Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing

Ejaz Ghani, Arti Grover Goswami, and William R. Kerr* †

June 2014

Abstract

We investigate the impact of transportation infrastructure on the organization and efficiency of manufacturing activity. The Golden Quadrangle (GQ) project upgraded a central highway network in India. Manufacturing activity grew disproportionately along the network. These findings hold in straight-line IV frameworks and are not present on a second highway that was planned to be upgraded at the same time as GQ but subsequently delayed. Both entrants and incumbents facilitated the output growth, with scaling among entrants being important. The upgrades facilitated better industrial sorting along the network and improved the allocative efficiency of industries initially positioned on GQ.

JEL Classification: L10, L25, L26, L60, L80, L90, L91, L92, M13, O10, R00, R10, R11, R14

Keywords: Highways, roads, infrastructure, development, entry, growth, misallocation, India, manufacturing.

*Author institutions and contact details: Ghani: World Bank, eghani@worldbank.org; Grover Goswami: World Bank, agrover1@worldbank.org; Kerr: Harvard University, Bank of Finland, and NBER, wkerr@hbs.edu.

†Acknowledgments: We are grateful to Ahmad Ahsan, Nate Baum-Snow, Rachel Griffith, Partha Mukhopadhyay, Stephen O'Connell, Amil Petrin, Jagadeesh Sivadasan, Hyoung Gun Wang, Chris Woodruff, seminar participants, and two referees for helpful suggestions/comments. We are particularly indebted to Katie McWilliams, Sarah Elizabeth Antos, and Henry Jewell for excellent data work and maps. Funding for this project was graciously provided by a Private Enterprise Development in Low-Income Countries grant by the Centre for Economic Policy Research, Harvard Business School, and the World Bank's Multi-Donor Trade Trust Fund. The views expressed here are those of the authors and not of any institution they may be associated with.

1 Introduction

Adequate transportation infrastructure is an essential ingredient for economic development and growth. Rapidly expanding countries like India and China face severe constraints on their transportation infrastructure. Business leaders, policy makers, and academics describe infrastructure as a critical hurdle for sustained growth that must be met with public funding, but to date there is a limited understanding of the economic impact of those projects. We study how proximity to a major new road network affects the organization of manufacturing activity, especially the location of new plants, through industry-level sorting and the efficiency of resource allocation.

We exploit a large-scale highway construction and improvement project in India, the Golden Quadrilateral (GQ) project. The analysis compares districts located 0-10 km from the GQ network to districts 10-50 km away, and we utilize time series variation in the sequence in which districts were upgraded and differences in the characteristics of industries and regions that were affected. Our study employs establishment-level data that provide new insights into the sources of growth and their efficiency improvements.

The GQ upgrades stimulated significant growth in organized manufacturing (formal sector) in the districts along the highway network, even after excluding the four major cities that form the nodal points of the quadrangle. Long-differenced estimations suggest output levels in these districts grew by 49% over the decade after the construction began. This growth is not present in districts 10-50 km from the GQ network nor in districts adjacent to another major Indian highway system that was scheduled for a contemporaneous upgrade but subsequently delayed. We further confirm this growth effect in a variety of robustness checks, including dynamic analyses and straight-line instrumental variables (IV) based upon minimal distances between nodal cities. As the 0-10 km districts contained a third of India's initial manufacturing base, this output growth represented a substantial increase in activity that would have easily covered the costs of the upgrades.

Decomposing these aggregate effects, districts along the highway system experienced a significant boost in the rate of new output formation by young firms, roughly doubling pre-period levels. These entrants were drawn from industries intensive in land and buildings, suggesting the GQ upgrades facilitated sharper industrial sorting between the major nodal cities and the districts along the highway. Despite a substantial increase in entrant counts, the induced entrants maintained comparable size and productivity to control groups. The young cohorts, moreover, demonstrated a post-entry scaling in size that is rare for India and accounted for an important part of the output growth. We also observe heightened output levels from incumbent firms that existed in these 0-10 km districts before the reforms commenced. This growth combines slightly higher survival rates with increases in plant size. Despite this aid to incumbent growth, the incumbent share of local activity declines due to the stronger entry effects.

Looking at industries as a whole, the GQ upgrades improved the allocative efficiency (e.g., Hsieh and Klenow 2009) for industries that were initially positioned along the GQ network. Similar improvements were not present in earlier periods nor for industries that were mostly aligned on the placebo highway system. These results suggest that the GQ upgrades shifted activity towards more productive plants in the most affected industries. Among district traits, the GQ upgrades helped activate intermediate cities of medium population density, where some observers believe India's development has underperformed compared to China. We also find that local education levels were important for explaining the strength of the changes, but that various other potential adjustment costs (e.g., labor regulations) were not.

Our project contributes to the literature on the economic impacts of transportation networks in developing economies, which is unfortunately quite small relative to its policy importance. Two studies consider India and the GQ upgrades specifically. Datta (2011) finds evidence of improved inventory efficiency and input sourcing for manufacturing establishments located on the GQ network almost immediately after the upgrades commenced. These results connect to our emphasis on the GQ upgrade's impact for the organization of formal manufacturing activity. Khanna (2014) examines changes in night-time luminosity around the GQ upgrades, finding evidence for a spreading-out of economic development. Both studies are further discussed below. In related work, Ghani et al. (2012) identify how within-district infrastructure and road quality aid the allocative efficiency of manufacturing activity in local areas between rural and urban sites.

Beyond India, several recent studies find mixed evidence regarding economic effects for non-targeted locations due to transportation infrastructure in China or other developing economies.[1] These studies complement the larger literature on the United States and those undertaken in historical settings.[2] This study is the first to bring plant-level data to the analysis of these highway projects. This granularity is not feasible in the most-studied case of the United States as the major highway projects mostly pre-date the United States' detailed Census data. As a consequence, state-of-the-art work like Chandra and Thompson (2000) and Michaels (2008) utilize aggregate data and broad sectors. The later timing of the Indian reforms affords data that can shed light on many margins like entry behavior, misallocation, and distributions of activity. Moreover, prior work mostly identifies how the existence of transportation networks impacts activity, but we can quantify the impact from investments into improving road networks compared to placebo networks that are not enhanced. This provides powerful empirical identification, and the comparisons are informative for the economic impact of road upgrade investments, which are very large and growing.[3]

The remainder of this paper is as follows: Section 2 gives a synopsis of highways in India and the GQ project. Section 3 describes the data used for this paper and its development. Section 4 presents the empirical work of the paper, determining the impact of highway improvements on economic activity. Section 5 provides a discussion of the results and concludes.

[1] For example, Brown et al. (2008), Ulimwengu et al. (2009), Baum-Snow et al. (2012), Banerjee et al. (2012), Roberts et al. (2012), Baum-Snow and Turner (2013), Aggarwal (2013), Xu and Nakajima (2013), Qin (2014), and Faber (2014).

[2] For example, Fernald (1998), Chandra and Thompson (2000), Lahr et al. (2005), Baum-Snow (2007), Michaels (2008), Holl and Viladecans-Marsal (2011), Hsu and Zhang (2011), Duranton and Turner (2012), Donaldson and Hornbeck (2012), Duranton et al. (2013), Fretz and Gorgas (2013), Holl (2013), and Donaldson (2014).

Related literatures consider non-transportation infrastructure investments in developing economies (e.g., Duflo and Pande 2007, Dinkelman 2011) and the returns to public capital investment (e.g., Aschauer 1989, Munell 1990, Otto and Voss 1994). Several studies evaluate the performance of Indian manufacturing, especially after the liberalization reforms (e.g. Kochhar et al. 2006, Ahluwalia 2000, Besley and Burgess 2004). Some authors argue that Indian manufacturing has been constrained by inadequate infrastructure and that industries that are dependent upon infrastructure have not been able to reap the maximum benefits of the liberalization's reforms (e.g. Gupta et al. 2008, Gupta and Kumar 2010, Mitra et al. 1998).

[3] Through 2006 and including the GQ upgrades, India invested USD 71 billion for the National Highways Development Program to upgrade, rehabilitate, and widen India's major highways to international standards. A recent Committee on Estimates report for the Ministry of Roads, Transport and Highways suggests an ongoing investment need for Indian highways of about USD 15 billion annually for the next 15 to 20 years (*The Economic Times*, April 29, 2012).

2 India's Highways and the Golden Quadrilateral Project

Road transportation accounts for 65% of freight movement and 80% of passenger traffic in India. National highways constitute about 1.7% of this road network, carrying more than 40% of the total traffic volume.[4] To meet its transportation needs, India launched its National Highways Development Project (NHDP) in 2001. This project, the largest highway project ever undertaken by India, aimed at improving the GQ network, the North-South and East-West (NS-EW) Corridors, Port Connectivity, and other projects in several phases. The total length of national highways planned to be upgraded (i.e., strengthened and expanded to four lanes) under the NHDP was 13,494 km; the NHDP also sought to build 1,500 km of new expressways with six or more lanes and 1,000 km of other new national highways. In most cases, the NHDP sought to upgrade a basic infrastructure that existed, rather than build infrastructure where none previously existed.[5]

The NHDP evolved to include seven different phases, and we focus on the first two stages. NHDP Phase I was approved in December 2000 with an initial budget of Rs 30,300 crore (about USD 7 billion in 1999 prices). Phase I planned to improve 5,846 km of the GQ network (its total length), 981 km of the NS-EW highway, and 671 km of other national highways. Phase II was approved in December 2003 at an estimated cost of Rs 34,339 crore (2002 prices). This phase planned to improve 6,161 km of the NS-EW system and 486 km of other national highways. About 442 km of highway is common between the GQ and NS-EW networks.

The GQ network connects the four major cities of Delhi, Mumbai, Chennai, and Kolkata and is the fifth-longest highway in the world. Panel A of Figure 1 provides a map of the GQ network. The GQ upgrades began in 2001, with a target completion date of 2004. To complete the GQ upgrades, 128 separate contracts were awarded. In total, 23% of the work was completed by the end of 2002, 80% by the end of 2004, 95% by the end of 2006, and 98% by the end of 2010. Differences in completion points were due to initial delays in awarding contracts, land acquisition and zoning challenges, funding delays[6], and related contractual problems. Some have also observed that India's construction sector was not fully prepared for a project of this scope. One government report in 2011 estimated the GQ upgrades to be within the original budget.

The NS-EW network, with an aggregate span of 7,300 km, is also shown in Figure 1. This network connects Srinagar in the north to Kanyakumari in the south, and Silchar in the east to Porbandar in the west. Upgrades equivalent to 13% of the NS-EW network were initially planned to begin in Phase I alongside the GQ upgrades, with the remainder scheduled to be completed by 2007. However, work on the NS-EW corridor was pushed into Phase II and later, due to issues with land acquisition, zoning permits, and similar. In total, 2% of the work was completed by the end of 2002, 4% by the end of 2004, and 10% by the end of 2006. These figures include the overlapping portions with the GQ network that represent about 40% of the NS-EW progress by 2006. As of January 2012, 5,945 of the 7,300 kilometers in the NS-EW project had been completed.

[4] Source: National Highway Authority of India website: http://www.nhai.org/. The Committee on Infrastructure continues to project that the growth in demand for road transport in India will be 1.5-2 times faster than that for other modes. Available at: http://www.infrastructure.gov.in. By comparison, highways constitute 5% of the road network in Brazil, Japan, and the United States and 13% in Korea and the United Kingdom (World Road Statistics 2009).

[5] The GQ program in particular sought to upgrade highways to international standards of four- or six-laned, dual-carriageway highways with grade separators and access roads. This group represented 4% of India's highways in 2002, and the GQ work raised this share to 12% by the end of 2006.

[6] The initial two phases were about 90% publicly funded and focused on regional implementation. The NHDP allows for public-private partnerships, which it hopes will become a larger share of future development.

3 Data Preparation

We employ repeated cross-sectional surveys of manufacturing establishments carried out by the government of India. Our work studies the organized sector surveys that were conducted in 1994-95 and in the 11 years stretching from 1999-00 to 2009-10. In all cases, the survey was undertaken over two fiscal years (e.g., the 1994 survey was conducted during 1994-1995), but we will only refer to the initial year for simplicity. This time span allows us three surveys before the GQ upgrades began in 2001, annual observations for five years during which the highway upgrades were being implemented, and annual data from this point until 2009. Estimations typically use 1994 or 2000 as a reference point to measure the impact of GQ upgrades. This section describes some key features of these data.

The organized manufacturing sector of India is composed of establishments with more than ten workers if the establishment uses electricity. If the establishment does not use electricity, the threshold is 20 workers or more. These establishments are required to register under the India Factories Act of 1948. The unorganized manufacturing sector is, by default, comprised of establishments which fall outside the scope of the Factories Act. The organized sector accounts for over 80% of India's manufacturing output, while the unorganized sector accounts for a high share of plants and employment (Ghani et al. 2012). The results reported in this paper focus on the organized sector.[7]

The organized manufacturing sector is surveyed by the Central Statistical Organization through the Annual Survey of Industries (ASI). Establishments are surveyed with state and four-digit National Industry Classification (NIC) stratification. For most of our analysis, we use the provided sample weights to construct population-level estimates of organized manufacturing activity at the district level. Districts are administrative subdivisions of Indian states or union territories that provide more-granular distances from the various highway networks. We also construct population-level estimates of three-digit NIC industries for estimations of allocative efficiency.[8]

ASI surveys record economic characteristics of plants like employment, output, capital, raw materials, and land and building value. For measures of total manufacturing activity in locations, we aggregate the activity of plants up to the district level. We also develop measures of labor productivity and total factor productivity (TFP). Weighted labor productivity is simply the total output divided by the total employment of a district. Unweighted labor productivity is calculated through averages across plants and is used in robustness checks. TFP is calculated primarily through the approach of Sivadasan (2009), who modifies the Olley and Pakes (1996) and Levinsohn and Petrin (2003) methodologies for repeated cross-section data.[9]

[7] In a companion piece, Ghani et al. (2013) also consider the unorganized sector and find a very limited response to the GQ upgrades. There are traces of evidence of the organized sector findings repeating themselves in the unorganized sector (e.g., heightened entry rates, forms of industry sorting discussed below), but the results are substantially diminished in economic magnitudes. These null patterns also hold true regardless of the gender of the business owner in the unorganized sector. This differential is reasonable given the greater optimization in location choice that larger plants conduct and the ability of these plants to trade inputs and outputs at a distance.

[8] For additional detail on the manufacturing survey data, see Nataraj (2011), Kathuria et al. (2010), Fernandes and Pakes (2008), Hasan and Jandoc (2010), and Ghani et al. (2014).

[9] As the Indian data lack plant identifiers, we cannot implement the Olley and Pakes (1996) and Levinsohn and Petrin (2003) methodologies directly since we do not have measures of past plant performance. The key insight from Sivadasan (2009) is that one can restore features of these methodologies by instead using the average productivity in the previous period for a closely matched industry-location-size cell as the predictor for firm productivity in the current period. Once the labor and capital coefficients are recovered using the Sivadasan correction, TFP is estimated as the difference between the actual and the predicted output. This correction removes the simultaneity bias of input choices and unobserved firm-specific productivity shocks. We also consider a residual regression approach as an alternative. For every two-digit NIC industry and year, we regress log value-added (output minus

Repeated cross-sectional data do not allow panel analyses of firms or accurate measures of exiting plants. The data do, however, allow us to measure and study entrants. Plants are distinguished by whether or not they are less than four years old. We will use the term "young" plant to describe the activity of these recent entrants. Estimations also consider incumbent establishments operating in districts from 2000 or earlier.

The sample for long-differenced estimations contains 311 districts. This sample is about half of the total number of districts in India of 630, but it accounts for over 90% of plants, employment, and output in the organized manufacturing sector throughout the period of study. The reductions from the 630 baseline occur due to the following reasons. First, the ASI surveys only record data for about 400 districts due to the lack of organized manufacturing (or its extremely limited presence) in many districts. Second, we drop states that have a small share of organized manufacturing.[10] Finally, we require manufacturing activity be observed in the district in 2000 and 2007/9 to facilitate the long-differenced estimations over a consistent sample.[11]

We measure the distance of districts to various highway networks using official highway maps and ArcMap GIS software. Reported results use the shortest straight-line distance of a district to a given highway network, measured from the district's edge. We find very similar results when using the distance to a given highway network measured from the district centroid. The Empirical Appendix provides additional details on data sources and preparation, with the most attention given to how we map GQ traits that we ascertain at the project level to district-level conditions for pairing with ASI data.[12]

Empirical specifications use a non-parametric approach with respect to distance to estimate treatment effects. We define indicator variables for the shortest distance of a district to the indicated highway network (GQ, NS-EW) being within a specified range. Most specifications use four distance bands: nodal districts, districts located 0-10 km from a highway, districts located 10-50 km from a highway, and districts over 50 km from a highway. In an alternative setup, the last distance band is further broken into 50-125 km, 125-200 km, and over 200 km.

Our focus is on the non-nodal districts of a highway. We measure effects for nodal districts, but the interpretation of these results is difficult as the highway projects are intended to improve the connectivity of the nodal districts. For the GQ network, we follow Datta (2011) in defining the nodal districts as Delhi, Mumbai, Chennai, and Kolkata. In addition, Datta (2011) describes several contiguous suburbs (Gurgaon, Faridabad, Ghaziabad, and NOIDA for Delhi; Thane for Mumbai) as being on the GQ network as "a matter of design rather than fortuitousness." We include these suburbs in the nodal districts. As discussed later when constructing our instrument variables, there is ambiguity evident in Figure 1 about whether Bangalore should also be considered a nodal city. The base analysis follows Datta (2011) and does not include Bangalore, but we return to this question. For the NS-EW network, we define Delhi, Chandigarh, NOIDA, Gurgaon, Faridabad, Ghaziabad, Hyderabad, and Bangalore to be the nodal districts using similar criteria to those applied to the GQ network.

raw materials) of plants on their log employment and log capital, weighting plants by their survey multiplier. The residual from this regression for each plant is taken as its TFP. We then take the average of these residuals across plants for a district.

[10] These excluded states are Andaman and Nicobar Islands, Dadra and Nagar Haveli, Daman and Diu, Jammu and Kashmir, Tripura, Manipur, Meghalaya, Nagaland and Assam. The average share of organized manufacturing from these states varies from 0.2% to 0.5% in terms of establishment counts, employment or output levels. We exclude this group to ensure reasonably well measured plant traits, especially with respect to labor productivity and plant TFP. With respect to the latter, we also exclude plants that have negative value added.

[11] As described below, our dynamic estimations focus on a subset of non-nodal districts continuously observed across all 12 surveys (1994, 1999-2009) and within 50 km of the GQ network.

[12] Appendix materials and tables identified in this paper are available online at http://www.people.hbs.edu/wkerr/.

Table 1 presents simple descriptive statistics that portray some of the empirical results that follow. As we do not need the panel nature of districts for these descriptive exercises, we retain some of the smaller districts that are not continuously measured to provide as complete a picture as possible. The total district count is 363, with the following distances from the GQ network: 9 districts are nodal, 76 districts are 0-10 km away, 42 districts are 10-50 km away, and 236 districts are over 50 km away.

Panel A provides descriptive tabulations from the 1994/2000 data that precede the GQ upgrades, and Panel B provides similar tabulations for the 2005/2007/2009 data that follow the GQ upgrades. Columns 1-3 report aggregates of manufacturing activity within each spatial grouping, averaging the grouped surveys, and Columns 4-6 provide similar figures for young establishments. Columns 7 and 8 document means of productivity metrics. One important observation from these tabulations is that non-nodal districts in close proximity to the highway networks typically account for around 40% of Indian manufacturing activity.

Panels C and D provide some simple calculations. Panel C considers the simple ratio of average activity in 2005/2007/2009 to 1994/2000, combining districts within spatial range. Panel D instead tabulates the change in the share of activity accounted for by that spatial band. Shares of productivity metrics are not a meaningful concept. Starting with the top row of Panel C, the study is set during a period in which growth in manufacturing output exceeds that of plant counts and employment. Also, growth of entrants exceeds that for total firms. Looking at differences in growth patterns by distance from the GQ network, 0-10 km districts exceed 10-50 km districts in every column but total employment growth. Moreover, in most cases, the growth in these very proximate districts also exceeds that in districts over 50 km away. The associated share changes in Panel D tend to be quite strong considering the big increases in the nodal cities that are factored into these share changes.[13]

4 Empirical Analysis of Highways' Impact on Economic Activity

We first consider long-differenced estimations that compare district manufacturing activity before and after the GQ upgrades. We use this approach as well for our placebo analyses and IV estimations. We then turn to dynamic estimations that consider annual data throughout the 1994-2009 period, followed by the industry-level sorting analyses and examinations of allocative efficiency.

4.1 Long-Differenced Estimations

Long-differenced estimations compare district activity in 2000, the year prior to the start of the GQ upgrades, with district activity in 2007 and 2009 (average across the years). About 95% of the GQ upgrades were completed by the end of 2006. We utilize two surveys after the conclusion of most of the GQ upgrades, rather than just our final data point of 2009, to be conservative. Dynamic estimations below find that the 2009 results for many economic outcomes are the largest in districts nearby the GQ network. An average across 2007 and 2009 is a more conservative approach under these conditions. These estimations will also show that benchmarking 1994 or

[13] Appendix Table 1 provides a comparable tabulation organized around distance from the NS-EW highway system. Districts have the following distances from the NS-EW network: 11 districts are nodal, 90 districts are 0-10 km away, 66 districts are 10-50 km away, and 196 districts are over 50 km away. The abnormal growth associated with districts along the GQ network is weaker in districts nearby the NS-EW network, with the districts within 0-10 km of the NS-EW system only outperforming districts 50+ km away in two of the six metrics. Likewise, a direct comparison of the districts within 10 km of the GQ network to those within 10 km of the NS-EW network favors the former in four of the six metrics.

1999 as the reference period would deliver very similar results given the lack of pre-trends surrounding the GQ upgrades.

Indexing districts with i, the specification takes the form:

$$\Delta Y_i = \sum_{d \in D} \beta_d \cdot (0,1) GQDist_{i,d} + \gamma \cdot X_i + \varepsilon_i. \quad (1)$$

The set D contains three distance bands with respect to the GQ network: a nodal district, 0-10 km from the GQ network, and 10-50 km from the GQ network. The excluded category includes districts more than 50 km from the GQ network. The β_d coefficients measure by distance band the average change in outcome Y_i over the 2000-2009 period compared to the reference category.

Most outcome variables Y_i are expressed in logs, with the exception of TFP, which is expressed in unit standard deviations. Estimations report robust standard errors, weight observations by log total district population in 2001, and have 311 observations representing the included districts. We winsorize outcome variables at the 1%/99% level to guard against outliers. Our district sample is constructed such that employment, output, and establishment counts are continuously observed. We do not have this requirement for young plants, and we assign the minimum 1% value for employment, output, and establishment entry rates where zero entry is observed in order to model the extensive margin and maintain a consistent sample.

The long-differenced approach is transparent and allows us to control easily for long-run trends in other traits of districts during the 2000-2009 period. All estimations include as a control the initial level of activity in the district for the appropriate outcome variable Y_i to flexibly capture issues related to economic convergence across districts. In general, however, estimates show very little sensitivity to the inclusion or exclusion of this control. In addition, the vector X_i contains other traits of districts: national highway access, state highway access, broad-gauge railroad access, and district-level measures from 2000 Census of log total population, age profile, female-male sex ratio, population share in urban areas, population share in scheduled castes or tribes, literacy rates, and an index of within-district infrastructure. The variables regarding access to national and state highways and railroads are measured at the end of the period and thus include some effects of the GQ upgrades. The inclusion of these controls in the long-differenced estimation is akin to including time trends interacted with these initial covariates in a standard panel regression analysis.

The column headers of Table 2 list dependent variables. Columns 1-3 present measures of total activity, Columns 4-6 consider new entrants, Columns 7 and 8 document productivity outcomes, and Columns 9 and 10 report wage and labor cost metrics. Panel A reports results with a form of specification (1) that only includes initial values of the outcome variable as a control variable. The first row shows increases in nodal district activity for all metrics. The higher standard errors of these estimates, compared to the rows beneath them, reflect the fact that there are only nine nodal districts. Yet, many of these changes in activity are so substantial in size that one can still reject that the effect is zero. We do not emphasize these results much given that the upgrades were built around the connectivity of the nodal cities. Because the β_d coefficients are being measured for each band relative to districts more than 50 km from the GQ network, the inclusion or exclusion of the nodal districts does not impact results regarding non-nodal districts.

Our primary emphasis is on the highlighted row where we consider non-nodal districts that are 0-10 km from the GQ network. To some degree, the upgrades of the GQ network can be taken as exogenous for these districts. Columns 1-3 find increases in the aggregate activity of these districts. The coefficient on output is particularly strong and suggests a 0.4 log point increase in output levels for districts within 10 km of the GQ network in 2007/9 compared to 2000, relative to districts more than 50 km from the GQ system. As foreshadowed in Table 1, estimates for establishment counts and output in districts 0-10 km from the GQ network exceed the employment responses. These employment effects fall short of being statistically significant at a 10% level, and this is not due to small sample size as we have 76 districts within this range. Generally, the response around the GQ changes favored output over employment, which we trace out further below with industry-level analyses.

Columns 4-6 examine the entry margin by quantifying levels of young establishments and their activity. We find much sharper entry effects than the aggregate effects in Columns 1-3, and these entry results are very precisely measured. The districts within 0-10 km of GQ have a 0.8-1.1 log point increase in entry activity after the GQ upgrade compared to districts more than 50 km away.

Columns 7 and 8 report results for the average labor productivity and TFP in the districts 0-10 km from the GQ network. These average values are weighted and thus primarily driven by the incumbent establishments. Labor productivity for the district increases (also evident in a comparison of Columns 2 and 3). On the other hand, we do not observe TFP growth using the Sivadasan (2009) approach, and unreported estimations find limited differences between the TFP growth of younger and older plants (relative to plants of similar ages in the pre-period). This general theme is repeated below with continued evidence of limited TFP impact but a strong association of the GQ upgrades with higher labor productivity. Columns 9-10 finally show an increase in wages and average labor costs per employee in these districts.

For comparison, the third row of Panel A provides the interactions for the districts that are 10-50 km from the GQ network. None of the effects on the allocation of economic activity that we observe in Columns 1-6 for the 0-10 km districts are observed at this spatial band. This isolated spatial impact provides a first assurance that these effects can be linked to the GQ upgrades rather than other features like regional growth differences. By contrast, Columns 7-10 suggest we should be cautious about placing too much emphasis on the productivity and wage outcomes as being special for districts neighboring the GQ network, since the patterns look pretty similar for all plants within 50 km of the GQ network. On the other hand, it is important to recognize that the productivity/wage growth in Columns 7-10 for the districts 10-50 km are coming from relative declines in activity that are evident in Columns 1-6. That is, the labor productivity of 0-10 km districts is increasing because output is expanding more than employment, but in the 10-50 km districts the labor productivity is increasing due to employment contracting more than output. The different foundations for the productivity and wage changes suggest that we should not reject the potential benefits of the GQ network on these dimensions, and we return to this issue below with a detailed analysis of productivity distributions for entrants and incumbents.

The remaining panels of Table 2 test variations on these themes. Panel B next introduces the longer battery of district traits described above. The inclusion of these controls substantially reduces the coefficients for the nodal districts. More important, they also diminish somewhat the coefficients for the 0-10 km districts, yet these results remain quite statistically and economically important. The controls, moreover, do not explain the differences that we observe between districts 0-10 km from the GQ network and those that are 10-50 km away. Appendix

Table 2 reports the coefficients for these controls for the estimation in Panel B. From hereon, this specification becomes our baseline estimate, with future analyses also controlling for these district covariates.

Panel C further adds in state fixed effects. This is a much more aggressive empirical approach than the baseline estimations as it only considers variation within states (and thus we need to have districts located on the GQ network and those farther away together in individual states). This reduces the economic significance of most variables, and raises the standard errors. Yet, we continue to see evidence suggestive of the GQ upgrades boosting manufacturing activity.

Panel D presents results about the differences in the types of GQ work undertaken. Prior to the GQ project, there existed some infrastructure linking these cities. In a minority of cases, the GQ project built highways where none existed before. In other cases, however, a basic highway existed that could be upgraded. Of the 70 districts lying near the GQ network, new highway stretches comprised some or all of the construction for 33 districts, while 37 districts experienced purely upgrade work. In Panel D, we split the 0-10 km interaction variable for these two types of interventions. The entry results are slightly stronger in the new construction districts, while the labor productivity results favor the road upgrades. This latter effect is strong enough that the total output level grows the most in the road upgrade districts. Despite these intriguing differences, the bigger message from the breakout exercise is the degree to which these two groups are comparable overall.

Panel E extends the spatial horizons studied in Panel B to include two additional distance bands for districts 50-125 km and 125-200 km from the GQ network. These two bands have 48 and 51 districts, respectively. In this extended framework, we measure effects relative to the 97 districts that are more than 200 km from the GQ network. Two key observations can be made. First, the results for districts 0-10 km are very similar when using the new baseline. Second, the null results generally found for districts 10-50 km from the GQ network mostly extend to districts 50-200 km from the GQ network. Even from a simple association perspective, the manufacturing growth in the period surrounding the GQ upgrades is localized in districts along the GQ network.

It is tempting to speculate that the steeper negative point estimates in Columns 4-6 suggest a "hollowing-out" of new entry towards districts more proximate to the GQ system after the upgrades. This pattern would be similar to Chandra and Thompson's (2000) finding that U.S. counties that were next to counties through which U.S. highways were constructed were adversely affected. Chandra and Thompson (2000) described their results within a theoretical model of spatial competition whereby regional highway investments aid the nationally-oriented manufacturing industries and lead to the reallocation of economic activity in more regionally-oriented industries like retail trade. Unreported estimations suggest that this local reallocation is not happening for Indian manufacturing, at least in a very tight geographic sense.[14] For India, the evidence is more consistent with potential diversion of entry coming from more distant points. Either way, the lack of statistical precision for these estimations prevents strong conclusions in this regard.

Appendix Table 3 provides several robustness checks on these results. We first show very similar results when not weighting districts and including dropped outlier observations. We obtain even stronger results on most dimensions when just comparing the 0-10 km band to all districts more than 10 km apart from the GQ network,

[14] This exercise considers districts that lie between 10 and 200 km of the GQ network. Using the long-differenced approach, we regress the change in a district's manufacturing activity and entry rates on the average change in entry rates for the 0-10 km segments within the focal district's state. There is a positive correlation, which is inconsistent with a "hollowing-out" story operating at a very local level.

which is to be expected given the many negative coefficients observed for the 10-50 km band. We also show results that include an additional 10-30 km band. These estimations confirm a very rapid attenuation in effects. The appendix also shows similar (inverted) findings when using a linear distance measure over the 0-50 km range. Appendix Table 4 documents alternative approaches to calculating labor productivity and TFP consequences.

4.2 Comparison of GQ Upgrades to NS-EW Highway

The stability of the results in Table 2 is encouraging, especially to the degree to which they suggest that proximity to the GQ network is not reflecting other traits of districts that could have influenced their economic development. There remains some concern, however, that we may not be able to observe all of the factors that policy makers would have known or used when choosing to upgrade the GQ network and designing the specific layout of the highway system. For example, policy makers might have known about the latent growth potential of regions and attempted to aid that potential through highway development.

We examine this feature by comparing districts proximate to the GQ network to districts proximate to the NS-EW highway network that was not upgraded. The idea behind this comparison is that districts that are at some distance from the GQ network may not be a good control group if they have patterns of evolution that do not mirror what districts immediately on the GQ system would have experienced had the GQ upgrades not occurred. This comparison to the NS-EW corridor provides perhaps a stronger foundation in this regard, especially as its upgrades were planned to start close to those of the GQ network before being delayed. The identification assumption is that unobserved conditions such as regional growth potential along the GQ network were similar to those for the NS-EW system (conditional on covariates).

The upgrades scheduled for the NS-EW project were to start contemporaneous to and after the GQ project. To ensure that we are comparing apples to apples, we identified the segments of the NS-EW project that were to begin with the GQ upgrades and those that were to follow in the next phase. We use separate indicator variables for these two groups so that we can compare against both. Of the 90 districts lying within 0-10 km of the NS-EW network, 40 districts were to be covered in the 48 NS-EW projects identified for Phase I. The empirical appendix provides greater detail on this division.

Table 3 repeats Panel B of Table 2 and adds in four additional indicator variables regarding proximity to the NS-EW system and the planned timing of upgrades. In these estimations, the coefficients are compared to districts more than 50 km from both networks. None of the long-differenced outcomes evident for districts in close proximity to the GQ network are evident for districts in close proximity to the NS-EW network, even if these latter districts were scheduled for a contemporaneous upgrade. The placebo-like coefficients along the NS-EW highway are small and never statistically significant. The lack of precision is not due to too few districts along the NS-EW system, as the district counts are comparable to the distance bands along the GQ network and the standard errors are of very similar magnitude. The null results continue to hold when we combine the NS-EW indicator variables. Said differently, with the precision that we estimate the positive responses along the GQ network, we estimate a lack of change along the NS-EW corridor.

4.3 Straight-Line Instrumental Variables Estimations

Continuing with potential identification challenges, a related worry is that perhaps the GQ planners were better able to shape the layout of the network to touch upon India's growing regions (and maybe the NS-EW planners were not as good at this or had a reduced choice set). Tables 4a and 4b consider this problem using IV techniques. Rather than use the actual layout of the GQ network, we instrument for being 0-10 km from the GQ network with being 0-10 km from a (mostly) straight line between the nodal districts of the GQ network.

The identifying assumption in this IV approach is that endogenous placement choices in terms of weaving the highway towards promising districts (or struggling districts[15]) can be overcome by focusing on what the layout would have been if the network was established based upon minimal distances only. This approach relies on the positions of the nodal cities not being established as a consequence of the transportation network, as the network may have then been developed due to the intervening districts. This is a reverse causality concern, and an intuitive example is the development of cities at low-cost points near to mineral reserves that are accessed by railroad lines. Similar to the straight-line IV used in Banerjee et al. (2012), the four nodal cities of the GQ network were established hundreds or thousands of years ago, making this concern less worrisome in our context.

The exclusion restriction embedded in the straight-line IV is that proximity to the minimum-distance line only affects districts in the post-2000 period due to the likelihood of the district being on the GQ network and experiencing the highway upgrade. This restriction could be violated if the districts along these lines possess characteristics that are otherwise connected to growth during the post-2000 period. For example, these districts could generally have had more-skilled workforces than other districts, and perhaps these educational qualities became more important after 2000. The districts may also have possessed more favorable spatial positions. To guard against these concerns, we will estimate the IV with and without the battery of covariates for district traits in 2000.[16]

Panel B of Figure 1 shows the implementation. IV Route 1 is the simplest approach, connecting the four nodal districts outlined in the original Datta (2011) study. We allow one kink in the segment between Chennai and Kolkata to keep the straight line on dry land. IV Route 1 overlaps with the GQ layout and is distinct in places. We earlier mentioned the question of Bangalore's treatment, which is not listed as a nodal city in the Datta (2011) work. Yet, as IV Route 2 shows, thinking of Bangalore as a nodal city is visually compelling. We thus test two versions of the IV specification, with and without the second kink for Bangalore.

Panel A of Table 4a provides a baseline OLS estimation similar to Panel A of Table 2. For these IV estimations, we drop nodal districts (sample size of 302 districts) and measure all effects relative to districts more than 10 km from the GQ network. This approach only requires us to instrument for a single variable—being within 10 km of

[15] As Duranton and Turner (2011) highlight, endogenous placement could bias findings in either direction. Infrastructure investments may be made to encourage development of regions with high growth potential, which would upwardly bias measurements of economic effects that do not control for this underlying potential. However, there are many cases where infrastructure investments are made to try to turn around and preserve struggling regions. They may also be directed through the political process towards non-optimal locations (i.e., "bridges to nowhere"). These latter scenarios would downward bias results.

[16] Banerjee et al. (2012) provide an early application and discussion of the straight-line IV approach, and Khanna (2014) offers a recent application to India. Faber (2014) provides an important extension to this methodology. Faber uses data on local land characteristics and their impact on construction costs to define a minimum-cost way of connecting 54 key cities that were to be linked by the development of China's highway network. We do not replicate Faber's approach due to our focus on particular segments of India's network and the difficulty assembling the very detailed geographic data necessary for calculating minimum-cost paths (versus minimum distance). We hope in future research to examine the whole system of India's highway network similar to Faber (2014).

the GQ network. Panel B shows the reduced-form estimates, with the coefficient for each route being estimated from a separate regression. The reduced-form estimates resemble the OLS estimates for many outcomes.

The first-stage relationships are quite strong. IV Route 1, which does not connect Bangalore directly, has a first-stage elasticity of 0.43 (0.05) and an associated F-statistic of 74.5. IV Route 2, which treats Bangalore as a connection point, has a first-stage elasticity of 0.54 (0.05) and an associated F-statistic of 138.1. Panel C presents the second-stage results. Not surprisingly, given the strong fit of the first-stage relationships and the directionally similar reduced-form estimates, the IV specifications generally confirm the OLS findings. In most cases, we do not statistically reject the null hypothesis that the OLS and IV results are the same. Wage and labor productivity are the two exceptions, where the IV indicates that OLS underestimates the true impact.

In Table 4b, we repeat this analysis and further introduce the district covariates measured in 2000 that we modelled in Panel B of Table 2.[17] When doing so, the first-stage retains reasonable strength. IV Routes 1 and 2 have associated F-statistics of 13.9 and 20.9, respectively. The covariates have an ambiguous effect on the reduced-form estimates, being very similar for aggregate outcomes, generally lower for entry growth, and generally higher for productivity and wage effects. Most results carry through, although the second-stage coefficients for employment and output entry are substantially lower. Among the controls added, the inclusion of the total population control is the most important for explaining differences between Table 4a and 4b. We again do not statistically reject the null hypothesis that the OLS and IV results are the same for most outcomes. We obtain similar outcomes when also including controls for distance from India's 10 largest cities.

On the whole, we find general confirmation of the OLS findings with these IV estimates, which help with particular concerns about the endogenous weaving of the network towards certain districts with promising potential. IV estimates indicate that there may be an upward bias in the entry findings, perhaps due to endogenous placement towards districts that could support significant new plants in terms of output. A second alternative is that the GQ upgrades themselves had a particular feature that accentuates these metrics (e.g., high output levels of contracted plants to support the actual construction of the road). This latter scenario seems unlikely, however, given the industry-level patterns documented below.

4.4 Dynamic Specifications

Dynamic patterns around these reforms provide additional assurance about the role of the GQ upgrades in these economic outcomes and insight into their timing. A first step is to estimate our basic findings in a pre-post format. We estimate this panel regression using non-nodal districts within 50 km of the GQ network. We thus estimate effects for 0-10 km districts compared to those 10-50 km apart from the GQ highways.[18] Indexing districts with i and time with t, the panel specification takes the form:

$$Y_{i,t} = \beta \cdot (0,1)GQDist_{i,d<10km} \cdot (0,1)PostGQ_t + \phi_i + \eta_t + \varepsilon_{i,t}. \quad (2)$$

[17] We do not include in these estimates the three road and railroad access metrics variables, since these are measured after the reform period, and we want everything in this analysis to be pre-determined. These variables can be included, however, with little actual consequence for Table 4b's findings.

[18] We will be interacting these distance variables with annual metrics, and the reduced set of coefficients is appealing. Our NBER working paper contains earlier results that show similar patterns when several distance bands are interacted with time variables.

The distance indicator variable takes unit value if a district is within 10 km of the GQ network, and the $PostGQ_t$ indicator variable takes unit value in the years 2001 and afterwards. The panel estimations include a vector of district fixed effects ϕ_i and a vector of year fixed effects η_t. The district fixed effects control for the main effects of distance from the GQ network, and the year fixed effects control for the main effects of the post-GQ upgrades period. Thus, the β coefficient quantifies differences in outcomes after the GQ upgrades for those districts within 10 km of the GQ network compared to those 10-50 km away.

Table 5 implements this approach using the 1994, 2000, 2005, 2007, and 2009 data. These estimates cluster standard errors by district, weight districts by log population in 2001, and include 530 observations from the cross of 5 periods with 106 districts where manufacturing plants, employment, and output are continually observed in all five surveys. The results are quite similar to the earlier work, especially for the entry variables. The total activity variables in Columns 1-3 are somewhat diminished, however, and we will later describe the time path of the effects that is responsible for this deviation. The productivity and wage estimations show weaker patterns, which is to be expected given how close the two bands looked in Table 2's analysis. We report them for completeness, but we do not discuss their dynamics further.[19]

Panel B studies the actual completion dates of the GQ upgrades. Due to the size of the GQ project, some sections were completed earlier than other sections. We model this by extending the 0-10 km indicator variable to also reflect whether the district's work was completed by March 2003 (27 districts), between March 2003 and March 2006 (27 districts), or later (16 districts). Columns 1-6 find that the relative sizes of the effects by implementation date are consistent with the project's completion taking hold and influencing economic activity. The results are strongest for sections completed by March 2003, closely followed by those sections completed by March 2006. On the other hand, there is a drop-off in many findings for the last sections completed.

Figure 2 further extends specification (2) to take a non-parametric dynamic format:

$$Y_{i,t} = \sum_{t \in T} \beta_t \cdot (0,1) GQDist_{i,d<10km} \cdot (0,1) Year_t + \phi_i + \eta_t + \varepsilon_{i,t}. \tag{3}$$

Rather than introduce post-GQ upgrades variables, we introduce separate indicator variables for every year starting with 1999. We interact these year indicator variables with the indicator variable for proximity to the GQ network. The vectors of district and year fixed effects continue to absorb the main effects of the interaction terms. Thus, the β_t coefficients in specification (3) quantify annual differences in outcomes for 0-10 km districts compared to those 10-50 km away, with 1994 serving as the reference period. These estimations include 1188 observations as the cross of 12 years with the 99 non-nodal districts within 50 km of the GQ network for which we can always observe their activity.

By separately estimating effects for each year, we can observe whether the growth patterns appear to follow the GQ upgrades hypothesized to cause them. Conceptually, we also believe this dynamic approach is a better way of characterizing the impact of the GQ upgrades than the specific completion dates of segments. Once the upgrades started, work began all along the GQ network and proceeded in parallel. Every state along the GQ network had at least one segment completed within the first two years of the program. Work continued thereafter

[19] Our young plant variables recode entry to the 1% observed value by year if no entry activity is recorded in the data. The 1% value is the winsorization level generally imposed. Appendix Table 5 shows similar results when using a negative binomial estimation approach to model plants and employments as count variables where zero values have meaning.

across all states, with the average spread of completion times between the first and last segments for states being 6.4 years. Since manufacturing activity and location choice decisions can easily be influenced by upgrades on nearby segments (and even anticipation of future upgrades to a segment), we believe it more appropriate to model the GQ event as a whole, timing the impact of all segments from 2001.

Panels A and B of Figure 2 plot the coefficient values for log entrant counts and log new output, respectively, and their 90% confidence bands. These panels include vertical lines to mark when the GQ upgrades began and when they reached the 80% completion mark. The entrant patterns are pretty dramatic. Effects are measured relative to 1994, and we see no differences in 1999 or 2000 for non-nodal districts within 10 km of the GQ compared to those 10-50 km apart for either entrant measure. Once the GQ upgrades commence, the log entry counts in neighboring districts outpace those a bit farther away. These gaps increase throughout the period and are statistically significant in 2004 and 2009. In Panel B, output rises more dramatically and increases up until the upgrades are mostly complete. The differences begin to diminish in 2005 and then stabilize for 2006-2009. New output (and employment) growth substantially lead the new establishment effects, a pattern reflective of large plants being the earliest to respond to the GQ upgrades.

Panels C and D show the series for log total plant counts and output. Aggregate plant counts are very stable before the upgrades start. There is some measure of a downward trend in output levels for 0-10 km districts before the reform, but these pre-results are not statistically different from each other nor from 1994's levels. After the GQ upgrades start, total plant counts and output also climb and then stabilize, before climbing again as the sample period closes. At all points during this post period, the coefficient values are positive, indicating an increase over 1994 levels, but the differences are not statistically significant until the end.

The paths depicted in these figures provide important insights. The young entrant measures in Panels A and B are in essence flow variables into the district. Thus, comparing the post-2006 period to 2004, it is not that the earlier cohort of young firms is shrinking. Instead, a surge of entry occurred as the GQ upgrades made areas more accessible, and with time this surge abated into a lower sustained entry rate that still exceeded pre-reform levels. By contrast, the metrics in Panels C and D are stock variables. Thus, their gradual development over time as more entrants come in and the local base of firms expands makes intuitive sense.

We began in Table 2 by considering long-differenced specifications that compare activity in 2000 with 2007/9. Figure 2 and the appendix material highlight the position of these long-differenced years.[20] The choice of 2000 as a base year is theoretically appropriate as it is immediately before the upgrades began. This choice, however, is not a sensitive point for the analysis. Utilizing 1994 or 1999 delivers a very similar baseline, while the 2001 period would generally lead to larger effects due to the dip in some variables. To this end, the appendix shows that the downward shift in output in Figure 2 is by far the largest pre-movement among the outcomes considered. Encouragingly, there is no evidence of a pre-trend that upward biases our work with any outcome variable.

The choice to average 2007 and 2009 is also illuminated. The dynamics of most aggregate outcomes provide a similar picture to Figure 2. The common themes are a general increase in activity across the post-2002 period, with individual years not statistically significant, and then a run-up as 2009 approaches. By averaging 2007 and 2009, we give a better representation of the aggregate impact than using 2009 alone. The entry margin—where

[20] Appendix Table 6a documents dynamic estimations for all of our outcome variables. Appendix Table 6b also provides comparable results that utilize time since segment completion on an annual basis.

location choices are being made at present—adjusts much faster to the changing attractiveness of regions, and thus registers sharper effects in the short- to medium-run. We return to projections about future impacts from the GQ upgrades in the closing section.

4.5 Entrants and Incumbents

Plant-level data offer the opportunity to examine the roles of entrants and incumbents in aggregate growth—whether the growth is mainly through the displacement of older plants by new entrants, within-plant productivity growth, or some combination of the two. The ideal scenario for this analysis is to have panel data on plants (e.g., Glaeser et al. 2013). While we unfortunately lack this panel structure, we can use information on the ages of plants to consider cohorts over time.

Columns 1-3 of Table 6 consider the role of incumbents in this growth by estimating the log activity in 2007/9 due to plants that have been alive at least 10 years compared to the total initial activity of the district in 2000. The positive coefficient in Column 1 for the 0-10 km group suggests that a greater fraction of the firms already present in the 0-10 km districts by 2000 (i.e., before the GQ upgrades began) survived to 2007/9 than firms in their peer cohorts in districts farther away from the GQ network. Columns 2 and 3 further show that employment and output increased disproportionately for these incumbent firms. Moreover, the relative magnitudes of Columns 1-3 emphasize a point made earlier about the productivity results. For the 0-10 km districts, output is rising at a faster pace than employment, leading towards higher labor productivity at the same time that plant survival is also growing. By contrast, incumbents in the 10-50 km districts are closing at a similar rate or even faster than the control group. These more-distant plants are also shedding employment faster than output. As a result, their labor productivity is also rising, but the origin of this productivity growth is very different from the districts near to the GQ network.

Columns 4-6 quantify the role of entrants by considering as the outcome variable the log activity in 2007/9 due to plants that have been alive fewer than 10 years compared to the total initial activity of the district in 2000. This young firm activity is measured against the same baseline as in Columns 1-3, but is important to note that the relative coefficient sizes measure proportionate effects and thus do not directly rank order aggregate effects (further discussed in the closing section). The outcome measures are all very strong for the 0-10 km districts. There is also some evidence suggestive of larger entrants being less likely to locate in the 10-50 km band.

Thus, both entrants and incumbents contribute to the aggregate growth evident in Table 2. The last three columns consider as an outcome variable the share of activity in each district in 2007/9 contained in firms that have been alive ten or more years. Despite their better survival rates and growth compared to distant incumbent peers, the share of activity accounted for by incumbent firms in districts along the GQ network declines.[21]

Table 7 further analyzes the productivity distributions and selection margins in districts by distance from the GQ network. We first normalize the plant-level productivity estimates developed using the Sivadasan (2009)

[21] We find similar results if grouping firms by whether or not they were specifically alive in 2000, but the ten-year bar on firm age allows us to apply a consistent threshold across the 2007 and 2009 surveys. In a small number of districts, activity is not observed for either entrants or incumbents (but at least one group is always observed). In these cases, we recode the zero value in 2007/9 with the lowest observed proportion among districts with reported data on that margin. That is, if the lowest observed incumbent employment proportion is 5% of the initial 2000 district size, we use this 5% estimate for the districts where zero incumbents are observed. This approach maintains a consistent sample. We find very similar results when excluding these cases, with the one change being that the output contributions of incumbents and entrants become substantially closer in size.

methodology by dividing by the employment-weighted average productivity estimate for an industry-year. We then calculate in Column 1 the average normalized TFP in 2000 for plants within districts by distance from the GQ network. These entries sum over all industries and plants within each district group, weighting individual observations by employment levels. Normalized productivity levels are naturally centered around one and are somewhat higher in nodal districts—a typical finding in urban productivity patterns—with the further initial differences over the other distance bands being marginal.

Column 2 provides a similar calculation in 2007/9. The normalization process again centers values around one, such that aggregate TFP growth is removed at the industry level. The percentage listed next to each entry in Column 2 is the average value in 2007/9 compared to that in 2000. Overall, there is limited movement for any of the groups; the 0-10 km range increases slightly, while the other three ranges show very small declines. This pattern is possible because the 0-10 km group is becoming larger during the period of study.

The more interesting tabulations are in Columns 3 and 4. In Column 3, we calculate these TFP averages for plants that are at least ten years old in 2007/9, while Column 4 presents the comparable figures for plants less than ten years old. Productivity rises with plant age, such that the values in Column 3 are higher than in Column 4. This may be due to differences in technical efficiency. As pointed out in Foster et al. (2008), these differences could also result from using revenues for TFP calculations rather than physical products, if, for example, young firms have lower prices to build demand. Either way, our focus is on the relative comparisons back to the initial 2000 period that are expressed in the accompanying percentages.

Column 3 shows that surviving incumbent plants in the 10-50 km range from the GQ network have substantially higher TFP compared to initial values than in the other district ranges. These districts have reduced entry rates, and the entrants have lower TFPs compared to the other bands. By contrast, the TFP distributions for the 0-10 districts have a more homogeneous adjustment over entrants and incumbents. The stability in entrant TFP in these districts is important given the massive increase in entry rates associated with the GQ upgrades. Despite these surges, the TFP positions are not weakening compared to districts that are 50 km or more from the GQ network. This comparison shows again the very different sources of productivity development in the 0-10 km versus 10-50 km ranges surrounding the GQ network with the upgrades.

4.6 District and Industry Heterogeneity

The remaining analyses quantify heterogeneity in effects for districts and industries. Table 8 considers district heterogeneity using the long-differenced specification (1). Articulating this heterogeneity is challenging empirically because the data variation becomes very thin as one begins to partition the sample by additional traits beyond proximity to the GQ network. We take a simple approach by allowing the coefficient on 0-10 km districts to vary by whether the district is above or below the median value for a trait. Panel A reports the baseline estimation, and we include unreported main effects for interactions in Panels B-E.

Panels B and C document the two key dimensions that we have identified. Districts along the GQ network with higher population density and literacy rates show a stronger response. Given that these density levels are less than in nodal cities that are excluded from the analysis, this response provides some support for the hypothesis that intermediate-sized districts were particularly aided by the GQ infrastructure.[22] By contrast, Panels D and E do not find prominent differences when looking at within-district infrastructure levels or distances along the GQ network from nodal cities.[23] Finally, unreported analyses investigate whether labor regulations play a role in these adjustment patterns. Using the employment protection and industrial dispute resolution laws from Ahsan and Pages (2008), we do not find evidence that districts located in states with above or below average stringency to their labor regulations respond differently to each other. At least on this widely discussed policy dimension (Besley and Burgess 2004), local policy conditions display a weaker connection than workforce factors like literacy rates, which could perhaps themselves be seen as barriers of adjustment. We do not push this interpretation strongly given that we are unable to assess other possible dimensions like entry regulations or corruption levels.

Table 9a describes a key feature of the industry heterogeneity in entry that occurred after the GQ upgrades. We focus specifically on the land and building intensity of industries. We select this intensity due to the intuitive inter-relationship that non-nodal districts may have with nodal cities along the GQ network due to the general greater availability of land outside of urban centers and its cheaper prices. This general urban-rural or core-periphery pattern is evident in many countries and is associated with efficient sorting of industry placement. Moreover, this feature has particular importance in India due to government control over land and building rights, leading some observers to state that India has transitioned from its "license Raj" to a "rents Raj" (e.g., Subramanian 2012a,b). Given India's distorted land markets, the heightened connectivity brought about by the GQ upgrades may be particularly important for efficient sorting of industry across spatial locations. We measure land and building intensity at the national level in the year 2000 through the industry's land and building value per unit of output (listed in Appendix Table 7).

In Table 9a, we repeat the entry specifications isolating activity observed for industries in three bins: those with low land intensity (the bottom quartile of intensity), medium intensity (the second quartile), and high intensity (the top two quartiles). These estimations use the long-differenced approach in specification (1). The 0-10 km districts show a pronounced growth in entry by industries that are land and building intensive. Especially for young firm establishments and output, the adjustment is weaker among plants with limited land and building intensities compared to the top half (there are no important differences between the two quartiles in the top half). As remarkable, the opposite pattern is generally observed in the top row for nodal districts—where nodal districts are experiencing heightened entry of industries that are less land and building intensive after the GQ upgrades—and no consistent patterns are observed for districts 10-50 km from the GQ network. Table 9b shows a similar picture after including district controls and state fixed effects, and Appendix Tables 8a and 8b show instead a weak or opposite relationship is evident with labor and materials intensity. Using capital intensity to group industries not surprisingly gives similar results to land and building intensity. These patterns suggest that

[22] Our NBER working paper contains further evidence regarding the intermediate city dimension. This pattern would be similar to Baum-Snow et al. (2012) for China and Henderson et al. (2001) for Korea. See also World Development Report (2009), Henderson (2010), Desmet et al. (2012), and McKinsey Global Institute (2010, 2012). Related work on spatial ranges includes Duranton and Puga (2001, 2004), Rosenthal and Strange (2004), Ellison et al. (2010), and Gill and Goh (2010).

[23] Ghani et al. (2013) document a nuance of this latter effect that compares urban and rural portions of districts along the GQ network. The study finds that the organized sector's uniform advancement along the GQ system in Table 8 is composed of greater advancement in urban areas in districts closer to the nodal cities, while rural areas are more activated in districts distant from nodal cities. Thus, it appears that different types of industry were able to take advantage of the development of the GQ network in different ways. Urban places close to nodal cities became more attractive to avoid the higher rents and regulations, while rural places also became increasingly attractive for very land-intensive industries.

the GQ upgrades aided the efficient sorting of industries across locations along the network.

4.7 Changes in Allocative Efficiency

Our final exercise takes up directly the allocative efficiency of the Indian economy. In a very influential paper, Hsieh and Klenow (2009) describe the degree to which India and China have a misallocation of activity toward unproductive manufacturing plants. That is, India has too much employment in plants that have low efficiency, and it has too little employment in plants with high efficiency levels. We evaluate whether the GQ upgrades are connected with improvements in allocative efficiency for industries that were mostly located near the GQ network in 2000, compared to those that were mostly off of the GQ network. The hypothesis is that allocative efficiency will improve most in industries that were initially positioned around the GQ network. This could be due to internal plant improvements in operations, increases in competition and the entry/exit of plants, and adjustments in price distortions.

Quantifying improvements in allocative efficiency is quite different than the district-level empirics undertaken thus far as we must look at the industry's production structure as a whole. We thus calculate for the 55 three-digit industries in the manufacturing sample a measure of their allocative efficiency in 1994, 2000, and 2007/9. This measure is calculated as the negative of the standard deviation of TFP across the plants in an industry. Thus, a reduction in the spread of TFP is taken as an improvement in allocative efficiency.[24]

Panel A of Figure 3 plots the change in allocative efficiency (larger numbers being improvements in unit standard deviations) from 2000 to 2007/9 for industries against the share of employment for the industry that was within 200 km of the GQ network in 2000. There is an upward slope in this relationship, providing some broad confirmation for the hypothesis. Panel B shows that this relationship is not evident in terms of proximity to the NS-EW system. Panels C and D repeat these graphs using the share of output within 200 km of the two highway systems. Industries that were in closer proximity to the GQ system in 2000 exhibit sharper improvements in allocative efficiency from 2000 to 2007/9.

Table 10 provides variants of these figures. Panel A considers proximity to the GQ network, while Panel B considers industrial proximity to the NS-EW network. Column headers indicate outcome variables. Column 1 continues with our baseline estimates of changes from 2000 to 2007/9 with the underlying TFP estimates using the Sivadasan (2009) methodology. Column 2 instead substitutes a measure of growth in allocative efficiency that builds from a residual TFP calculation approach at the plant level. Column 3 considers the change in allocative efficiency across the earlier period of 1994 to 2000, before the GQ upgrades began.

Each entry in the table is from a separate regression with the row header describing the metric used to estimate proximity to the GQ network. The point estimates from these various techniques are reasonably similar, generally being larger for estimations that consider employment or output proximity. The measures built upon total output proximity or upon young firm activity in 2000 show the strongest statistical precision. They suggest that each

[24] Hsieh and Klenow (2009) calculate their TFP measures as revenue productivity (TFPR) and physical productivity (TFPQ). In their model, revenue productivity (the product of physical productivity and a firm's output price) should be equated across firms in the absence of distortions. Hsieh and Klenow (2009) use the extent that TFPR differs across plants as a metric of plant-level distortions. When TFPQ and TFPR are jointly log normally distributed, there is a simple closed-form expression for aggregate TFP. In this case, the negative effect of distortions on aggregate TFP can be summarized by the variance of log TFPR. Intuitively, the extent of misallocation is worse when there is greater dispersion of marginal products. The standard deviation measure picks up this feature.

10% increase in the share of an industry's activity near the GQ network in 2000 is associated with about a 0.07 unit standard deviation increase in allocative efficiency to 2007/9.

Column 3 does not find evidence of a link before 2000, and we find null results in Panel B's focus on proximity to the NS-EW network. These results are robust to controlling for the land and building intensity of an industry, calculating proximity to the GQ network using a 50 km range, and similar exercises. With 55 data points, there are natural limits on the extensions and robustness checks that can be undertaken, but these exercises provide some confidence in the conclusion that the GQ upgrades had a positive impact for the allocative efficiency of India's manufacturing sector. This impact may have been particularly strong for industries where new activity was already occurring in a modest band around the network.[25]

5 Discussion and Conclusions

This study finds that the GQ upgrades led to a substantial increase in manufacturing activity. This growth included higher entry rates, incumbent productivity expansion, adjustments in the spatial sorting of industries, and improved allocative efficiency in the manufacturing industries initially located along the GQ network. We close this paper by further discussing the economic magnitudes of the GQ reforms by the end of the sample period and what might lie ahead. These discussions, by their nature, require stepping beyond the econometric analyses conducted, and the assumptions made below are important for the insights derived. We focus on output for this analysis. The organized sector accounts for over 80% of Indian manufacturing output, and much of the work in this paper points to a central connection of the GQ upgrades to output growth.

We start with the impact of the upgrades for the 0-10 km districts. Our preferred estimates of output impacts to the end of the sample period are the 0.43 (0.16) and 0.37 (0.21) coefficients in Panels B and C of Table 2, respectively. These estimates are quite robust to specification checks, and they grow somewhat in magnitude in the IV analyses. Taking a mid-point of 0.4 suggests a 49% overall output increase from initial values for the average district located near the GQ network. Compared to pre-period levels, this would be an increase of output levels for the average district from USD 1.8 billion to USD 2.6 billion. The actual increase, for reference, was to USD 3.8 billion. The estimates would thus credit about 43% of the observed increase to the GQ upgrades, with the rest due to general expansion of Indian manufacturing and the accentuated development of manufacturing along other dimensions that these districts possess (e.g., education levels).

Looking forward, it seems that the scope for further increase in output is relatively modest. While Panel D of Figure 2 shows sustained growth from 2006-2009 in total output levels, we expect this upward trend to flatten and stabilize for the long-term. This forecast comes from consideration of the drivers of growth: 1) continued new entry along the GQ network, 2) the growth to scale of past entrants to these districts since the upgrades, and 3) the growth of incumbents present from before 2000. The second and third effects combine multiple elements—selection and exit among firms, output expansion among survivors and reallocation, and within-firm

[25] Unreported estimations suggest that increases in price competition may have played a role. We generally find accentuated declines in output price dispersion over the 2000 to 2007/9 period for industries located closer to the GQ network. As one example, every 10% increase in initial industry employment within 200 km of the GQ network is associated with a -0.14 (0.09) change in output price dispersion, with the latter measured in unit standard deviations. These results, however, are not statistically significant, and we do not observe quality dispersions nor consumer prices by region. As such, we do not strongly emphasize this channel but note suggestive evidence in this regard.

productivity growth. Repeated cross-sectional data cannot perfectly separate these effects, but we can provide some informative calculations.

The first effect—the ongoing heightened rate of new output formation—appears to have reached its long-term level. The dynamic estimation in Panel B of Figure 2 suggests that the increment of new entry over initial levels for districts along the GQ network stabilized by 2006-2009. The heightened output formation at this time is 40% less than at its 2004 peak. Similarly, the share of establishments that are young firms is stable in these districts at 17%-18% in 2007-2009, after an increase from 14% in 2000. This effect added about 11% output overall to the average district compared to initial levels.

The second driver—growth to scale of young cohorts of entrants—is often thought to be weak in the Indian context (e.g., Hsieh and Klenow 2014). In this context, however, we observe some growth due to scaling in the 0-10 km districts. Yet, there does not appear to be much scope for further growth in this regard either. The average size of all entrants since 2000 is 86% of all remaining pre-2000 incumbents by 2009; for the most recent cohort, the size ratio is 93%. Thus, while some further growth stimulus may occur through growth of recent entrants, it seems unlikely from these conditions that such increments will be very large.

The third factor—growth of the surviving incumbents—is also unlikely to be a source of further growth. By 2009, this group accounts for about 68% of output in these districts. While their output level increased, it has not kept pace with the district as a whole. This surviving group has been strongly selected during the 2000-2009 period, so we do not anticipate a mass exodus nor rapid growth for them.

Having reviewed these three components, we believe GQ's output growth may increase beyond 49%, but it is difficult to project it being much higher. Most of the long-term impact of the GQ upgrades on the 0-10 km districts appears to have been achieved. One uncertainty in this forecast is the role of entrants and reallocation. The data suggest stability is setting in on the rate of reallocations, but it is possible that reallocations could accelerate again in the future if some industries or plants face very large moving costs (e.g., recently built plants, agglomeration economies) that further weaken. Another source of uncertainty is how continued development of other infrastructure projects in India will cut into the connectivity advantage of districts along the GQ network (recall that the GQ upgrades were the first of seven phases planned).

With repeated cross-sectional data, we are not able to precisely decompose the aggregate growth gains into the three channels described above. Some rough calculations suggest a balance between entrants (the first two channels) and incumbents (the third channel). The 2009 output level of pre-2000 incumbents is 148% of the district's size in 2000 for the 0-10 km districts; as a comparison point, the similar ratio for the 50+ km districts is 138%. These percentages include initial levels, and thus the net growth among pre-2000 incumbents is 48% and 38%, respectively. The 2009 output level of post-2000 entrants is 70% of the district's size in 2000 for the 0-10 km districts, with the comparison point now being 67%. Thus, in both distance bands, entrants account for a majority of net output growth, and the increment for the 0-10 km districts is roughly comparable for both incumbents and entrants, suggestive of mostly balanced growth roles that somewhat favor entrants.

With this long-term portrait, we can also make a cautious ballpark estimate of the overall effect by considering the four spatial bands. We first make the important assumption that the GQ upgrades had no effect on the 236 districts that are 50 km beyond the GQ network. This group, which accounted for 43% of India's output in the initial period, has served as a frequent control group and does not appear to be impacted substantially by the GQ network. Some evidence for this assumption is found in the NS-EW placebo that compares distant districts from the GQ network by their proximity to the NS-EW network. Finding no major differences across sub-groups of districts in this analysis suggests that aggregate effects in the distant spatial range were unlikely to be material.

With this assumption in place, we next conclude from our analytical work that weak effects accrued to districts 10-50 km from the GQ network, which accounted for a little under 10% of initial output levels. There may have been some declines, as evidenced by the small negative coefficients in Table 2, but these results are not statistically significant. Even taking the point estimates as true values, the overall magnitude would still be small (1%-2% of total Indian manufacturing) given that the declines would be applied against a small share of the total manufacturing output in India.

The biggest growth comes from the 0-10 km districts, which accounted for 34% of initial levels. If long-term output development is modest from its current level, as argued above, we can use the 49% growth estimate. This would suggest a 17% output growth contribution overall for India's manufacturing sector.

Finally, we come to the question of the nodal districts, which represented 15% of initial output. Strictly speaking, the estimations would suggest that this group added a further 9% output expansion. Some portion of this growth may be due to GQ upgrades, but it is likely the case that a majority of the credit belongs elsewhere. For representative purposes, a value of 2% is assigned.

Thus, on a whole, our ballpark calculation would suggest that the GQ upgrades increased manufacturing output by 15%-19%. For reference, Indian manufacturing output doubled during this period, and 37% of this growth occurred in the non-nodal 0-10 km districts along the GQ network. The estimates would thus credit something less than a fifth of the organized sector growth to better connectivity provided by the enhanced GQ network, with all of that impact concentrated on adjacent districts. This calculation makes clear the key assumption underlying the analysis—the only factors pushing forward these nearby districts, where the majority of this manufacturing growth occurred, were the GQ upgrades or the other traits explicitly modeled.[26]

These powerful effects may be localized to organized manufacturing. Khanna (2014) examines changes in night-time luminosity around the GQ upgrades. He finds evidence that the upgrades yielded weaker spatial differentials in terms of night-time lights, suggestive of a spreading out of economic development. There are sufficient overlaps in our methodologies that the differences between the papers are not likely to be due to choices regarding empirical technique. Instead, it seems more likely that the organized manufacturing sector reacted differently from other forms of economic activity, and Section 3 noted the significantly weaker responses observed for unorganized manufacturing. Our findings thus point to a particular connection of the highway upgrades to the spatial development, sorting, and allocative efficiency of large-scale manufacturing plants that ship inputs and outputs at a distance. Improvements in this sector's efficiency and spatial organization can generate substantial economic gains.

These outcomes are an important input into policy choices in India and other developing economies. This paper provides quantitative estimates of the likely economic growth associated with highway upgrades and the

[26] We do not attempt a full cost-benefit analysis given our uncertainty about the full costs of the upgrades. Using the initial cost estimates of the project and government reports that the project remained within budget, it is highly likely the benefits exceeded the costs.

spatial impacts of these infrastructure projects. These estimates are most applicable to similar settings. For example, the GQ upgrades were the first phase of an overhaul to the inadequate transportation infrastructure of India. Thus, the impact may be larger than what would be observed from marginal investments into more developed settings. Similarly, the projects were undertaken during a period of economic growth, possibly allowing for greater rates of entry and more adjustment in the location choice decisions of new entrants than in economies with stagnant growth (where infrastructure investment can also be seen as a way to boost the economy). On the whole, our project speaks to the severe constraints that inadequate infrastructure can have for the development of manufacturing in emerging economies and the potential growth that may follow from alleviating that constraint.

References

Aggarwal, S. (2013). "Do Rural Roads Create Pathways out of Poverty?: Evidence from India". Working Paper.

Ahluwalia, M. (2000). "Economic Performance of States in the Post Reforms Period". *Economic and Political Weekly*, 35(19):1637-1648

Ahsan, A. and C. Pages. (2008). "Are All Labor Regulations Equal? Evidence from Indian Manufacturing". IZA Working Paper 3394.

Aschauer, D.A. (1989). "Is Public Expenditure Productive?". *Journal of Monetary Economics*, 23(2): 177-200.

Banerjee, A., E. Duflo and N. Qian. (2012). "On the Road: Access to Transportation Infrastructure and Economic. Growth in China". NBER Working Paper No. 17897.

Baum-Snow, N. (2007). "Did Highways Cause Suburbanization?". *Quarterly Journal of Economics*, 122(2): 775-805.

Baum-Snow, N., L. Brandt, V. Henderson, M. Turner and Q. Zhang. (2012). "Roads, Railroads and Decentralization of Chinese Cities". Working Paper.

Baum-Snow, N. and M. Turner. (2013). "Transportation and the Decentralization of Chinese Cities". Working Paper.

Besley, T. and R. Burgess. (2004). "Can Labor Regulation Hinder Economic Performance? Evidence from India?". *Quarterly Journal of Economics*, 119(1): 91-134.

Brown, D., M. Fay, J. Felkner, S. Lall and H. Wang. (2008). "The Death of Distance? Economic Implications of Infrastructure Improvement in Russia". *EIB Papers*, 13(2): 126-147.

Chandra, A. and E. Thompson. (2000). "Does Public Infrastructure Affect Economic Activity? Evidence from the Rural Interstate Highway System". *Regional Science and Urban Economics*, 30(4): 457-490.

Datta, S. (2011). "The Impact of Improved Highways on Indian Firms". *Journal of Development Economics*, 99(1): 46-57.

Desmet, K., E. Ghani, S. O'Connell and E. Rossi-Hansberg. (2012). "The Spatial Development of India". World Bank Policy Research Paper No. 6060.

Dinkelman, T. (2011). "The Effects of Rural Electrification on Employment: New Evidence from South Africa". *American Economic Review*, 101(7): 3078-3108.

Donaldson, D. (2014). "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure". *American Economic Review*, forthcoming.

Donaldson, D. and R. Hornbeck. (2012). "Railroads and American Economic Growth: New Data and Theory". Working Paper.

Duflo, E. and R. Pande. (2007). "Dams". *Quarterly Journal of Economics*, 122(2): 601-646.

Duranton, G., P. Morrow and M. Turner. (2013). "Roads and Trade: Evidence from the US". *Review of Economic Studies*, forthcoming.

Duranton, G. and D. Puga. (2004). "Micro-Foundations of Urban Agglomeration Economies". In V. Henderson and J.F. Thisse (eds.) *Handbook of Regional and Urban Economics*, Volume 4 (Amsterdam: North-Holland): 2063-2117.

Duranton, G. and D. Puga (2001). "Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products". *American Economic Review*, 91(5): 1454-1477.

Duranton, G. and M. Turner. (2011). "The Fundamental Law of Road Congestion: Evidence from US Cities". *American Economic Review*, 101(6): 2616-2652.

Duranton, G. and M. Turner. (2012). "Urban Growth and Transportation". *Review of Economic Studies*, 79: 1407-1440.

Ellison, G., E. Glaeser and W. Kerr. (2010). "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns", *American Economic Review*, 100(3): 1195-1213.

Faber, B. (2014). "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System". *Review of Economic Studies*, forthcoming.

Fernald, J.G. (1998). "Roads to Prosperity? Assessing the Link between Public Capital and Productivity". *American Economic Review*, 89(3): 619-638.

Fernandes, A. and A. Pakes. (2008). "Factor Utilization in Indian Manufacturing: A Look at the World Bank Investment Climate Survey Data". NBER Working Paper No. 14178.

Foster, L., J. Haltiwanger and C. Syverson. (2008). "Reallocation, Firm Turnover and Efficiency: Selection on Productivity or Profitability?", *American Economic Review*, 98: 394-425.

Fretz, S. and C. Gorgas. (2013). "Regional Economic Effects of Transport Infrastructure Expansions: Evidence from the Swiss Highway Network." Working Paper.

Ghani, E., A. Goswami and W. Kerr. (2012). "Is India's Manufacturing Sector Moving Away from Cities?". NBER Working Paper No. 17992.

Ghani, E., A. Goswami and W. Kerr. (2013). "The Golden Quadrilateral Highway Project and Urban/Rural Manufacturing in India". World Bank Working Paper No. 6620.

Ghani, E., W. Kerr and S. O'Connell. (2014). "Spatial Determinants of Entrepreneurship in India". *Regional Studies*, 48(6).

Glaeser, E., S. Kerr and W. Kerr. (2013). "Entrepreneurship and Urban Growth: An Empirical Assessment with Historical Mines". *Review of Economics and Statistics*, forthcoming.

Gill, I. and C.-C. Goh. (2012). "Scale Economies and Cities". *World Bank Research Observer*, 25(2): 235-262.

Gupta, P., R. Hasan and U. Kumar. (2008). "What Constrains Indian Manufacturing?". ICRIER Working Paper No. 211.

Gupta, P. and U. Kumar. (2010). "Performance of Indian Manufacturing in the Post Reform Period". Working Paper.

Hasan, R. and K. Jandoc. (2010). "The Distribution of Firm Size in India: What Can Survey Data Tell Us?". ADB Economics Working Paper No. 213.

Henderson, V. (2010). "Cities and Development". *Journal of Regional Science*, 50(1): 515-540.

Henderson, V., T. Lee and Y.J. Lee. (2001). "Scale Externalities in Korea". *Journal of Urban Economics*, 49(3): 479-504.

Holl, A. (2013). "Highways and Productivity in Urban and Rural Locations". Universitat de Barcelona-Institut d'Economia de Barcelona Working Paper.

Holl, A. and E. Viladecans-Marsal. (2011). "Infrastructure and Cities: The Impact of New Highways on Urban Growth". Universitat de Barcelona-Institut d'Economia de Barcelona Working Paper.

Hsieh, C. and P. Klenow. (2009). "Misallocation and Manufacturing TFP in China and India". *Quarterly Journal of Economics*, 124(4): 1403-1448.

Hsieh, C. and P. Klenow. (2014). "The Life Cycle of Plants in India and Mexico". *Quarterly Journal of Economics*, forthcoming.

Hsu, W.-T. and H. Zhang. (2011). "The Fundamental Law of Highway Congestion: Evidence from Japanese Expressways". Chinese University of Hong Kong Working Paper.

Kathuria, V., S. Natarajan, R. Raj and K. Sen. (2010). "Organized versus Unorganized Manufacturing Performance in India in the Post-Reform Period". MPRA Working Paper No. 20317.

Khanna, G. (2014). "The Road Oft Taken: Highways to Spatial Development". Unpublished mimeo.

Kochhar, K., U. Kumar, R. Rajan, A. Subramanian and I. Tokatlidis. (2006). "India's Pattern of Development: What Happened, What Follows?". IMF Working Paper No. 06/22.

Lahr, M., R. Duran and A. Varughese. (2005). "Estimating the Impact of Highways on Average Travel Velocities and Market Size". Unpublished mimeo.

Levinsohn, J. and A. Petrin. (2003). "Estimating Production Functions Using Inputs to Control for Unobservables". *Review of Economic Studies*, 70(2): 317-342.

McKinsey Global Institute. (2010). "India's Urban Awakening: Building Inclusive Cities, Sustaining Economic Growth". McKinsey & Company Report.

McKinsey Global Institute. (2012). "The Shifting Urban Economic Landscape: What Does it Mean for Cities?". McKinsey & Company Report.

Michaels, G. (2008). "The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System". *Review of Economics and Statistics*, 90(4): 683-701.

Mitra, A., A. Varoudakis and M. Véganzonès. (1998). "State Infrastructure and Productive Performance in Indian Manufacturing". OECD Working Paper No. 139.

Munell, A. (1990). "Why has Productivity Growth Declined? Productivity and Public Investment". *New England Economic Review*, January/February: 3-22.

Nataraj, S. (2011). "The Impact of Trade Liberalization on Productivity: Evidence from India's Formal and Informal Manufacturing Sectors". *Journal of International Economics*, 85(2): 292-301.

Olley, S. and A. Pakes. (1996). "The Dynamics of Productivity in the Telecommunications Equipment Industry". *Econometrica*, 64(6): 1263-1297.

Otto, G. and G. Voss. (1994). "Public Capital and Private Sector Productivity". *The Economic Record*, 70(209): 121-132.

Qin, Y. (2014). "No County Left Behind? The Distributional Impact of High-Speed Rail Upgrade in China". Unpublished mimeo.

Roberts, M., U. Deichmann, B. Fingleton and T. Shi. (2012). "Evaluating China's Road to Prosperity: A New Economic Geography Approach". *Regional Science and Urban Economics*, 42(4): 580-594.

Rosenthal, S. and W. Strange. (2004). "Evidence on the Nature and Sources of Agglomeration Economies". In V. Henderson and J. F. Thisse (eds.) *Handbook of Regional and Urban Economics*, Volume 4 (Amsterdam: North-Holland): 2119-2171.

Sivadasan, J. (2009). "Barriers to Competition and Productivity: Evidence from India". *The BE Journal of Economic Analysis & Policy*, 9(1): Article 42.

Subramanian, A. (2012a). "The Ideas India Must Shed". June 4th, 2012.

Subramanian, A. (2012b). "What is India's Real Growth Potential". May 23, 2012.

The Economic Times. (2012). "Highway Development Requires Rs 200 cr Investment Every Day". April 29, 2012.

Ulimwengu, J., J. Funes, D. Headey and L. You. (2009). "Paving the Way for Development? The Impact of Transport Infrastructure on Agricultural Production and Poverty Reduction in the Democratic Republic of Congo". IFPRI Discussion Paper No. 00944.

World Bank. (2012). "Planning, Connecting, and Financing Cities—Now". Urbanization Review Flagship Report, The World Bank, Washington DC.

World Development Report. (2009). Reshaping Economic Geography, The World Bank, Washington DC.

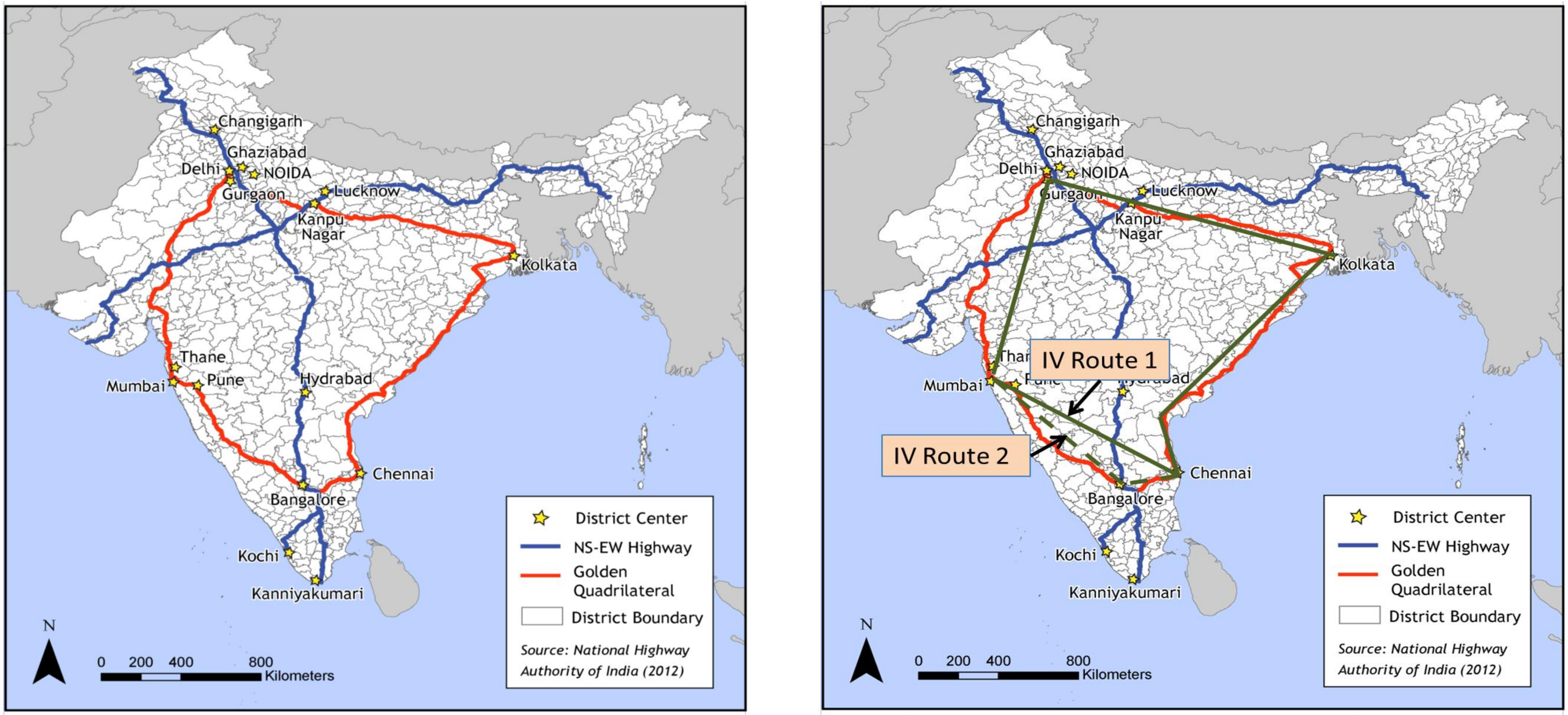
World Road Statistics. (2009). World Road Statistics 2009: Data 2002-2007. International Road Federation, Geneva.

Xu, H. and K. Nakajima. (2013). "Highways and Development in the Peripheral Regions of China". Working Paper.

Figure 1: Map of the Golden Quadrangle and North-South East-West Highway systems in India

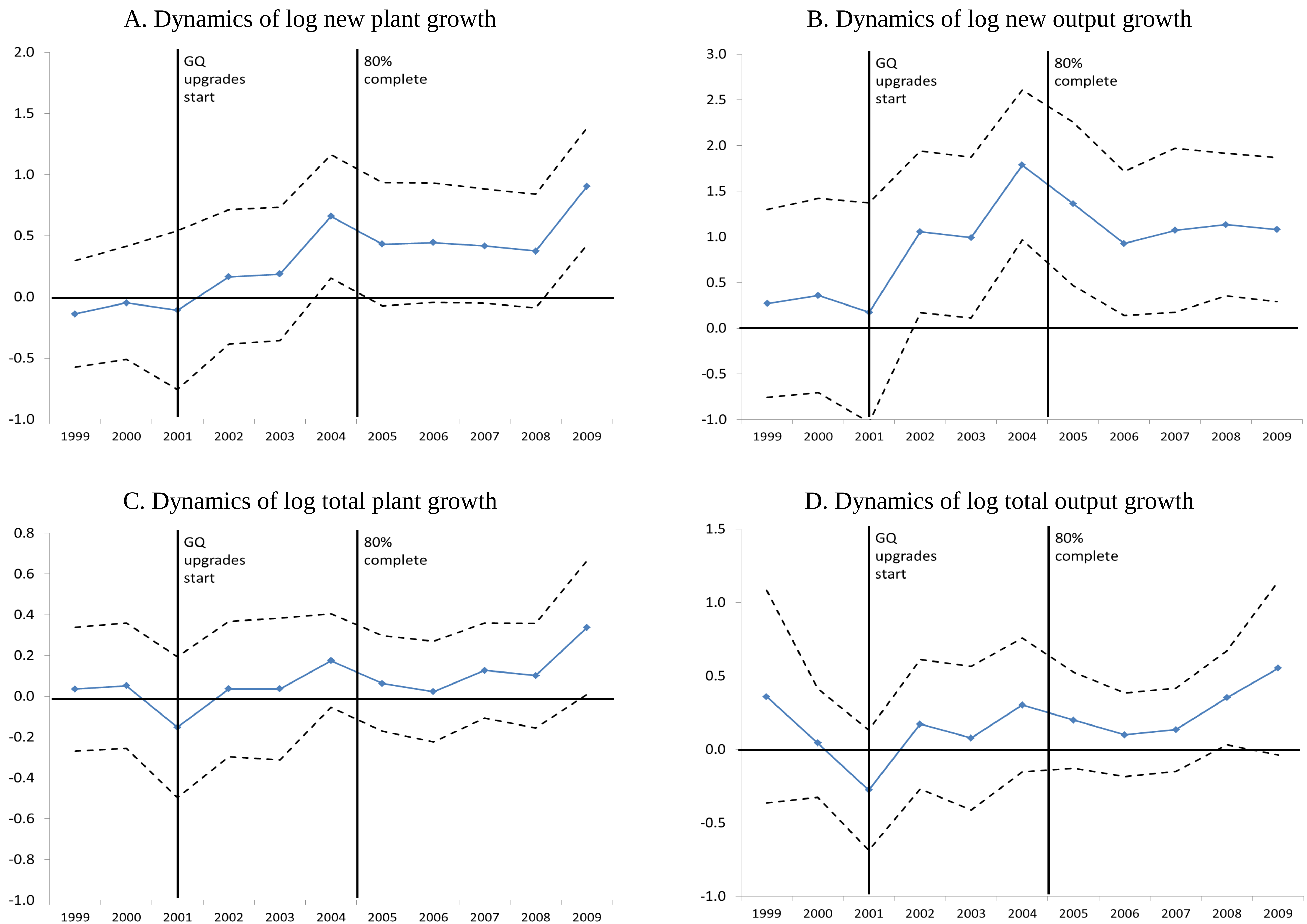
A. Highway route structure

B. Overlay of straight-line IV strategy



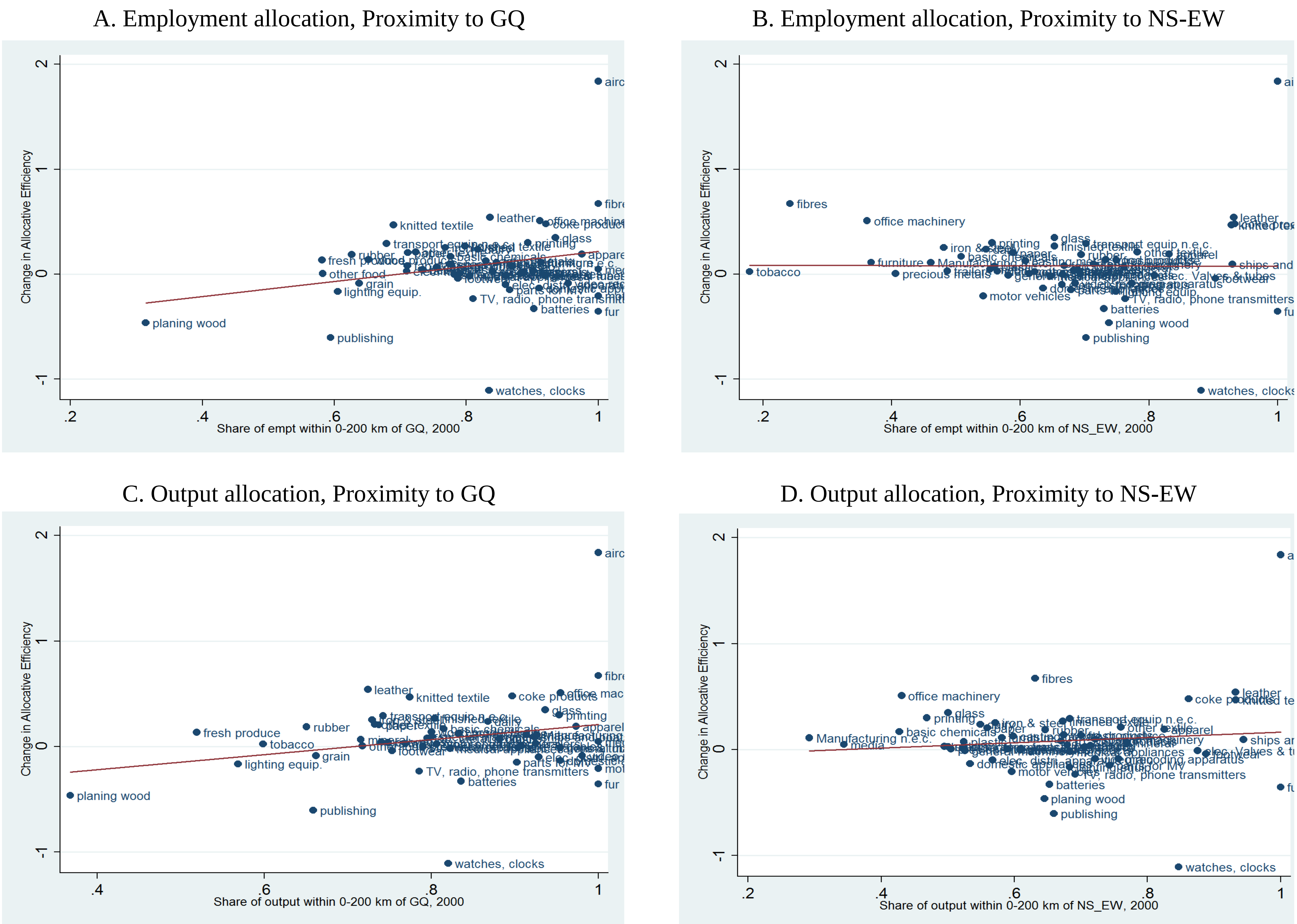
Notes: Panel A plots the Golden Quadrangle and North-South East-West Highway systems. Panel B plots the instrumental variables route formed through the straight-line connection of the GQ network's nodal cities: Delhi, Mumbai, Kolkata, and Chennai. IV Route 2 also considers Bangalore as a fifth nodal city.

Figure 2: Dynamics of plant count and output growth around the GQ upgrades



Notes: Panels A and C illustrate the dynamics of young entrant and total plant count growth for non-nodal districts located 0-10 km from the GQ network relative to districts 10-50 km from the GQ network. The solid line quantifies the differential effect for the GQ upgrades by year, with 1994 as the reference year. Dashed lines present 90% confidence intervals, with standard errors clustered by district. Panels B and D consider comparable output estimations. Appendix Table 6a reports complete regression results.

Figure 3: Change in allocative efficiency for Indian organized sector industries from 2000 to 2007/9



Notes: Panels A and C plot the change in allocative efficiency for 55 industries from 2000 to 2007/9 based upon the initial share of activity in those industries along the GQ network in 2000. A 200 km radius is employed, and Panel A considers employment and Panel C considers output. Panels B and D plot comparable graphs based upon the proximity of industries to the NS-EW highway system.

Table 1: Descriptive statistics

	Levels of total activity			Levels of young firm activity			Labor	Total factor
	Plants	Employment	Output	Plants	Employment	Output	productivity	productivity
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A. Average levels of activity in 1994 and 2000, combining districts within spatial range							
Total	81,884	5,915,323	4.0E+11	12,035	556,463	4.5E+10	67,109	n.a.
Nodal district for GQ	11,416	729,312	5.9E+10	1,404	72,022	5.3E+09	80,420	0.158
District 0-10 km from GQ	24,897	2,109,045	1.3E+11	3,999	193,342	1.5E+10	63,230	-0.132
District 10-50 km from GQ	6,017	377,902	3.4E+10	1,058	43,959	5.8E+09	90,336	-0.081
District over 50 km from GQ	39,554	2,699,064	1.7E+11	5,573	247,140	1.9E+10	63,291	-0.082
	B. Average levels of activity in 2005, 2007 and 2009 combining districts within spatial range							
Total	95,678	7,621,581	8.1E+11	14,986	1,008,038	1.1E+11	106,385	n.a.
Nodal district for GQ	12,921	991,419	1.2E+11	1,989	145,347	1.6E+10	120,522	0.167
District 0-10 km from GQ	31,492	2,635,072	2.9E+11	5,184	348,214	4.0E+10	108,331	-0.099
District 10-50 km from GQ	7,019	475,986	6.7E+10	1,069	57,066	6.2E+09	141,099	-0.055
District over 50 km from GQ	44,246	3,519,104	3.4E+11	6,744	457,411	5.2E+10	96,249	-0.129
	C. Ratio of activity in 2005/2007/2009 to 1994/2000 (Change for TFP)							
Total	1.168	1.288	2.043	1.245	1.812	2.541	1.585	n.a.
Nodal district for GQ	1.132	1.359	2.037	1.416	2.018	2.921	1.499	0.009
District 0-10 km from GQ	1.265	1.249	2.141	1.296	1.801	2.712	1.713	0.033
District 10-50 km from GQ	1.166	1.260	1.967	1.010	1.298	1.072	1.562	0.026
District over 50 km from GQ	1.119	1.304	1.983	1.210	1.851	2.750	1.521	-0.048
	D. Change in share of activity between 2005/2007/2009 and 1994/2000							
Nodal district for GQ	-0.004	0.007	0.000	0.016	0.015	0.018	n.a.	n.a.
District 0-10 km from GQ	0.025	-0.011	0.016	0.014	-0.002	0.022		
District 10-50 km from GQ	0.000	-0.001	-0.003	-0.017	-0.022	-0.075		
District over 50 km from GQ	-0.021	0.005	-0.013	-0.013	0.010	0.035		

Notes: Descriptive statistics calculated from Annual Survey of Industries (ASI). There are 363 included districts with the following allocation: 9 are nodal, 76 are 0-10 km away, 42 are 10-50 km away, and 236 are over 50 km away. Districts are local administrative units that generally form the tier of local government immediately below that of India's subnational states and territories. These are the smallest entities for which data is available with ASI. Nodal districts include Delhi, Mumbai, Kolkata, and Chennai and their contiguous suburbs (Gurgaon, Faridabad, Ghaziabad, and NOIDA for Delhi; Thane for Mumbai). Distance is calculated taking the minimum straight line from the GQ network to the district edge. Labor productivity is total output per employee. Appendix Table 1 reports comparable descriptive statistics for the NS-EW highway system.

Table 2: Long-differenced estimations of the impact of GQ improvements, comparing 2007-2009 to 2000

DV: Change in manufacturing trait listed in column header	Log levels of total activity			Log levels of young firm activity			Log labor productivity	Total factor productivity	Log average wage	Log cost per employee
	Plants	Employment	Output	Plants	Employment	Output				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A. Base spatial horizon measuring effects relative to districts 50+ km from the GQ network										
(0,1) Nodal district	1.467+++	1.255+++	1.413+++	1.640+++	2.004+++	2.468+++	0.138	1.971+++	0.382+++	0.393+++
	(0.496)	(0.464)	(0.480)	(0.499)	(0.543)	(0.621)	(0.111)	(0.195)	(0.065)	(0.069)
(0,1) District 0-10 km from GQ	0.364+++	0.235	0.443+++	0.815+++	0.882+++	1.069+++	0.199+++	0.163	0.121++	0.130++
	(0.128)	(0.144)	(0.163)	(0.161)	(0.198)	(0.277)	(0.074)	(0.195)	(0.055)	(0.056)
(0,1) District 10-50 km from GQ	-0.199	-0.325	-0.175	-0.238	-0.087	-0.281	0.157	0.286	0.098	0.095
	(0.185)	(0.222)	(0.293)	(0.237)	(0.314)	(0.455)	(0.126)	(0.280)	(0.091)	(0.094)
B. Panel A including covariates for initial district conditions and additional road and railroad traits										
(0,1) Nodal district	0.541	0.468	0.493	0.831	0.964	0.927	0.004	1.367+++	0.239++	0.249++
	(0.591)	(0.657)	(0.677)	(0.718)	(0.858)	(0.957)	(0.151)	(0.280)	(0.096)	(0.100)
(0,1) District 0-10 km from GQ	0.312++	0.233+	0.427+++	0.616+++	0.555+++	0.680++	0.241+++	0.112	0.169+++	0.185+++
	(0.124)	(0.129)	(0.157)	(0.174)	(0.201)	(0.286)	(0.085)	(0.215)	(0.060)	(0.062)
(0,1) District 10-50 km from GQ	-0.117	-0.202	-0.024	-0.115	-0.025	-0.194	0.177	0.403	0.151+	0.155+
	(0.161)	(0.196)	(0.271)	(0.207)	(0.279)	(0.416)	(0.127)	(0.288)	(0.087)	(0.090)
C. Panel B including state fixed effects										
(0,1) Nodal district	0.773	0.671	0.661	1.110	1.087	1.033	-0.011	1.292+++	0.256++	0.259++
	(0.643)	(0.718)	(0.728)	(0.797)	(0.963)	(1.062)	(0.157)	(0.342)	(0.114)	(0.117)
(0,1) District 0-10 km from GQ	0.334++	0.194	0.370+	0.503++	0.361	0.490	0.189+	0.235	0.160++	0.177++
	(0.147)	(0.172)	(0.211)	(0.208)	(0.246)	(0.345)	(0.113)	(0.262)	(0.073)	(0.075)
(0,1) District 10-50 km from GQ										
	-0.145	-0.275	-0.147	-0.190	-0.178	-0.382	0.113	0.424	0.123	0.126
	(0.186)	(0.237)	(0.320)	(0.224)	(0.309)	(0.463)	(0.147)	(0.324)	(0.102)	(0.106)

Notes: Long-differenced estimations consider changes in the location and productivity of organized-sector manufacturing activity in 311 Indian districts from 2000 to 2007-2009 from the Annual Survey of Industries (ASI). Explanatory variables are indicators for distance from the GQ network that was upgraded starting in 2001. Estimations consider the effects relative to districts more than 50 km from the GQ network. Column headers list dependent variables. Young plants are those less than four years old. Labor productivity is total output per employee in district, and TFP is weighted average of the Sivadasan (2009) approach to Levinsohn-Petrin estimations of establishment-level productivity with repeated cross-section data. Outcome variables are winsorized at their 1% and 99% levels, and entry variables are coded at the 1% level where no entry is observed to maintain a consistent sample. Estimations report standard errors, have 311 observations, control for the level of district activity in 2000, and weight observations by log total district population in 2001. Initial district conditions include variables for national highway access, state highway access, broad-gauge railroad access and district-level measures from 2000 Census of log total population, age profile, female-male sex ratio, population share in urban areas, population share in scheduled castes or tribes, literacy rates, and an index of within-district infrastructure. Appendix Table 2 reports the coefficients for these controls for the estimation in Panel B. +, ++, and +++ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 2: Long-differenced estimations, continued

DV: Change in manufacturing trait listed in column header	Log levels of total activity			Log levels of young firm activity			Log labor	Total factor	Log average	Log cost per
	Plants	Employment	Output	Plants	Employment	Output	productivity	productivity	wage	employee
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	D. Panel B separating new construction vs. improvements of existing roads									
(0,1) Nodal district	0.539	0.470	0.487	0.833	0.975	0.928	-0.003	1.377+++	0.243++	0.253++
	(0.594)	(0.659)	(0.681)	(0.720)	(0.860)	(0.961)	(0.153)	(0.281)	(0.096)	(0.101)
(0,1) District 0-10 km from GQ *	0.295++	0.253	0.382++	0.636+++	0.633++	0.692++	0.194++	0.181	0.199+++	0.211+++
(0,1) New construction district	(0.129)	(0.156)	(0.171)	(0.203)	(0.258)	(0.332)	(0.083)	(0.197)	(0.065)	(0.066)
(0,1) District 0-10 km from GQ *	0.328+	0.215	0.468++	0.598+++	0.484++	0.669+	0.285++	0.046	0.140+	0.160+
(0,1) Road upgrade district	(0.179)	(0.175)	(0.236)	(0.227)	(0.238)	(0.368)	(0.121)	(0.311)	(0.084)	(0.085)
(0,1) District 10-50 km from GQ	-0.117	-0.203	-0.023	-0.115	-0.028	-0.195	0.178	0.401	0.151+	0.154+
	(0.161)	(0.196)	(0.271)	(0.208)	(0.280)	(0.417)	(0.127)	(0.289)	(0.087)	(0.090)
	E. Panel B with extended spatial horizon measuring effects relative to districts 200+ km from the GQ network									
(0,1) Nodal district	0.450	0.425	0.549	0.718	0.847	0.853	0.102	1.433+++	0.334+++	0.353+++
	(0.597)	(0.662)	(0.687)	(0.733)	(0.871)	(0.978)	(0.166)	(0.307)	(0.105)	(0.110)
(0,1) District 0-10 km from GQ	0.226	0.196	0.490++	0.509++	0.445+	0.612+	0.344+++	0.175	0.259+++	0.284+++
	(0.145)	(0.156)	(0.190)	(0.213)	(0.236)	(0.342)	(0.113)	(0.245)	(0.075)	(0.077)
(0,1) District 10-50 km from GQ	-0.208	-0.242	0.043	-0.227	-0.141	-0.265	0.283+	0.470	0.247++	0.260++
	(0.176)	(0.212)	(0.282)	(0.235)	(0.312)	(0.465)	(0.146)	(0.319)	(0.098)	(0.101)
(0,1) District 50-125 km from GQ	-0.268+	-0.165	-0.043	-0.301	-0.355	-0.292	0.143	0.151	0.233++	0.252++
	(0.150)	(0.173)	(0.242)	(0.221)	(0.265)	(0.391)	(0.167)	(0.322)	(0.097)	(0.099)
(0,1) District 125-200 km from GQ	-0.068	0.018	0.286	-0.115	-0.072	0.032	0.247+	0.095	0.114	0.131
	(0.159)	(0.191)	(0.219)	(0.245)	(0.331)	(0.454)	(0.143)	(0.323)	(0.091)	(0.094)

Notes: See notes above. Panel D splits local effects along the GQ network by whether the development is new highway construction or the improvement of existing highways. Panel E includes extended spatial rings to measure effects relative to districts 200 km away from the GQ network.

Table 3: Long-differenced estimations comparing the impact of GQ improvements to districts along the NS-EW network

DV: Change in manufacturing trait listed in column header	Log levels of total activity			Log levels of young firm activity			Log labor productivity	Total factor productivity	Log average wage	Log cost per employee
	Plants	Employment	Output	Plants	Employment	Output				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Effects for districts based upon distance from the GQ network:										
(0,1) Nodal district	0.377	0.188	0.208	0.581	0.688	0.720	0.036	1.147+++	0.237++	0.253+++
	(0.513)	(0.565)	(0.584)	(0.699)	(0.878)	(0.997)	(0.147)	(0.306)	(0.094)	(0.095)
(0,1) District 0-10 km from GQ	0.338+++	0.259+	0.457+++	0.626+++	0.548++	0.663++	0.248+++	0.109	0.192+++	0.209+++
	(0.127)	(0.135)	(0.168)	(0.186)	(0.221)	(0.312)	(0.093)	(0.234)	(0.064)	(0.066)
(0,1) District 10-50 km from GQ	-0.085	-0.161	0.025	-0.098	-0.014	-0.202	0.185	0.410	0.169+	0.173+
	(0.158)	(0.193)	(0.265)	(0.210)	(0.285)	(0.425)	(0.128)	(0.287)	(0.087)	(0.090)
Effects for districts based upon distance from the NS-EW network:										
(0,1) Nodal district for NS-EW	0.456	0.807	0.840	0.649	0.676	0.559	-0.058	0.403	0.110	0.097
	(0.521)	(0.575)	(0.600)	(0.713)	(0.914)	(1.001)	(0.136)	(0.249)	(0.087)	(0.086)
(0,1) District 0-10 km from NS-EW	0.059	0.193	0.226	0.089	0.109	0.198	0.017	-0.142	0.105	0.101
section scheduled for Phase I	(0.158)	(0.156)	(0.189)	(0.224)	(0.248)	(0.325)	(0.120)	(0.283)	(0.076)	(0.079)
(0,1) District 0-10 km from NS-EW	0.232	0.283	0.367	0.062	0.081	-0.136	0.094	0.046	0.115	0.110
section scheduled for Phase II	(0.142)	(0.184)	(0.236)	(0.239)	(0.303)	(0.424)	(0.155)	(0.331)	(0.103)	(0.106)
(0,1) District 10-50 km from	0.073	-0.026	-0.084	0.056	-0.162	-0.206	-0.034	0.120	0.053	0.062
NS-EW	(0.167)	(0.173)	(0.230)	(0.238)	(0.282)	(0.390)	(0.129)	(0.284)	(0.086)	(0.089)

Notes: See Table 2. Long-differenced estimations compare results from proximity to the GQ network to the NS-EW highway network that was planned for partial upgrade at the same time as the GQ project but was then delayed. Phase I portions of the NS-EW upgrade were planned to overlap with the GQ upgrades but were postponed. The regressions control for the initial district conditions listed in Table 2.

Table 4a: Instrumental variable estimations using distance from a straight line between nodal districts

DV: Change in manufacturing trait listed in column header	Log levels of total activity			Log levels of young firm activity			Log labor productivity	Total factor productivity	Log average wage	Log cost per employee
	Plants	Employment	Output	Plants	Employment	Output				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A. Base OLS estimation that excludes nodal districts and measures effects relative to districts 10+ km from the GQ network										
(0,1) District 0-10 km from GQ	0.362+++	0.264+	0.458+++	0.840+++	0.881+++	1.100+++	0.174++	0.116	0.104+	0.115++
	(0.122)	(0.139)	(0.158)	(0.156)	(0.191)	(0.270)	(0.070)	(0.186)	(0.053)	(0.054)
B. Reduced-form estimates for distance from a straight-line between nodal districts										
(0,1) District 0-10 km from line	0.168	-0.015	0.256	0.406++	0.310	0.358	0.253+++	0.132	0.146++	0.162+++
ROUTE 1	(0.122)	(0.136)	(0.168)	(0.176)	(0.218)	(0.310)	(0.085)	(0.210)	(0.061)	(0.062)
(0,1) District 0-10 km from line	0.195	0.056	0.315+	0.450++	0.418+	0.448	0.220+++	0.319	0.175+++	0.186+++
ROUTE 2	(0.123)	(0.139)	(0.170)	(0.179)	(0.221)	(0.312)	(0.085)	(0.199)	(0.059)	(0.060)
C. IV estimates using distance from a straight-line between nodal districts										
(0,1) District 0-10 km from line	0.343	-0.030	0.513	0.818++	0.622	0.713	0.490+++	0.256	0.282++	0.313++
ROUTE 1	(0.236)	(0.280)	(0.322)	(0.323)	(0.408)	(0.585)	(0.172)	(0.405)	(0.122)	(0.125)
Exogeneity test p-value	0.928	0.207	0.853	0.947	0.498	0.487	0.039	0.714	0.083	0.058
(0,1) District 0-10 km from line	0.320+	0.092	0.509+	0.726+++	0.675++	0.717	0.348++	0.503	0.276+++	0.294+++
ROUTE 2	(0.193)	(0.226)	(0.266)	(0.259)	(0.330)	(0.471)	(0.136)	(0.316)	(0.098)	(0.100)
Exogeneity test p-value	0.791	0.336	0.824	0.644	0.483	0.371	0.151	0.161	0.028	0.024

Notes: See Table 2. Panel A modifies the base OLS estimation to exclude nodal districts and measure effects relative to districts 10+ km from the GQ network. This sample contains 302 districts. Panel B reports reduced-form estimations of whether or not a district edge is within 10 km of a straight line between nodal districts. Panel C reports IV estimations that instrument being within 10 km from the GQ network with being within 10 km of the straight line between nodal districts. Route 1 does not connect Bangalore directly, with the first-stage elasticity of 0.43 (0.05) and the associated F-statistic of 74.5. Route 2 treats Bangalore as a connection point, with the first-stage elasticity of 0.54 (0.05) and the associated F-statistic of 138.1. The null hypothesis in the exogeneity tests is that the instrumented regressor is exogenous.

Table 4b: Table 4a including district controls

DV: Change in manufacturing trait listed in column header	Log levels of total activity			Log levels of young firm activity			Log labor productivity	Total factor productivity	Log average wage	Log cost per employee
	Plants	Employment	Output	Plants	Employment	Output				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A. Base OLS estimation that excludes nodal districts and measures effects relative to districts 10+ km from the GQ network										
(0,1) District 0-10 km from GQ	0.319+++	0.246++	0.381++	0.628+++	0.541+++	0.663++	0.186++	0.030	0.120++	0.136++
	(0.117)	(0.123)	(0.150)	(0.172)	(0.197)	(0.279)	(0.076)	(0.203)	(0.057)	(0.058)
B. Reduced-form estimates for distance from a straight-line between nodal districts										
(0,1) District 0-10 km from line ROUTE 1	0.165	0.016	0.275+	0.298+	0.106	0.125	0.299+++	0.183	0.185+++	0.204+++
	(0.112)	(0.114)	(0.155)	(0.165)	(0.201)	(0.291)	(0.096)	(0.212)	(0.065)	(0.066)
(0,1) District 0-10 km from line ROUTE 2	0.153	0.046	0.264	0.276	0.101	0.051	0.250+++	0.327	0.209+++	0.225+++
	(0.116)	(0.118)	(0.162)	(0.175)	(0.211)	(0.300)	(0.096)	(0.204)	(0.065)	(0.067)
C. IV estimates using distance from a straight-line between nodal districts										
(0,1) District 0-10 km from line ROUTE 1	0.374	0.038	0.623+	0.667+	0.239	0.280	0.660+++	0.402	0.409+++	0.452+++
	(0.238)	(0.256)	(0.339)	(0.344)	(0.434)	(0.635)	(0.225)	(0.464)	(0.153)	(0.157)
Exogeneity test p-value	0.803	0.382	0.474	0.905	0.457	0.536	0.019	0.408	0.026	0.016
(0,1) District 0-10 km from line ROUTE 2	0.274	0.083	0.471+	0.485+	0.179	0.089	0.438++	0.571	0.368+++	0.395+++
	(0.197)	(0.208)	(0.279)	(0.285)	(0.360)	(0.519)	(0.171)	(0.364)	(0.122)	(0.124)
Exogeneity test p-value	0.793	0.376	0.739	0.573	0.252	0.233	0.113	0.103	0.014	0.011

Notes: See Table 4a. Estimations include district controls from Panel B of Table 2 other than road and railroad access variables. Route 1 does not connect Bangalore directly, with a first-stage elasticity of 0.38 (.05) and associated F-statistic of 13.9. Route 2 treats Bangalore as a connection point, with a first-stage elasticity of 0.49 (.05) and associated F-statistic of 20.9.

Table 5: Estimations of the impact of GQ improvements by completion date, districts within 50 km of GQ network

DV: Levels of manufacturing trait listed in column header	Log levels of total activity			Log levels of young firm activity			Log labor productivity	Total factor productivity	Log average wage	Log cost per employee
	Plants	Employment	Output	Plants	Employment	Output				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	A. Base estimation measuring effects relative to districts 10-50 km from the GQ network									
(0,1) Post GQ upgrades *	0.184	0.190	0.376	0.581++	0.541	1.021++	0.185	-0.016	0.015	0.049
(0,1) District 0-10 km from GQ	(0.154)	(0.182)	(0.244)	(0.243)	(0.355)	(0.509)	(0.117)	(0.147)	(0.087)	(0.096)
	B. Panel A using timing of GQ section completions									
(0,1) Post GQ upgrades *	0.209	0.295	0.414	0.689++	0.680+	1.162++	0.111	-0.114	0.036	0.042
(0,1) District 0-10 km from GQ	(0.192)	(0.215)	(0.288)	(0.277)	(0.400)	(0.585)	(0.146)	(0.166)	(0.098)	(0.103)
and completed by March 2003										
(0,1) Post GQ upgrades *	0.218	0.203	0.357	0.571+	0.549	0.916	0.153	0.051	-0.040	0.006
(0,1) District 0-10 km from GQ	(0.196)	(0.223)	(0.285)	(0.301)	(0.410)	(0.593)	(0.131)	(0.146)	(0.104)	(0.111)
and completed 2003-2006										
(0,1) Post GQ upgrades *	0.077	-0.027	0.340	0.399	0.274	0.952	0.375+++	0.039	0.076	0.141
(0,1) District 0-10 km from GQ	(0.212)	(0.232)	(0.325)	(0.380)	(0.518)	(0.700)	(0.141)	(0.279)	(0.121)	(0.133)
and completed after March 2006										

Notes: See Table 2. Estimations consider the location and productivity of organized-sector manufacturing activity in non-nodal Indian districts within 50 km of the GQ network for 1994, 2000, 2005, 2007 and 2009 from the Annual Survey of Industries. Panel A repeats the base specification in the narrower range. Estimations in Panel B separate upgrade by completion date. Estimations report standard errors clustered by district, include district and year fixed effects, have 530 observations, and weight observations by log total district population in 2001.

Table 6: Long-differenced estimations of the relative role of incumbents versus entrants in districts

	Log levels of activity in 2007/9 due to incumbents alive for at least 10 years compared to total initial district activity in 2000			Log levels of activity in 2007/9 due to firms less than 10 years old compared to total initial district activity in 2000			Share of activity in 2007/9 in district that is contained in incumbent plants at least 10 years in age		
	Plants	Employment	Output	Plants	Employment	Output	Plants	Employment	Output
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(0,1) Nodal district	0.302	0.252	0.266	0.926	0.793	0.827	0.016	-0.010	0.113
	(0.569)	(0.645)	(0.690)	(0.938)	(1.043)	(1.058)	(0.136)	(0.077)	(0.356)
(0,1) District 0-10 km from GQ	0.220+	0.216	0.451+++	0.710+++	0.633+++	0.830+++	-0.282+++	-0.136+++	-0.282
	(0.121)	(0.137)	(0.161)	(0.206)	(0.244)	(0.295)	(0.074)	(0.052)	(0.250)
(0,1) District 10-50 km from GQ	-0.130	-0.232	-0.035	-0.284	-0.366	-0.477	-0.101	-0.033	-0.151
	(0.189)	(0.228)	(0.311)	(0.354)	(0.434)	(0.472)	(0.087)	(0.051)	(0.177)

Notes: See Table 2. Estimations compare activity among incumbents and entrants in districts along the GQ network. Total activity for the district in 2000 is taken as the baseline for all estimations. Columns 1-3 compare the log levels of activity in firms at least 10 years old in 2007/9 to the 2000 baseline. Positive values indicate relatively higher survival and/or within-firm growth. Columns 4-6 compare the log levels of activity in firms less than 10 years old in 2007/9 to the 2000 baseline. Positive values indicate greater accumulated entry at the end of the sample period. Columns 7-9 consider as the outcome variable the raw share of activity among older incumbent firms. Estimations include the covariates for initial district conditions and additional road and railroad traits used in Panel B of Table 2.

Table 7: Productivity distributions among incumbents and entrants

	Average of normalized TFP metric in 2000	Average of normalized TFP metric in 2007/9	Average of normalized TFP metric in 2007/9, Plants 10+ years old	Average of normalized TFP metric in 2007/9, Plants less than 10 years
	(1)	(2)	(3)	(4)
Nodal district for GQ	1.0349	1.0274, 99%	1.0344, 100%	1.0096, 98%
District 0-10 km from GQ	0.9998	1.0011, 100%	1.0068, 101%	0.9797, 98%
District 10-50 km from GQ	1.0044	1.0038, 100%	1.0346, 103%	0.9006, 90%
District 50+ km from GQ	0.9915	0.9912, 100%	0.9982, 101%	0.9654, 97%

Notes: Normalized TFP metrics divide plant-level TFP values developed with the Sivadasan (2009) approach by their industry-year average value (weighted by employment in plant). Entries on the table are then employment-weighted averages over these normalized metrics across all plants located in the indicated districts (aggregating all districts and industries). Column 1 reports initial values in 2000. Column 2 reports averages in 2007/9 and their relative percentage ratio to 2000. Column 3 reports the value for plants at least ten years of age and their ratio to the initial value in 2000. Column 4 reports a similar statistic for subsequent entrants and their ratio to the initial value in 2000. Districts in the 10-50 km show a very strong selection effect towards incumbent plants, while districts in the 0-10 km range show more homogeneous adjustments over entrants and incumbents.

Table 8: Interactions with district traits

DV: Change in manufacturing trait listed in column header	Log levels of total activity			Log levels of young firm activity			Log labor productivity	Total factor productivity	Log average wage	Log cost per employee
	Plants	Employment	Output	Plants	Employment	Output				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A. Base OLS estimation that excludes nodal districts and measures effects relative to districts 10-50 km from the GQ network										
(0,1) District 0-10 km from GQ	0.490++	0.501++	0.583++	1.017+++	0.932+++	1.343+++	0.020	-0.128	0.016	0.030
	(0.205)	(0.234)	(0.284)	(0.265)	(0.327)	(0.480)	(0.126)	(0.292)	(0.093)	(0.096)
B. Panel A with interaction split using median of district population density										
(0,1) District 0-10 km from GQ Above median	0.756+++	0.807+++	0.825+++	1.269+++	1.311+++	1.784+++	0.010	0.122	0.084	0.102
	(0.243)	(0.275)	(0.286)	(0.312)	(0.349)	(0.492)	(0.130)	(0.309)	(0.098)	(0.100)
(0,1) District 0-10 km from GQ Median value and below	0.323	0.315	0.405	0.832+++	0.663+	0.992+	0.030	-0.382	-0.055	-0.044
	(0.229)	(0.268)	(0.359)	(0.294)	(0.373)	(0.557)	(0.141)	(0.341)	(0.104)	(0.107)
C. Panel A with interaction split using median of district literacy										
(0,1) District 0-10 km from GQ Above median	0.514++	0.776+++	0.848+++	1.092+++	1.273+++	1.875+++	0.060	0.199	0.115	0.134
	(0.232)	(0.264)	(0.299)	(0.299)	(0.334)	(0.485)	(0.127)	(0.301)	(0.098)	(0.100)
(0,1) District 0-10 km from GQ Median value and below	0.469++	0.273	0.331	0.940+++	0.602	0.783	-0.032	-0.546	-0.116	-0.108
	(0.236)	(0.269)	(0.354)	(0.289)	(0.386)	(0.564)	(0.147)	(0.354)	(0.104)	(0.109)
D. Panel A with interaction split using median of district infrastructure index										
(0,1) District 0-10 km from GQ Above median	0.503++	0.591++	0.566++	1.048+++	1.154+++	1.481+++	-0.044	0.049	0.043	0.055
	(0.238)	(0.239)	(0.266)	(0.336)	(0.375)	(0.522)	(0.123)	(0.293)	(0.097)	(0.098)
(0,1) District 0-10 km from GQ Median value and below	0.482++	0.444	0.595+	0.995+++	0.776++	1.234++	0.086	-0.305	-0.011	0.004
	(0.228)	(0.274)	(0.353)	(0.267)	(0.362)	(0.540)	(0.146)	(0.362)	(0.103)	(0.108)
E. Panel A with interaction split using median of district distance from nodal city										
(0,1) District 0-10 km from GQ Above median	0.474++	0.281	0.546+	1.039+++	0.769+	1.095+	0.059	-0.201	-0.002	0.010
	(0.228)	(0.251)	(0.323)	(0.310)	(0.390)	(0.596)	(0.151)	(0.352)	(0.112)	(0.117)
(0,1) District 0-10 km from GQ Median value and below	0.499++	0.623++	0.604+	1.006+++	1.023+++	1.478+++	0.004	-0.095	0.024	0.039
	(0.226)	(0.258)	(0.312)	(0.286)	(0.349)	(0.504)	(0.127)	(0.315)	(0.098)	(0.100)

Notes: See Table 2. Long-differenced estimations consider changes in the location and productivity of organized-sector manufacturing activity for the time period starting from 2000 to 2007-2009 in 106 non-nodal districts located within 50 km of GQ. Panel A repeats the base estimation for this group. In Panels B-E, the base effect is interacted with indicator variables for above or below median values for indicated district traits. Estimations control for unreported main effects of district traits.

Table 9a: Interactions with industry land/building intensity

DV: Change in manufacturing trait listed in column header	Log new establishment counts by industry land/building intensity			Log new employment levels by industry land/building intensity			Log new output levels by industry land/building intensity		
	0-25th	25th-50th	>50th	0-25th	25th-50th	>50th	0-25th	25th-50th	>50th
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(0,1) Nodal district	1.937+++	1.766+++	1.226++	3.077+++	2.510+++	1.431++	3.457+++	2.642+++	2.238+++
	(0.477)	(0.354)	(0.527)	(0.631)	(0.473)	(0.596)	(0.779)	(0.586)	(0.777)
(0,1) District 0-10 km from GQ	0.425++	0.769+++	0.794+++	0.802+++	0.974+++	0.907+++	0.859++	1.162+++	1.473+++
	(0.165)	(0.150)	(0.190)	(0.298)	(0.222)	(0.248)	(0.379)	(0.294)	(0.339)
(0,1) District 10-50 km from GQ	-0.144	-0.187	-0.186	0.056	-0.093	-0.185	-0.011	-0.181	-0.118
	(0.164)	(0.221)	(0.213)	(0.312)	(0.324)	(0.288)	(0.412)	(0.431)	(0.424)

Notes: See Table 2. Long-differenced estimations consider entry rates grouping industries by their land and building intensity in 2000 at the national level. These three bins include those with low land intensity (the bottom quartile of intensity), medium intensity (the second quartile), and high intensity (the top two quartiles).

Table 9b: Table 9a with district controls and state fixed effects

DV: Change in manufacturing trait listed in column header	Log new establishment counts by industry land/building intensity			Log new employment levels by industry land/building intensity			Log new output levels by industry land/building intensity		
	0-25th	25th-50th	>50th	0-25th	25th-50th	>50th	0-25th	25th-50th	>50th
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(0,1) Nodal district	1.183+	1.347++	0.534	1.876++	1.624+	0.387	2.149+	1.350	0.728
	(0.623)	(0.642)	(0.817)	(0.888)	(0.930)	(1.028)	(1.116)	(1.066)	(1.286)
(0,1) District 0-10 km from GQ	0.219	0.372++	0.448++	0.560	0.301	0.348	0.592	0.333	0.861++
	(0.185)	(0.186)	(0.214)	(0.359)	(0.282)	(0.289)	(0.447)	(0.378)	(0.414)
(0,1) District 10-50 km from GQ	-0.104	-0.204	-0.072	0.217	-0.244	-0.157	0.191	-0.434	-0.046
	(0.162)	(0.192)	(0.201)	(0.320)	(0.312)	(0.287)	(0.417)	(0.434)	(0.442)

Notes: See Table 9a.

Table 10: Estimations of the impact of GQ improvements on allocative efficiency

	Change in allocative efficiency 2000-2007/9, Sivadasan, L-P	Change in allocative efficiency 2000-2007/9, Residual	Change in allocative efficiency 1994-2000, Sivadasan, L-P
	(1)	(2)	(3)
	A. Share of industry activity within 200 km of GQ, initial period		
Total plants	0.510	0.612	-0.295
	(0.467)	(0.450)	(0.214)
Total employment	0.718	0.796+	-0.233
	(0.436)	(0.419)	(0.210)
Total output	0.710+	0.810++	-0.242
	(0.415)	(0.398)	(0.293)
Young plants	0.319	0.380	-0.263
	(0.351)	(0.344)	(0.172)
Young employment	0.675++	0.792++	-0.255++
	(0.325)	(0.313)	(0.125)
Young output	0.579++	0.668+++	-0.275+
	(0.255)	(0.241)	(0.149)
	B. Share of industry activity within 200 km of NS-EW, initial period		
Total plants	0.352	0.471	0.059
	(0.676)	(0.654)	(0.200)
Total employment	-0.012	0.087	0.145
	(0.468)	(0.462)	(0.197)
Total output	0.248	0.270	0.001
	(0.519)	(0.502)	(0.210)
Young plants	-0.042	-0.067	-0.063
	(0.391)	(0.384)	(0.231)
Young employment	0.068	0.040	0.106
	(0.282)	(0.284)	(0.178)
Young output	0.037	-0.013	0.126
	(0.239)	(0.238)	(0.150)

Notes: Each table entry is from a separate estimation. Estimations in Columns 1 and 2 of Panel A consider the change in the allocative efficiency of organized-sector manufacturing from 2000 to 2007/9 by initial proximity to the GQ network in 2000. Column 3 presents a similar exercise from 1994 to 2000 during the pre-period before construction began. Panel B considers the placebo case of proximity to NS-EW system. Allocative efficiency is calculated using TFP estimates described in the column header. Regressors are expressed in shares. Estimations report robust standard errors, have 55 observations, and are unweighted.