HARVARD Kennedy School
JOHN F. KENNEDY SCHOOL OF GOVERNMENT

Using Satellite Imagery to Detect the Impacts of New Highways: An Application to India

Faculty Research Working Paper Series

Kathryn Baragwanath Vogel
University of California, San Diego

Gordon H. Hanson
Harvard Kennedy School

Amit Khandelwal
Yale University

Chen Liu
National University of Singapore

Hogeun Park
The World Bank

January 2024

RWP23-003

Visit the **HKS Faculty Research Working Paper Series** at: https://ken.sc/faculty-research-working-paper-series

The views expressed in the **HKS Faculty Research Working Paper Series** are those of the author(s) and do not necessarily reflect those of the John F. Kennedy School of Government or of Harvard University. Faculty Research Working Papers have not undergone formal review and approval. Such papers are included in this series to elicit feedback and to encourage debate on important public policy challenges. Copyright belongs to the author(s). Papers may be downloaded for personal use only.

Using Satellite Imagery to Detect the Impacts of New Highways: An Application to India
Kathryn Baragwanath Vogel, Gordon H. Hanson, Amit Khandelwal, Chen Liu, and Hogeun Park
NBER Working Paper No. 32047
January 2024
JEL No. O1,R1

ABSTRACT

This paper integrates daytime and nighttime satellite imagery into a spatial general-equilibrium model to evaluate the returns to investments in new motorways. Our approach has particular value in developing-country settings in which granular data on economic activity are scarce. To demonstrate our method, we use multi-spectral imagery—publicly available across the globe—to evaluate India's varied road construction projects in the early 2000s. Estimating the model requires only remotely-sensed data, while evaluating welfare impacts requires one year of population data, which are increasingly available through public sources. We find that India's road investments from this period improved aggregate welfare, particularly for the largest and smallest urban markets. The analysis further reveals that most welfare gains accrued within Indian districts, demonstrating the potential benefits of using of high spatial resolution of satellite images.

Kathryn Baragwanath Vogel
University of California, San Diego
9500 Gilman Dr
La Jolla, CA 92093
kbaragwa@ucsd.edu

Gordon H. Hanson
Harvard Kennedy School
Harvard University
79 John F. Kennedy St.
Cambridge, MA 02138
and NBER
gordon_hanson@hks.harvard.edu

Amit Khandelwal
Yale University
Economics Department, Jackson School
55 Hillhouse Avenue
New Haven, CT 06511
and NBER
amit.khandelwal@yale.edu

Chen Liu
National University of Singapore
Department of Economics
AS2 #0504, 1 Arts Link,
Singapore 117570
ecsliuc@nus.edu.sg

Hogeun Park
The World Bank
701 18th street NW
J6-601
Washington, DC 20433
hpark2@worldbank.org

For acknowledgments and financial disclosure information, see:
http://www.nber.org/papers/w32047

1 Introduction

This paper develops a method for applying daytime and nighttime satellite imagery to analyze the economic impacts of infrastructure projects. Our approach combines high-resolution imagery with analytical tools refined in the quantitative spatial economics literature (Redding and Rossi-Hansberg, 2017). Our goal is to provide a tractable framework for studying both the aggregate and highly local effects of infrastructure investments that would be applicable in contexts in which conventional data sources are limited or unavailable, for instance in many low-income countries.

Motivating our work is a multi-decade investment boom in transportation infrastructure in emerging and developing economies. Since the late 1990s, India has constructed 6,000 km of new expressways to connect its largest cities as part of the Golden Quadrilateral (GQ) network, while upgrading 19,000 km of minor roads to national highways and building an extensive network of rural roads (ADB, 2017). India's investment boom has recently accelerated, with the country currently adding 10,000 km of highways *each year* (The Economist, 2022). China's efforts are even more ambitious, involving over 40,000 km of new motorways within China and a Belt-Road Initiative that is creating new transport links throughout Asia and Africa. Since 2000, the World Bank has devoted one-fifth of its total lending to transportation infrastructure, with much of this funding going to low-income countries (Woetzel et al., 2017).

A major challenge in studying these investments is the coarseness of conventional data. To evaluate impacts within a nation, one is typically left to work with Census samples, which tend to appear with long lags and to limit geographic identifiers to municipalities or districts. Yet, in theory the impacts and distributional consequences of investments are highly localized (Allen and Arkolakis, 2022). The high-resolution imagery we exploit lets us evaluate infrastructure investments from the national to the granular level. As a demonstration case, we study the expansion of India's highways during the late 1990s and 2000s. Because this episode has been studied using conventional data (Datta, 2012; Ghani et al., 2014; Asturias et al., 2016; Alder, 2016), we are able to benchmark our results using satellite imagery with results using standard approaches.

Our analysis proceeds in four steps. First, we identify the footprints of local markets in India using the approach developed in Baragwanath et al. (2019), which applies an algorithm to define the geographic unit of analysis—a local market—as a cluster of adjacent pixels whose properties in daytime multi-spectral satellite imagery indicate the presence of builtup landcover. We measure economic activity in these granular markets using the intensity of nighttime lights captured by DMSP-OLS satellite sensors. The result is data on economic activity in 2001 and 2011 for 13,387 markets, which are either small towns unto themselves or the smaller cities that comprise a larger metropolitan area. Within the metropolitan borders of New Delhi, for example, we identify 579 local markets. In 2001, these markets collectively contained 34.8% of India's population. As a comparison, World Development Indicators reports India's 2001 urbanization rate at 27.9%.

Next, we measure travel times between each pair of markets by digitizing official road maps for India in 1996, prior to the highway construction, and 2011, by which point most construc-

tion from this phase was complete.[1] The maps, which identify five categories of intercity roadways from national highways to simple roads, reveal that India's recent highway construction was more extensive than the Golden Quadrilateral (GQ), the national artery connecting Delhi, Mumbai, Chennai, and Kolkata, and the primary focus of earlier studies. The GQ accounted for just one-fifth of new highway miles created by road upgrades or new construction between 1996 and 2011, and suggests that a comprehensive analysis of India's road expansion should include the full extent of the intercity road construction during this period. Using Dijkstra algorithm to calculate minimum travel time between each market pair, the new roads reduced travel times for 200-500km distances on average by about 1 hour.

Third, we demonstrate how to apply satellite imagery to an otherwise standard quantitative spatial model (Redding and Rossi-Hansberg, 2017; Allen and Arkolakis, 2023). The procedure requires calculating the market access function for each market in each year, using our data on market boundaries, nightlights, and bilateral travel times, and estimates of the trade cost elasticity and the elasticity of income with respect to nightlights (Henderson et al., 2012). We then estimate the impact of changes in market access on market income, accounting for endogenous placement of the roads using a least-cost-spanning-tree instrument (Faber, 2014). Finally, we use our parameter estimates to construct counterfactual changes in welfare based on alternative highway construction plans, which requires that we allocate population across our local markets in one year of our data. Although this last step requires non-satellite imagery, it is only for the purposes of providing welfare impacts. Moreover, population for at least one recent year are available for many low-income countries (e.g., through Ipums.org or Gridded Population of the World).

To preview our results, we find that improved market access from road construction over 1996 to 2011 increased the net present value of Indian GDP by 1.08%. Nearly all of this increase comes from the construction of the GQ. Although the expansion of non-GQ national highways increased real GDP by 0.77%, this gain was offset by construction costs, leaving a small negative net impact. Overall impacts are far from uniform across markets: initially larger and smaller markets experienced larger net gains than medium-size markets. Specifically, our findings reveal a welfare gain of 2.36% and 2.37% for markets situated in the 5th and 95th percentiles of nightlight intensities, respectively. However, for markets at the median, the welfare gain is 1.85%.

When we decompose the overall variance in welfare gains by level of spatial aggregation, we find that 47.6% of unequal impacts accrued to markets within Indian districts and the remaining 52.4% across the districts. Thus, the granular spatial resolution of satellite imagery reveals highly localized impacts of national infrastructure investment, a finding difficult to uncover in India's (spatially coarse) administrative data.

Our work contributes to an expanding literature that applies satellite imagery to the study of roads and economic growth. Storeygard (2016) uses nightlights to estimate impacts of road networks across 289 metropolitan regions in 15 Sub-Saharan Africa countries; and Jedwab and Storeygard (2020) examine spatial impacts of road construction in 3000 larger cities across 39 Sub-

[1]Surprisingly, we were unable to obtain digitized road maps for India in the 1990s or 2000s. We therefore digitized official road maps using PDFs of the decadal Survey of India, with help from a digital mapping company.

Saharan African countries. Relative to this work, our framework detects impacts for much finer geographic units and embeds the analysis in a quantitative spatial model; relative to broader literature in economics that uses satellite imagery (Henderson et al., 2012; Donaldson and Storeygard, 2016), we demonstrate how the *combination* of nightlight data, which captures the intensive margin of economic activity, with multi-spectral output from daytime images, which defines the extensive margin of markets, can be leveraged for analysis at very high spatial resolutions.[2]

Road construction in India has attracted substantial scholarly interest. In related work, Alder (2016) supplements administrative data with nightlights to study the impact of the GQ on Indian districts and finds that aggregate real GDP (net of construction and maintenance costs) would have been 0.8 percent lower in 2012 if the GQ had not been built.[3] Our framework, which arrives at similar aggregate impacts, reveals substantial within-district variation in gains. Other work finds positive impacts of Indian highways on manufacturing income and productivity (Datta, 2012; Asturias et al., 2019), and farm income Allen and Atkin (2022), while adjacent studies examine how the expansion of rural roads have affected Indian villages (Aggarwal, 2018; Asher and Novosad, 2020). Because we detect builtup towns and cities and not unilluminated hamlets, our approach is not well suited for the analysis of rural villages.[4]

Beyond the economic consequences of road construction in India, our work demonstrates how to apply satellite imagery to "off-the-shelf" spatial quantitative models, which have been developed and refined in recent years (Allen and Arkolakis, 2014; Ahlfeldt et al., 2015; Monte et al., 2018). In so doing, our goal is to provide a low-cost, tractable, and rigorous framework for spatial impact analysis in general equilibrium. It is difficult to imagine, for instance, studying transnational infrastructure investments, such as the Belt-Road Initiative, using administrative data. Moreover, unlike conventional statistical sources, satellite imagery is available on a nearly real-time basis, which allows one, for instance, to evaluate India's current road investments. Naturally, our approach cannot assess outcomes for individual units within a market, such as firms or consumers, but our results do suggest that remotely-sensed data are useful for assessing national policies in data-poor environments.

2 Data

2.1 Detecting Markets with Satellite Imagery

We use satellite imagery to construct markets by aggregating across adjoining pixels that meet given thresholds for economic activity. Baragwanath et al. (2019) develop a clustering algorithm that defines a boundary of a market as the set of contiguous or near-contiguous pixels that contain built-up economic activity as indicated by their day-time spectral properties. We apply this method to detect markets in India, and measure the intensive margin of economic activity of these

[2]The use of daytime satellite imagery in spatial economic analysis, which remains less common than the use of nightlights, was pioneered by Burchfield et al. (2006) in their study of urban sprawl.

[3]Alder (2016) studies 612 of India's 639 districts in 2001 compared to the 531 districts that contain detectible urban markets in our analysis.

[4]See Redding and Rossi-Hansberg (2017); Allen and Arkolakis (2023) for a discussion of the of quantitative spatial models and Coşar et al. (2021) for a recent application to Turkey.

markets through nightlight intensity. We summarize the procedure in Appendix A, and refer the reader to Baragwanath et al. (2019) for more details.

Market Boundaries

To define the boundary of markets, we rely on builtup landcover from the Moderate Resolution Imaging Spectroradiometer (MODIS) layer constructed by Friedl and Sulla-Menashe (2015), a publicly available dataset widely applied in remote sensing. We use their "Urban and Builtup" pixels to indicate builtup landcover for 2001 (Sulla-Menashe and Friedl, 2018). We assign adjoining pixels with builtup landcover to a cluster, and form the market boundary by combining proximate clusters whose boundaries are within 1km of each other. Finally, in order to capture possible changes in market footprints over time, we draw 0.5km buffers around the markets. This process creates 13,387 markets which cover 2.7% of India's surface area. Markets have mean and median areas of 6.5km^2 and 3.4km^2. The standard deviation across market sizes is 23.9km^2, and the largest market – New Delhi – has a boundary of 1,607.6km.

Baragwanath et al. (2019) also generate "super-markets" based on an coarser buffers that nest disaggregate markets defined at smaller buffers: 1km markets $\subset$ 4km markets $\subset$ 8km markets. These "super-markets" represent data-driven concentrations of economic activity, as opposed to those based on administrative district borders, which can be highly idiosyncratic in size. Relative to the 13,387 1km markets, we identify 7,539 4km markets, and 3,483 8km markets. As a reference, the 2011 India Census recognizes 6171 "towns" that were home to India's 377 million urban residents (31% of total population). Table A.1 provides summary statistics for the 1km, 4km and 8km markets. Our unit of analysis will be the 1km markets, but we will also be interested in assessing welfare gains within and across the coarser market buffers.

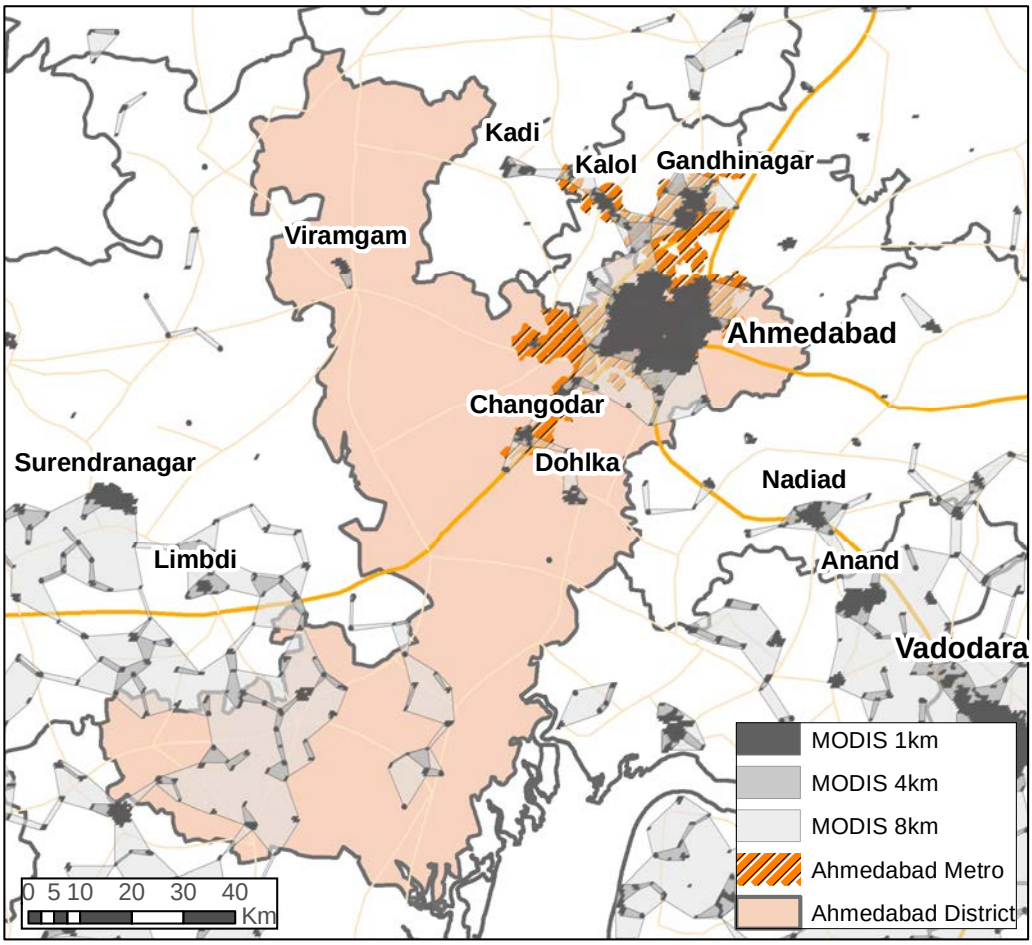
To illustrate our approach, Figure 1 shows local markets around the Ahmadabad district. In 2011, these markets contained 7.1 million inhabitants in the overall district (shown in grey) and 5.6 million inhabitants in the metro area (shown in cross-hatched yellow). Local markets are shown in black. The main city of Ahmedabad is evident, as are its principal satellite cities of Gandhinagar, Kalol, Kadi, Changodar, and Dohlka, and many smaller satellite towns and cities, some of which lie inside and some of which lie outside district boundaries. Our method detects 69 markets within the district of Ahmadabad and a total of 383 1km markets in the image.

Market Economic Activity

We measure economic activity within each market using the intensity of light emitted at night from the Defense Meteorological Satellite Program Operational Linescan System (DMSP-OLS), which detects visible and near-infrared light. DMSP-OLS pixels have a spatial resolution of 1km. For each market, we identify the pixels that overlap with its borders and aggregate light intensity across these pixels to create a measure of economic activity for the market. Appendix A discusses various approaches to address well-known issues with DMSP-OLS data: bottom-coding, top-coding, and blooming in nightlight data.

Although we do not directly observe GDP at the level of our 1km markets, we have GDP data

Figure 1: Satellite-Detected Markets in Ahmedabad Metro Area, 2001



Notes: The figure shows markets that lie in the metropolitan area and district of Ahmedabad. Major highways as of 2011 are shown in bold lines; minor roads as of 2011 are shown in faint lines. Within the figure, we observe 383 1km markets, 183 4km markets and 42 8km markets.

at the district level from the Indian Planning Commission in 2001,[5] which allows us to determine the relationship between nightlights and GDP for these geographic units. Figure A.1 plots nightlights against district GDP (left panel), revealing a strong positive correlation. We take this as supportive evidence of our use of nightlight intensity to proxy for local economic activity. To estimate a precise connection between nightlights and economic activity, we assume the following relationship between log GDP, $\log y_{rt}$, log nightlight intensity, $\log \mathrm{N}_{rt}$,

$$\log y_{rt} = C + \alpha \log \mathrm{N}_{rt} + \beta \mathrm{Land}_{rt} + v_{rt} \tag{1}$$

where v_{rt} is an i.i.d. disturbance associated with measurement error in the use of nightlights to

[5]India does not produce GDP figures at the district level or sub-district level on a regular basis. We could only obtain estimates for 2001. Of India's 639 districts in 2001, 531overlapped with one of our markets in 2001. Of these, we have GDP estimates for 326 districts.

capture GDP. We estimate equation (1) using 2001 data for 326 Indian districts by projecting log GDP on the district's log nightlight value, controlling for land area (which adjusts for top-coding of large cities; see Storeygard 2016). The regression coefficients are: 0.53 for log nightlights, -0.14 for log land area, and 5.59 for the constant, with an R^2 of 0.58. We then use the estimated coefficient to project GDP for each 1km market-year:

$$y_{rt} = \exp\left(5.59 + 0.53 \log \mathrm{N}_{rt} - 0.14 \mathrm{Land}_{rt}\right) \cdot \left(\tag{2}$$

Market Population

A key contribution of this paper is to demonstrate that satellite imagery can be used to estimate the elasticity of GDP to changes in market access. In order to conduct convert this elasticity to welfare impacts and to perform s, we further need an estimate of market population (in at least one year, either at the baseline or endline). This is because nightlights provide an estimate of GDP, whereas welfare in standard spatial models rests on changes in real income. Population data are therefore necessary to separately pin down market changes in population versus income. If an analyst was interested in short-run consequences of infrastructure investments and was comfortable with no (or very limited) labor mobility across space, one would not need population data. In environments where population estimates are not available, one could potentially use elasticities of nightlights with respect to population, as shown above, to predict a market's population size.

Census data are available at the town or village level, which are coarser geographic units than the markets we detect in satellite imagery. Following Baragwanath et al. (2019), to allocate India's population across our markets, we overlay the footprints of the markets with Census shapefiles. We assign any town or village that overlaps with one of our markets to that market in its entirety. Where a town/village overlaps with two or more markets, we divide the village/town population equally between the markets that overlap with it. Using this method, we find that 32.7% of India's total population in 2011 were assigned to one of our 1km markets.[6] Table A.1 reports that the average number of people living inside a 1km market increased from 26.8 thousand in 2001 to 29.6 thousand in 2011. See Appendix A for details.

2.2 India's Highway System

2.2.1 National Highway Development Project

The Golden Quadrilateral was the centerpiece of India's National Highways Development Project (NHDP). Commencing in 2001 and reaching completion by 2012, the GQ upgraded existing roads to four or six-lane expressways in a circular route connecting New Delhi, Mumbai, Chennai, and Kolkata, at a total cost of $5.4 billion (Ghani et al., 2014). The NHDP also upgraded many more roads to national highways and constructed other roadways de novo. To detect highway upgrades and new road construction, we digitize road maps from the Survey of India, which prior to 2015 mapped intercity roads for all of India on a roughly decadal basis. We digitize maps for 1996 and 2011, which span a period from five years before the NHDP to just before its end.

[6] World Development Indicators report India's urbanization rate was 31.3% in 2011.

Appendix B provides more details explaining the road digitization process and validation.

Figure 2 shows the results of our digitization exercise. The GQ appears in red, new highways (whether upgrades or new construction) appear as dashed lines, and existing national highways appear in black. Other roads appear as faint lines. The newly constructed portions of the grid are more dense in India's north, east, and south, and less dense in the center. The figure also shows urban agglomerations with populations above 300,000. To a first approximation, the existing grid connected India's largest cities (populations of 1 million or more), while the newly constructed portions of the grid connected smaller cities. As shown in Figure 1, within an individual region (in this case, Ahmedabad), there is considerable heterogeneity in access to roads according to market size. Figure 2 summarizes our measures of India's highway and intercity road construction. The GQ added 5,714km of new expressways, and national highways expanded by 19,250km. Although the total increase in road mileage between 1996 and 2011 was relatively modest—the 22,371km of additional roads represented only a 10.0% increase in total road length—the improvement in road quality was significant. India's length of highway-quality roads increased by a substantial 62.3% over the period.

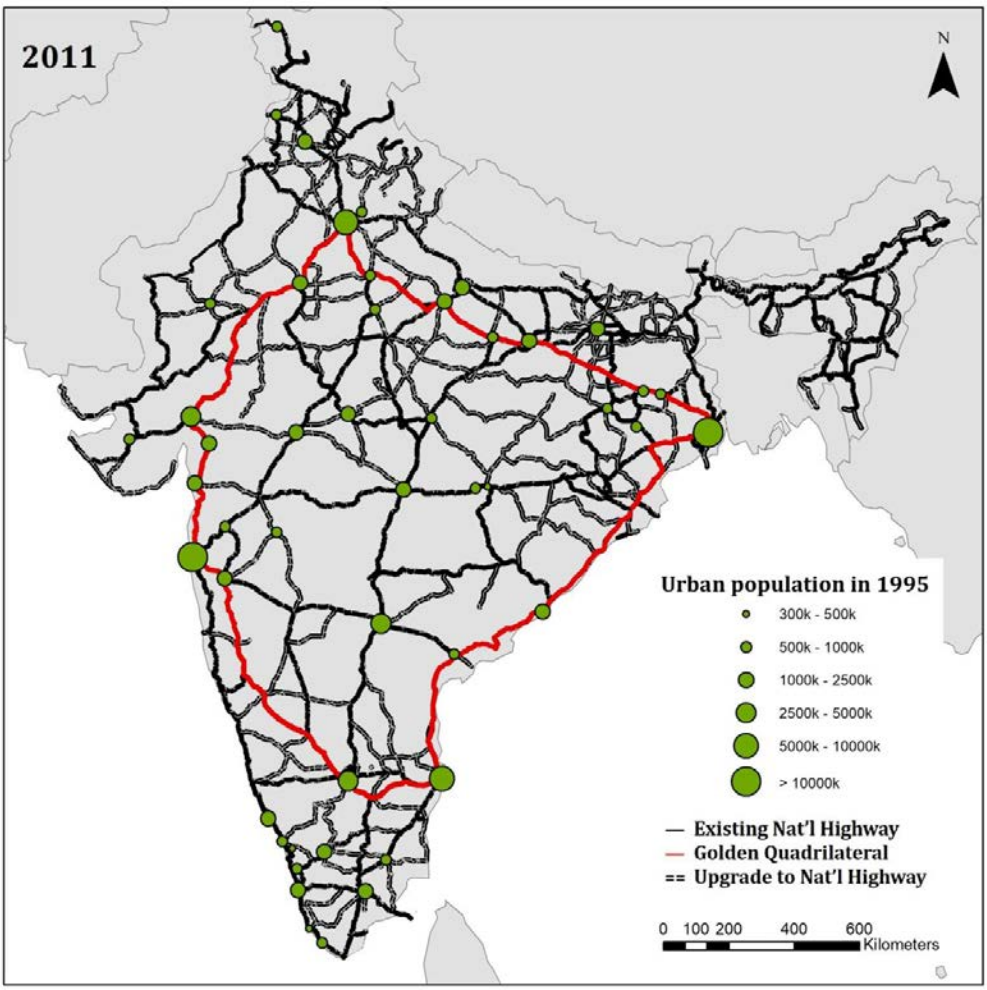
2.2.2 Travel Time between Markets

The increase in road miles understates the full benefits of the construction through improved speed, as new highways have two to six lanes whereas the roads being upgraded had one to two lanes. To translate road construction into changes into transport costs, we estimate changes in the time required to travel between each pair of markets in India (see Appendix B). Travel may occur along a portion of the GQ (four to six lanes, paved), a national highway (two to four lanes, paved), and lesser road types (one to two lanes, paved, semi-paved, or unpaved). As a starting point, we use estimates in Allen and Atkin (2016), which indicate average speeds of 20mph for the lowest-quality roads (roads where delays may occur) and 60mph for top-quality roads (GQ). For intermediate roadways, we assume speeds of 40mph for national highways, 30mph for all-weather motorable roads, and 20mph for fair-weather motorable roads.

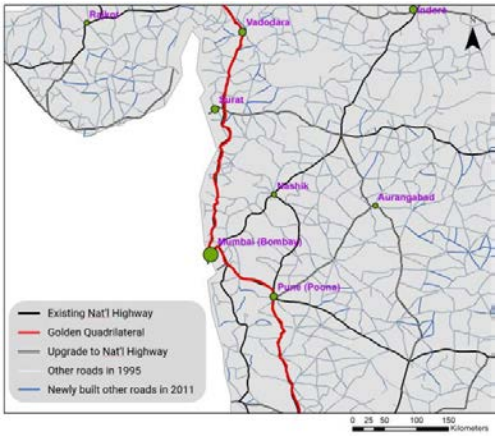
Because the majority of markets are small, most of them do not have an identifiable road going directly through them. The median market in our data is 2.4km away from an identifiable road in 1996 (interquartile range 3.9km) and 2.2km away from an identifiable road in 2011 (interquartile range 3.4km). Although we cannot detect connecting roads to these identifiable roads, we assume that they exist and that they follow the minimum distance from the market centroid to the nearest road, assuming a speed of 20mph along these roads.

Given assumed travel speeds, we estimate the minimum travel time between each pair of markets by implementing the Dijkstra algorithm (Dijkstra, 1959), which finds the shortest path from a starting node to a target node through a weighted graph, which is a data structure composed on a set of nodes and edges (Donaldson and Hornbeck, 2016; Donaldson, 2018) (see Appendix B). Between 1996-2011, Table B.1 indicates that travel times between markets 50-200km away fell by about 0.5 hours and travel times between markets 200-500km fell, on average, by 1 hour while travel times between markets more than 500km away from each other fell by more than 3 hours.

Figure 2: Highway Construction in India, 1996 to 2011



Road Construction near Mumbai, 1996 to 2011



Total Road Length by Road Type, India 1996 to 2011

Category of Road	1996	2011
Golden Quadrilateral	--	5,714
National Highway	29,422	49,172
All Weather Motorable	165,617	164,022
Motorable in Fair Weather	15,353	13,067
Where delay may occur	2,838	3,625
Total	213,230	235,601

Note: Table reports road construction estimates from digitized road maps from the 1996 and 2011 Survey of India.

Notes: The figure presents: (i) a map of India in 2011, in which lines represent highways (solid black for existing national highways, solid red for the Golden Quadrileral, dashed black for upgrades to national highways), and green dots represent major cities (dot size is proportional to city population in 1995), (ii) a close up of road construction near Mumbai, where white lines represent non-highway roads in 1996 and blue lines represent newly built such roads in 2011, and (iii) a table showing total road length by type of road in India in 1996 and 2011, in kilometers.

3 Quantitative Framework

We implement a standard spatial general equilibrium model, which embodies the market-access approach (see, e.g., Donaldson and Hornbeck, 2016; Bartelme, 2018). As noted by Allen and Arkolakis (2023), these spatial models are characterized by a system of regional consumer demand and labor supply equations. Consumer demand has a CES structure, and labor supply derives from workers' idiosyncratic preferences across locations. We present estimating equations and welfare formulas based on this framework. For complete details, see Appendix C.

3.1 Equilibrium Conditions

The framework delivers two expressions that link log changes in wages (w) and population (L) to log changes in a region r's market access:

$$\hat{w}_{rt} = \varepsilon_w \hat{\Phi}_{rt} + \hat{\chi}^w_{rt} \qquad \hat{L}_{rt} = \varepsilon_l \hat{\Phi}_{rt} + \hat{\chi}^l_{rt} \tag{3}$$

where $\hat{x}_t \equiv \ln x_t - \ln x_{t-1}$ denotes log changes, ε_w and ε_l represent the reduced-form elasticities, each of which depends on the trade elasticity parameter θ and the inverse labor supply elasticity μ.[7] $\hat{\chi}^w_{rt}$ and $\hat{\chi}^l_{rt}$ are functions of i.i.d. shocks to regional amenities and productivity.

The change in market access $\hat{\Phi}_{rt}$ is determined by:

$$\hat{\Phi}_{rt} = \ln \left(\sum_n \frac{w_{nt} L_{nt}}{\Phi_{nt}} \tau^{-\theta}_{rn,t} \right) - \ln \left(\sum_n \frac{w_{n,t-1} L_{n,t-1}}{\Phi_{n,t-1}} \tau^{-\theta}_{rn,t-1} \right) \tag{5}$$

where $\tau_{rn,t}$ is the symmetric iceberg trade cost between regions r and n at time t.[8] The level of market access at year t is solved from the following system of R equations,

$$\Phi_{rt} = \sum_n \frac{w_{nt} L_{nt} \tau^{-\theta}_{rnt}}{\Phi_{nt}}. \tag{6}$$

In order to compute market access for each region, equation (6) reveals that we need data for $w_{nt}L_{nt}$, which is local GDP predicted as a function of nighttime lights, and the iceberg trade cost composite, $\tau^{-\theta}_{rnt}$, which following the literature we assume is a function of travel times (see Redding and Rossi-Hansberg, 2017; Monte et al., 2018; Allen and Arkolakis, 2022):

$$\tau^{-\theta}_{rn} = \left(\text{travel time}_{rn} + \frac{\kappa \times (d_{rn} + d_{nr})}{\text{speed}} \right)^{-\phi}, \tag{7}$$

where ϕ measures the trade elasticity to travel time. We set $\phi = 1.5$, following Redding and Rossi-Hansberg (2017). The variables d_{rn} and d_{nr} are the last miles traveled, from r to n and from n to r, to connect markets to observed roadways. We set $\kappa = 0.000621$ which converts meters into miles.

[7] Specifically,

$$\varepsilon_w = \frac{\mu\theta - 1}{\theta[\mu\theta + \mu + 1]}, \qquad \varepsilon_l = \frac{2\theta + 1}{\theta[\mu\theta + \mu + 1]} \tag{4}$$

[8] Given symmetric trade costs, consumer market access and firm market accesses are equal up to a factor of proportion, which simplifies the analysis (Donaldson and Hornbeck, 2016).

We set last-mile travel speeds equal to 20mph.

3.2 The Estimating Equation

Assuming that GDP equals labor income, $\hat{y}_{rt} \equiv \hat{w}_{rt} + \hat{L}_{rt}$, equation (3) implies

$$\hat{y}_{rt} = \delta \hat{\Phi}_{rt} + \hat{\Gamma}_{rt}. \tag{8}$$

where $\delta \equiv (\varepsilon_w + \varepsilon_l)$ and $\hat{\Gamma}_{rt} = \hat{\chi}^w_{rt} + \hat{\chi}^l_{rt} + \hat{v}_{rt}$. For the dependent variable, equation (1) implies that the change in GDP can be measured as

$$\hat{y}_{rt} = \alpha \hat{N}_{rt} + \hat{v}_{rt}, \tag{9}$$

where we set $\alpha = 0.53$ based on estimation results for equation (2). The independent variable, the change in market access, $\hat{\Phi}_{rt}$, is derived by substituting y_{rt} for $w_{nt}L_{nt}$ in equation (6), solving the system of R equations in R unknowns in (6) for Φ_r at both time t and time $t-1$, and then taking the time difference between these two values.

3.3 Identification

The change in market access contains GDP and trade costs in period t, which embody endogenous responses to shocks between $t-1$ and t. This requires an instrument to identify δ in equation (8) (Allen and Arkolakis, 2023). We follow Faber (2014) by constructing an instrument based on a least-cost approach to connect markets:

$$\hat{\Phi}^{\text{mst}}_{rt} = \ln \left(\sum_{r'} \frac{y_{n,t-1}}{\Phi^{mst}_{nt}} \tau^{\text{mst}}_{rn,t} \left(\Omega_{\prime t-1}\right)^{-\theta} \right) - \ln \left(\sum_{n} \frac{y_{n,t-1}}{\Phi_{n,t-1}} \tau^{-\theta}_{rn,t-1} \right) \Bigg(\tag{10}$$

This instrument uses GDP at time $t-1$ to calculate market access in both time periods. Additionally, bilateral trade costs, $\tau^{mst}_{rn,t}\left(\Omega_{\prime t-1}\right)$, are constructed by a least-cost spanning tree algorithm to connect India's 180 largest cities using information as of time $t-1$, denoted as $\Omega_{\prime t-1}$. See Appendix B for details. Φ^{mst}_{nt} is given by the solution to the $R \times R$ system of equations,

$$\Phi^{\text{mst}}_{rt} = \sum_{n} \frac{y_{n,t-1}}{\Phi^{\text{mst}}_{n,t}} \tau^{\text{mst}}_{rn,t} \left(\Omega_{\prime t-1}\right)^{-\theta}. \tag{11}$$

We note that estimating (8) only requires information from satellite imagery—the market boundaries that determine r, and the value of nightlights within r that proxy for y_{rt} and $y_{r,t-1}$—and the changes in travel times across markets. Thus, if one were interested solely in how changes in market access affect local GDP, no additional sources of data would be required.

3.4 Equilibrium Changes in Welfare

To perform counterfactual exercises, we solve the model in changes with 2011 as the baseline equilibrium (Dekle et al., 2008).[9] We use $\hat{X}$ to denote the ratio between the counterfactual and

[9]Regarding whether to use the first period or last period for the baseline equilibrium, the existing literature is mixed. Caliendo and Parro (2015) calibrate the model to the beginning period (1996 in our case), whereas Adao,

the actual economy in 2011 of variable X. Since $\hat{X}$ no longer represents changes over time, we suppress the subscript t. These exercises call for additional data on the population of each market, which we obtain as described above.

Given data on local population L_r, the market access value Φ_r in 2011, our measure of market-level GDP, and counterfactual values of trade costs τ'_{rn}, we solve the counterfactual market access in changes $\hat{\Phi}_r$ from the following system of equations:

$$\hat{\Phi}_r \Phi_r = \left[\sum_n \hat{\Phi}_n^{\frac{1}{\mu}(\varepsilon_w + \frac{1}{\theta})} \frac{L_n}{L} \right]^{-\frac{\theta}{\theta+1}} \left[\sum_n \frac{\hat{\Phi}_n^{\varepsilon_w + \varepsilon_l} y_n}{\hat{\Phi}_n \Phi_n} \tau'^{-\theta}_{rn} \right], \tag{12}$$

Once we obtain $\hat{\Phi}_r$, we calculate the changes in local wage and population as

$$\hat{w}_r = \hat{\Phi}_r^{\varepsilon_w} \left[\sum_n \hat{\Phi}_n^{\frac{1}{\mu}(\varepsilon_w + \frac{1}{\theta})} \frac{L_n}{L} \right]^{\frac{1}{\theta+1}}, \qquad \hat{L}_r = \hat{\Phi}_r^{\varepsilon_l} \left[\sum_n \hat{\Phi}_n^{\frac{1}{\mu}(\varepsilon_w + \frac{1}{\theta})} \frac{L_n}{L} \right]^{-1}. \tag{13}$$

The changes in the local price index and welfare are given by:

$$\hat{P}_r = \hat{\Phi}_r^{-\frac{1}{\theta}}, \qquad \widehat{\text{Welfare}}_r = \frac{\hat{w}_r}{\hat{P}_r}. \tag{14}$$

Solving the model requires values for the elasticity on iceberg trade costs (θ) and the inverse labor elasticity (μ). Given our estimate of δ and assuming $\theta = 8$ (see Donaldson and Hornbeck, 2016; Allen and Arkolakis, 2022), we can recover the inverse labor supply elasticity μ as (see Appendix C.7 for details):

$$\mu = \frac{\varepsilon_w}{\delta - \varepsilon_w} \times \frac{2-\delta}{1-\delta}, \quad \text{where } \varepsilon_w = \frac{1-\delta}{\theta}. \tag{15}$$

4 Empirical Analysis

In this section, we first estimate δ from (8) using $\hat{\Phi}_{rt}^{\text{mst}}$ as an instrument. We then use our model to evaluate the cost-benefit and distributional consequences of India's highway construction.

4.1 Estimation Results for Reduced-Form Parameters

Panel A of Table 1 presents estimation results for equation (8). Column 1 reports the OLS estimate, δ =0.31 (s.e. 0.07). This estimate rises to 0.82 (s.e. 0.37) when using IV, as seen in column 2.[10] The IV results thus indicate an elasticity of GDP with respect to market access of 0.82. All models control for 8km-buffer fixed effects and initial log GDP to account for mean reversion. We cluster the standard errors at the 4km-buffer level. With reference to the literature, Alder (2016) estimates $\delta = 0.6$ (s.e. 0.23), when using district-level nightlight data to measure GDP.

Costinot, and Donaldson (2017) calibrate the model to the end period (2011 in our case). We chose 2011 for the baseline equilibrium because we have local population data for that year.

[10] To verify that our estimation results are not the by-product of longer-term trends in India's regional urban growth, we conduct a falsification exercise by regressing past nightlight changes (between 1994 and 1999) on changes in market access for 2001 to 2011. Column 3 shows a small and statistically insignificant elasticity: δ =0.40 (s.e. 0.56).

Table 1: Baseline Estimates and Cost-benefit Calculations

	Panel A: Parameter Estimates		
	(1)	(2)	(3)
	OLS	2SLS	Pre-trend
δ	0.31	0.82	0.40
	(0.07)	(0.37)	(0.56)
Observations	11,153	11,153	9,295
R^2	0.36	0.08	0.06
F		446.1	367.5
Fixed-effects	8km markets	8km markets	8km markets
	All Roads	**GQ**	**NH**
NPV Increase in GDP	2.32	1.47	0.77
NPV of Construction Costs	0.39	0.13	0.26
NPV of Maintenance Costs	0.86	0.28	0.57
NPV of Total Costs	1.24	0.40	0.83
Increase in the NPV of GDP	1.08	1.06	-0.06

Notes: The table presents results from the estimation and cost-benefit analysis. Panel A presents the baseline estimates for the effects of changes in market access (due to improvements in the road network) on GDP as measured by nightlights. Columns 1 and 2 of panel A show the elasticity of GDP with respect to structural market access (in changes between 1996 and 2011), using OLS and 2SLS, respectively. Column 3 of panel A presents the pre-trend elasticity of GDP on structural market access, based on changes between 1994 and 1999. The fixed effects of 8km-markets are controlled for all models. The instrumental variable for the 2SLS estimation is constructed using 180 cities in the least-cost spanning tree algorithm. Standard errors (clustered at 4km markets) are in parenthesis. Panel B presents a cost-benefit analysis of the improvement in the road network, which disentangles the increase in GDP and the costs of road construction and maintenance. We differentiate the effects of all roads (panel B, column 1), GQ construction only (panel B, column 2), and national-highway construction only (panel B, column 3). GQ counterfactual: keep the road structure as in 2011, but reduce lower the speed of GQ roads to 40mph (speed of national highways); NH counterfactual: keep the road structure as in 2011, but reduce the speed of newly built national highways to 20mph (speed of other roads).All values are the NPV values computed over a 50-year time horizon using a 5% discount rate, and then normalized by the NPV of the 2011 GDP.

4.2 Welfare Analysis

Given parameter values of $\theta = 8$ and $\delta = 0.82$, equation (15) implies $\mu = 0.18$. We now use the parameter estimates and model structure to evaluate the impact of India's highway build-up over 1996 to 2011. We emphasize three results. First, we evaluate the costs and benefits of the full highway construction enterprise. We calculate the net present value (NPV) of the increase in GDP from the entire road-building program. We also separately estimate the effects of the GQ construction and the national highway expansion. Second, we study the spatial distribution of these effects, analyzing which types of markets benefited more from road construction. Third, we decompose the spatial variation in welfare changes into between and within-district components to illustrate the benefits of working with granular satellite imagery.

Our benchmark counterfactual is an environment in which the entire transportation network remains the same as it was in 1996. To implement this, we take our model to the data in 2011 and

evaluate highway travel times at the level of those for 1996.

4.2.1 Cost-Benefit of India's Highway Investment

To annualize construction costs, we conduct cost-benefit analysis in terms of NPV (Tsivanidis, 2023). NPV values are computed over a 50-year time horizon using a 5% discount rate. Panel B of Table 1 displays the NPV estimates for gross GDP, costs, and net GDP. Since all NPVs are normalized by the NPV of GDP in 2011, they can be interpreted as the annualized percentage of GDP.

We report results comparing the actual economy with the observed road network in 2011, our benchmark scenario, to three different counterfactual scenarios. Column 1 reports the impacts of the entire road network build-up, which we obtain by comparing the benchmark scenario to an economy without any road construction. Column 2 isolates the effects of the GQ by comparing our benchmark scenario to a counterfactual economy with the 2011 road network, except for the GQ upgraded sections (i.e., GQ speeds equal regular highway speeds). Column 3 reports the effects of only expanding national highways, which we obtain by comparing our benchmark scenario to a counterfactual economy with the 2011 road network that has the upgraded national highway sections removed (i.e., national highway speeds equal regular motorway speeds).

The first row of Panel B of Table 1 reports that India's highway investment during this period increased annual gross GDP by 2.32%. Column 2 indicates that the contribution of the GQ to overall welfare is 1.47%, or 62.9% of the overall increase. The expansion of non-GQ national highways accounts for a 0.77% increase.[11] Row 2 shows the NPV of the total costs, which has two parts. The first is the construction cost.[12] Table B.2 details the overall construction costs in 1996 USD, which translates to an NPV of 129.93 billion USD in 2011. The annual maintenance costs are assumed to be 12% of the total construction costs (Allen and Arkolakis, 2014); row 4 reports the associated NPV. Row 5 reports the NPV of GDP net of the costs, equaling the difference between rows 1 and 2. We estimate that the GQ delivered a net GDP increase of 1.06%, but national highways, by contrast, do not appear to have delivered a net benefit, which is surprising given the large expansion in this road type reported in Figure 2.

Our results are similar to the gross and net GDP gains of 1.30% and 0.80%, respectively, that (Alder, 2016) estimates for the GQ. This suggests that granular high-resolution satellite data can similar aggregate estimates uncovered from district-level nightlight data.

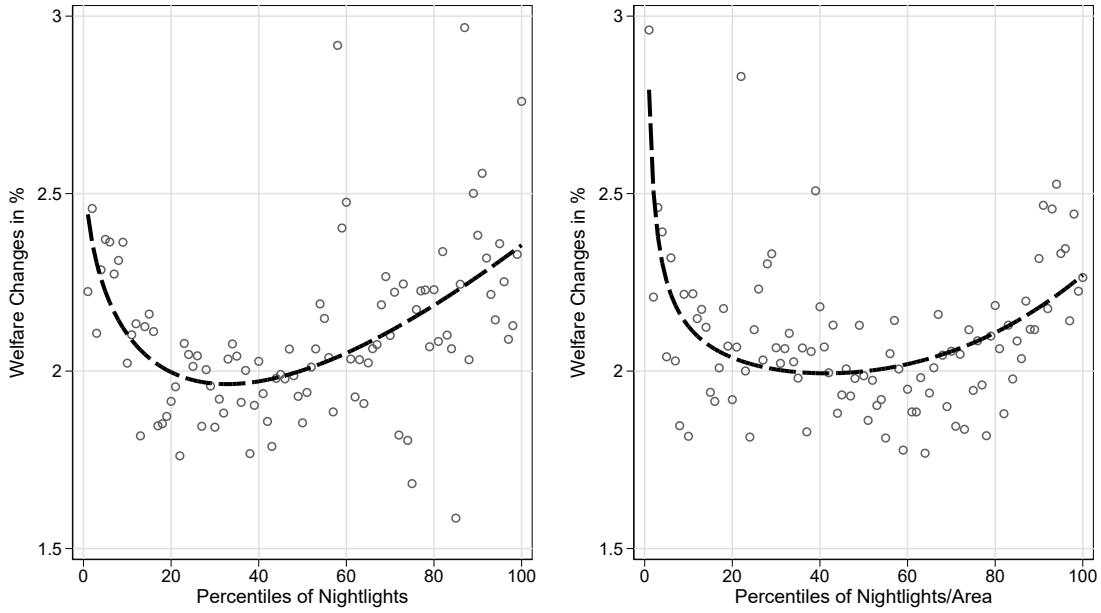
4.2.2 Welfare Impacts By Market Size

In Figure 3, we analyze heterogeneity in welfare gains across markets. Specifically, we examine the relationship between the percentile of nightlight values (a proxy for GDP) and the welfare impacts of road upgrades. The left panel of the figure displays the binscatter plot along with a polynomial best-fit line. The plot reveals a U-shaped pattern, indicating that the smallest and largest markets experienced greater welfare gains from road investments compared to medium-

[11] Note that, because the model is non-linear, the joint effect of GQ and NH does not necessarily equal the sum of the two individual effects. The differences reflect the interactive effects between the two, which are small.

[12] The construction cost is from the World Bank's Project Appraisal Document and the Implementation Completion and Results Report. The average construction cost is calculated based on data from 11 transportation projects that were executed in India between 2008 and 2017.

Figure 3: Welfare impacts by Market Characteristics



Notes: This figure shows the welfare impacts of road construction by market characteristics. Each panel has 100 bins, one for each percentile. Circles represent average welfare changes by percentile of either nightlight (left) or nightlight/area (right). Dashed line is the polynomial fit.

sized markets. The ratio of welfare gains between the 95th percentile and the 50th percentile is 1.27, which is similar to the 1.28 ratio between the 5th percentile and the 50th percentile.

The right panel examines the relationship between welfare gains and the percentile of nightlight intensity per land area. We again see a U-shaped pattern that is steep at the bottom end of the distribution, indicating that the less developed markets (in terms of nightlight/area) experienced greater welfare gains from road investments.

4.2.3 Decomposing Welfare Changes

Figure 3 illustrates large heterogeneity in the welfare impacts of road construction across the 1km markets. In Table 2, we decompose the variance between and across geographic units. Panel A performs this decomposition by allocating the 1km markets into standard Indian administrative units—districts and states—and decomposes aggregate welfare impacts across and within units for each level. Specifically, we decompose the log variance of welfare impacts into the sum of between and within-district components as follows:

$$\text{Total} = \text{Between} + \text{Within}. \tag{16}$$

Table 2: Decomposing Welfare Changes

	A. States and Districts
Between state	20.1%
Within state	79.9%
Between district	32.4%
Within district	47.6%

	B. Super-markets
Between 8km markets	70.8%
Within 8km markets	29.2%
Between 4km markets	19.7%
Within 4km markets	9.5%
Between 2km markets	5.2%
Within 2km markets	4.2%

Notes: The table presents a decomposition of welfare between and within districts. The left panel decomposes aggregate welfare across and within Indian states and districts. The right panel performs an alternative decomposition using the 8km, 4km, and 2km boundaries detected from satellite imagery.

where

$$\text{Total} = \frac{1}{N}\sum_i (w_i - \bar{w})^2, \quad \text{Between} = \frac{1}{N}\sum_d N_d(\bar{w}_d - \bar{w})^2 \tag{17}$$

where w_i is the log welfare impact of market i, $\bar{w}$ is the simple average of w_i across all markets, and $\bar{w}_d$ is the simple average over markets within administrative unit d (states or districts). N_d is the number of markets within d, and N is the total number of markets. The within component is the difference between the two.

The first row reports that 20.1% of the overall welfare gains from the roads accrue across states, and the second row shows 79.9% accrue across markets *within* states. We further decompose the within-state variation using district administrative boundaries. Here, we continue to see substantial variation within the component. While 32.4% of the overall variance is driven by differences across districts (within-state), 47.6% of the overall variance comes from welfare differences across 1km markets within districts. Analysis performed at the district level may thus smooth over nearly half of the spatial variation in welfare impacts.

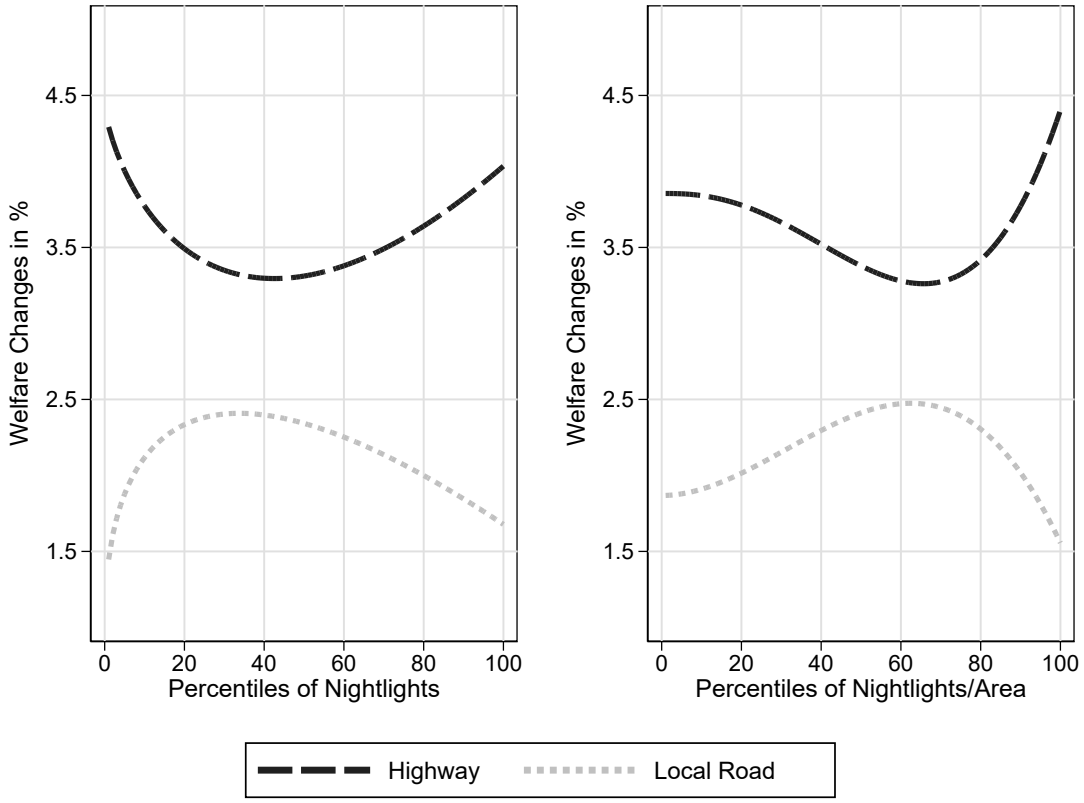
Panel B of Table 2 reports an alternative decomposition that uses the delineation of 8km, 4km, and 2km markets. Recall that the satellite imagery detect 3,483 8km markets in India, and the decomposition shows that variation across these markets accounts for 70.8% of aggregate welfare. This leaves the remaining 29.2% variation within 8km markets, again indicating that large heterogeneity in impacts of the roads at the sub-district level. Of this 29.2% variation, 19.7% accrues to the within-8km but across 4km markets. This panel again suggests that administrative boundaries, which do not overlap directly with transactions in urban environments, may miss important variations in the impacts of market access.

4.3 Further Upgrade of Highway versus Local Roads

Finally, we consider the potential impact of two further road upgrades. First, we examine an upgrade of national highways that maintains the existing road structure from 2011 but increases the speed to 60mph (equivalent to the speed of GQ roads). Second, we explore an upgrade of local roads that keeps the same road structure as in 2011 but improves the speed to 40mph (equivalent to the speed of national highways).

Figure 4 presents the polynomial best-fit lines, with the black dashed line representing the national highways and the light dotted line representing the local roads. The left panel displays the welfare impacts against the nightlight values. Upgrading national highways would lead to a substantial welfare improvement (over 3%) and benefit the smallest and largest markets the most. The ratio of welfare gains between the 95th percentile and the 50th percentile is 1.08, as is the ratio between the 5th percentile and the 50th percentile. On the other hand, upgrading local roads primarily benefits medium-sized markets, and the impacts are small in magnitude. Moving to the right panel, we examine the unequal impacts across markets with varying nightlight/land areas, finding similar patterns for further upgrades to highways and local roads.

Figure 4: Welfare Impacts by Market Characteristics, Highway vs Local Roads



Notes: The figure shows the welfare impacts of road construction by market characteristics. Dark dashed lines are the polynomial fit for changes due to highway construction. Light dashed lines are the polynomial fit for changes due to local road construction.

5 Conclusion

Many developing country policymakers appear to take it as an article of faith that building new highways will raise national income per capita. They act on these beliefs by dedicating substantial public funds to transportation infrastructure, often with the encouragement of international financial institutions and countries eager to finance such investments. With standard administrative

data, analyzing the economic impacts of highway construction in the typical lower income country would be infeasible until a decade or more after project completion. We provide a method that permits rapid and low-cost evaluation of infrastructure investments. Analysts need only freely available satellite imagery, to evaluate impacts on GDP, and a baseline or endline year of population data, to perform additional welfare analysis. With the ongoing digitalization of road maps globally (e.g., OpenStreetMap), our method should have wide applicability.

References

ADAO, R., A. COSTINOT, AND D. DONALDSON (2017): "Nonparametric counterfactual predictions in neoclassical models of international trade," *American Economic Review*, 107, 633–89.

ADB (2017): *Meeting Asia's Infrastructure Needs*, Asian Development Bank.

AGGARWAL, S. (2018): "Do rural roads create pathways out of poverty? Evidence from India," *Journal of Development Economics*, 133, 375–395.

AHLFELDT, G. M., S. J. REDDING, D. M. STURM, AND N. WOLF (2015): "The Economics of Density: Evidence From the Berlin Wall," *Econometrica*, 83, 2127–2189.

ALDER, S. (2016): "Chinese roads in India: The effect of transport infrastructure on economic development," *Available at SSRN 2856050*.

ALLEN, T. AND C. ARKOLAKIS (2014): "Trade and the Topography of the Spatial Economy *," *The Quarterly Journal of Economics*, 129, 1085–1140.

——— (2022): "The welfare effects of transportation infrastructure improvements," *The Review of Economic Studies*, 89, 2911–2957.

——— (2023): "Economic Activity across Space: A Supply and Demand Approach," *Journal of Economic Perspectives*, 37, 3–28.

ALLEN, T. AND D. ATKIN (2016): "Volatility and the Gains from Trade," Tech. rep., National Bureau of Economic Research.

——— (2022): "Volatility and the Gains From Trade," *Econometrica*, 90, 2053–2092.

ANDERSON, J. E. (1979): "A theoretical foundation for the gravity equation," *The American economic review*, 69, 106–116.

ARKOLAKIS, C., A. COSTINOT, AND A. RODRIGUEZ-CLARE (2012): "New Trade Models, Same Old Gains?" *American Economic Review*, 102, 94–130.

ASHER, S. AND P. NOVOSAD (2020): "Rural Roads and Local Economic Development," *American Economic Review*, 110, 797–823.

ASTURIAS, J., M. GARCÍA-SANTANA, AND R. RAMOS (2016): "Competition and the welfare gains from transportation infrastructure: Evidence from the Golden Quadrilateral of India," *Journal of the European Economic Association*.

ASTURIAS, J., M. GARCÃA-SANTANA, AND R. RAMOS (2019): "Competition and the Welfare Gains from Transportation Infrastructure: Evidence from the Golden Quadrilateral of India," *Journal of the European Economic Association*, 17, 1881–1940.

BARAGWANATH, K., R. GOLDBLATT, G. HANSON, AND A. K. KHANDELWAL (2019): "Detecting Urban Markets with Satellite Imagery: An Application to India," *Journal of Urban Economics*.

BARTELME, D. (2018): "Trade costs and economic geography: evidence from the US," *mimeo Univ. Michigan*.

BURCHFIELD, M., H. OVERMAN, D. PUGA, AND M. TURNER (2006): "Causes of Sprawl: A Portrait from Space," *The Quarterly Journal of Economics*, 121, 587–633.

CALIENDO, L. AND F. PARRO (2015): "Estimates of the Trade and Welfare Effects of NAFTA," *The Review of Economic Studies*, 82, 1–44.

CHANEY, T. (2008): "Distorted gravity: the intensive and extensive margins of international trade," *American Economic Review*, 98, 1707–21.

COŞAR, A. K., B. DEMIR, D. GHOSE, AND N. YOUNG (2021): "Road capacity, domestic trade and regional outcomes," *Journal of Economic Geography*, 22, 901–929.

DATTA, S. (2012): "The impact of improved highways on Indian firms," *Journal of Development Economics*, 99, 46–57.

DEKLE, R., J. EATON, AND S. KORTUM (2008): "Global Rebalancing with Gravity: Measuring the Burden of Adjustment," *IMF Staff Papers*, 55, 511–540.

DIJKSTRA, E. W. (1959): "A note on two problems in connexion with graphs," *Numerische mathematik*, 1, 269–271.

DINGEL, J. I., A. MISCIO, AND D. R. DAVIS (2019): "Cities, Lights, and Skills in Developing Economies," Working Paper 25678, National Bureau of Economic Research.

DONALDSON, D. (2018): "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, 108, 899–934.

DONALDSON, D. AND R. HORNBECK (2016): "Railroads and American Economic Growth: A Market Access Approach," *The Quarterly Journal of Economics*, 131, 799–858.

DONALDSON, D. AND A. STOREYGARD (2016): "The View from Above: Applications of Satellite Data in Economics," *Journal of Economic Perspectives*, 30, 171–98.

EATON, J. AND S. KORTUM (2002): "Technology, geography, and trade," *Econometrica*, 70, 1741–1779.

ELVIDGE, C., D. ZISKIN, K. BAUGH, B. TUTTLE, T. GHOSH, D. PACK, E. ERWIN, AND M. ZHIZHIN (2009): "A fifteen year record of global natural gas flaring derived from satellite data," *Energies*, 2, 595–622.

ELVIDGE, C. D., K. BAUGH, M. ZHIZHIN, F. C. HSU, AND T. GHOSH (2017): "VIIRS night-time lights," *International Journal of Remote Sensing*, 38, 5860–5879.

ELVIDGE, C. D., K. E. BAUGH, J. B. DIETZ, T. BLAND, P. C. SUTTON, AND H. W. KROEHL (1999): "Radiance Calibration of DMSP-OLS Low-Light Imaging Data of Human Settlements," *Remote Sensing of Environment*, 68, 77 – 88.

ELVIDGE, C. D., F.-C. HSU, K. E. BAUGH, AND T. GHOSH (2014): "National trends in satellite-observed lighting," *Global urban monitoring and assessment through earth observation*, 23, 97–118.

FABER, B. (2014): "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System," *The Review of Economic Studies*, 81, 1046–1070.

FALLY, T. (2015): "Structural gravity and fixed effects," *Journal of International Economics*, 97, 76–85.

FRIEDL, M. AND D. SULLA-MENASHE (2015): "MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006," distributed by nasa eosdis land processes daac, distributed by NASA EOSDIS Land Processes DAAC.

GHANI, E., A. G. GOSWAMI, AND W. R. KERR (2014): "Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing," *The Economic Journal*, 126, 317–357.

GOLDBLATT, R., M. STUHLMACHER, B. TELLMAN, N. CLINTON, G. HANSON, M. GEORGESCU, C. WANG, F. SERRANO-CANDELA, A. KHANDELWAL, W. CHENG, AND R. BALLING (2018): "Using Landsat and nighttime lights for supervised pixel-based image classification of urban land cover," *Remote Sensing of Environment*, 205, 253–275.

GUO, W., D. LU, Y. WU, AND J. ZHANG (2015): "Mapping impervious surface distribution with integration of SNNP VIIRS-DNB and MODIS NDVI data," Remote Sensing.

HARARI, M. (2020): "Cities in Bad Shape: Urban Geometry in India," *American Economic Review*, 110, 2377–2421.

HENDERSON, J. V., A. STOREYGARD, AND D. N. WEIL (2012): "Measuring Economic Growth from Outer Space," *American Economic Review*, 102, 994–1028.

HUANG, X., A. SCHNEIDER, AND M. FRIEDL (2016): "Mapping sub-pixel urban expansion in China using MODIS and DMSP/OLS nighttime lights," Remote Sensing of Environment.

JEDWAB, R. AND A. STOREYGARD (2020): "The Average and Heterogeneous Effects of Transportation Investments: Evidence from Sub-Saharan Africa 1960-2010," Working Paper 27670, National Bureau of Economic Research.

KRUGMAN, P. (1991): "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99, 483–99.

KRUSKAL, J. B. (1956): "On the shortest spanning subtree of a graph and the traveling salesman problem," *Proceedings of the American Mathematical society*, 7, 48–50.

MELITZ, M. J. (2003): "The impact of trade on intra-industry reallocations and aggregate industry productivity," *Econometrica*, 71, 1695–1725.

MERTES, C., A. SCHNEIDER, D. SULLA-MENASHE, A. TATEM, AND B. TAN (2015): "Detecting change in urban areas at contintental scales with MODIS data," Remote Sensing of Environment.

MONTE, F., S. J. REDDING, AND E. ROSSI-HANSBERG (2018): "Commuting, Migration, and Local Employment Elasticities," *American Economic Review*, 108, 3855–90.

NELSON, A. (2008): "Estimated travel time to the nearest city of 50,000 or more people in year 2000," Global Environment Monitoring Unit - Joint Research Centre of the European Commission, Ispra Italy.

PESARESI, M., D. EHRILCH, A. J. FLORCZYK, S. FREIRE, A. JULEA, T. KEMPER, P. SOILLE, AND V. SYRRIS (2015): "GHS built-up confidence grid, derived from Landsat, multitemporal (1975, 1990, 2000, 2014)," European Commission, Joint Research Centre (JRC).

REDDING, S. J. (2016): "Goods trade, factor mobility and welfare," *Journal of International Economics*, 101, 148 – 167.

REDDING, S. J. AND E. ROSSI-HANSBERG (2017): "Quantitative spatial economics," *Annual Review of Economics*, 9, 21–58.

ROZENFELD, H. D., D. RYBSKI, X. GABAIX, AND H. A. MAKSE (2011): "The Area and Population of Cities: New Insights from a Different Perspective on Cities," *American Economic Review*, 101, 2205–25.

STOREYGARD, A. (2016): "Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa," *The Review of economic studies*, 83, 1263–1295.

SULLA-MENASHE, D. AND M. A. FRIEDL (2018): "User guide to collection 6 MODIS land cover (MCD12Q1 and MCD12C1) product," *USGS: Reston, VA, USA*, 1, 18.

THE ECONOMIST (2022): "India is getting an eye-wateringly big transport upgrade," .

TSIVANIDIS, N. (2023): "Evaluating the impact of urban transit infrastructure: Evidence from Bogota's Transmilenio," .

TUTTLE, B. T., S. ANDERSON, C. ELVIDGE, T. GHOSH, K. BAUGH, AND P. SUTTON (2014): "Aladdin's Magic Lamp: Active Target Calibration of the DMSP OLS," *Remote Sensing*, 6, 12708–12722.

WOETZEL, J., N. GAREMO, J. MISCHKE, P. KAMRA, AND R. PALTER (2017): "Bridging infrastructure gaps: Has the world made progress," *McKinsey & Company*, 5.

WORLD BANK GROUP (2017a): "P050649: India - Tamil Nadu Road Sector Project," World Bank.

——— (2017b): "P067606: India - Uttar Pradesh State Roads Project," World Bank Group.

——— (2017c): "P069889: India - Mizoram State Roads Project," World Bank Group.

——— (2017d): "P070421: India - Karnataka State Highways Improvement Project," World Bank Group.

——— (2017e): "P072539: India - Kerala State Transport Project," World Bank Group.

——— (2017f): "P073776: India - Allahabad Bypass Project," World Bank Group.

——— (2017g): "P077856: India - Lucknow-Muzaffarpur National Highway Project," World Bank Group.

Online Appendix

A Detecting Markets with Satellite Imagery **22**

A.1 Market Boundaries . . . 22
A.2 Market Economic Activity . . . 23
A.3 Market Population . . . 24
A.4 Statistics . . . 25

B Road Digitization **26**

B.1 Digitization . . . 26
B.2 Calculating Travel Time between Markets . . . 26
B.3 Construction Costs . . . 27
B.4 Spanning Tree Network . . . 27

C Model **29**

C.1 Spatial Equilibrium Model . . . 30
C.2 Estimating Equations . . . 31
C.3 Market Access and Travel Times . . . 32
C.4 Instrumentation Strategy . . . 33
C.5 Welfare Impacts . . . 33
C.6 Deriving Equations (C.19), (C.20), and (C.21). . . . 34
C.7 Linking Structural Parameters with the Reduced-Form Parameters . . . 35

A Detecting Markets with Satellite Imagery

A.1 Market Boundaries

We define markets based on the methods in Baragwanath et al. (2019), who develop a clustering algorithm that defines a market as the set of contiguous or near-contiguous pixels that contain built-up economic activity as indicated by their spectral properties. This section summarizes that procedure.

The approach uses data on builtup landcover from the Moderate Resolution Imaging Spectroradiometer (MODIS) layer constructed by Friedl and Sulla-Menashe (2015). The MODIS sensor captures data in 36 spectral bands (ranging in wavelength from 0.4 µm to 14.4 µm) at varying spatial resolutions (from 250m to 1km). These raw data indicate the reflectance intensity within each spectral band for each pixel at a given moment in time. To convert data on the spectral properties of pixels into classifications of land use, Friedl and Sulla-Menashe (2015) apply a supervised machine learning method, which takes advantage of a global database of training sites extracted from high-resolution imagery.[13] The resulting data, which have a spatial resolution of 500m, are publicly available on Google Earth Engine and have been widely used in the remote sensing literature (e.g., Huang et al. 2016, Mertes et al. 2015, Guo et al. 2015).

There are six versions of the MODIS land-use classification layer. We use the University of Maryland version MCD12Q1 V006, which classifies global landcover at yearly intervals from 2001 to 2016. Using data for 2001 and 2011, we take Urban and Builtup pixels (classification 13) to indicate builtup landcover from MCD12Q1 V6 product (see Sulla-Menashe and Friedl 2018).

We combine adjoining pixels with builtup landcover into a cluster, where we set the minimum number of pixels in a cluster to be one ($0.25km^2$). We then form markets by combining proximate clusters. This method combines any pair of clusters into a market for which the minimum distance between their boundaries is less than $1km$, motivated by the method in Rozenfeld et al. (2011) for agglomerating neighboring administrative units into cities. Baragwanath et al. (2019) also construct markets for buffers of 2km, 4km, and 8km, where for a given threshold larger buffers nest smaller buffers (i.e., 1km markets $\subseteq$ 2km markets $\subseteq$ 4km markets $\subseteq$ 8km markets). For our analysis, we define market boundaries using 2001 images.

To combine clusters of built up pixels, we use the "Aggregate Polygons" function in ArcGis. This function combines polygons within a specified buffer to form larger polygons. Baragwanath et al. (2019) explain this procedure in depth. Finally, in order to capture changes in footprint growth between 2001 and 2011, we draw 0.5km buffers around the markets. Figure 1 illustrates

[13] There are alternatives to MODIS imagery for classifying land use. Baragwanath et al. (2019) also detect urban land using the Global Human Settlements Layer (GHSL) of Pesaresi et al. (2015), which is based on Landsat imagery; a recent layer produced by Goldblatt et al. (2018), which applies machine learning to Landsat imagery and nightlight data to detect builtup landcover in India; and a minimum threshold for nightlight intensity, similar to Harari (2020) and Dingel et al. (2019). Results based on GHSL and Goldblatt et al. (2018) are similar to those based on MODIS. As explained in that paper, we elect to use the MODIS layer over these options because it is more widely used in the literature and covers a longer time span; we elect not to use nightlights to define the extent of markets because of the well-known blooming effect of lights (Donaldson and Storeygard, 2016), which produces markets that are overly smooth and expansive relative to the ragged, amoeba-like clusters of settlements that compose metropolitan areas, as visible in daytime satellite imagery.

the result of this exercise for 1km, 4km and 8km MODIS markets around the Ahmedabad district area. While the 1km markets identify pockets of intense economic activity, the 8km markets do a good job of tracing the borders of the larger metro area. In the figure, we identify 383 1km markets, 183 4km markets and 42 8km markets.

A.2 Market Economic Activity

To measure the intensive margin of activity within each market, we use the intensity of light emitted at night from the Defense Meteorological Satellite Program Operational Linescan System (DMSP-OLS). DMSP-OLS sensors detect visible and near-infrared emissions at night from different sources on Earth, such as city lights, auroras, gas flares, and fires. Pixels in DMSP-OLS have a spatial resolution of approximately $1km$. For each pixel, the digital number (DN) of calibrated light intensity ranges from 0 to 63. We use the stable light band of sensor F14 for 2001 and sensor F18 for 2011, both of which discard ephemeral events, such as fires, but remain sensitive to persistent lighting, including from gas flares or volcanoes.[14] Since India has no active volcanoes or gas flares on land (Elvidge et al. 1999), it is safe to assume that highly lit pixels indicate builtup activity. For each urban market, we identify the DMSP-OLS pixels that overlap with its borders and aggregate light intensity across these pixels to create a measure of economic activity for the market.

DN values are subject both to top coding—because the sensor becomes saturated at high levels of light intensity, there is a maximum detectable DN value—and bottom coding—because of how the stable light band of the satellite sensors operate, at low levels of light intensity the sensor tends to bunch DN values at either 0 or 5.[15] Top coding is unlikely to be a concern in our data, owing to the infrequency of pixels hitting the maximum DN value of 63, which in India occurs only in the central business districts of the largest cities. In 2011, the DN value of the pixel at the 99.5th percentile was 60. Bottom coding, however, is a potential concern. In 2001, we find that 1,124 (8.4%) markets have zero aggregate nightlight intensity, while in 2011 no markets register zero nightlight intensity. Part of this over-time decrease in zero values may reflect an increase in economic activity. However, zero values in 2001 may also be the result of the sensor for that year being more prone to bunching DN values at zero. To deal with bottom coding, we proceed as follows: 1 using the sample of pixels with non-zero DN values in urban markets for 2001, we calculate the 5th percentile DN value, 2 we apply this 5th-percentile DN value to all pixels in urban markets that have zero DN readings in 2001, and 3 we sum DN values across pixels in each market to create aggregate light intensity at the market level in 2001. Our estimation results are insensitive to simply dropping markets with zero DN values in 2001.

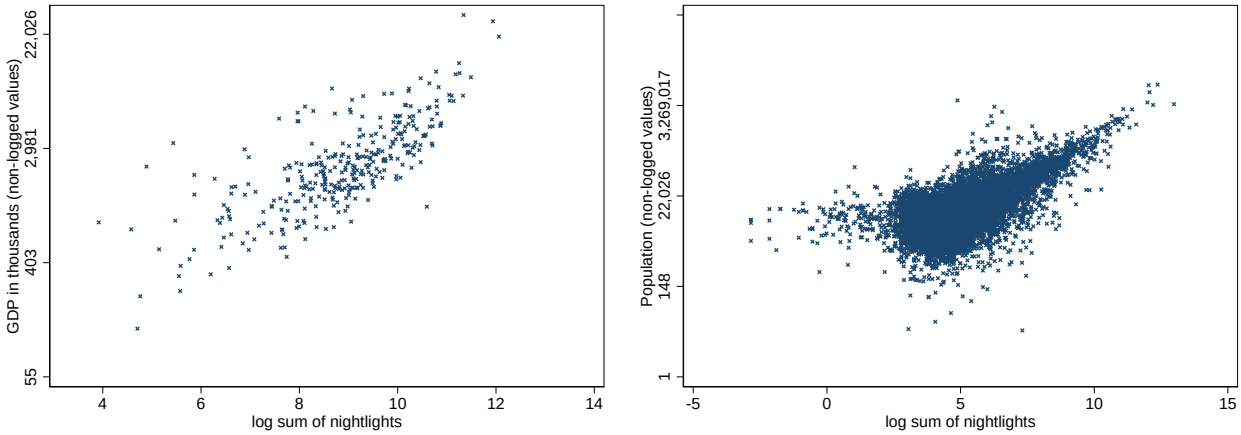
[14] The annual composites of DMSP-OLS cover 1992 to 2013 and include data from six different satellite sensors. New satellites are launched to account for the gradual degradation of sensors, where pairs of successive sensors overlap in space for periods of up to three years. To address differences in radiometric performance between sensors, we apply the calibration methods developed in Elvidge et al. (2009) and Elvidge et al. (2014), which uses the F12 image from Sicily in 1999 as a reference and estimates a series of second-order polynomial regression functions for paired pixels from successive sensors (e.g., F12-F13, F13-F14, etc.) and then recursively matches each sensor back to the range of F12 in 1999.

[15] Tuttle et al. (2014) develop a mapping of DNs to wattage by placing portable high-pressure sodium lamps at uninhabited sites in Colorado and New Mexico to check the DN recorded by the F16 and F18 sensors. They find that ninety-three 100-watt incandescent lamps could be detected (DN=1) at both fine (0.6km) and coarse (2.7km) resolutions. Eight times as many bulbs would saturate (DN=63) the sensor at the fine resolution but not at the coarser resolution.

The saturation effect, blooming effect, and other sources of noise in nightlight intensity are less pronounced in data from the Visible Infrared Imaging Radiometer Suite (VIIRS), imagery from which is available from 2012 forward (Elvidge et al., 2017). Here, we use DMSP-OLS imagery in order to create methods for measuring markets that can be extended backward in time. Our approach is easily adaptable to VIIRS data.

Figure A.1: Nightlight Intensity versus Economic Activity

(a) Log GDP (b) Log Population



Notes: The figure presents the relationship between nightlights and two measures of economic activity, GDP and population. Panel (a) shows the bin-scatter plot for log sum of nightlights and log GDP in 2001 at the district level. GDP for 326 of India's districts is from the Indian Planning Commission; nightlights data are from DMSP-OLS Nighttime Lights Time Series Version 4, Defense Meteorological Program Operational Linescan System and aggregated up to the district level. We run OLS regression of log GDP on log nightlights, log land area, and a regression constant. We obtain a coefficient of 0.53 (sd 0.05) for log nightlights, and an R^2 of 0.58 . Panel (b) shows the bin-scatter plot for log sum of nightlights and log population in 2001 for 13,234 markets (all markets with non-zero population). Population data is from the 2001 Indian Census; DMSP-OLS nightlights are aggregated up to the market level. For interpretability, the non-logged values are presented on the Y-axis for both plots.

A.3 Market Population

In order to measure the number of people inside each market, we rely on the census conducted by the Indian government in 2001. Census data are available at the town or village level for the whole of India, which represent the most dissaggregated level at which information is available. There is an order of magnitude more towns and villages in the census compared to the number of local markets we identify. In 2001, India was composed of 638,588 villages and 5,161 towns, with a total population of 1,028,737,436; in 2011, India was composed of 649,481 villages and 7,935 towns, with a total population of 1,210,854,977.

Since villages and towns are, on average, significantly smaller than our urban markets, we allocate the population living in these administrative units by spatially overlaying the census shapefiles with our urban markets shapefile. Most villages and towns overlap only one urban market.

For these, we allocate the entire population living inside the town or village to the urban market that they overlap with. When a town or village overlaps with more than one urban market, the population inside that town or village is divided by the number of urban markets it overlaps with, and equally distributed to each of these markets.

A.4 Statistics

Table A.1 reports the statistics of the markets detected from the procedures described above. We detect 13,387 markets at the 1km buffer.

Table A.1: Summary Statistics of Markets

	N	Mean	Median	SD	Min	Max
Panel A: 1km Markets						
Area (Km2)	13,387	6.5	3.4	23.9	1.7	1,607.6
Population 2001 (in thousands)	13,368	26.8	7.8	176.6	0.0	10,366.0
Population 2011 (in thousands)	13,386	29.6	8.7	198.2	0.0	11,647.7
Mean Nightlights 1996	13,387	10.7	6.9	10.0	0.0	62.7
Mean Nightlights 2011	13,387	11.6	7.7	10.6	1.9	60.1
Percent δ Ntl Mean 1996-2011	13,387	13.8	12.5	47.5	-175.6	724.5
Sum Nightlights 1996	13,387	709.4	108.4	6,092.4	0.1	435,154.8
Sum Nightlights 2011	13,387	733.5	122.4	6,188.0	16.1	441,145.0
Percent δ Ntl Sum 1996-2011	13,387	13.7	10.8	47.6	-175.6	724.5
Panel B: 4km Markets						
Area (Km2)	7,539	14.9	4.5	54.7	1.7	2,402.5
Population 2001 (in thousands)	7,531	48.2	11.6	307.0	0.0	13,884.3
Population 2011 (in thousands)	7,538	53.3	13.1	339.9	0.0	15,625.0
Mean Nightlights 1996	7,539	10.8	7.6	9.7	0.0	56.4
Mean Nightlights 2011	7,539	11.4	7.9	9.9	1.9	59.5
Percent δ Ntl Mean 1996-2011	6,895	9.6	6.0	56.9	-169.3	681.0
Sum Nightlights 1996	7,539	1,428.5	171.7	10,056.6	0.0	528,017.8
Sum Nightlights 2011	7,539	1,525.5	180.9	11,113.0	16.6	598,270.5
Percent δ Ntl Sum 1996-2011	6,895	9.6	6.0	56.9	-169.3	681.0
No. 1km Markets	7,539	1.8	1.0	2.9	1.0	89.0
Panel C: 8km Markets						
Area (Km2)	3,483	62.1	6.5	574.5	1.7	23,384.8
Population 2001 (in thousands)	3,481	113.1	19.8	705.0	0.0	18,374.9
Population 2011 (in thousands)	3,482	125.9	22.5	803.1	0.0	28,184.2
Mean Nightlights 1996	3,483	12.7	9.8	10.3	0.0	53.9
Mean Nightlights 2011	3,483	13.3	9.7	10.5	1.9	56.5
Percent δ Ntl Mean 1996-2011	3,245	8.8	5.3	52.7	-169.3	527.1
Sum Nightlights 1996	3,483	4,161.1	329.8	31,546.7	0.0	1085618.0
Sum Nightlights 2011	3,483	4,604.7	352.0	35,723.9	16.6	1180177.5
Percent δ Ntl Sum 1996-2011	3,245	8.8	5.3	52.7	-169.3	527.1
No. 1km Markets	3,483	3.8	1.0	29.1	1.0	1,263.0
No. 4km Markets	3,483	2.2	1.0	11.5	1.0	562.0

Notes: The table presents summary statistics for MODIS-based markets. Panel A displays summary statistics for 1km markets while panel B displays summary statistics for 4km markets and Panel C for 8km markets. Area information is based on the footprints derived from MODIS satellite data. Population data is from the 2001 and 2011 Indian Census; nightlights data is computed from DMSP-OLS satellite images. Mean nightlights represents the average DN value inside the markets, and can vary between 0 and 63. Sum of Nightlights is the sum of DN values within a market.

B Road Digitization

B.1 Digitization

The digitization process involved five steps. In the first two, we created georeferenced maps for 1996 and 2011,[16] with the assistance of the digital mapping company CyberSwift, and then identified the type of each intercity road in each year. The raw road maps identify five road types: the Golden Quadrilateral, national highways, all-weather motorable roads, fair-weather motorable roads, and roads where delays may occur. To begin, we assigned roads their given classification for the corresponding year. We also use the OpenStreetMap (OSM) for India in 2018 to visually inspect the accuracy of our classification method.

The next three steps involved identifying the presence of new and upgraded roads. We began by comparing maps for 1996, 2011, and 2018. This exercise revealed (i) roads that existed in 2011 and 2018 but not 1996, which indicate new construction; (ii) roads that existed in all years but had a higher type in 2011 and 2018 than in 1996, which indicate road upgrades; and (iii) anomalies in road placement or type, which included cases in which the 2011 road was of a lower type than that of 1996 or in which the georeferenced maps for the two years identify different locations for the same road.[17] To resolve the anomalies, we used the 2018 map as ground truth. We aligned the georeferenced map for 2011 with OSM for 2018. Where the two maps agreed on road placement, we made no changes. Where they disagreed, we realigned the road location for 2011 to match than in 2018. Similarly, we realigned the 1996 map to the adjusted map for 2011, where they disagreed on road placement. To resolve anomalies in road type, if a road had a higher type in 1996 and 2018 than in 2011, we upgraded the 2011 type; if a road had a higher type in 1996 than in 2011 and 2018, we downgraded the 1996 road type. In the final step, we made other corrections to the georeferenced maps, accounting for "dangling" roads (which begin at one intersection and connect to no other road) and "orphan" roads (which begin and end without connecting to another road), by manually comparing the georeferenced and raw maps.

B.2 Calculating Travel Time between Markets

We use Dijkstra's algorithm to compute bilateral travel times between markets. This algorithm, developed by Dijkstra in 1959, finds the shortest path between a starting node and a target node through a weighted graph. The graph is a data structure consisting of a set of nodes and edges, where the edges represent travel times between the nodes. By comparing the sum of distances, such as physical distance, travel costs, and travel times, the algorithm can find the shortest path from an origin market to a destination market. Given our setting, the nodes can be interpreted as intersections of digitized roads and the edges as travel times between nodes.

We applied multiple settings to generate a series of travel times with different combinations of travel speeds. Since major roads do not have directional flows, we defined the graph as undirected. To convert urban markets and digitized roads into a weighted graph, we utilized the R package shp2graph developed by Lu et al. (2018). This allowed us to calculate travel times for a

[16] To align the maps, we used the Geographic Coordinate System and Lambert_Conformal_Conic for projection.

[17] We ignore new road construction or road upgrades after 2011.

Table B.1: Reduction in Travel Times Between 1996 and 2011

	Travel Time 1996 (hrs) (1)	Travel Time 2011 (hrs) (2)	Difference (3)
Distance between markets			
0-50km	2.6	2.2	-.37
50-200km	4.8	4.2	-.56
200-500km	9.3	8.2	-1.03
500km +	22.7	19.4	-3.26

Notes: The table shows average travel times (in hours) between markets that are 0-50km, 50-200km, 200-500km or more than 500km away from each other. Column 1 presents average travel times in 1996, column 2 in 2011, and column 3 presents the difference between 2011 and 1996. Negative values in column 3 indicate reductions in travel times between the two years. Travel times were calculated using the Dijkstra algorithm.

series of urban markets based on the spatial network, while incorporating the weight of each edge to obtain a more accurate representation of the spatial network.

Although urban markets are generally located along major roads, most of the market centroids are not situated directly on the edges, which consist of digitized roads. Additionally, our original data source, the Road Map of India, does not contain information on tertiary roads due to the scale of the maps. To address this issue, we constructed a series of pseudo roads between urban markets and their nearest geometric point on the edges of the digitized roads. This allowed us to factor in additional travel times that were not accounted for in the original data set. By doing so, we were able to obtain a more accurate representation of travel times between urban markets.

B.3 Construction Costs

Table B.2 reports the construction costs from the World Bank's Project Appraisal Document, and the Implementation Completion and Reports (World Bank Group, 2017a,b,c,d,e,f,g).

B.4 Spanning Tree Network

This section describes the construction of the least-cost minimum spanning tree algorithm for the Golden Quadrilateral (GQ) and national highways contracted in India between 1996 and 2001. Since the primary objective of highway construction in India was to establish effective transportation networks among major cities, we used the minimum spanning tree algorithm to create the least-cost connectivity network. We executed the computation in ArcGIS using Kruskal's algorithm (Kruskal, 1956). This algorithm finds edges with the least possible weights to connect all nodes in a undirected weighted graph setting.

Since we aim to reproduce the least-cost connectivity network based on real-world features, it is critical that we use multiple layers of geographical constraints, rather than a simple digital elevation model (as is commonly used). We thus applied a friction surface developed by Nelson (2008), in which we aggregated data of multiple layers such as land cover, elevation, slope,

Table B.2: The Length of Digitized Roads and Associated Costs of Road Construction

	Length (km)			Unit Costs ($ million/km)	Total Costs ($ billion)
	1996	2011	Build-up		
Golden Quadrilateral (GQ)	0	5714	5714	3.55	20.3
National Highway (NH)	29422	49172	19750	2.12	41.9
Other: Where delay may occur	2838	3625	787	0.42	0.33

Notes: The unit cost for the GQ project is based on that of the Allahabad Bypass Expressway (P073776) (World Bank Group, 2017f), which is part of the GQ project. The unit cost for national highway construction is obtained by taking the average of construction costs for the Lucknow-Muzaffarpur national highway project (P077856) (World Bank Group, 2017g) and the Grand Trunk road projects (P069889) (World Bank Group, 2017c). The unit cost for local roads is based on average of widening and reconstruction costs of state road projects (P050649, P067606, P070421, and P072539) (World Bank Group, 2017a,b,d,e).

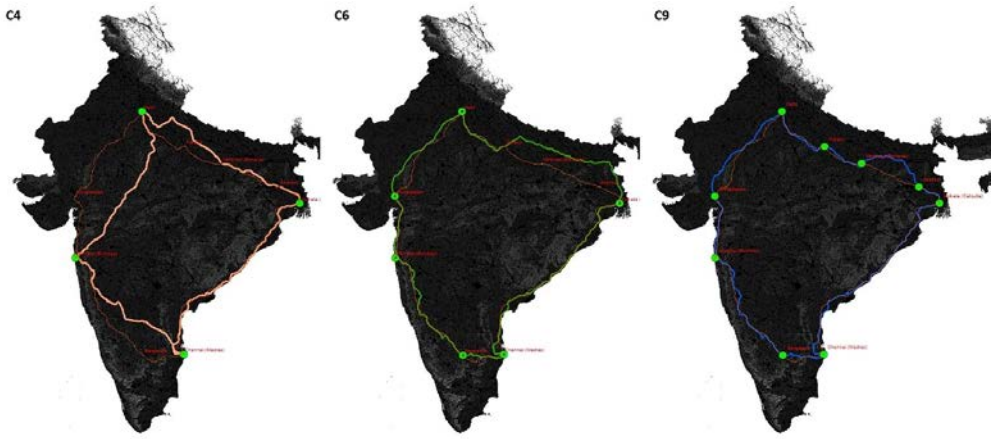
national borders, and existing road, railway, and water bodies as of 2000, which was originally created to calculate global travel times. This friction surface is a composite of $1km$ spatial resolution pixels, which represents the time required to cross the pixel in minutes. We use these data as the cost layer of the spanning tree algorithm.

We separately represent construction of the GQ and other national highways. For the GQ, we targeted connecting India's very largest cities; for other highways, we targetted connecting large and middle-tier cities. For the construction of the GQ, we tested three versions: 1) connecting the four largest cities (Chennai, Delhi, Kolkata, and Mumbai), 2) connecting the six cities (adding Ahmadabad, Bangalore), and 3) connecting the nine largest cities: adding Asansol, Kanpur, and Varanasi).

The solutions to these three approaches is seen in Figure B.1, with option 1 for four cities in pink, option 2 for six cities in green, and option 3 for nine cities in blue. Option 2, with a total length of 6,198km delivers the closest approximation of the GQ (5,714km) and we use this solution to construct the GQ portion of our instrument India's highway expansion.

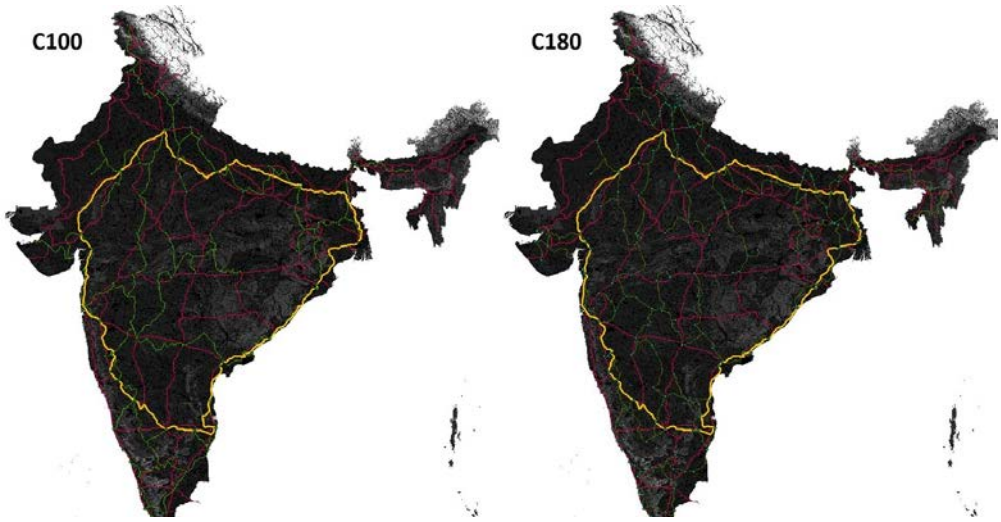
For the construction of other highways, we considered a spanning tree to connect India's 100 largest cities (minimum population in $1991 > 500,000$) or India's 180 largest cities (minimum population in $1999 > 200,000$), where we use populations in 1991, ten years before the National Highway Development Project began. Using the same friction surface as in the GQ construction, Figure B.2 shows the results for the two approaches, where existing roads in 1996 are in red, the spanning tree for the GQ is in yellow, and the spanning tree for other highways is in green. Both versions deliver close approximations of India's highway expansion. We use the 100-city version as our baseline and the 180-city version in further analysis.

Figure B.1: Spanning Tree Version of GQ Construction



Notes: Three versions of GQ construction based on the minimum spanning algorithm: 1) four cities (pink), 2) six cities (green), and 3) nine cities (blue). The orange color is an original path of GQ for the reference. The background imagery is a cost layer constructed based on a friction surface.

Figure B.2: Spanning Tree Version of Highway Construction



Notes: On top of the four cities version of the GQ, two versions of NH construction were implemented: 1) hundred cities version (left) and hundred eighty version (right). The background imagery is a cost layer constructed based on a friction surface. GQ is yellow line; the existing NH in 1996 is red; and newly constructed NH by MST is green line.

C Model

This section presents a spatial equilibrium model of trade and labor mobility, which we use to derive estimating equations and perform counterfactual welfare analysis.

C.1 Spatial Equilibrium Model

Gravity trade and prices. We extend the framework in Eaton and Kortum (2002) to model the spatial equilibrium for Indian markets. There is a continuum of goods ω. Under perfect competition, each Indian market can potentially produce any type of good. For each variety ω, importers will source from their cheapest exporter. Let $z_r(\omega)$ denote market r's productivity in producing variety ω, which follows an i.i.d. Fréchet distribution across product varieties ω and markets r, with cumulative distribution function $F_r(z) = \exp\big(-T_r z^{-\theta}\big)$, where $\theta > 1$ is the shape parameter and T_r is the scale parameter. Productivity is more dispersed the smaller is θ.

In this setting, the share of city n's expenditure on goods produced by city r, Π_{rn}, is given by,

$$\Pi_{rn} = \frac{T_r w_r^{-\theta} \tau_{in}^{-\theta}}{\sum_\ell z_\ell w_\ell^{-\theta} \tau_{\ell n}^{-\theta}}, \tag{C.1}$$

where the price index of city n is,

$$P_n = C\Big(\sum_r T_r w_r^{-\theta} \tau_{rn}^{-\theta}\Big)^{-\frac{1}{\theta}}, \tag{C.2}$$

Equations (C.1) and (C.2) characterize a standard constant elasticity of substitution (CES) demand system (Arkolakis et al., 2012). Similar equilibrium conditions can be generated by other trade models, including Anderson (1979), Krugman (1991), and Melitz (2003)-Chaney (2008), where each model differs in its interpretations of the parameters T_r and θ.

Labor Supply. We build on Allen and Arkolakis (2014), Redding (2016), and Redding and Rossi-Hansberg (2017) to model labor mobility. We assume that each worker has an idiosyncratic taste parameter, η_n, which is distributed i.i.d. Fréchet across locations, with cumulative distribution function $G(\eta) = \exp\big(-\eta^{-1/\mu}\big)$. $\frac{1}{\mu} > 1$ is the dispersion parameter, where a higher value of $\frac{1}{\mu}$ implies less dispersion in idiosyncratic tastes across workers. μ captures the inverse labor supply elasticity. Individuals choose the location that maximizes their utility, which is the product of real wages and location preferences, $\frac{w_n}{P_n}\eta_n$. In equilibrium, the fraction of workers choosing market n is,

$$\frac{L_n}{L} = \frac{(w_n/P_n)^{\frac{1}{\mu}}}{\sum_\ell (w_\ell/P_\ell)^{\frac{1}{\mu}}}.^{18} \tag{C.3}$$

We assume that there is no international migration. We restate equation (C.3) as,

$$w_n = U P_n L_n^{\mu}. \tag{C.4}$$

where $U = \Big[\sum_\ell (w_\ell/P_\ell)^{\frac{1}{\mu}} \frac{1}{L}\Big]^{\mu}$ captures aggregate worker utility and is constant across n. In the regression analysis, U is absorbed by the constant term, but does vary with changes in the general equilibrium.

[18] This upward-sloping labor supply model can alternatively be generated via constraints on housing supply, see, e.g., (Bartelme, 2018), in which case $\mu = \frac{1-\alpha}{\alpha}$, where $1-\alpha$ is household expenditure share on housing.

Market Access and Trade balance. Below, we carry out our analysis using market access, following Donaldson and Hornbeck (2016). Consumer market access of city n is given by,

$$\Phi_n = \sum_r T_r w_r^{-\theta} \tau_{rn}^{-\theta}.$$

A higher value of consumer market access indicates that consumers of city n have access to cheaper goods.

Balance trade indicates that city r's income equals the sum of exports to all cities,

$$w_r L_r = \sum_n \Pi_{rn} w_n L_n = T_r w_r^{-\theta} \underbrace{\sum_n \frac{w_n L_n \tau_{in}^{-\theta}}{\Phi_n}}_{\text{Firm market access}}. \tag{C.5}$$

In the last term, the summation measures the firm market access to consumers—how easy it is for firms in city n to sell their products to consumers across all markets. Under symmetric trade costs, consumer and firm market accesses are equal up to a factor of proportion (Donaldson and Hornbeck, 2016), such that

$$\Phi_r = \sum_n \frac{w_n L_n \tau_{rn}^{-\theta}}{\Phi_n}. \tag{C.6}$$

Below, we carry out our analysis using consumer market access, which allows us to analyze changes in general equilibrium without data on bilateral trade (given data on bilateral trade costs).

C.2 Estimating Equations

As shown in Fally (2015) and Bartelme (2018), equations (C.4) and (C.5) imply that,

$$w_r = \Phi_r^{\varepsilon_w} \chi_r^w \tag{C.7}$$

$$L_r = \Phi_r^{\varepsilon_l} \chi_r^l \tag{C.8}$$

where $\chi_r^w = U^{\frac{1}{\mu(\theta+1)+1}} T_r^{\frac{\mu}{\mu(\theta+1)+1}}$ and $\chi_r^l = U^{\frac{-\theta-1}{\mu(\theta+1)+1}} T_r^{\frac{1}{\mu(\theta+1)+1}}$ are functions of aggregate utility and regional productivities and ε_w and ε_l are reduced-form wage and population elasticities with respect to market access, and are functions of the structural parameters, μ and θ.[19] We write these conditions in log changes as, (where $\hat{x}_t = \ln(x_t / x_{t-1})$) as,

$$\hat{w}_{rt} = \varepsilon_w \hat{\Phi}_{rt} + \hat{\chi}_{rt}^w \tag{C.9}$$

$$\hat{L}_{rt} = \varepsilon_l \hat{\Phi}_{rt} + \hat{\chi}_{rt}^L \tag{C.10}$$

[19] Appendix C.7 shows that

$$\varepsilon_w = \frac{\mu\theta - 1}{\theta\left[\mu\theta + \mu + 1\right]}, \qquad \varepsilon_l = \frac{2\theta + 1}{\theta\left[\mu\theta + \mu + 1\right]}.$$

and the change in market access as,

$$\hat{\Phi}_{rt} = \ln \left(\sum_n \frac{w_{nt} L_{nt}}{\Phi_{nt}} \tau_{rn,t}^{-\theta} \right) - \ln \left(\sum_n \frac{w_{n,t-1} L_{n,t-1}}{\Phi_{n,t-1}} \tau_{rn,t-1}^{-\theta} \right) \quad \text{(C.11)}$$

We will perform the empirical analysis as though satellite imagery is our only source of data for economic activity in each market. Following Henderson et al. (2012), we presume that the intensity of light emitted at night is a function of total economic activity in a location. We assume in particular that the log change in income in market r, $\hat{y}_{rt} \equiv \hat{w}_{rt} + \hat{L}_{rt}$, has a linear relationship with the log change in nightlight intensity given by

$$\hat{y}_{rt} = \alpha \hat{\mathrm{N}}_{rt} + v_{rt} \quad \text{(C.12)}$$

where $\hat{\mathrm{N}}_{rt}$ is the log change in the intensity of light emitted in market r, α is the inverse elasticity of nightlight intensity with respect to income, and v_{rt} is an i.i.d. disturbance associated with measurement error in the use of nightlights to capture GDP. Combining equations (C.9) to (C.12), we obtain a modified set of equilibrium conditions for the general spatial model, given by

$$\hat{y}_{rt} = \delta \hat{\Phi}_{rt} + \hat{\Gamma}_{rt} \quad \text{(C.13)}$$

$$\hat{\Phi}_{rt} = \ln \left(\sum_n \frac{y_{nt}}{\Phi_{nt}} \tau_{rn,t}^{-\theta} \right) - \ln \left(\sum_n \frac{y_{n,t-1}}{\Phi_{n,t-1}} \tau_{rn,t-1}^{-\theta} \right) \quad \text{(C.14)}$$

where $\delta \equiv (\varepsilon_w + \varepsilon_l)$; $\hat{\Gamma}_{rt} = \hat{\chi}_{rt}^w + \hat{\chi}_{rt}^l + v_{rt}$ is a regional growth shock that incorporates the shocks to amenities and productivities in (C.9) and (C.10), and measurement error in using nightlights to proxy for income. Given θ, α, data on NTL_{rt} and $\tau_{rn,t}$ across time, and the functional form of $\hat{\Phi}_{r,t}$ in (C.14), we can use (C.13) to estimate the reduced-form parameter δ.

C.3 Market Access and Travel Times

In taking equations (C.13) and (C.14) to the data, we confront two estimation issues. First, the change in market access in (C.14) is a function of unknown levels of market access in periods t and $t-1$.

We measure the structural market access by solving Φ_n from the following system of R equations in R unknowns at each time t,

$$\Phi_{rt}^{\text{structural}} = \sum_n \frac{y_{nt}}{\Phi_{nt}} \tau_{rn,t}^{-\theta}.^{20} \quad \text{(C.15)}$$

In (C.14), we exclude Φ_{nt} and $\Phi_{n,t-1}$ in constructing the expressions in the two bracketed terms that appear on the right-hand side of the equation.

Given values of α and θ, the approach computes the market access as a function of nightlight intensity and the iceberg trade costs. We follow an extensive literature that assumes the iceberg

[20] In so doing, we subsume into the solution for Φ_{rt} the time-specific constant, $ce^{\gamma_t + 0.5\sigma_\eta}$.

trade cost as a function of travel times below

$$\tau_{rn}^{-\theta} = \left(\text{travel times}_{rn} + \frac{\kappa \times (d_{rn} + d_{nr})}{\text{speed}}\right)^{-\phi}, \tag{C.16}$$

where θ is the trade elasticity to iceberg costs, and ϕ measures the trade elasticity to travel time or distance. We set $\phi = 1.45$ which implies a trade elasticity to the distance of 1.45 used in the literature (see Allen and Arkolakis, 2022). Counterfactual trade costs used in section 4 are calculated analogously.

C.4 Instrumentation Strategy

The change in market access in (C.14) contains nightlights and trade costs in period t. These values embody endogenous responses to shocks between $t-1$ and t, creating the need for an instrumentation strategy. We specify the following instrument for $\hat{\Phi}_{rt}$,

$$\hat{\Phi}_{rt}^{iv} = \ln \left(\sum_{r'} \frac{y_{n,t-1}}{\Phi_{nt}^{mst}} \tau_{rn,t}^{mst} \left(\Omega_{r't-1}\right)^{-\theta} \right) - \ln \left(\sum_{n} \frac{y_{n,t-1}}{\Phi_{n,t-1}} \tau_{rn,t-1}^{-\theta} \right) \Bigg(\tag{C.17}$$

Following Faber (2014), we project the value of bilateral trade costs in period t, $\tau_{rn,t}$, using $\tau_{rn,t}^{mst}\left(\Omega_{r't-1}\right)$, which is constructed using a least-cost spanning tree algorithm. We begin with India's road and highway network as of the initial period and upgrade roads by connecting India's 180 largest cities. Following Donaldson and Hornbeck (2016), in the first bracketed term we use the lagged value $y_{n,t-1}$ in place of y_{nt}. We further replace Φ_{nt} with Φ_{nt}^{mst}, which is given by the solution to the $R \times R$ system of equations,

$$\Phi_{rt}^{mst} = \sum_{n} \frac{y_{n,t-1}}{\Phi_{n,t}^{mst}} \tau_{rn,t}^{mst} \left(\Omega_{r't-1}\right)^{-\theta}. \tag{C.18}$$

In this way, the instrument in (C.17) is based purely on information as of time $t-1$.

C.5 Welfare Impacts

We solve the model in changes using (Dekle, Eaton, and Kortum, 2008), and set the observed economy of 2011 as the initial equilibrium.[21] Since productivity is unchanged between equilibria, $\hat{T}_r = 1$. Given data on local population L_n, the initial market access value Φ_r in 2011, local GDP measured as N_r^{α}, and the counterfactual values in trade costs τ'_{rn}, we solve $\hat{\Phi}_r$ from the following equation systems

$$\hat{\Phi}_r \Phi_r = \left[\sum_{n} \hat{\Phi}_n^{\frac{1}{\mu}(\varepsilon_w + \frac{1}{\theta})} \frac{L_n}{L} \right]^{-\frac{\theta}{\theta+1}} \left[\sum_{n} \frac{\hat{\Phi}_n^{\varepsilon_w + \varepsilon_l} \mathrm{N}_n^{\alpha}}{\hat{\Phi}_n \Phi_n} \tau'^{-\theta}_{rn} \right]. \tag{C.19}$$

[21] Regarding the choice of the initial equilibrium, the existing literature is mixed. For example, Caliendo and Parro (2015) calibrate the model to the beginning period (1996 in our case), whereas Adao, Costinot, and Donaldson (2017) calibrate the model to the ending year (2011 in our case). We choose year 2011 as the initial equilibrium because the local population data is only available in year 2011. Different choices of the initial equilibrium differ in the states of economy we are conditional on.

See Appendix C.6 for the derivation. Once we obtain $\hat{\Phi}_i$, we calculate the changes in local wage as

$$\hat{w}_r = \hat{\Phi}_r^{\varepsilon_w} \left[\sum_n \hat{\Phi}_n^{\frac{1}{\mu}(\varepsilon_w + \frac{1}{\theta})} \frac{L_n}{L} \right]^{\frac{1}{\theta+1}}. \tag{C.20}$$

We calculate changes in local population as

$$\hat{L}_r = \hat{\Phi}_r^{\varepsilon_l} \left[\sum_n \hat{\Phi}_n^{\frac{1}{\mu}(\varepsilon_w + \frac{1}{\theta})} \frac{L_n}{L} \right]^{-1}, \tag{C.21}$$

and changes in local price and welfare as

$$\hat{P}_r = \hat{\Phi}_r^{-\frac{1}{\theta}}, \qquad \widehat{\text{Welfare}}_r = \frac{\hat{w}_r}{\hat{P}_r}. \tag{C.22}$$

C.6 Deriving Equations (C.19), (C.20), and (C.21).

Let $\hat{X} = \frac{X'}{X}$ be variables in proportional changes between two equilibria, where X' denotes variables in the counterfactual equilibrium and X denoted variables in the initial equilibrium (which we treat as in the year 2011). The counterfactual deviates from the initial equilibrium only by changing trade costs from τ_{in} to τ'_{in}, while productivity is unchanged, such that $\hat{T}_r = 1$.

We express (C.32)and (C.33) in changes to have

$$\hat{w}_i = \hat{\Phi}_i^{\varepsilon_w} \hat{\chi}_w, \tag{C.23}$$

$$\hat{L}_i = \hat{\Phi}_i^{\varepsilon_l} \hat{\chi}_l. \tag{C.24}$$

Assuming $\hat{T}_r = 1$, $\hat{\chi}_w = \hat{U}^{\frac{1}{\mu(\theta+1)+1}}$, $\hat{\chi}_l = \hat{U}^{\frac{-(\theta+1)}{\mu(\theta+1)+1}}$. $\hat{U}$ captures the changes in aggregate utility, summing over all locations and therefore common across locations. $\hat{U}$ is absorbed by the constant regressor in our reduced-form regression but might change in general equilibrium. Recall that the labor supply curve in market n is

$$L_n = \frac{(w_n/P_n)^{\frac{1}{\mu}}}{\sum_n (w_n/P_n)^{\frac{1}{\mu}}} L \Longrightarrow w_n = P_n L_n^{\mu} \Big[\sum_n (w_n/P_n)^{\frac{1}{\mu}} \frac{1}{L} \Big]^{\mu} \tag{C.25}$$

Here, $U = \left[\sum_n (w_n/P_n)^{\frac{1}{\mu}} \frac{1}{L} \right]^{\mu}$. Supposing foreign workers cannot migrate to India, $L' = L$. The proportional changes is

$$\hat{U} = \left[\frac{\sum_n (w'_n/P'_n)^{\frac{1}{\mu}} \frac{1}{L'}}{\sum_n (w_n/P_n)^{\frac{1}{\mu}} \frac{1}{L}} \right]^{\mu} = \Big[\sum_n \big(\frac{\hat{w}_n}{\hat{P}_n}\big)^{\frac{1}{\mu}} \frac{L_n}{L} \Big]^{\mu} \tag{C.26}$$

It turns out $\hat{U}$ can be expressed as functions of the vector of market access changes $\{\hat{\Phi}_1, \hat{\Phi}_2, ..., \hat{\Phi}_n\}$, and therefore, $\hat{w}_i$, $\hat{L}_i$ are also functions of $\{\hat{\Phi}_1, \hat{\Phi}_2, ..., \hat{\Phi}_n\}$. This allows us to simplify the equation system. First, the proportional changes in U can be expressed as

$$\hat{U} = \Big[\sum_n \big(\frac{\hat{w}_n}{\hat{P}_n}\big)^{\frac{1}{\mu}} \frac{L_n}{L}\Big]^{\mu} = \Big[\sum_n \hat{\Phi}_n^{\frac{1}{\mu}(\varepsilon_w+\frac{1}{\theta})} \hat{\chi}_w^{\frac{1}{\mu}} \frac{L_n}{L}\Big]^{\mu} = \hat{U}^{\frac{1}{\mu(\theta+1)+1}} \Big[\sum_n \hat{\Phi}_n^{\frac{1}{\mu}(\varepsilon_w+\frac{1}{\theta})} \frac{L_n}{L}\Big]^{\mu}. \tag{C.27}$$

$$\iff \hat{U} = \Big[\sum_n \hat{\Phi}_n^{\frac{1}{\mu}(\varepsilon_w+\frac{1}{\theta})} \frac{L_n}{L}\Big]^{\frac{\mu\theta+\mu+1}{\theta+1}} \tag{C.28}$$

Substituting equation (C.28) into equation (C.23) and (C.24), we have that

$$\hat{w}_i = \hat{\Phi}_i^{\varepsilon_w} \Big[\sum_n \hat{\Phi}_n^{\frac{1}{\mu}(\varepsilon_w+\frac{1}{\theta})} \frac{L_n}{L}\Big]^{\frac{1}{\theta+1}}, \tag{C.29}$$

$$\hat{L}_i = \hat{\Phi}_i^{\varepsilon_l} \Big[\sum_n \hat{\Phi}_n^{\frac{1}{\mu}(\varepsilon_w+\frac{1}{\theta})} \frac{L_n}{L}\Big]^{-1}, \tag{C.30}$$

Finally, we can write the market access equation (C.5) at the counterfactual equilibrium value as

$$\Phi_i' = \sum_n \frac{w_n' L_n' \tau_{in}'^{-\theta}}{\Phi_n'} \Longrightarrow \hat{\Phi}_i \Phi_i = \sum_n \frac{\hat{w}_n \hat{L}_n w_n L_n}{\hat{\Phi}_n \Phi_n} \tau_{in}^{-\theta} \hat{\tau}_{in}^{-\theta}. \tag{C.31}$$

Given $w_i L_i = \mathrm{N}_i^{\alpha}$, we can substitute (C.29) and (C.30) into equation (C.31) to obtain equation (C.7).

C.7 Linking Structural Parameters with the Reduced-Form Parameters

We express θ and μ as a function of ε_w and ε_l. First, we express wage and population in terms of market access as

$$w_i = \Phi_i^{\varepsilon_w} \chi_w \tag{C.32}$$

$$L_i = \Phi_i^{\varepsilon_l} \chi_l \tag{C.33}$$

where $\chi_w = U^{\frac{1}{\mu(\theta+1)+1}} z_n^{\frac{\mu}{\mu(\theta+1)+1}}$, $\chi_l = U^{\frac{-\theta-1}{\mu(\theta+1)+1}} z_n^{\frac{1}{\mu(\theta+1)+1}}$. $U = \Big[\sum_n (w_n/P_n)^{\frac{1}{\mu}} \frac{1}{L}\Big]^{\mu}$ captures the aggregate utility and is constant across all locations. Note U is absorbed by the constant regressor in reduced-form regression, but does change in general equilibrium. We take into account the changes in U in the counterfactual analysis.

We can express ε_w and ε_l as functions of our structural parameters as follows

$$\varepsilon_w = \frac{\mu\theta - 1}{\theta\big[\mu\theta + \mu + 1\big]}, \qquad \varepsilon_l = \frac{2\theta + 1}{\theta\big[\mu\theta + \mu + 1\big]}. \tag{C.34}$$

Taking the ratio between the two to have

$$\frac{\varepsilon_w}{\varepsilon_l} = \frac{\mu\theta - 1}{2\theta + 1} \Longrightarrow \mu = \frac{1}{\theta} + \frac{2\theta + 1}{\theta} \frac{\varepsilon_w}{\varepsilon_l}.$$

Then, we substitute the above expression into (C.34) to have

$$\begin{aligned}
\varepsilon_w &= \frac{(2\theta+1)\frac{\varepsilon_w}{\varepsilon_l}}{(2\theta+1)+\theta(2\theta+1)\frac{\varepsilon_w}{\varepsilon_l}+(2\theta+1)\frac{\varepsilon_w}{\varepsilon_l}}. \\
&\Longrightarrow \quad \theta(2\theta+1)\varepsilon_w+(2\theta+1)\varepsilon_w+(2\theta+1)\varepsilon_l=2\theta+1. \\
\text{when } 2\theta+1\neq 0 \quad &\Longrightarrow \quad \theta\varepsilon_w+\varepsilon_w+\varepsilon_l=1. \\
&\Longrightarrow \theta=\frac{1-\varepsilon_w-\varepsilon_l}{\varepsilon_w}.
\end{aligned}$$

Substituting $\theta=\frac{1-\varepsilon_w-\varepsilon_l}{\varepsilon_w}$ into $\mu=\frac{1}{\theta}+\frac{2\theta+1}{\theta}\frac{\varepsilon_w}{\varepsilon_l}$, we obtain

$$\mu=\frac{\varepsilon_w}{\varepsilon_l}\frac{2-\varepsilon_w-\varepsilon_l}{1-\varepsilon_w-\varepsilon_l}.$$

Since $\theta>0$ and $\mu>0$, we need parameter restriction $\varepsilon_w+\varepsilon_l<1$. According to (Allen and Arkolakis, 2014), $\varepsilon_w+\varepsilon_l<1$ is a sufficient condition for the model to have a unique equilibrium.