

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

Simon Alder

University of North Carolina at Chapel Hill

December 3, 2016

## Abstract

India and China followed different strategies in the design of their recent highway network projects. India 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 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, Economic Growth, Regional Development, Trade, Network Design, 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 Brutti, Marius Brühlhart, Adrien Bussy, 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, Peter Kondor, Michael König, Rafael Lalive, Sergi Jiménez-Martín, Omar Licandro, Thierry Mayer, Andreas Müller, Alessandro Pavan, Michelle Rendall, José-Víctor Ríos-Rull, Dominic Rohner, Esteban Rossi-Hansberg, Lin Shao, Kjetil Storesletten, Viktor Tsyrennikov, Ashish Vachhani, Valentin Verdier, Rainer Winkelmann, Christoph Winter, Gabriel Zucman, Josef Zweimüller, Robert Zymek, and participants in presentations at the University of Zurich, University of Lausanne, University of Oslo, 2014 NBER Summer Institute Workshop on Urban Economics, Barcelona GSE Summer Forum, SNF - CEPR Conference, the Midwest Macro Meeting, and the World Congress and Asia Meeting of the Econometric Society 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) is gratefully acknowledged.

# 1 Introduction

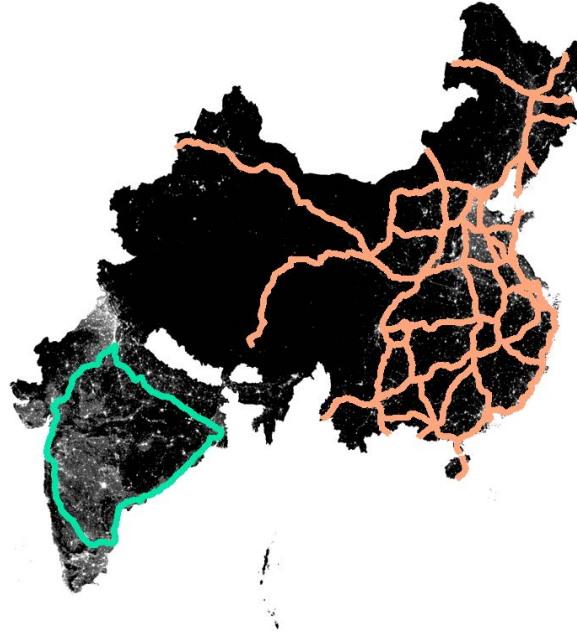
China and India, the two most populous countries in the world, are developing at unprecedented rates. Yet, their spatial, or regional, development patterns are surprisingly different. Throughout China, new clusters of economic activity are emerging and there is a stronger pattern of convergence across Chinese counties. 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 Calcutta 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 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 mimics the Chinese strategy of connecting intermediate-sized cities. More specifically, I construct a counterfactual network that approximately maximizes aggregate income net of road construction costs, while connecting all Indian cities with a population above 500,000 and all state capitals as suggested by the Chinese policy.

The actual and counterfactual networks are evaluated in a Ricardian general equilibrium trade framework based on Donaldson and Hornbeck (2016). The model allows for trade among Indian districts 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 of changes in the transport network are

Figure 1: Indian and Chinese highway projects



The figure shows the 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.

captured by a measure of market access. A location's market access is the sum over the income of all districts, discounted by the bilateral trade costs and by the other districts' market access. The model predicts a relationship between income and market access with a constant elasticity that can be estimated using panel data on districts' incomes and market access.

Since official district-level GDP data are not available for the entire period, I use night light data as in Henderson et al. (2012) to measure real income. The market access measures are general equilibrium outcomes that I obtain from the model for each set of trade costs. For a given transport network, these trade costs can be derived from the computed shortest path between all district centroids. Hence, the bilateral trade costs can be calculated for the transport network in 2000 (before the construction of the GQ), in 2012 (after completion of the GQ), and for the counterfactuals (replicating the Chinese network).

The observed changes in market access from 2000 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 obtain an elasticity that is consistent with the prediction of the model for reasonable

parameter assumptions. With this estimated elasticity, I then solve the general equilibrium 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 2.1 percent lower in 2012 if the GQ had not been built.

Second, I predict the aggregate effect of the counterfactual transport infrastructure that replicates the Chinese network in India by connecting intermediate-sized cities in a way that approximately maximizes net aggregate income in the general equilibrium framework.<sup>1</sup> The resulting network would cost more than five times as much as the GQ, but it would lead to a net increase in real income (relative to the GQ) equivalent to 2.24 percent of GDP in 2012. I also compare the effects from 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 four times larger than the GQ and connects most intermediate-sized cities. It would imply an increase in net income equivalent to 2.25 percent of GDP. While these counterfactual network designs are computed with a heuristic algorithm and they don't necessarily represent the global optimum, the results provide a lower bound for the net gains that could have been achieved with a denser network that targets intermediate-sized cities according to the Chinese policy.<sup>2</sup>

The third contribution is to evaluate the distributional consequences of the actual and counterfactual networks. The results show that initially less developed regions would gain from the counterfactual. The reason is that the counterfactual network, by connecting cities with a population above 500,000 and state capitals, also reaches into regions that previously had low growth and were neglected by the GQ. Thus, a

---

<sup>1</sup>This is based on an iterative procedure that starts from a fully connected network and iteratively removes the least beneficial links, while recomputing the bilateral shortest paths and general equilibrium market access measures in each iteration. I show that this heuristic algorithm leads to a relatively similar result when starting from the empty network and sequentially adding the most beneficial links.

<sup>2</sup>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 network based on the iterative procedure is substantially better in terms of net income.

transport network that follows the strategy of connecting intermediate-sized cities would increase growth particularly in India's lagging regions.

The remainder of the paper is structured as follows. Section 2 reviews the related literature. Section 3 discusses the transport infrastructure in India and China. Section 4 presents the conceptual framework 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 combines economic theory with geographic information such as the exact location of transport infrastructure.<sup>3</sup> My methodology for evaluating the impact of infrastructure builds on Donaldson and Hornbeck (2016) who estimate the aggregate effect of the expansion of the American railway network in the 19th century. Donaldson and Hornbeck (2016) also compare the effect of the actually built network to counterfactual scenarios in which railways are replaced by roads and canals. My counterfactual analysis differs from theirs by designing a network that targets different cities than the actual network and that connects these cities in a way that approximately maximizes income net of construction costs. Furthermore, I compare the aggregate and distributional consequences of various network designs.

To construct the networks, I use an iterative procedure that adds and removes links based on construction costs and income implied by trade costs in the general equilibrium framework.<sup>4</sup> Iterative algorithms to search for the optimal network in equilibrium have previously been applied in Jia (2008) and Antras et al. (2016)

---

<sup>3</sup>See for example recent surveys by Breinlich et al. (2013), Redding and Turner (2015), and Donaldson (2015). The general decline of transport costs for goods and its implication for urban and regional development is discussed in Glaeser and Kohlhase (2004).

<sup>4</sup>Similar heuristic algorithms have been applied in the network design literature, such as Gastner and Newman (2006), but not in a general equilibrium trade framework. Gastner (2005) discusses this literature and also outlines the heuristic algorithm used here. In the economics literature, Burgess et al. (2015) and Balboni (2016) construct counterfactual networks by ranking cities by initial market potential based on Euclidean distance and connecting those with the highest rank, but they do not select links based on general equilibrium effects.

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. Felbermayr and Tarasov (2015) model the endogenous distribution of transport infrastructure on a line and consider the density of transport infrastructure when approaching national borders. Fajgelbaum and Schaal (2016) propose a trade model with congestion in transport that leads to a convex optimization problem when congestion is sufficiently strong. This allows them to compute the globally optimal transport network and they illustrate the approach in a number of environments. My approach is based on a heuristic algorithm that finds local optima, but I show that different starting points lead to similar networks. It is to the best of my knowledge the first design of a national transport network that approximately maximizes net income in a general equilibrium gravity trade framework applied to an actual economy.

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. (2015) study the effects of the GQ on firms located in the proximity of the new highways and find positive effects on manufacturing activity.<sup>5</sup> An important aspect of these studies is the identification of exogenous sources of variation in transport infrastructure.<sup>6</sup> They rely on an identification strategy similar to the one proposed by Chandra and Thompson (2000) and Michaels (2008) who estimate the effect of U.S. highways on counties that lie between two larger 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 also rely on the identification strategy based on non-nodal districts and 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

---

<sup>5</sup>Both studies use firm surveys to evaluate the effect of the GQ. Khanna (2016) applies the identification strategies in Ghani et al. (2015) based on nodal cities and straight lines to estimate the effect of the GQ on Indian sub-district using light data. Asher and Novosad (2016) estimate the effect of rural roads on structural transformation in Indian villages.

<sup>6</sup>See Redding and Turner (2015) for a discussion of different identification strategies to estimate the effect of transport infrastructure on various outcomes.

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 (forthcoming) estimates the effect of railways in colonial India. My analysis differs from the above studies by estimating the effect of transport infrastructure in India with the market access approach 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 optimal 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. (2016) use a similar model as in Donaldson and Hornbeck (2016) and estimate the effect of Chinese highways on growth of prefectures. They also consider the role of access to international markets and find differences between reduced form estimation and general equilibrium modeling. 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.<sup>7</sup> I differ from these studies by comparing the aggregate and distributional effects of the actual network to a counterfactual network that connects targeted cities

---

<sup>7</sup>See also Atack et al. (2008), Herrendorf et al. (2012), and Jaworski and Kitchens (2016) for studies of intercity transport infrastructure in the U.S.. Bird and Straub (2015) and Morten and Oliveira (2016) study the effects highways in Brazil and Storeygard (2016) estimate the effect of transport costs on sub-Saharan African cities using light data. A number of studies that do not focus on transport infrastructure have also used light data, see for example Elvidge et al. (1997), Ghosh et al. (2010), Chen and Nordhaus (2011), Henderson et al. (2012), Ma et al. (2012), and Hodler and Raschky (2014).

in an approximately optimal 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; and Redding, 2016). 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.<sup>8</sup> 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.

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 effectiveness of infrastructure projects (see World Bank, 2007a for an overview). While those concepts have advantages in capturing project-specific aspects such as safety and road deterioration, the methodology applied in this paper is able to capture the general equilibrium effects at a large scale, which allows evaluating and comparing national infrastructure strategies.

---

<sup>8</sup>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.



### 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.<sup>9</sup> 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.<sup>10</sup> 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.<sup>11</sup>

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

---

<sup>9</sup>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).

<sup>10</sup>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.

<sup>11</sup>A description of the history of the Chinese highway network and its different components is provided by ACASIAN. See [www.acasian.com](http://www.acasian.com) for further details.

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).<sup>12</sup> 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.<sup>13</sup> 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).

### 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

---

<sup>12</sup>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. (2015).

<sup>13</sup>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.

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.<sup>14</sup> 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 substantially 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 severe 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 had a transport infrastructure like China. To answer this question, I propose a counterfactual road network for India that mimics the Chinese strategy of connecting intermediate-sized cities. The exact routes are chosen to approximate the optimal network in the general equilibrium framework by balancing the gains from market access 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 compare the effects of actual and counterfactual networks.

## 4 Conceptual Framework

The setup is a general equilibrium trade model based on Eaton and Kortum (2002). Donaldson and Hornbeck (2016) derive from a variation of that model a reduced form expression for the impact of transport infrastructure on income.<sup>15</sup> 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

---

<sup>14</sup>The share of highways in the total freight turnover is even higher in India than in China (KPMG, 2013).

<sup>15</sup>The presentation in this section focuses on the key aspects of the model. The details are discussed in the appendix.

the trading partners. They use this framework to estimate the effect of the expansion of the American railway network on land prices. I follow this approach to estimate the effect of the Indian transport network on income by adapting their framework to a version which can be estimated with light data as a measure for real income.

#### 4.1 A Ricardian Model of Trade

The basic setup is a Ricardian trade model with the immobile production factors land and labor and the mobile factor capital.<sup>16</sup> 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),

$$x_o(j) = z_o(j)(L_o(j))^\alpha (H_o(j))^\gamma (K_o(j))^{1-\alpha-\gamma}, \quad (1)$$

where  $z_o(j)$  is an exogenous probabilistic productivity shifter as in Eaton and Kortum (2002).<sup>17</sup> The production function implies marginal costs

$$MC_o(j) = \frac{q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma}}{z_o(j)} \quad (2)$$

where  $q_o$  is the land rental rate,  $w_o$  is the wage, and  $r_o$  is the interest rate.

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 such that  $p_{oo}(j) = MC_o(j)$ , which implies

$$p_{od}(j) = \tau_{od}MC_o(j) = \tau_{od} \frac{q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma}}{z_o(j)} \quad (3)$$

$$z_o(j) = \tau_{od} \frac{q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma}}{p_{od}(j)}. \quad (4)$$

---

<sup>16</sup>In their study of the effect of railways on American counties between 1870 and 1890, Donaldson and Hornbeck (2016) assume that labor is mobile. Over the sample period of 12 years that I have available to quantify the effect of the Indian GQ, the assumption of immobile labor is more appropriate. However, the analysis could also be carried out when assuming labor is mobile.

<sup>17</sup>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 within districts (comparative advantage) and  $T_o$  is a district’s state of technology (absolute advantage).

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 of the following form:

$$P_d = \gamma \left( \sum_o [T_o (\tau_{od} q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma})^{-\theta}] \right)^{-\frac{1}{\theta}}. \quad (5)$$

I follow the notation in Donaldson and Hornbeck (2016) and define the sum over origins' factor costs as "consumer market access", because it measures district  $d$ 's access to cheap goods,<sup>18</sup>

$$\begin{aligned} P_d^{-\theta} &= \kappa_1 \sum_o [T_o (\tau_{od} q_o^\alpha w_o^\gamma)^{-\theta}] \\ &\equiv CMA_d. \end{aligned} \quad (6)$$

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

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

$$\frac{X_{od}}{X_d} = \frac{T_o (q_o^\alpha w_o^\gamma)^{-\theta} \tau_{od}^{-\theta}}{\sum_o [T_o (q_o^\alpha w_o^\gamma)^{-\theta} \tau_{od}^{-\theta}]}. \quad (7)$$

Assuming that a district's expenditure equals income ( $X_d = Y_d$ ),<sup>19</sup> this can be rearranged to

$$\begin{aligned} X_{od} &= \underbrace{T_o (q_o^\alpha w_o^\gamma)^{-\theta}}_{\text{Origin's productivity and factor costs}} \times \underbrace{\tau_{od}^{-\theta}}_{\text{Trade costs}} \times \underbrace{Y_d}_{\text{Destination's income}} \\ &\times \underbrace{\kappa_1 CMA_d^{-1}}_{\text{Destination's CMA}}. \end{aligned} \quad (8)$$

This is a gravity equation where the amount of trade from  $o$  to  $d$  depends positively on the origin's competitiveness (productivity) and the destination's income,

<sup>18</sup>Using the fact that the rental rate for capital is equalized everywhere to  $r_o = r$ , we can define the constant  $\kappa_1 \equiv \gamma^{-\theta} r^{-(1-\alpha-\gamma)\theta}$  where  $\gamma = \left[ \Gamma \left( \frac{\theta+1-\sigma}{\theta} \right) \right]^{\frac{1}{1-\sigma}}$  and  $\Gamma$  is the gamma function.

<sup>19</sup>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.

but negatively on the consumer market access of the destination and on the bilateral trade costs. This feature of an income gravity equation is shared by a large class of models and has found strong support in the data.

#### 4.1.2 Income and Market Access

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

$$Y_o = \sum_d X_{od} = \kappa_1 T_o (q_o^\alpha w_o^\gamma)^{-\theta} \sum_d [\tau_{od}^{-\theta} CMA_d^{-1} Y_d], \quad (9)$$

where Donaldson and Hornbeck (2016) define “firm market access” of district  $o$  as

$$FMA_o \equiv \sum_d \tau_{od}^{-\theta} CMA_d^{-1} Y_d. \quad (10)$$

If trade costs are symmetric, then 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

$$MA_o = \rho \sum_d \tau_{od}^{-\theta} MA_d^{-1} Y_d. \quad (11)$$

Equation (9) for income then becomes

$$Y_o = \kappa_1 T_o (q_o^\alpha w_o^\gamma)^{-\theta} MA_o. \quad (12)$$

Equations (11) and (12) are the key model equations and they summarize how trade costs affect income. While Equation (12) implies a relationship between income and market access, Equation (11) 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. In particular, the market access approach takes into account that a reduction in bilateral trade costs  $\tau_{di}$  between two trading partners  $d$  and  $i$  can affect market access in  $o$ . This can be seen in Equation 11, where an increase in  $MA_d$  (for example due to a decrease in  $\tau_{di}$ ) could reduce  $MA_o$ .

#### 4.1.3 Prices and Real Income

The framework summarized by Equations (11) and (12) suggests a relationship between transport costs and income that can be estimated with suitable data. For the empirical analysis, I will use night lights as a measure of real income and therefore

rewrite the model accordingly. To this aim, I use the property in Eaton and Kortum (2002) that the price index is related to market access,

$$P_d = (\rho^{-1} MA_d)^{-\frac{1}{\theta}}. \quad (13)$$

This allows rewriting Equation (11) as

$$MA_o = \rho^{\frac{1+\theta}{\theta}} \sum_d \tau_{od}^{-\theta} MA_d^{\frac{-(1+\theta)}{\theta}} Y_d^r, \quad (14)$$

where  $Y_d^r = \frac{Y_o}{P_o}$  denotes real income.

Equation (12) can also be written in terms of real income by using the price index in Equation (13). Furthermore, the wage and land rental rates can be substituted using the factor income shares to obtain<sup>20</sup>

$$Y_o^r = (\kappa_2 T_o)^{\frac{1}{1+\theta\alpha+\theta\gamma}} \left(\frac{\alpha}{L_o}\right)^{\frac{-\theta\alpha}{1+\theta\alpha+\theta\gamma}} \left(\frac{\gamma}{H_o}\right)^{\frac{-\theta\gamma}{1+\theta\alpha+\theta\gamma}} (MA_o)^{\frac{1+\theta(1+\alpha+\gamma)}{(1+\theta\alpha+\theta\gamma)\theta}}, \quad (15)$$

where  $\kappa_2 = \kappa_1 \rho^{-\frac{1+\theta\alpha+\theta\gamma}{\theta}}$ . As pointed out by Donaldson and Hornbeck (2016), the effect of transport infrastructure on income is captured by the market access measures. The elasticity of income with respect to market access in Equation (15) will be estimated using panel data on real income and market access measures that solve Equation (14). This estimate will then be used for the counterfactual analysis.

## 4.2 Empirical Strategy

Estimating Equation (15) 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 (16) shows the different components of Equation (15) over time:

$$\begin{aligned} \ln(Y_{o,t}^r) &= \underbrace{-\frac{\theta\alpha}{1+\theta\alpha+\theta\gamma} \ln\left(\frac{\alpha}{L_o}\right) - \frac{\theta\gamma}{1+\theta\alpha+\theta\gamma} \ln\left(\frac{\gamma}{H_o}\right)}_{\text{Constant over time}} \\ &+ \underbrace{\frac{1}{1+\theta\alpha+\theta\gamma} \ln(\kappa_{2,t})}_{\text{Country characteristics}} + \underbrace{\frac{1}{1+\theta\alpha+\theta\gamma} \ln(T_{o,t})}_{\text{Productivity}} \\ &+ \underbrace{\frac{1+\theta(1+\alpha+\gamma)}{(1+\theta\alpha+\theta\gamma)\theta} \ln(MA_{o,t})}_{\text{Market access}}. \end{aligned} \quad (16)$$

<sup>20</sup>With the Cobb-Douglas production function and competitive markets, we have  $q_o = \frac{\alpha Y_o}{L_o}$  and  $w_o = \frac{\gamma Y_o}{H_o}$ .

The right-hand side of the first line in Equation (16) collects parameters and factor endowments, which are assumed to be constant over time and can thus be absorbed by the district fixed effects. The second line includes country characteristics (the interest 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 (16) shows the effect of market access.

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

$$\beta = \frac{1 + \theta(1 + \alpha + \gamma)}{(1 + \theta\alpha + \theta\gamma)\theta}. \quad (17)$$

The panel fixed effects specification corresponding to Equation (16) to estimate  $\beta$  is

$$\ln(Y_{o,s,t}^r) = \phi_o + \delta_{s,t} + \beta \ln(MA_{o,t}) + X_{o,t}\pi_t + \varepsilon_{o,s,t}, \quad (18)$$

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.

While I identify only  $\beta$  and cannot estimate the individual parameters in Equation (17) separately, one can verify whether the resulting estimate for  $\beta$  is close to what would be predicted by certain parameter values. Using 0.3 for the capital share (and therefore 0.7 for the sum of the land and labor shares) and 8 for the trade elasticity, would imply an elasticity of income with respect to market access of 0.28. This value is not rejected by the estimates in Section 5.2 below.<sup>21</sup>

### 4.3 Identification

Identifying the causal effect of infrastructure on income is challenging for several reasons.<sup>22</sup> 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.

<sup>21</sup>Note that I use light to measure real income both as the dependent variable and in the market access terms. The correlation between light and GDP is discussed below.

<sup>22</sup>Redding and Turner (2015) provide a discussion of the identification issues related to transport infrastructure.



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 which were accidentally affected by the GQ. This identification strategy was proposed by Chandra and Thompson (2000) and Michaels (2008) and similar strategies have been applied to China and India by Banerjee et al. (2012), Datta (2012), Ghani et al. (2015), and Asturias et al. (2016). I follow this strategy and exclude the nodal cities and the corresponding districts.

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 14 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 (19) shows this version of the market access equation,<sup>23</sup>

$$MA_{o,t} = \rho \sum_d \tau_{od,t}^{-\theta} MA_{d,t}^{\frac{-(1+\theta)}{\theta}} Y_{d,1999}^T. \quad (19)$$

I estimate the elasticity  $\beta$  based on measured variation in market access due to the GQ and then predict income for counterfactual transport networks.

## 4.4 Counterfactual Predictions

$\beta$  represents the elasticity of income with respect to market access, as shown in Equation (15). The general equilibrium market access measures implied by Equation (14) themselves also depend on income. I jointly solve this system of Equations (14)

---

<sup>23</sup> 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. They first use an approximation  $MA_o \approx \sum_d \tau_{od}^{-\theta} N_d$ , where  $N$  denotes population, which in their case is endogenous. They then compare the results to those based on the numerical solution for market access implied by an equation similar to Equation (14). They find that the effects are about of the same magnitude. I focus on the second approach and use the market access measures based on the numerical solution of Equation (14).

and (15) for each version of the transport network, holding constant the immobile production factors, districts' technology levels, and the interest rate (see Appendix A.8). Another approach in the literature is to use the estimate of  $\beta$  in Equation (18) 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 (14) and (15). Hence, market access changes over the counterfactuals only due to the trade costs and does not account for the reallocation of income.<sup>24</sup> I will focus on the first approach and solve Equations (14) and (15) jointly to predict income. With the alternative approach that assumes income to be constant in the market access equation, the aggregate effect of the GQ would be about 30 percent lower but the distributional effects are qualitatively similar.

For the computation of the market access measures that solve Equation (14), a value for the trade elasticity  $\theta$  is required. As a benchmark, I use a value of 8. This is higher than values that have previously been derived for India but it provides a better fit and it is similar to the value used in Donaldson and Hornbeck (2016).<sup>25</sup> In the robustness section, I show the results for the estimated elasticity with other values of the trade elasticity.

## 5 Data and Estimation

The data required for the estimation of Equation (18) 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 can be found in Appendix B.

---

<sup>24</sup>Donaldson and Hornbeck (2016) show the results from different approaches to estimate the general equilibrium effects of changes in transport infrastructure, including the ones above, but assuming mobile labor. See also Head and Mayer (2013) for a comparison of different approaches to quantify the effect of trade costs.

<sup>25</sup>I have experimented with different values for  $\theta$  to test what value provides the best fit in the 1999 cross section. I find that a value of 8 provides a better fit than lower values such as 3.8, which was found for the case of India in Donaldson (forthcoming), although in a different context of the colonial railway network. Simonovska and Waugh (2014) found lower values than 8 using international trade data, but I prefer to use the value that provides a good fit in my context and is in line with other work focusing on intra-national trade.

## 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 Night Lights as a Measure for Real Income

I use Indian districts with administrative boundaries as defined in the 2011 census. Official GDP data is not available for the full period and I therefore use night lights as a measure for income.<sup>26</sup> Growth in light at night measured by weather satellites has been shown to be a good proxy for income growth (Henderson et al., 2012).<sup>27</sup> 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.<sup>28</sup> The data source and details on the light data preparation are provided in Appendix B.1.

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 based on the 2011 census maps. This allows me to construct a panel of districts with light data from 1992 to 2013.

While the estimation of the elasticity  $\beta$  is based on light data, the results of the quantitative exercise 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. The true relationship between GDP and light in India may differ from this elasticity and this would affect the magnitude of the aggregate GDP results

---

<sup>26</sup>The Indian Planning Commission publishes district level GDP data for the years 1999 - 2005, but not for the full period that I study here. See also Section 7.2 where the early data is used to obtain the districts' trade to GDP ratio in 1999.

<sup>27</sup>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).

<sup>28</sup>See also Ma et al. (2012) and Hodler and Raschky (2014) for an application of light data at the sub-national level.

reported here, but it would not affect the comparison across the various networks nor the distributional consequences.

There are some caveats when using light data. First, the light intensity is top-coded for very bright pixels at 63, 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.<sup>29</sup> 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 as reported in the census.

### 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 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,

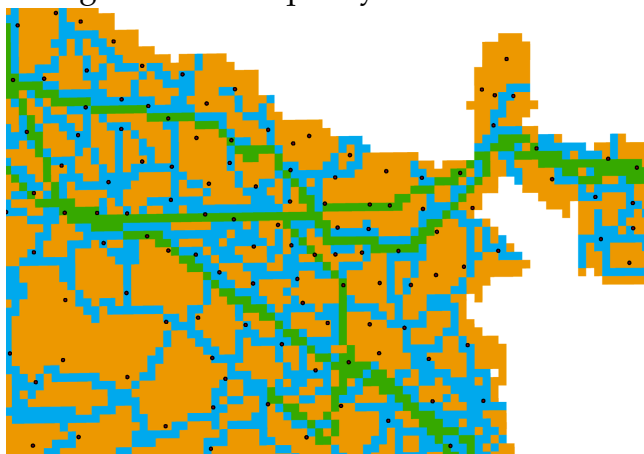
---

<sup>29</sup>The alternative would be to use a radiance-calibrated version of the data, but this is unfortunately not available in all years.

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.2.

With these inputs, it is possible to construct a grid of India where the value of each  $1 \times 1$  km cell represents the costs of traveling through this cell. These travel costs are the inverse of the speed on the best road inside of each cell. Such a grid of transport costs is shown in Figure 2. I use the fast marching algorithm as in Allen and Arkolakis (2014) and Allen and Atkin (2016) to compute bilateral travel times.<sup>30</sup> The algorithm then calculates the cheapest way to travel from one location (district centroids, represented by dots in Figure 2) 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.

Figure 2: Road quality and travel costs



The figure shows a part of the Indian landscape, where the colors of different cells represent differences in travel costs 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 cost grid.

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.<sup>31</sup> More precisely, I calculate iceberg trade costs between an origin  $o$  and a

<sup>30</sup>I would like to thank the authors for making their code available.

<sup>31</sup>This is a common assumption, see for example also Au and Henderson (2006) who assume that transport costs increase less than proportionally in *distance*.

destination  $d$  as

$$TradeCosts_{od} = 1 + \gamma TransportTime_{od}^{0.8}. \quad (20)$$

$\gamma$  is chosen such that the median iceberg trade cost is 1.25 for the network without the GQ.<sup>32</sup> The transport costs within districts is 1.

Although the analysis undertaken here captures a key aspect of the modern transport infrastructure in India and China, some caveats must be pointed out. The first concern is the omission of other types of domestic transport infrastructure such as railways or urban transport systems such as subways. Second, access to international markets via 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 demand 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. The above caveats would therefore 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 above 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:

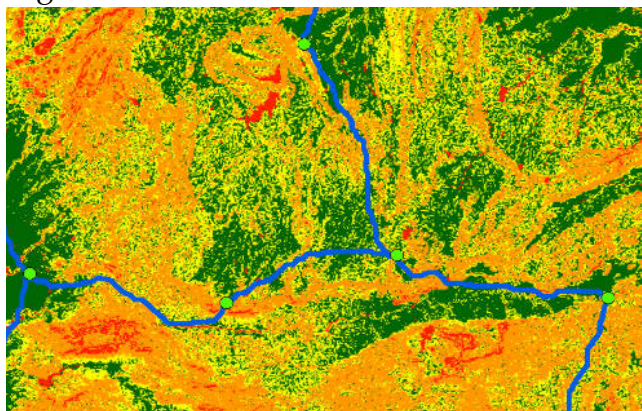
$$ConstructionCosts_c = 1 + Slope + 25 \times Builtup + 25 \times Water. \quad (21)$$

---

<sup>32</sup>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. (2016) for a similar calculation for China.

*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.<sup>33</sup> Applying this formula using detailed terrain data produces a  $1 \times 1$  km grid of construction costs for the entire Indian landscape. Given this grid of construction costs, one can in a second step apply the Dijkstra (shortest path) algorithm to find the cheapest connection between any two given points through the cost grid.<sup>34</sup> The procedure is illustrated in Figure 3, where the cells represent different construction costs (based on Equation 21) and the lines are the least-cost paths to connect the locations (shown as circles).

Figure 3: Terrain and road construction costs



The figure 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.

For each resulting counterfactual network, one can calculate the total road construction costs based on the topography. In order to obtain an estimate for the construction costs in USD, I use the ratio between the USD costs of the GQ and its costs based on the topography.

---

<sup>33</sup>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.

<sup>34</sup>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 (forthcoming). It is similar to the fast marching algorithm, but it allows keeping track of the path, which is essential to build the counterfactual network.

### 5.1.4 Descriptive Statistics

The descriptive statistics for the main variables market access and light growth are shown in Table 1. 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 from 1999. The third variable shows the implied log change in market access, i.e. 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.

Table 1: Descriptive Statistics

Variable	Obs	Mean	Std. Dev.
Market access without GQ, light in 1999	636	869.75	216.57
Market access with GQ, light in 1999	636	938.93	231.26
Log difference in market access from GQ, constant light	636	.08	.06
Market access with GQ, light in 2012	636	1209.1	287.8
Log difference in market access from GQ, variable light	636	.33	.07
Log difference in light	630	.56	.54

The table shows the summary statistics for the general equilibrium market access measures and light data. The market access measures are derived from equations (14) and (19) 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  $\theta = 8$ . The sample consists of 636 Indian mainland districts from the 2011 census maps.

## 5.2 Estimation

The estimation is based on Equation (16) of the model, suggesting a log-linear relationship between real income and market access. This equation is estimated using variation in income and market access over time. The corresponding empirical specification is shown in Equation (18).

The estimate of  $\beta$  represents the elasticity of light with respect to market access, where the time variation in market access is due to the construction of the GQ. The



results are shown in Table 2. As a benchmark, column 1 uses the full sample of main-land 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 dummy). Observations are weighted by the logarithm of districts' initial light in 1999 and standard errors are clustered at the state-level. All results are based on instrumental variable regressions where the market access measures from Equation (14) are instrumented for with market access measures from Equation (19) that vary only based on changes in trade costs due to the construction of the GQ. The first stage is strong and the elasticity between the two market access measures is 1.06.

Table 2: Elasticity of light with respect to market access

	(1)	(2)	(3)	(4)	(5)
Log Market Access	0.626*	0.655*	0.662*	0.668*	0.668**
	(0.340)	(0.362)	(0.365)	(0.389)	(0.300)
Excluded nodal districts	None	4	5	5	5
Weighting	Yes	Yes	Yes	No	No
Standard errors	Cluster	Cluster	Cluster	Cluster	Robust
N	626	613	612	612	612
Rsq.	0.508	0.502	0.502	0.495	0.495

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. The explanatory variable is market access computed based on Equation (14) and instrumented with the market access with constant light in (19). 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 are clustered at the state-level except in column 5.

The estimated coefficient for the effect of market access on light implies that a one percent increase in market access is associated with a 0.626 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. To address these challenges, I apply the identification strategy discussed in Section 4.3.

First, it should be noted that the panel structure of the data helps addressing concerns due to unobserved location characteristics, which here are absorbed by the district fixed effects. It is in principal possible that there are time-varying district characteristics, but it is less likely – given the national scope of 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 the identification strategy by Chandra and Thompson (2000) and Michaels (2008), 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 0.655. This suggests that the observed correlation between market access and light is not driven by the endogenous location of infrastructure in the four nodes.<sup>35</sup> 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 similar when excluding all five nodes. The regressions shown in column 1 - 3 weighted 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.<sup>36</sup>

Section 7 discusses the robustness of the results to alternative values for the trade elasticity, pre-trends, labor mobility, and international trade. For the counterfactual exercise below, I rely on the estimates of column 3 in Table 2, which weighs observations by the log of the initial sum of light and excludes the five nodes.

---

<sup>35</sup>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.

<sup>36</sup>The previous columns 1 - 4 allow errors to be correlated within states, but the number of clusters is relatively small with 32 states in the sample.

## 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. The estimate of this effect can be used to predict how much each district's income changes for various counterfactual networks. Because the framework captures general equilibrium effects of transport infrastructure, this allows analyzing the aggregate and distributional consequences. The results are derived from solving Equations (14) and (15) for each version of the actual and counterfactual transport networks. The estimate for the elasticity of income with respect to market access is taken from the third column in Table 2. The aggregate effects are summarized in Tables 3 and 4 and discussed in detail below.

### 6.1 Actual Network

I evaluate the effect of the actually built infrastructure by constructing a network in 2012 without the GQ. I first discuss the aggregate effects by summing over all districts and then consider the distributional implications.

#### 6.1.1 Aggregate Effect of GQ

Based on the estimated elasticity of GDP with respect to market access and the market access measures without the GQ, the result suggest that aggregate light in 2012 would be 7.34 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 7.34 percent change in light corresponds to a 2.2 percent change in GDP. Aggregate GDP in India in 2012 was USD 1,099 billion (in 1999 prices) such that a 2.2 percent difference would correspond to an annual change in GDP of roughly USD 24 billion.

The total costs of the GQ (including interest until 2012) amounted to about USD 6.2 billion (1999 prices).<sup>37</sup> In order to quantify the annual cost, I need to assume a

---

<sup>37</sup>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. (2015) as USD 7 billion (1999 prices). 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 and that the annual

Table 3: Aggregate effects of transport networks in percent of 2012 GDP

	Costs	Income	Net income
Removing GQ	-0.10	-2.20	-2.10
Approximation of optimum with all 68 cities	0.44	2.68	2.24
Approximation of unconstrained optimum	0.35	2.60	2.25
Least cost network with all 68 nodes	0.21	0.38	0.17
Rays & corridors with all 68 nodes	0.39	2.04	1.64
Approximation with same budget as least cost	0.20	1.76	1.56
Approximation based on sequential increase	0.35	2.47	2.12

The table summarizes the 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 GDP in 2012. Annual costs are based on 5% cost of capital and 12% maintenance costs. The counterfactual networks are assumed to replace the GQ and the construction costs of the GQ are subtracted from the construction costs of the counterfactual.

cost of capital and maintenance. I use a cost of capital of 5%, which is a conservative assumption since during the past decade the actual cost of capital was lower. Furthermore, I assume maintenance costs of 12% of the construction costs, which is approximately what Allen and Arkolakis (2014) report for the U.S. interstate highway network.<sup>38</sup> Based on these assumptions, the annual net effect on income, shown in Table 4, is around USD 23 billion.

It should be noted that this result is based on several assumptions and estimates. In particular, it depends on the two elasticities (GDP with respect to light and light with respect to market access) and on the assumptions on costs made above. Furthermore, Ghani et al. (2015) point out that there is some uncertainty regarding the exact costs of the GQ, which could also affect the result, although the assumptions on cost of capital and maintenance are conservative. While the estimate of  $\beta$  is somewhat imprecise, it should be noted that it has a similar magnitude as predicted by the model for reasonable parameter values once the elasticity of GDP with respect to light is taken into account. Furthermore, the finding that the GQ had a positive effect on economic activity is in line with the existing literature such as Ghani et al. (2015) and Asturias et al. (2016).

---

interest rate was 2% (based on the central government bond rates and inflation). In what follows, the base year is omitted, but all values are in USD at 1999 prices.

<sup>38</sup>This is also a conservative estimate if we assume that maintenance is more labor intensive than construction and that the labor costs are relatively lower in India compared to the US.

Table 4: Aggregate effects of transport networks in USD 1999 prices

	Costs	Income	Net income
Removing GQ	-1.05	-24.19	-23.13
Approximation of optimum with all 68 cities	4.81	29.48	24.67
Approximation of unconstrained optimum	3.89	28.61	24.72
Least cost network with all 68 nodes	2.31	4.19	1.88
Rays & corridors with all 68 nodes	4.31	22.39	18.08
Approximation with same budget as least cost	2.14	19.32	17.18
Approximation based on sequential increase	3.86	27.14	23.28

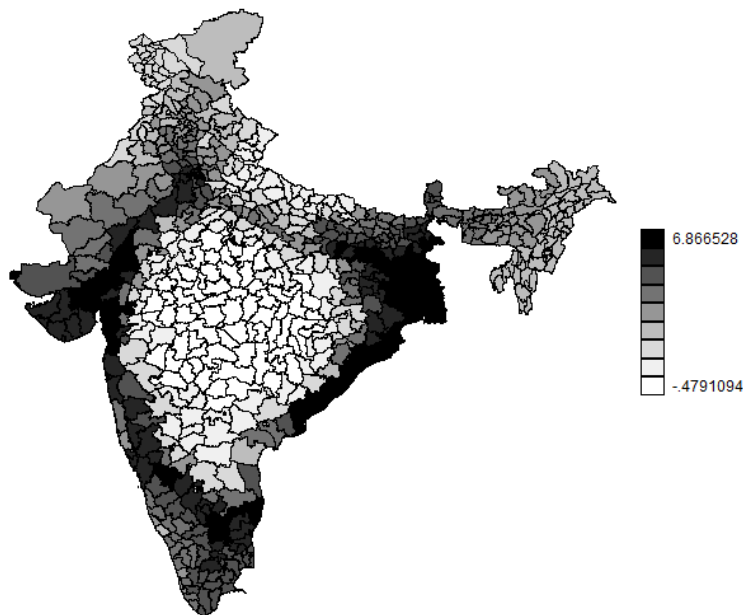
The table summarizes the effects of the actual and counterfactual networks. The changes in construction costs, income, and net income due to each network are shown in billion USD (1999 prices). Annual costs are based on 5% cost of capital and 12% maintenance costs. The counterfactual networks are assumed to replace the GQ and the construction costs of the GQ are subtracted from the construction costs of the counterfactual.

### 6.1.2 Distributional Effects of GQ

China and India have shown significant differences in their regional development. In particular, India had less convergences 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 approach used here is that the effects of different transport networks can be assessed both at the aggregate and at the local level, allowing to analyze the distributional consequences and the regional development patterns under each scenario. 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 4 shows the estimated 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 paths of the newly built or upgraded highways. The largest beneficiaries had an almost 7 percent higher GDP level in 2012 than they would have had in the absence of the GQ. Although transport costs did not increase anywhere (it is assumed that the GQ was added to an otherwise unchanged transport network), there are some districts with small losses from the infrastructure investments due to general equilibrium effects.

Figure 4: Percent increase in GDP generated by the GQ



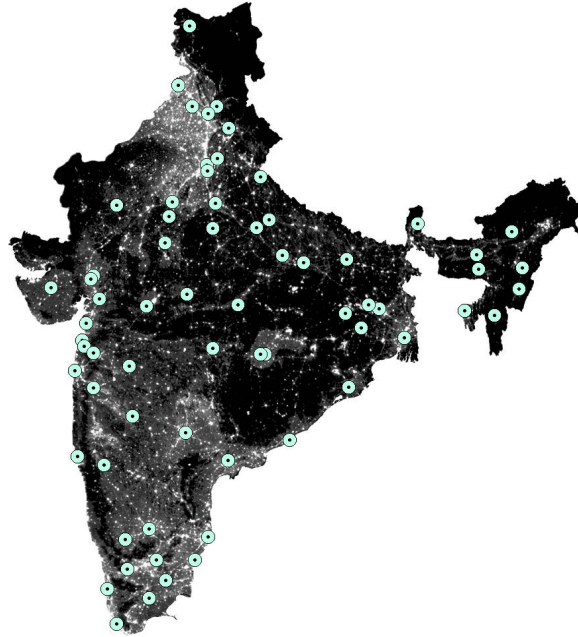
The map shows the boundaries of Indian districts. Darker color represents higher percentage difference in GDP generated by the GQ relative to the network without the GQ.

## 6.2 Counterfactual Network Based on Chinese Strategy

The counterfactual exercise asks how India would develop if it had built a highway network like the Chinese NEN. To replicate the Chinese network in India, 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 location of these cities is shown in Figure 5.

I then design a network to connect the cities through the Indian terrain with the same quality of road as the GQ. I approximate the income-maximizing network among all targeted cities using an iterative procedure that starts from the fully connected network and sequentially removes the least important links, while in each loop recomputing the bilateral driving times and the resulting general equilibrium market access measures to predict income net of road construction costs. I then compare the results to an alternative counterfactual that approximates the income-

Figure 5: Intermediate-sized cities in India.



The figure shows the cities in India with more than 500,000 residents and all state capitals. The image in the background shows night-time lights.

maximizing network without the constraint that all 68 cities are connected.<sup>39</sup> I also compare the results to a network that connects cities in a way that minimizes construction costs as in Faber (2014). Finally, I use additional policy objectives by the Chinese government that specified that a certain number of corridors should be built to connect the targeted cities. I implement this in an ad-hoc way and compare the results to the approximation of the optimal network. The iterative procedure to approximate the optimal network is explained below and further details on the different counterfactual networks can be found in Appendix C.

### 6.2.1 Designing Network that Approximately Maximizes Income

Based on the location of the 68 Indian cities and the road construction costs between them, I build a counterfactual network that connects these cities in an approximately optimal way. While the problem of finding the global optimum for a transport network that connects a large set of nodes in a general equilibrium gravity model has

---

<sup>39</sup>Similar to Jia (2008) and Antras et al. (2016), I search for the network once by starting from the fully connected network and sequentially removing links, and once by starting from the empty network and sequentially adding links.

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.<sup>40</sup> 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 travel costs through the network and they use a heuristic algorithm that finds local optima. I also rely on a heuristic algorithm, but with an objective function based on income in the general equilibrium framework of Section 4.<sup>41</sup>

Starting from the full network, the algorithm removes each link individually from the network and recomputes the resulting shortest paths among all 68 cities. The implied new travel costs after removing a link are then used to solve for the general equilibrium market access measures as given in Equation (14). Based on these market access measures, Equation (16) predicts, for each removed link, the change in income in all locations.<sup>42</sup> Using data on the topography, I can approximate the construction costs and thus calculate the change in aggregate real income net of construction costs. This allows identifying the links with the smallest (most negative) net effect.

The algorithm first reduces the network by removing the links with the most negative effect on net income. 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. 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 6. As we can see, this is a substantially larger network than the GQ and it would cost more than five times as much.

The above network strictly implements the Chinese policy of connecting all tar-

---

<sup>40</sup>Fajgelbaum and Schaal (2016) are able to compute the globally optimal transport network in a general equilibrium spatial model with congestion. Allen and Arkolakis (2016) develop a general equilibrium gravity trade model to characterize the impact of transport infrastructure investment on welfare. They point out that this framework in combination with observed trade flows could be used as a step towards characterizing the optimal network in gravity trade models.

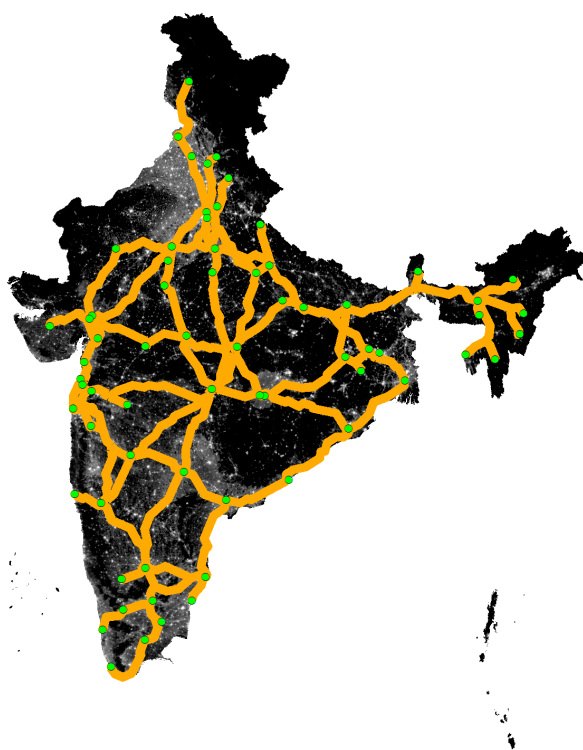
<sup>41</sup>Gastner (2005) discusses various heuristic algorithms to design transport networks, including the iterative procedure that is used here.

<sup>42</sup>For the design of the network, I do not jointly solve Equations (14) and (15) because it is computationally more costly. Instead, I solve Equation (14) for constant income and then use the resulting market access measures to predict changes in income. However, all resulting final network designs are evaluated and compared based on solving both equations jointly.



geted cities and the network may be too large in terms of net aggregate income. I therefore also find the approximation of the optimal 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 equalized. The resulting network, shown in Figure 7, is still more than four times larger than the GQ. 9 of the 68 targeted cities are not connected, but the overall structure of the network is similar.

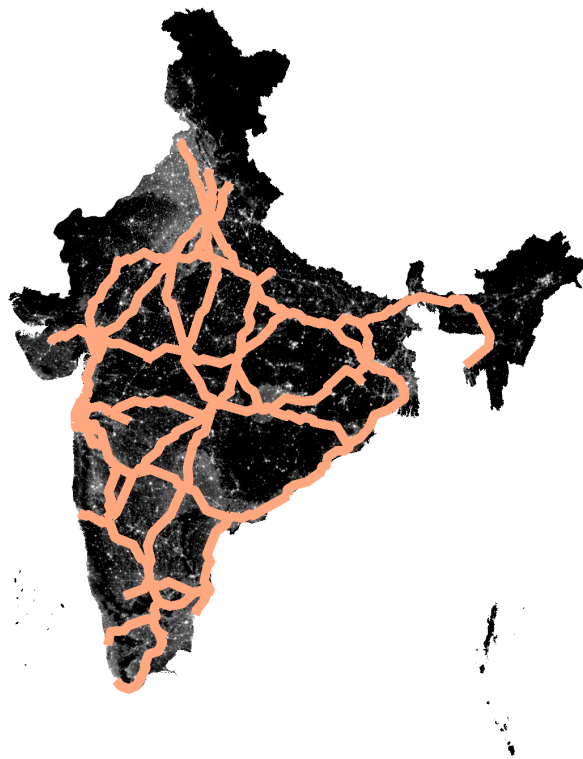
Figure 6: Counterfactual network that connects all 68 cities



The map shows a counterfactual highway network that connects all 68 cities in an approximately optimal way with the iterative procedure.

A caveat of the heuristic algorithm used here based on the sequential elimination of links is that there is no guarantee that it converges to the global optimum. To partially address this concern, I compare the above approximation of the optimal 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 Figure A1. The two networks appear to have a similar structure, relatively similar welfare implications (0.13 percentage points higher for the sequential

Figure 7: Counterfactual network that equalizes marginal costs and benefits



The map shows a counterfactual highway network that is designed to equalize marginal costs and benefits of highway construction with the iterative procedure.

reduction), and they share 77% of the individual links. 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.<sup>43</sup>

### 6.2.2 Aggregate Effects of the Counterfactual Infrastructure

The second rows in Tables 3 and 4 show the aggregate effects of the counterfactual network that connects all 68 cities in an approximately optimal way. GDP in 2012 would have been 2.68% higher with this counterfactual network than with the GQ. Based on the topography, the counterfactual network is predicted to be more than five times as expensive to build as the GQ but it would imply annual net gains

---

<sup>43</sup>Jia (2008) and Antras et al. (2016) can show that in their case the global optimum is within the bounds obtained from sequential reduction and sequential increase of the network. This does not apply here because roads could be complements or substitutes.

of 2.24% of GDP, roughly USD 24.67 billion. Although this is a sizable increase, it should be noted that even the dense counterfactual network only upgrades a fraction of the existing highways to the standard of modern highways, such that the infrastructure overall would still be below the standards of developed countries.

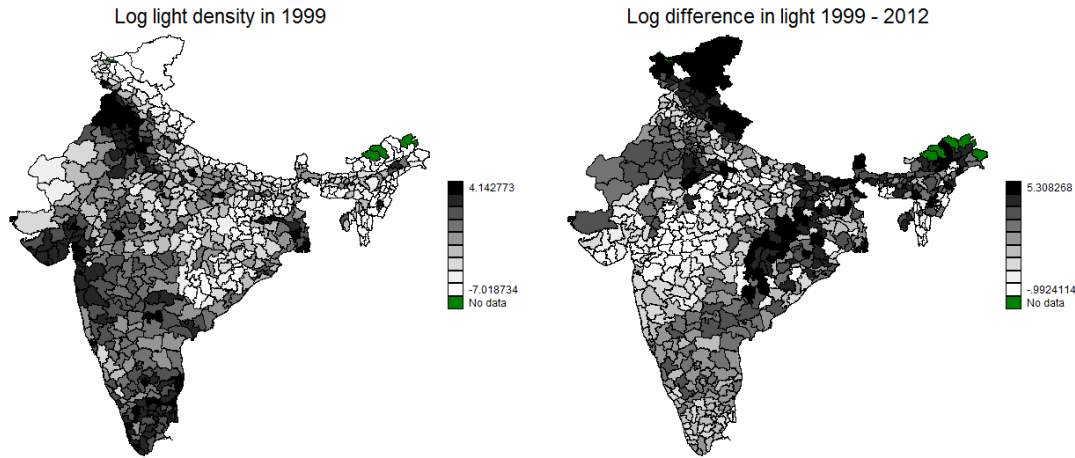
I then consider a counterfactual network without the constraint that all 68 cities must be connected. The network is therefore designed such that the marginal costs and benefits are approximately equalized. This counterfactual network is somewhat smaller than the previous network, but it does not strictly implement the Chinese policy because nine intermediate-sized cities are not connected. As shown in the third row, the aggregate net gain from this approximation of the optimal network is 2.25 percent of GDP. This is only slightly higher than the network that fulfills the Chinese policy objective, which could suggest that the strategy of connecting all intermediate-sized cities is relatively good.

There are some important assumptions that I made for the above cost 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 optimal network in different parts of the country. However, given the magnitudes, it seems unlikely that this would overturn the basic results.

### **6.2.3 Distributional Effects of the Counterfactual Infrastructure**

Replacing the GQ with the counterfactual network that connects intermediate-sized cities in an approximately optimal 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 regions were states with low growth during the past decade (see Figure 8). As there are several cities in that area that would fulfill the criteria of the NEN, they would become better connected by the counterfactual network and experience increases in market access and higher income. Figure 9 shows the distributional effects of the counterfactual relative to the GQ.

Figure 8: Initial Density and Growth in India



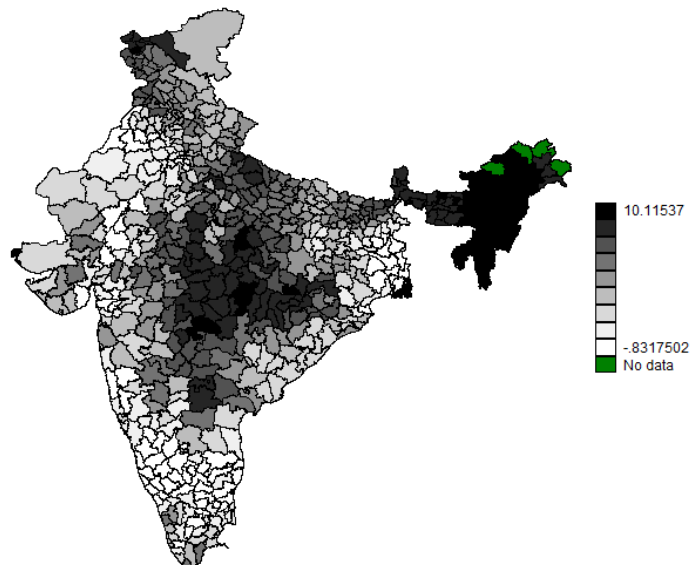
The two maps show the initial density (left map) and growth (right map) in light in India. The initial density is the logarithm of average light intensity per pixel in the year 1999 (averaging 1998, 1999, and 2000). Growth is approximated by the log difference in light intensity between 1999 and 2012. The units are 636 Indian districts. Darker colors refer to higher density or higher growth rates. The green areas represent missing observations due to zeros in the initial light per district.

#### 6.2.4 Other Counterfactual Networks

I also consider alternative ways of implementing the Chinese policy of connecting intermediate-sized cities in India. The comparison of these networks to the above counterfactuals that connect cities in an approximately optimal way provides insights on how different the effects are from alternative ways of designing a network. In particular, the results show that these alternative networks have lower net gains, but the distributional implications are qualitatively similar to the counterfactual networks that connect intermediate-sized cities in an approximately optimal way.

**Least-Cost Network** Faber (2014) uses the minimum spanning tree to predict the 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 10. It represents the cheapest way to formally fulfill the Chinese policy objective (connecting cities which have a population above 500'000

Figure 9: Percent increase in GDP from replacing GQ with counterfactual network that connects 68 cities



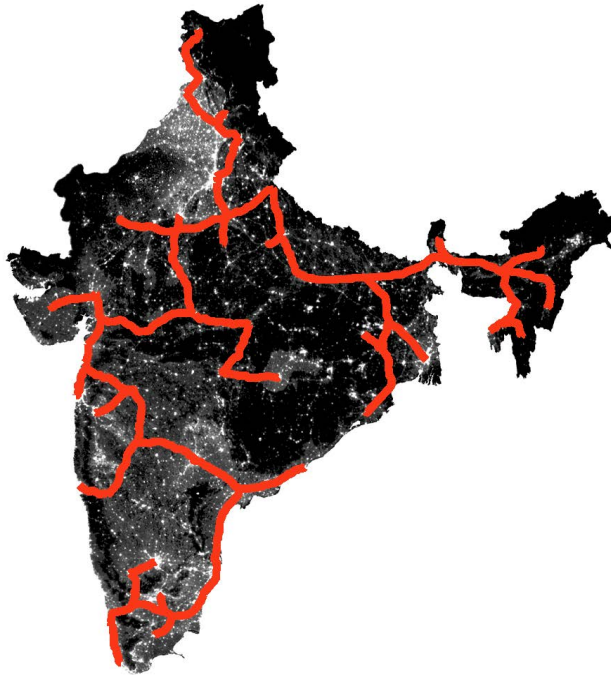
The map shows the boundaries of Indian districts. Darker color represents higher percentage difference in GDP generated by replacing the GQ with the counterfactual network that connects all 68 cities in a way that approximately maximizes income.

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 costs.

As shown in the fourth column of Tables 3 and 4, the least-cost network in the aggregate implies about a 0.38% higher income, or USD 4.19 billion per year. It would cost overall approximately USD 20 billion and imply an annual net gain of about 0.17% of GDP, or USD 1.88 billion, when replacing the GQ. The least-cost network therefore is somewhat better than the GQ in terms of aggregate net income, but the other counterfactuals yield even larger gains. However, the distributional consequences, shown in Appendix Figure A2, are qualitatively similar to the previous counterfactual networks. The least-cost network connects all intermediate-sized cities and thus reaches regions that did not benefit from the GQ.

The least-cost network leads to a much lower increase in net income than the approximation of the optimal 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

Figure 10: Least-cost counterfactual network



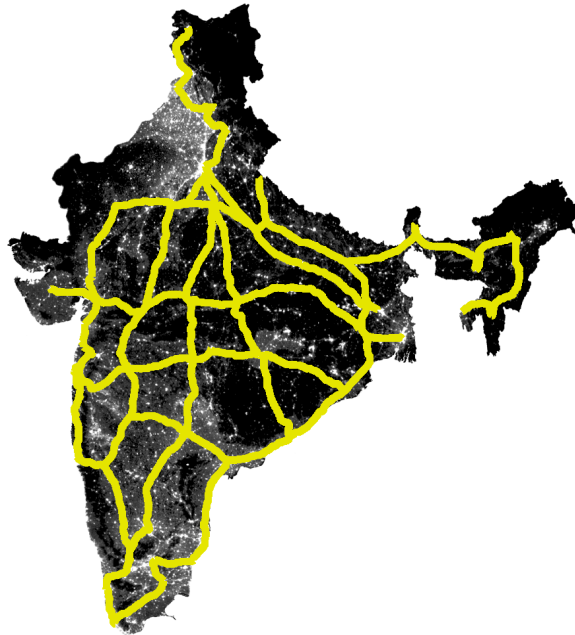
The map shows a counterfactual highway network in India that connects all 68 targeted cities at the least-cost (minimum spanning tree).

compute a network that has the same costs as the least-cost network and maximizes the net income gain under this constraint. The resulting network is shown in Figure A3 and the structure differs substantially from the least-cost network. As reported in Tables 3 and 4, this network leads to a larger increase in net income than the least-cost network, suggesting that the design of the network is important.

**Counterfactual Network Based on Additional Policy Objectives** 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 11 shows an example where these additional criteria are implemented. 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.

As shown in the last columns of Tables 3 and 4, the network based on rays and corridors would generate 2.04% higher income in 2012 and yield annual net gains of 1.64% of GDP, or about USD 18 billion. It is therefore not as effective as the approx-

Figure 11: Counterfactual network with rays and corridors



The map shows a counterfactual highway network in India that connects all 68 targeted cities with rays, horizontal corridors, and vertical corridors following the Chinese strategy.

imation of the optimal network, but still implies a substantial net gain. Appendix Figure A5 shows that the distributional implications are qualitatively similar to the approximation of the optimal 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.

**Comparison to North-South and East-West Corridors** The above evidence on the impact of transport infrastructure illustrates that there are both aggregate and distributional consequences. By choosing a network that connects the four largest economic centers, India has been able to achieve an increase in income that is well above the construction costs of the new or upgraded highways. However, the strategy also implied that lagging regions were neglected by the new infrastructure investments. These distributional consequences are particularly relevant in view of the unequal regional development of India. Policy makers seem to be aware of the need to connect other parts of the country and the NHDP did include plans for other highway connections besides the GQ. In particular, the government planned the NS-EW which cross through regions that were not reached by the GQ. However, these other projects were delayed (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 A6 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.

## 7 Robustness

For the results presented so far, I had to make a number of assumptions to estimate the effect of transport infrastructure and to perform the counterfactual analysis. This section discusses the robustness of the results to different choices for the value of the trade elasticity, inclusion of demand from international trade, pre-trends, and population growth.

### 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 targeted but non-nodal regions. One hypothesis 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 2 again but use as the dependent variable the growth in light between 1992 and 1999. If it was 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 Table 5. 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.



Table 5: Light growth prior to road investment

	(1)	(2)	(3)	(4)	(5)
Log Market Access	-0.453 (0.373)	-0.528 (0.349)	-0.512 (0.351)	-0.444 (0.363)	-0.444 (0.289)
Excluded nodal districts	None	4	5	5	5
Weighting	Yes	Yes	Yes	No	No
Standard errors	Cluster	Cluster	Cluster	Cluster	Robust
N	620	607	606	606	606
Rsq.	0.388	0.385	0.386	0.354	0.354

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 (3 year averages). The explanatory variable is market access computed based on Equation (14) and instrumented with the market access with constant light in (19) with  $\theta = 8$ . 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 are clustered at the state-level except in column 5.

## 7.2 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.<sup>44</sup> I use the total of exports and imports of overseas traffic in 1999 and 2012 (averaging over three years).<sup>45</sup> Similar to the approach in Donaldson and Hornbeck (2016) based on population, I then inflate the income of the districts that host the ports by the international trade flows. Since I measure income using the light data but trade flows are in tons (converted to USD), I first obtain the Gross District Product in 2000 from the Indian Planning

<sup>44</sup>The ports are Kamarajar, Chennai, New Mangalore, Chidambaranar, Kandla, Mormugao, Vishakhapatnam, Jawahar Lal Nehru, Paradip, Mumbai, Cochin, Kolkata and Haldia. See <https://data.gov.in/catalog/traffic-handled-major-ports-india>.

<sup>45</sup>For the average value per ton, I use \$668, which was the average value per ton of imports to the US (United States Department of Transportation, 2002) in the year 2000.

commission.<sup>46</sup> This yields the ratio between the value of trade and each port’s income, which I then use to inflate the light measures in each year. The results are shown in Table 6. The estimates don’t change substantially when these alternative income measures are used.

Table 6: Including trade in 12 major ports

	(1)	(2)	(3)	(4)	(5)
Log Market Access	0.586*	0.608*	0.614*	0.593*	0.593**
	(0.309)	(0.330)	(0.332)	(0.355)	(0.299)
Excluded nodal districts	None	4	5	5	5
Weighting	Yes	Yes	Yes	No	No
Standard errors	Cluster	Cluster	Cluster	Cluster	Robust
N	626	613	612	612	612
Rsquared	0.507	0.500	0.500	0.494	0.494

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. The explanatory variable is market access computed based on Equation (14) and instrumented with the market access with constant light in (19). Exports and imports from major ports is 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 are clustered at the state-level except in column 5.

### 7.3 Alternative Values for the Trade Elasticity

When solving the system of equations in (14) and (19) numerically to derive the general equilibrium market access measures, it is necessary to choose a value for the trade elasticity parameter  $\theta$ . The value of 8 was chosen based on the fit in the 1999 cross-section and it is close to the value used in Donaldson and Hornbeck (2016). However, alternative values for the trade elasticity have been found in the literature

<sup>46</sup>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. See <https://data.gov.in/catalog/district-wise-gdp-and-growth-rate-current-price1999-2000>.

as well. Appendix tables A1 - A3 report the estimated effect of market access on light with values for the trade elasticity  $\theta$  of 4, 6, and 10. The point estimates for the preferred specifications in columns 3 range from 0.556 to 1.181 and are generally significant at the 10 percent level. Note that the elasticity predicted by the model also changes in the trade elasticity.

## 7.4 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 decadal change in each district's population between the 2001 and 2011 census on the change in market access due to transport investments. In other words, the dependent variable in specification (18) is replaced by population. The results in Table 7 show that there is no significant effect of market access on population.

Table 7: Effect of transport infrastructure investments on population

	(1)	(2)	(3)	(4)	(5)
Log Market Access	0.122 (0.113)	0.0975 (0.116)	0.0619 (0.126)	0.0523 (0.126)	0.0523 (0.0986)
Excluded nodal districts	None	4	5	5	5
Weighting	Yes	Yes	Yes	No	No
Standard errors	Cluster	Cluster	Cluster	Cluster	Robust
N	627	614	613	614	614
Rsqr.	0.198	0.200	0.205	0.246	0.246

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 (14) and instrumented with the market access with constant light in (19) with  $\theta = 8$ . 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 are clustered at the state-level except in column 5.

## 8 Conclusion

Investments in transport infrastructure are often at the heart of efforts to foster economic development, as it is generally agreed that insufficient transport infrastructure is an important constraint in many countries. However, the impact of these investments is difficult to identify due to the general equilibrium consequences of transport networks. Furthermore, we often lack sources of exogenous variation in infrastructure.

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 implements the Chinese highway network in India. 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 targets intermediate-sized cities as suggested by the Chinese policy and connects them in an approximately optimal way. Such a network is substantially larger than the existing network but 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. Furthermore, the results suggest that a substantially denser network of modern highways would be beneficial, since the approximation of the optimal network shows that large additional gains could be obtained from increasing the network of modern highways by a factor of four to five relative to the GQ.

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 the later are important. This suggests that the debates about infrastructure investments in other countries should give careful consideration to the distributional consequences of alternative networks. 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 I plan to consider how distributional objectives could be taken into account directly in the design of the network.

## 9 References

- Allen, Treb and Costas Arkolakis** forthcoming. "Trade and the Topography of the Spatial Economy." *Quarterly Journal of Economics* 129(3): 1085-1140.
- Allen, Treb and Costas Arkolakis** 2016. "The Welfare Effects of Transportation Infrastructure Improvements." Working Paper.
- Allen, Treb, and David Atkin** 2016. "Volatility and the Gains from Trade." *NBER Working Paper* 22276.
- Anderson, James E. and Eric van Wincoop** 2003. "Gravity with Gravitas: A Solution to the Border Puzzle." *American Economic Review*, 93(1): 170-192.
- Antras, Pol, Teresa C. Fort, and Felix Tintelnot** 2016. "The margins of global sourcing: theory and evidence from US firms." Working Paper.
- Asher, Sam, and Paul Novosad** 2016. "Market Access and Structural Transformation: Evidence from Rural Roads in India." Working Paper.
- Asian Development Bank** 2007. "Retrospective Analysis of the Road Sector, 1997-2005." *Asian Development Bank*.
- Asturias, Jose, Manuel García-Santana, and Roberto Ramos** 2016. "Competition and the welfare gains from transportation infrastructure: Evidence from the Golden Quadrilateral of India." Working Paper.
- Atack, Jeremy, Michael R. Haines, and Robert A. Margo** 2008. "Railroads and the Rise of the Factory: Evidence for the United States, 1850-70." *NBER Working Paper* 14410.
- Au, Chun-Chung and J. Vernon Henderson** 2006. "Are Chinese Cities Too Small?" *Review of Economic Studies*, 73, p. 549-576.
- Australian Consortium for the Asian Spatial Information and Analysis Network (ACASIAN)** 2013. "People's Republic of China Spatio-Temporal Expressway Database." <http://acasian.com/price.html#TRs>.
- Balboni, Clare** 2016. "Living on the Edge: Infrastructure Investments and the Persistence of Coastal Cities." Working Paper.

- Banerjee, Abhijit, Esther Duflo, and Nancy Qian** 2012. "On the Road: Access to Transport Infrastructure and Economic Growth in China." *NBER Working Paper* 17897.
- Baum-Snow, Nathaniel** 2007. "Did Highways Cause Suburbanization?" *The Quarterly Journal of Economics*, 122(2): 775-805.
- Baum-Snow, Nathaniel, Loren Brandt, J. Vernon Henderson, Matthew A. Turner, and Qinghua Zhang** forthcoming. "Roads, Railroads and Decentralization of Chinese Cities." *Review of Economics and Statistics*.
- Baum-Snow, Nathaniel, Loren Brandt, J. Vernon Henderson, Matthew A. Turner, and Qinghua Zhang** 2016. Transport Infrastructure, Urban Growth and Market Access in China. Working Paper.
- Bird, Julia, and Stéphane Straub** 2015. "The Brasilia experiment: road access and the spatial pattern of long-term local development in Brazil." Working Paper.
- Breinlich, Holger, Gianmarco I. P. Ottaviano, and Jonathan R.W. Temple** 2013. "Regional Growth and Regional Decline" *Handbook of Economic Growth*.
- Burgess, Robin, Remi Jedwab, Edward Miguel, and Ameet Morjaria** 2015. "The value of democracy: evidence from road building in Kenya." *The American Economic Review* 105, no. 6: 1817-1851.
- Census of India** 2001. "Migration." [http://censusindia.gov.in/Census\\_And\\_You/migrations.aspx](http://censusindia.gov.in/Census_And_You/migrations.aspx)
- Chandra, Amitabh and Eric Thompson** 2000. "Does Public Infrastructure Affect Economic Activity? Evidence from the Rural Interstate Highway System." *Regional Science and Urban Economics*, 30(4): 457-490.
- Chaudhuri, Shubham and Martin Ravallion** 2006. "Partially Awakened Giants: Uneven Growth in China and India" *World Bank Policy Research Working Paper* 4069.
- Chen, Xi and William D. Nordhaus** 2011. "Using Luminosity Data as a Proxy for Economic Statistics." *Proceedings of the National Academy of Sciences* 108(21): 8589-8594.



- Chinese Ministry of Transportation** 2004. "Plans for the National Expressway Network" (in Chinese).
- Datta, Saugato** 2012. "The Impact of Improved Highways on Indian Firms." *Journal of Development Economics* 99: 46-57.
- Dell, Melissa** 2015. "Trafficking Networks and the Mexican Drug War." *American Economic Review*, 105(6): 1738-1779.
- DHL** 2007. "Logistics in India." DHL, <http://www.dhl-discoverlogistics.com/cms/en/course/trends/asia/india.jsp>
- Desmet, Klaus, Ejaz Ghani, Stephen O'Connell, and Esteban Rossi-Hansberg** 2013. "The Spatial Development of India" *Journal of Regional Science* 55, 10-30.
- Desmet, Klaus and Esteban Rossi-Hansberg** 2014. "Spatial Development." *American Economic Review* 104(4): 1211-1243.
- Donaldson, Dave** forthcoming. "Railroads and the Raj: Estimating the Impact of Transport Infrastructure" *American Economic Review*.
- Donaldson, Dave** 2015. "The gains from market integration." *Annual Review of Economics* 7: 619-647.
- Donaldson, Dave and Richard Hornbeck** 2016. "Railroads and American Economic Growth: A Market Access Approach" *The Quarterly Journal of Economics*, 131 (2): 799-858.
- Duranton, Gilles and Matthew A. Turner** 2012. "Urban growth and transportation." *The Review of Economic Studies* 79(4): 1407-1440.
- Duranton, Gilles, Peter M. Morrow, and Matthew A. Turner** 2014. "Roads and Trade: Evidence from the US." *The Review of Economic Studies* 81(2): 681-724.
- Earth Observation Group** 2015. "Version 4 DMSP-OLS Nighttime Lights Time Series." National Oceanic and Atmospheric Administration, <http://ngdc.noaa.gov/eog/index.html>.
- Eaton, Jonathan and Samuel Kortum** 2002. "Technology, Geography, and Trade" *Econometrica* 70(5): 1741-1779.

- Elvidge, Christopher D., Kimberly E. Baugh, Eric A. Kihn, Herbert W. Kroehl, Ethan R. Davis, and C. W. Davis** 1997. "Relation Between Satellite Observed Visible-Near Infrared Emissions, Population, Economic Activity and Electric Power Consumption." *International Journal of Remote Sensing* 18(6): 1373-1379.
- Ernst and Young** 2013. "Infrastructure 2013: Global Priorities, Global Insights".
- Esri** 2013. "World Street Map" Redlands, CA: Esri. <http://www.esri.com/software/arcgis/arcgis-online-map-and-geoservices/map-services>.
- Faber, Benjamin** 2014. "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System." *The Review of Economic Studies*.
- Fajgelbaum, Pablo D. and Edouard Schaal** 2016. "Optimal Transport Networks in Spatial Equilibrium." Working Paper.
- Felbermayr, Gabriel J., and Alexander Tarasov** 2015. "Trade and the Spatial Distribution of Transport Infrastructure." Available at SSRN: <https://ssrn.com/abstract=2706533>.
- Frankfurter Allgemeine Zeitung (FAZ)** 2013. "Indien wird China überholen." [www.faz.net](http://www.faz.net) (in German).
- Fujita, Masahisa, Paul Krugman, and Anthony J. Venables** 1999. "The Spatial Economy. Cities Regions, and International Trade." MIT Press, Cambridge.
- Gastner, Michael T.** 2005. "Spatial Distributions: Density-Equalizing Map Projections, Facility Location, and Two-Dimensional Networks", Thesis, The University of Michigan.
- Gastner, Michael T., and M. E. J. Newman** 2006. "Optimal design of spatial distribution networks", *Physical Review*, E 74, no. 1.
- Ghani, Ejaz, Arti Grover Goswami, and William R. Kerr** 2015. "Highway to success: The impact of the Golden Quadrilateral project for the location and performance of Indian manufacturing." *The Economic Journal*.
- Ghosh, Tilottama, Rebecca L. Powell, Christopher D. Elvidge, Kimberly E. Baugh, Paul C. Sutton, and Sharolyn Anderson** 2010. "Shedding Light on the Global Distribution of Economic Activity." *The Open geography Journal* 3: 148-61.

- Glaeser, Edward L. and Janet E. Kohlhase** 2004. "Cities, regions and the decline of transport costs." *Papers in Regional Science* 83(1): 197-228.
- Global Land Cover Facility** 2013. "UMD Land Cover Classification." University of Maryland Department of Geography, <http://glcf.umd.edu/data/landcover/>
- Hanson, Gordon H.** 2005. "Market potential, increasing returns and geographic concentration." *Journal of International Economics* 67: 1-24.
- Harral, Clel , Jit Sondhi, and Guang Zhe Chen** 2006. "Highway and Railway Development in India and China, 1992-2002." *Transport Note TRN-32*. The World Bank, Washington, D.C.
- Head, Keith and Thierry Mayer** 2011. "Gravity, Market Potential and Economic Development." *Journal of Economic Geography* 11: 281-294.
- Head, Keith and Thierry Mayer** 2013. "Gravity Equations: Workhorse, Toolkit, and Cookbook." *Handbook of International Economics* 4.
- Henderson, J. Vernon, Adam Storeygard, and David N. Weil** 2012. "Measuring Economic Growth from Outer Space." *American Economic Review* 102(2): 994-1028.
- Herrendorf, Berthold, James A. Schmitz Jr., and Arilton Teixeira** 2012. "The Role of Transportation in U.S. Economic Development: 1840-1860." *International Economic Review* 53(3): 693-715.
- Hodler, Roland and Paul A. Raschky** 2014. "Regional Favoritism". *The Quarterly Journal of Economics*, 129(2): 995-1033.
- Indian Ministry of Railways** 2012. "Indian Railway: Facts and Figures."
- Jarvis, Andy, Hannes Isaak Reuter, Andrew Nelson, and Edward Guevara** 2008. "Hole-filled SRTM for the globe Version 4" CGIAR-CSI SRTM 90m Database, <http://srtm.csi.cgiar.org>.
- Jaworski, Taylor, and Carl T. Kitchens** 2016. "National Policy for Regional Development: Evidence from Appalachian Highways." *NBER Working Paper* 22073.
- Jia, Panle** 2008. "What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry," *Econometrica*, 76 (6), 1263-1316.

- Khanna, Gaurav** 2016. "The Road Oft Taken: The Route to Spatial Development." SSRN: <https://ssrn.com/abstract=2426835>.
- KPMG** 2013. "Logistics Games Changers: Transforming India's Logistic Industry."
- Kruskal, Joseph B.** 1956. "On the shortest spanning subtree of a graph and the traveling salesman problem." *Proceedings of the American Mathematical society* 7(1): 48-50.
- Limao, Nuno and Anthony J. Venables** 2001. "Infrastructure, Geographical Disadvantage, Transport Costs, and Trade." *The World Bank Economic Review* 15(3): 451-479.
- Ma, Ting, Chenghu Zhou, Tao Pei, Susan Haynie, and Junfu Fan** 2012. "Quantitative Estimation of Urbanization Dynamics Using Time Series of DMSP/OLS Nighttime Light Data: A Comparative Case Study from China's Cities." *Remote Sensing of Environment* 124: 99-107.
- Michaels, Guy** 2008. "The Effect of Trade on The Demand for Skill: Evidence from the Interstate Highway System" *The Review of Economics and Statistics* 90(4): 683-701.
- Morten, Melanie, and Jaqueline Oliveira** 2016. "Paving the Way to Development: Costly Migration and Labor Market Integration." *NBER Working Paper* 22158.
- National Highway Authority of India (NHAI)** 2010. "National Highways Development Project Phase." Prepared by Information Technology & Planning Division.
- National Highway Authority of India (NHAI)** 2013. "National Highways Development Project Phase - I,II, & III." Prepared by Information Technology & Planning Division.
- Puga, Diego** 2002. "European Regional Policies in Light of Recent Location Theory." *Journal of Economic Geography* 2: 373-406.
- Redding, Stephen J.** 2016. "Goods Trade, Factor Mobility and Welfare." *Journal of International Economics*, 101, 148-167.

- Redding, Stephen J. and Daniel M. Sturm** 2008. "The Costs of Remoteness: Evidence from German Division and Reunification." *American Economic Review* 98(5): 1766-1797.
- Redding, Stephen J. and Matthew A. Turner** 2015. "Transportation costs and the spatial organization of economic activity." *Elsevier Handbook of Urban and Regional Economics* 5.
- Redding, Stephen J. and Anthony Venables** 2004. "Economic geography and international inequality." *Journal of International Economics* 62: 53-82.
- Roberts, Mark, Uwe Deichmann, and Bernard Fingleton, Tuo Shi** 2012. "Evaluating China's road to prosperity: A new economic geography approach." *Regional Science and Urban Economics* 42: 580-594.
- Simonovska, Ina and Michael E. Waugh** 2014. "The Elasticity of Trade: Estimates and Evidence." *Journal of International Economics* 94: 34-50.
- Storeygard, Adam** 2016. "Farther on down the Road: Transport costs, trade and urban growth in Sub-Saharan Africa." *Review of Economic Studies*, 83(3): 1263-1295.
- United States Department of Transportation** 2002. "Value per metric ton of U.S. waterborne imports and exports."
- Van Leemput, Eva** 2015. "A Passage to India: Quantifying Internal and External Barriers to Trade." Working Paper.
- World Bank** 2002. "India's Transport Sector: The Challenges Ahead."
- World Bank** 2005. "India: Road Transport Service Efficiency Study." Energy & Infrastructure Operations Division.
- World Bank** 2007a. "A Decade of Action in Transport. An Evaluation of World Bank Assistance to the Transport Sector, 1995-2005." The World Bank, Washington, D.C.
- World Bank** 2007b. "An Overview of China's Transport Sector - 2007" *EASTE Working Paper* 15.

**World Bank** 2008. "India: Accelerating Growth and Development in the Lagging Regions of India." *Poverty Reduction and Economic Management* Report No. 41101-IN.

**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

December 3, 2016

## A Model Details

This appendix 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 and digitized geographic data alone, 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 9-year period which I consider.

The basic setup is a Ricardian 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  $j$ ,

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

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 a vector of 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),

$$x_o(j) = z_o(j) (L_o(j))^\alpha (H_o(j))^\gamma (K_o(j))^{1-\alpha-\gamma},$$

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 within districts (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:

$$\begin{aligned} p_{oo}(j) &= MC_o(j) = \frac{q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma}}{z_o(j)} \\ p_{od}(j) &= \tau_{od}MC_o(j) = \tau_{od} \frac{q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma}}{z_o(j)} \\ z_o(j) &= \tau_{od} \frac{q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma}}{p_{od}(j)} \end{aligned} \tag{A1}$$

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

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

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

$$G_d(p) = P[P_d \leq p] = 1 - \prod_o \{1 - G_{od}(p)\}.$$

Inserting for  $G_{od}(p)$  yields

$$\begin{aligned} G_d(p) &= 1 - \prod_o \{ \exp [ -T_o (\tau_{od} q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma})^{-\theta} p^\theta ] \} \\ &= 1 - \exp \left[ - \sum_o [T_o (\tau_{od} q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma})^{-\theta}] p^\theta \right] \\ &= 1 - \exp \left[ - \Phi_d p^\theta \right] \end{aligned}$$

where the destination-specific parameter  $\Phi_d = \sum_o [T_o (\tau_{od} q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma})^{-\theta}]$  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

$$P_d = \mu \Phi_d^{-\frac{1}{\theta}} \quad (\text{A2})$$

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

$$\begin{aligned} P_d^{-\theta} &= \kappa_1 \sum_o [T_o (\tau_{od} q_o^\alpha w_o^\gamma)^{-\theta}] \\ &= \kappa_1 \sum_o [T_o (q_o^\alpha w_o^\gamma)^{-\theta} \tau_{od}^{-\theta}] \equiv CMA_d. \end{aligned} \quad (\text{A3})$$

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

$$\begin{aligned} \frac{X_{od}}{X_d} &= \frac{T_o (q_o^\alpha w_o^\gamma r^{1-\alpha-\gamma})^{-\theta} \tau_{od}^{-\theta}}{\Phi_d} \\ &= \frac{T_o (q_o^\alpha w_o^\gamma r^{1-\alpha-\gamma})^{-\theta} \tau_{od}^{-\theta}}{\sum_o [T_o (q_o^\alpha w_o^\gamma r^{1-\alpha-\gamma})^{-\theta} \tau_{od}^{-\theta}]}. \end{aligned} \quad (\text{A4})$$

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

$$\begin{aligned} X_{od} &= \underbrace{T_o (q_o^\alpha w_o^\gamma)^{-\theta}}_{\text{Origin's productivity and factor costs}} \\ &\times \underbrace{Y_d}_{\text{Destination's income}} \\ &\times \underbrace{\left( \sum_o [T_o (q_o^\alpha w_o^\gamma)^{-\theta} \tau_{od}^{-\theta}] \right)^{-1}}_{\text{Destination's CMA}} \underbrace{\tau_{od}^{-\theta}}_{\text{Trade costs}}. \end{aligned}$$

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

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

which yields

$$\begin{aligned} X_{od} &= \underbrace{T_o(q_o^\alpha w_o^\gamma)^{-\theta}}_{\text{Origin's productivity and factor costs}} \\ &\times \underbrace{Y_d}_{\text{Destination's income}} \\ &\times \underbrace{\kappa_1 CMA_d^{-1}}_{\text{Destination's CMA}} \underbrace{\tau_{od}^{-\theta}}_{\text{Trade costs}} \end{aligned} \quad (\text{A5})$$

## 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  $o$ ,

$$Y_o = \sum_d X_{od} = \kappa_1 T_o(q_o^\alpha w_o^\gamma)^{-\theta} \sum_d [\tau_{od}^{-\theta} CMA_d^{-1} Y_d]. \quad (\text{A6})$$

Donaldson and Hornbeck define "firm market access" of district  $o$  as

$$FMA_o \equiv \sum_d \tau_{od}^{-\theta} CMA_d^{-1} Y_d, \quad (\text{A7})$$

such that

$$Y_o = \kappa_1 T_o(q_o^\alpha w_o^\gamma)^{-\theta} FMA_o. \quad (\text{A8})$$

$FMA_o$  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

$$\begin{aligned}
CMA_d &= \kappa_1 \sum_o T_o (q_o^\alpha w_o^\gamma)^{-\theta} \tau_{od}^{-\theta} \\
&= \sum_o \tau_{od}^{-\theta} FMA_o^{-1} Y_o \\
CMA_o &= \sum_d \tau_{od}^{-\theta} FMA_d^{-1} Y_d.
\end{aligned} \tag{A9}$$

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”.<sup>47</sup> 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 equation 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

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

---

<sup>47</sup>Note that symmetric trade costs ( $\tau_{od} = \tau_{do}$ ) imply that  $\frac{FMA_o}{CMA_o} = \frac{\sum_d \tau_{od}^{-\theta} CMA_d^{-1} Y_d}{\sum_d \tau_{od}^{-\theta} FMA_d^{-1} Y_d} = \frac{\sum_o \tau_{od}^{-\theta} CMA_o^{-1} Y_o}{\sum_o \tau_{od}^{-\theta} FMA_o^{-1} Y_o} = \frac{FMA_d}{CMA_d}$ , such that the ratio is a constant.

## 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

$$Y_o = \kappa_1 T_o (q_o^\alpha w_o^\gamma)^{-\theta} MA_o \quad (\text{A12})$$

where income is a function of productivity, factor prices, and market access. The constant  $\kappa_1$  includes the interest rate, which is equalised 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

$$\begin{aligned} q_o L_o &= \alpha Y_o \\ w_o H_o &= \gamma Y_o. \end{aligned}$$

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

$$Y_o = (\kappa_1 T_o)^{\frac{1}{1+\theta\alpha+\theta\gamma}} \left(\frac{\alpha}{L_o}\right)^{\frac{-\theta\alpha}{1+\theta\alpha+\theta\gamma}} \left(\frac{\gamma}{H_o}\right)^{\frac{-\theta\gamma}{1+\theta\alpha+\theta\gamma}} (MA_o)^{\frac{1}{1+\theta\alpha+\theta\gamma}} \quad (\text{A13})$$

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

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

to obtain

$$Y_o^r = \frac{Y_o}{P_o} = (\kappa_2 T_o)^{\frac{1}{1+\theta\alpha+\theta\gamma}} \left(\frac{\alpha}{L_o}\right)^{\frac{-\theta\alpha}{1+\theta\alpha+\theta\gamma}} \left(\frac{\gamma}{H_o}\right)^{\frac{-\theta\gamma}{1+\theta\alpha+\theta\gamma}} (MA_o)^{\frac{1+\theta(1+\alpha+\gamma)}{(1+\theta\alpha+\theta\gamma)\theta}} \quad (\text{A14})$$

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

$$\begin{aligned} \ln(Y_o^r) &= \underbrace{\frac{1}{1+\theta\alpha+\theta\gamma} \ln(\kappa_2) - \frac{\theta\alpha}{1+\theta\alpha+\theta\gamma} \ln\left(\frac{\alpha}{L_o}\right) - \frac{\theta\gamma}{1+\theta\alpha+\theta\gamma} \ln\left(\frac{\gamma}{H_o}\right)}_{\text{Constant over districts or time}} \\ &+ \underbrace{\frac{1}{1+\theta\alpha+\theta\gamma} \ln(T_o)}_{\text{Productivity}} + \underbrace{\frac{1+\theta(1+\alpha+\gamma)}{(1+\theta\alpha+\theta\gamma)\theta} \ln(MA_o)}_{\text{Market access}}. \end{aligned} \quad (\text{A15})$$

This equation suggests a log-linear relationship between real income and transport infrastructure, where the effect of transport infrastructure goes through the measure of market access. The elasticity of income with respect to market access,

$$\beta = \frac{1 + \theta(1 + \alpha + \gamma)}{(1 + \theta\alpha + \theta\gamma)\theta},$$

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 and then solve for the new equilibrium. This requires jointly solving Equations (A11) and (A14). However, productivity, immobile factor production factors, and  $\kappa_2$  (which includes the interest rate) are unobserved in Equation (A14). In order to solve the two equations for all counterfactual networks, I first back out the unobserved term and then solve for equilibrium outcomes of the counterfactuals while holding this term constant.<sup>48</sup>

More precisely, I first solve the market access Equation (A11) for the actual transport network and income as observed in 2012. Using these measures in Equation (A14) together with observed income in 2012 then allows backing out the unobserved term, denoted as  $B$ ,

$$Y_o^r = \underbrace{(\kappa_2 T_o)^{\frac{1}{1+\theta\alpha+\theta\gamma}} \left(\frac{\alpha}{L_o}\right)^{\frac{-\theta\alpha}{1+\theta\alpha+\theta\gamma}} \left(\frac{\gamma}{H_o}\right)^{\frac{-\theta\gamma}{1+\theta\alpha+\theta\gamma}}}_{B} (MA_o)^{\frac{1+\theta(1+\alpha+\gamma)}{(1+\theta\alpha+\theta\gamma)\theta}}. \quad (\text{A16})$$

The unobserved term  $B$  is assumed to be constant over the counterfactuals and I can therefore jointly solve Equations (A11) and (A14) for all counterfactuals. The effect of each counterfactual on income is then expressed relative to actual income in 2012.

Holding  $B$  constant at its 2012 value implies that productivities and the interest rate are constant. The latter implies a perfectly elastic supply of capital from international markets. The former implies that overall productivity levels are not directly affected by improvements in transport infrastructure. This is a strong assumption,

---

<sup>48</sup>Donaldson and Hornbeck (2016) use this approach in a setting with mobile labor. In that case utility is equalized across locations and they present the results when holding utility constant (i.e., there is international migration in response to the transport infrastructure) or total national population constant (i.e., there is no international migration).

but one may argue that such effects would be positive and thus the effect of transport infrastructure could be underestimated. It should also be noted that predicting changes in income based on changes in market access alone, i.e. holding income fixed in the market access equation), relies on the same assumption.

## **B Data**

### **B.1 Administrative Boundaries and Light Data**

The district boundaries are from the 2011 census 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.<sup>49</sup> 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 1.

### **B.2 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 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 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)

---

<sup>49</sup>The last two years have recently been made available and they have not been included in the present analysis.

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.<sup>50</sup> 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 taken to be the same as for the the Chinese expressways and the GQ, which according to google maps is 75 km/h. To take into account the usual waiting time of trucks at state borders, I include in the digital network a cost of crossing state borders. 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.

## C Construction of Counterfactual Networks

### C.1 Approximation of Income-Maximizing Network

This section discusses the procedure that is used to approximate the optimal network among the 68 cities.

**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 market access and aggregate income. These steps are explained in detail below.

---

<sup>50</sup>These estimates are in line with more recent numbers by KPMG (2013).



**Step 2a: Shortest Paths** I recompute all bilateral driving times based on a shortest path algorithm. I do this with a traditional 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.<sup>51</sup>

**Step 2b: General Equilibrium Market Access Measures** For each network and implied iceberg trade costs, I recompute the general equilibrium market access measures using Equation (19),<sup>52</sup> where  $Y_d^r$  are the real incomes of the 68 cities measured by light in the year 1999,  $\tau_{od}$  are iceberg trade costs among these cities, and  $\theta$  is the trade elasticity which I set to 8.<sup>53</sup>

**Step 2c: Predicted Income** Equation (16) in the conceptual framework predicts a relationship between a district's income and its market access, with a constant elasticity  $\beta$ . Income for any given market access can therefore be predicted based on the initial income and the log change in market access between the initial and the new market access:

$$\ln Y_{o,n} = \ln Y_{o,n_0} + \beta(\ln MA_{o,n} - \ln MA_{o,n_0}), \quad (\text{A17})$$

where  $n_0$  denotes the initial income and market access, and  $n$  denotes the current ones (based on the network in this iteration). As initial real income, I take the national GDP in 2000 that is assumed to be distributed among the 68 cities in proportion to their light (calculated within a 30 km buffer around each city). To calculate initial market access, I take the transport network with only conventional highways (the network in 2000, before the GQ was built). The elasticity of light with respect to market access is estimated in Section 5.2 based on the GQ as  $\hat{\beta} = 0.662$  and the elasticity of GDP with respect to light is assumed to be 0.3 as in Henderson et al. (2012). Hence, I can write income predicted by each version of a modern highway network

---

<sup>51</sup>Different from Section 5.1.2, driving times here are computed using a version of Dijkstras algorithm based on a graph instead of the Fast Marching Algorithm on a discrete cost grid as in Allen and Arkolakis (2014). The implementation of Dijkstras algorithm is obtained from the Octave Network Toolbox, which can be obtained at <http://aeolianine.github.io/octave-networks-toolbox>.

<sup>52</sup>I have also used an approximation of market access,  $MA_o = \sum_{d \neq o} \frac{Y_d^r}{\tau_{od}^\theta}$ . The resulting network based on the approximation is larger, which is to be expected because it does not account for the general equilibrium effect in Equation (19) of other location's market access.

<sup>53</sup>The iceberg trade costs are again obtained from the driving times following Equation (20) where  $\gamma$  is chosen such that the median iceberg trade costs in 2000 is 1.25.

based on initial income, the change in market access, the estimate for the elasticity  $\beta$ , and the elasticity of GDP with respect to light. I then sum over the incomes of each district in order to obtain aggregate income.

**Step 2d: Net Change in Aggregate Income** The net effect of removing a highway link is the change in aggregate real income per year net of capital costs from the construction and the maintenance costs,

$$\Delta I = Y_{o,n}^r - Y_{o,n-1}^r + b, \quad (\text{A18})$$

where I use  $n$  to denote the network after removing the link and  $n - 1$  the network before removing the link.  $b$  is the annual cost due to the construction of the link, including construction and maintenance.

One challenge is that the construction costs in the data are in units based on the topographical features as described in Equation 21, while the market access measures and income are based on light data. In order to weigh these two terms appropriately in the objective function, I calculate the ratio of the GQ's construction costs in USD to its construction costs based on the topography. This ratio then allows me to express each 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 constructing each road segment. The resulting change in net income from removing a link then is

$$\Delta I = Y_{o,t}^r - Y_{o,t-1}^r + 0.17\omega c, \quad (\text{A19})$$

where  $\omega$  is the ratio of construction costs of the GQ in USD and to the construction costs based on the topography ( $c$ ).

**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 considered. I then select those links that lead to the largest increase in net income when 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 new links and then adds the most useful one if it implies an increase in net income.<sup>54</sup>

**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. I therefore add a last 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 income. The resulting network, shown in Figure 6, is then used as an approximation of the optimal network. I combine this network with the existing network of conventional highways in the year 2000 and use the fast marching algorithm in order to compute the driving times among all 636 Indian districts, as described in Section 5.1.2.

**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 when no links can be added or removed to increase net income. Nine cities would not be connected without this step, all of them in periphery regions in the north and east. This network is shown in Figure 7 and in the counterfactual analysis this network is used as an approximation of the unconstrained optimum.

**Starting from Empty 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 and sequentially adding links.<sup>55</sup> The result is shown in Figure A1. The welfare implications of the two solutions are similar. The networks also appear to have a similar structure, although 23% of the individual links are different.

---

<sup>54</sup>Gastner and Newman (2006) also include this intermediate step.

<sup>55</sup>The comparison of the solution when starting from the full network and from the empty network is similar to Jia (2008) and Antras et al. (2016).

**Approximation of National Income Based on 68 Cities** A further caveat is that changes in market access and income are computed only for the 68 cities that are targeted by the Chinese policy. By using the Indian aggregate GDP in 2000 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 2000). 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 many parts of the country.

## **C.2 Approximation of the 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. Figure A3 shows the network that has the same construction costs as the minimum spanning tree. Since the minimum spanning tree is the cheapest way of connecting the 68 cities in one common network, it is clear that 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 algorithm tends to provide better connections to locations with higher income, at the cost of not including certain cities in the network, as can be seen in Figure A3. However, the resulting network more closely resembles a transport network than the least-cost network as the solution to the minimum spanning tree.

Figure A4 shows a network that has about the same overall length as the network in Figure 11 (based on ad-hoc rays and corridors), but it approximates the constrained optimal 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 market access and road construction costs, the algorithm tends to build more connections to locations with higher income, since they affect market access more.

## D Additional Tables and Figures

Table A1: Effect of Actual Transport Infrastructure Investments with Trade Elasticity ( $\theta$ ) of 4

	(1)	(2)	(3)	(4)	(5)
Log Market Access	1.125*	1.167*	1.181*	1.121	1.121**
	(0.589)	(0.631)	(0.636)	(0.695)	(0.545)
Excluded nodal districts	None	4	5	5	5
Weighting	Yes	Yes	Yes	No	No
Standard errors	Cluster	Cluster	Cluster	Cluster	Robust
N	626	613	612	612	612
Rsqr.	0.507	0.500	0.501	0.494	0.494

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. The explanatory variable is market access computed based on Equation (14) and instrumented with the market access with constant light in (19) with  $\theta = 4$ . 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 are clustered at the state-level except in column 5.

Table A2: Effect of Actual Transport Infrastructure Investments with Trade Elasticity ( $\theta$ ) of 6

	(1)	(2)	(3)	(4)	(5)
Log Market Access	0.792*	0.826*	0.835*	0.818*	0.818**
	(0.422)	(0.451)	(0.455)	(0.490)	(0.382)
Excluded nodal districts	None	4	5	5	5
Weighting	Yes	Yes	Yes	No	No
Standard errors	Cluster	Cluster	Cluster	Cluster	Robust
N	626	613	612	612	612
Rsquared	0.507	0.501	0.501	0.494	0.494

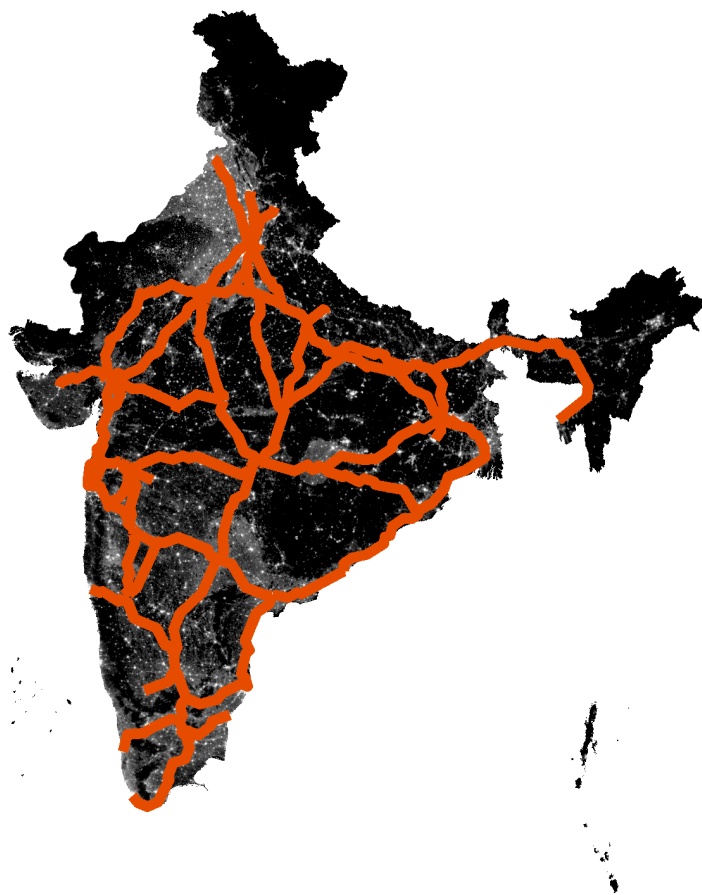
The table shows the 2SLS estimates of the elasticity of light with respect to market access. The dependent variable is the logarithm of the sum of light in each district in the years 1999 and 2012. The explanatory variable is market access computed based on Equation (14) and instrumented with the market access with constant light in (19) with  $\theta = 6$ . 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 are clustered at the state-level except in column 5.

Table A3: Effect of Actual Transport Infrastructure Investments with Trade Elasticity ( $\theta$ ) of 10

	(1)	(2)	(3)	(4)	(5)
Log Market Access	0.524*	0.551*	0.556*	0.577*	0.577**
	(0.291)	(0.308)	(0.310)	(0.328)	(0.251)
Excluded nodal districts	None	4	5	5	5
Weighting	Yes	Yes	Yes	No	No
Standard errors	Cluster	Cluster	Cluster	Cluster	Robust
N	626	613	612	612	612
Rsquared	0.509	0.502	0.502	0.495	0.495

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. The explanatory variable is market access computed based on Equation (14) and instrumented with the market access with constant light in (19) with  $\theta = 10$ . 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 are clustered at the state-level except in column 5.

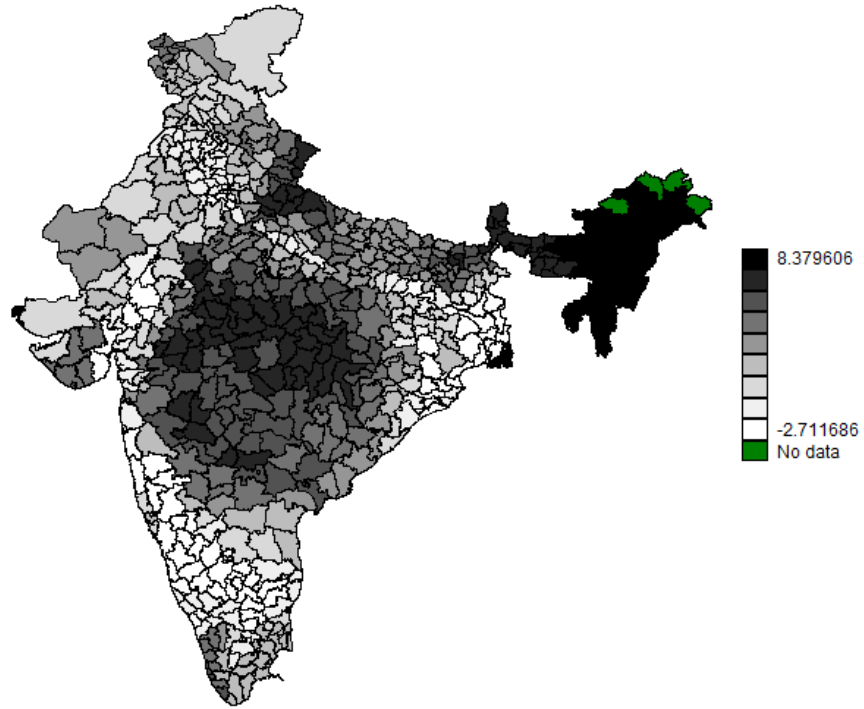
Figure A1: Approximation of the optimal network by sequentially adding links



The map shows a counterfactual highway network that equalizes marginal costs and benefits of highway construction by starting from the empty network and sequentially adding links.

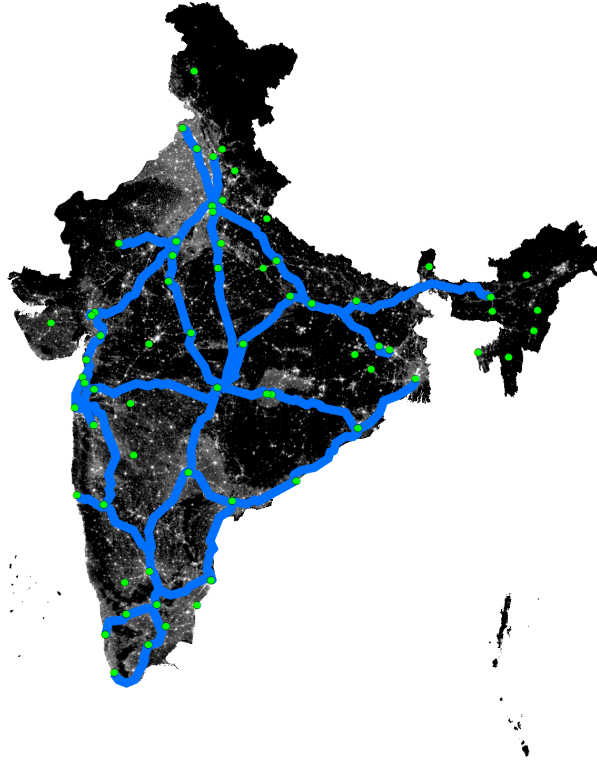


Figure A2: Percent increase in GDP from replacing GQ with least-cost network to connect all 68 cities



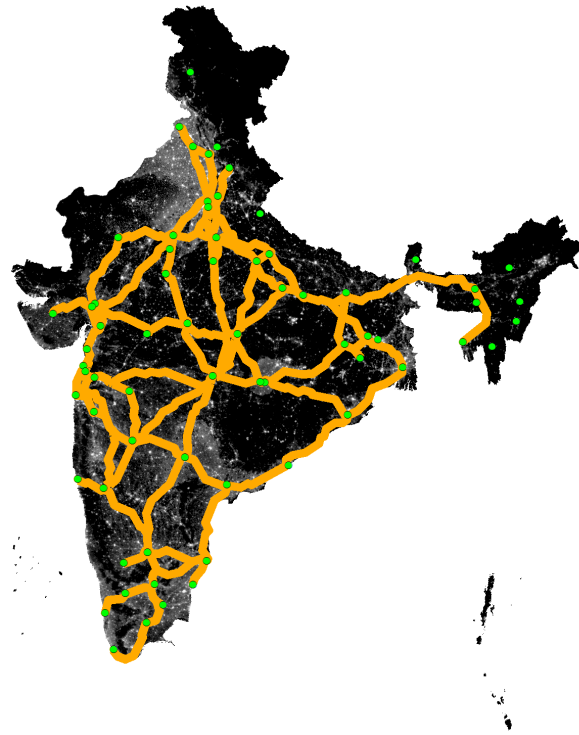
The map shows the boundaries of Indian districts. The colors represent percentage difference in GDP generated by replacing the GQ with the counterfactual network that connects targeted cities in a network with the least costs.

Figure A3: Approximating an optimal network of the same cost as minimum spanning tree in Figure 10



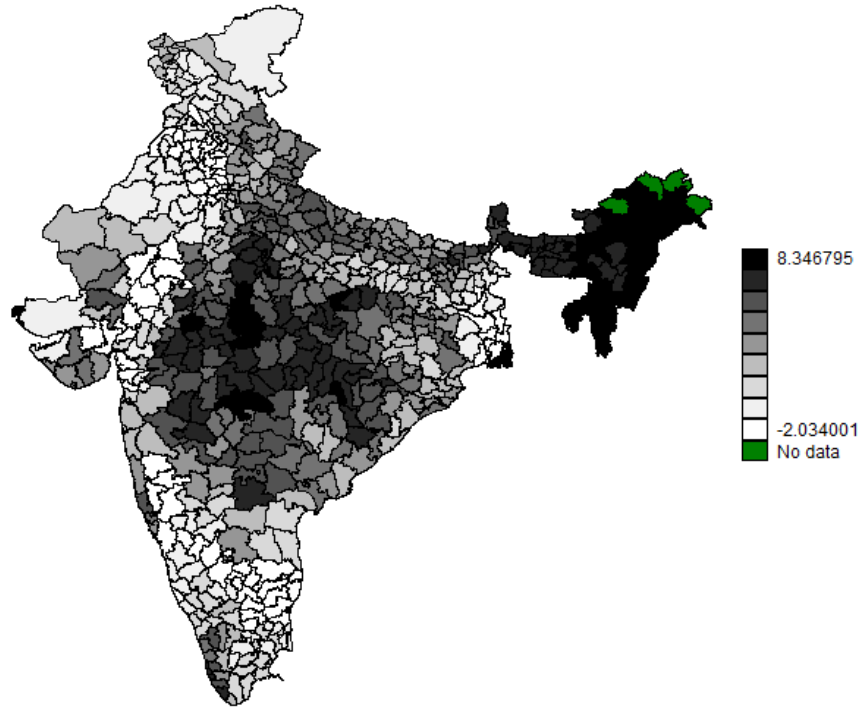
The map shows a counterfactual highway network that is constructed with the iterative procedure but imposing a budget constraint equal to the cost of the network in Figure 10. The green dots show the 68 largest cities and state capitals.

Figure A4: Approximating an optimal network of the same length as Figure 11



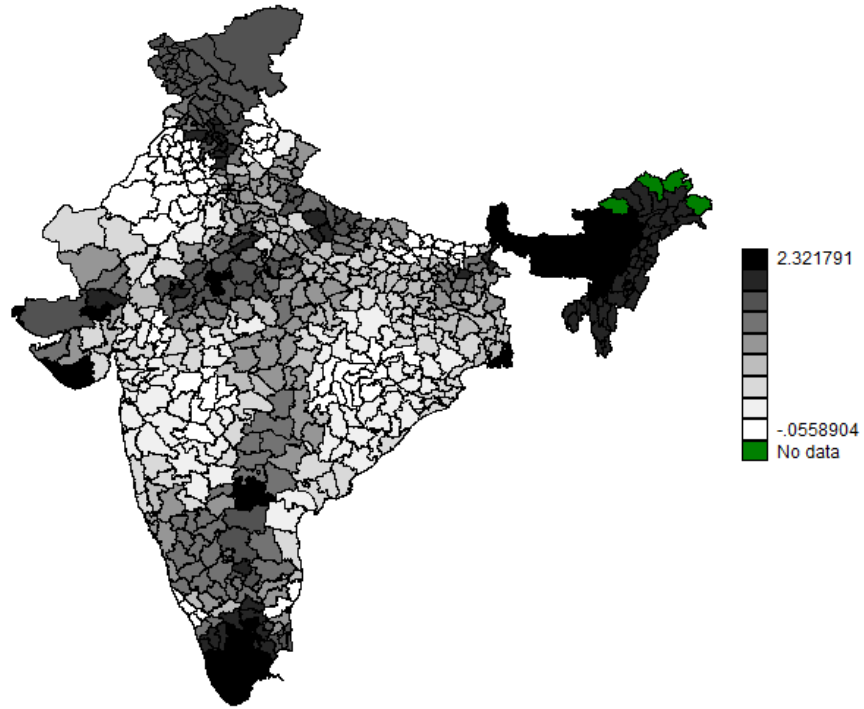
The map shows a counterfactual highway network that is constructed with the iterative procedure but imposing a budget constraint equal to the cost of the network in Figure 11. The green dots show the 68 largest cities and state capitals.

Figure A5: Percent increase in light from replacing GQ with counterfactual of rays and corridors



The map shows the boundaries of Indian districts. The colors represent percentage difference in GDP generated by replacing the GQ with the counterfactual network that is constructed by connecting cities with rays and corridors.

Figure A6: 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.