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Standard errors</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Robust</td>
</tr>
<tr>
<td>N</td>
<td>626</td>
<td>613</td>
<td>612</td>
<td>612</td>
<td>612</td>
</tr>
<tr>
<td>F statistics</td>
<td>1433.55</td>
<td>1367.88</td>
<td>1356.33</td>
<td>1376.27</td>
<td>13245.1</td>
</tr>
</tbody>
</table>

This table shows the first stage regression for the specifications in Table 1 and regresses market access computed based on Equation (7) on market access with constant light in Equation (12). All regressions include district fixed effects, state-year fixed effects, and controls for distance to the coast and the level of electrification in 2001 interacted with a year fixed effect. Column 1 shows the effect in the full sample and columns 2 - 5 exclude four or five nodal districts as stated in the table. Columns 1 - 3 weigh by the logarithm of initial sum of light. Standard errors (in parentheses) are clustered at the state-level except in column 5.
Table A14: Elasticity of light with respect to market access (OLS)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Market Access</td>
<td>1.096</td>
<td>1.129</td>
<td>1.139</td>
<td>1.162</td>
<td>1.162</td>
</tr>
<tr>
<td></td>
<td>(0.508)</td>
<td>(0.526)</td>
<td>(0.529)</td>
<td>(0.542)</td>
<td>(0.313)</td>
</tr>
<tr>
<td>Excluded nodal districts</td>
<td>None</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Weighting</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Standard errors</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Cluster</td>
<td>Robust</td>
</tr>
<tr>
<td>N</td>
<td>626</td>
<td>613</td>
<td>612</td>
<td>612</td>
<td>612</td>
</tr>
<tr>
<td>Rsq.</td>
<td>0.509</td>
<td>0.503</td>
<td>0.503</td>
<td>0.496</td>
<td>0.496</td>
</tr>
</tbody>
</table>

The table shows OLS estimates of the elasticity of light with respect to market access. The dependent variable is the logarithm of the sum of light in each district in the years 1999 and 2012 (averaged over three years). The explanatory variable is market access computed based on Equation (7). All regressions include district fixed effects, state-year fixed effects, and controls for distance to the coast and the level of electrification in 2001 interacted with a year fixed effect. Column 1 shows the effect in the full sample and columns 2 - 5 exclude four or five nodal districts as stated in the table. Columns 1 - 3 weigh by the logarithm of initial sum of light. Standard errors (in parentheses) are clustered at the state-level except in column 5.
Figure A1: Road type and driving speed

The image shows a part of India, where the colors of different cells represent differences in the driving speeds due to different roads. The green lines represent highways of the NHDP and the blue lines highways of lower quality. The dots represent the centroids of Indian districts between which bilateral trade costs are computed as the least-cost path through the cells.

Figure A2: Terrain and road construction costs

The image shows the road construction costs as a function of slope and land cover. Dark red refers to high construction costs, orange and yellow to intermediate costs, and light green to low costs. The green circles represent cities in India which fulfill one of the two criteria of the Chinese NEN. The blue connections between the cities represent the cheapest construction routes.
Figure A3: Intermediate-sized cities in India.

The map shows the cities in India with more than 500,000 residents and all state capitals. The image in the background shows night-time lights.
Figure A4: Counterfactual networks

(a) Starting from empty or full network
(b) Budget of least-cost network
(c) Budget of GQ and NS-EW corridors
(d) Same cost as rays and corridors

Figure A4a shows the income-maximizing network when starting from the empty network (yellow) together with the network from Figure 3b (blue). Figure A4b shows the income-maximizing network with the budget of the least-cost network. Figure A4c shows the income-maximizing network with the budget of the planned GQ and NS-EW corridors. Figure A4d shows the income-maximizing network with the budget of the counterfactual rays and corridors network in Figure 3d.
Figure A5: Random starting points for network design algorithm

The figures illustrates how the network design algorithm converges when random starting points are used. 100 random networks are used as starting points in the algorithm to find the income-maximizing network that connects all 68 cities. The lines show how the number of links, construction costs, income, and income net of road construction costs evolve in each iteration.
The two maps show the initial density (left map) and growth (right map) in light in India. The initial density is the logarithm of average light intensity per pixel around the year 1999 (averaging 1998, 1999, and 2000). Growth is approximated by the log difference in light intensity between 1999 and 2012 (averaging the start and end years). The units are 636 Indian districts. Darker areas refer to higher density or higher growth rates.
Figure A7: Percent increase in GDP from replacing GQ with least-cost network to connect all 68 cities

The map shows the boundaries of Indian districts. Darker areas have a higher percentage difference in GDP generated by replacing the GQ with the counterfactual network that connects targeted cities in a network with the least costs.
The map shows the boundaries of Indian districts. Darker areas have a larger percentage difference in GDP generated by replacing the GQ with the counterfactual network that is constructed by connecting cities with rays and corridors.
Figure A9: Percent difference between NS-EW corridors and GQ

The map shows the boundaries of Indian districts. Darker color represents higher percentage difference in GDP between a network that includes the completed parts of the NS-EW and the GQ.