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

Kathryn Baragwanath
Harvard

Gordon H. Hanson
Harvard & NBER

Amit K. Khandelwal
Yale & NBER

Chen Liu
NUS

Hogeun Park
World Bank

April 2024

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 spatially granular economic data are scarce. To demonstrate our method, we use publicly available imagery to evaluate India's road construction projects in the early 2000s. Estimating the model and evaluating welfare impacts only requires remotely-sensed data. We find that India's road investments improved aggregate welfare, particularly for the largest and smallest urban markets. Most welfare gains accrued *within* Indian districts—the unit of analysis in conventional approaches—which demonstrates the benefits of using of high-resolution satellite imagery.

*We thank Treb Allen for valuable comments and seminar participants at the 2019 Banco de Mexico Conference on International Economics. We acknowledge funding from the Center on Global Transformation at UC San Diego, the World Bank, and the International Growth Centre (Project 89448).

1 Introduction

This paper integrates 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* (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 at lower spatial resolution (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, India's urban markets collectively contained 34.8% of the national 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 construction 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 much 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. A comprehensive analysis of India's road expansion should thus include the full extent of the intercity road construction during this period. Using Dijkstra algorithm to calculate the 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. To do so, we use publicly available population rasters from the Gridded Population of the World for the year 2010.

We find that improved market access from road construction over 1996 to 2011 increased the net present value of Indian GDP by 1.60%. Nearly all of this increase comes from the construction of the GQ. Although the expansion of non-GQ national highways increased real GDP by 0.63%, this gain was offset by construction costs,

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

leaving a small negative net impact. Impacts are far from uniform across markets: we uncover a hockey-stick-shaped pattern where the (initially) smallest and largest markets experienced larger gains than medium-size markets. Welfare rose by 2.2% for markets in the first quartile in terms of size (as measured by the sum of nightlights), followed by gains of 2.0% for markets in the second quartile, 2.2% for markets in the third quartile, 2.3% for markets in the fourth quartile, and, notably, of 2.5% and 3.3% for the top 50 markets and top 10 markets, respectively.

We then decompose the variance in welfare gains for different levels of spatial aggregation: 42.6% of the total variance in welfare effects accrues within Indian districts, and the remaining 57.4% accrues across districts. Satellite imagery thus reveals highly localized impacts of national infrastructure investment, which would not be detectable using India's district-level administrative data.

When we disentangle the welfare impacts from the construction of the GQ versus the upgrading of national highways, we find opposing results. While the GQ disproportionately benefited the largest markets in terms of nightlight intensities, the gains from national highway upgrades concentrated on the smallest markets. The GQ produced a welfare gain of 1.0% for the first quartile of markets, compared to a welfare gain of 2.8% for the 10 largest ten markets. By contrast, the expansion of national highways produced a welfare gain of 1.2% in the first quartile of markets, and a smaller 0.4% gain in the top 10 markets. These findings highlight the varying distributional consequences of major road construction projects.

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 the 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-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 regional markets, can be leveraged for analysis at very high spatial resolutions.[2]

Road construction in India has attracted much scholarly interest. In related work,

[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.

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 has affected Indian villages (Aggarwal, 2018; Asher and Novosad, 2020). Because we detect builtup towns and cities and not un-illuminated 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 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. Overall, our results 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 markets through nightlight intensity. We summarize the procedure in Appendix A, and refer the reader to Baragwanath et al. (2019) for more details.

[3] Alder (2016) studies 612 of India's 639 districts in 2001 compared to the 532 districts that contain detectable 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.

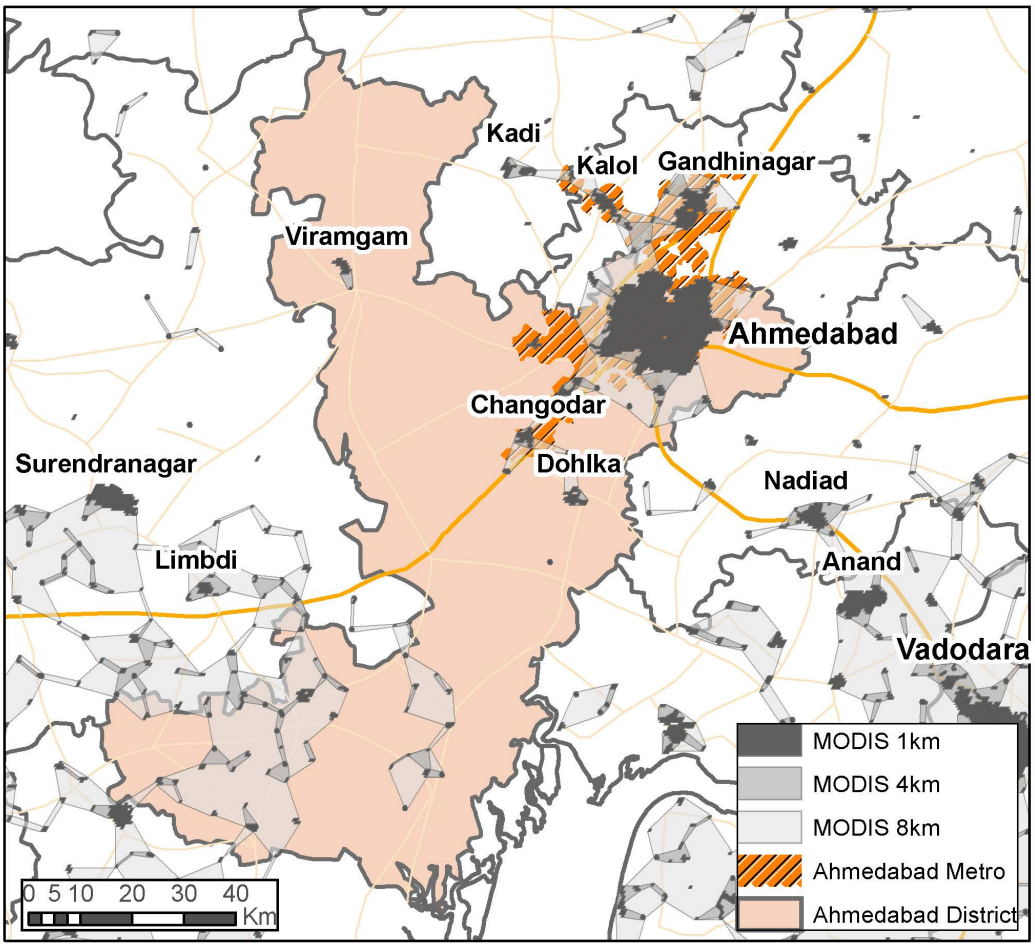
Market Boundaries

To define the boundary of markets, we rely on builtup landcover from the Moderate Resolution Imaging Spectroradiometer (MODIS) 500m resolution 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 larger "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 in order to evaluate the analytical value of having observations at higher versus lower spatial resolutions.

To illustrate our approach, Figure 1 shows local markets around the Ahmedabad 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 Ahmedabad and a total of 383 1km markets in the image.

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



Notes: The figure shows markets 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.

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 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, $\log \mathbf{N}_{rt}$,

$$\log y_{rt} = C + \alpha \log \mathbf{N}_{rt} + \upsilon_{rt} \tag{1}$$

where υ_{rt} is an i.i.d. disturbance associated with measurement error in the use of nightlights to capture GDP. We estimate equation (1) using 2001 data for 326 Indian districts by projecting log GDP on the district's log nightlight value. The regression coefficients are: 0.42 for log nightlights, and 3.91 for the constant, with an R^2 of 0.57. We then use the estimated coefficients to project GDP for each 1km market-year:

$$y_{rt} = \exp\left(3.91 + 0.42 \log \mathbf{N}_{rt}\right). \tag{2}$$

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 convert this elasticity to welfare impacts, 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 GDP per person. Population data are therefore necessary to separately pin down market changes in population versus income.

Population data are publicly available across the globe in datasets such as the Gridded Population of the World - GPW (Center for International Earth Science Information Network - CIESIN - Columbia University, 2016). GPW population data are available at a spatial resolution of 30 arc-seconds (approximately 1 km at the equator), for the years 2000, 2010, 2015, and 2020. We use data from year 2010 in our exercise[6]. We process the GPW data in Google Earth Engine to obtain the total num-

[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, 532 districts overlapped with one of our markets in 2001. Of these, we have GDP estimates for 326 districts.

[6]In a previous draft, we used population data from the 2011 Indian Census rather the Gridded Population of the World (GPW). The results were similar to those presented in the current version. Specifically, we observed a total increase in the Net Present Value (NPV) of GDP by approximately 1.08% (compared to 1.60% in the current version), primarily driven by gains from the construction of

ber of people in each market by adding the population inside each pixel that overlaps with a given market. Table A.1 reports that the average number of people living inside a 1km market increased from 8.9 thousand in 2000 to 10.4 thousand in 2010. As a validation, the largest 1km market in our sample, (corresponding to New Delhi and covering 1,607.6km^2) was assigned a population of 13.1 million in 2000 and 16.0 million in 2010, while the census reported population of New Delhi was 13.8 million in 2001 and 16.8 million in 2011.

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. 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 digitized road maps from the Survey of India, which, prior to 2015, mapped intercity roads for all of India on a roughly decadal basis. We also digitized maps for 1996 and 2011, which span a period from five years before the NHDP to just before its end. 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 above 1 million), while new 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 further summarizes our measures of India's highway and intercity road construction. The GQ added 5,714km of new expressways, and national highways expanded by 19,750km. Although the total increase in road mileage between 1996 and 2011 was relatively modest—the 22,371km of additional roads represented only a 10% increase

the GQ, with a minor negative NPV attributable to NH expansion. The decision to transition to the GPW dataset was motivated by the desire to employ a model that leverages only public (and globally available) data.

in total road length—the quality of roads improved significantly.[7] India's length of highway-quality roads increased by a substantial 86.5% over the period.[8]

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 in 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.

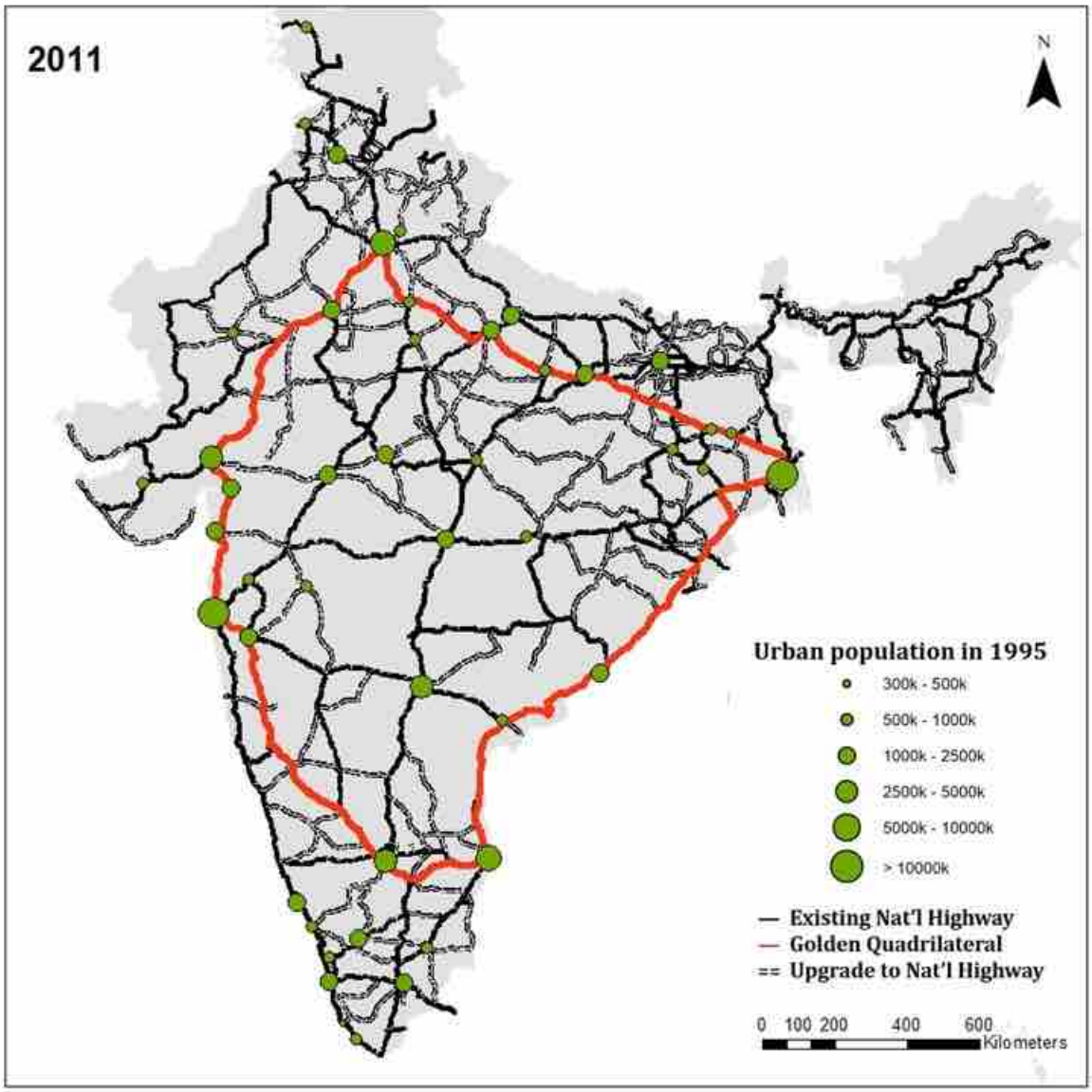
Because the majority of markets are small, most do not have an identifiable road going directly through them. The median market in our data is 2.4km away from an identifiable road in 1996 (interquartile range 3.9km) and 2.2km away from an identifiable road in 2011 (interquartile range 3.4km). Because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

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 apart fell by about 0.5 hours, travel times between markets 200-500km apart fell by about 1 hour, and travel times between markets more than 500km apart fell by more than 3 hours.

[7] Most of the improvements came from upgrading the quality of existing roads to highway-quality roads, not the construction of new roads that would add to total road length. This is why the 22,371km of new roads does not coincide with the 5,714km added by the GQ plus the 19,750km added by the expansion of national highways.

[8] Total highway-quality roads (which include the GQ and other national highways) in 2011 was equal to 5,714km + 49,172km = 54,886km while total highway quality roads in 1996 was equal to 29,442km (see Figure 2 for more details).

Figure 2: Highway Construction in India, 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	235,600

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 Quadrilateral, dashed black for upgrades to national highways), and green dots represent major cities (dot size is proportional to city population in 1995), and (ii) 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 recently 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. We denote the trade elasticity parameter as θ and the inverse labor supply elasticity as μ. For complete details, see Appendix C.

3.1 Equilibrium Conditions in Market Access

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 structural parameters θ and μ. $\hat{\chi}^w_{rt}$ and $\hat{\chi}^l_{rt}$ are functions of unobserved regional amenities and productivity shocks.

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

$$\hat{\Phi}_{rt} = \ln\left(\sum_n \frac{w_{nt}L_{nt}}{\Phi_{nt}}\tau^{-\theta}_{rn,t}\right) - \ln\left(\sum_n \frac{w_{n,t-1}L_{n,t-1}}{\Phi_{n,t-1}}\tau^{-\theta}_{rn,t-1}\right) \tag{4}$$

where $\tau_{rn,t}$ is the symmetric iceberg trade cost between regions r and n at time t.[9] The level of market access at year t is solved from the following system of R equations,

$$\Phi_{rt} = \sum_n \frac{w_{nt}L_{nt}\tau^{-\theta}_{rnt}}{\Phi_{nt}}. \tag{5}$$

In order to compute market access for each region, equation (5) 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):

$$\tau^{-\theta}_{rn} = \left(\text{travel time}_{rn} + \frac{\kappa \times (d_r + d_n)}{\text{speed}}\right)^{-\phi}. \tag{6}$$

[9] 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).

Here, ϕ measures the elasticity of iceberg trade costs with respect to travel time. We set $\phi = 1.5$, following Redding and Rossi-Hansberg (2017). The variable d_r is the last miles traveled that connect market r to observed roadways, and d_n is the last miles traveled that connect market n to the observed roadways. We set $\kappa = 0.000621$ which converts meters into miles. 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

$$\hat{y}_{rt} = \delta \hat{\Phi}_{rt} + \hat{\Gamma}_{rt}. \tag{7}$$

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

$$\hat{y}_{rt} = \alpha \hat{\mathbf{N}}_{rt} + \hat{\upsilon}_{rt}, \tag{8}$$

where we set $\alpha = 0.42$ 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 (5), solving the system of R equations in R unknowns in (5) 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 (7). We follow Faber (2014) by constructing an instrument based on a least-cost approach to connect markets:

$$\hat{\Phi}^{\text{mst}}_{rt} = \ln\left(\sum_{r'} \tfrac{y_{n,t-1}}{\Phi^{mst}_{nt}} \tau^{\text{mst}}_{rn,t} \left(\Omega_{t-1}\right)^{-\theta}\right) - \ln\left(\sum_{n} \tfrac{y_{n,t-1}}{\Phi_{n,t-1}} \tau^{-\theta}_{rn,t-1}\right). \tag{9}$$

This instrument uses GDP at time $t-1$ to calculate market access in both time periods. Additionally, bilateral trade costs, $\tau^{mst}_{rn,t}\left(\Omega_{t-1}\right)$, are constructed by a least-cost spanning tree algorithm to connect India's largest cities using information as of time $t-1$, denoted as Ω_{t-1}. See Appendix B for details. We use India's 180 largest cities in the baseline (cities with population above 100 thousand in 2000). We find similar estimates when using India's 100 largest cities (cities with a population above 350 thousand in 2000), as reported in Appendix D.

The value Φ^{mst}_{nt} is given by the solution to the $R \times R$ system of equations,

$$\Phi_{rt}^{\text{mst}} = \sum_n \frac{y_{n,t-1}}{\Phi_{n,t}^{\text{mst}}} \tau_{rn,t}^{\text{mst}} \left(\Omega_{t-1}\right)^{-\theta}. \tag{10}$$

The instrument maps roads a central planner would have constructed if the only objective was to connect all targeted points (i.e., India's 100 or 180 largest cities) in a network subject to global cost minimization (Faber, 2014). The identifying assumption for this instrument is that the location of non-targeted peripheral markets along the least cost spanning tree network affects changes of market-level economic outcomes only through new highway connections, conditional on district fixed effects.

3.4 Equilibrium Changes in Welfare

To perform counterfactual exercises, we solve the model in changes with 2011 as the baseline (Dekle et al., 2008).[10] We use $\hat{X}$ to denote the ratio between the counterfactual and the actual economy in 2011 of variable X (and henceforth suppress the subscript t). These exercises call for additional data on the population of each market and a parameter value for θ, 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:

$$\hat{\Phi}_r \Phi_r = \left[\sum_n \hat{\Phi}_n^{\frac{1}{\mu}(\varepsilon_w + \frac{1}{\theta})} \frac{L_n}{L}\right]^{-\frac{\theta}{\theta+1}} \left[\sum_n \frac{\hat{\Phi}_n^{\varepsilon_w+\varepsilon_l} y_n}{\hat{\Phi}_n \Phi_n} \tau'^{-\theta}_{rn}\right], \tag{11}$$

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

$$\hat{w}_r = \hat{\Phi}_r^{\varepsilon_w} \left[\sum_n \hat{\Phi}_n^{\frac{1}{\mu}(\varepsilon_w + \frac{1}{\theta})} \frac{L_n}{L}\right]^{\frac{1}{\theta+1}}, \qquad \hat{L}_r = \hat{\Phi}_r^{\varepsilon_l} \left[\sum_n \hat{\Phi}_n^{\frac{1}{\mu}(\varepsilon_w + \frac{1}{\theta})} \frac{L_n}{L}\right]^{-1}. \tag{12}$$

where ε_w and ε_l are functions of the structural parameters μ and θ as

$$\varepsilon_w = \frac{\mu\theta - 1}{\theta\left[\mu\theta + \mu + 1\right]}, \qquad \varepsilon_l = \frac{2\theta + 1}{\theta\left[\mu\theta + \mu + 1\right]}. \tag{13}$$

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

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

[10] Regarding whether to use the first or last period for the baseline, existing literature is agnostic. Caliendo and Parro (2015) calibrate the model to the beginning period (1996 in our case), whereas Adao et al. (2017) calibrate the model to the end period (2011 in our case).

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

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

3.5 Intuition

In our model, the inverse labor elasticity μ governs the extent of population relocation across markets. Two extreme cases are worth mentioning. The first is when the labor supply is elastic—i.e., when μ approaches one. The reduced-form parameters become $\varepsilon_l = (2\theta+1)/(\theta^2+2\theta)$ and $\varepsilon_w = (\theta-1)/(\theta^2+2\theta)$. As seen in equation (3), an improvement in market access translates into higher wages and higher city population, where impacts on population (ε_l) are larger than impacts on wages (ε_w), owing to labor supply being elastic. The second case is when the labor supply is inelastic—i.e., μ approaches infinity. In this case, equation (13) implies that the reduced-form parameters are $\varepsilon_l = 0$ and $\varepsilon_w = 1/(\theta+1)$. Naturally, all changes in market access are absorbed by changes in wages, with no impact on population.

4 Empirical Analysis

In this section, we first present baseline estimates of δ from (7) using $\hat{\Phi}_{rt}^{\text{mst}}$ as an instrument and then discuss estimates under alternative model specifications.

4.1 Baseline Elasticity Estimates

Panel A of Table 1 presents estimation results for equation (7). Column 1 reports the OLS estimate, δ =0.16 (s.e. 0.03). This estimate rises to 0.46 (s.e. 0.16) when using IV, as seen in column 2. The IV results thus indicate an elasticity of GDP with respect to market access of 0.46. All models control for district-fixed effects and initial log GDP to account for mean reversion. We cluster the standard errors at the 4km-buffer level. Given the inclusion of district fixed effects, markets that are the only unit within a district do not contribute to the estimation, which means that the model is estimated using 13,342 out of the 13,387 markets. Our estimate based on within-district variation across markets is comparable to Alder (2016) who estimates δ =0.6 (s.e. 0.23) using district-level variation.

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 for 1994 to 1999 on changes in market access for 2001 to 2011. In Column 3, the elasticity is small and statistically insignificant (δ =0.18 , s.e. 0.18).

Table 1: Baseline Estimates

	(1) OLS	(2) 2SLS	(3) Pre-trend
δ	0.16	0.46	0.18
	(0.03)	(0.16)	(0.18)
Observations	13,342	13,342	11,220
R^2	0.34	0.16	0.18
F		592.8	607.5
Fixed-effects	district	district	district

Notes: The table presents baseline estimates for the effects of changes in market access (due to improvements in India's road network) on market-level GDP as measured by nightlights. Columns 1 and 2 show the elasticity of GDP with respect to structural market access (in changes between 1996 and 2011), using OLS and 2SLS, respectively. Column 3 presents the pre-trend elasticity of GDP on structural market access, based on changes in GDP between 1994 and 1999. All models include district-level fixed effects. The instrumental variable for the 2SLS estimation is constructed based on connecting India's largest 180 cities using a least-cost spanning tree algorithm. Standard errors (clustered by 4km markets) are in parentheses.

4.2 Variation Within Different Geographic Clusters

Our baseline estimates are identified from within-district variation. We compare the estimates for δ using variation within different levels of geographic clusters. Table 2 reports the IV estimates without any fixed effects in column 1, with state-level fixed effects in column 2, and 8km-market fixed effects in column 3.[11]

When comparing across all markets, we find an elasticity of δ =0.47 (s.e. 0.09), which drops to δ =0.29 (s.e. 0.10) when comparing markets within the same states, and becomes δ =0.43 (s.e. 0.17) when comparing markets within the same 8km markets. These estimates are comparable to our baseline estimates, although it appears that including only state fixed effects, the largest geographic unit, leads to an underestimation of the effects. The similar and statistically significant coefficients identified across different levels of geographic variation underscore the stability of the positive relationship between GDP and market access that we estimate.

[11]The number of observations varies across models because some markets drop out of the regression when they are the only unit inside a higher order geographic cluster. Thus, we end up with 13386 markets in the regression without fixed effects, 13385 markets in the regression with state fixed effects, and 11153 markets in the regression with 8km-market fixed effects.

Table 2: 2SLS Estimates under Different Fixed-Effects

	(1) 2SLS	(2) 2SLS	(3) 2SLS
δ	0.47	0.29	0.43
	(0.09)	(0.10)	(0.17)
Observations	13,386	13,385	11,153
R^2	0.18	0.17	0.16
F	1906.5	1093.5	667.8
Fixed-effects	no	state	super8

Notes: The table presents estimates for the effects of changes in market access (due to improvements in India's road network) on market-level GDP as measured by nightlights using alternative sets of regional fixed effects. We set $\theta = 8$. Column 1 shows estimates without fixed effects, column 2 includes state fixed effects, and column 3 includes 8km-market fixed effects. The instrumental variable is constructed based on connecting India's 180 largest cities using a least-cost spanning tree algorithm. Standard errors (clustered by 4km markets) are in parentheses.

4.3 Alternative Specifications

We perform robustness checks using alternative model specifications (for complete results, see Appendix D). First, we use a reduced-form market access measure (Donaldson and Hornbeck, 2016), defined as, $\Phi_{rt}^{\text{reduced-form}} = \sum_n y_{nt} \tau_{rn,t}^{-\theta}$. Results appear in Appendix Table D.1. Estimates are similar to our baseline results. Second, Table D.2 reports estimates using a least-cost spanning tree instrument based on connecting India's 100 largest cities, rather than its largest 180. Results are again similar to our baseline estimates. Finally, we construct market access using alternative values of the elasticity of iceberg trade costs with respect to travel time, ϕ. As a lower bound, we set $\phi = 1.3$, following Monte et al. (2018); as an upper bound, we set $\phi = 1.7$ following Allen and Arkolakis (2022). Appendix Table D.3 reports the estimates of δ. While the estimates of δ vary with the assumed value of ϕ, the results remain statistically significant across specifications.

5 Welfare Analysis

Given the assumed parameter value of θ =8 and the estimated parameter value of δ =0.46, equation (15) implies that the inverse labor supply elasticity is given by μ =0.49.[12] We use these parameter values and the model structure to evaluate the

[12]The welfare impacts are insensitive to values of ϕ but are sensitive to values of θ. In our baseline calibration, we set θ =8 . We also report the results when setting $\theta = 4$ (Simonovska and Waugh, 2014). Compared to our baseline estimates reported in Table 1, the welfare improvements become

impact of India's highway build-up over 1996 to 2011. First, we evaluate the welfare effects of the full transportation network, and study the spatial distribution of these effects. Second, we decompose spatial variation in welfare changes into between and within-region components, which illustrate the benefits of working with granular satellite imagery. Third, we separately evaluate the effects of the construction of the GQ and of India's national highway expansion, which allows us to disentangle the impact of different types of road construction on welfare. Fourth, we evaluate impacts of further upgrading of national highways and local roads. Finally, we incorporate the costs of road construction to perform cost-benefit analysis, and thus evaluate the net return of the infrastructure project as a whole. To implement the analysis, we take our model to the data in 2011, evaluate highway travel times at the level of those for 1996, and then consider alternative changes in India's highway infrastructure.

5.1 Effects of the Full Transportation Upgrade on Welfare by Market Size

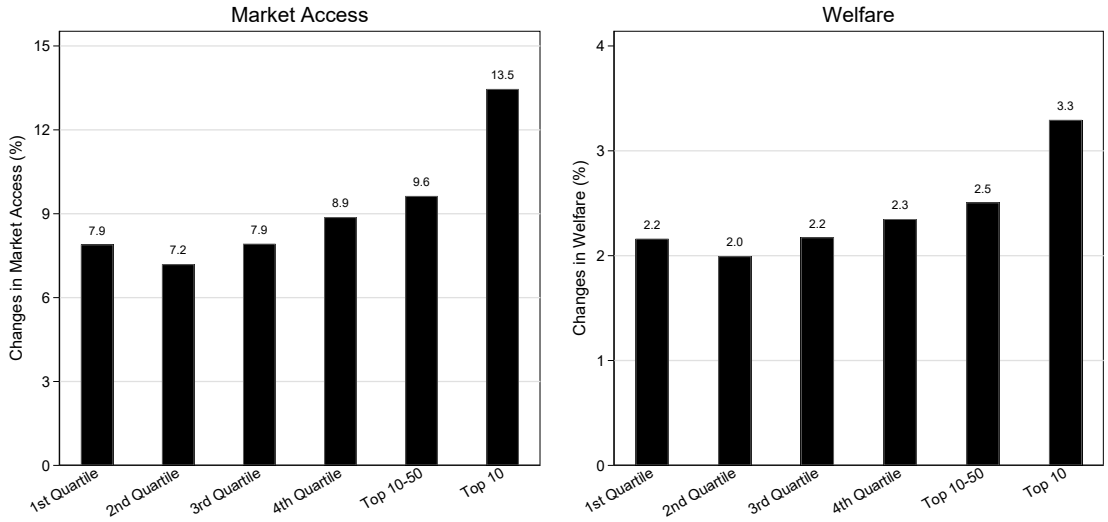
We begin by considering changes in market access and welfare across markets resulting from the full transportation network build-up (i.e., the GQ construction and the expansion of national highways). We group markets into six categories based on their nightlight intensities: the first three groups represent the first, second and third quartiles of markets (which together represent 25.2% of the population in 2000), the fourth group represents the fourth quartile of markets excluding the top 50 markets (29.3% population in 2000), the fifth group represents the top 11-50 markets (31.6% population in 2000), and the sixth group represents the top 10 markets in terms of nightlight intensities (13.9% population in 2000).

In the left panel of Figure 3, which displays percentage changes in market access, it is evident that the gain in market access is greatest for India's very largest urban areas. Market access rises by 13.5% for the top 10 markets, 9.6% for the top 11-50 markets, and 7.2-8.9% for the remaining markets. The right panel of Figure 3 displays percentage changes in welfare. Mirroring the market access results, welfare gains are most pronounced for the largest markets: welfare increases by 3.3% for top 10 markets, 2.5% for top 11-50 markets, and 2.0-2.3% for the remaining markets.

Figure 4 displays the geographic distribution of welfare changes across markets. Panel (a) shows dispersion across all markets, while Panel (b) zooms in on the area surrounding New Delhi, and Panel (c) zooms in on the area surrounding Ahmedabad. Smaller welfare changes are represented by lighter shades of yellow; larger welfare changes are represented by darker shades of red. The thick blue line represents the GQ, while the thinner and lighter blue lines represent national highways.

much larger as θ becomes smaller.

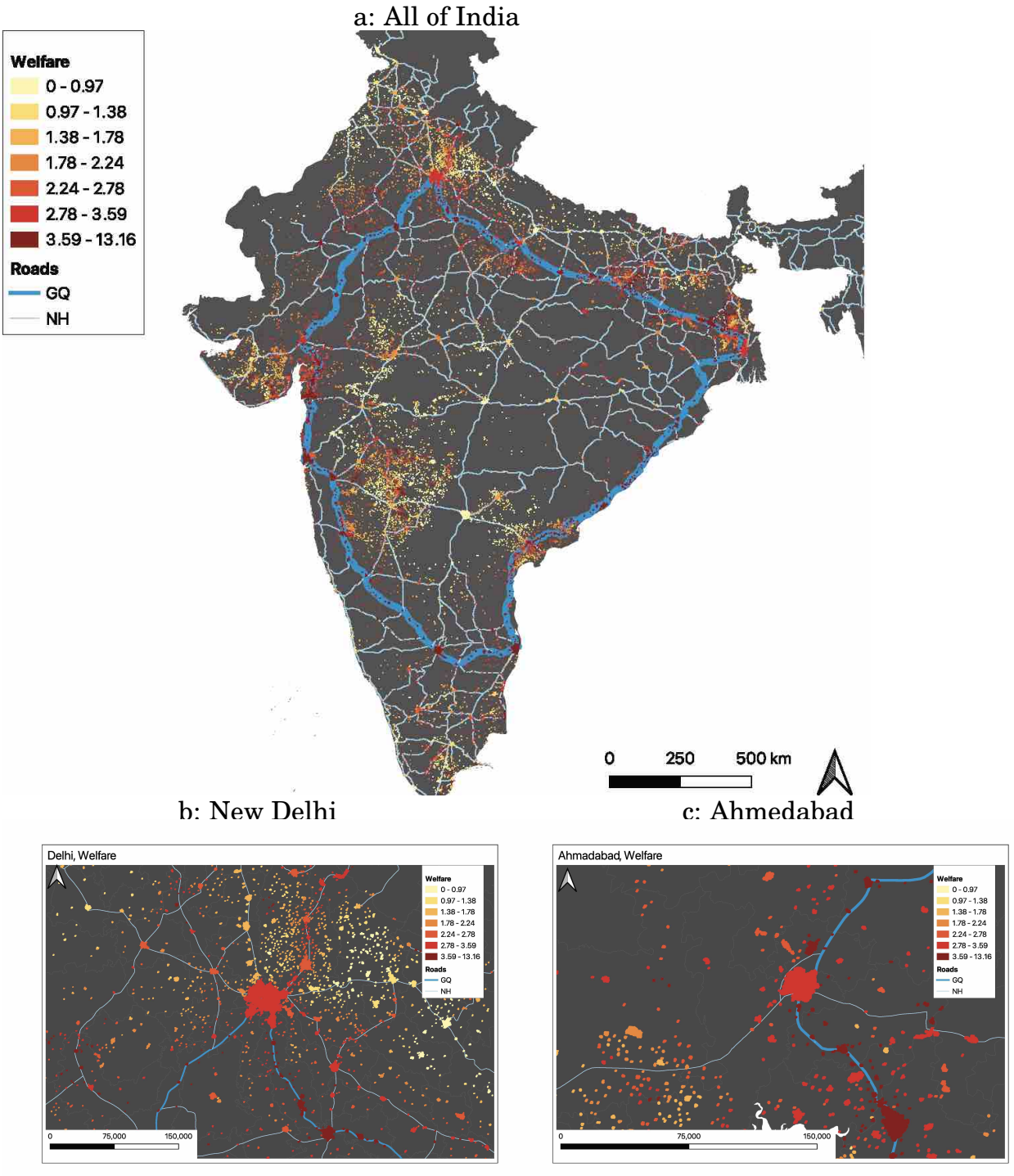
Figure 3: (%) Change in Market Access and Welfare by Market Size



Notes: This figure shows the average change in market access and welfare by markets grouped according to size as measured by nightlight intensity for India's complete transportation upgrade (GQ plus national highway expansion). The 6 groups are the top 10 markets, the top 11-50 markets, and the four quartiles of markets, with the 4th quartile excluding the top 50 markets.

Upon inspection of Figure 4, it is clear that markets along the GQ display the largest welfare changes (shown in dark red), some markets along national highways also show large welfare changes, and less connected markets in the center of India display the smallest welfare changes (shown in pale yellow). Additionally, there is substantial variation in welfare changes across 1km markets within the geographic regions that are typically the unit of analysis in conventional data. For instance, if we focus on Figure 4b, which shows the New Delhi region, we see that the average welfare change for the 1km market corresponding to New Delhi was 3.04%. The market with the smallest welfare change in Panel (b) had a meager 0.54% increase, while the market with the largest change had a sizable 6.36% increase. Similarly, when we zoom in on Ahmedabad in Figure 4c, we see that welfare changes for the market representing Ahmedabad city was 3.56%. The market with the smallest welfare change in the region had a 0.71% increase, while the market with the largest change showed a 7.34% gain. This variation in welfare impacts across markets within a region (5.82 percentage point difference for the New Delhi area, 6.63 percentage point difference for the Ahmedabad area) represents almost half of the difference between markets across all India (13.16 percentage points), a fact which underscores the analytic value of capturing welfare impacts in granular spatial data.

Figure 4: Geographic Distribution of Welfare Changes



Notes: This figure shows the geographic distribution of welfare changes across all urban markets in India for the complete transportation upgrade (GQ plus national highway expansion). Smaller welfare effects are represented by lighter shades of yellow; larger welfare effects are represented by darker shades of red. The thicker blue line represents the GQ; the lighter blue lines represent national highways. Panel (a) shows the geographic distribution for all markets in India, Panel (b) shows effects for the area around New Delhi, and Panel (c) shows effects for the area around Ahmedabad.

5.2 Decomposing Welfare Changes Between vs. Within Regions

Next, we decompose the variance of log changes in welfare into the sum of between and within-region components as follows:

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

where

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

where W_i is change in log welfare in market i, $\bar{\mathrm{W}}$ is the average of W_i across all markets, $\bar{\mathrm{W}}_d$ is the average across markets within administrative unit d (e.g., state or district), N_d is the number of markets within d, and N is the total number of markets. The Within component is calculated as the difference between the Total and Between components. Across all 13,387 1km markets, welfare changes have a mean of 2.04% and a standard deviation of 1.30%.

Table 3: Decomposing Welfare Changes

	A. States and Districts
Between state	18%
Within state	82%
Between district	39.4%
Within district	42.6%

	B. Super-markets
Between 8km markets	63.1%
Within 8km markets	36.9%
Between 4km markets	21.8%
Within 4km markets	15.1%
Within 2km markets	5.6%
Between 2km markets	9.5%

Notes: The table presents a decomposition of changes in welfare between and within regional units in India for the complete transportation upgrade (GQ plus national highway expansion). 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 market boundaries detected from satellite imagery.

In Table 3, we display the variance decomposition. Panel A performs the decomposition by allocating 1km markets into standard Indian administrative units—districts and states. The first row indicates that 18.0% of the overall variation in welfare gains from India's highway infrastructure expansion accrue across states, and the second row shows that 82.0% of the variance accrues across markets *within* states. We further decompose the within-state variation using district administrative boundaries. Here again, much of the variance accrues to the within component. While 39.4% of the overall variance is driven by differences across districts (within-state), 42.6% of

the overall variance comes from welfare differences across 1km markets within districts. Analysis performed at the district level thus may smooth over nearly half of the spatial variation in welfare impacts.

Panel B of Table 3 reports an alternative decomposition that uses the delineation of 8km, 4km, and 2km markets. Recall that the satellite imagery detects 3,483 8km markets in India, and the decomposition shows that 63.1% of variation in welfare changes is between 8km markets. This leaves the remaining 36.9% variation within 8km markets, again indicating that large heterogeneity in impacts of the roads at the sub-district level. Of this 36.9% variation, 21.8% 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 smooth over variation in the impacts of changes in market access.

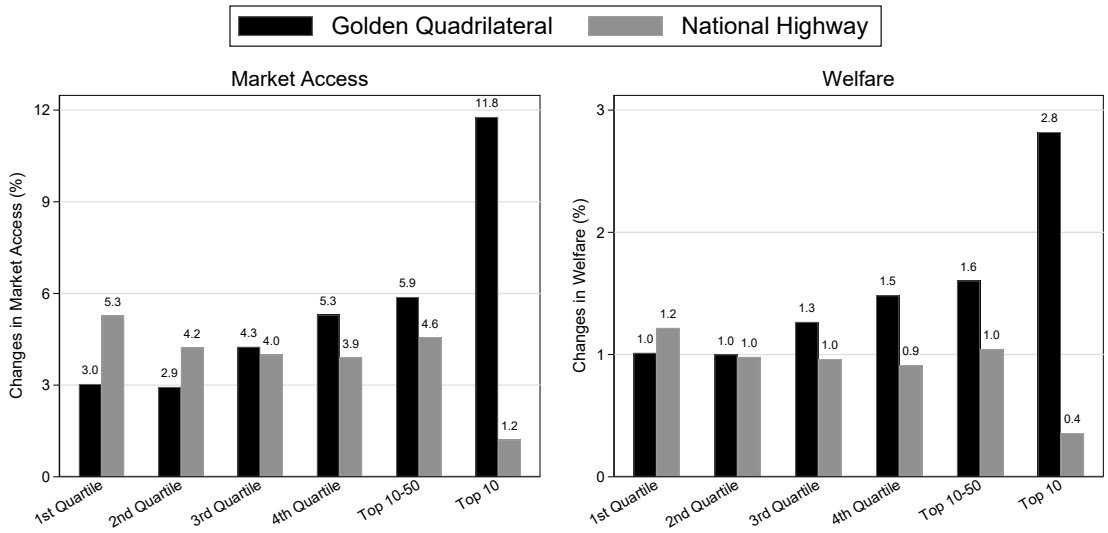
5.3 Disentangling the Effects of the GQ and National Highways

So far, we have analyzed the impacts of India's entire transportation network (the GQ plus national highways). While the effects of the GQ have been previously analyzed using district-level data (Alder, 2016), we are not aware of a welfare comparison between the GQ and national highways. Evaluating the relative economic benefits of the two— the first of which connected India's very largest cities and the second of which connected India's next tier of cities—provides insight into how alternative strategies to expand road networks affect well-being. Worthy of note in this comparison is that the expansion of national highways (in kilometers of lanes built) was larger than that of the GQ (see Section 2.2), where the cost of the national highway buildup was almost double that of the GQ (see Appendix Table B.2).

We compare a counterfactual in which only the GQ is built against one in which only national highways are expanded. Figure 5 displays percentage changes in market access (left panel) and welfare (right panel) across markets grouped according to size. The GQ had larger impacts on welfare than the national highway expansion for all market-size groups except the bottom group, which represents the first quartile of markets in terms of nightlight intensity. The GQ-national highway difference is most pronounced for the very largest markets. Whereas the difference in welfare gains between the GQ and nationals highways ranges from -0.2% (1st quartile of markets by size) to 0.6% (11-50 top markets), the difference is 2.4% for the top 10 markets. Overall, the GQ alone would have led to a 2.2% increase in welfare, while the national highway expansion alone would have led to a 0.6% increase in welfare.

To further explore these differences, Figure 6 displays geographic variation in wel-

Figure 5: Market Access and Welfare Impacts by Highway Type and Market Size



Notes: This figure shows the average change in market access and welfare by markets grouped by size based on one scenario in which only the GQ is built (darker bars) and another scenario in which only national highways are upgraded (lighter bars). The 6 groups are organized according to their nightlight intensities (top 10 markets, top 11-50 markets, and four quartiles with the 4th quartile excluding the top 50 markets).

fare improvements due to the GQ (left panels) and national highways (right panels) for New Delhi (top panels) and Ahmedabad (bottom panels). As expected, welfare gains for the GQ are concentrated in markets that lie close to the construction of the GQ, whereas welfare improvements from national highways accrue mostly in more remote markets that lie along the construction of the network. Overall, the absolute increases in welfare from the GQ tend to be larger than those of national highways, as evidenced by a higher presence of lighter shades of yellow in the right panels of Figure 6. Figures E.2 and E.3 in Appendix E show the geographic distribution of welfare from GQ and NH for all of India, which mirrors the patterns uncovered in Figure 6.

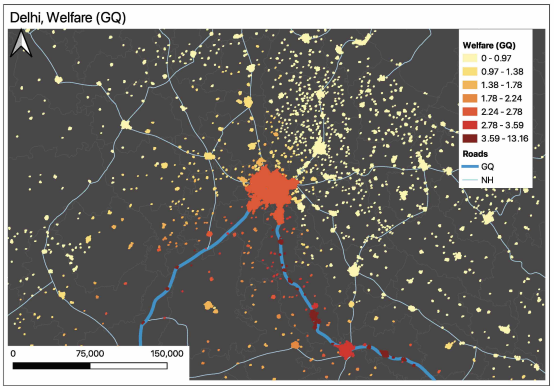
5.4 Cost-Benefit of India's Highway Investment.

Until this point, we have focused on welfare effects without consideration of the costs incurred in the construction of these roads. We next perform a cost-benefit analysis to evaluate the net return of building roads, explicitly considering the costs of construction and maintenance. We calculate the net present value (NPV) of the increase in GDP from the entire road-buildup program, considering benefits from welfare improvements and costs of construction and maintenance.

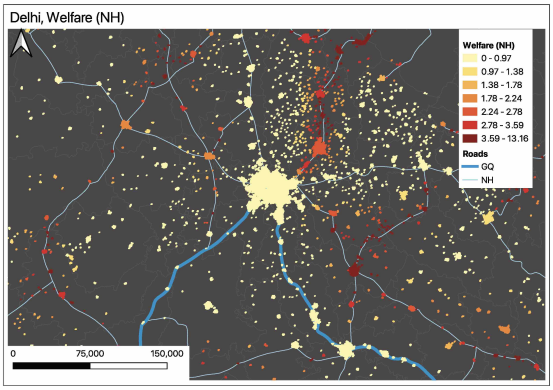
To annualize construction costs, we conduct cost-benefit analysis in terms of NPV (Tsivanidis, 2023). NPV values are computed over a 50-year time horizon using a 5%

Figure 6: Geographic Distribution of Welfare from the GQ and the NH

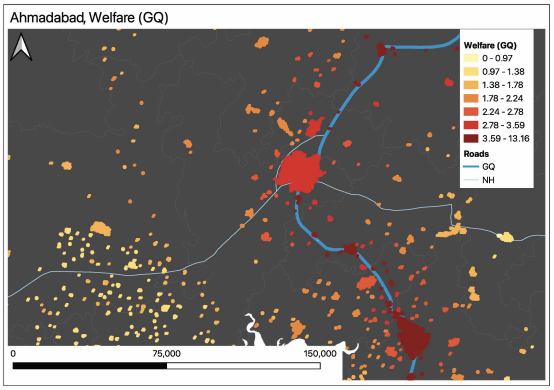
a: New Delhi: Welfare GQ



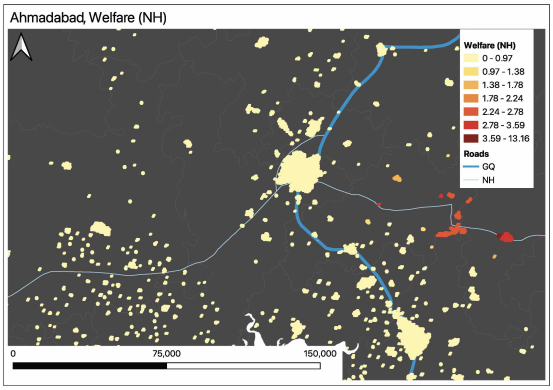
c: New Delhi: Welfare NH



b: Ahmedabad: Welfare GQ



d: Ahmedabad: Welfare NH



Notes: This figure shows the geographic distribution of changes in welfare from the GQ and the national highway expansion (NH) in markets around New Delhi (top two panels) and Ahmedabad (bottom two panels). Smaller welfare changes are represented by lighter shades of yellow while larger welfare changes are represented by darker shades of red. The thicker blue line represents the GQ, while the lighter blue lines represent national highways.

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 compare the actual economy with the observed road network in 2011, our benchmark scenario, to three counterfactual scenarios. Column 1 reports the impacts of the entire road network, which we obtain by comparing the benchmark scenario to an economy with no highway 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).

Table 4: Cost-benefit Calculations By Different Roads

	All Roads	**GQ**	**NH**
NPV Increase in GDP	2.84	2.17	0.63
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	1.60	1.76	-0.21

Notes: The table 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 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.

The first row of Panel B of Table 1 reports that India's highway investment during this period increased annual gross GDP by 2.84%. Column 2 indicates that the contribution of the GQ to overall welfare is 2.17%, or 76.3% of the overall increase. The expansion of non-GQ national highways increased the NPV by 0.63%.[13] Row 2 shows the NPV of the total costs, which has two parts. The first is the construction cost.[14] Table B.2 details the overall construction costs in 1996 USD, which translates to an NPV of $129.93 billion 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 1.76%, 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, although slightly larger in magnitude, to those estimated by Alder (2016), who finds that the construction of the GQ led to gross and net GDP

[13]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.

[14]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.

gains of 1.30% and 0.80%, respectively. This suggests that granular high-resolution satellite data can similarly aggregate estimates uncovered from (district-level) night-light data alone. Additionally, one necessarily requires our granular spatial to evaluate the distribution of welfare changes inside districts.

6 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 digitization of road maps globally (e.g., OpenStreetMap) and the availability of high-resolution population data through such resources as Gridded Population of the World, our method should have wide applicability.

In the much studied Indian context, our analysis delivers similar estimates of aggregate welfare changes to those that use conventional data. This is to be expected, considering that the market access approach is well-tested. Even for India, however, our approach delivers insights that analyses using conventional, lower-resolution data cannot. Our approach reveals the distributional consequences of new infrastructure across spatial markets that vary enormously in size and in their proximity to new roads and highways. We expect that such results would be important to policymakers, considering that both efficiency and equity considerations weigh in how large projects are evaluated and sold to the public.

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.

CENTER FOR INTERNATIONAL EARTH SCIENCE INFORMATION NETWORK - CIESIN - COLUMBIA UNIVERSITY (2016): "Gridded Population of the World, Version 4 (GPWv4): Population Count." .

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.

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

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.

SIMONOVSKA, I. AND M. E. WAUGH (2014): "The elasticity of trade: Estimates and evidence," *Journal of international Economics*, 92, 34–50.

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.

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

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.[15] 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 builtup 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 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

[15]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.

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.[16] 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 operates, at low levels of light intensity, the sensor tends to bunch DN values at either 0 or 5.[17] 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.

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

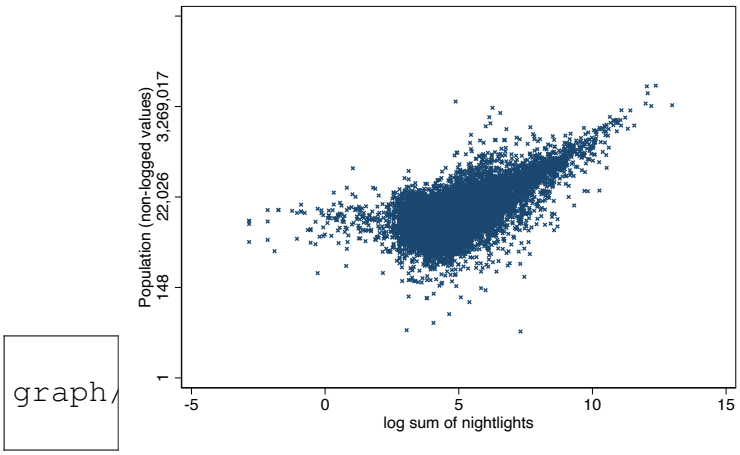
[16]The annual composites of DMSP-OLS cover 1992 to 2013 and include data from six different satellite sensors. New satellites are launched to account for the gradual degradation of sensors, where pairs of successive sensors overlap in space for periods of up to three years. To address differences in 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.

[17]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.

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 0.42 (sd 0.02) for log nightlights, and an R^2 of 0.57 . 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

To calculate population inside each market, we rely on publicly available worldwide population rasters produced by the Gridded Population of the World (GPW). GPW is a "minimally-modeled gridded population data collection that incorporates census population data from the 2010 round of censuses. Population estimates are created by extrapolating the raw census counts to estimates for target years 2000, 2005, 2010, 2015, and 2020" (Center for International Earth Science Information Network - CIESIN - Columbia University, 2016). GPW relies only on population count data from administrative censuses around the world and spatially explicit administrative boundaries to extrapolate population counts at the 30 arcsecond pixel level.

We process the GPW data in Google Earth Engine to obtain the total number of people in each market by adding the population inside each pixel that overlaps with a market. Table A.1 reports the average number of people living inside a 1km market increased from 8.9 thousand in 2000 to 10.4 thousand in 2010. As a validation, the largest market (corresponding to Delhi) was assigned a population of 13.1 million in 2000 and 16.0 million in 2010, while the population of Delhi reported by the Census was 13.8 million in 2001 and 16.8 million in 2011.

In order to check that our markets and population data are capturing real variation, we also use the census conducted by the Indian government in 2001. This step is used as validation. 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 $1,210,854,977$.

Following Baragwanath et al. (2019), to allocate India's population across our markets, we overlay the footprints of the markets with Census shapefiles. 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.

Using this method, we find that 32.7% of India's total population in 2011 were assigned to one of our 1km markets.[18] Additionally, using this method, the average number of people living inside a 1km market increased from 26.8 thousand in 2001 to 29.6 thousand in 2011.

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.

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

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
GPW Population 2000 (in thousands)	13,387	8.9	1.2	203.4	0.0	13,070.8
GPW Population 2010 (in thousands)	13,387	10.4	1.4	232.3	0.0	15,970.1
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
GPW Population 2000 (in thousands)	7,539	18.4	1.6	304.8	0.0	15,014.9
GPW Population 2010 (in thousands)	7,539	21.5	1.9	352.6	0.0	18,538.5
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
GPW Population 2000 (in thousands)	3,483	56.4	2.3	669.6	0.0	26,286.2
GPW Population 2010 (in thousands)	3,483	65.9	2.6	787.1	0.0	32,176.5
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,[19] 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.[20] 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 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

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

[20] 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.

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 an 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, 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

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).

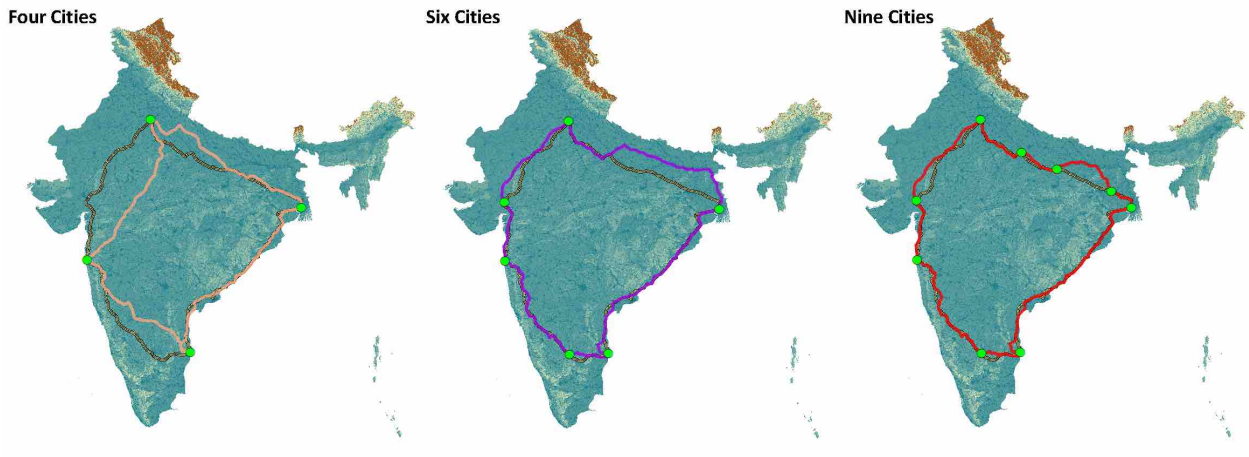
algorithm.

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 Ahmedabad, 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.

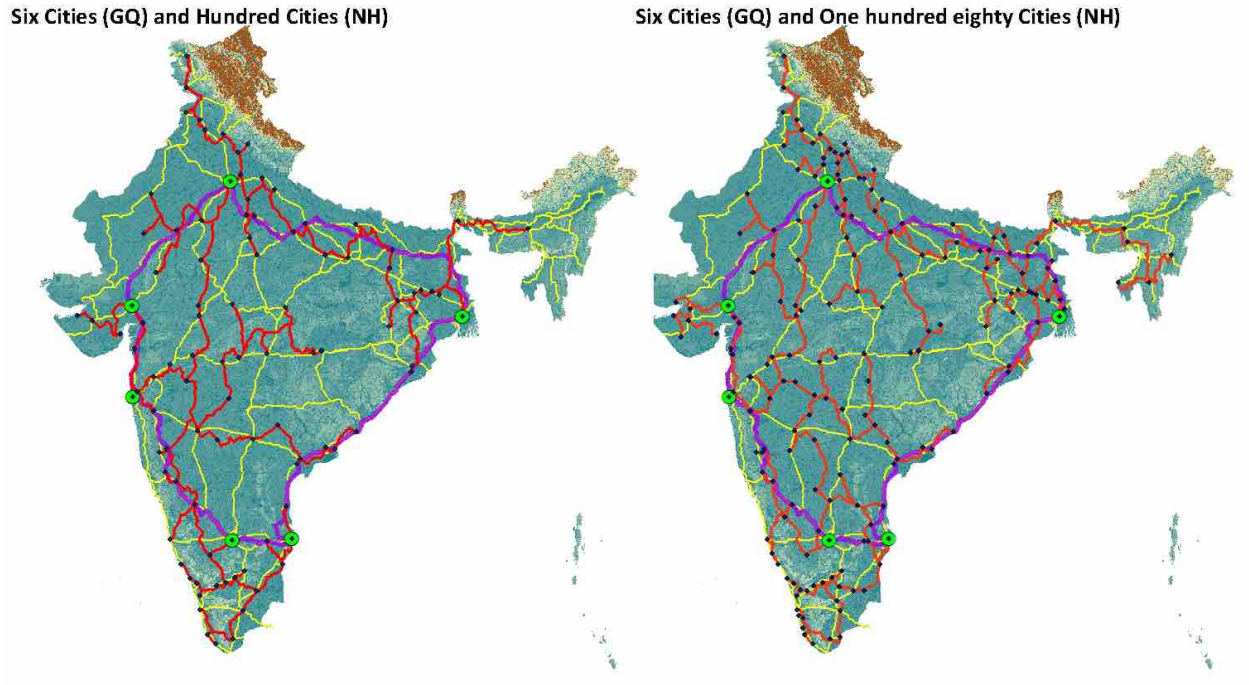
For the construction of other highways, we considered a spanning tree to connect India's 100 largest cities (minimum population in 1991 $> 500,000$) or India's 180 largest cities (minimum population in 1999 $> 200,000$), where we use populations in 1991, ten years before the National Highway Development 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 180-city version as our baseline and the 100-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 (purple), and 3) nine cities (red). For reference, the yellow-black strip is an original path of GQ. 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 six 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 purple line; the existing NH in 1996 is yellow; 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,

$$\frac{L_n}{L} = \frac{(w_n/P_n)^{\frac{1}{\mu}}}{\sum_\ell (w_\ell/P_\ell)^{\frac{1}{\mu}}}.^{21} \tag{C.3}$$

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.

[21] 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 market access, Φ_r, 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.[22]

We add time subscript t for each steady-state equilibrium, and write these conditions in log changes as, (where $\hat{x}_t = \ln(x_t/x_{t-1})$),

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

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

and the change in market access as,

[22] Appendix C.7 shows that

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

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

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

$$\hat{y}_{rt} = \alpha \hat{\mathrm{N}}_{rt} + \upsilon_{rt} \tag{C.12}$$

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 υ_{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

$$\hat{y}_{rt} = \delta \hat{\Phi}_{rt} + \hat{\Gamma}_{rt} \tag{C.13}$$

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

where $\delta \equiv (\varepsilon_w + \varepsilon_l)$; $\hat{\Gamma}_{rt} = \hat{\chi}_{rt}^w + \hat{\chi}_{rt}^l + \upsilon_{rt}$ 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,

$$\Phi_{rt}^{\text{structural}} = \sum_n \tfrac{y_{nt}}{\Phi_{nt}}\tau_{rn,t}^{-\theta}.^{23} \tag{C.15}$$

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 trade cost as a function of travel times below

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

where θ is the trade elasticity to iceberg costs, and ϕ measures the trade elasticity to travel time or distance. We set $\phi = 1.5$ which implies a trade elasticity to the distance of 1.5 used in the literature (see

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

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 $\hat{\Phi}_{rt}$,

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

Following Faber (2014), we project the value of bilateral trade costs in period t, $\tau_{rn,t}$, using $\tau_{rn,t}^{mst}\left(\Omega_{'t-1}\right)$, 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,

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

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 et al., 2008), and set the observed economy of 2011 as the initial equilibrium.[24] 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

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

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

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

[24] 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 et al. (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.

and changes in local price and welfare as

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

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.7)and (C.8) in changes to have

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

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

$\hat{\chi}_i^w = \hat{U}^{\frac{1}{\mu(\theta+1)+1}}$, $\hat{\chi}_i^l = \hat{U}^{\frac{-(\theta+1)}{\mu(\theta+1)+1}}$. $\hat{U}$ 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, $U = \Big[\sum_n (w_n/P_n)^{\frac{1}{\mu}} \frac{1}{L} \Big]^{\mu}$. 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

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

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

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

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

Given $w_n L_n = y_n$, where y_n is measured according to equation (2), we can substitute (C.29) and (C.30) into equation (C.31) to obtain equation (C.19).

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 in equations (C.7) and (C.8), with $\chi_i^w = U^{\frac{1}{\mu(\theta+1)+1}} z_n^{\frac{\mu}{\mu(\theta+1)+1}}$, $\chi_i^l = U^{\frac{-\theta-1}{\mu(\theta+1)+1}} z_n^{\frac{1}{\mu(\theta+1)+1}}$. $U = \left[\sum_n (w_n/P_n)^{\frac{1}{\mu}} \frac{1}{L}\right]^{\mu}$ 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

$$\varepsilon_w = \frac{\mu\theta - 1}{\theta\big[\mu\theta + \mu + 1\big]}, \qquad \varepsilon_l = \frac{2\theta + 1}{\theta\big[\mu\theta + \mu + 1\big]}. \tag{C.32}$$

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 (13) to have

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

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.

D Sensitivity Analysis

D.1 Reduced-form Estimates

We perform sensitivity analysis on our reduced-form elasticity estimates δ under alternative model specifications. First, we reestimates δ using a reduced-form market access measure following Donaldson and Hornbeck (2016), defined as,

$$\Phi_{rt}^{\text{reduced-form}} = \sum_{n} y_{nt} \tau_{rn,t}^{-\theta}. \tag{C.33}$$

In Table D.3, we obtain similar results on δ as the baseline model. Second, Table D.2 reports the estimates using 100 cities in the least-cost spanning tree algorithm. The estimated results are similar. Third, our mapping of travel time to trade costs depends on parameter ϕ, as in Equation (6). In Table D.3, we set $\phi = 1.3$ as the lower bound and $\phi = 1.7$ as the upper bound, following Monte et al. (2018) and (Allen and Arkolakis, 2022). While the estimates of δ vary under different values of ϕ, in both cases, we find statistically significant estimates for δ, while smaller and statistically insignificant estimates for δ using pre-trend.

Table D.1: Estimates Using Reduced-form Market Access

	(1) OLS	(2) 2SLS	(3) Pre-trend
δ	0.17	0.40	0.11
	(0.03)	(0.13)	(0.15)
Observations	13,342	13,342	11,220
R^2	0.34	0.16	0.18
F		698.7	714.1
Fixed-effects	district	district	district

Notes: The table shows estimates using reduced form market access. We set $\theta = 8$. Column 1 shows coefficient estimates on changes in night-lights between 1996 and 2011 using OLS; column 2 shows corresponding coefficient estimates using 2SLS, where the instrumental variable is constructed using 180 cities in the least-cost spanning tree algorithm; column 3 shows coefficients estimates on changes in nightlights between 1994 and 1999 as a test of pre-trends. The fixed effects of 8km-markets are controlled for all models. Standard errors (clustered at 4km markets) are in parenthesis.

Table D.2: Etimates using 100 Largest cities to construct the IV

	(1)	(2)	(3)
	OLS	2SLS	Pre-trend
A. Structural Market Access.			
δ	0.16	0.43	0.21
	(0.03)	(0.21)	(0.32)
Observations	13,342	13,342	11,220
R^2	0.34	0.16	0.18
F		236.5	228.2
Fixed-effects	district	district	district
	(1)	(2)	(3)
	OLS	2SLS	Pre-trend
B. Reduced-form Market Access.			
δ	0.17	0.35	0.11
	(0.03)	(0.16)	(0.24)
Observations	13,342	13,342	11,220
R^2	0.34	0.16	0.18
F		338.9	329.0
Fixed-effects	district	district	district

Notes: The table shows estimates using reduced form market access. We set $\theta = 8$. Column 1 shows coefficient estimates on changes in nightlights between 1996 and 2011 using OLS; column 2 shows corresponding coefficient estimates using 2SLS, where the instrumental variable is constructed using 180 cities in the least-cost spanning tree algorithm; column 3 shows coefficients estimates on changes in nightlights between 1994 and 1999 as a test of pre-trends. The fixed effects of 8km-markets are controlled for all models. Standard errors (clustered at 4km markets) are in parenthesis.

Table D.3: Estimates with Alternative Values of ϕ

	(1)	(2)	(3)
	OLS	2SLS	Pre-trend
	A. Setting $\phi = 1.3$.		
δ	0.22	0.69	0.16
	(0.04)	(0.21)	(0.25)
Observations	13,342	13,342	11,220
R^2	0.34	0.15	0.18
F		592.2	625.2
Fixed-effects	district	district	district
	(1)	(2)	(3)
	OLS	2SLS	Pre-trend
	B. Setting $\phi = 1.7$.		
δ	0.12	0.30	0.17
	(0.02)	(0.12)	(0.14)
Observations	13,342	13,342	11,220
R^2	0.34	0.16	0.18
F		592.9	596.2
Fixed-effects	district	district	district

Notes: The table shows estimates using reduced form market access. We set $\theta = 8$. Column 1 shows coefficient estimates on changes in nightlights between 1996 and 2011 using OLS; column 2 shows corresponding coefficient estimates using 2SLS, where the instrumental variable is constructed using 180 cities in the least-cost spanning tree algorithm; column 3 shows coefficient estimates on changes in nightlights between 1994 and 1999 as a test of pre-trends. The fixed effects of 8km-markets are controlled for all models. Standard errors (clustered at 4km markets) are in parenthesis.

Table D.4: Cost-Benefit Analysis $\theta = 4$.

	All Roads	**GQ**	**NH**
NPV Increase in GDP	5.59	4.27	1.24
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	4.35	3.87	0.41

Notes: The table 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, differentiating the effects of all roads (panel B, column 1), only GQ construction (panel B, column 2) and only national highway construction (Panel B, column 3). GQ counterfactual: keep the road structure as in 2011, but reduce the speed of GQ roads to 40mph (the speed of national highways); NH counterfactual: keep the road structure as in 2011, but reduce the speed of newly built national highway roads to 20mph (the speed of other roads). We set $\theta = 4$, $\delta = 0.46$, $\phi = 1.5$.

E Geographic Distribution of Effects

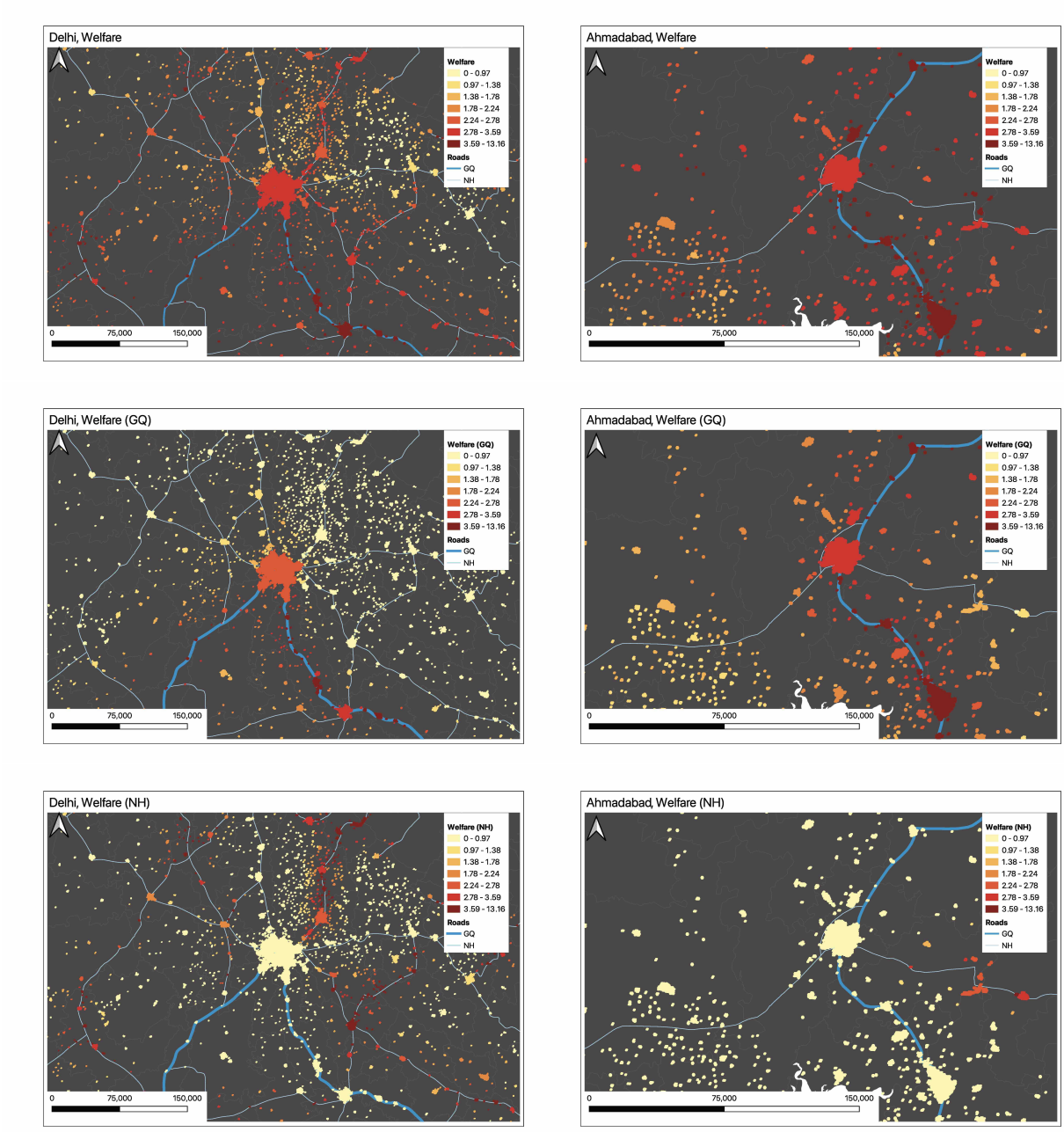


Figure E.1: Welfare Effects in Delhi and Ahmedabad

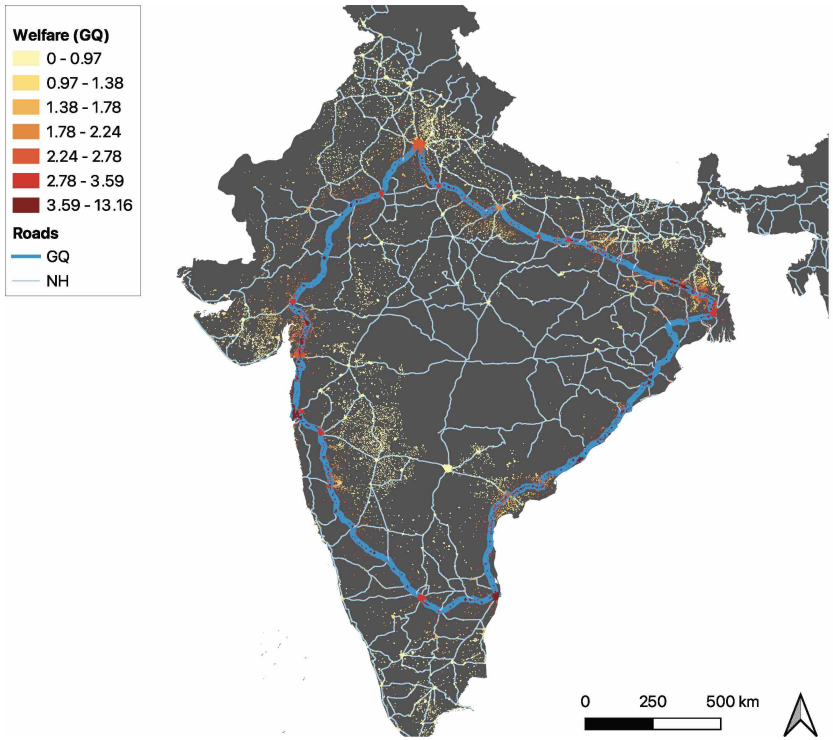


Figure E.2: Welfare Effects from GQ

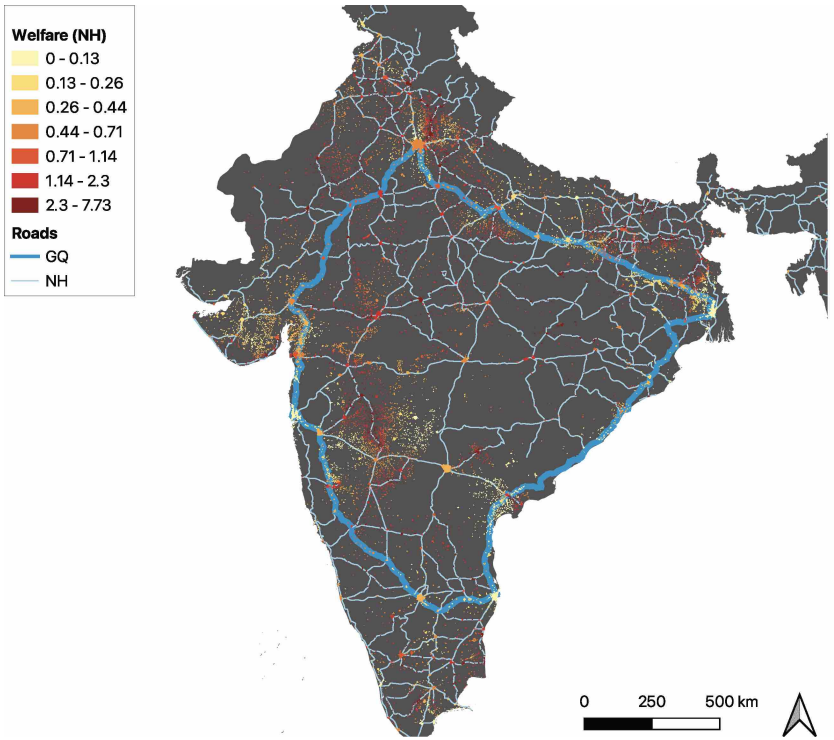


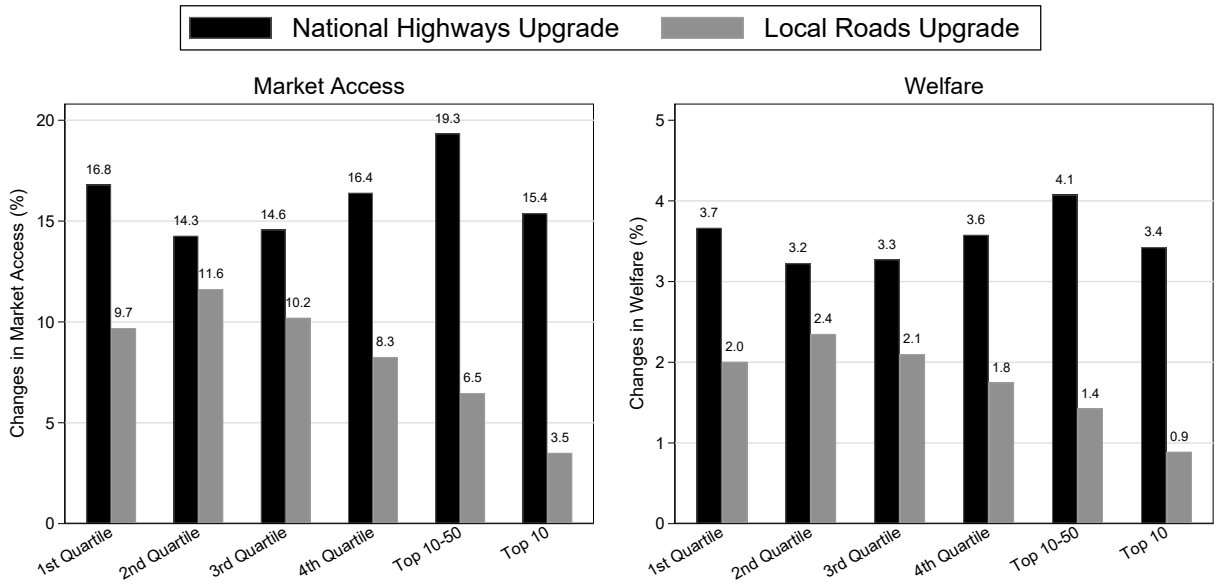
Figure E.3: Welfare Effects from NH

Notes: This figure shows the geographic distribution of welfare effects from the GQ (left panel) and the NH (right panel) for all our markets in India. Lower welfare effects are represented by lighter shades of yellow while higher welfare effects are represented by darker shades of red. The thicker blue line represents the GQ, while the lighter blue lines represent the NH.

F Further Highway Upgrades versus Local Roads Improvements

We explore the potential impact of two further road upgrades. First, we examine the upgrading of all national highways to the quality of GQ roads, which maintains the existing road structure in 2011 but increases the speed of national highways to 60mph (equivalent to that of the GQ). Second, we explore upgrading of all local roads to the quality of national highways, thus keeping the same road structure as in 2011 but improving the speed of local roads to 40mph (equivalent to the speed of existing national highways).

Figure E.4: Impacts of Upgrading Highways vs Local Roads



Notes: This figure shows the average increases in market access and welfare by market-size groups for additional road upgrades. The 6 market size groups are clustered based on their nightlight intensities (top 10 markets, top 11-50 markets, and four quartiles with the 4th quartile excluding the top 50 markets).

Figure E.4 displays the percentage increase in market access and welfare for markets according to their size. A further upgrade of national highways to GQ quality roads is projected to improve market access from 14.28% for the 2nd quartile of markets to 19.35% for the top 10-50 markets, with the most improvement for the top 10-50 markets. If we exclude the top 10 markets, the welfare improvements produced by such a policy present a U-shaped curve, whereby the smallest and largest markets benefit the most from such a road buildup.

When we run the counterfactual exercise for upgrading local roads to NH quality, we see that such a policy would lead to smaller overall welfare improvements compared to a further upgrade of national highways to GQ. Additionally, this type of policy would benefit smaller markets relatively more than larger markets, showing the different distributional effects of these alternative road construction projects.