

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

Simon Alder[†]

University of Zurich

April 3, 2014

Abstract

This paper uses a general equilibrium framework as in Eaton and Kortum (2002) to estimate the contribution of transport infrastructure to regional development. I apply the analysis to India, a country with a notoriously weak and congested transportation infrastructure. I first analyze the development effects of a recent Indian highway project that improved connections between the four largest economic centers. I estimate the effect of this new infrastructure on income across districts using satellite data on night lights. The results show aggregate net gains from the Indian highway project, but very unequal effects across regions. China has followed a different highway construction strategy and has experienced more significant convergence across regions than India. I therefore use the model to gauge the effects of a counterfactual highway network for India that replicates the Chinese strategy of connecting intermediate-sized cities. I find that this counterfactual network would have significantly benefited the lagging regions of India, inducing more equal regional development. However, the effects on average national growth would have been modest. I also construct additional counterfactuals and discuss their effects on economic development.

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

Keywords: Transport Infrastructure, Economic Growth, Regional Development, Trade, India, China, Geographic Information System, Satellite Data, Luminosity.

*I would like to thank Fabrizio Zilibotti for his support during this project. I also thank Simeon D. Alder, George-Marios Angeletos, Marco Bassetto, Timo Boppert, Filippo Brutti, Marius Brühlhart, Adrien Bussy, Guido Cozzi, Gregory Crawford, Mariacristina De Nardi, David Dorn, 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, Rainer Winkelmann, Christoph Winter, Gabriel Zucman, Josef Zweimüller, Robert Zymek, and seminar participants at the University of Zurich, the University of Lausanne, and the Zurich Workshop on Economics for helpful comments. Sebastian Ottinger provided outstanding research assistance. Ronald Schmidt and Larry Crissman provided valuable support with GIS software and data.

[†]University of Zurich, Department of Economics, Schönberggasse 1, 8001 Zurich, Switzerland.
E-Mail: simonalder11@gmail.com.

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 emerge and there is a strong 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., forthcoming; 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 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. 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 first identify the effect of a major highway project in India, the construction of the GQ. Then, in light of the stark difference in the transport infrastructure strategies of India and China, I ask how India would have developed if it had built a network like the Chinese NEN. To this aim, I construct a counterfactual Indian highway network that mimics the Chinese approach of connecting intermediate-sized cities. The counterfactual is built based on the precise location of cities and on the topographic features of India, which are modeled using a geographic information system (GIS). The paths of the counterfactual highway connections are chosen to minimize the construction costs based on slope and land cover. The resulting road network then allows me to

compute bilateral transport costs between all 590 mainland districts in India using a shortest path algorithm. In the empirical analysis, the bilateral transport costs are related to income, which is measured using data on luminosity at night. This data is available from satellite images and captures human economic activity at a high spatial resolution and over time. Luminosity has been shown to be a good proxy for income growth (Henderson et al., 2012) and can be aggregated with digitized maps to the level of districts, for which official GDP data is not available.

The empirical analysis builds on general equilibrium trade theory. I follow Donaldson and Hornbeck (2013) who derive from a Ricardian trade model a reduced-form measure for the aggregate impact of transport infrastructure on income.¹ This reduced form captures the *market access* of a location by summing over the income of trading partners, discounted by the bilateral trade costs and by the destination's market access.² Transport infrastructure determines bilateral trade costs such that changes in the infrastructure over time generate variation in market access. More precisely, the bilateral transport costs can be computed for the transport network in 2000 (before the start of the recent Indian highway project), in 2009 (after completion of the first phases), and for the counterfactual (replicating the Chinese network). The resulting bilateral transport costs are then used to derive districts' market access for each version of the transport network.

The model predicts a log-linear relationship between income and market access. The time variation in market access between the actual networks in 2000 and 2009 allows me to estimate this relationship. Given the resulting estimate for the elasticity of income with respect to market access, I predict each district's income based on the market access implied by the counterfactual network. Importantly, market access captures the general equilibrium consequences of transport infrastructure and the resulting predictions therefore represent aggregate effects.³

¹Donaldson and Hornbeck (2013) estimate the effect of American railways on land value, while I estimate the effect of Indian highways on real income. Sections 4 and 5 will discuss the differences in more detail.

²A related measure is *market potential*, which has been derived from models in the new economic geography literature. See for example Redding and Venables (2004) and Hanson (2005).

³An increase in market access of the trading partner (e.g. because it is better connected to a third district) can reduce the market access of an origin. Market access therefore captures general equilibrium consequences such as trade diversion (see Section 5).

The empirical analysis makes three contributions. First, I quantify the aggregate effect of the realized GQ, India's major highway investment project between 2001 and 2009. The results suggest that aggregate income was 2.4 - 3.5 percent higher in 2009 than it would have been if the GQ had not been built. This implies a gain after one decade that is more than three times the construction costs. Second, I predict the aggregate effects of the counterfactual transport infrastructure, which replicates the salient aspects of the Chinese network in India in a way that minimizes road construction costs. Taking into account the construction costs of counterfactual roads, the results imply in the aggregate a modest difference relative to the existing infrastructure. 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 substantially from the counterfactual. The reason is that this network, by connecting all intermediate-sized cities, also reaches into regions that previously had low growth and were neglected by the GQ. Thus, a transport network that follows the Chinese strategy would increase growth particularly in India's lagging regions. This provides an explanation for the weaker convergence in India compared to China. Two alternative ways to replicate the Chinese network in India lead to qualitatively similar results.

The distributional consequences are particularly relevant in light of the unequal regional development of India. Policy makers are aware of this and the national highway development strategy did include plans for other highway connections besides the GQ. In particular, the government planned the North-South and East-West Corridors which cross through regions that were not reached by the GQ. However, these other projects were delayed and by 2009 only a small part has been finished. In an additional counterfactual exercise, I find that the completion of these corridors would indeed increase income in some of the lagging states. However, the explicit strategy of connecting all intermediate-sized cities would have larger aggregate effects and benefit more lagging districts.

The remainder of the paper is structured as follows. Section 2 reviews the related literature. Section 3 discusses the spatial development and transport infrastructure in India and China. Section 4 shows how the counterfactual network is constructed and what data is used. Section 5 presents the conceptual frame-

work and Section 6 discusses the results. Section 7 shows alternative ways to replicate the Chinese network and discusses the robustness of the results. Section 8 concludes.

2 Related Literature

The role of transport infrastructure for development has been the subject of a large literature.⁴ A recent increase in this literature was triggered by a combination of economic theory with geographic information such as the exact location of transport infrastructure. My methodology for evaluating the impact of infrastructure builds on Donaldson and Hornbeck (2013) who estimate the aggregate effect of the expansion of the American railway network in the 19th century. They derive market access as a reduced form measure for the impact of transport infrastructure in a general equilibrium trade model as in Eaton and Kortum (2002). Donaldson and Hornbeck (2013) also compare the effect of the actually built network to counterfactual scenarios in which railways are replaced by roads and canals.⁵ I adapt their framework to the use of light data as a measure of real income in India. To determine the precise paths of the counterfactual roads, I use the least-cost network that connects a given set of cities. Such a network has previously been used by Faber (2013) within China in order to construct an instrument for the actually built highways. I follow the approach of connecting cities which fulfill the criteria of the NEN in a way that minimizes road construction costs, but I apply it to Indian cities and to the local terrain in order to construct a counterfactual network.

The empirical exercise of this paper is also related to recent studies on the local effects of transport infrastructure. Datta (2012) and Ghani et al. (2012) study the effects of the GQ and find a positive impact on firms located in the proximity of the new highways.⁶ An important aspect of these studies is the identification of exogenous sources of variation in transport infrastructure. They rely on an

⁴See for example Redding (2010) and a recent survey by Breinlich et al. (2013).

⁵Such a counterfactual exercise was also proposed in the seminal work by Fogel (1964).

⁶Both studies use firm surveys to evaluate the effect of the GQ. Ghani et al. (2012) point out that it would be valuable to estimate the effect with luminosity data.

identification strategy similar to the one proposed by Chandra and Thompson (2000) and Michaels (2008) who estimate the effect of US highways on counties that lie between two important nodal cities. This is based on the observation that the highways were built to connect larger cities and thereby passed through other counties which consequently obtained access to the new transport infrastructure without being targeted. The results of Datta (2012) and Ghani et al. (2012) suggest positive effects on firms located close to the GQ. I use this strategy in order to estimate the effect of the GQ on non-nodal districts, excluding the four cities that were targeted by the GQ. Several related studies have analyzed the Chinese transport network (see Banerjee et al., 2012; Baum-Snow et al., 2013; and Faber, 2013).⁷

The above studies focus on identifying the local effects in the proximity of new roads. Two recent contributions that estimate the general equilibrium consequences of the national transport infrastructure in India and China are Donaldson (forthcoming) and Roberts et al. (2012). Donaldson (forthcoming) estimates the effect of railways in colonial India and finds that historical income levels of Indian districts had increased by 16 percent when they were connected to the railway network. He also shows that there is a sufficient statistic for the general equilibrium effect of transport infrastructure on income, which explains most of the variation due to transport infrastructure. Roberts et al. (2012) use a structural new economic geography model to measure the aggregate effect of the expansion of the NEN in China and find that aggregate income was 6 percent higher in 2007 due to the NEN. My analysis differs from the above studies by estimating the aggregate effect of transport infrastructure through market access, as was proposed by Donaldson and Hornbeck (2013), and using this estimate to predict income under various counterfactual infrastructures. To identify exogenous variation in market access, I apply the identification strategy proposed by Chandra and Thompson (2000) and Michaels (2008).

The market access approach used in this paper is closely related to models in the new economic geography literature. Several authors analyze the role of market access (or market potential), which can be affected by transport costs (Puga,

⁷Transport infrastructure in other countries has recently been studied by Atack (2008), Baum-Snow (2007), Gollin and Rogerson (2010), Herrendorf et al. (2012), and Storeygard (2013).

2002; Redding and Venables, 2004; Hanson, 2005; Redding and Sturm, 2007; 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 (Anderson and van Wincoop, 2003; Allen and Arkolakis, 2013, Atkin and Donaldson, 2013; Coşar and Fajgelbaum, 2013; Redding, 2012). 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. A contribution of this paper is that the digital transport network 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. While these models are static, Desmet and Rossi-Hansberg (forthcoming) 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 also relates to cost-benefit analyses of individual infrastructure investments. 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.

3 Spatial Development and Transport Infrastructure

This paper makes a link between regional growth and transport infrastructure. India and China provide an interesting context to study this relationship. While both countries are growing fast, they also show substantial differences in their regional development patterns and in their transport infrastructure. This section first reviews the evidence on the spatial distribution of income and growth in the

two countries and then discusses their transport infrastructure.

3.1 The Spatial Development of India and China

During the past two decades, real GDP per capita in India has been growing at an average rate of 4.8 percent (World Bank, 2013). China's growth, averaging at 9.2 percent, has been even more spectacular and its income per capita overtook India's in the early 1990s. Although there is substantial variation in the regional growth rates within both countries, previous studies found that China has overall seen more convergence (Desmet et al., forthcoming; Chaudhuri and Ravallion, 2006). The same finding emerges when using light as a measure of income.⁸ Figure 1 shows the spatial distribution of light in the year 2000. Not surprisingly, there is a strong clustering of income in both countries. A similar picture arises when aggregating the light pixels to the sub-national units of Chinese prefectures and Indian districts.

Interestingly, there are substantial differences in the spatial development over time. To illustrate this, Figures 2 and 3 show the spatial distribution of initial density and growth in the two countries. While China has seen the highest growth rates in prefectures with initially low density (measured as average light intensity per pixel in 2000), this has not been the case for Indian districts. In particular, the right panel of Figure 3 suggest that the districts with initially low or intermediate light density had experienced surprisingly low growth. This is consistent with the evidence presented in Desmet et al. (forthcoming) and Chaudhuri and Ravallion (2006) who found stronger convergence patterns in China than in India.⁹

While the finding that regional growth patterns differ between the two countries is well known, we still lack precise answers for what is driving these differences. Transport infrastructure is a potential candidate since it is an important determinant of the spatial distribution of economic activity. India and China indeed have followed different strategies for how to invest in their transport networks

⁸Section 4 discusses the light data in more detail. Henderson et al. (2012) show that light correlates strongly with GDP in a panel of 188 countries.

⁹The observation is confirmed when regressing light growth of each prefecture or district on its initial density. The slope coefficient is significantly smaller in China than in India, suggesting stronger convergence.

and these differences will be outlined next.

3.2 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). In this section, I first review the key elements of the infrastructure investments in the two countries and then discuss the construction of a counterfactual network for India which mimics the Chinese strategy.

In the early 1990s, the Indian road infrastructure was superior to the Chinese in terms of total km length and km 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.¹⁰ Over the 1990s, China's highway and railway network developed significantly faster than the Indian counterpart. In particular, it built the National Expressway Network (NEN) with the explicit objective of connecting all cities with more than 500,000 people and all provincial capitals in a modern highway system.¹¹ 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, shown in red in Figure 4, had reached a length of 40,000 km by 2007 and it continued to be expanded. It consists of four-lane limited access highways that allowed significantly higher driving speed than the existing roads.¹²

India also invested in its road infrastructure, but about ten times less than

¹⁰The 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).

¹¹This 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 (2013) for a discussion.

¹²A description of the history of the Chinese highway network and its different components is provided by ACASIAN. See www.acasian.com for further details.

China and with a focus on the main economic centers. In particular, it launched a National Highways Development Project (NHDP) in 2001 and the first achievement of that project was the GQ, which connects the four major economic centers with four-lane highways (shown in Figure 4). 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 a cost of USD 6 billion (1999 prices).¹³ 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, these projects were delayed and not fully completed by 2010. Figure 5 shows the parts which were completed by 2010.

The GQ in India, like the NEN in China, has significantly reduced the transport times between places with access to 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), while the driving speed on the GQ is around 75 km/h.¹⁴ However, there is ample evidence that, even today, insufficient transport infrastructure is a severe constraint for the Indian economy. Raghuram Rajan, the current Governor of the Reserve Bank of India, recently stated that India needs to improve its infrastructure with the same discipline 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, 2011; Ernst and Young, 2013; KPMG, 2013).

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

¹³Most parts were already completed by 2007. See the webpage of the National Highway Authority of India (<http://www.nhai.org/index.asp>) for details. The cost estimates are based on Ghani et al. (2012).

¹⁴The 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.

expenditures on railways in total transport infrastructure declined from 50% in the 1990s to 30% by the end of the 2000s (Ministry of Railways, 2012). Today, roads are by a large margin the most important transport mode in India, carrying 60% of the freight turnover compared to 31% for railways.¹⁵ The highway projects undertaken in the two countries are therefore the crucial parts of their transport strategies and of high importance for the development of the two countries.

Although the analysis undertaken here captures a key aspect of the modern transport infrastructure in India and China, some caveats must be pointed out that concern possible changes in other types of infrastructure. But I will show that these concerns are mitigated by my empirical strategy which exploits exogenous variation in transport infrastructure and controls for location and time fixed effects. The first concern is the omission of other types of domestic transport infrastructure such as railways or urban systems such as subways. Second, access to international markets via sea ports or airports is not modeled as part of the transport network here. Third, villages' access to the transport infrastructure via rural roads is not considered due to a lack of precise data.¹⁶ Finally, non-transport infrastructure such as electricity and water also affect economic development. However, these caveats would limit the validity of the exercise here only if the omitted factors were time-varying at the district level and correlated with the explanatory variable market access. Section 5 discusses in detail how I address this with a suitable empirical strategy.

The above discussion, and the illustration in Figure 4, makes clear that India and China have followed different strategies to improve their road infrastructure. While India's highway investments focused on connecting its largest economic centers with better highways, China has built a network that connects all cities that have a registered population of more than 500,000 and all provincial capitals. Furthermore, it is clear that highways play a key role in the overall transport in-

¹⁵The share of highways in the total freight turnover is even higher in India than in China (KPMG, 2013).

¹⁶According to Harral et al. (2006), India has prior to the start of the NHDP in 2001 focused its infrastructure investments on the improvement of roads which provide access to highways, while China has from the start of its program in 1992 put the emphasis on investments in arterial highways to connect cities. While I cannot observe the upgrades of local roads prior to my sample period, the NHDP was a large infrastructure program to which the Chinese NEN can be directly compared.

frastructure. India currently faces severe constraints due to insufficient transport infrastructure, which is less the case for China. 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 by applying the policy objective of the Chinese government to identify the Indian cities which would be connected with the Chinese network. The exact routes are chosen such that the costs of building the roads, which depend on characteristics of the Indian terrain, are minimized. The next section will present the data that is required to build such a counterfactual network and to evaluate its effect on economic development.

4 Counterfactual Roads

In this section I build a counterfactual network of highways and compute bilateral transport costs. First, I use topographic features of India in order to determine the precise paths of the counterfactual roads such that road construction costs are minimized. Based on the geographically explicit road network, I then calculate the bilateral transport costs between all Indian districts using a shortest path algorithm. The data and methods required for this procedure are described next.

4.1 Data

In order to analyze the effect of transport infrastructure on economic development, I construct a dataset that incorporates the precise geographic location of different types of roads and of local income. I add to this data set the boundaries of Indian states and districts and multiple maps of the Indian terrain.

Transport Infrastructure and Terrain

I use geographic information system (GIS) methods to process the spatial data.¹⁷ Digital maps with the location of the actual Indian transport infrastructure are

¹⁷The applications used here are ArcGIS Desktop 10.1 and the `spmat` functions in Stata 13.

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 Golden Quadrilateral and the completed parts of the North-South and East-West Corridors. The average driving speed on existing roads are taken from several transport efficiency studies. World Bank (2005) reports that the typical driving speed on the existing Indian national and state highways is between 30 and 40 km/h and I therefore assume a speed of 35 km/h for all highways built before the start of the NHDP.¹⁸ Roads of lower or unknown quality are also controlled for and the assumed travel speed is 25 km/h, which is suggested by surveys of rural Indian transport infrastructure (Liu, 2000). 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 Chinese expressways and the GQ, which according to google maps is 75 km/h. 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.

Political Boundaries and Luminosity

The units of analysis in this paper are Indian districts. I focus on mainland districts, of which there are 590. Luminosity measured by weather satellites has

¹⁸These estimates are in line with more recent numbers by KPMG (2013).

been shown to be a good proxy for income (Henderson et al., 2012). Two important advantages of the light data are that it has a high spatial resolution and is independent of countries' statistical capacity. It is particularly useful when official GDP figures are not available, for example for subnational administrative units such as Indian districts.

The digitized district boundaries are provided by Global Administrative Areas (2012) and the light data is available from the Earth Observation Group (2013) 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.¹⁹ The raster are 30 arc second grids, spanning -180 to 180 degrees longitude and -65 to 75 degrees latitude. To derive a measure of real income I for each district, I aggregate light within district boundaries using an equal area projection. The light summary statistics of the sample of mainland Indian districts are presented in Table 1.

In order to interpret the light data in terms of GDP, one first has to analyse the relationship between GDP and light. GDP data is not available for Indian districts such that the relationship has to be inferred from other samples. GDP at the level of sub-national units such as districts is also in other countries difficult to obtain. An exception is China where one can rely on GDP data at the prefecture level based on official statistical yearbooks.²⁰ This data allows estimating the relationship with light at a similar level of aggregation as Indian districts. However, there are some caveats when using this approach. First, light intensity is measured by various satellites over the years. These satellites can have somewhat different calibrations such that observations from different satellite years cannot directly be compared. Second, GDP data is measured within administrative boundaries and these may also change over time.²¹ I therefore rely only on the year 2010 for which I have GDP figures for Chinese prefectures and the digital maps of admin-

¹⁹The last two years have recently been made available and they have not been included in the present analysis.

²⁰See Alder et al. (2013) for a more detailed description of the GDP data in the Chinese statistical yearbooks.

²¹The Chinese statistical yearbooks also report the land area. These figures show that subnational borders have indeed been changing.

istrative boundaries that match the units in the statistical yearbooks. This allows me to measure light and GDP on the same land area and with the same satellite, thereby addressing the two concerns above. I then regress the logarithm of GDP on the logarithm of light within Chinese prefectures in order to estimate the relationship. The estimated elasticity is 1.05 and significant, but indistinguishable from 1. In the empirical analysis, I therefore interpret the magnitude of an effect on light as an equally large effects in terms of GDP.²² While the elasticity that is used to translate the estimate from light to GDP clearly affects the magnitude of the aggregate effects, the distributional implications are not affected by this choice.

4.2 Building a Counterfactual Highway Network for India

To replicate the Chinese network in India, I first identify the cities in India which would have been chosen by the Chinese policy and then build a counterfactual network to connect them through the Indian terrain in a way that minimizes construction costs.²³ 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 6.

In order to determine the network which connects all targeted cities in a least-cost manner, one first needs to obtain a measure for road construction costs on the Indian terrain. I follow Faber (2013) 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:

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

Slope is measured in percent and *Builtup* and *Water* are binary indicators which

²²Note that to estimate the relationship between GDP and light I do not exploit time variation. An alternative, which has also been used in Henderson et al. (2012), is to estimate the relationship in a panel or in long differences. In the period from 2000 to 2009, the result for the later suggests an elasticity of roughly 0.5. However, it is subject to the two concerns raised above and in this context I prefer to rely on the estimate based on the 2010 cross section.

²³This approach has previously been applied by Faber (2013) in China to construct an instrument for the actually built expressways. In the steps below, I adapt this approach to replicate the Chinese network in India.

take the unit value if the majority of the cell is built up or water, respectively.²⁴ Applying this formula using detailed terrain data produces a fine 1×1 km grid of construction costs for the entire Indian landscape. Given this grid of construction costs, one can in a second step apply the Dijkstra algorithm to find the cheapest connection between any two given points through the cost grid.²⁵ The procedure is illustrated in Figure 7, where the cells represent different construction costs (based on Equation 1) and the lines are the least-cost paths to connect the cities (shown as circles). The third step in order to obtain the counterfactual network is to find the cheapest possible way to connect all targeted cities to the network.²⁶ This is achieved by 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. This produces the least-cost network. Once it is determined which bilateral connections must be made, the counterfactual highways can be drawn with GIS software following the least-cost path computed above (illustrated by lines in Figure 7). The resulting counterfactual network is shown in Figure 9. It represents the cheapest way to formally fulfill the Chinese policy objective (connecting cities which have a population above 500'000 or are state capitals) in India.²⁷

Despite its immediate link to the official policy objective, this minimalistic network is not one that would typically be implemented by governments. One reason is that planners would most likely complement it with additional connections between "loose ends" created by the algorithm. For example, if two cities are indirectly connected through other cities, then connecting them is redundant from the perspective of the policy objective, even if the cities are close to each

²⁴The 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 (2013), my formulation does not include wetlands.

²⁵This algorithm is implemented in the ArcGIS Network Analyst extension. The same algorithm can be used to compute the least-cost transport route (instead of least cost construction path). The algorithm has already been widely used in the economics literature, for example in Dell (2012), Faber (2013), Donaldson and Hornbeck (2013), and Donaldson (forthcoming).

²⁶Note that the previous step computed the least-cost construction path between all bilateral pairs of cities. Most of these paths are redundant because a given city may already be connected to the network through another city.

²⁷More precisely, it is the cheapest way to connect a given set of nodes with bilateral links. The procedure does not allow endogenous nodes or hubs.

other. However, in reality, the additional link may possibly be effective for reducing transport times. This illustrates that the least-cost network is not minimizing transport costs nor maximizing aggregate income. However, the least-cost network is a useful benchmark for the counterfactual analysis because it improves transport infrastructure for a particular set of cities which would have been targeted by the NEN. Importantly, it is an objective way to replicate the Chinese network because the least-cost network is unique. The counterfactual network therefore allows an interesting comparison between a network that focused on the four largest economic centers (the Indian GQ) and one that connects all intermediate sized cities (the Chinese NEN). In the robustness section, I discuss the results for alternative counterfactual networks in India. In particular, I propose an alternative network exploiting that the Chinese government also specified that the targeted cities should be connected with rays out of the capital city and with horizontal and vertical corridors. These alternative networks resemble more the structure of the network which was actually built in China. The disadvantage is that they are not unique as there would be several ways to make the connections. In the main part of this paper, I will therefore focus on the least-cost network and discuss the alternatives in the robustness section.

4.3 Computing Transport Costs Through a 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.

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 can be derived from automated searches on applications like google maps, this is not the case for past or counterfactual networks. My approach is to use the Dijkstra algorithm that finds the shortest path (in terms of transport time) between

any two locations on a 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 (described in Section 4.1) are the geographically referenced roads and the transport speed on different types of roads. With these inputs, it is possible to construct a grid of India where the value of each 10×10 km cell represents the costs of traveling through this cell. These travel costs depend on the quality of the road inside of each cell, i.e. the travel costs are high if there are only roads of poor quality with low travel speeds. Such a grid of transport costs is shown in Figure 8. The Dijkstra algorithm then calculates the cheapest way to travel from one location (district centroids, represented by dots in Figure 8) to another location. Depending on the road infrastructure and thus on the transport costs in each cell, the cheapest path may not be the shortest in terms of distance. More importantly, the transport times associated with the cheapest path change when the infrastructure is improved, thus generating time variation in the transport costs. Following Roberts et al. (2012), I assume that there are economies of scale in transport, such that transport costs increase less than proportionally in transport times.²⁸ More precisely, I calculate transport costs between an origin o and a destination d as

$$TransportCosts_{od} = 1 + TransportTime_{od}^{0.6}, \quad (2)$$

where the exponent of 0.6 is an average value that Roberts et al. (2012) derived for the Chinese network. A particular case is when $o = d$, i.e. transport costs within a district. Although a district is represented here by its centroid and the iceberg assumption in such a case implies transport costs of 1, this would not be accurate for actual Indian district which differ substantially in size. One solution which has been used in the literature is to normalize it to the observation with the smallest land area (Au and Henderson, 2006). I use the distance between the district centroid and the nearest district border as a measure for within-district costs and normalize all travel costs to the smallest distance in the sample.

By applying the Dijkstra algorithm to all versions of the transport networks (past, current, and counterfactual network), one can derive the bilateral trade

²⁸This is a common assumption, see for example also Au and Henderson (2006) who assume that transport costs increase less than proportionally in *distance*.

costs of any pair of districts for each version. In the empirical analysis, this variation in trade costs over time is related to growth in income. The channel through which this relationship works is shown by the conceptual framework in the next section, which serves as guide for the empirical analysis.

5 Conceptual Framework

This section first presents a conceptual framework that illustrates how transport infrastructure affects economic development.²⁹ Then I discuss how the light data and the transport networks presented in Sections 3 and 4 are used to estimate this effect. The setup is a general equilibrium trade model as in Eaton and Kortum (2002). Donaldson and Hornbeck (2013) derive from a variation of that model a reduced form expression for the impact of transport infrastructure on income. That expression captures the “market access” of a location, which is the sum over trading partners’ income, discounted by the bilateral trade costs and by the market access of the trading partners. They use this framework to estimate the effect of the expansion of the American railway network on land prices. I 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.

5.1 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.³⁰ 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

²⁹The presentation in this section focuses on the key aspects of the model. The details are discussed in the appendix.

³⁰Donaldson and Hornbeck (2013) assume that labor is mobile. The motivation for and implication of the assumption that labor is immobile will be discussed below. The two immobile production factors cannot actually be separated, but I distinguish them in order to allow a comparison to the model with mobile labor as in Donaldson and Hornbeck (2013).

Cobb-Douglas technology using land (L), labor (H), and capital (K),

$$x_o(j) = z_o(j)(L_o(j))^\alpha (H_o(j))^\gamma (K_o(j))^{1-\alpha-\gamma}, \quad (3)$$

where $z_o(j)$ is an exogenous probabilistic productivity shifter as in Eaton and Kortum (2002).³¹ The production function implies marginal costs

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

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

$$p_{od} = \tau_{od}MC_o(j) = \tau_{od} \frac{q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma}}{z_o(j)} \quad (5)$$

$$z_o(j) = \tau_{od} \frac{q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma}}{p_{od}}. \quad (6)$$

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:³²

$$P_d = \gamma \left(\sum_o [T_o (\tau_{od} q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma})^{-\theta}] \right)^{-\frac{1}{\theta}}. \quad (7)$$

I follow the notation in Donaldson and Hornbeck (2013) and define the sum over origins’ factor costs as “consumer market access”, because it measures district d ’s

³¹Each 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).

³²Using the fact that the rental rate for capital is equalized everywhere to $r_o = r$, we can define the constant $\kappa \equiv \gamma r^{-\frac{1}{1-\sigma}}$ where $\gamma = \left[\Gamma \left(\frac{\theta+1-\sigma}{\theta} \right) \right]^{\frac{1}{1-\sigma}}$ and Γ is the gamma function.

access to cheap goods,

$$\begin{aligned} P_d^{-\theta} &= \kappa \sum_o [T_o (\tau_{od} q_o^\alpha w_o^\gamma)^{-\theta}] \\ &\equiv CMA_d. \end{aligned} \quad (8)$$

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

5.2 Trade Flows and Gravity

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

$$\frac{X_{od}}{X_d} = \frac{T_o (q_o^\alpha w_o^\gamma)^{-\theta} \tau_{od}^{-\theta}}{\sum_o [T_o (q_o^\alpha w_o^\gamma)^{-\theta} \tau_{od}^{-\theta}]} \quad (9)$$

Assuming that a district's expenditure equals income ($X_d = Y_d$),³³ this can be rearranged to

$$\begin{aligned} X_{od} &= \underbrace{T_o (q_o^\alpha w_o^\gamma)^{-\theta}}_{\text{Origin's productivity and factor costs}} \times \underbrace{\tau_{od}^{-\theta}}_{\text{Trade costs}} \times \underbrace{Y_d}_{\text{Destination's income}} \\ &\times \underbrace{\kappa CMA_d^{-1}}_{\text{Destination's CMA}}. \end{aligned} \quad (10)$$

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, but negatively on the consumer market access of the destination and on the bilateral trade costs. Summing the gravity equation over destinations d yields total income of origin o ,

$$Y_o = \sum_d X_{od} = \kappa T_o (q_o^\alpha w_o^\gamma)^{-\theta} \sum_d [\tau_{od}^{-\theta} CMA_d^{-1} Y_d], \quad (11)$$

³³Capital is the only mobile production factor such that output and expenditure may differ within districts. Since I cannot distinguish the two in the data, I assume that capital income is spent where it is used for production which implies that expenditure equals income.

where Donaldson and Hornbeck (2013) define “firm market access” of district o as

$$FMA_o \equiv \sum_d \tau_{od}^{-\theta} CMA_d^{-1} Y_d. \quad (12)$$

If trade costs are symmetric, then a solution must satisfy $CMA_o = FMA_o$. Donaldson and Hornbeck (2013) refer to this term as “market access” (MA),

$$MA_o = \sum_d \tau_{od}^{-\theta} MA_d^{-1} Y_d. \quad (13)$$

The equation for income (11) then becomes

$$Y_o = \kappa T_o (q_o^\alpha w_o^\gamma)^{-\theta} MA_o. \quad (14)$$

Equations (13) and (14) summarize how trade costs affect income. While Equation (14) provides a relationship between income and market access, Equation (13) shows that this market access measure is the channel through which transport costs affect income. This framework resembles a gravity equation in the sense that locations which are better connected influence each other more and the influence is increasing in the size of their markets. This feature is shared by models in the new economic geography literature (see Fujita et al., 1999) and has found strong support in the data. A second appealing property of the model is that it is a general equilibrium setup and thus allows to quantify aggregate effects. In particular, the market access approach takes into account that a reduction in bilateral trade costs τ_{di} between two trading partners d and i can affect market access in o . This can be seen in Equation 13, where an increase in MA_d (due to a decrease in τ_{di}) reduces MA_o .

5.3 Empirical Implementation

The framework summarized by Equations (13) and (14) suggests a relationship between transport costs and income that can be estimated with the appropriate data. To this aim, I rewrite the model in terms of real income (which can be measured by light in each district) and use an appropriate identification strategy

to estimate the causal effect of market access on income.

Market Access

In order to incorporate luminosity as a measure for real income, I use the property in Eaton and Kortum (2002) that the price index is related to market access in the form

$$P_d = MA_d^{-\frac{1}{\theta}}. \quad (15)$$

This allows to rewrite Equation (13) as

$$MA_o = \sum_d \tau_{od}^{-\theta} MA_d^{-\frac{(1+\theta)}{\theta}} Y_d^r, \quad (16)$$

where $Y_d^r = \frac{Y_o}{P_o}$ denotes real income. For given income (measured by light), bilateral transport costs (computed from the transport network), and trade elasticity θ (taken from Donaldson, forthcoming), I solve this system of non-linear equations and obtain each districts market access measure.³⁴

Real Income

Equation (14) can be written in terms of real income. The wage and land rental rates can be substituted with the factor income shares to obtain

$$Y_d^r = (\kappa T_o)^{\frac{1}{1+\theta\alpha+\theta\gamma}} \left(\frac{\alpha}{L_o}\right)^{\frac{-\theta\alpha}{1+\theta\alpha+\theta\gamma}} \left(\frac{\gamma}{H_o}\right)^{\frac{-\theta\gamma}{1+\theta\alpha+\theta\gamma}} (MA_o)^{\frac{1+\theta(1+\alpha+\gamma)}{(1+\theta\alpha+\theta\gamma)\theta}}. \quad (17)$$

Estimating this equation in a cross-section would require to control for relevant district characteristics which are difficult to obtain. It is therefore useful to exploit the panel structure to identify the causal relationship. Therefore, the above equation will be estimated with a fixed effect panel regression that relies on the time

³⁴Donaldson and Hornbeck (2013) estimate the effect of market access on land values. They measure market access with population and their benchmark estimation uses a first order approximation $MA_o \approx \sum_d \tau_{od}^{-\theta} N^d$, where N denotes population. They compare their results to those based on the numerical solution for MA implied by the model and find similar effects (Column 2 in their Table 1). I focus on the second approach and use the market access measures based on the numerical solution of Equation (16).

variation within districts. This allows to account for the unobserved heterogeneity across districts. Equation (18) shows the analogue of Equation (17) in logs and over time.

$$\begin{aligned}
\ln(Y_{o,t}^r) &= \underbrace{-\frac{\theta\alpha}{1+\theta\alpha+\theta\gamma}\ln\left(\frac{\alpha}{L_o}\right) - \frac{\theta\gamma}{1+\theta\alpha+\theta\gamma}\ln\left(\frac{\gamma}{H_o}\right)}_{\text{Constant over time}} \\
&+ \underbrace{\frac{1}{1+\theta\alpha+\theta\gamma}\ln(\kappa_t)}_{\text{Country characteristics}} + \underbrace{\frac{1}{1+\theta\alpha+\theta\gamma}\ln(T_{s,t})}_{\text{Productivity}} \\
&+ \underbrace{\frac{1+\theta(1+\alpha+\gamma)}{(1+\theta\alpha+\theta\gamma)\theta}\ln(MA_{o,t})}_{\text{Market access}}. \tag{18}
\end{aligned}$$

Time variation in real income is measured by light in the years 2000 (the year before the NHDP started) and 2009. Variation in the market access measures, obtained from solving Equation (16), can either be due to changes in the transport infrastructure (differences in the bilateral transport costs τ_{od}) or in income. The corresponding panel fixed effects specification is

$$\ln(\text{light}_{o,s,t}) = \phi_o + \delta_{s,t} + \beta \ln(MA_{o,t}) + \varepsilon_{o,s,t}, \tag{19}$$

where ϕ_o is a location fixed effect and $\delta_{s,t}$ is a state-year fixed effect. The link to Equation (18) is as follows. The first line in Equation (18) collects parameters and factor endowments, which are assumed to be constant over time and are thus absorbed by the location fixed effects ϕ_o . The second line includes country characteristics (the interest rate inside of κ) and productivity of each district. While changes in the national interest rate are absorbed by the state-year fixed effects, local productivity can potentially vary over time *and* districts. As will be argued below, my identification strategy uses exogenous variation in transport infrastructure such that there is no effect of unobserved productivity changes on transport infrastructure. Furthermore, part of this variation is absorbed by the state-year fixed effects. The last line in Equation (18) shows the effect of market access, which is computed in Equation (16) for different transport infrastructures.

A caveat of the specification in Equation (19) is that the state-time fixed ef-

fects may absorb differences in the growth rates across states that may be driven by some states being more exposed to infrastructure investments. As an alternative specification, I will also report the results for the case when the state-year fixed effects are replaced by predetermined state-characteristics such as state-level growth trends prior to the start of the NHDP.

The model predicts an elasticity of real income with respect to market access of

$$\beta = \frac{1 + \theta(1 + \alpha + \gamma)}{(1 + \theta\alpha + \theta\gamma)\theta}. \quad (20)$$

While I identify only β and cannot estimate the individual parameters separately, one can verify whether the resulting estimate for β 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 3.8 for the trade elasticity would imply an elasticity of income with respect to market access of 0.54. This value is not rejected by the estimates in Section 6 below.

Model and Data

The data required for the estimation of Equation (19) are income of each district (obtained from light data) and bilateral trade costs (computed based on a shortest path algorithm). I consider the period from 2000 to 2009 in order to capture the infrastructure investments of the NHDP, which started in 2001. The GQ and NSW were built in different phases during which fragments of new or upgraded highways were added to the network. I abstract from the yearly variation and only consider the total constructions between 2000 and 2009. Using data for the years 2000 and 2009, Equation (19) can be estimated in first differences. For the computation of the market access measures in Equation (16), a value for the trade elasticity θ is assumed. Ideally this would be estimated with bilateral trade data between Indian districts. While I am not aware of an estimate for the trade elasticity for contemporaneous India, Donaldson (forthcoming) has estimated θ using bilateral trade data between Indian districts for his study on the colonial railway system and I use his estimate of 3.8. This estimate is also in line with recent results by Simonovska and Waugh (2013) using international trade data.

Identification

Identifying the causal effect of infrastructure on income is challenging for several reasons. First, the choice of where to build infrastructure is not exogenous. In particular, the GQ had the explicit goal of connecting the four largest economic centers. This raises the concern that infrastructure may have been built where high growth could be expected. But the clear objective of the GQ also poses an advantage for identification. By connecting the four largest centers with fairly straight lines, it affected districts which happened to be in between two important cities. By excluding the nodes of a 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), and Ghani et al. (2012). I follow this strategy and exclude the nodal cities and the corresponding states Maharashtra, Delhi, West Bengal, and Tamil Nadu.

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 hold income fixed in 2000 when computing market access and exploit only the variation due to changes in transport infrastructure (and thus bilateral trade costs), as is shown in Equation (21),

$$MA_{o,t} = \sum_d \tau_{od,t}^{-\theta} MA_{d,t}^{\frac{-(1+\theta)}{\theta}} Y_{d,2000}^r. \quad (21)$$

According to the model, the elasticity of income with respect to market access, β , is constant over the counterfactual. Therefore, given an identification strategy that allows to estimate this β , income can be predicted for counterfactual transport networks that imply different market access values. The results of these two steps are discussed in the next section.

6 Results

The results of the empirical analysis are presented in three steps. First, I estimate the elasticity of income with respect to market access based on Equation 18 of the model. I exploit time variation in market access due to actually built infrastructure in India. The market access measures are computed from Equation (21) and they capture general equilibrium effects. Second, I will use the estimated elasticity to predict income levels in 2009 for various (actual and counterfactual) transport networks and assess their aggregate effects. The distributional consequences of the different transport networks are discussed in the third step.

6.1 Elasticity of Income with Respect to Market Access

The estimate of β in Equation (19) represents the elasticity of income with respect to market access, where the time variation in market access is due to the construction of the GQ and parts of the NS-EW between 2000 and 2009. The results are shown in Table 2. As a benchmark, column 1 uses the full sample of mainland Indian districts and regresses the logarithm of income on the logarithm of market access, controlling for district fixed effects and state-year fixed effects. The estimated coefficient implies that a one percent increase in market access is associated with a 0.39 percent increase in income. For the subsequent counterfactual exercise, it is crucial to estimate the causal effect of market access on income. In particular, one has to address concerns related to omitted variables and reversed causality. I first discuss how I address omitted variables in the context of column 1 of Table 2 and then turn to the problem of reverse causality (addressed in column 2).

The problem of omitted variables can be mitigated by the panel structure of the data.³⁵ More precisely, the inclusion of district fixed effects in the panel esti-

³⁵This is an important advantage over a cross-sectional analysis, where unobservable location characteristics would likely be correlated both with income and transport infrastructure. An example for this issue is the economic or political situation in a state or district, which may determine whether (and how effectively) investments in roads are made. The same characteristics may also affect the economic performance through the overall investment climate or other government actions. As it is difficult to control for all relevant location characteristics, transport infrastructure could be correlated with the error term in a cross-sectional analysis.

mation in column 1 absorbs all time-invariant district characteristics such as the initial level of development. The estimation only exploits the changes in income and in transport infrastructure over time. Although in principal it could also be that there are time-varying district characteristics (for example economic or political shocks) that affect *changes* in income, it is unlikely – 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.

A second source of endogeneity is due to reverse causality of economic performance on transport infrastructure. As mentioned in the discussion of the identification strategy in Section 5, it is evident that the GQ was not built randomly, but targeted the four nodal cities Delhi, Calcutta, Chennai, and Mumbai. It is possible that the choice of this network structure was driven by the economic performance (or the anticipation of which) in certain regions. Following the identification strategy by Chandra and Thompson (2000) and Michaels (2008), I exclude the nodal states and only exploit the variation in districts that were in between these nodes and therefore accidentally affected by the roads connecting the four largest centers. This identification strategy is exploited in column 2, which re-estimates the specification of column 1 for the sub-sample of non-nodal districts. The point estimate changes only marginally to 0.395 and continues to be significant at 5 percent. This suggests that the observed correlation between market access and income is not driven by the endogenous location of infrastructure.³⁶

The first two specifications include state-year fixed effects to control for potential differences in the time trends across states. The identification therefore comes from variation in districts over time and within states. The inclusion of the state-year fixed effects implies that cross-state differences in the growth rates from 2000 to 2009 are absorbed. This could be a concern if these differences are related to changes in the transport infrastructure during this period, such that part of the effect of transport infrastructure may be attributed to the state-year fixed effect instead of to an increase in market access. This is a likely scenario,

³⁶Further 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 transport infrastructure between 2000 and 2009 are uncorrelated with districts’ growth trends prior to the start of the NHDP.

since some states experienced larger average reductions in travel costs than others and may have had higher growth for this reason. Furthermore, the estimated fixed effects are unlikely to be constant over the counterfactual networks. An alternative approach would be to find controls for intrinsic district trends which are independent of the road investments. Since the light data goes back to 1992, it is possible to include the pre-2000 growth in light to capture state-specific trends prior to the road investment program. I also include initial levels of light and the shares of services and industry in state income.³⁷ As column 3 shows for the full sample, when the state-year fixed effects are replaced with the state-level control variables, then the point estimate increases to 0.537 and continues to be statistically significant at the 5 percent level. In column 4, I re-estimate the specification of column 3 for the sub-sample of non-nodal states and find a similar estimate of 0.574.

For the counterfactual exercise, I rely on the sub-sample of non-nodal districts to estimate the causal effect of market access on income. I therefore report the results of the counterfactual exercise based on the estimates of columns 2 (state fixed effects) and 4 (state controls). Each of the two specifications has its advantages and the two estimates provide a credible range for the true effect. Column 2 appears to be the more conservative specification to estimate β since it absorbs the state-level effects of roads. I therefore view this estimate as a lower bound. Column 4 provides an alternative for controlling for state trends and its estimate for β of 0.574 is close to the elasticity predicted by the model (for a reasonable choice of parameter values). However, in this case the identification of the causal effect relies fully on the randomness in the treatment of non-nodal states.

6.2 Aggregate Effects of Transport Infrastructure

The results above established an effect of transport infrastructure on income through the channel of market access. The estimate of this effect can be used to predict each district's income for various counterfactual networks. Because market access captures the general equilibrium effects of transport infrastructure, this al-

³⁷More precisely, I include interactions of the year indicator with these variables, as the district fixed effects absorb the time-invariant characteristics.

lows to analyze the aggregate and the distributional consequences.

Aggregate Effects of the Existing Infrastructure

In order to interpret the magnitude of the impact of a counterfactual infrastructure, I first evaluate the effect of the actually built GQ and NS-EW. To this aim, I construct a network in 2009 in the absence of the GQ and NS-EW and ask how India would have developed if the GQ and the NS-EW had not been built.³⁸ In the conceptual framework used here, this would imply that there was no change in the measures for market access because the actually built GQ and finished parts of the NS-EW are the only sources of variation in the road infrastructure that I consider. Based on the specification in the second column of Table 2, the results suggest that aggregate income in 2009 would be 3.4 percent lower if the GQ and the NS-EW had not been built. The result based on the specification with state controls instead of state fixed effects would imply an effect of 4.7 percent of income. Most of this effect is due to the GQ, which alone led to a 2.4 - 3.5 percent higher income by 2009. Aggregate GDP in India in 2009 was USD 905 billion (1999 prices) such that a 3.4 percent difference would correspond to roughly USD 31 billion.³⁹

Aggregate Effects of the Counterfactual Infrastructure

In this section I ask how India would have developed if it had built a transport infrastructure like the Chinese NEN. There are two scenarios which may be considered in this context. First, I *add* a network that mimics the Chinese NEN to the Indian transport infrastructure. Given the evidence that the current transport system does not have enough capacity, this is an interesting scenario to analyze. However, the counterfactual network will create certain redundancies with the already improved infrastructure. The results based on column 2 of Table 2 sug-

³⁸In order to avoid confusion with the terminology, I do not use the term "counterfactual" when considering the 2009 network without the GQ and NS-EW, although it is evidently a counterfactual network. The term "counterfactual" will only be used for the network proposed in Section 4.2.

³⁹In what follows the base is omitted, but all values are in USD at 1999 prices. The costs of the investments are discussed below.

gest that income in 2009 would have been 2.2 percent higher if in addition to the existing infrastructure also the network discussed in Section 4 had been built, which translates into a USD 20 billion benefit. The increase in income amounts to 3.2 percent when the specification of column 4 of Table 2 is used, or USD 29 billion.

In a second scenario, I ask how India would have developed if instead of building the GQ it had chosen a similar strategy as the Chinese NEN. Therefore, I consider a highway network where the GQ is *replaced* by the network proposed in Section 4. The total length of the GQ is 5,846 km, while the length of the counterfactual is 12,840 km. Given this difference, it is not surprising that the aggregate effect of replacing the GQ with the counterfactual is an increase in aggregate income relative to the actual network. This increase is (based on column 2 of Table 2) roughly 1.5 percent of income in 2009 (USD 13 billion). The increase is 2.2 percent based on the specification with state controls, which implies a USD 20 billion benefit. Although the counterfactual network has more than twice the length of the GQ, the aggregate effects are relatively small in comparison to the 2.4 - 3.5 percent increase due to the GQ. The likely reason is that the GQ connected the four largest economic centers and could thereby generate a more substantial increase in market access than connections to intermediate-sized cities would in the case of the counterfactual. However, there are important distribution consequences, which will be discussed further below.

Construction Costs of the Counterfactual Infrastructure

In order to evaluate the net aggregate gain of the counterfactual networks as described above, it is necessary to compute the construction costs of the counterfactual network. The challenge is that the normalized units of Equation (1) do not correspond to actual monetary costs and only represent the relative increase in construction costs due to the various topographic features. In order to derive the expected costs of the counterfactual in USD, I first replicate the GQ on the cost grid discussed in Section 4, which yields the costs of the GQ in terms of the units in Equation (1). Since the actual cost of the GQ are known, one can derive the ratio between the costs in units of Equation (1) and the actual USD costs.

This ratio then allows to predict the USD costs of the counterfactual roads by scaling the costs in Equation (1). Using this approach, the costs of the full counterfactual network are approximately USD 17.5 billion. However, when adding the counterfactual network to the existing network, then a number of connections are redundant because they have already been made by the GQ or NS-EW. When excluding these connections from the cost calculations, then the total costs amount to USD 14.3 billion.⁴⁰ To compare costs and benefits, I have to make assumptions on the interest rate and on the year in which the construction costs have to be paid. For the later I assume that the costs were evenly spread over the years 2001-2009. For the real interest rate during the period 2001-2009, I take 2 percent.⁴¹ With accumulated interests until 2009, the costs amount to USD 19 for the full network and to USD 16 billion for the network without redundancies.

Comparing these costs to the benefits suggests aggregate net gains from *adding* the counterfactual of about USD 4 billion when using state fixed effects and USD 13 billion when using state control variables. The net gains from *replacing* the GQ with the counterfactual would lie between zero and USD 7 billion.⁴² Besides these gains over the first 9 years of the project, there may be higher benefits of such infrastructure projects on a longer horizon, although in that case the discount rate plays a key role and one has to consider additional costs for maintenance.

Overall, I conclude that by connecting the largest economic centers, the GQ had achieved relatively large aggregate effects at a moderate cost, i.e. there were net gains from the GQ. Taking into account the construction costs, the additional net gains of the counterfactual are small. However, the counterfactual network also includes the intermediate-sized cities and thereby reaches into more districts, including the less developed. This has important distributional consequences, which will be quantified below.

⁴⁰The costs of the GQ were USD 6 billion (Ghani et al., 2012), but it had less than half the length of the counterfactual.

⁴¹According to the Reserve Bank of India, the average nominal interest rate on central government bonds for the period 2001-2009 was 7.5 percent. However, the benefits and costs above are denoted in 1999 constant prices and the average inflation rate between 2001 and 2009 was 5.7 percent. Using, instead of an average value of 2 percent, the yearly real interest rates based on government bonds and inflation would imply costs of USD 15.3.

⁴²In this case the costs of the full counterfactual network have to be paid, but USD 6 billion would have been saved on the GQ.

6.3 Distributional Effects of Transport Infrastructure

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). In this light, 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 over the various versions of the transport networks. I will start by discussing the effects of the actually implemented parts of the NHDP and then consider the effects of the two counterfactual scenarios. To be able to discuss the differences across states and districts, the analysis builds on the specification using state-level control variables (column 4 of Table 2) instead of state-year fixed effects.

Distributional Effects of the Existing Infrastructure

Figure 10 shows the estimated effects of the GQ and the finished parts of the NS-EW (the network in Figure 5) at the level of Indian districts. The numbers represent the percentage gains from building the GQ and finished parts of the NS-EW.⁴³ As expected, the effects are strongest along the paths of the newly built or upgraded highways. The largest beneficiaries had a 33 percent higher income level in 2009 than they would have had in the absence of the GQ and NS-EW. Although transport costs did not increase anywhere, there are some losers from the infrastructure investments, which is due to trade diversion.

Distributional Effects of the Counterfactual Infrastructure

Adding the counterfactual network to the existing infrastructure increases particularly the market access of regions in the center and in the east which have been

⁴³More precisely, the numbers correspond to differences between the prediction of income with the actual network and the prediction with the counterfactual network, relative to the prediction with the actual network.

neglected by the recent improvements in infrastructure. Particularly the central regions were also the ones with low growth during the past decade. As there are several cities in that area which would fulfill the criteria of the NEN, they become better connected by the counterfactual network and thus experience increases in market access. The results from adding the counterfactual network to the existing infrastructure, illustrated in Figure 11, also show that the growth benefits are generally more equally distributed across the country compared to the existing network with the GQ and the NS-EW. The largest beneficiaries gain 32 percent of income, but there are also some districts with moderate losses. This is purely due to trade diversion, since trade costs do not increase anywhere when transport infrastructure is added.

When replacing the GQ with the counterfactual network, then the above results in terms of distribution are strengthened and the largest beneficiaries are still regions with low growth during the past decade (Figure 12). But in this scenario, there are also significant losses because, by replacing the GQ with an alternative network, those cities which were well connected by the GQ have higher (or at best the same) trade costs. In particular, the four large cities which were targeted by the GQ experience declines. The largest beneficiaries have gains of up to 36.8% and the largest loss is 17.4%.

Planned Highway Projects and Policy Implications

The above evidence on the impact of transport infrastructure makes clear 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 largely 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 this and the NHDP did include plans for other highway connections besides the GQ that reach into more states. In particular, the government planned the North-South and East-West Corridors which cross through regions that were not reached by the GQ. However, these

other projects were delayed and by 2009 only a small part has been finished (see also Ghani et al, 2012). A complementary exercise is therefore to compare the effects which could be expected from completing the planned corridors to the effects from the counterfactual network that follows the Chinese strategy. The results suggest that the counterfactual network generates about 1 percent higher income than the completion of the NS-EW corridors. Clearly, districts along the planned but not yet implemented routes reap larger gains from these corridors than from the counterfactual. There are therefore losers if the instead of the planned NS-EW corridors a network with Chinese characteristics is built. But, as is shown in Figure 13, the counterfactual network reaches more regions and especially parts of the country that have experienced low growth rates.

7 Robustness

The results presented so far rely on a particular counterfactual network (the least-cost network), Furthermore, I had to make a number of assumptions to estimate the effect of transport infrastructure and to perform the counterfactual analysis. This section first presents the results for further counterfactual networks and then discusses robustness to pre-investment trends, weighting, and alternative choices for parameter values.

7.1 Alternative Counterfactual Highway Networks

The least-cost network presented in Section 4 implements the cheapest network that connects the cities which would fulfill the criteria of the Chinese NEN. The advantage of this network is that it represents an objective way to replicate the Chinese policy. The disadvantage is that it has some features that we do not observe in reality. In particular, there are some cities which are relatively close but not directly connected. The reason for this is that the Kruskal algorithm has as its only objective to connect all cities in the cheapest way. It does not take into account that a connection between two cities which are already part of the network could be made in order to further reduce transport costs. This is particularly evident when considering the "loose" ends in Figure 9.

In order to evaluate the effect of networks that are closer to those we observe in reality, I construct two alternative counterfactual networks. I first "close" the network of Figure 9 by adding connections between the loose ends. This network is shown in Figure 14. The results suggest that replacing the GQ with this counterfactual network would increase income by 1.9 - 2.8 percent, roughly 0.5 percentage points more than the least-cost network. For a second counterfactual network, I exploit that the Chinese policy not only specified which types of cities should be connected, but also how (World Bank, 2007b). More precisely, the policy stated that the cities with a population above 500,000 and state capitals should be connected in a network that consists of rays spreading out from the capital city Beijing and with horizontal and vertical corridors. I apply this approach to India by connecting the targeted cities with rays from Delhi and horizontal and vertical connections, while the precise routes of these connections are again determined by the terrain. I follow the rough rule that the number of connections should be proportional to the total land area of the two countries, thus making about three times fewer rays and corridors in India than in China. The resulting network is illustrated in Figure 15. This network resembles the actually built NEN and is significantly denser than the least-cost network proposed in Section 4. It is therefore not surprising that this network generates large gains when replacing the GQ. The estimated effects are between 3.3 and 4.9 percent. The gains correspond to between USD 30 and 44 billion, while the estimated construction costs of this network amount to USD 31 billion.⁴⁴ This result is consistent with the finding for the least-cost network that there may be some net gains but that they are not necessarily large. However, the maybe more important implication is again due to the distributional effect. As is shown in Figure 16, some of the largest gains are precisely in those regions which previously experienced low growth.

7.2 Trends in District Growth Prior to Road Investment

The identification strategy used in Section 6 relied on the assumption that non-nodal districts were randomly affected by the GQ that connected the four largest

⁴⁴I again assume that the costs were evenly spread over the 9 years and I use an average annual real interest rate of 2 percent.

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. 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 columns 2 and 4 of Table 2 for non-nodal states 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 3. In neither of the two specifications is the estimate significant. Furthermore, the point estimate is in both cases substantially lower than the ones in Table 2 or even negative. This provides compelling evidence against the hypothesis that the GQ connected non-nodal districts that were already growing faster.

7.3 Weighting by Initial Income

A further concern may be that the results presented so far are driven by districts with little initial light, because small absolute changes in light in these districts may generate large growth rates. Although it is not evident that smaller districts should be given less weight, one may worry that their measurement error is larger. Table 4 repeats the regressions of Table 2 but weights observations by the logarithm of initial light in 2000. While not being more precise, the estimates are of lower magnitude when weighting by initial light. Using these estimates for the counterfactual analysis yields somewhat smaller aggregate effects, but does not change the distributional implications of the different transport networks.

7.4 Alternative Values for the Trade Elasticity

When solving the system of equations in (21) numerically to derive the market access measures, it was necessary to make a choice for the trade elasticity pa-

parameter θ . The value of 3.8 was chosen based on Donaldson (forthcoming) who estimated the trade elasticity using bilateral trade data between Indian districts during the colonial time. Although not estimated from current trade data, this estimate is in line with recent evidence by Simonovska and Waugh (2013). Other estimates in the literature tend to be higher. Therefore, Tables 5 - 7 report the estimated effect of market access on income with values for the trade elasticity θ of 3, 5, and 7. The point estimates for the preferred specifications in columns 3 and 4 range from 0.41 to 0.65. With one exception (column 4 in Table 5), the estimates are significant at the 10 percent level. The preceding analysis in Section 6, by using a value of 3.8 for the trade elasticity, estimated the effect of income with respect to market access around 0.54 or 0.57. These appear to be intermediate values compared to the estimates from the various alternative θ .

7.5 Light per Capita

In the conceptual framework and in the empirical analysis so far, I have abstracted from changes in population across districts that may lead to changes in income per capita. In a further robustness check, I therefore use census data on population size in each district in 2001 and 2011 in order to estimate the effect on light per capita.⁴⁵ The decadal growth rates of light per capita across mainland Indian districts are shown in Figure 17. The elasticities of income per capita with respect to market access (using the baseline specification in Table 2) are shown in Table 8. The estimates suggest that the result is similar when using aggregate or per capita figures. For example, the fixed effects regression without nodal states (column 2) suggests a small increase in the point estimate from 0.395 to 0.408 and the standard errors are also relatively similar.

⁴⁵Data collection for the 2011 census began in 2010. Although there is still a small discrepancy to the period for the light data (2000 to 2009), it is unlikely that this would change the results substantially.

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 a key 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, which is needed to estimate the causal effects.

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 model. 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 integration of precise geographic data and economic theory allows me to perform detailed counterfactual analyses. The findings suggest that the recent Indian highway investments led to positive aggregate net gains, but unequal effects across districts. A comparison to a counterfactual network that connects intermediate-sized cities shows modest aggregate effects relative to the actual network. However, there are important distributional effects. In particular, previously less developed regions would benefit substantially from such a network that integrates regions of intermediate density. The results of this counterfactual stand in stark contrast to the ones of the actual highway investments in India, which have focused on connecting the four largest economic centers. The comparison of the regional development effects under the actual and counterfactual networks therefore provides an explanation for the lack of convergence by India's lagging regions.

The implications of the findings above extend to other countries. The theoretical framework allows to quantify the aggregate and distributional effects and I find that particularly the later are sizable. 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 methods applied in this paper – the effects of infrastructure projects could be evaluated in many other settings.

The possibilities from integrating economic theory with geographic data also raise interesting new questions. For example, it would be fascinating to identify the optimal network with respect to policy objectives such as aggregate income or distributional goals. There could also be important interactions of regional economic policies with the transport network. Since infrastructure investments are often tied to economic policy, it would be important to consider them jointly for the policy analysis. Finally, the decision of how to invest in transport infrastructure can hardly be separated from environmental considerations. This connection is apparent in many policy debates and there may be tradeoffs between different objectives. The present paper contributes to this debate by quantifying the economic impact of transport infrastructure and thereby providing a sound basis for a more general assessment.

9 References

- Anderson, James E. and Eric van Wincoop** 2003. "Gravity with Gravitas: A Solution to the Border Puzzle." *American Economic Review*, 93(1): 170-192.
- Alder, Simon, Lin Shao, and Fabrizio Zilibotti** 2013. "Economic Reforms and Industrial Policy in a Panel of Chinese Cities." *CEPR Discussion Paper* 9748.
- Allen, Treb and Costas Arkolakis** 2013. "Trade and the Topography of the Spatial Economy." *NBER Working Paper* 19181.
- Asian Development Bank** 2007. "Retrospective Analysis of the Road Sector, 1997-2005." *Asian Development Bank*.
- Atak, 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.
- Atkin, David and Dave Donaldson** 2013. "Who's Getting Globalized? The Size and Nature of Intranational Trade Costs." mimeo, MIT.
- 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>.
- 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** 2013. "Roads, Railroads and Decentralization of Chinese Cities." mimeo, Brown University.

- Breinlich, Holger, Gianmarco I. P. Ottaviano, and Jonathan R.W. Temple** 2013. "Regional Growth and Regional Decline" *CEP Discussion Paper* 1232.
- 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.
- Chinese Ministry of Transportation** 2004. "Plans for the National Expressway Network" (in Chinese).
- Center for International Earth Science Information Network (CIESIN)** 2013 "Global Roads Open Access Data Set, Version 1 (gROADSv1)." Columbia University, and Information Technology Outreach Services - ITOS - University of Georgia. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). <http://sedac.ciesin.columbia.edu>.
- Coşar, A. Kerem and Pablo D. Fajgelbaum** 2013. "Internal Geography, International Trade, and Regional Specialization." *NBER Working Paper* 19697.
- Datta, Saugato** 2012. "The Impact of Improved Highways on Indian Firms." *Journal of Development Economics* 99: 46-57.
- Dell, Melissa** 2012. "Trafficking Networks and the Mexican Drug War." mimeo, Harvard University.
- Démurger, Sylvie** 2001. "Infrastructure Development and Economic Growth: An Explanation for Regional Disparities in China?" *Journal of Comparative Economics* 29: 95-117.
- DHL** 2007. "Logistics in India."
- Desmet, Klaus, Ejaz Ghani, Stephen O'Connell, and Esteban Rossi-Hansberg** forthcoming. "The Spatial Development of India" *Journal of Regional Science*.

- Desmet, Klaus and Esteban Rossi-Hansberg** forthcoming. "Spatial Development." *American Economic Review*.
- Donaldson, Dave** forthcoming. "Railroads and the Raj: Estimating the Impact of Transport Infrastructure" *American Economic Review*.
- Donaldson, Dave and Richard Hornbeck** 2013. "Railroads and American Economic Growth: A 'Market Access' Approach." *NBER Working Paper Series* 19213.
- Earth Observation Group** 2013. "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.
- 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** 2013. "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System." mimeo, University of California, Berkeley.
- Fogel, Robert W.** 1964. "Railroads and American Economic Growth: Essays in Econometric History." Johns Hopkins University Press, Baltimore.
- Frankfurter Allgemeine Zeitung (FAZ)** 2013. "Indien wird China überholen." www.faz.net (accessed on 19.10.2013).
- Fujita, Masahisa, Paul Krugman, and Anthony J. Venables** 1999 "The Spatial Economy. Cities Regions, and International Trade." MIT Press, Cambridge.
- Ghani, Ejaz Arti Grover Goswami, and William R. Kerr** 2012. "Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing." *NBER Working Paper Series* 18524.

- Global Administrative Areas** 2012. "GADM Version 2." <http://www.gadm.org/country>.
- Global Land Cover Facility** 2013. "UMD Land Cover Classification." University of Maryland Department of Geography, <http://glcf.umd.edu/data/landcover/>
- Gollin, Douglas and Richard Rogerson** 2010 "Agriculture, Roads, and Economic Development in Uganda" *NBER Working Paper Series* 15863.
- 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." *CEPR Discussion Paper* 9322.
- 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.
- 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>.

- 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.
- Liu, Z** 2000. "Economic Analysis of a Rural Basic Access Road Project. Case Study: Andhra Pradesh, India." Rural Transport Knowledge Base.
- 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.
- 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.** 2010. "The Empirics of New Economic Geography." *Journal of Regional Science* 50(1): 297-311.
- Redding, Stephen J.** 2012. "Goods Trade, Factor Mobility and Welfare." *NBER Working Paper* 18008.
- 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 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** forthcoming. "The Elasticity of Trade: Estimates and Evidence." *Journal of International Economics*.
- Storeygard, Adam** 2013. "Farther on down the road: Transport costs, trade and urban growth in Sub-Saharan Africa." *World Bank Policy Research Working Paper* 6444.
- 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** 2013. "World Development Indicators." The World Bank.

10 Tables

Table 1: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Sum of light 2000	17884.53	18965.65	0	138936.5	590
Mean light intensity 2000	4.35	6.52	0	62.91	590
Sum of light 2009	19552.13	22124.29	0	151527	590
Mean light intensity 2009	4.89	7.41	0	63	590
Within-sat yearly growth rate	0.03	0.11	-0.22	1.85	588
District area (km2)	5329.64	4676.99	18	47855	590

The table shows the summary statistics of the light data. Sum of light is the total light measured within a district's boundaries. Mean light intensity is light per pixel. The within-satellite yearly growth rate from 2000 to 2009 is reported to address differences in satellite calibration. This figure is the average growth rate when only growth observations from the same satellite are used. The total sample consists of 590 Indian mainland districts.

Table 2: Income and Actual Transport Infrastructure Investments

	(1)	(2)	(3)	(4)
Log Market Access	0.392**	0.395**	0.537**	0.574**
	(2.58)	(2.36)	(2.54)	(2.29)
Excluding nodal states	No	Yes	No	Yes
State-year fixed effects/State controls	FE	FE	State controls	State controls
District fixed effects	Yes	Yes	Yes	Yes
N	585	501	582	498
R2	0.497	0.492	0.212	0.209

The table shows the elasticity of income with respect to market access (based on within-variation). The dependent variable is the logarithm of the sum of light in each district in the years 2000 and 2009. The explanatory variable is market access computed based on Equation (21). All regressions include district fixed effects. Columns 1 and 2 include state-year fixed effects. In columns 3 and 4 the state-year fixed effects are replaced by a set of state-level control variables (interactions of year with initial log light in 1992, growth of state-level light from 1992-2000, and shares of services and industry). T-statistics are shown in brackets. Standard errors are clustered at the state-level.

Table 3: Income Prior to Road Investments

	(1)	(2)
Log Market Access	0.128 (0.43)	-0.0834 (-0.33)
Excluding nodal states	Yes	Yes
State-year fixed effects/State controls	FE	State controls
N	489	486
R2	0.412	0.232

The table shows the results from regressing districts' income growth between 1992 and 1999 on changes in market access between 2000 and 2009. Column 1 includes state fixed effects. In column 2 the state fixed effects are replaced by a set of state-level control variables (interactions of year with initial log light in 1992, growth of state-level light from 1992-2000, and shares of services and industry). T-statistics are shown in brackets. Standard errors are clustered at the state-level.

Table 4: Income and Actual Transport Infrastructure Investments when Weighting by Initial Light

	(1)	(2)	(3)	(4)
Log Market Access	0.282*** (2.88)	0.268** (2.57)	0.385** (2.36)	0.391* (1.91)
Excluding nodal states	No	Yes	No	Yes
State-year fixed effects/State controls	FE	FE	State controls	State controls
District fixed effects	Yes	Yes	Yes	Yes
N	585	501	582	498
R2	0.520	0.515	0.212	0.206

The table shows the elasticity of income with respect to market access (based on within-variation). The dependent variable is the logarithm of the sum of light in each district in the years 2000 and 2009. The explanatory variable is market access computed based on Equation (21) with a trade elasticity (θ) of 3.8. All regressions include district fixed effects (not shown). Columns 1 and 2 include state-year fixed effects. In columns 3 and 4 the state-year fixed effects are replaced by a set of state-level control variables (interactions of year with initial log light in 1992, growth of state-level light from 1992-2000, and shares of services and industry). Observations are weighted by the log of districts' light in 2000. T-statistics are shown in brackets. Standard errors are clustered at the state-level.

Table 5: Income and Actual Transport Infrastructure Investments with Trade Elasticity (θ) of 3

	(1)	(2)	(3)	(4)
Log Market Access ($\theta = 3$)	0.297**	0.277**	0.414*	0.416
	(2.61)	(2.27)	(1.85)	(1.54)
Excluding nodal states	No	Yes	No	Yes
State-year fixed effects/State controls	FE	FE	State controls	State controls
District fixed effects	Yes	Yes	Yes	Yes
N	585	501	582	498
R2	0.493	0.488	0.204	0.201

The table shows the elasticity of income with respect to market access (based on within-variation). The dependent variable is the logarithm of the sum of light in each district in the years 2000 and 2009. The explanatory variable is market access computed based on Equation (21) with a trade elasticity (θ) of 3. All regressions include district fixed effects (not shown). Columns 1 and 2 include state-year fixed effects. In columns 3 and 4 the state-year fixed effects are replaced by a set of state-level control variables (interactions of year with initial log light in 1992, growth of state-level light from 1992-2000, and shares of services and industry). Observations are weighted by the log of districts' light in 2000. T-statistics are shown in brackets. Standard errors are clustered at the state-level.

Table 6: Income and Actual Transport Infrastructure Investments with Trade Elasticity (θ) of 5

	(1)	(2)	(3)	(4)
Log Market Access ($\theta = 5$)	0.453** (2.70)	0.464** (2.54)	0.598*** (3.00)	0.648*** (2.82)
Excluding nodal states	No	Yes	No	Yes
State-year fixed effects/State controls	FE	FE	State controls	State controls
District fixed effects	Yes	Yes	Yes	Yes
N	585	501	582	498
R2	0.499	0.494	0.213	0.212

The table shows the elasticity of income with respect to market access (based on within-variation). The dependent variable is the logarithm of the sum of light in each district in the years 2000 and 2009. The explanatory variable is market access computed based on Equation (21) with a trade elasticity (θ) of 5. All regressions include district fixed effects (not shown). Columns 1 and 2 include state-year fixed effects. In columns 3 and 4 the state-year fixed effects are replaced by a set of state-level control variables (interactions of year with initial log light in 1992, growth of state-level light from 1992-2000, and shares of services and industry). Observations are weighted by the log of districts' light in 2000. T-statistics are shown in brackets. Standard errors are clustered at the state-level.

Table 7: Income and Actual Transport Infrastructure Investments with Trade Elasticity (θ) of 7

	(1)	(2)	(3)	(4)
Log Market Access ($\theta = 7$)	0.413***	0.422***	0.501***	0.531***
	(3.08)	(2.98)	(4.67)	(4.52)
Excluding nodal states	No	Yes	No	Yes
State-year fixed effects/State controls	FE	FE	State controls	State controls
District fixed effects	Yes	Yes	Yes	Yes
N	585	501	582	498
R2	0.495	0.491	0.205	0.203

The table shows the elasticity of income with respect to market access (based on within-variation). The dependent variable is the logarithm of the sum of light in each district in the years 2000 and 2009. The explanatory variable is market access computed based on Equation (21) with a trade elasticity (θ) of 7. All regressions include district fixed effects (not shown). Columns 1 and 2 include state-year fixed effects. In columns 3 and 4 the state-year fixed effects are replaced by a set of state-level control variables (interactions of year with initial log light in 1992, growth of state-level light from 1992-2000, and shares of services and industry). Observations are weighted by the log of districts' light in 2000. T-statistics are shown in brackets. Standard errors are clustered at the state-level.

Table 8: Income per Capita and Actual Transport Infrastructure Investments

	(1)	(2)	(3)	(4)
Log Market Access	0.398** (2.37)	0.408** (2.20)	0.485** (2.20)	0.520* (2.01)
Excluding nodal states	No	Yes	No	Yes
State-year fixed effects/State controls	FE	FE	State controls	State controls
District fixed effects	Yes	Yes	Yes	Yes
N	584	500	582	498
R2	0.505	0.502	0.222	0.223

The table shows the elasticity of income per capita with respect to market access (based on within-variation). The dependent variable is the logarithm of the sum of light in each district in the years 2000 and 2009 relative to the population size in 2001 and 2011. The explanatory variable is market access computed based on Equation (21). All regressions include district fixed effects. Columns 1 and 2 include state-year fixed effects. In columns 3 and 4 the state-year fixed effects are replaced by a set of state-level control variables (interactions of year with initial log light in 1992, growth of state-level light from 1992-2000, and shares of services and industry). T-statistics are shown in brackets. Standard errors are clustered at the state-level.

11 Figures

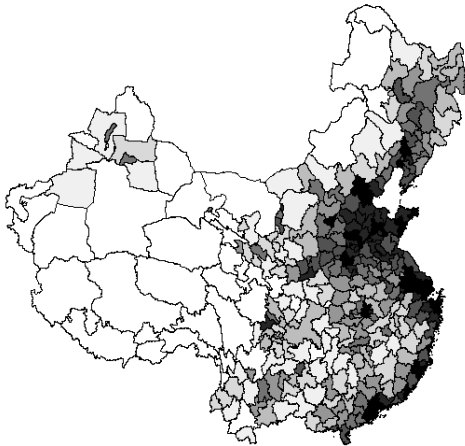
Figure 1: Distribution of Economic Activity in 2000



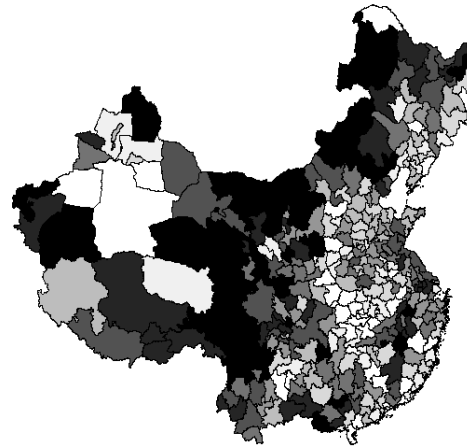
The figure shows the light intensity of each pixel in China and India in the year 2000.

Figure 2: Initial Density and Growth in China

Light Density in 2000
China

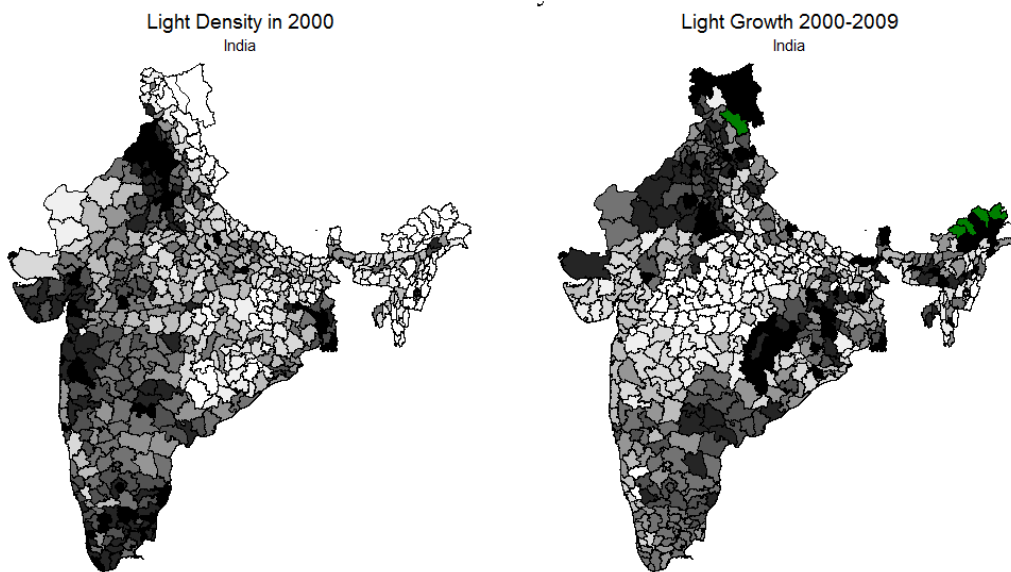


Light Growth 2000-2009
China



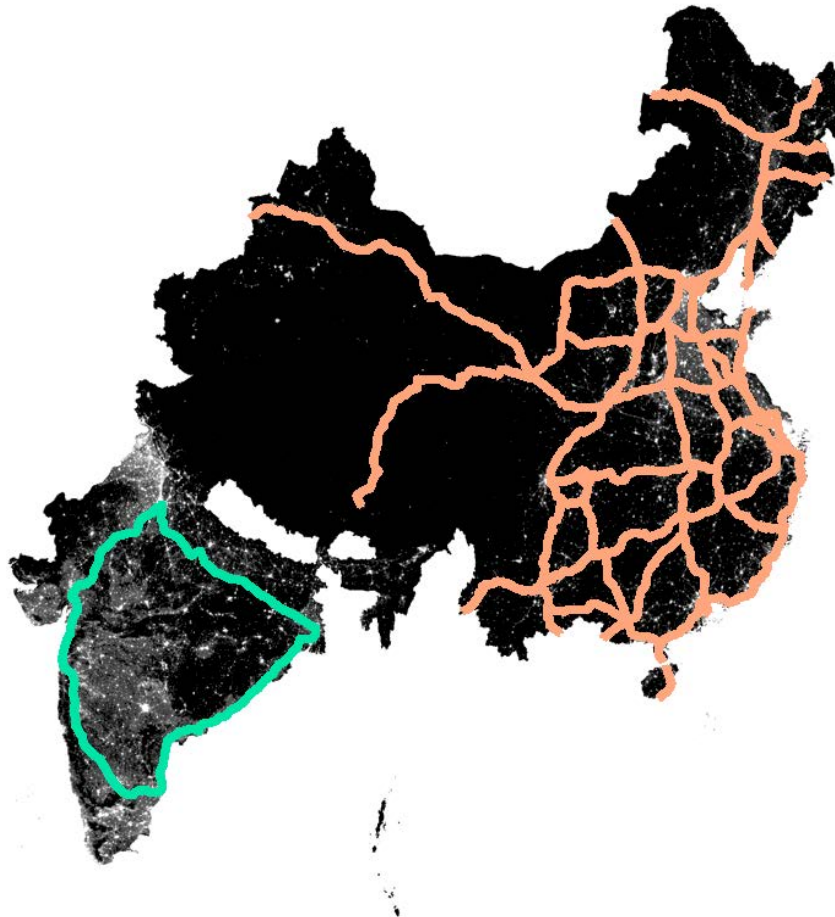
The two maps show the initial density (left map) and growth (right map) in light in China. The initial density measures the average light intensity per pixel in the year 2000. The growth rate measures the long difference in light intensity between 2000 and 2009. The units are 341 Chinese prefectures. Darker colors refer to higher density or higher growth rates.

Figure 3: Initial Density and Growth in India



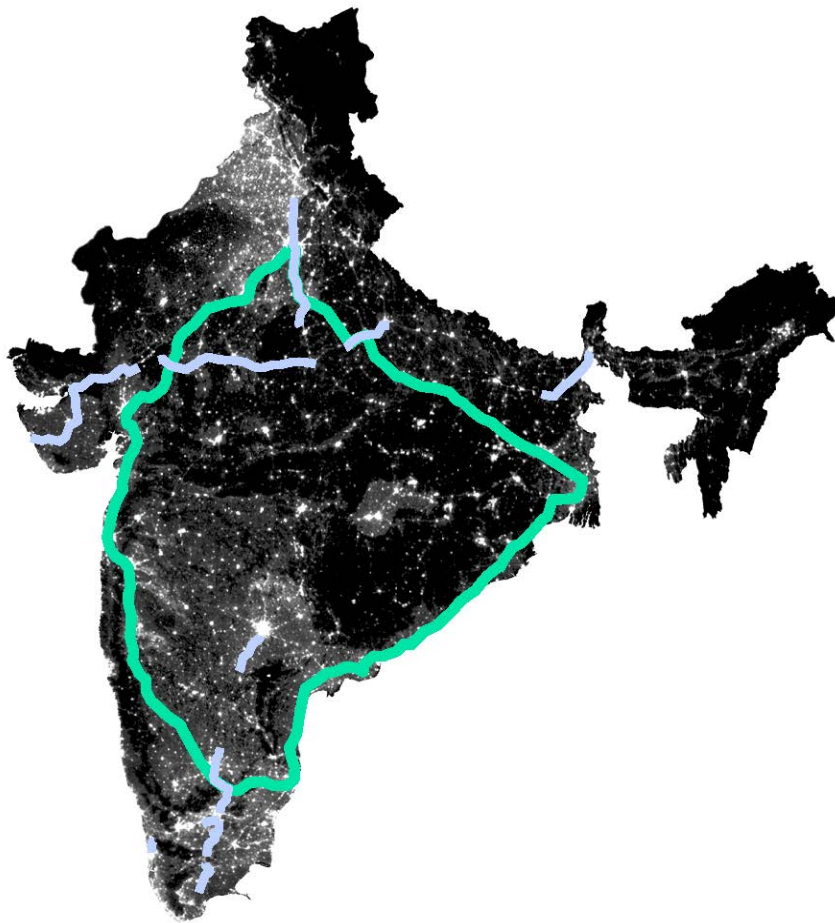
The two maps show the initial density (left map) and growth (right map) in light in India. The initial density measures the average light intensity per pixel in the year 2000. The growth rate measures the long difference in light intensity between 2000 and 2009. The units are 590 Indian districts. Darker colors refer to higher density or higher growth rates. The small green areas in the north and east represent missing observations due to zeros in the initial light per district.

Figure 4: Indian GQ and the Chinese NEN.



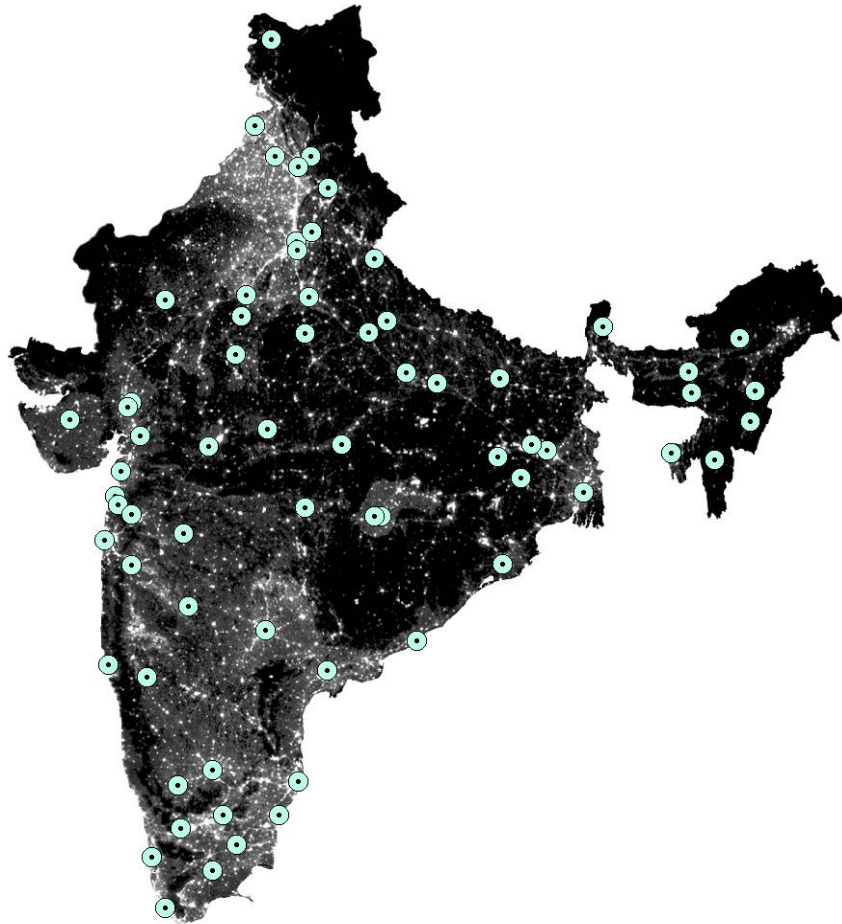
The figure shows the major highway investment programs in China (NEN, in red) and India (QG, in green).

Figure 5: Completed parts of the NHDP by 2010



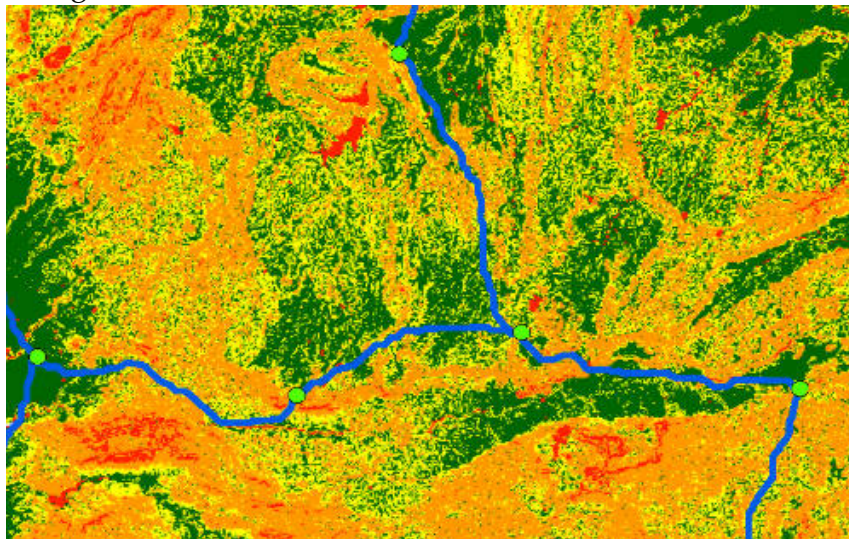
The map shows the parts of the NHDP which were completed by 2010. The GQ is shown as a thick green line and the finished NS-EW corridor parts are shown as thin blue lines.

Figure 6: Cities in India with more than 500,000 residents and all state capitals.



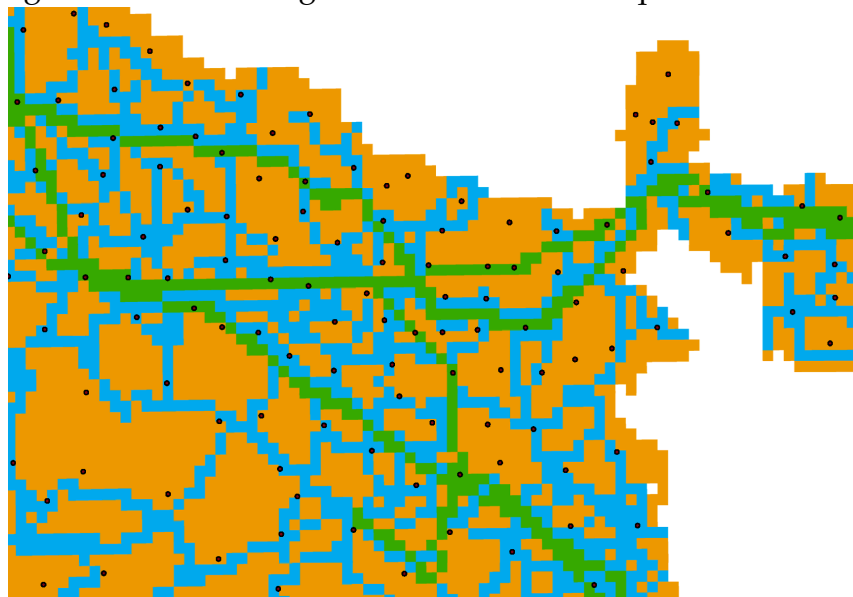
The figure shows the cities in India with more than 500,000 residents and all state capitals. The image in the background shows luminosity in the year 2009.

Figure 7: Road construction costs on the Indian terrain



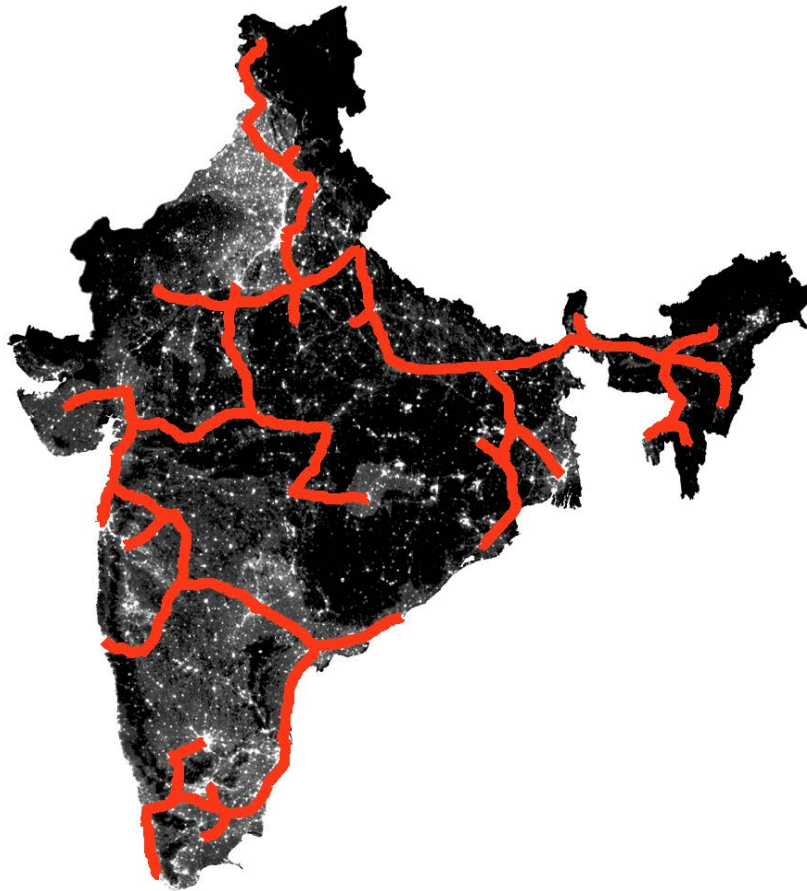
The figure shows the road construction costs as a function of slope and land cover. Dark red refers to high construction costs, orange and yellow to intermediate costs, and light green to low costs. The green circles represent cities in India which fulfill one of the two criteria of the Chinese NEN.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

Figure 8: Travel cost grid based on different qualities of roads



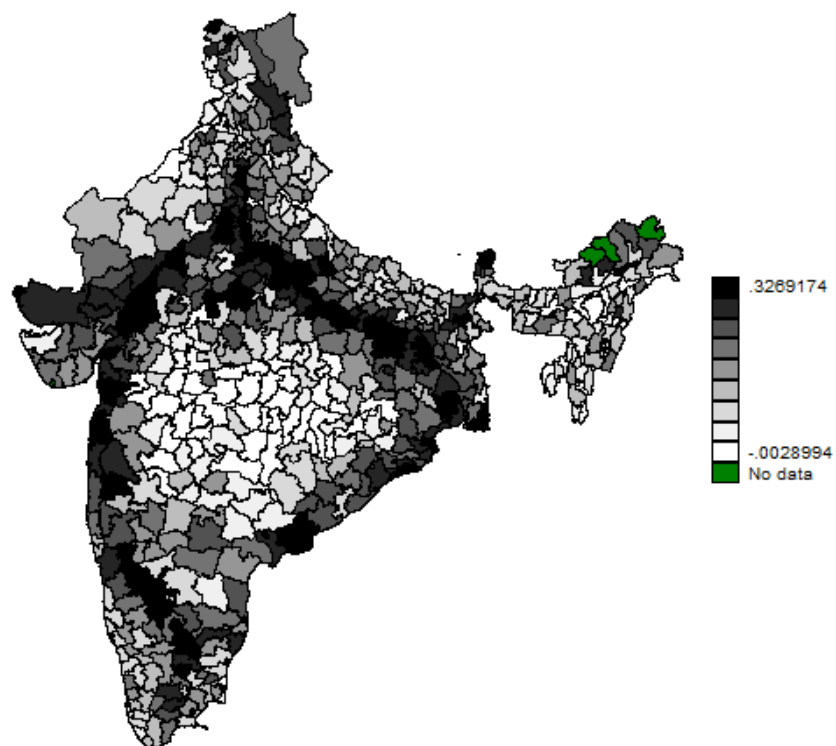
The figure shows a part of the Indian landscape, where the colors of different cells represent differences in travel costs. The green lines represent highways of the NHDP and the blue lines highways of lower quality. The dots represent the centroids of Indian districts between which bilateral trade costs are computed as the least-cost path through the cost grid.

Figure 9: Least-cost counterfactual network



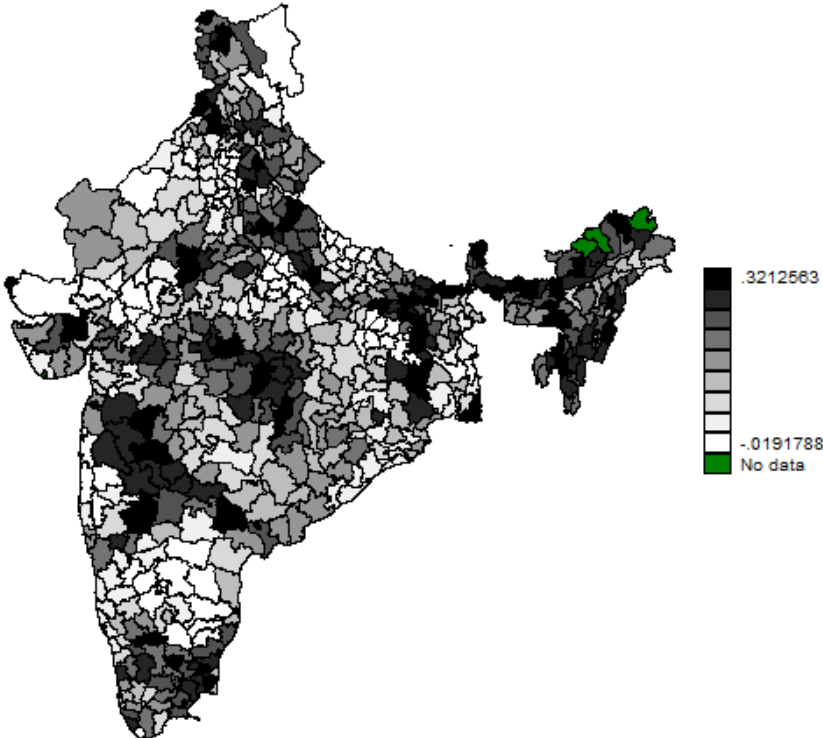
The map shows the counterfactual highways in India which connect all 68 targeted cities in a least-cost network.

Figure 10: Percent increase in income generated by the NHDP



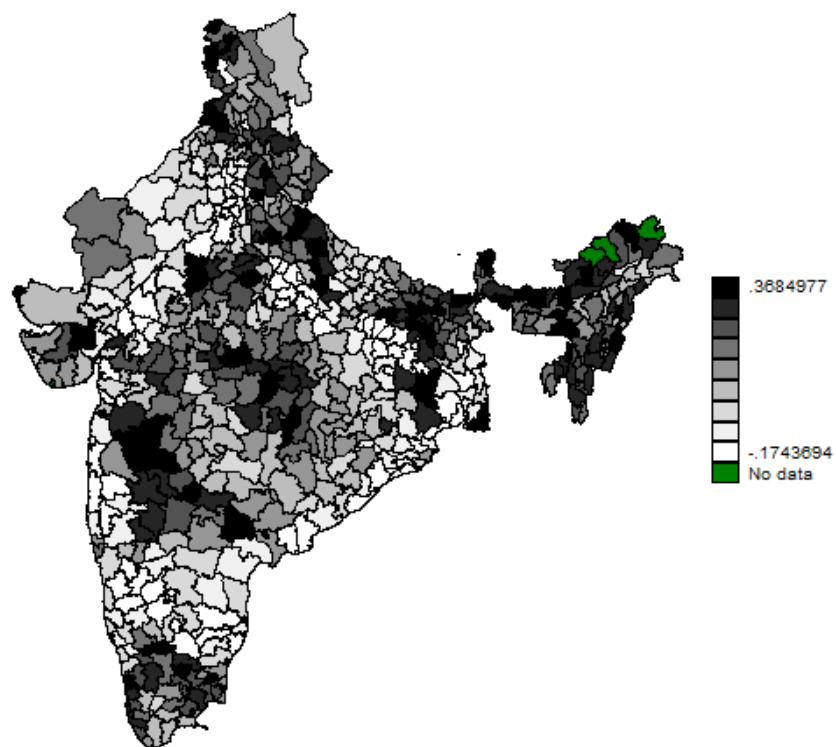
The map shows the boundaries of Indian districts. Darker color represents higher percentage difference in income generated by the NHDP until 2009 compared to a network without the NHDP.

Figure 11: Percent increase in income from adding counterfactual network



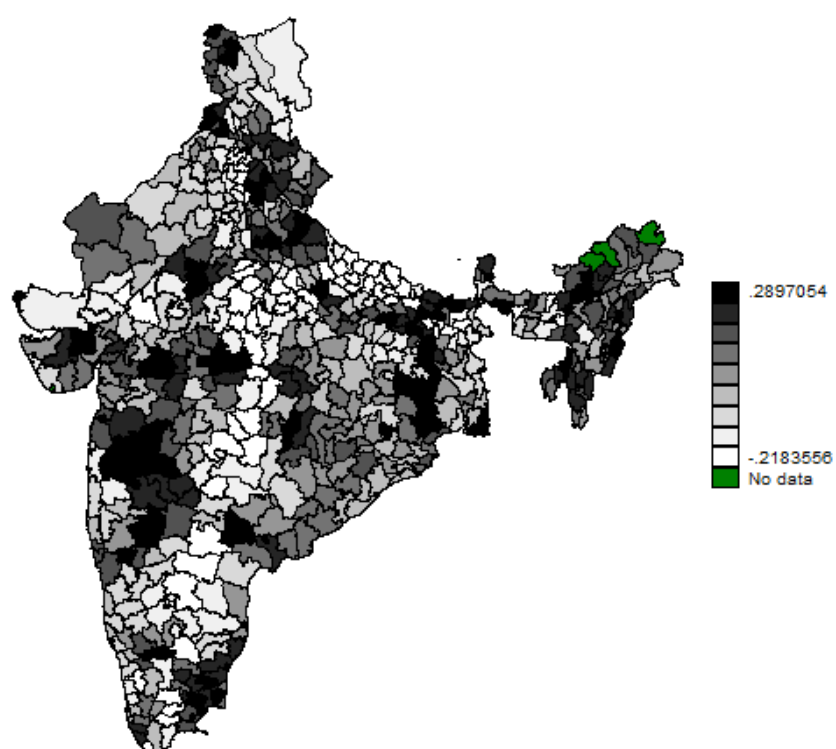
The map shows the boundaries of Indian districts. Darker color represents higher percentage difference in income generated by adding the counterfactual to the existing network.

Figure 12: Percent increase in income from replacing GQ with counterfactual



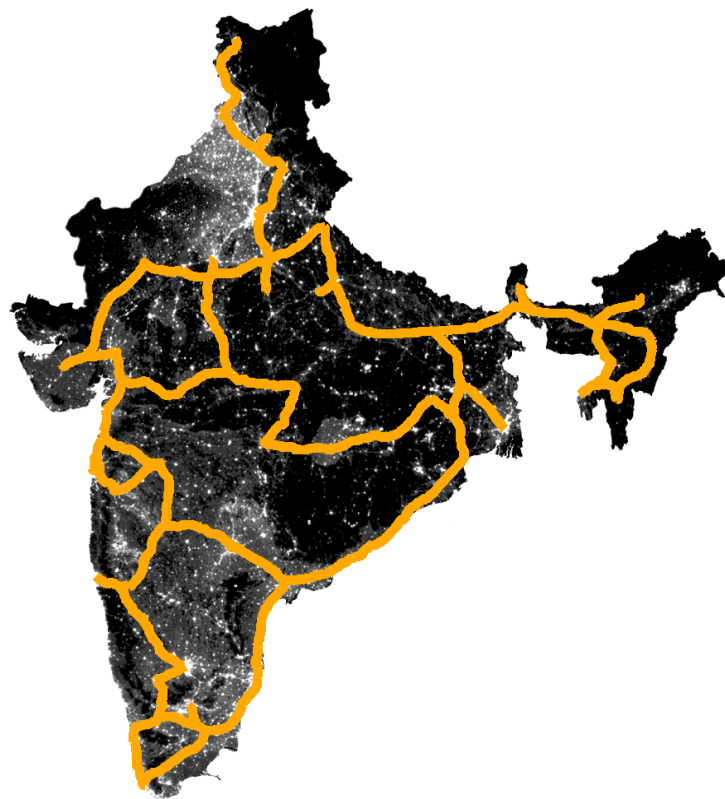
The map shows the boundaries of Indian districts. Darker color represents higher percentage difference in income generated by replacing the GQ with the counterfactual network.

Figure 13: Percent difference between counterfactual and completion of NS-EW corridors



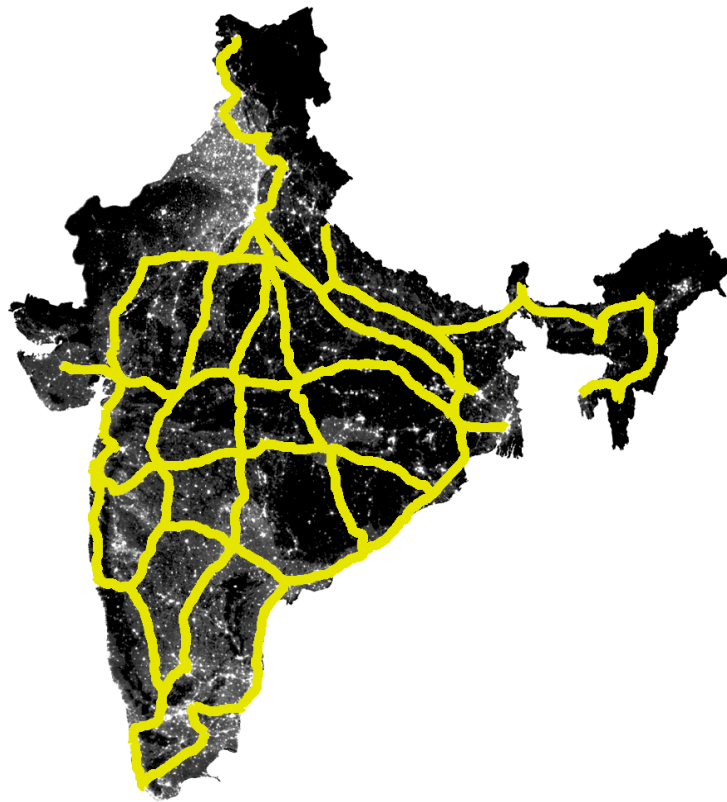
The map shows the boundaries of Indian districts. Darker color represents higher percentage difference in income between the counterfactual network and the completion of the planned NS-EW corridors.

Figure 14: Counterfactual network when connecting loose ends of the least-cost network



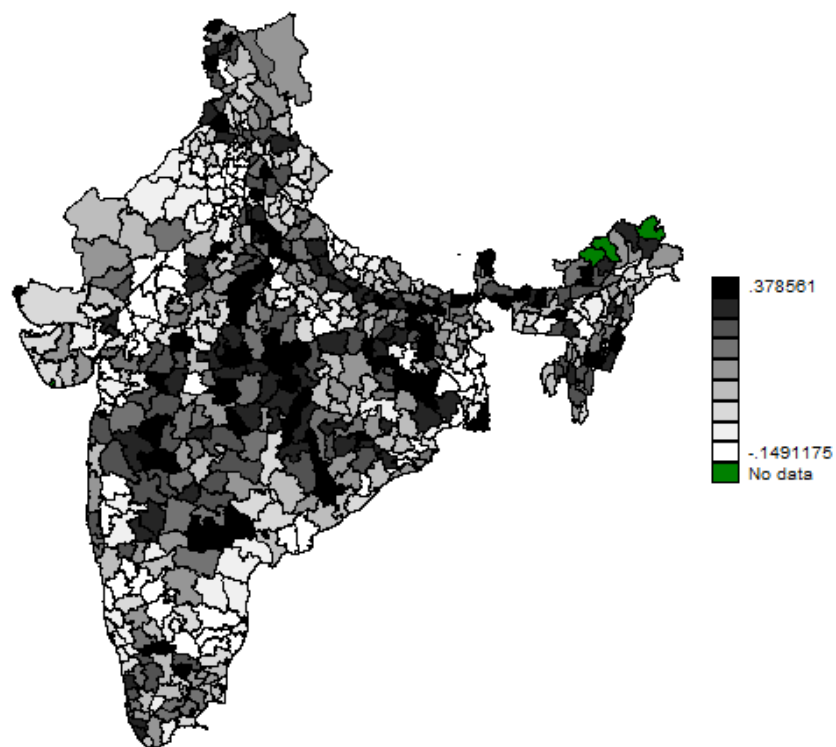
The map shows the counterfactual highways in India which connect all 68 targeted cities in a least-cost network and in addition connects loose ends.

Figure 15: Counterfactual network with rays and corridors



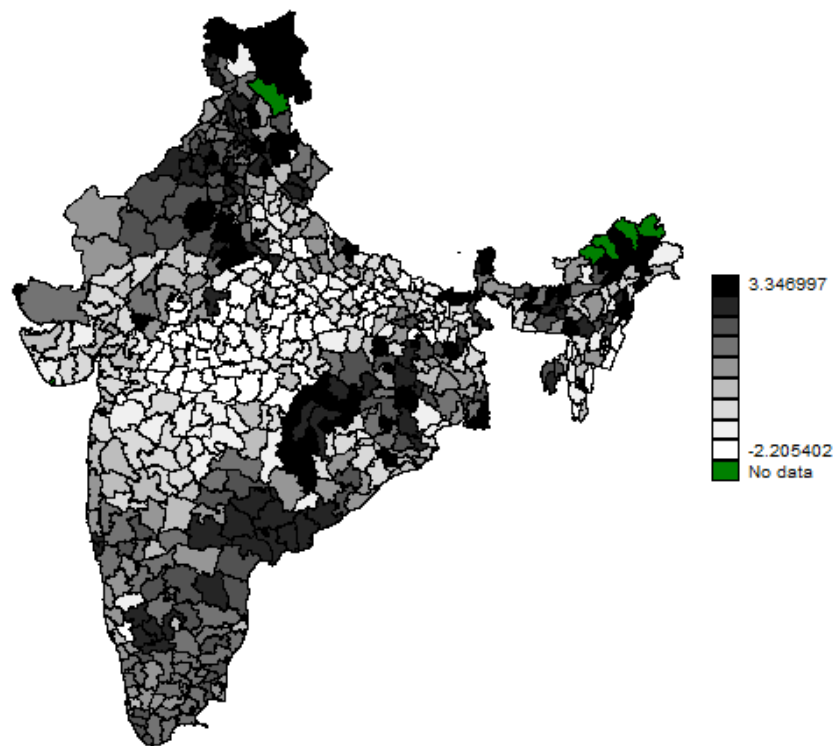
The map shows the counterfactual highways in India which connect all 68 targeted cities with rays, horizontal corridors, and vertical corridors following the Chinese strategy.

Figure 16: Percent increase in income from replacing GQ with counterfactual of rays and corridors



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

Figure 17: Growth of Light per Capita in India



The map show the growth of light per capita in India. Light is measured in the years 2000 and 2009. Population in each district is obtained from the Census in 2001 and 2011. The units are 590 Indian districts. Darker colors refer to higher growth rates. The small green areas in the north and east represent missing observations due to zeros in the initial light per district.

A Model Details

This appendix provides a detailed discussion of the model presented in Section 5. The framework is based on Donaldson and Hornbeck (2013) and Eaton and Kortum (2002).

Donaldson and Hornbeck (2013) 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 590 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 σ 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

$$x_o(j) = \frac{y_o}{P_o} \left(\frac{p(j)}{P} \right)^{-\sigma},$$

where P is the a CES price index of the form

$$P_o = \left(\int_j p_o(j) dj \right)^{\frac{1}{1-\sigma}}.$$

Indirect utility of a consumer who has income y_o and faces a vector of prices \mathbf{P}_o then is

$$V(\mathbf{P}_o, y_o) = \frac{y_o}{P_o}.$$

A.2 Production Technology

Each district produces intermediate goods 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:

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

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 international technology, factor costs, and trade costs.

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

$$P_d = \gamma \Phi_d^{-\frac{1}{\theta}} \tag{A2}$$

with

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

where Γ is the Gamma function. The rental rate for capital is equalized everywhere to $r_o = r$ because capital is perfectly mobile. Donaldson and Hornbeck (2013) then define

$$\kappa_1 = \gamma r^{-\frac{1}{1-\sigma}}$$

and rearrange Equation (A2) to

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

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

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

Noting that aggregate expenditures must equal aggregate income ($X_d = Y_d$), this can be rearranged to

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

This is a gravity equation where the amount of trade from o to d depends positively on the origin's competitiveness and the destination's GDP, but negatively on the competition in the destination's market and on the trade costs. 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

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

A.5 Consumer market access and firm market access

Donaldson and Hornbeck (2013) rearrange Equation (A4) to

$$X_{od} = \kappa_1 T_o (q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma})^{-\theta} CMA_d^{-1} Y_d \tau_{od}^{-\theta}.$$

This 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 and in consumer market access of the destination. Summing the gravity equation over destinations d yields total income of origin o ,

$$\begin{aligned} Y_o = \sum_d X_{od} &= \kappa_1 T_o (q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma})^{-\theta} \sum_d [\tau_{od}^{-\theta} CMA_d^{-1} Y_d] \\ &= \kappa_1 T_o (q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma})^{-\theta} FMA_o \end{aligned} \quad (\text{A5})$$

where Donaldson and Hornbeck define “firm market access” of district o as

$$FMA_o \equiv \sum_d \tau_{od}^{-\theta} CMA_d^{-1} Y_d. \quad (\text{A6})$$

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 (A5), one can obtain

$$\frac{Y_o}{\kappa_1 FMA_o} = T_o (q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma})^{-\theta}$$

which can be substituted into the definition of CMA_d to obtain

$$\begin{aligned} CMA_d &= \kappa_1 \sum_o T_o (q_o^\alpha w_o^\gamma r_o^{1-\alpha-\gamma})^{-\theta} \tau_{od}^{-\theta} \\ &= \sum_o \tau_{od}^{-\theta} FMA_o^{-1} Y_o \\ CMA_o &= \sum_d \tau_{od}^{-\theta} FMA_d^{-1} Y_d. \end{aligned} \quad (\text{A7})$$

Following Donaldson and Hornbeck (2013), if trade costs are symmetric, then a solution to the Equations (A6) and (A7) must satisfy $CMA_o = FMA_o$ and they

refer to this term as “market access”, such that

$$MA_o = \sum_d \tau_{od}^{-\theta} MA_d^{-1} Y_d. \quad (\text{A8})$$

This system of non-linear equation captures the general equilibrium effects of the bilateral trade costs τ_{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 (2013) by incorporating light as a measure for real income. The starting point is Equation (A8). 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

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

Using the formula for the price index,

$$P_d = MA_d^{-\frac{1}{\theta}},$$

I obtain

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

A.7 Income and Market Access with Immobile Labor

Donaldson and Hornbeck (2013) proceed to solve Equation (A5) for land prices. I instead solve for real income, which in the empirical analysis I can approximate with luminosity. Using that result that firm market access equals consumer mar-

ket access, this yields

$$Y_o = \kappa_1 T_o (q_o^\alpha w_o^\gamma)^{-\theta} M A_o \quad (\text{A9})$$

where income is a function of productivity, factor prices, and market access. Note that the constant κ 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 (A9) and solving for income yields

$$Y_o = (\kappa_1 T_o)^{\frac{1}{1+\theta\alpha+\theta\gamma}} \left(\frac{\alpha}{L_o}\right)^{\frac{-\theta\alpha}{1+\theta\alpha+\theta\gamma}} \left(\frac{\gamma}{H_o}\right)^{\frac{-\theta\gamma}{1+\theta\alpha+\theta\gamma}} (M A_o)^{\frac{1}{1+\theta\alpha+\theta\gamma}} \quad (\text{A10})$$

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

$$P_o = M A_o^{-\frac{1}{\theta}},$$

to obtain

$$\frac{Y_o}{P_o} = (\kappa_1 T_o)^{\frac{1}{1+\theta\alpha+\theta\gamma}} \left(\frac{\alpha}{L_o}\right)^{\frac{-\theta\alpha}{1+\theta\alpha+\theta\gamma}} \left(\frac{\gamma}{H_o}\right)^{\frac{-\theta\gamma}{1+\theta\alpha+\theta\gamma}} (M A_o)^{\frac{1+\theta(1+\alpha+\gamma)}{(1+\theta\alpha+\theta\gamma)\theta}}$$

After taking logs, the determinants of real income can be grouped into

$$\begin{aligned} \ln(Y_o^r) &= \underbrace{\frac{1}{1+\theta\alpha+\theta\gamma} \ln(\kappa_1) - \frac{\theta\alpha}{1+\theta\alpha+\theta\gamma} \ln\left(\frac{\alpha}{L_o}\right) - \frac{\theta\gamma}{1+\theta\alpha+\theta\gamma} \ln\left(\frac{\gamma}{H_o}\right)}_{\text{Constant}} \\ &+ \underbrace{\frac{1}{1+\theta\alpha+\theta\gamma} \ln(T_o)}_{\text{Productivity}} + \underbrace{\frac{1+\theta(1+\alpha+\gamma)}{(1+\theta\alpha+\theta\gamma)\theta} \ln(M A_o)}_{\text{Market access}}. \end{aligned} \quad (\text{A11})$$

This equation suggests a log-linear relationship between real income and trans-

port 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.