# The Road Oft Taken: The Route to Spatial Development

Gaurav Khanna *
University of Michigan

August 2014

**Abstract**

The relationship between transportation networks and spatial development has been important for the understanding of the geographic spread of economic activity. And how economic activity spreads within a country depends on the extent of spatial spillovers across regions. This paper studies the impact of being close to transit networks in India, by looking at regions near straight-line paths that connect the four nodal cities. Being connected to these cities leads to greater economic activity till the early 1990s, but this relationship gradually dissipates over the next two decades. The two decades were a period of high growth and development, and more importantly, an upgrading of the transit networks. One possible explanation is that upgrading the transit networks actually allowed regions that were further away to develop, suggesting that there may be positive regional externalities for policies that invest in infrastructure projects. Another is that while economic activity was first concentrated on these transit networks, over time it has been spreading away from well-connected regions to their neighbors that are farther away. The results show that despite there initially being significant spatial inequalities, there has been rapid convergence across regions in the 1990s. This rate of convergence has been faster than usual rates found in the cross-country literature because of neighborhood spillovers in economic activity. The paper tries to identify crucial parameters that determine spatial spillovers and rates of convergence that depend on connectivity to these transit networks.

JEL: O10, O18, R40

Keywords: Roads, highways, transportation networks, spatial development, economic geography, India, Golden Quadrilateral, convergence, regional spillovers

*Department of Economics. 611 Tappan Street, Ann Arbor, MI 48109. email: gkhanna@umich.edu. I would like to thank Raj Arunachalam, John Bound, Charlie Brown, David Lam and Dean Yang for insightful advice. And Austin Davis, Gretchen Lay, Katie Lim, Ben Thompson and seminar participants at the University of Michigan for helpful comments.

# 1 Introduction

One of the reasons why some regions within a country are rich while others are poor is the stark difference in access to infrastructure. This is especially true in developing countries, where large regional disparities in income levels can arise as a result of infrastructure differences. This paper focuses on how access to transit networks can affect the level of income and growth across regions. Intra-regional growth is often augmented by spillovers across regions that generally do not exist across countries. Such positive externalities are particularly significant in facilitating a faster rate of convergence in incomes across disparate regions. This paper studies the impact of transit networks connecting the four major Indian cities, and the pattern of regional development produced due to access to these networks and the corresponding spatial spillovers in economic activity.

Better infrastructure is generally thought to aid development by reducing the costs of trade and migration, equalizing prices and facilitating the spread of ideas and technology. However, causal impacts of roads are hard to find because of a few reasons. First, more developed regions have funds to build better roads and upgrade their infrastructure. Second, roads may be built to connect regions that were already developed or expected to develop in the near future. It is easy to see that richer regions have better roads, but this is not only because roads may lead to more development, but also because these regions have the capability to build better roads and many of the roads were built in regions that were starting out on a path to prosperity. On the other hand, roads may have been built to help struggling regions recover, and are more likely to be built in regions with suitable terrain and easier land acquisition policies. Therefore, to understand the causal impacts of investments in transportation infrastructure it is important to get around these issues, else we would be crediting roads with differences in development unrelated to road-construction.

The focus of this paper is to determine the impact of access to transportation infrastructure on long-term regional development in India, and especially for the two decades starting in 1992. This was a period of rapid growth, economic liberalization, and more importantly, a period of upgrading the highways that connected major cities. In order to look for causal estimates, I use straight-line paths between nodal cities to identify the existence of transit networks. I combine this with a multi-period differences-in-differences specification to find that being connected to a transportation networks causes a region to be more developed in the long run, but this relationship dissipates over time.

Consistent with the dissipation in income-differentials for regions along the road, I find that while significant spatial inequalities existed at the beginning of the period, there was a rapid convergence across regions in the 1990s. One possibility is that economic activity spreads from these transit networks to neighboring regions allowing them to catch-up. The paper then attempts to estimate the parameters that determine these spatial spillovers and how they affect the rates of convergence in economic activity. The presence of spatial externalities have significant implications for infrastructure policies, and an understanding of regional development within a country.

By looking at the cross-section, I find that distance to transit networks do have an impact on long-term regional development, where districts closer to the transit networks have higher levels of economic activity. Furthermore, regions closer to the highway display higher levels of within-region spatial inequality, suggesting that the pattern of development is one where there exists a few active towns in the midst of low-activity areas. This pattern is consistent with a theory of agglomeration economies, where activity concentrates within some pockets of a region but is absent in other areas.

Changes over time show that upgrading the already existing highway system is actually coincident with the period when economic activity shifts further away from the network. This shift away from the highway network could be because of eventual geographic dispersion - while regions near the network are first to develop, they help spur growth in neighboring districts by expanding the network. This would indicate that the growth trends for regions near and far away from the highway are different. A different hypothesis is that the highway actually 'causes' the shift away from regions directly connected to the highway, by lowering transportation costs and allowing economic activity to shift further away. A better transportation infrastructure, then would allow economic activity to be located in areas that had less activity (and plausibly lower wages and rent). This would then indicate that building transit networks not only affects regions that are directly connected, but also has impacts on regions that are further away.

The extent of these spillovers becomes more evident when we see that neighboring regions that are closer to the highway have stronger impacts on a region, that neighboring regions farther away from the highway. Furthermore, there is a high level of $\beta-$convergence across regions, largely driven by these spatial spillovers from neighboring regions. Documenting such spillovers can be valuable for policy analysis, because positive externalities across regions suggest that the benefits of infrastructure projects can be larger than previously thought. This paper will argue that while the early construction of transit networks lead to regional development and significant spillovers to other regions, there were diminishing returns to continue investing in and upgrading these highways. It would then be a better policy to connect other regions with transit networks, rather than upgrade already existing networks.

The road-map of this paper is as follows: the next few sections discuss the literature and the background motivating the identification strategy. Before discussing the results, there is a discussion of the data and the empirical methods used. Then a few models are proposed to explain the results, before concluding. In the models section, the paper explores various mechanisms that may cause such patterns, and highlight the extent of neighborhood externalities and the impact it has on convergence across regions.

## 2 Literature

Two papers in the Chinese context are crucial to this discussion. In a large contribution to the literature, Banerjee, Duflo and Qian (2004, 2012) use 'straight-lines' to connect historical cities in China, and are the first to use this to predict the existence of transportation networks. They find that this can explain moderate differences in GDP per-capita, but has no effect on the growth in GDP per-capita. While they find it difficult to distinguish between transportation networks (roads, rails and other forms of connectivity), a recent paper by Faber (2014) uses the construction of the Chinese National Highway system and combines it with a spatial instrument based on the "least-cost" path of connectivity between between capital cities. Similarly, I connect historical cities in India, which were selected as nodal cities for a large highway-upgrading project, and study the impact on indicators that are closely related to economic activity, like the amount of night-time lights emitted (luminosity) captured by satellites in this period.

Unlike other papers, this paper will focus on the mechanisms that determine spatial inequalities in regional development, and how the spillovers across neighboring regions lead to faster rates of convergence. One crucial difference between the Chinese and Indian cases is the mobility of factors - while migration was highly regulated for many decades in the Chinese context,[1] there is free labor mobility in India, which would be especially useful in an analysis of spillovers across neighboring regions. In so far as access to transportation will have large impacts via migration, the Indian context would allow for studying these effects and be more relevant to contexts that do not have migration restrictions. While mobility in India is low (Munshi and Rosenzweig 2009), this may be due to the high costs of migration for regions that are not well connected to transportation networks. The second big difference is the data used - similar to the Chinese case, the Indian data on GDP at a sub-regional level is poor and problematic. Changes in data-collection methodology over time and across regions may be well co-related with regional development and access. Banerjee et al. (2012) also highlight other issues with the Chinese context - that a non-random sample of regions report GDP numbers, and which years those regions chose to report is also endogenous. It may therefore be better to use data collected from an external source - like the night-time lights data used by this paper. Lastly, unlike the Banerjee et al. (2012) paper, I conduct a before-after analysis to look at the impact of a large National Highway construction project that connected the four major cities and formed a region called the Golden Quadrilateral.

Starting in 1999, the Golden Quadrilateral (GQ) project upgraded about 5,846km of already existing highways in India. In an interesting contemporaneous paper Ghani, Goswami and Kerr (2013) look at manufacturing firms and find an increase in entry-rates for firms

[1] From 1958 to 1978 it was restricted, and then some reforms were put into place to loosen but regulate mobility till the late 1990s

within 10km of the highway, and modest impacts on other indicators. They are careful to point out that OLS results may be affected by the fact that the route of the highway may have been chosen to connect regions that were (a) expected to develop or attract businesses, or (b) were struggling and needed investments to turn them around, or (c) had other systematic differences like lesser land acquisition issues (and therefore less agricultural regions), etc. This paper is, therefore, complementary to the Ghani et al (2013) work, while their paper does an in-depth analysis of manufacturing firms, this paper tries to get at overall economic activity by focusing on night-time lights, and finds somewhat different results. Furthermore, since the highway project updated an already existing network of highways, unlike the Ghani et al (2013) paper, this paper focuses on the long-run economic impacts of being connected to these historical transportation networks, and the eventual dispersion and dissipation of these impacts, especially after the highways are upgraded. Lastly, unlike the other papers in the literature, this paper will study and quantify the extent of neighborhood spillovers and show how they affect the rates of convergence in incomes across regions by formalizing regional development in a growth-model framework.

By combining the empirical strategies used by Banerjee et al. (2012), Faber (2014) and Ghani et al. (2013) one can study causal impact of being connected to a better transport network, and an upgrading of the highway system. These empirical strategies, are not restricted to the developing country context however - papers in the US context (Atack et al (2009), Michaels (2007)) use similar methodologies to look at urbanization, population movements and the demand for skill across different regions. In the Indian context, there has also been other work using other identification strategies, like using the historical expansion of railroads (Donaldson 2010) to look at price equalization and regional development, or more recent rural road construction programs (Asher and Novosad 2014) to look at village employment. While the literature on neighborhood spillovers in economic activity is scant, there is a growing literature on Solow-style convergence within countries (for an analysis of US counties see Higgins et al (2006)) which this paper also seeks to contribute to.

# 3 Background

The paper uses an empirical strategy that first connects the four nodal cities in India with straight lines, and then calculates the distance from each sub-district to the closest straight line. It also uses actual distance from rail and road networks to see how well these straight-lines predict access to transportation networks. Then it compares the impacts of distance before and after a major infrastructure upgrade of the highway system. Lastly, it studies the mechanisms that produce such patterns and quantifies the extent of such neighborhood spillovers and their impact on regional convergence in incomes.

The highway system studied here connects the four nodal cities forming what is called the Golden Quadrilateral (GQ). Three of the four cities (Mumbai, Kolkata and Chennai) were chosen to be capitals of the British Presidencies because they were natural harbors and therefore could be used as ports for trade. There was little economic activity in these three regions prior to the British, and were therefore not on any pre-existing road network. The fourth (Delhi) was a major historical capital of various pre-Colonial empires, and was a British cantonment during the Raj.[2]

The period under focus in this analysis begins in 1992 and goes on till 2012. While the decades leading up to this period was burdened with sluggish growth, these two decades were a time of high and rapid development following the reforms of 1991 that came under a proclamation of 'Liberalization, Privatization, Globalization.' The reforms opened up major sectors of the economy to foreign trade and eventually some sectors to foreign investment, privatized many industries and cut down what is well known as the 'license-permit raj.' Halfway through this period, starting in 1998, the Government proposed an upgrade of the highway system to connect the four major nodal cities under the National Highways

[2]Job Chanock, a member of the British East India Company arrived near modern day Kolkata in 1690, and the British established Fort William in 1698, which gave rise to the modern Kolkata. A few decades before that, in 1639, the British had set up Fort St. George which grew into modern day Chennai. While on the other side of the peninsula, Francisco de Almeida, a Portuguese explorer, sailed into the deep natural harbor of the Mumbai islands in 1508, and the Portuguese acquired the islands in 1534. In 1661 the islands were given to the British as part of the dowry for Catherine of Braganza's wedding to Charles II. These three cities were then chosen to be major British capitals for their natural harbors and strategic positioning of coastal forts. Delhi, on the other hand, had a rich pre-Colonial history, and it wasn't till 1803 when it passed over to British hands. From 1858 to 1911, Kolkata was the capital of the British empire, but this was shifted to Delhi in 1911.

Development Project (NHDP), making it the fifth-longest highway in the world.

The NHDP invested about US $71 billion in order to widen the national highways, and strengthen them for heavy traffic and truck transportation. While the proposal was approved in 1998, many projects started only as late as 2001. Most of the delays had to do with issues of land acquisition, which makes the placement of the final roads endogenous, prompting the use of the 'straight-lines' between the nodal cities.[3] The bulk of the projects were over by the end of 2006, but some alterations on additional phases of the project continue even as late as 2014. For the purposes of this analysis the 1999 to 2006 period will be considered to be the 'construction stage,' but all results will be showing effects over time and are thus flexible to what this stage actually should be. It is, however, crucial to remember that a highway network already existed between these four cities, and the project merely upgraded the system by strengthening and widening the roads rather than building new routes.

## 4 Data

The primary dependent variable of interest is night-time lights as measured by satellite imagery. This has been used as an indicator for economic development, especially in developing countries that have issues with disaggregated income data (Henderson, Storeygard and Weil (2012), Michalopolous and Papaioannou (2013)). Researchers at the National Oceanic and Atmospheric Administrations (NOAA) National Geophysical Data Center (NGDC) process data from weather satellites that circle the Earth 14 times a day and take pictures between 2030 and 2200 hours at night. They use algorithms to filter out other sources of natural light using information about the lunar cycles, sunset times and the northern lights, and other occurrences like forest fires and cloud cover. While Figure 1a shows an example of the night-time lights image along with the straight-lines between the four nodal cities, Figure 1b zooms in on the region connecting Delhi-Mumbai and Delhi-Kolkata where a stream of

[3]The junior Highways Minister, Tushar Chaudhary told the Parliament that "Projects have been delayed mainly due to problems associated with land acquisition, shifting of utilities, obtaining environment and forest clearance, approval for road over bridges, poor performance of some contractors due to cash flow constraints and law and order problems in some states"

lights is associated with the National Highway that connects the nodal cities. The Golden Quadrilateral Highways and the sub-districts used in the analysis are shown in Figure 2. Distance to the nearest straight line is calculated using standard ArcGIS software, and in some regressions I also use data on actual highways and roads that are obtained from the Digital Chart of the World (DCW) database.

The lights data is calculated at approximately every one square kilometer, but I aggregate the results to the sub-district level in order to account for issues of spatial correlation.[4] I regress this lights data on distance to the nearest straight line connecting the nodal cities. [5] Histograms and kernel density plots of the primary variables of interest are presented in Figure 3.

In all calculations, I drop the nodal cities and 26 neighboring sub-districts so as to not capture the impact of being a neighbor - doing so only slightly attenuates the results.[6] In all specifications, I also control for other geographic features like distance to closest nodal city, coastline, latitude and longitude. While the NHDP upgraded the highways to make them stronger and wider, there was a road and rail network that existed prior to the project. Therefore, even in the period before 1999, cities on the straight line would be well connected to the nodal cities and the transit network of the country. The identification strategy can be broken up into a couple of stages: first, just examining the long-run impact of 'connectivity' till the 1990s, and second, looking at how this impact changes over time. The second approach helps to isolate if there are any direct impacts of upgrading the highway system from the general effects of connectivity to transportation networks. After which, I set up certain models in the Solow-framework that help determine the parameters of convergence and the

[4]Sub-districts are third largest administrative unit of aggregation with a population of about 460,000 people on average. Results will be presented with standard errors clustered at even higher levels of aggregation. There are about 2253 sub districts in 594 districts which are in 35 states and union territories. Results are statistically significant even at standard errors clustered at the 35 states level.

[5]Following the conventions established in the literature (Michalopolous and Papaioannou (2013)), the lights data is transformed to be of the form: $Log(0.01 + Luminosity)$ to account for the fact that some areas have no luminosity. About 1.6% of the total sample, and less than 1% of the last 3 years of the sample have sub-districts that had no luminosity. Furthermore, the results are robust and significant in using Poisson regression specifications of luminosity. I also present results of $Log(Luminosity)$ for the regions that never have 0 recorded lights - which can be interpreted as an impact on purely the intensive margin.

[6]These include Delhi and surrounding areas like Ghaziabad and Gurgaon, Mumbai and greater-Mumbai, Kolkata and Haora, and Chennai and it's neighbors.

extent to spatial spillovers across regions.

### 4.1 The Elasticity between Light Density and Domestic Product

In order to get at spatial development at a finer level, this paper studies the impacts at the administrative level of the 2253 sub-districts. Gross Domestic Product (GDP) is not calculated at this level, but there are GDP numbers available for about half the 594 districts in the country, and for all the 32 states. The elasticity between State Domestic Product (SDP) and light-density will be an underestimate of the true elasticity between sub-district domestic product and light-density because of the measurement error introduced in aggregating the 2253 sub-districts into 32 states. The relationship between District Domestic Product (DDP) and light-density, while suffering from some measurement error as well, will however bring us closer to the true parameter. Appendix Tables XV and XVI show the state-level and district-level elasticities respectively. While the state-level elasticities are a little below 0.2, the district-level elasticities are a little above 0.3 in general, and there is no trend over time in the elasticities. Due to measurement error in aggregating lights and domestic product to a higher administrative unit, the true sub-district level elasticities should be higher. The cross-country literature (Henderson, Storeygard and Weil (2012), Michalopolous and Papaioannou (2013)) has elasticities of about 0.3 for sub-samples of low and middle income countries, suggesting that 0.3 would be a reasonable lower bound for the sub-district level elasticities.

## 5 Connectivity to Transit Networks

### 5.1 Identification Strategy

While the actual path of roads is endogenous to regional characteristics, being on a straight-line between two major cities should not be correlated with anything other than being close to transit networks connecting them. In order to examine the long-term general impact of connectivity to historically determined transportation networks, one can look at the impact on lights for cities closer to the straight-lines using the following regression specification, for sub-district $i$:

$$LogLights_i = \alpha_i + \beta\mathbf{X} + \beta_d Distance_i + \epsilon_i \tag{1}$$

Equation 1 is the "reduced form" specification where $Distance_i$ is the distance[7] between the sub-district and the nearest straight line, and $\mathbf{X}$ are geographic controls (distance to nearest nodal city, coastline, latitude and longitude). Similarly the OLS formulation of this same equation would replace the $Distance_i$ variable with distance to the nearest highway, or distance to the nearest rail-line. The OLS estimates, however, will be biased because the highways and rail-lines will be laid according to where cities and economic centers are located, or lagging regions where the government wishes to induce economic activity. It is interesting, however, to study the direction of the OLS bias. If road and rail-lines are laid to be closer to economic centers, then the OLS estimates will show large impacts of being close to a highway or rail-line. If, however, land acquisition for construction of lines and roads forces the government to move away from economic centers, then the impact of distance to these lines would be attenuated.

Finally, one can derive the upper-bound of the effect of roads and rail lines by performing a two-staged least squares exercise of the following form:

$$\begin{aligned} DistanceToRoad_i &= \pi_d Distance_i + \mu_i \\ LogLights_i &= \alpha_i + \delta_i + \beta\mathbf{X} + \beta_d \widehat{DistanceToRoad}_i + \epsilon_i \end{aligned} \tag{2}$$

Since the impact of the distance to the straight-lines will work through both rail lines and roads, the coefficient $\beta_d$ is strictly not the "instrumental variables" estimator. Later, the paper will seek to isolate the impact of upgrading the highway system by comparing districts close to and far away from the straight-lines, before and after the highways were built.

Table I shows the relationship between distance to the straight-lines and distance to

[7] Some results are presented as distance in kms to be comparable to the Ghani et al (2013) paper, and other results are presented as Log(distance) to calculate the elasticity as in the Banerjee et al (2007) paper.

transit networks. While the distance to the straight-line is a good predictor of distance to the nearest GQ highway, it only does moderately well in predicting distance to closest rail line. This is hardly surprising, since while the GQ highways were built in order to connect the nodal cities, the rail lines were built to connect other cities as well.[8] Throughout the paper, the results will be clustered at higher-level administrative units like districts or states in order to account for possibilities in spatial correlation.

## 5.2 Results: Distance to GQ highway and railways

The OLS relationship between lights and proximity to the nearest GQ highway is shown in Table II, for the beginning and the end of the sample period (i.e. the years 1992 and 2012). The most commonly used metric for luminosity as a predictor of development is the 'Lights per area' or light-density variable (Henderson, Storeygard and Weil (2012), Michalopolous and Papaioannou (2013)). In 1992, this relationship had a coefficient of -3.913. Since the distance variables are in 1000km, this means that a 100 km increase in distance from the highway was associated with a fall in light-density of 0.3913 log points. As discussed previously, a reasonable lower-bound for the elasticity between sub-district domestic product and lights is 0.3. Furthermore, in the literature, a 1 log point increase in light density is usually related to the a 0.3 log point increase in income for the sample of low and middle income countries (Henderson, Storeygard and Weil (2012), Michalopolous and Papaioannou (2013) appendix). Therefore, a 100 km increase in distance from the highway would be correlated with a 12% fall in income. By the year 2012, this had halved to about a 6% difference in income (again assuming that the elasticity between light-density and income is 0.3). The variables are statistically significant even at higher levels of clustering standard errors. In order to look at the extensive margin, the last column in the table is a linear probability model (LPM) of the probability of having the majority of recorded light-emission pixels in a sub-district be greater than 0. This relationship is greater in 1992, again suggesting that over the two decades the relationship between distance to the highway and development has

[8] Interestingly enough, the first set of railway lines laid in India (in the 1860s) were built by Lord Dalhousie to connect the five major provincial capitals - four of which are the current nodal cities

weakened.

This OLS relationship, however, could be biased as the exact path of the highway will depend on the government's wish to connect some areas, and the ease with which land could be acquired for construction. Table III shows the reduced-form relationship between lights and the distance to the straight-lines that connect the nodal cities. Once again, the relationship is much larger in magnitudes in 1992 than in 2012, and of similar size as the OLS relationship.

The two-staged least squares (2SLS) estimates are presented in Table IV. The excluded distance-to-line variable has an extremely high F-stat no matter what the level of clustering, displaying a strong first-stage relationship. The 2SLS estimates are slightly more negative than the OLS estimates in some cases - for example, in 1992, the 2SLS results say that a 100km increase in distance from the highway leads to a 0.5 log point fall in light density. Assuming the same elasticity between lights and income, this is a 0.15 log point or about a 16% fall in income. The 2012 light-density coefficient, however, is identical to the OLS result.

Tables V and VI show the analogous OLS and 2SLS results for distance to the nearest rail-way line. Keeping in mind that distance to the straight line is only a moderately good predictor of distance to the rail line, we can see that the 2SLS results are much larger than the OLS results. The reduced form is in Table III, and the 2SLS results are magnified by the fact that the excluded distance-to-line variable is not a very good predictor of the distance to nearest railway line.[9]

In Table VII, I estimate the elasticity between distance and economic activity over time. By using a log-log specification in each year, I calculate how this elasticity between lights and distance to the nearest straight line is changing between 1992 and 2012. Assuming that a lower bound for the elasticity between lights and GDP per capita is constant at 0.3 (as shown previously), the results indicate that the elasticity between GDP per capita and distance is falling over this period from about 0.15 to 0.06. The range subsumes the Banerjee et al. (2012) elasticity of 0.07, but for much of the period is higher. The results therefore indicate that while in 1992 the elasticity was high, it more than halves to fall to an economically

[9]The distance-to-line is a "weak instrument" for rail-lines, and the bulk of the impact is coming via roads.

insignificant relationship by the end of the period. This dissipation of the impacts hasn't been investigated much in the literature, and is studied below in more detail, where large spatial spillovers can lead to a dissipation in the differential increases in income for the regions along the road.

# 6 Upgrading the highways

The NHDP projects were first finalized in 1998, and the foundation stone was laid by the Prime Minister on January 6, 1999. The first couple of years, however, were plagued with delays in certain areas - because of contractual issues and problems with land acquisition. About 20% of the projects started between 1998 and 2000, whereas almost 50% of projects started in 2001.[10] While Phase I of this project officially ended in 2006, about 8% of the projects ended a few years later. Later Phases added some additional upgrades, and work continued on the GQ till the end of 2011. This timing allows for a before-after analysis of this highway construction, since the lights data spans from 1992 to 2012. The period 1999 to 2006 in the sample will be considered to be the 'construction' phase, while the years after that will be the post-project phase.

In order to see how the impact of distance changes with time, one can run the following regression:

$$LogLights_{it} = \alpha+\tau_t+\beta\mathbf{X}+\delta_{1i}Distance_i+\delta_{2i}Distance_i*Construction_t+\delta_{3i}Distance_i*Post_t+\epsilon_{it} \tag{3}$$

where $\delta_{1i}$ is the impact of distance on lights in the pre- construction period, $\delta_{1i}+\delta_{2i}$ is the impact in the construction period, and $\delta_{1i}+\delta_{3i}$ is the effect in the post-construction period:

Table X shows the impact of distance over these three periods. In the pre-construction period, light density would fall by 0.3389 log points for every increase in 100km from the straight line, but once construction starts, this falls to about 0.2434 log points, and in the post-construction period it's even lower at about 0.2 log points. Except for the extensive mar-

[10]Source: National Highways Authority of India http://www.nhai.org/completed.asp

gin metric (the probability of the majority of lights being positive) in the construction phase period, all other results are statistically significant, and show that the impact of distance on development falls around the turn of the century, after the highway construction begins. That does not necessarily indicate that the highway construction caused the relationship to dissipate, as this relationship may have already been weakening over time.

To see at what distances the change in impact appears, I split up all positive distances into 8 equal quantiles $\Psi_i$, and interact them with indicators for being in the 'construction' phase or the 'post-project' phase. In the regression equation below $\tau_t$ are year fixed effects and $\mathbf{X}$ is a vector of geographic controls:

$$LogLights_{it} = \alpha + \tau_t + \beta\mathbf{X} + \psi_{\mathbf{1i}}\mathbf{\Psi_i} + \psi_{\mathbf{2i}}\mathbf{\Psi_i} * Construction_t + \psi_{\mathbf{3i}}\mathbf{\Psi_i} * Post_t + \epsilon_{it} \tag{4}$$

The omitted category in this regression are the sub-districts that are on the straight-line (a little more than 5% of all sub-districts). $\psi_{\mathbf{1i}}$ traces out the impact of distance from these sub-districts in the pre-construction period, whereas $\psi_{\mathbf{1i}} + \psi_{\mathbf{2i}}$ is the impact during the construction phase. These coefficients can be plotted for each distance quintile to look at the the semi-parametric impact of distance, and how that changes in the three time periods.

The lines in Figure 10 show the impact on lights by distance quintiles, relative to sub-districts that touch the straight-lines. The blue lines are for the pre-construction period, the orange for the construction period, and the green lines are for when the project is over. Looking at the pre-construction period in panel (b) we can see that a district in the eighth distance quintile has about 2 log points less light density than a district that is on the line. But once construction begins, these lines start flattening out. Together these results seem to suggest that while an increase in distance from the straight line leads to less development, this relationship weakens in the later period, and especially after the construction of the highway.

The panel on the standard-deviation of lights in Figure 10 shows that there is a larger dispersion of lights within a region that is closer to the road. This gives us some indication towards the pattern of development in these regions - that in sub-districts near the road there are a few large towns with a lot of activity, and then areas with very little activity. In regions away from the road however, there is an equal amount of low economic activity. This pattern is consistent with developed regions reflecting agglomeration economies, where activity is concentrated in certain areas but is sparse in other regions (Krugman 1991).

While the time periods in the above analysis was split into three phases, it is also important to see when the change in impacts started setting it. In order to do so, I perform the following analysis for sub-district $i$ in year $t$:

$$LogLights_{it} = \alpha + \delta_i + \tau_t + \beta\mathbf{X} + \gamma_{\mathbf{t}} Distance_i * \tau_t + \epsilon_{it} \tag{5}$$

The regression includes year fixed-effects $\tau_t$, region fixed effects $\delta_i$ and the usual geographic controls $\mathbf{X}$, similar to a multi-period difference-in-differences specification. The coefficient of interest $\gamma_{\mathbf{t}}$ is on the interaction term between $Distance_i$ and $\tau_t$. In the regressions, the omitted year is the first year of the sample - 1992. One can plot this coefficient $\gamma_{\mathbf{t}}$ to look at how the impact of distance to the line is changing over this period, relative to the year 1992.

Figures 6 shows how the effect of distance on lights changes relative to 1992. Positive values of $\gamma_{\mathbf{t}}$ indicate that the impact of distance on lights is falling relative to 1992. In the figures it is clear that this differential impact is indeed falling over time, and especially after the construction begins. It is however, not possible to reject the possibility that the dissipation of the relationship between distance and development would not have happened if the highways were not upgraded. The panel on the standard-deviation of lights shows that even though inequality within a sub-district was higher for regions closer to the roads, this difference has been shrinking over time.

Robustness to different specifications is shown in Appendix Figures 11 to 13. Each figure has four panels allowing for two-different levels of clustering errors and doing robustness checks with dropping any sub-district that ever had 0 recorded luminosity, and dropping any subdistrict that ever had any pixel with the maximum possible luminosity value. Appendix Figures 10 reproduces the main results after excluding regions that are a significant distance away from the road.[11]

While Table VII shows the 'reduced form' elasticities between light and distance to the straight-line, in Figure 7 I show the OLS and 2SLS elasticities over time. While the elasticities are similar towards the end of the period, there are stark differences in the beginning of the period. One explanation for this convergence is that in the early 1990s, the 'distance to line' could be picking up other transit networks as well in the 2SLS regression, but by the 2000s the highways seem to become the dominant channel. And this shift in importance of which transportation networks are used may be due to the GQ upgrades.

Ghani et al. (2013), show that upgrading the highways induced new manufacturing firms to enter in regions close to the highway. The difference in results are not due to the methodology,[12] and must therefore be because of different outcome variables under analysis. Their paper looks at the organized manufacturing sector, for about half the districts in the country[13] and shows that there was an increase in entry for such firms in regions within 10kms of the highway. Together our papers, therefore show that while there was an increase in the entry of organized manufacturing enterprises, overall economic activity was still shifting away from the highways.[14]

This spread in economic activity may have been stemmed if upgrading the highways induced enterprises to stay or enter in regions on the highway at a higher rate than other regions, like it happened for firms in the organized manufacturing sector. There is however, little evidence in this paper to show that upgrading the highway system actually turned around the trends that were already visible in the data for overall economic activity. If anything it may have spurred the spread to regions further away by cutting down costs of transportation.

[11]The excluded regions include the states of Jammu and Kashmir, Sikkim, Assam, Arunachal Pradesh, Meghalaya, Mizoram, Tripura, Nagaland, Andaman and Nicobar Islands and Lakshwadeep.

[12]While Ghani et al (2013) rely on the OLS specification for their results, they show that their main specification is robust to using the "instrumental variables" approach. Unlike this paper, instead of using a continuous "distance" measure, they use two discrete categories - compare districts between 0 and 10km near the road to districts further away with a 1/0 indicator for whether you are within 10kms of the highway. When I use their methodology, I still find that the relationship between night-lights and distance to the road dissipates over time.

[13]for states that had enough manufacturing activity, in districts that were observed over their entire panel

[14]One possibility is the change in the composition of economic activity - for example, service sector firms could have shifted further away from the highway and manufacturing firms closer to the highway.

# 7 Models of Spatial Development

Better access to transportation networks can induce development in connected regions by facilitating trade and migration, the spread of technology and ideas, and reducing price volatility. So far from the results it is clear that till the early 1990s, being near the transit network that connects the four major cities had significant impacts on regional development. However, the impact of being close to the straight lines dissipates over time, especially after the highway system is upgraded. There are a few models of spatial development that would help explain these patterns.

The first is a location choice model, where better connectivity on the highway system allows firms and households to locate further away from the highway. A firm or household enterprise wishes to be connected to the four nodal cities for purposes of trade and exchange. For locations that are not near the nodal cities, the cost of being connected to the city is the sum of the cost of being connected to the highway $c_1(d)$ and the cost of using the highway to get to the city $c_2$, where $c_1(d)$ is an increasing function of the distance between the region and the closest highway $d$.[15] Firms have returns to investment that is drawn from a distribution $R \sim F(.)$ and choose to locate in region $d$ if $R \geq c_1(d) + c_2$. The fraction of firms that locate there are therefore $1 - F(c_1(d) + c_2)$. Since the cost functions are increasing in the distance, this would indicate that more firms will locate closer to the highways. Once the highway is upgraded, this reduces $c_2$, thereby allowing the same firm to locate at a distance $d'$ further away from the highway. After the NHDP upgrades, firms and households can therefore move into regions that would earlier have been too costly for them to locate in.

Another model is that of lowering costs of seasonal migration. Once the highways are upgraded, this allows people from near the highways to migrate to the city at lower costs. If the migration is permanent, there will be a fall in population and economic activity in the region close to the highway. However, a lot of the migration in the Indian context is seasonal in nature, where people work in the cities during the agricultural slack season but on the farms in the peak season. For a city wage $W_c$ that lies between the peak season $w_p$ and slack

[15] $c_2$ can also be made to depend on the distance on the highway from the city, but is irrelevant for this analysis

season $w_s$ wages, after taking into account the cost of migrating $c_2$, a laborer word work in the city during the agricultural slack season and return to his fields in the peak season. If a worker doesn't migrate in the slack season (he lives in a region where $c_2$ is high), then he would engage in non-agricultural economic activity instead, but if he does migrate ($c_2$ is low) then there is no need for non-agricultural enterprises to exist in these regions since in the slack season workers would easily migrate to the city. A lowering of migration costs by upgrading the highways would then facilitate seasonal migration and thereby reduce any non-agricultural enterprises from setting up in rural areas near the highways.

Having said that, the trends over time may not necessarily be due to investments in the highway system. While transit networks may be important for regional development, these impacts may dissipate over time on their own. After the four nodal cities were established, the regions that had the least-cost connections to these cities started developing. This helped build up a network structure whereby regions connected to these growing regions started growing. Since regions connected to the roads started growing earlier, they are at any point of time closer to their steady-state level of development than regions further away which will hence be growing faster. If we then look at long-term development, the first set of regions would have more economic activity, but the their neighbors and their neighbors' neighbors would be catching up over time. This would then produce the trends seen in the data, and is consistent with the results found in Banerjee et al (2009) where they find that Chinese regions near the straight-lines had reached a higher level of GDP per capita but were not necessarily growing faster than other regions.

One way to try and determine whether the GQ upgrades caused these trends is to look at another route that did not have heavy investments in the highway system: that between Mumbai and Kolkata. Table IX shows that the elasticity between economic activity and distance to a straight line connecting these two cities has been dissipating over time, despite the fact that there was no projects under the National Highway Authority of India to upgrade the routes between these two cities. If this is an indication of the trends seen along routes connecting the other nodal cities, then the construction of the highway may not have had much of an impact on the pre-existing trends under which economic activity was already spreading geographically away from the well developed regions connected to the highway. These mechanisms are explored further when studying the geographic spillovers within the confines of a Solow model set-up.

## 7.1 Migration: Changes in Population

The question of what kind of economic activity the night-lights are picking up will provide more information on the kind of development. If the regions further away from the roads were previously uninhabited and the roads allowed people to locate there, then we should see a rise in population for those regions. If, however, it is merely the composition of the population and the kind of economic activity undertaken by them, then for a given population, the increase in night-lights will represent more wealth per capita being generated.

To get at the question of whether there are population changes to less inhabited regions or whether the changes over time are picking up increases in economic activity per capita, I use the LandScan data on population estimates. The data compiled by the US Department of Energy's Oak Ridge National Laboratory, uses sub-national Census counts and primary geospatial ancillary datasets, including land cover, roads, slope, urban areas, village locations, and high resolution imagery analysis of settlements, to predict the populations at a finer geographic level than available elsewhere.

Table VIII presents the elasticity between population and distance to the nearest straight-line connecting the nodal cities over time (starting with the data in 2002). While the elasticity is high (between 0.18 and 0.2) for this period, there is not much change in the elasticity over time. This may be indicative of the fact that the change in the night-lights elasticity over time may be driven by something other than the number of people migrating to newer areas, and instead be due to changes in per-capita economic activity.

## 7.2 The Nature of Development

While light-density may be a good proxy for overall economic activity, it tells us little about what the nature of this economic activity. Unfortunately, the only source of data that provides counts at the sub-district level is the Census of India which has a limited number of outcome variables and is only compiled once every 10 years. Table XII shows results from the 2001 and 2011 Censuses. While regions away further away from the road have less population overall, they are more likely to have a higher concentration of Scheduled Tribe (ST) persons. Along with the Scheduled Castes (SCs) these are among the most socio-economically disadvantaged groups in the country. Furthermore, regions further away also have a higher concentration of cultivators, but lower concentration of persons engaged in household industry. While distant regions also have lower literacy rates [16], they seem to have a more equitable distribution of literacy across genders (i.e. the gender-gap defined as the difference in the male and female literacy rates is lower). Unfortunately, without more detailed data at the sub-district level, it is hard to discern any intricate patterns in the nature of economic activity, but it is clear that regions farther away from the road have a higher concentration of Scheduled Tribe disadvantaged populations, and have more cultivators but less persons engaged in household enterprises.

# 8 Development Through the Lens of Growth Theory

## 8.1 Beta and Sigma Convergence in Light-Density

In the results so far, we see that while regions along the road are richer, there is a catching-up of regions further away. The Solow (1956) model's predictions of a conditional (on parameters) convergence of per-capita incomes may be used as a framework to study these patterns in the data. Barro and Sala-i-Martin (1992) are careful to distinguish between the different notions of convergence. If $\beta-$convergence holds then poorer sub-districts would be growing

[16] literacy rates are not statistically significantly different from 0 for standard errors clustered at higher levels

faster than their richer counterparts (i.e. in the equation below, we would expect $\beta < 0$):

$$Log(y_{it}) - Log(y_{it-1}) = \alpha + \beta Log(y_{it-1}) + \epsilon_{it} \tag{6}$$

In column 1 of Table XI, it can be seen that the $\beta-$convergence parameter is an economically and statistically significant -0.108 or about 10%. This is a lot higher than the cross country literature (about 2%), but is close to estimates found for panels of US-counties of about 6 to 8% (Higgins et al. 2006). The section below discusses how consistent with these studies, within-country $\beta-$convergence can be faster than cross-country convergence because of spillovers across regions within a country. Positive externalities of this nature can have significant implications for policies that consider investments in regions. They indicate that regions should be targeted keeping in mind the connectivity and levels of economic activity in surrounding regions.

$\beta-$convergence, however, is not a sufficient condition for $\sigma-$convergence since income shocks may lead to a higher variance in income overall. $\sigma-$convergence therefore, holds when the variance in incomes is falling over time. Figure 8 shows that between 1992 and 2002 the variance in light density had shrunk a lot, but this did not necessarily continue for the next decade.

## 8.2 Distance to the Road and the Solow Framework

The impact of distance to transit networks on regional development can be looked at using the Solow model. One possibility is that distance to networks affects the steady-state level of development, which would then predict regions closer to the highways would grow to a higher level of income than regions further away. Another possibility is that the distance to transit networks only affects the initial level of income, and all regions converge to a similar steady-state level of economic activity. This would then be consistent with a result that shows that regions further away from the road have higher growth rates. To formalize this framework, we can modify the empirical predictions of the Solow model in the following way:

Let $y_d = Y_d/A_d$ be income per effective worker for a given level of per-capita income $Y_d$

and technology $A_d$ in sub-district $d$. If the distance to transit network affects the steady state level of income, then we can define steady state in sub-district $d$ to be:

$$y_d^* = D^\lambda y^* \tag{7}$$

Where $D$ is the distance to the transit network and $y^*$ is steady-state level of income for a region on the road. This framework is similar to gravity-models in trade theory, where the distance would be picking up trade-costs and other frictions. If on the other hand, the distance affects the initial level of income, then we can define initial per capita income to be:

$$y_{0d} = D^\psi y_0 \tag{8}$$

Where $y_0$ is the initial level of income for a region on the road. Technology grows at a rate $g$, such that $A_{dt} = A_{0d}e^{gt}$, and from the Solow model, we know that regions will grow by closing the gap between their initial level and steady-state levels of income (i.e. $\beta < 0$):

$$Log(y_{dt}) - Log(y_{d0}) = \beta(Log(y_d^*) - Log(y_{d0})) \tag{9}$$

Combining the above equations we can derive an estimatable equation:

$$Log(Y_{dt}) - Log(Y_{d0}) - gt = \beta((\lambda LogD + Log(y^*)) - (\psi LogD + Log(y_0))) \tag{10}$$

$$Log(Y_{dt}) - Log(Y_{d0}) = \beta(Log(y^*) - Log(y_0)) + gt + \beta(\lambda - \psi)LogD \tag{11}$$

Estimating the above equation, and evaluating the parameter on $Log(Distance)$ can help determine whether distance to the networks produce different steady states ($\lambda > \psi$) or different initial levels of income ($\psi > \lambda$). Column 2 of Table XI shows that the coefficient on $LogD$ is positive. Since we know that $\beta < 0$ from the Solow model, this would indicate that $\psi > \lambda$, or that distance from transit networks are more likely to affect the initial levels of per-capita income but less so the steady-states. Furthermore, in Column 3 of the same table, we can see that there is little change in the value of this parameter after the road construction began. An implication of this analysis is that transit networks may influence initial conditions directly, but find it difficult to change the final steady-state value of development. This would then mean diminishing returns to investment in these networks. However, investing early on can help increase incomes sooner rather than later, and have significant spillovers to neighboring regions which would overall increase the returns to such investments.

# 9 Spillovers and the Direction of the Spread of Development

If transit networks affect the initial level of development, then it is possible that economic activity starts spreading from regions along the roads to neighboring regions. In Figure 4 we can follow the regions around the Mumbai-Chennai over time. The pictures show the regions along the highway (depicted by a blue line) and along the straight-line path (red line) every 5 years, in a Green to Red spectrum (where deeper green reflects lesser economic activity, and red indicates more activity). It does seem like the regions to develop first are the ones along the road, after which economic activity seems to fan out to neighboring areas, eventually reaching areas farther away.

If however, economic activity spreads from regions close to highways to regions away from the highway, then a given region ($k$) should be affected more by economic activity in neighboring regions that are closer to the highway ($k-1$) than neighboring regions that are further away from the highway ($k+1$). Let $Log(LightDensity)_{t,k}$ represent the lights in region $k$ at time $t$, $Log(LightDensity)_{t,k-1}$ be the mean of the lights for all it's neighboring regions that are closer to the highway, and $Log(LightDensity)_{t,k+1}$ be the mean of lights for neighboring regions farther away from the highway. In the regression below, we should then expect $\beta > \gamma > 0$ if economic activity is spreading from regions closer to the highway to regions away:

$$Log(LightDensity)_{t,k} = \alpha_k + \beta Log(LightDensity)_{t,k-1} + \gamma Log(LightDensity)_{t,k+1} + \epsilon_{t,k} \tag{12}$$

While this regression formulation shows a contemporaneous impact, the effect could also have a period lag, when activity in a region today can affect economic activity in a neighboring region tomorrow. Furthermore, the true relationship could be one of changes, where changes in a region's economic activity affect changes in it's neighbor's activity. It is important to stress however, that the true test of the model is one where $\beta > \gamma$, and not merely if $\beta > 0$ since $\beta > 0$ can also be consistent with a model of spatial correlation in income shocks. The regressions for contemporaneous spillover effects, as well as the one-period lagged effect and the changes over time specification are shown in Table XIV, where it can be seen that $\beta$ is always statistically significantly greater than $\gamma$.

A stronger test of this model is to see how much of the effect of distance can be explained by these neighborhood spillover effects. To formalize this, let $L_k$ represent light-density in region $k$ and $L_{k-1}$ be the mean light-density for all it's neighbors that are closer to the road than region $k$. For a given distance from the road $D_k$, let the true relationship between light-density and distance be:

$$L_k = \chi L_{k-1} + \mu D_k \tag{13}$$

In this relationship, $\chi$ represents the neighborhood spillover effect of economic activity from bordering regions closer to the highways than the given region, and $\mu$ represents the direct effect of being further away from a transit network. This equation can be solved forward recursively, to:

$$L_k = (\sum_{j=0}^{k} \chi^j)\mu D_k = \alpha D_k \tag{14}$$

Equations 13 and 14 can be used to estimate parameters $\chi$,$\mu$ and $\alpha$, and we can test if $\alpha = (\sum_{j=0}^{k} \chi^j)\mu$. For this, we need to use the fact that on average regions have about five to six degrees of separation between the road and themselves (i.e. $k$ is approximately six). Table XIII shows results for both Equations 13 and 14. The distance-spillover parameter $(\sum_{j=0}^{k} \chi^j)\mu$ according to equation 13 is also presented, and is very similar to $\alpha$ as seen in equation 14. This is a strong test of the model that substantial neighborhood spillovers across regions exist in this context.

How this relationship changes over time tells us about how these spillovers can actually lead to convergence across regions. In the cross-country version of the Solow model, the convergence across regions could be generated without any spillovers, and often leads to rates of $\beta-$convergence of about 2% (Sala-i-Martin 1996). Within a country, however, spillovers across regions can speed up convergence, and generate rates like 6 to 8% for US counties (Higgins et al 2006), or 10% as in the case of this paper.

Studying the pattern of lights in regions along the road, and their neighbors would help answer the question of how spillovers are leading to convergence. In order to study this, let us define a 'degree-of-seperation' ($s$) as how many regions lie between your region and the road. For example, $s = 1$ means the region neighbors a sub-district that lies on the road, and $s = 3$ means that the region is a neighbor of a neighbor of a neighbor of a region that lies on the road. Plotting the coefficients ($\beta_{st}$) of the regression below for each $s$ and over time $t$ allows us to study whether convergence takes place across neighboring regions. In the specification below, $\mathbf{1}_{s=S}$ is an indicator function that depends on the 'degree of separation' of the given region, and $Year_t$ is an indicator variable for the year:

$$Log(lightdensity)_{dst} = \alpha + \beta_{st}\mathbf{1}_{s=S} * Year_t + \epsilon_{dt} \tag{15}$$

In Figure 9 each point is the differential impact of light density on that region compared to other regions in a given year. It can be seen that regions on the straight-line connecting two major cities have the highest light-density compared to other regions, and the regions that are one-degree of separation away are only slightly worse off. The ordering in term of 'degrees of separation' is maintained, whereby regions closer to the road have higher light-density, and even though there is convergence over time, we don't seem to see an 'over-taking' by the regions further away. Lastly, while it seems like the 1990s were a period where convergence was rapid, this has slowed down in the later half of the 2000s. At the end of the period, regions still maintain their initial ordering in concentration of light-density, an ordering which directly depends on their 'degree-of-separation' from the road.

# 10 Conclusion

The impact of transportation infrastructure on regional development has been a long debated discussion. In general, better transit networks have been thought to facilitate trade, migration, the spread of ideas and technology, credit and other financial opportunities, and decrease price differentials and volatility. Studying infrastructure projects in different contexts have however provided contradictory evidence. Fogel's (1964) study of US historical development argues that there were limited impacts of railways on growth relative to the transportation networks that used waterways, whereas Hirschman's (1969) treatise posits that social overhead capital, like railways, have significant linkages that promote growth in industries. For Hirschman, infrastructure projects would have forward linkages (promote industries that need roads and railways), backward linkages (promote industries that supply materials for road and rail construction) and lateral linkages (connect industries together). The Fogelian view, on the other hand, supports the idea that much of US historical investment in railways was misguided and therefore did not have impacts on development because of governmental policies that subsidized railway construction. The natural experiment under analysis in the Indian context, however, is that some regions happened to be on the path of shortest distance connecting major centers of economic activity, and we are hence less likely to find 'misguided' investments in this context.

The results in this paper indicate that while distance to the straight line may have significant impacts on regional development till the early 1990s, this relationship dies out slowly over the next two decades. If one was to analyze the relationship at the end of this period, they would come to the conclusion that there is little economically significant impacts of being near a transportation network, which would support the Fogelian view. However, the relationship in the early 1990s shows how this may not have necessarily been the case. The question then arises, as to what happened in the two decades that weakened this relationship.

The period of study was one of rapid economic growth and development after reforms that liberalized the market structure, cut down on the license-permit bureaucracy and integrated various industries with world markets. It was also a period of upgrading the existing transportation network by strengthening the highway system. One possibility is that these changes allowed economic activity to shift away from the existing network and move further away by lowering transportation costs. If that is indeed the case, then transportation networks can have impacts on development even for regions that are far away from them, and policies that promote investing in these infrastructure projects should keep in mind these externalities. Another possibility is that it induced migration from well connected regions by cutting down on the costs of migration. The policy prescription, in that case, would depend on whether the government is trying to promote or stem this rural-urban migration.

A last explanation would be that while the first set of regions were to benefit from being directly connected to the cities by being on the least-cost path of connectivity, over time other regions would establish indirect connections via these already connected regions. This would then lead to a protruding network structure that would link regions and spread development by lowering the costs of trade and exchange. This explanations can be ensconced in a growth-model framework by formalizing and estimating the extent of the spatial spillovers. The results in the paper indicate that a substantial amount of the high-rates of $\beta-$convergence can be explained by spatial spillovers in economic activity across neighboring regions.

The implications for policy in the light of such spillovers can be crucial. While the initial transit networks did a lot to encourage economic activity in connected regions, future investments in upgrading these highways did little to help these regions indicating that investments in these highways had reached a portion of diminishing returns. However, the initial investments in the highways not only helped develop connected regions, but also lead to spillovers in activity to neighboring regions. Together, these results indicate that policy-makers should try to connect more regions rather than upgrade routes on already connected regions.

# References


Banerjee, Abhijit, Esther Duflo and Nancy Qian (2012) "On the Road: Access to Transportation Infrastructure and Economic Growth in China" (unpublished)

Barro, Robert J., and Xavier X. Sala-i-Martin. (1992) "Convergence." Journal of Political Economy, 100, 22351.

Donaldson, David (2014) "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure, forthcoming AER

Faber, Benjamin (2014) "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System" - Forthcoming at the Review of Economic Studies

Fogel, Robert (1964) *Railroads and American Economic Growth: Essays in Econometric History*, Baltimore: Johns Hopkins Press

Ghani, Ejaz, Arti Goswami and William Kerr (2013) "Highway to Success in India: The Impact of the Golden Quadrilateral Project for the Location and Performance of Manufacturing" World Bank Working Paper

Henderson, J. Vernon, Adam Storeygard, and David N. Weil (2012) "Measuring Economic Growth from Outer Space." American Economic Review, 102(2): 994-1028.

Higgins, Matthew J., Daniel Levy, and Andrew T. Young. (2006) "Growth and Convergence Across the U.S.: Evidence from County-Level Data. Review of Economics and Statistics, 88, 67181

Hirschman, Albert O. (1969) "The Strategy of Economic Development," in Agarwal, A.N. and Singh, S.P.(eds), Accelerating Investment in Developing Economies (London Oxford Press)

Krugman, P. (1991): "Increasing Returns and Economic Geography," Journal of Political Economy, 99, 483499.

Michaels, Guy (2008) “The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System, The Review of Economics and Statistics, 07 2008, 90 (4), 683701.

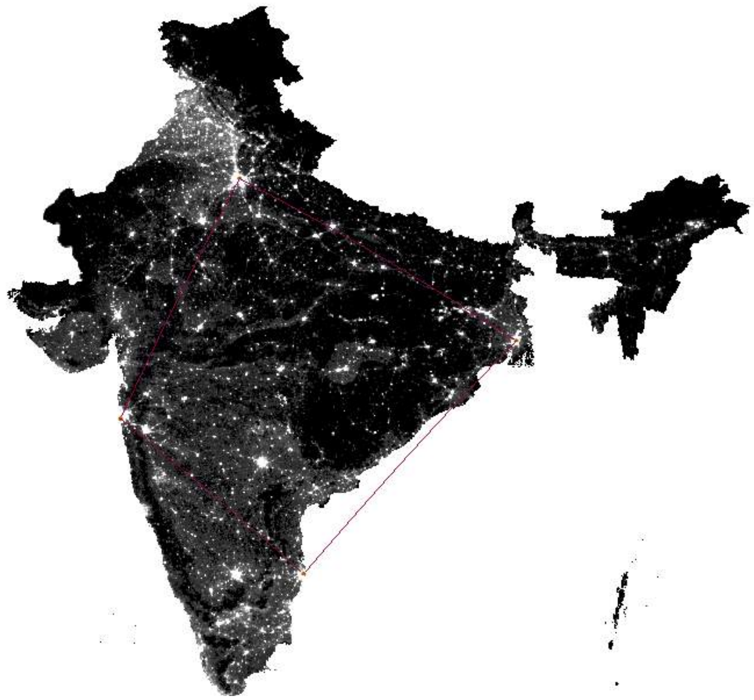
Michalopoulos, Stelios and Elias Papaioannou (2013) “Pre-colonial Ethnic Institutions and Contemporary African Development,” Econometrica, 2013, 81(1): 113152

Munshi, Kaivan and Mark Rosenzweig, 2009. ”Why is Mobility in India so Low? Social Insurance, Inequality, and Growth,” NBER Working Papers 14850, National Bureau of Economic Research, Inc.

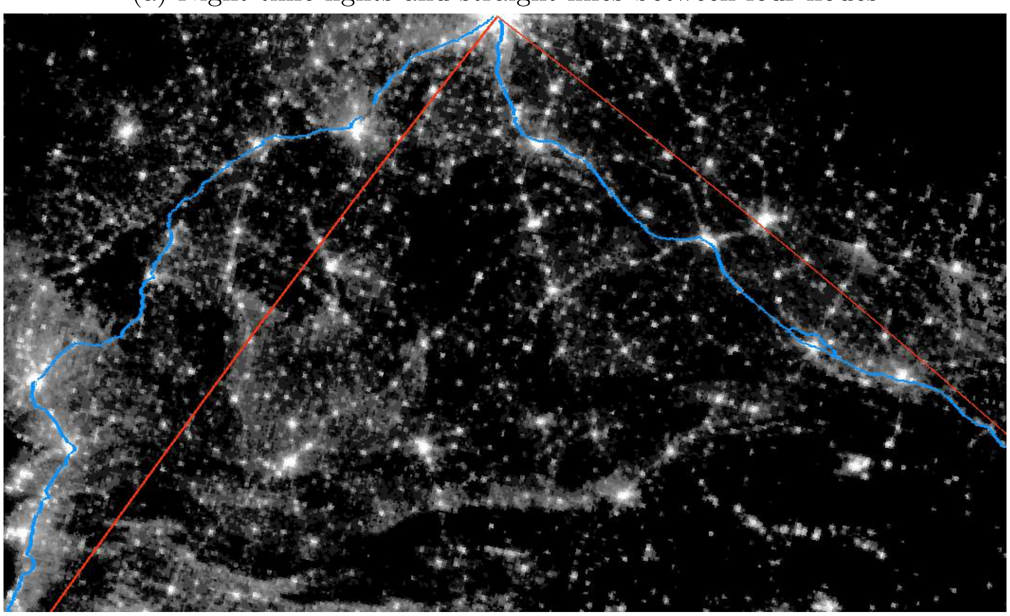
Sala-i-Martin, Xavier X. (1996) “Regional Cohesion: Evidence and Theories of Regional Growth and Convergence.” European Economic Review, 40, 132552

Solow, Robert M (1956). “A Contribution to the Theory of Economic Growth”. Quarterly Journal of Economics (The MIT Press) 70 (1): 6594 [1]

## 11 Tables and Figures



(a) Night-time lights and straight-lines between four nodes



(b) Night-time lights, highways and straight-lines between Mumbai-Delhi and Delhi-Kolkata (The blue-line traces the actual route of the highway, and the red-line indicates the straight-line path between the major cities.

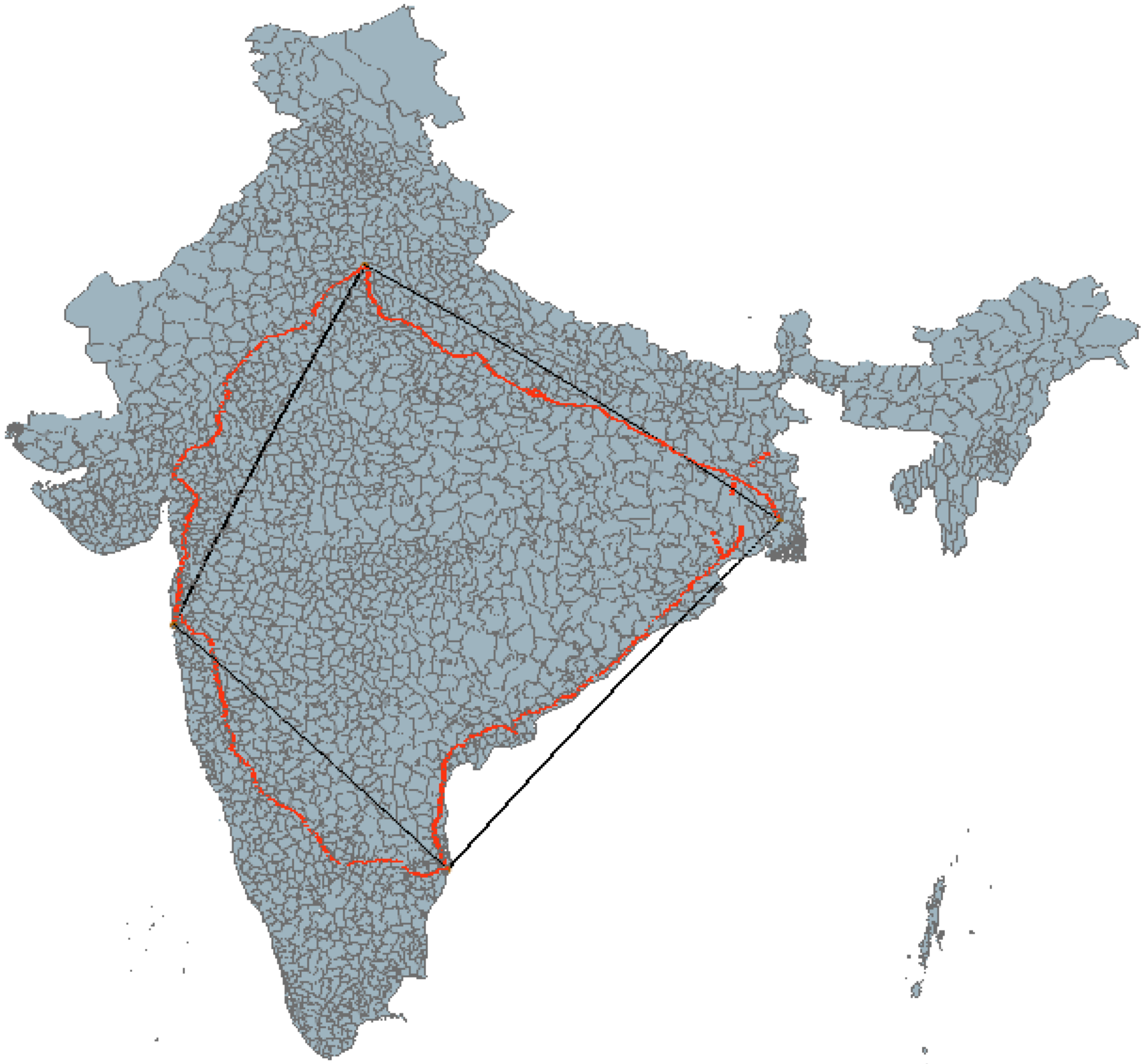
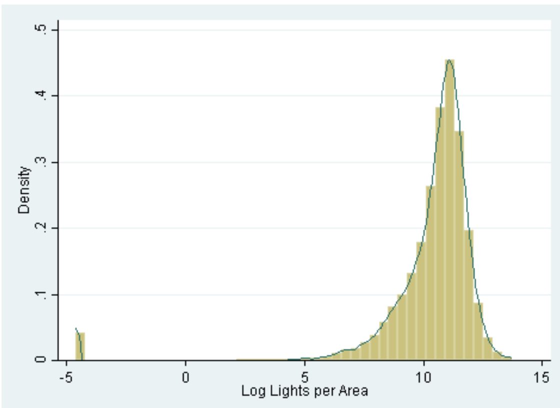
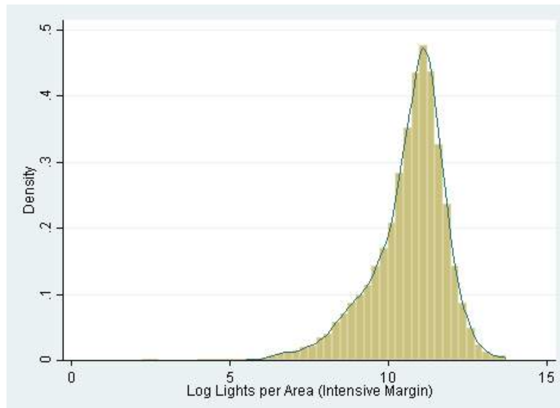


Figure 2: Golden Quadrilateral Highways and boundaries of the sub-districts used in the analysis

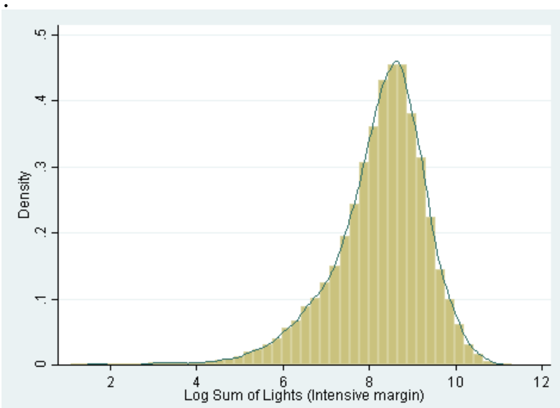
Figure 3: Distribution of Lights and Distance from nearest straight-line



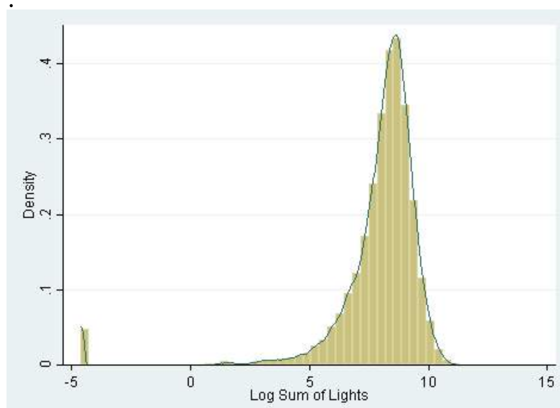

(a) Distribution of Log Lights per Area




(b) Log Lights per Area (no 0s)




(c) Log Sum of Lights (no 0s)

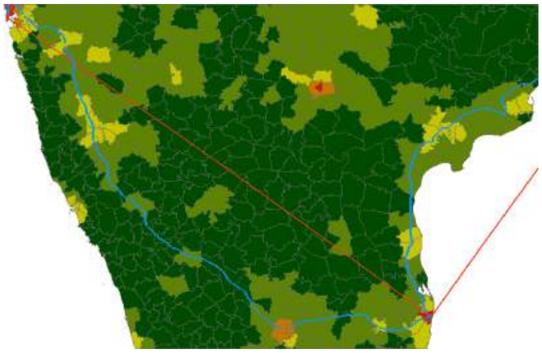



(d) Distribution of Log Sum of Lights

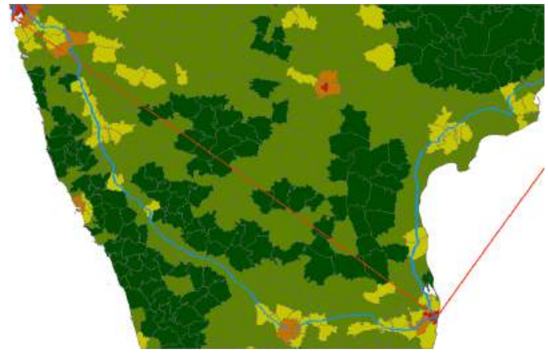
Source: Lights data from NOAA's night-time lights data. Distances calculated using ArcGIS

.

Figure 4: Spread of Lights From the Mumbai-Chennai Highway

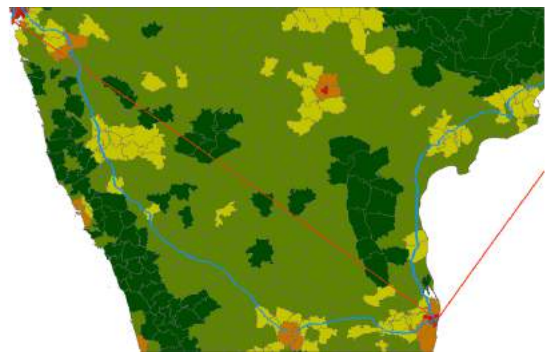


(a) 1992

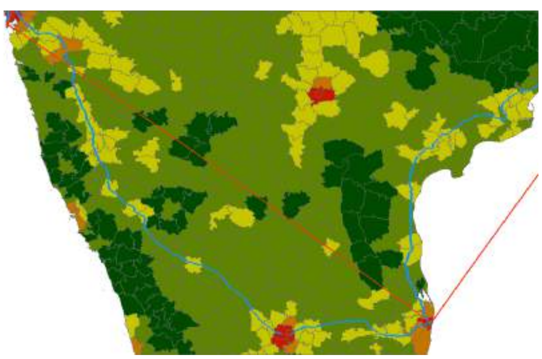


(b) 1997

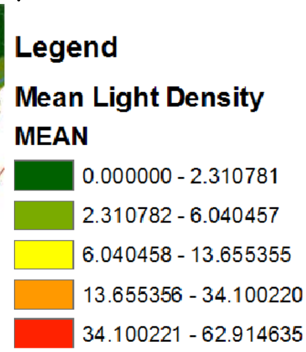
(c) 2003



(d) 2008



(e) 2012

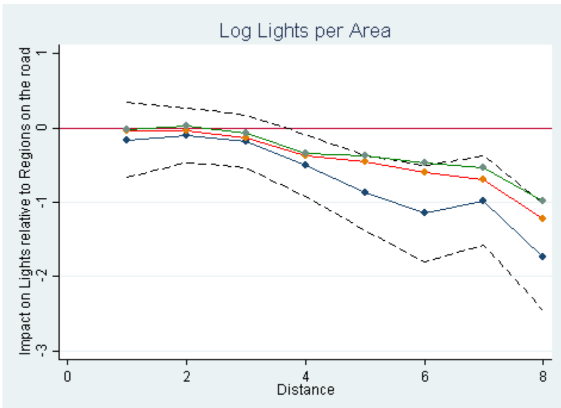



(f)

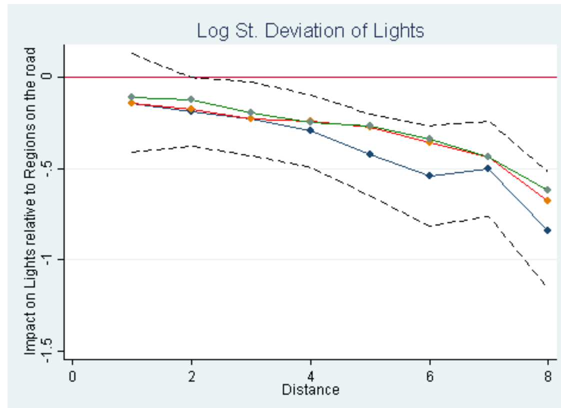
Displaying the spread of night-time light density every five years. The Blue Line indicates the actual path of the Highway, and the Red Line indicates the straight line connecting 2 major cities. The legend of light-densities is on a green to red spectrum, where green is relatively less light-density and red is a higher level of light-density

.

Figure 5: Impact of distance on lights



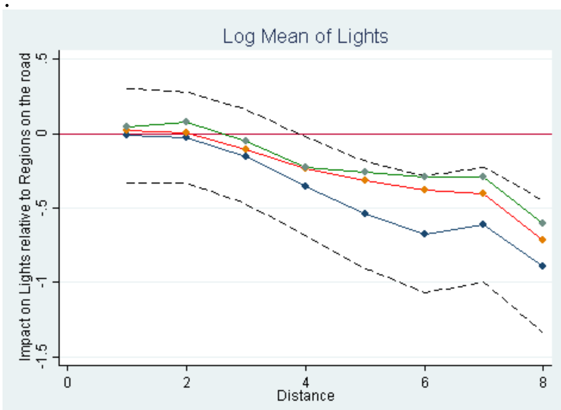

(a) Light Density (per sq miles)



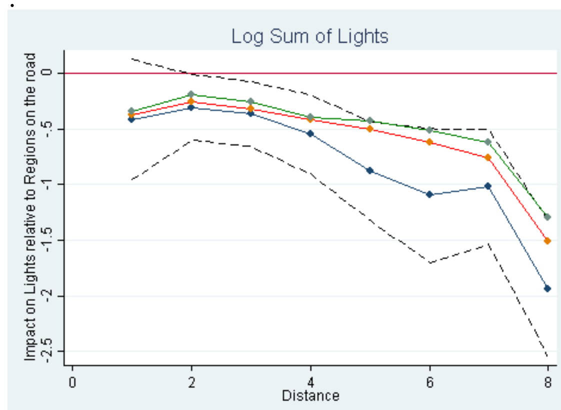

(b) Standard Deviation of Lights

.

.




(c) Mean Lights




(d) Sum of lights

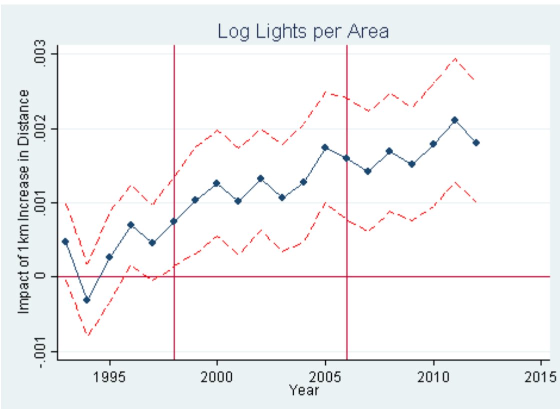
.

.

The graphs show the impact of distance on night-time lights relative to sub-districts that have any portion of the sub-district touching the straight-lines. The blue lines are for the pre-construction period, the orange lines for the construction period and the green lines for the post-construction period. The standard error bands are for the pre-construction (blue) lines and clustered at the district level. The 'Distance' axis consists of 8 quintiles of equal size. The distance quintile cutoffs are roughly as follows: 0 to 40kms, 40 to 90 kms, 90 to 135 kms, 135 to 200kms, 200 to 260kms, 260 to 340 kms, 340 to 440 kms, and above 440kms.

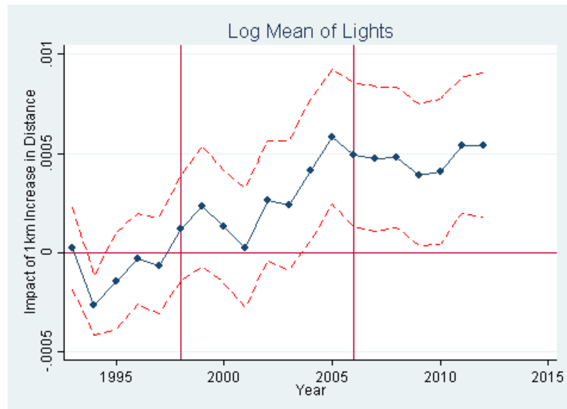
.

Figure 6: How the Impact of Distance on Light-Density changes over time (relative to 1992)

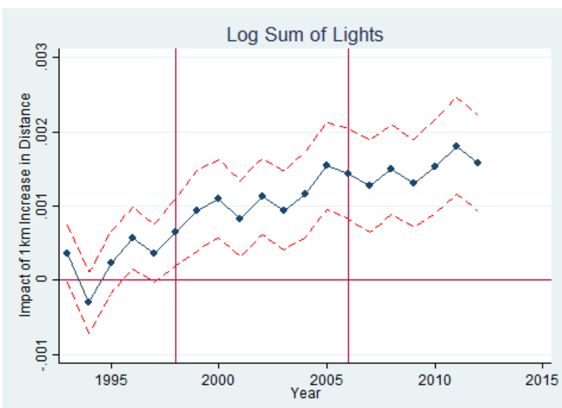



.

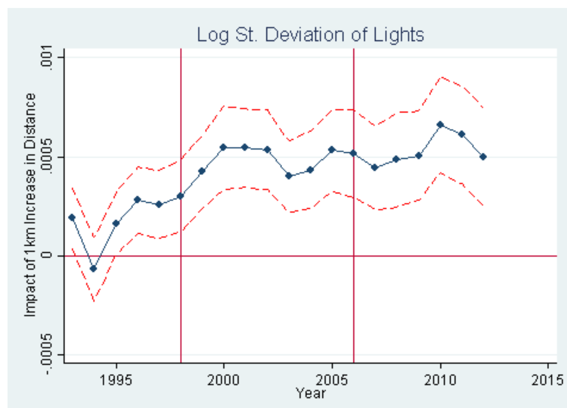
(a) Light-Density:1992 coefficient -0.00406




(b) Mean-lights. 1992 cofficient is -0.00186




(c) Sum of lights: 1992 coefficient is -0.0038




(d) Standard Deviation of Lights -0.00174

Coefficients of change in impact relative to 1992. Standard errors calculated at the district level. Vertical lines represent the phases of construction - 1999 is when the highways started being built. There were delays till 2001 when most work started, and 2006 is when most work was completed. To interpret the graph: the mean impact of a 1km increase in distance from the highway was a 0.00406 fall in light-density, and this impact has been dissipating over time. By 2012 the impact of a 1km increase in distance from the highway had become -0.00406+0.00205, or about -0.00201.

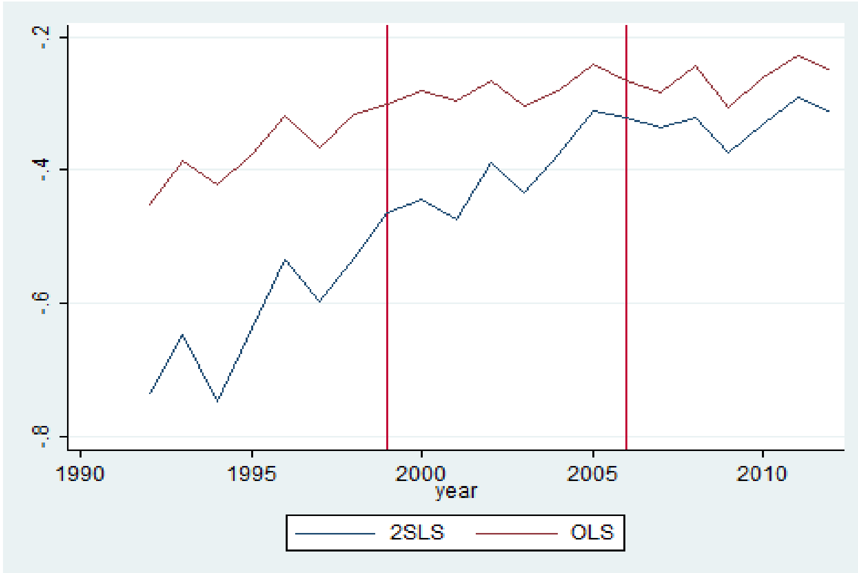



Figure 7: Change in Elasticities Over Time

Elasticities calculated by running a log-log relationship between lights and distance. Vertical lines represent the phases of construction - 1999 is when the highways started being built. There were delays till 2001 when most work started, and 2006 is when most work was completed.

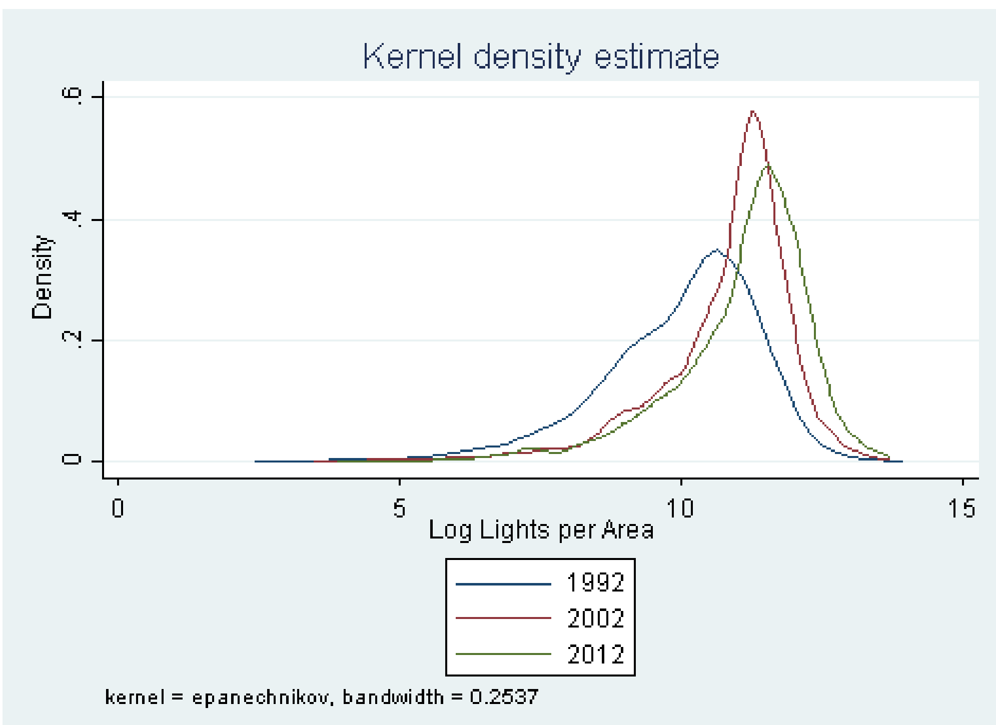



Figure 8: Sigma Convergence in Light Density

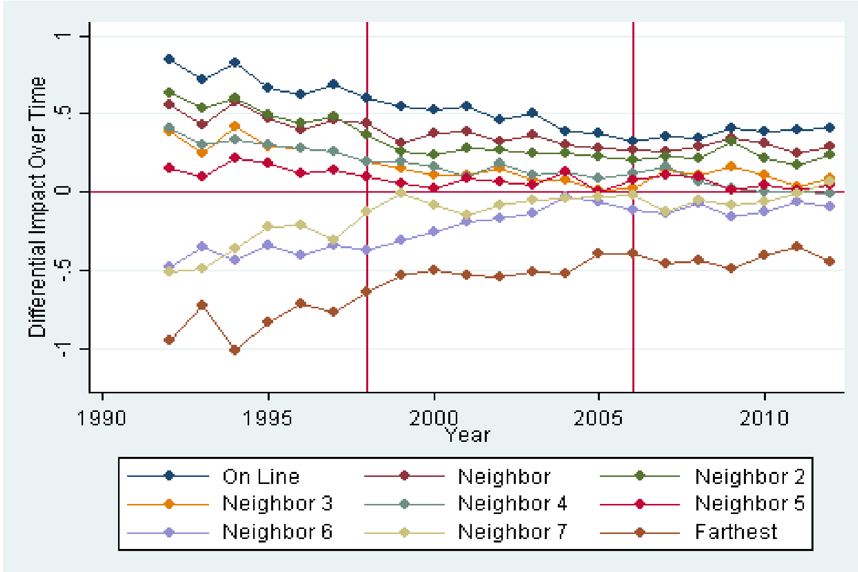



Figure 9: Relative light density for regions on the line, and their neighbors

Relative light density calculated as Log(light density) for that region relative to all other regions. "On Line" represents regions on the straight-line path between two major cities. "Neighbor" represents sub-districts that are neighbors of "On Line" sub-districts. "Neighbors 2" represents neighbors of neighbors of "On Line" regions, and so on. "Farthest" represents regions that are removed from "On Line" regions by more than seven-degrees of separation.

.

Table I: Predicting distance to transit networks with distance to straight-lines

| | Distance to GQ Highway | Distance to Railroad |
|---|---|---|
| Distance to Line | 0.81 | 0.055 |
| SE clusters: | | |
| Sub-district | (0.00983) | (0.0215) |
| District | (0.0205) | (0.0238) |
| State | (0.0372) | (0.0334) |
| R-squared | 0.791 | 0.071 |
| Observations | 2253 | 2253 |
| Controls | Y | Y |

Level of observation - sub-district in the year 2010
Dependent variable 'Distance to GQ Highway' is the nearest geo-distance between the the sub-district and the closest Golden Quadrilateral highway
Dependent variable 'Distance to Railroad' is the nearest geo-distance between the sub-district and the closest railway line
Independent variable 'Distance to Line' is the nearest geo-distance between the sub-district and closest straight-line connecting nodal cities
Controls include distance to nearest nodal city, coastline, latitude and longitude. Results are robust to excluding controls.

.

Table II: OLS relationship between night-time lights and distance to GQ highway

| **Year: 2012** | Sum of Lights | Mean Lights | Lights per area | P(Majority Lights>0) |
|---|---|---|---|---|
| Distance to GQ Highway | -2.016 | -1.287 | -1.929 | -0.0536 |
| | | | | |
| SE clusters: | | | | |
| Sub-district | (0.302) | (0.176) | (0.316) | (0.0614) |
| District | (0.469) | (0.289) | (0.457) | (0.107) |
| State | (0.858) | (0.577) | (0.849) | (0.190) |
| | | | | |
| R-squared | 0.084 | 0.220 | 0.156 | 0.201 |

| **Year: 1992** | Sum of Lights | Mean Lights | Lights per area | P(Majority Lights>0) |
|---|---|---|---|---|
| Distance to GQ Highway | -3.680 | -1.654 | -3.913 | -0.217 |
| | | | | |
| Standard Errors | | | | |
| Level of Clustering: | | | | |
| Sub-district | (0.434) | (0.398) | (0.492) | (0.0582) |
| District | (0.701) | (0.731) | (0.784) | (0.0958) |
| State | (1.592) | (1.213) | (1.773) | (0.178) |
| | | | | |
| R-squared | 0.108 | 0.163 | 0.144 | 0.051 |
| Controls | Y | Y | Y | Y |
| Observations | 2,253 | 2,253 | 2,253 | 2,253 |

Level of observation - sub-district. Distances in 1000 kms.
Independent variable 'Distance to GQ Highway' is the nearest geo-distance between the sub-district and the closest Golden Quadrilateral highway
Dependent variables 'Sum of Lights', 'Mean Lights' and 'Lights per area' are of the form Log(0.01 + Lights). 'Sum of Lights' is the sum of pixels in a sub-district. 'Mean Lights' is the mean light intensity for pixels in a sub-district. 'Lights per area' normalizes the sum by the surface area of the district. Dependent variable P(Majority Lights>0) is a 1/0 indicator variable for if the sub-district has any visible lights emitted in that year
Controls include distance to nearest nodal city, coastline, latitude and longitude. Results are robust to excluding controls.

.

Table III: Reduced-form relationship between Lights and straight-lines

| **Year: 2012** | Sum of Lights | Mean Lights | Lights per area | P(Majority Lights>0) |
|---|---|---|---|---|
| Distance to Line | -1.919 | -1.210 | -1.876 | -0.133 |
| SE clusters: | | | | |
| Sub-district | (0.305) | (0.171) | (0.316) | (0.0568) |
| District | (0.464) | (0.268) | (0.432) | (0.0942) |
| State | (0.778) | (0.517) | (0.782) | (0.136) |
| R-squared | 0.085 | 0.220 | 0.157 | 0.202 |

| **Year: 1992** | Sum of Lights | Mean Lights | Lights per area | P(Majority Lights>0) |
|---|---|---|---|---|
| Distance to Line | -3.805 | -1.861 | -4.060 | -0.224 |
| SE clusters: | | | | |
| Sub-district | (0.420) | (0.216) | (0.479) | (0.0548) |
| District | (0.668) | (0.377) | (0.749) | (0.0923) |
| State | (1.402) | (0.794) | (1.588) | (0.167) |
| R-squared | 0.117 | 0.173 | 0.152 | 0.052 |
| Controls | Y | Y | Y | Y |
| Observations | 2,253 | 2,253 | 2,253 | 2,253 |

Level of observation - sub-district. Distances in 1000 kms.
Independent variable ‘Distance to Line’ is the nearest geo-distance between the sub-district and the closest straight-line connecting nodal cities
Dependent variables ‘Sum of Lights’, ‘Mean Lights’ and ’Lights per area’ are of the form Log(0.01 + Lights). ‘Sum of Lights’ is the sum of pixels in a sub-district. ‘Mean Lights’ is the mean light intensity for pixels in a sub-district. ‘Lights per area’ normalizes the sum by the surface area of the district. Dependent variable P(Majority Lights>0) is a 1/0 indicator variable for if the sub-district has any visible lights emitted in that year
Controls include distance to nearest nodal city, coastline, latitude and longitude. Results are robust to excluding controls.

.

Table IV: Two-staged least squares relationship between lights and distance to GQ highways

| **Year: 2012** | Sum of Lights | Mean Lights | Lights per area | P(Majority Lights>0) |
|---|---|---|---|---|
| Distance to GQ Highway | -2.368 | -1.494 | -1.929 | -0.164 |
| SE clusters: | | | | |
| Sub-district | (0.377) | (0.211) | (0.391) | (0.0699) |
| District | (0.572) | (0.331) | (0.535) | (0.117) |
| State | (0.951) | (0.638) | (0.965) | (0.167) |
| Pagan-Hall Het Test | 72.25 | 129.2 | 68.56 | 183.6 |
| p-value of Pagan-Hall | 0 | 0 | 0 | 0 |
| **Year: 1992** | Sum of Lights | Mean Lights | Lights per area | P(Majority Lights>0) |
| Distance to GQ Highway | -4.697 | -2.297 | -5.011 | -0.277 |
| SE clusters: | | | | |
| Sub-district | (0.521) | (0.266) | (0.593) | (0.0675) |
| District | (0.830) | (0.468) | (0.933) | (0.114) |
| State | (1.702) | (0.969) | (1.940) | (0.202) |
| Pagan-Hall Het Test | 215.6 | 164.4 | 200.0 | 161.9 |
| p-value of Pagan-Hall | 0 | 0 | 0 | 0 |
| Controls | Y | Y | Y | Y |
| Observations | 2,253 | 2,253 | 2,253 | 2,253 |
| Level of Clustering: | F-Stat | Prob>F | Hansen J | Partial R-sq |
| Sub-district | 6796 | 0 | 0 | 0.738 |
| District | 1569 | 0 | 0 | 0.738 |
| State | 474.7 | 0 | 0 | 0.738 |

Level of observation - sub-district. Distances in 1000 kms.

Independent variable 'Distance to GQ Highway' is the nearest predicted geo-distance between the sub-district and the closest GQ highway, predicted by the distance to the closest straight-line connecting nodal cities

Dependent variables 'Sum of Lights', 'Mean Lights' and 'Lights per area' are of the form Log(0.01 + Lights). 'Sum of Lights' is the sum of pixels in a sub-district. 'Mean Lights' is the mean light intensity for pixels in a sub-district. 'Lights per area' normalizes the sum by the surface area of the district. Dependent variable P(Majority Lights>0) is a 1/0 indicator variable for if the sub-district has any visible lights emitted in that year

Controls include distance to nearest nodal city, coastline, latitude and longitude. Results are robust to excluding controls.

Table V: OLS relationship between lights and rail-lines

| **Year: 2012** | Sum of Lights | Mean Lights | Lights per area | P(Majority Lights>0) |
|---|---|---|---|---|
| Distance to Railways | -9.321 | -6.253 | -8.82 | -1.069 |
| SE clusters: | | | | |
| Sub-district | (5.341) | (2.885) | (4.656) | (0.400) |
| District | (5.536) | (2.954) | (4.816) | (0.403) |
| State | (7.584) | (3.997) | (6.552) | (0.530) |
| R-squared | 0.096 | 0.232 | 0.164 | 0.207 |
| **Year: 1992** | Sum of Lights | Mean Lights | Lights per area | P(Majority Lights>0) |
| Distance to Railways | -14.32 | -7.884 | -15.37 | -0.862 |
| SE clusters: | | | | |
| Sub-district | (8.247) | (3.813) | (8.535) | (0.404) |
| District | (8.569) | (3.882) | (8.827) | (0.416) |
| State | (11.86) | (5.328) | (12.20) | (0.542) |
| R-squared | 0.108 | 0.175 | 0.145 | 0.051 |
| Controls | Y | Y | Y | Y |
| Observations | 2,253 | 2,253 | 2,253 | 2,253 |

Level of observation - sub-district. Distances in 1000 kms.

Independent variable 'Distance to Railroad' is the nearest geo-distance between the sub-district and the closest Railway line

Dependent variables 'Sum of Lights', 'Mean Lights' and 'Lights per area' are of the form Log(0.01 + Lights). 'Sum of Lights' is the sum of pixels in a sub-district. 'Mean Lights' is the mean light intensity for pixels in a sub-district. 'Lights per area' normalizes the sum by the surface area of the district. Dependent variable P(Majority Lights>0) is a 1/0 indicator variable for if the sub-district has any visible lights emitted in that year

Controls include distance to nearest nodal city, coastline, latitude and longitude. Results are robust to excluding controls.

Table VI: Two staged least squares relationship between lights and distance to railways

| **Year: 2012** | Sum of Lights | Mean Lights | Lights per area | P(Majority Lights>0) |
|---|---|---|---|---|
| Distance to Railways | -34.86 | -21.99 | -34.09 | -2.412 |
| | | | | |
| SE clusters: | | | | |
| Sub-district | (14.51) | (8.781) | (14.32) | (1.271) |
| District | (15.63) | (9.903) | (15.91) | (1.856) |
| State | (22.23) | (14.34) | (22.31) | (2.669) |
| | | | | |
| Pagan-Hall Het Test | 24.15 | 18.87 | 23.54 | 52.75 |
| p-value of Pagan-Hall | 0.000203 | 0.00203 | 0.000266 | 0 |
| **Year: 1992** | Sum of Lights | Mean Lights | Lights per area | P(Majority Lights>0) |
| Distance to Railways | -69.14 | -33.82 | -73.77 | -4.076 |
| | | | | |
| SE clusters: | | | | |
| Sub-district | (28.20) | (13.48) | (30.25) | (1.831) |
| District | (30.19) | (15.06) | (33.09) | (2.293) |
| State | (42.09) | (21.36) | (45.97) | (3.562) |
| | | | | |
| Pagan-Hall Het Test | 42.88 | 25.49 | 38.05 | 12.99 |
| p-value of Pagan-Hall | 0 | 0.000112 | 0 | 0.0235 |
| Controls | Y | Y | Y | Y |
| Observations | 2,253 | 2,253 | 2,253 | 2,253 |
| Level of Clustering: | F-Stat | Prob > F | Hansen J | Partial R-sq |
| Sub-district | 6.570 | 0.0104 | 0 | 0.0516 |
| District | 5.348 | 0.0211 | 0 | 0.0516 |
| State | 2.707 | 0.109 | 0 | 0.0516 |

Level of observation - sub-district. Distances in 1000 kms.
Independent variable 'Distance to Railroad' is the nearest predicted geo-distance between the sub-district and the closest rail-line, predicted by the distance to the closest straight-line connecting nodal cities
Dependent variables 'Sum of Lights', 'Mean Lights' and 'Lights per area' are of the form Log(0.01 + Lights). 'Sum of Lights' is the sum of pixels in a sub-district. 'Mean Lights' is the mean light intensity for pixels in a sub-district. 'Lights per area' normalizes the sum by the surface area of the district. Dependent variable P(Majority Lights>0) is a 1/0 indicator variable for if the the majority of recorded lights was greater than 0 in that year
Controls include distance to nearest nodal city, coastline, latitude and longitude. Results are robust to excluding controls.

Table VII: Reduced Form: Elasticity of Lights, Distance and GDP

| Log(0.01+Light Density) | 1992 | 1996 | 2000 | 2004 | 2008 | 2012 |
|---|---|---|---|---|---|---|
| Log(0.01+Distance to line) | -0.497 | -0.362 | -0.3 | -0.254 | -0.217 | -0.212 |
| SE clusters: | | | | | | |
| Sub-district | (0.0591) | (0.0462) | (0.0379) | (0.0365) | (0.0365) | (0.0364) |
| District | (0.106) | (0.0747) | (0.0571) | (0.0567) | (0.0543) | (0.0545) |
| State | (0.226) | (0.151) | (0.116) | (0.109) | (0.0944) | (0.0921) |
| R-Squared | 0.135 | 0.160 | 0.170 | 0.196 | 0.154 | 0.147 |
| Observations | 2,253 | 2,253 | 2,253 | 2,253 | 2,253 | 2,253 |
| Controls | Y | Y | Y | Y | Y | Y |
| Estimated GDP-distance elasticity* | 0.1491 | 0.1086 | 0.09 | 0.0762 | 0.0651 | 0.0636 |

*Estimates of GDP-distance elasticity rely on elasticity of GDP-lights being 0.3 for low-middle income countries. Therefore to find elasticity of GDP-distance, multiply the coefficient with 0.3.
Level of observation - sub-district.
Independent variable 'Log (0.01+Distance to Line)' is the nearest geo-distance between the sub-district and the closest straight-line connecting Mumbai and Kolkata - two historical cities that have not had direct transit networks connecting them.
Dependent variables is Log(0.01 + Light density). 'Lights per area' normalizes the sum by the surface area of the district.
Controls include distance to nearest nodal city, coastline, latitude and longitude. Results are robust to excluding controls.

Table VIII: Elasticity between Population and Distance to Line

| Year | 2002 | 2004 | 2006 | 2008 | 2010 | 2012 |
|---|---|---|---|---|---|---|
| Log(0.01+Distance to line) | -0.226 | -0.182 | -0.178 | -0.179 | -0.176 | -0.191 |
| SE clusters: | | | | | | |
| Sub-district | (0.0221) | (0.0293) | (0.0290) | (0.0290) | (0.0305) | (0.0305) |
| District | (0.0404) | (0.0560) | (0.0543) | (0.0545) | (0.0562) | (0.0559) |
| State | (0.0838) | (0.0991) | (0.0968) | (0.0969) | (0.0977) | (0.0970) |
| R-Squared | 0.091 | 0.056 | 0.053 | 0.053 | 0.044 | 0.046 |
| Observations | 2,246 | 2,246 | 2,246 | 2,246 | 2,246 | 2,246 |
| Controls | Y | Y | Y | Y | Y | Y |

Population Data from LandScan, US Department of Energy

Table IX: Elasticity For non-GQ Route: Mumbai to Kolkata

| Year | 1992 | 1996 | 2000 | 2004 | 2008 | 2012 |
|---|---|---|---|---|---|---|
| Log(0.01+Distance to Line) | -0.248 | -0.147 | -0.0735 | -0.0822 | -0.0273 | -0.0520 |
| SE clusters: | | | | | | |
| Sub-district | (0.0463) | (0.0351) | (0.0307) | (0.0300) | (0.0293) | (0.0294) |
| District | (0.0805) | (0.0520) | (0.0436) | (0.0425) | (0.0405) | (0.0404) |
| State | (0.179) | (0.118) | (0.0964) | (0.0845) | (0.0780) | (0.0743) |
| R-Squared | | | | | | |
| Observations | 2,253 | 2,253 | 2,253 | 2,253 | 2,253 | 2,253 |
| Controls | Y | Y | Y | Y | Y | Y |
| Estimated GDP-distance elasticity* | 0.0744 | 0.0441 | 0.02205 | 0.02466 | 0.00819 | 0.0156 |

*Estimates of GDP-distance elasticity rely on elasticity of GDP-lights being 0.3 for low-middle income countries. Therefore to find elasticity of GDP-distance, multiply the coefficient with 0.3.
Level of observation - sub-district.
Independent variable 'Log (0.01+Distance to Line)' is the nearest geo-distance between the sub-district and the closest straight-line connecting Mumbai and Kolkata - two historical cities that have not had direct transit networks connecting them.
Dependent variables is Log(0.01 + Light density). 'Lights per area' normalizes the sum by the surface area of the district.
Controls include distance to nearest nodal city, coastline, latitude and longitude. Results are robust to excluding controls.

Table X: The impact of distance changing over time

| | Sum of Lights | Mean Lights | Lights per area | P(Majority Lights>0) |
|---|---|---|---|---|
| Distance to Line | -3.257 | -1.796 | -3.389 | -0.275 |
| SE Level of Clusters: | | | | |
| Sub-district | (0.357) | (0.184) | (0.386) | (0.0497) |
| District | (0.539) | (0.302) | (0.561) | (0.0893) |
| State | (1.088) | (0.643) | (1.187) | (0.155) |
| | | | | |
| Distance*Construction Period | 0.867 | 0.35 | 0.955 | -0.0117 |
| SE Level of Clusters: | | | | |
| Sub-district | (0.136) | (0.0521) | (0.162) | (0.0259) |
| District | (0.161) | (0.0872) | (0.202) | (0.0445) |
| State | (0.351) | (0.232) | (0.397) | (0.117) |
| | | | | |
| Distance*Post Period | 1.232 | 0.526 | 1.388 | 0.0938 |
| SE Level of Clusters: | | | | |
| Sub-district | (0.173) | (0.0733) | (0.205) | (0.0306) |
| District | (0.206) | (0.120) | (0.247) | (0.0512) |
| State | (0.437) | (0.267) | (0.487) | (0.127) |
| | | | | |
| R-Squared | 0.144 | 0.285 | 0.203 | 0.208 |
| Year Fixed Effects | Y | Y | Y | Y |
| Controls | Y | Y | Y | Y |
| Observations | 47,313 | 47,313 | 47,313 | 47,313 |

Level of observation - sub-district. Distances in 1000 kms.
Independent variable 'Distance to Line' is the nearest predicted geo-distance between the sub-district and the closest straight-line connecting nodal cities
Dependent variables 'Sum of Lights', 'Mean Lights' and 'Lights per area' are of the form Log(0.01 + Lights). 'Sum of Lights' is the sum of pixels in a sub-district. 'Mean Lights' is the mean light intensity for pixels in a sub-district. 'Lights per area' normalizes the sum by the surface area of the district. Dependent variable P(Majority Lights>0) is a 1/0 indicator variable for if the the majority of recorded lights was greater than 0 in that year
Controls include distance to nearest nodal city, coastline, latitude and longitude, and year fixed effects. Results are robust to excluding controls.
Pre-construction period is 1992 to 1999, construction period is 1999 to 2006, and post-construction period is 2007 onwards.

.

Table XI: Beta-convergence and the Solow model

| Log(light(t)) - Log(light(t-1) | $\beta$ convergence | Solow model | Road construction |
|---|---|---|---|
| Growth in light density | $\Delta$ Log(light density) | $\Delta$ Log(light density) | $\Delta$ Log(light density) |
| Year | | -0.00440 | -0.0312 |
| | | (0.000703) | (0.00156) |
| Lagged Log(light density) | -0.108 | | -0.108 |
| | (0.0143) | | (0.0144) |
| Pre construction | | | 0.180 |
| | | | (0.0217) |
| Building period | | | 0.288 |
| | | | (0.0195) |
| Post construction | | | 0.651 |
| | | | (0.0233) |
| Log(distance to line)*pre | | | -0.0105 |
| | | | (0.00324) |
| Log(distance to line)*build | | | -0.00349 |
| | | | (0.00261) |
| Log(distance to line)*end | | | -0.00759 |
| | | | (0.00224) |
| Log(distance to line) | | 0.00397 | |
| | | (0.000948) | |
| Constant | 1.187 | 8.872 | 63.27 |
| | (0.148) | (1.410) | (3.061) |
| Observations | 45,060 | 45,060 | 45,060 |
| R-squared | 0.071 | 0.001 | 0.081 |

Level of observation - sub-district-year
Dependent variable $Log(0.01 + lightdensity)_t - Log(0.01 + lightdensity)_{t-1}$
Standard errors calculated at the district level (587 districts)
While column 1 tests the notion of $\beta-$convergence (10%) in this case; column 2 tests for whether the distance to the line changes the steady-state level of income or the initial level of income, as discussed in the text. Column 3 studies how the relationship in column 2 changes over time.

Table XII: 2001 and 2011 Census: population, workers, and literacy

| Log Population (2011 Census) | Rural | Total | Female | Male | SC | ST |
|---|---|---|---|---|---|---|
| Log(Distance) | -0.239 | -0.218 | -0.216 | -0.220 | -0.226 | 0.285 |
| | (0.0394) | (0.0402) | (0.0401) | (0.0403) | (0.0474) | (0.0737) |
| Constant | 16.85 | 17.86 | 17.16 | 17.17 | 13.91 | 7.284 |
| | (0.802) | (0.798) | (0.799) | (0.797) | (0.966) | (1.153) |
| Observations | 5,259 | 5,290 | 5,290 | 5,290 | 5,032 | 5,149 |
| R-squared | 0.211 | 0.205 | 0.203 | 0.207 | 0.064 | 0.025 |

| Log Population (2001 Census) | Rural | Total | Female | Male | SC | ST |
|---|---|---|---|---|---|---|
| Log(Distance) | -0.213 | -0.201 | -0.200 | -0.203 | -0.351 | 0.300 |
| | (0.0390) | (0.0397) | (0.0398) | (0.0397) | (0.0619) | (0.0763) |
| Constant | 16.65 | 17.56 | 16.88 | 16.86 | 17.76 | 6.358 |
| | (0.800) | (0.799) | (0.802) | (0.797) | (1.433) | (1.232) |
| Observations | 5,269 | 5,293 | 5,293 | 5,293 | 5,112 | 5,050 |
| R-squared | 0.195 | 0.197 | 0.196 | 0.198 | 0.134 | 0.027 |

| Census Year | 2001 Log(cultivators) per capita | 2011 Log(cultivators) per capita | 2001 Log(Ag Laborers) per capita | 2001 Log(workers in HH ind. per capita) | 2001 Literacy Rate | 2001 Gender Gap in literacy |
|---|---|---|---|---|---|---|
| Log(Distance) | 0.0612 | 0.0868 | -0.149 | -0.0698 | -0.111 | -1.478 |
| | (0.0227) | (0.0222) | (0.0336) | (0.0228) | (0.700) | (0.283) |
| Constant | 3.326 | 2.323 | 1.207* | -3.391 | 125.3 | 18.93 |
| | (0.455) | (0.544) | (0.632) | (0.360) | (8.507) | (4.073) |
| Observations | 5,264 | 5,286 | 5,278 | 5,284 | 4,809 | 4,809 |
| R-squared | 0.073 | 0.035 | 0.268 | 0.081 | 0.090 | 0.119 |

Level of observation - Census Teshils (also called Taluks, Mandals and Wards depending on the region)
Standard errors calculated at the district level (587 districts)
SC are known as Scheduled Castes, and STs are Scheduled Tribes - which the two most economically and socially disadvantage
sections.
Gender Gap is defined as the male-literacy rate minus the female-literacy rate.

.

Table XIII: Spillovers from Neighbors and the effect of distance

| Log(light density) | 1992 | 1997 | 2003 | 2007 | 2012 |
|---|---|---|---|---|---|
| Log(light density) of | 0.868 | 0.827 | 0.773 | 0.843 | 0.802 |
| neighbors closer to road | (0.0491) | (0.0656) | (0.0637) | (0.0560) | (0.0762) |
| Log(distance) | -0.104 | -0.101 | -0.0853 | -0.0652 | -0.0640 |
| | (0.0376) | (0.0312) | (0.0254) | (0.0235) | (0.0225) |
| Constant | 5.453 | 5.881 | 6.632 | 5.487 | 5.631 |
| | (1.304) | (1.418) | (1.319) | (1.197) | (1.415) |
| Observations | 2,219 | 2,219 | 2,219 | 2,219 | 2,219 |
| R-squared | 0.493 | 0.470 | 0.462 | 0.502 | 0.440 |
| Distance-spillover paramter | -0.4954 | -0.4294 | -0.3138 | -0.2896 | -0.2543 |

| Log(light density) | 1992 | 1997 | 2003 | 2007 | 2012 |
|---|---|---|---|---|---|
| Log(distance) | -0.497 | -0.404 | -0.294 | -0.227 | -0.212 |
| | (0.0591) | (0.0488) | (0.0388) | (0.0382) | (0.0364) |
| Constant | 22.12 | 22.38 | 22.59 | 21.71 | 20.68 |
| | (1.277) | (1.122) | (1.046) | (1.084) | (0.998) |
| Observations | 2,253 | 2,253 | 2,253 | 2,253 | 2,253 |
| R-squared | 0.135 | 0.174 | 0.203 | 0.180 | 0.147 |

Level of observation - Sub-district
Standard errors calculated at the district level (587 districts)
This table tests the model where light density $L_k$ depends on light-density of the neighbors closer to the highway $L_{k-1}$ and distance to the highway $D$, in the following way: $L_k = \chi L_{k-1} + \mu D$. This relationship is estimated in the top panel. Furthermore, we can recursively solve, to show that $L_k = (\sum_{j=0}^{k} \chi^j)\mu D$, which is the parameter evaluated as the "Distance-spillover parameter" between the two panels for the average number of degrees-of-separation (i.e. $k = 6$). The bottom panel then tests if this parameter is equal to the parameter obtained by regressing $L_k = \alpha D$.

.

Table XIV: Neighbors closer to the road vs. farther away from the road

| | Contemporaneous effect | Lagged effect | Changes* |
|---|---|---|---|
| **Log(light density) of neighbors closer to road** | 0.490 | | |
| | (0.0302) | | |
| Lagged | | 0.370 | |
| | | (0.0269) | |
| Changes | | | 0.332 |
| | | | (0.0427) |
| **Log(light density) of neighbors further from road** | 0.351 | | |
| | (0.0356) | | |
| Lagged | | 0.237 | |
| | | (0.0333) | |
| Changes | | | 0.265 |
| | | | (0.0287) |
| **Constant** | 1.648 | 4.138 | 0.0280 |
| | (0.304) | (0.387) | (0.00290) |
| Observations | 45,045 | 42,900 | 42,900 |
| R-squared | 0.341 | 0.194 | 0.106 |
| Fixed Effect Units | 2,145 | 2,145 | 2,145 |

Fized effects regressions - Level of observation - Sub-district-year
Standard errors calculated at the district level (587 districts)
This table tests whether neighbors closer to the highway have larger impacts than regions away from the highway.
* The 'Changes' version of the equation estimates $\Delta Log(lights)_{t,k} = \beta\Delta Log(lights)_{t,k-1} + \gamma\Delta Log(lights)_{t,k+1}$

# 12 Additional Tables and Figures

.

Table XV: Elasticity between Light-Density and State Domestic Product for 32 States

| Per capita | GDP at | 2005 | prices | | | | | |
|---|---|---|---|---|---|---|---|---|
| Log(per cap GDP) | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
| Log(light density) | 0.19 | 0.198 | 0.19 | 0.191 | 0.183 | 0.198 | 0.192 | 0.186 |
| | (0.0508) | (0.0503) | (0.0510) | (0.0545) | (0.0557) | (0.0617) | (0.0600) | (0.0589) |
| Constant | 8.422 | 8.409 | 8.563 | 8.541 | 8.72 | 8.548 | 8.695 | 8.798 |
| | (0.526) | (0.523) | (0.529) | (0.584) | (0.594) | (0.684) | (0.655) | (0.648) |
| Observations | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 |
| R-squared | 0.318 | 0.340 | 0.316 | 0.291 | 0.265 | 0.255 | 0.254 | 0.249 |

| Per capita | NDP at | 2005 | prices | | | | | |
|---|---|---|---|---|---|---|---|---|
| Log(per cap NDP) | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
| Log(light density) | 0.191 | 0.199 | 0.191 | 0.192 | 0.186 | 0.202 | 0.196 | 0.191 |
| | (0.0514) | (0.0508) | (0.0517) | (0.0549) | (0.0560) | (0.0626) | (0.0613) | (0.0606) |
| Constant | 8.294 | 8.275 | 8.428 | 8.406 | 8.558 | 8.375 | 8.519 | 8.612 |
| | (0.533) | (0.528) | (0.536) | (0.589) | (0.597) | (0.694) | (0.670) | (0.667) |
| Observations | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 |
| R-squared | 0.315 | 0.339 | 0.313 | 0.290 | 0.269 | 0.257 | 0.255 | 0.250 |

Regressions of Log(0.01+light density) on Log(per capita domestic product) at the state level.
State Domestic Product Sources: Reserve Bank of India
GDP indicates Gross Domestic Product of the State; and NDP is the Net Domestic Product

.

Table XVI: Elasticity between Light Density and Per Capita District Domestic Product

| Per capita GDP | at current | prices | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Log(per cap GDP) | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
| Log(light density) | 0.408 | 0.482 | 0.484 | 0.555 | 0.44 | 0.532 | 0.519 | 0.325 |
| | (0.0257) | (0.0294) | (0.0302) | (0.0321) | (0.0291) | (0.0376) | (0.0320) | (0.0502) |
| Constant | 5.375 | 4.654 | 4.684 | 3.8 | 5.172 | 3.954 | 4.287 | 6.485 |
| | (0.258) | (0.303) | (0.311) | (0.343) | (0.305) | (0.416) | (0.347) | (0.554) |
| Observations | 114 | 127 | 127 | 127 | 127 | 146 | 127 | 76 |
| R-squared | 0.692 | 0.682 | 0.673 | 0.706 | 0.647 | 0.582 | 0.678 | 0.361 |

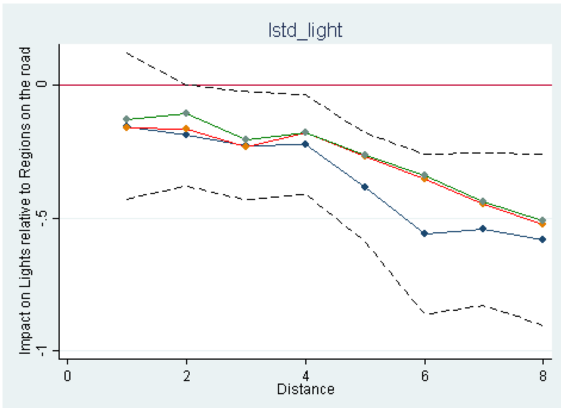
| Per capita GDP | at current | prices | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Log(per cap GDP) | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
| Log(light density) | 0.423 | 0.394 | 0.45 | 0.523 | 0.371 | 0.516 | 0.468 | 0.287 |
| | (0.0378) | (0.0376) | (0.0417) | (0.0412) | (0.0373) | (0.0485) | (0.0460) | (0.0492) |
| Constant | 5.307 | 5.731 | 5.221 | 4.392 | 6.266 | 4.526 | 5.311 | 7.371 |
| | (0.399) | (0.401) | (0.447) | (0.454) | (0.405) | (0.555) | (0.516) | (0.543) |
| Observations | 91 | 91 | 91 | 91 | 91 | 91 | 91 | 76 |
| R-squared | 0.585 | 0.553 | 0.566 | 0.644 | 0.526 | 0.560 | 0.538 | 0.316 |

| Per capita NDP | at current | prices | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Log(per cap NDP) | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
| Log(light density) | 0.332 | 0.357 | 0.347 | 0.371 | 0.354 | 0.371 | 0.394 | 0.388 |
| | (0.0236) | (0.0259) | (0.0264) | (0.0291) | (0.0249) | (0.0324) | (0.0296) | (0.0462) |
| Constant | 6.213 | 5.988 | 6.16 | 5.82 | 6.103 | 5.762 | 5.661 | 5.75 |
| | (0.241) | (0.268) | (0.274) | (0.312) | (0.264) | (0.361) | (0.323) | (0.518) |
| Observations | 209 | 222 | 222 | 222 | 222 | 222 | 190 | 96 |
| R-squared | 0.488 | 0.463 | 0.439 | 0.426 | 0.479 | 0.373 | 0.485 | 0.428 |

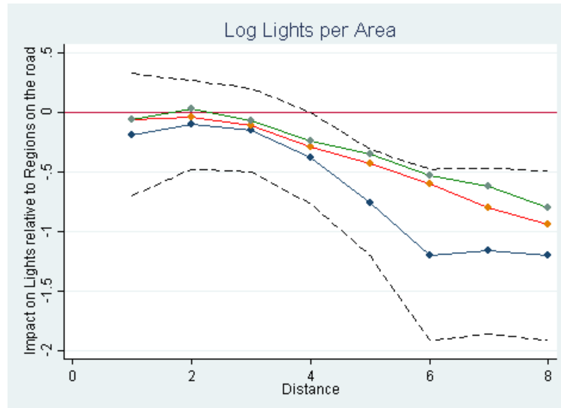
| Per capita NDP | at 2004-5 | prices | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Log(per cap NDP) | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
| Log(light density) | 0.363 | 0.354 | 0.344 | 0.365 | 0.344 | 0.362 | 0.385 | 0.357 |
| | (0.0230) | (0.0243) | (0.0261) | (0.0290) | (0.0244) | (0.0310) | (0.0287) | (0.0457) |
| Constant | 5.915 | 6.071 | 6.302 | 6.113 | 6.516 | 6.238 | 6.223 | 6.591 |
| | (0.238) | (0.252) | (0.271) | (0.312) | (0.260) | (0.347) | (0.315) | (0.511) |
| Observations | 249 | 249 | 222 | 249 | 249 | 249 | 217 | 97 |
| R-squared | 0.502 | 0.463 | 0.440 | 0.391 | 0.446 | 0.356 | 0.456 | 0.391 |

Regressions of Log(0.01+light density) on Log(per capita domestic product) at the district level.
District Domestic Product Sources: Department of Statistics and Programme Implementation, Government of West Bengal; Planning Commission; Directorate of Economics and Statistics Government of Uttar Pradesh; Department of Economics and Statistics Government of Tamil Nadu; Directorate of Economics and Statistics Government of Rajasthan; Department of Planning Government of Punjab; Planning and Coordination Government of Odisha; Directorate of Economics and Statistics Government of Maharashtra; Directorate of Economics and Statistics Government of Kerala; Planning Programme Monitoring and Statistics Department Government of Karnataka; Directorate of Economics and Statistics Government of Bihar; Directorate of Economics and Statistics Government of

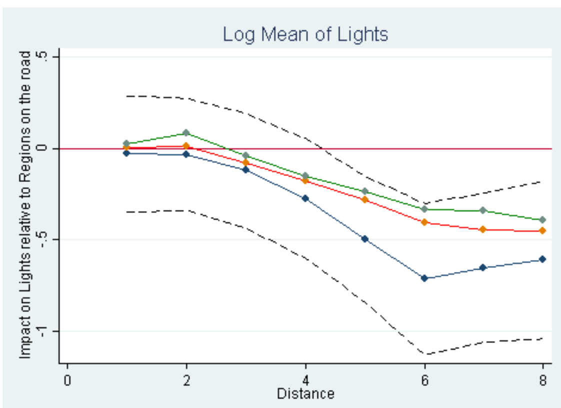
Figure 10: Robustness Checks: Impact of distance on lights excluding outlying states



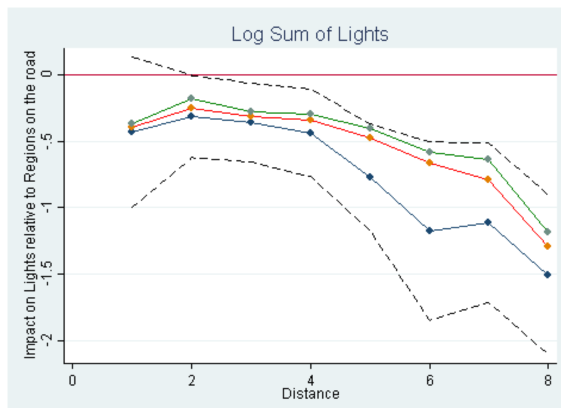

(a) Standard Deviation of Lights




(b) Light Density (per sq miles)




(c) Mean Lights

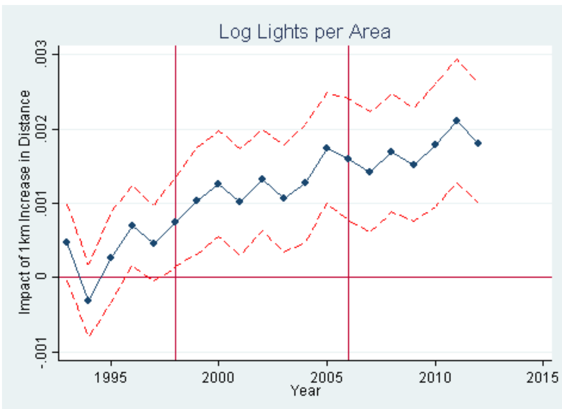



(d) Sum of lights

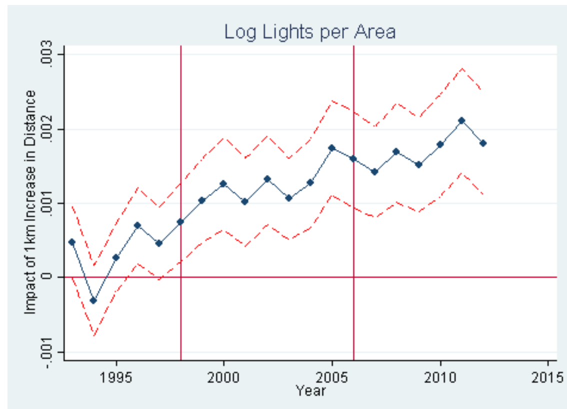
The excluded states in this robustness check include: Jammu and Kashmir, Manipur, Meghalaya, Tripura, Nagaland, Sikkim, Assam, Arunachal Pradesh, Mizoram, Andaman and Nicobar Islands and Lakshwadeep. The graphs show the impact of distance on night-time lights relative to sub-districts that have any portion of the sub-district touching the straight-lines. The blue lines are for the pre-construction period, the orange lines for the construction period and the green lines for the post-construction period. The standard error bands are for the pre-construction (blue) lines and clustered at the district level. The 'Distance' axis consists of 8 quintiles of equal size. The distance quintile cutoffs are roughly as follows: 0 to 40kms, 40 to 90 kms, 90 to 135 kms, 135 to 200kms, 200 to 260kms, 260 to 340 kms, 340 to 440 kms, and above 440kms.

.

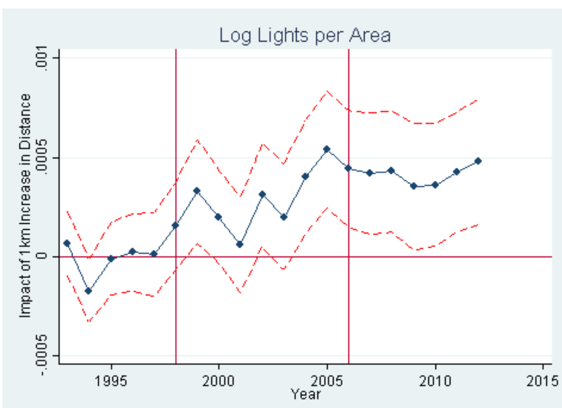
Figure 11: Robustness Checks: Different samples and specifications for: How the Impact of Distance on Light-Density changes over time



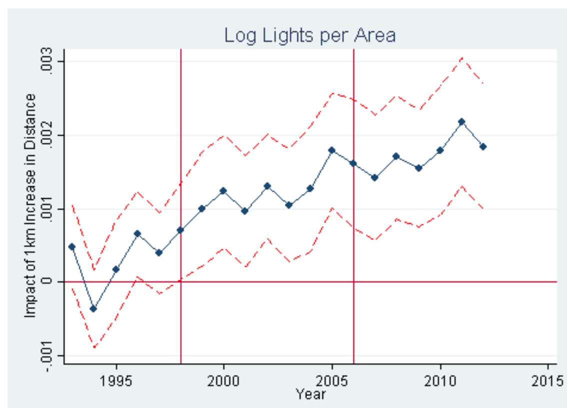

(a) District-level clusters




(b) Sub-district clusters



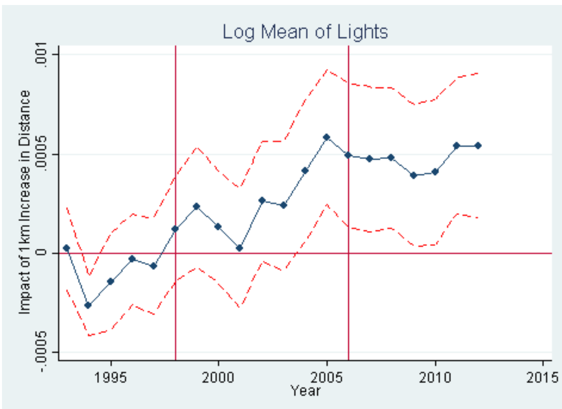

(c) Dropping 0s - Log(lights/area)



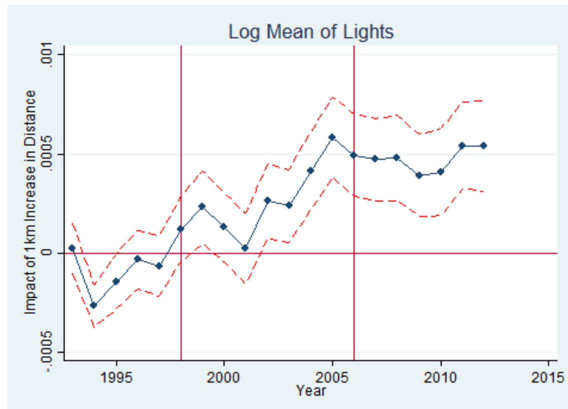

(d) Dropping extremely high light-density

Coefficients of change in impact relative to 1992. The impact of a 1km increase in distance in 1992 was a 0.00406. Vertical lines represent the phases of construction - 1999 is when the highways started being built. There were delays till 2001 when most work started, and 2006 is when most work was completed. The 'Dropping extremely high light-density' panel drops all sub-districts if they ever recorded a light pixel equal to the maximum possible value (63). This is 12% of the sample.

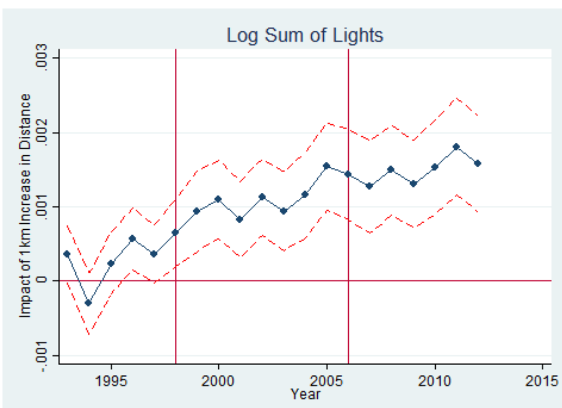
Figure 12: Robustness Checks: Different samples and specifications for: How the Impact of Distance on Mean Lights changes over time



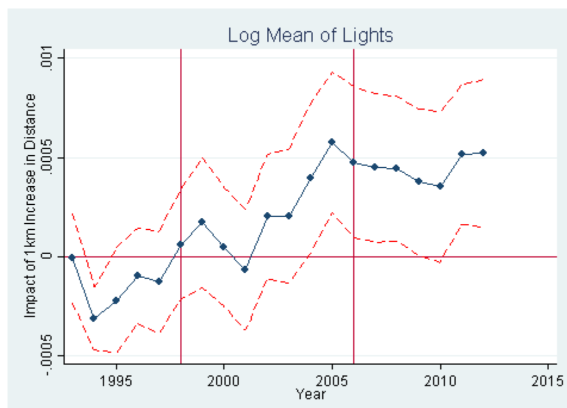

(a) District-level clusters




(b) Sub-district clusters



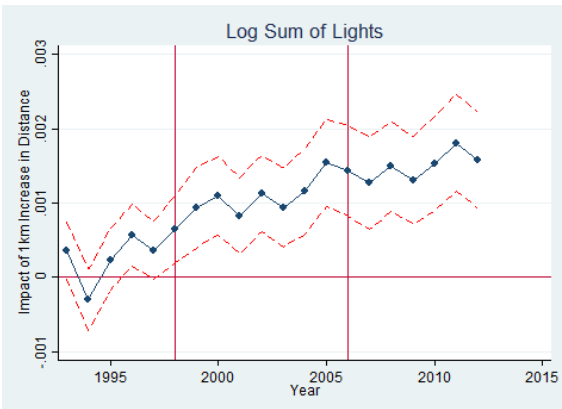

(c) District-level clusters



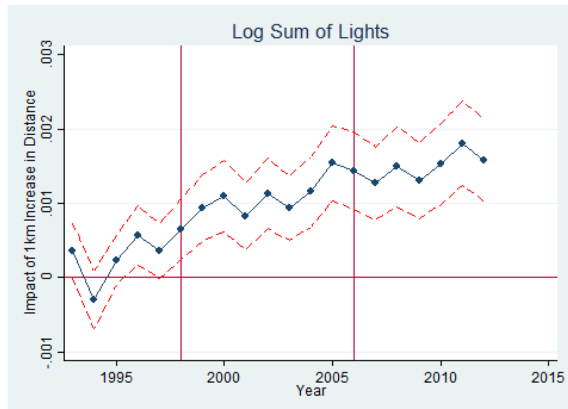

(d) Dropping extremely high light-density

Coefficients of change in impact relative to 1992. The impact of a 1km increase in distance in 1992 was a 0.00186. Vertical lines represent the phases of construction - 1999 is when the highways started being built. There were delays till 2001 when most work started, and 2006 is when most work was completed. The 'Dropping extremely high light-density' panel drops all sub-districts if they ever recorded a light pixel equal to the maximum possible value (63). This is 12% of the sample.

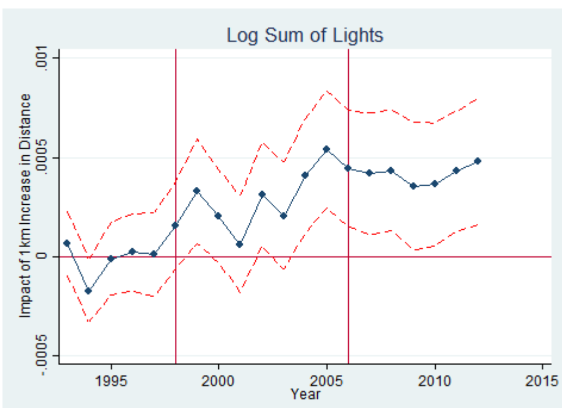
Figure 13: Robustness Checks: Different samples and specifications for: How the Impact of Distance on Sum of Density changes over time



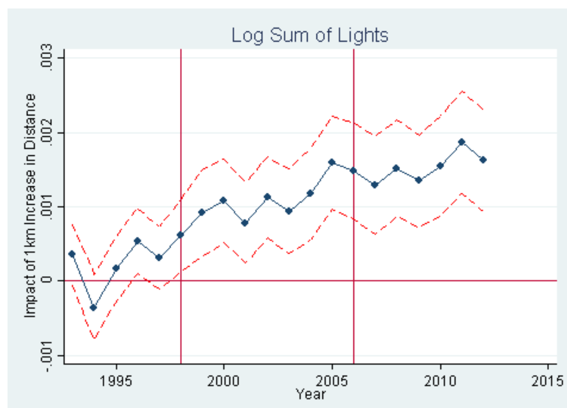

(a) District-level clusters




(b) Sub-district clusters




(c) Dropping 0s - Log(Sum of Lights)




(d) Dropping extremely high light-density

Coefficients of change in impact relative to 1992. The impact of a 1km increase in distance in 1992 was a 0.0038. Vertical lines represent the phases of construction - 1999 is when the highways started being built. There were delays till 2001 when most work started, and 2006 is when most work was completed. The 'Dropping extremely high light-density' panel drops all sub-districts if they ever recorded a light pixel equal to the maximum possible value (63). This is 12% of the sample.