Which Firms Drive the Gains from Connectivity and Competition? The Impact of India's Golden Quadrilateral across the Firm Life Cycle

Arti Grover* and William Maloney†

July 2025

Abstract: This paper uses the construction of India's Golden Quadrilateral (GQ) highway to explore the effects of an exogenous increase in market access and competition across the firm life cycle. While exit rates fall for all plants, aggregate increases in output, labor productivity and total factor productivity are driven by the expansion of an average young plant. Older plants stagnate or contract, consistent with the challenges of increased competition for incumbents. The benefits of connectivity to young plants depend on access to dense markets, finance and business conditions, although older plants increase labor productivity in districts with more distorted product and capital markets. Expanding young plants correspond to those that are intensive in value chains, and dependent on external finance. However, the weaker business environment forces expansion in industries that are not contract-intensive and do not require close coordination with final producers. Analysis of plant-level panel data confirms plant capabilities as central to both the magnitude of the response and to the composition of plants driving it. Aggregate expansion among young plants is driven by high skill plants while contraction of old plants is driven by low skill plants, consistent with frontier firms being able to escape competition.

Keywords: firm dynamics, connectivity, competition, productivity, infrastructure, highways, India, manufacturing.

JEL classification: D22, D24, O12, O14, O18, R12

Acknowledgements: We thank the World Bank's Research Support Budget for funding this paper. We are grateful to Leslie A. Martin, Shanthi Nataraj and Ann Harrison for sharing their code on matching ASI cross-sectional data with the ASI establishment panel data. We are indebted to Denis Medvedev and William Kerr for their support and insightful comments, and to Stephen D. O'Connell for collaboration on earlier versions of the paper.

* Corresponding author. International Finance Corporation, World Bank. Email: agrover1@worldbank.org
† International Bank for Reconstruction and Development, World Bank. Email: wmaloney@worldbank.org

I. Motivation and Context

The enhanced market access and increased competition facilitated by greater connectivity are critical drivers of firm dynamics and overall growth, although with theoretically ambiguous effects. On the one hand, increased connectivity through US railway expansion increased productivity and growth through market access and reallocation channels (Hornbeck and Rotenberg, 2019, 2021), whereas the high costs of shipping to distant markets may drive size differences of plants between the United States and India (Hsieh and Klenow, 2009, 2014).[1] On the other hand, increased competition may cause local firms and industries to contract (Banerjee et al., 2012; Fujita and Krugman, 1999; Chandra and Thompson, 2000).[2] In trade models of improved transport connectivity, the overall welfare effects of enhanced market access and higher competition can be decomposed into three effects: traditional Ricardian channel and pro-competitive changes in markups (Donaldson, 2018; Donaldson and Hornbeck, 2016); improved allocative efficiency (Asturias et al., 2019, Ghani et al., 2016a); and changes in relative terms of trade (e.g., Asturias et al., 2019).

This paper contributes to this literature by examining the impact of infrastructure interventions on firm dynamics *within* treated locations, focusing on the reallocation channel as the primary mechanism driving documented aggregate effects. While several studies have analyzed the effects of roads on the spatial distribution of economic activity *between* treated and untreated locations (e.g., Baum-Snow, 2007; Baum-Snow et al., 2016a; Donaldson and Hornbeck, 2016; Datta, 2012; Duranton and Turner, 2012; Faber, 2014; Gibbons et al., 2019; Michaels, 2008), fewer have focused on firm-level outcomes (e.g., Gibbons et al., 2019; Ghani et al., 2016a). This paper builds on Ghani et al. (2016a), which leveraged the natural experiment of India's Golden Quadrilateral (GQ) highway—connecting Delhi, Mumbai, Chennai, and Kolkata—to show that the net effect of the intervention was an expansion of manufacturing along the corridor, that is, across locations. We go a step further, using rich location-coded plant-level data to construct synthetic cohorts and a true plant panel which allows us to examine the adjustment *within* treated areas. We document how plants at different stages of their life cycle—young, mature, and old—responded to improved connectivity. Specifically, we analyze which cohorts expanded or contracted, the timeline of these adjustments, and how local initial conditions shaped these patterns of reallocation.

Theory provides little guidance on the types of firms that may benefit from improved connectivity and greater competition and the channels through which such effects manifest in prices, output, or inputs. Conceptually, the effects of connectivity could differ radically across these dimensions – older firms

[1] Transport connectivity impacts productivity through improved labor market matching as well and may exceed those from accessibility improvements (De Borger et al., 2025).
[2] For a review of the impact of transport infrastructure see Melo et al. (2013) and Foster et al. (2023).

unaccustomed to competition may contract, while firms born within a more competitive context may be prepared to take advantage of new markets. Alternatively, older firms may have established access to necessary complementary factors – finance, land, or even local product markets that, in line with de Loecker et al. (2016) findings of increased markups with trade liberalization in India, may permit them to benefit from cheaper inputs, while defending their markets, thereby even strengthening their incumbent position. Cutting across all groups is the literature pioneered by Aghion et al. (2005, 2009, 2014; Cusolito, et al. 2022; Bloom et al., 2022) that stresses how the proximity to the technological frontier determines the degree to which firms may innovate to escape new competition, versus stagnating or exiting, as well as the emerging literature suggesting that firm capabilities, such as managerial or technological practices raise firm performance and resilience to shocks (Bloom et al, 2007). Neither theory nor empirics offer a consensus on how long the sorting out process should take and for the reforms to come to fruition.

We study synthetic location-age firm cohorts and use a plant-level panel dataset to detail how plants of varying ages react to greater connectivity, focusing on young (0-5 years), mature (6-24 years) and older plants (25 and above years). Different from earlier work that focused on the extensive margin of reallocation across districts, our work follows firm cohorts since the GQ expansion to explore the reallocation dynamics in the exposed districts to attribute the improved outcomes to changes *within* plants versus *between* plant age cohorts. Our research empirically documents the following results.

First, our findings suggest improved resource allocation not only between treated and untreated districts, as demonstrated in Ghani et al. (2016a, b), but also *within* treated districts, where the more dynamic cohort of younger plants expands and benefits from improved connectivity. Specifically, we highlight the margins of adjustment *within* treated districts across plant cohorts (young, mature, and old), enabling us to identify the channels driving improved aggregate outcomes. Thus, our work contributes to the literature that decomposes firm growth into within-firm, between-firm, and entry-exit dynamics (Foster et al., 2001; Cusolito and Maloney, 2018). Importantly, our analysis differs from Ghani et al. (2016a), which attributes gains across districts primarily to the entry of new plants in treated districts compared to untreated locations. In contrast, we focus on the reallocation dynamics across existing plant cohorts *within* treated districts, explicitly excluding the entry margin by analyzing only plants that existed prior to the GQ upgrades. Our analysis reveals that, while exit rates decline across all plant age groups in treated districts (i.e., those with proximity to the GQ), the economic performance of mature and old plants shows an average insignificant or negative effect. These findings are reminiscent of the firm dynamics literature which debates the relative contributions of size versus age to firm growth (Haltiwanger et al., 2011; Grover et al., 2019). In particular, our results resonate with the "neither up nor out" aspect of business dynamism observed in developing

countries, where firms may neither grow significantly with age nor exit (Hsieh and Klenow, 2014; World Bank, 2024).[3] Examining the evolution of the impact of highway upgrades over time suggests that adjustment to the GQ shock unfolds over decades rather than years. This implies that the observed effects in the shorter time horizon are likely conservative, and a longer sample period would likely reveal an even greater divergence in performance between younger and older plants.[4]

Second, for young plants, in line with standard micro theory, the responsiveness to connectivity is blunted where factor and product markets are incomplete and where the enabling business environment shows a higher degree of distortion. Critically, then, building roads is not sufficient to stimulate local industry – other elements of the business climate, such as financial institutions must be in place as well. With the exception of plant capabilities, older plants generally show little such sensitivity to initial conditions suggesting that they have established alternative ways of resolving missing markets, and their performance is perversely *better* in districts with higher distortions, possibly suggesting that, as in de Loecker et al. (2016), distortions may allow them to defend their market share while they can take advantage of cheaper inputs.

Third, the young plants that benefit most from improved road infrastructure tend to operate in sectors that are more integrated into global value chains (GVCs), rely heavily on external finance, but in activities that do not require close monitoring by the final producer, as might be associated with a kind of internal "offshoring" along the GQ. Said differently, young plants expand in activities where physical connectivity complements existing sectoral attributes, enabling firms to capitalize on the new economic opportunities created by infrastructure investments. They also benefit in industries that are less dependent on localized aspects of the business environment that cannot be directly influenced by physical connectivity. This finding, in spirit, is aligned with the differentiated impact of connectivity across industry attributes noted in Cosar and Fajgelbaum (2016) by export-orientation and in Ghani et al. (2016a) by land-intensity.

Fourth, in line with the Aghion et al. and the Bloom et al. managerial quality literature, both (district-level) pseudo panel as well as true plant-level panel estimations suggest that more capable firms can escape competition and take advantage of new opportunities. Strikingly, older plants in high literacy districts show

[3] The firm dynamics literature contrasts the "up or out" pattern observed in the United States—where startups either grow rapidly or exit, driving resource reallocation and aggregate productivity growth—with the "neither up nor out" dynamics prevalent in developing countries like Colombia (Eslava and Haltiwanger, 2019) and India (Akcigit et al., 2016), as well as some OECD countries (Adalet McGowan et al., 2017). In these contexts, barriers to entry and distortions that inhibit growth of young firms hinder productivity and job creation (Tybout, 2000; Li and Rama, 2015).

[4] In a related study on Colombia, Barzin et al. (2018) find a one-year lag in firms' output responses to road improvements, highlighting the time needed for production adjustments. Our data enables us to track responses over a longer horizon.

a quantitatively similar positive response to GQ as that of the young plants, while it is those in low literacy districts that drive the aggregate negative trend for the cohort. This finding is broadly supported by the true plant-level panel analysis: young plants above median human capital and higher capital intensity drive the aggregate growth response while those below show no response to GQ; and below median older plants drive the contraction of the cohort while above median plants show no decline. Firm capabilities appear to be central to both the magnitude of the response to increased markets and competition, and to the composition of plants driving it.

The empirical challenge of estimating the causal impacts of roads is well known: their placement may be endogenous either to promote the growth of a lagging area or to aid the expansion of a growing region. While the earlier literature did not credibly address these endogeneity concerns (see e.g. Straub, 2011), recent papers make use of new data to devise identification strategies that use as instruments historical plans (e.g., Baum-Snow, 2007, 2010; Michaels, 2008; Duranton and Turner, 2011, 2012; Duranton et al., 2014; García-López et al., 2015) or physical geography (e.g., Banerjee et al., 2012; Faber, 2014; Jedwab and Moradi, 2016), with the assumption that original plans or physical geography were unrelated to current conditions. Some papers have also dealt with endogeneity using straight line distances connecting big cities as instruments assuming that the places that were incidental beneficiaries of new links were random (e.g., Chandra and Thompson, 2000; Holl, 2004a; Melo et al., 2010; Ahlfeldt and Feddersen, 2010; Ghani et al., 2016a). Following the latter technique, along with other identification strategies that rely on comparing new segments versus upgraded sections of the GQ network, or a planned highway that remained incomplete in the sample period as a placebo our key results remain robust.

The remainder of the paper is structured as follows: Section II details the data, while Section III outlines the methodology for long-difference estimation of the impact GQ on plant outcomes for the three cohort of plants within treated districts. It presents the long-difference estimation results, tracks the evolution of impact overtime and explores heterogeneity by district and industry traits. Section IV examines the impact using the true plant panel version of the data, and Section V concludes.

a

II. Data

Government records detail the timing of completion of specific segments of the 5,846 km (3,633 mi) of road connecting the major industrial, agricultural, and cultural centers of India. Work began in 2001 and 23%, 80%, 90%, 97% and 98% was completed by the end of 2002, 2004, 2005, 2007 and 2010 respectively. The distance of economic districts to the highway is calculated as the shortest straight-line distance using

official highway maps and ArcMap GIS software. Measuring to district edge or centroid yields similar results.[5]

The information on location and timing of GQ upgrades is combined with the plant-level Annual Survey of Industries (ASI), obtained from India's Ministry of Statistics and Program Implementation, to create two new data sets that allow us to study adjustment dynamics in a more detailed fashion than was previously possible. The ASI is broadly a census of establishments with 100 workers or more, and a rotating sample of one third (or one-fifth after 2004-05) of all other formal plants, defined as those with more than 10 workers.[6] One third (or one-fifth after 2004-05) of the units in each strata of state and 4-digit industries are systematically covered in each annual survey subject to a minimum sample size of 6 units in each stratum. The design ensures that the universe of units is covered in five years.[7] Each wave, referenced by its terminal year, was sampled over a two-year period from 1999-2000 to 2008-2009 which spans the entire period of construction of the GQ allowing before and after comparisons.[8] In any given wave there are between 20,000 and 30,000 manufacturing plants, covering all states and districts.

This allows the creation of a 283 district-level pseudo panel tabulating seven outcome measures and extensive covariates for 90% of the country's plants, employment and output. The reduction from a total of 630 districts arises from either the limited district presence of organized manufacturing, or incomplete series across the 2000 and 2009 period. In addition, given that not all plants are surveyed each year, we construct synthetic age cohorts of plants by grouping them appropriately by age. To this end, plants born after the GQ upgrades began in 2000 are dropped. This makes our analysis of young plants definitionally different from Ghani et al. who consider young as new entrants post the GQ upgrades, while in our case young cohorts are the plants that were young when the highway upgrades began.

All economic outcome variables are winsorized at the bottom 1 percentile to limit outliers and unavailable values in plant size and labor productivity coming from zeros in establishments counts, employment or output levels. Nine districts are categorized as *nodal* (Delhi, Mumbai, Chennai and Kolkata, and the several contiguous suburbs Gurgaon, Faridabad, Ghaziabad and Noida for Delhi; Thane for Mumbai); 67 districts

[5] For more details on data preparation, see Ghani et al. (2016a).
[6] The lists of registered factories maintained by the Chief Inspector of Factories (CIF) in each state forms the ASI sampling frame.
[7] As per Annual Survey of Industries Manual (2008, p. 12-13). a supplementary frame is prepared annually for new units, while closed factories only adjust the sampling weights of respondent units. At the end of the cycle, the frame is updated to account for new factories, closures, and changes in census and sample schemes.
[8] Our analysis is constrained by data limitations, as the ASI stopped publishing location identifiers after our sample period. While private data sources exist, they focus on large, listed firms rather than plants, making them unsuitable for geographic analysis. Similarly, the World Bank's Enterprise Surveys (WBES), as used in Dutta (2012), offer a small, non-representative sample, limiting their utility for studying the impact of highway treatment on firm life cycles.

located within 0 – 10 kilometers away from the GQ are the "treated" ones; 32 districts are 10–50 kilometers away; and 175 "control" districts located over 50 kilometers away are assumed to be minimally affected by the GQ, but track broad movements in the economy. Long differences in outcomes resulting from nearby GQ completion can thus be compared. To study adjustment patterns across plant ages, the data are further divided into age cohorts: 0-5 years (young), 6-24 years (mature) and plants with 25 and above years of age (old). Our results are not sensitive to moderate changes in category definition.

Uniquely for this paper, we construct a geo-coded plant-level panel from two available versions of the ASI data: (i) a repeated cross-section of plant-level data with district identifiers, but no plant identifiers ("common factory ID") and (ii) with plant identifiers, but without district identifiers. Using characteristics of firms such as industry code, ownership, establishment year, months of operation, capital assets, employment, wages, income from services etc., the two data sets are combined following Martin et al. (2017) with both district and plant identifiers with a match rate greater than 98%. This allows measuring year-to-year plant-level changes in revenue-based total factor productivity (TFPR), total revenues, output, labor and skill demand, and plant exit. Annex Table A.1 tabulates the summary statistics, splitting the outcomes by age groups and distance bands for pre-GQ and post-GQ years.

III. Estimation

(i) Long difference in difference estimation:

Our first analysis compares district activity in 2000, the year prior to the start of the GQ upgrades, with district-level activity in 2009, the year when GQ was nearly and totally complete. Indexing districts with *i*, the specification takes the form:

$$\Delta Y_i = \sum_{d \in D} \beta_d * GQ\ Distance\ band_{i,d} + \gamma * X_i + Y_{2000} + \varepsilon_i \qquad (1)$$

where ΔY_i , is the change in relative economic outcome observed for a district *i* during the period 2000-09 and captures seven outcome variables: three aggregate measures at the district level- the natural log of establishment counts, employment, and output; two capturing average plant size measured as employment and output; and changes in average labor productivity, defined as output per employee and Total Factor Productivity (TFP), estimated using LP-Sivadasan methodology (see Duranton et al., 2015 for details of the procedure). The set *D* contains the three distance bands from the GQ network: nodal district, 0–10 kilometers, and 10–50 kilometers. The coefficients, β_d measure for each distance band the change in outcome, Y_i over the 2000–09 period relative to the corresponding change for the 50+ control category. The specification is estimated separately for each of the three synthetic age cohorts and their outcomes are

compared with the corresponding plant cohort when the GQ was completed in 2009. For example, the outcomes of 0-5 year-old plant cohort in 2000 are compared with 9-14 year-old cohorts in 2009. All estimations include as a control the initial level of Y_i (that is, Y_{2000}) in the district. The vector X_i contains district-level controls: national highway access, state highway access, broad-gauge railroad access and district-level measures from the 2000 Census of log total population, age profile (measured as the inverse of the dependency ratio[9], 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 measuring local access to electricity, roads, telecom, and water/sanitation facilities (see Ghani et al., 2012 for details).

Panels A-C in Table 1 present the results from estimation of specification (1) across the three age groups and for each of the seven outcome variables, reported across columns in the three panels. Columns 1-3 report district-level aggregate changes in logged output, employment and plant count; columns 4 and 5 report changes in average plant size measured as employment or output; columns 6 and 7 report average plant performance measured as labor productivity and revenue TFP. These results are robust to splitting the residual category of 50 plus kilometer distance band into finer groupings (Annex Table A.2).

Panel A shows that the arrival of the GQ had a positive and significant impact for the young cohort on output, average plant size, measured by both employment and output levels (columns 4 and 5), and on median plant performance, measured as labor productivity and TFP (columns 6 and 7). By comparison, for mature and older cohorts we observe either no significant positive or adverse effects in Panels B and C.[10] The results suggest that the positive aggregate responses to the GQ observed by Ghani et al. (2016a, 2016b) were largely driven by the young cohort of plants.[11] Because of the imprecision of the estimates for the older plants, it is difficult to reject equivalence in the coefficients in most cases, with only three test rejecting at the 10% level, but the next section will present evidence that the differences are, indeed, important and likely significant over the longer term.

[9] We construct the demographic dividend measure as the ratio of working age population to non-working age population using 2001 population census counts.

[10] Effects in terms of employment for non-nodal districts within 0-10 kilometers of GQ are consistently positive and insignificant across all age classes. Thus, there is overall a minimal effect on labor demand among any age group, although these estimates are noisily measured. These results are in line with Sedláček and Sterk (2017) who estimate a general equilibrium firm dynamics model that suggests that aggregate conditions at birth, rather than post-entry choices, drive the majority of cohort-level employment variation.

[11] Notice that the results on new establishments (or activity cropping along the GQ network) in Ghani et al. (2016) are distinct from our results on the average effects being driven by the young cohort of plants because: (i) these plants are not young by the time of completion of GQ; (ii) Unlike Ghani et al. they *do not* represent new activity as our dataset drops all new plants that were born after the GQ upgrades began.

One exception to the null or negative results for mature and old cohort of plants is the large increase in plant count for mature and old plants. Since the synthetic cohort structure restricts our sample to the set of existing plants as of 2000, this necessarily implies a reduced exit rate among these firms. This is counter-intuitive given that we might expect the increased product market competition precisely to challenge weak incumbents. We return to this puzzle in section IV using plant-level data where exit decisions can be studied more directly.

Annex Tables undertake three tests of the possibly confounding effects of firm geographical selection and endogenous highway placement by exploiting new segments vs upgrades (A.3a), the unfinished NW-EW highway as a placebo (A.3b), and straight-line highway layouts as an instrument (A.3c) and confirms the findings of the previous literature that neither is importantly biasing the results.

Table 1: Long-differenced estimations of the impact of GQ improvements, comparing 2009 to 2000

DV - Change in logged values of manufacturing trait listed in column header:	**Output**	**Employment**	**Plant count**	**Avg. plant size (emp.)**	**Avg. plant size (output)**	**Labor productivity**	**TFPR**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				A. Young plants (0-5 years)			
(0,1) Nodal district	1.037	0.917	0.711	0.489	0.522	-0.049	0.284++
	(1.011)	(0.893)	(0.643)	(0.481)	(0.498)	(0.247)	(0.129)
(0,1) District 0-10 km from GQ	0.701++	0.270	0.088	0.311+	0.643++	0.409++	0.239++
	(0.290)	(0.233)	(0.202)	(0.185)	(0.272)	(0.179)	(0.120)
(0,1) District 10-50 km from GQ	0.062	-0.093	-0.046	-0.203	-0.040	0.226	0.121
	(0.469)	(0.324)	(0.230)	(0.266)	(0.397)	(0.276)	(0.120)
				B. Mature plants (6-24 years)			
(0,1) Nodal district	1.420	1.235+	0.813	0.467	0.623	-0.078	0.105
	(0.945)	(0.681)	(0.531)	(0.389)	(0.510)	(0.261)	(0.113)
(0,1) District 0-10 km from GQ	0.358	0.218	0.416++	-0.139	0.027	0.032	-0.014
	(0.333)	(0.225)	(0.176)	(0.172)	(0.253)	(0.150)	(0.088)
(0,1) District 10-50 km from GQ	-0.055	-0.108	-0.173	-0.133	0.070	0.143	-0.007
	(0.418)	(0.257)	(0.214)	(0.243)	(0.382)	(0.228)	(0.122)
				C. Old plants (25 years and above)			
(0,1) Nodal district	1.142	1.054	1.198+	-0.036	0.064	0.036	0.278++
	(0.809)	(0.787)	(0.651)	(0.364)	(0.359)	(0.166)	(0.126)
(0,1) District 0-10 km from GQ	0.335	0.197	0.342++	-0.013	0.012	0.157	0.048
	(0.273)	(0.203)	(0.172)	(0.161)	(0.207)	(0.144)	(0.082)
(0,1) District 10-50 km from GQ	-0.540	-0.553+	-0.239	-0.507++	-0.374	0.150	-0.032
	(0.415)	(0.288)	(0.260)	(0.233)	(0.314)	(0.192)	(0.099)

Notes: Long-differenced estimations consider changes in the location and productivity of organized-sector manufacturing activity in 283 Indian districts from 2000 to 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 Sivadasan (2009) approach to Olley-Pakes 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 283 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, and 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. Estimations include state fixed effects as well.

ii. Tracking the evolution of impact over time

Greater clarity both on the longer- term impact of the GQ, and differences across age cohorts emerges if we track the evolution of the coefficients with time elapsed both from the completion of the nearest highway segment (Figure 1), and the cumulative completion of the whole network (Annex Figure A.1). We use additional information arising from the actual completion dates of highway segments most relevant to each district and estimate:

$$Y_{i,t} = \sum_{tc=-3}^{tc=6} \beta_{tc} * (0,1)years\ to\ GQ\ completion_t * (0,1)GQDist_{i,d<10km} + \emptyset_i + \vartheta_t + \varepsilon_{i,t} \qquad (2)$$

Y_i is as before, β_{tc} is the estimated coefficient for the group of non-nodal districts located within 10 kilometers of GQ network with time to completion of GQ upgrades ranging from 3 years prior to six years or more after the completion of GQ segment. For example, if the outcome is observed in 2006, while the GQ upgrade for the district was completed in 2003, the regression measures the impact three years after the upgrade; in 2000, the coefficient measures the effect three years prior to the upgrade. $\emptyset_i$ is a vector of district and state fixed effects and ϑ_t is a vector of year fixed effects. Standard errors are clustered by district and we weight by district log population in 2001. The sample comprises 1,188 observations. For parsimony, we drop the lagged dependent variable which the above exercises suggest has little effect on the coefficients, and we use the 10-50 category as reference since it behaves similarly to the 50+, dropping the latter from the sample.

Figure 1 plots β_{tc} from equation (4) and reveals three striking patterns of evolution that cast additional light on the results above. We report only the young and old plants and confidence intervals for readability and because the mature follow an intermediate path between the two extremes.

First, as in Table 1, only the series of coefficient for the young is strongly statistically different from zero for output, plant size (output), labor productivity and TFPR. However, as the number of years after connectivity increases, the impact heightens dramatically positively for the young and usually negatively for the old. The exception is the number of plants, which, again, seems to suggest that there is progressively less exit for all cohorts. In all cases for the young, even if the coefficient may not become significantly different from zero, it does drift out of its previous confidence interval suggesting that the positive trend is, indeed, significant and the coefficient may become so.

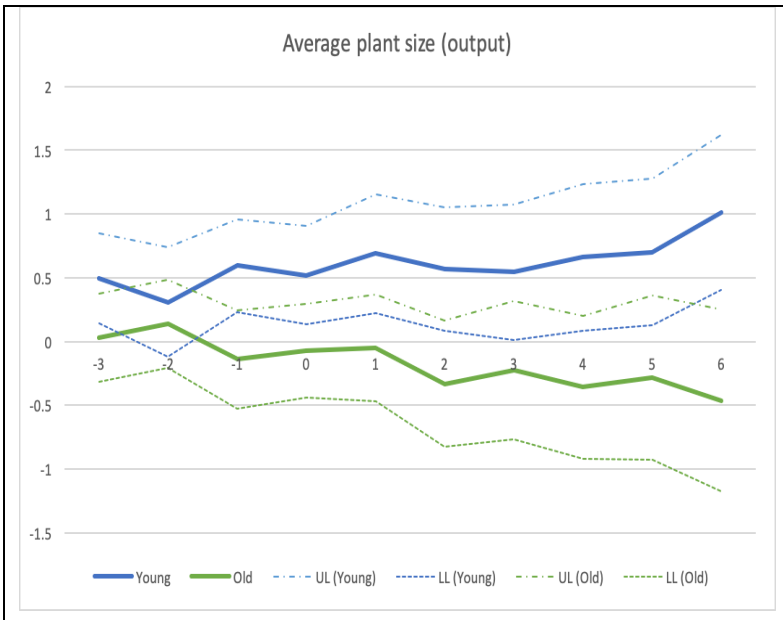
Second, this implies that, though formal tests of coefficient equivalence across age cohorts on only three occasions at the 10% level in table 2, the coefficients drift out of the own, but also the other age group

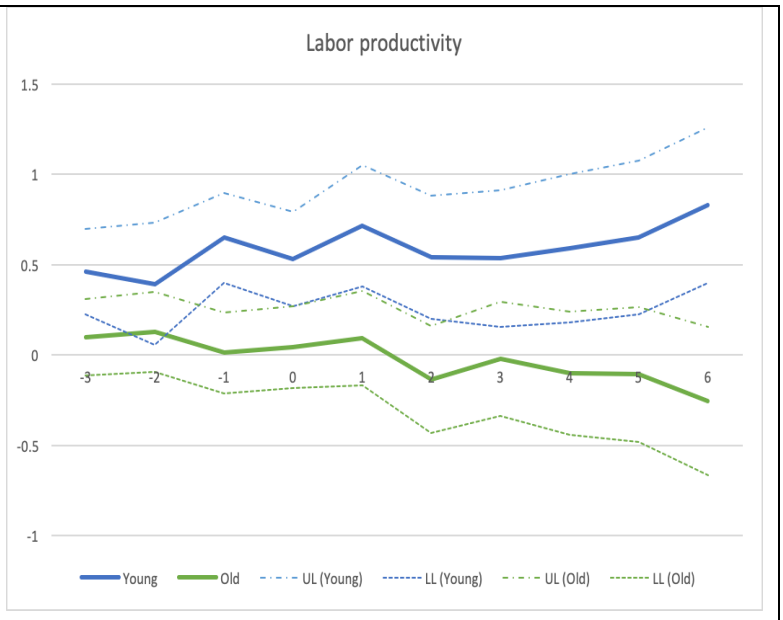
coefficient's confidence interval, strongly suggesting that eventually tests will reject equivalence over longer duration. The results for the cumulative systemic effect are broadly similar. It appears that, with time, the predominant role of young plants would become only more pronounced.

Third, this evolution shows few signs of convergence to a new value suggesting that the transition to the new steady state is a multi-decade process of which this analysis is capturing only the initial phase.

Figure 1: Impact of GQ network completion over time









iii) Initial conditions affecting the impact of GQ

Theory suggests that pre-existing complementary factors in treated districts shape the impact of infrastructure improvements. These include access to finance, access to markets, and the local business environment, which are key drivers of firm performance. For instance, districts with stronger banking networks may enable firms to leverage improved connectivity to access finance more effectively, while proximity to major cities, allows firms to expand their customer base. Similarly, distortions in input and output markets can influence how firms respond to new opportunities created by infrastructure, as areas with fewer distortions may be better positioned to capitalize on improved connectivity. Table 2 splits the

estimation of the long difference specification by initial district-level conditions derived from the 2001 census, capturing these factors, but again, disaggregated by age.

Table 2 reveals key patterns in firm responses to connectivity. Districts closer to nodal cities show stronger performance across most measures, except employment, likely due to lower transport and information costs or better access to inputs (Panel B). Access to bank finance consistently enhances performance (Panel C), with older plants showing no difference in low-finance areas, possibly due to established relationships with financial institutions or self-financing capabilities. Established relationships may also help preserve market power while reducing input costs, as suggested by de Loecker et al. (2016), potentially explaining the slower decline of incumbents. Panels D and E examine distortions in output and input markets, following Duranton et al., 2015a, 2015b, and proxies for incumbent power. Young plants respond strongly in less distorted districts, particularly in output-related measures and firm size, highlighting the importance of an enabling business environment. In contrast, old plants exhibit higher responsiveness in more distorted areas, especially in labor productivity and TFP, consistent with de Loecker et al. (2016). While some muted effects are visible for young plants, the findings suggest that young plants thrive in competitive environments with access to complementary inputs, while older plants may leverage distortions to sustain their markets.

iv. Industry characteristics correlated with responsiveness to GQ

The quality of road infrastructure is more critical for certain industries, depending on their reliance on physical connectivity. Relying again the theory of firm growth, we focus on industry attributes that are likely to be more (less) sensitive to improved connectivity, as highways can complement (substitute) the potential for their expansion. For example, plants in GVC-intensive industries can integrate into global markets when complemented with better physical connectivity, as observed in China (Cosar and Fajgelbaum, 2016). Likewise, plants in industries that rely heavily on external finance may benefit disproportionately from highways if it facilitates the flow of capital (Das et al., 2024), while those in industries that rely less on better contracting environment may do better because highways do not have a direct effect on business environment. The ASI data permits studying several industry related characteristics of the plants that could potentially be most responsive to connectivity. To this end, we replicate equation 1 splitting the regression with sub-samples partitioned by median industry characteristics. Critically, each age group is well-represented across industry characteristics (Annex Table A.4) so our results for young plants are not simply capturing newly arrived industries – the relevant characteristic is, in fact, youth.

Table 2: Long-differenced estimations of the impact of GQ improvements, comparing 2009 to 2000, interacted with initial conditions

	Young					Mature					Old				
DV - Logged values of	Output	Avg. plant size (emp.)	Avg. plant size (output)	Labor productivity	TFPR	Output	Avg. plant size (emp.)	Avg. plant size (output)	Labor productivity	TFPR	Output	Avg. plant size (emp.)	Avg. plant size (output)	Labor productivity	TFPR
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(12)	(13)	(14)	(15)	(16)
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	0.701++	0.311+	0.643++	0.409++	0.239++	0.358	-0.139	0.027	0.032	-0.014	0.335	-0.013	0.012	0.157	0.048
	(0.290)	(0.185)	(0.272)	(0.179)	(0.120)	(0.333)	(0.172)	(0.253)	(0.150)	(0.088)	(0.273)	(0.161)	(0.207)	(0.144)	(0.082)
B. Panel A with interaction split using median of district distance from nodal cities															
(0,1) District 0-10 km from GQ	0.781+	0.360	0.775++	0.543++	0.444++	0.113	-0.351	-0.107	0.032	0.044	0.291	0.063	0.096	0.235	-0.014
Above median	(0.405)	(0.262)	(0.352)	(0.241)	(0.180)	(0.436)	(0.242)	(0.360)	(0.226)	(0.154)	(0.284)	(0.203)	(0.273)	(0.189)	(0.117)
(0,1) District 0-10 km from GQ	0.661++	0.288	0.582++	0.344+	0.146	0.473	-0.044	0.087	0.032	-0.037	0.354	-0.047	-0.027	0.123	0.074
Median value and below	(0.312)	(0.201)	(0.294)	(0.183)	(0.124)	(0.379)	(0.187)	(0.266)	(0.155)	(0.093)	(0.320)	(0.185)	(0.228)	(0.148)	(0.088)
C. Panel A with interaction split using median of district access to banking services (% Households with access to Bank)															
(0,1) District 0-10 km from GQ	0.827++	0.436++	0.741++	0.434++	0.329+++	0.384	-0.018	0.084	-0.037	0.038	0.444	0.068	0.169	0.159	0.114
Above median	(0.340)	(0.218)	(0.340)	(0.216)	(0.118)	(0.315)	(0.206)	(0.260)	(0.155)	(0.097)	(0.311)	(0.181)	(0.222)	(0.151)	(0.098)
(0,1) District 0-10 km from GQ	0.596	0.208	0.561+	0.387+	0.160	0.337	-0.243	-0.022	0.099	-0.066	0.242	-0.081	-0.121	0.155	-0.012
Median value and below	(0.366)	(0.230)	(0.300)	(0.197)	(0.163)	(0.489)	(0.230)	(0.349)	(0.202)	(0.118)	(0.333)	(0.203)	(0.259)	(0.179)	(0.099)
D. Interaction split using median of Output market misallocation															
(0,1) District 0-10 km from GQ	0.428	0.007	0.483	0.581+++	0.264+	0.035	-0.373	-0.287	-0.025	-0.110	0.345	-0.204	0.025	0.358++	0.089
Above median	(0.381)	(0.236)	(0.345)	(0.221)	(0.155)	(0.480)	(0.227)	(0.339)	(0.208)	(0.118)	(0.325)	(0.208)	(0.256)	(0.160)	(0.098)
(0,1) District 0-10 km from GQ	0.905+++	0.542++	0.767+++	0.288	0.221+	0.613+	0.045	0.276	0.075	0.059	0.326	0.139	0.001	-0.000	0.016
Median value and below	(0.330)	(0.209)	(0.295)	(0.187)	(0.131)	(0.335)	(0.187)	(0.255)	(0.148)	(0.092)	(0.313)	(0.177)	(0.226)	(0.158)	(0.095)
E. Interaction split using median of Input market misallocation															
(0,1) District 0-10 km from GQ	0.522	0.114	0.482	0.480++	0.257++	0.313	-0.118	-0.005	0.016	-0.023	0.450	-0.175	0.05	0.332++	0.123
Above median	(0.359)	(0.211)	(0.319)	(0.212)	(0.130)	(0.424)	(0.207)	(0.309)	(0.184)	(0.103)	(0.299)	(0.178)	(0.227)	(0.152)	(0.091)
(0,1) District 0-10 km from GQ	0.914++	0.567++	0.852+++	0.322	0.216	0.416	-0.168	0.068	0.053	-0.003	0.182	0.205	-0.04	-0.083	-0.053
Median value and below	(0.354)	(0.240)	(0.318)	(0.197)	(0.156)	(0.371)	(0.210)	(0.277)	(0.161)	(0.107)	(0.344)	(0.210)	(0.258)	(0.163)	(0.103)

Notes: See Table 1. Long-differenced estimations consider changes in the location and productivity of organized-sector manufacturing activity in 106 non- nodal districts located within 0-50 km of GQ for the time period starting from 2000 to 2009 from the Annual Survey of Industries (ASI). 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. Estimation controls for unreported main effects of district traits. District level access to infrastructure, literacy, access to banking, population density is extracted from 2001 census, while distance from nodal cities to district is computed from district edge using India's district level GIS maps. Lastly, initial TFP is plant-level TFP for the year 2000-01 computed following LP-Sivadasan methodology on ASI data. misallocation matrices are computed using the methodology proposed in Duranton et al. (2015a).

This said, there are significant differences in the stimulative effect of connectivity across industry characteristics within age category, as shown in Table 3. When considering GVC-intensive sectors in Panel B, that is, those with strong backward and forward value chain linkages (e.g., textiles; wearing apparel; office computing; radio, television communication; motor vehicles), we find that output, employment and plant size of younger cohorts are significantly more sensitive to physical connectivity relative to their older counterparts.

Even though access to finance is an important concern for growth of firms, younger plants with a higher degree of dependence on *traditional* finance, per the Chor and Manova (2012) measures, show more responsiveness for output, employment, and plant size (Panel C) than those less dependent. Das et al. (2023) may shed light on this paradox by noting the strong supply response of credit to supply the new firms' demand along the GQ.[14] To square this finding with those of the previous section on the increased responsiveness in areas with a higher access to banks, Das et al. document that expansion along this extensive margin – the expansion of bank branches into areas where previously there was no service – took several years after the establishment of the GQ while increased flows through existing institutions ramped up quickly. Hence financially dependent plants moving into areas with established bank infrastructure may find accommodating financial flows, while plants in areas with low bank density find problems responding or rely on non-traditional financing mechanisms.

Panel D and E correspond to the dependence of an industry on business environment attributes. Young plants in sectors that are low in contractual intensity (Panel D), that is, the proportion of a sector's intermediate inputs that require relationship-specific investments and differentiated products, respond better when connected with highways. Likewise, young plants are also more likely to be responsive to connectivity in sectors where relationship stickiness, as measured by firm to firm relationship durations, is low (panel E) particularly for output and firm size in the young plants.[15] Together, these patterns are suggestive of a more "anonymous" internal offshoring behavior, that is, young plants can become parts of value chains that do not require proximity to clients to tailor intermediate inputs and may be substituted by other providers relatively easily.

[14] Here again, however, we document that this effect is entirely through young plants. Older plants may either have secured alternate sources long ago informally or rely on retained earnings.

[15] Based on Martin et al. (2023). This is a measure of stickiness in business relationships constructed at the product-level using firm-to-firm relationship duration data. The higher degree of stickiness translates into a lower probability of switching and longer firm-to-firm trade relationships, conditional on match quality.

Table 3: Long-differenced estimations of the impact of GQ improvements, comparing 2009 to 2000, by industry traits

	Young					Mature					Old				
DV - Logged values of	**Output**	**Avg. plant size (emp.)**	**Avg. plant size (output)**	**Labor productivity**	**TFPR**	**Output**	**Avg. plant size (emp.)**	**Avg. plant size (output)**	**Labor productivity**	**TFPR**	**Output**	**Avg. plant size (emp.)**	**Avg. plant size (output)**	**Labor productivity**	**TFPR**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(12)	(13)	(14)	(15)	(16)
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	0.701++	0.311+	0.643++	0.409++	0.239++	0.358	-0.139	0.027	0.032	-0.014	0.335	-0.013	0.012	0.157	0.048
	(0.290)	(0.185)	(0.272)	(0.179)	(0.120)	(0.333)	(0.172)	(0.253)	(0.150)	(0.088)	(0.273)	(0.161)	(0.207)	(0.144)	(0.082)
B. Panel A split regressions using median of industry GVC intensity															
(0,1) District 0-10 km from GQ	0.950++	0.540++	0.539+	0.103	0.124	0.662+	0.280	0.026	-0.131	0.022	0.585	0.408+	0.134	-0.141	-0.005
Above median industry	(0.388)	(0.264)	(0.320)	(0.134)	(0.119)	(0.372)	(0.220)	(0.245)	(0.140)	(0.084)	(0.363)	(0.226)	(0.260)	(0.158)	(0.111)
(0,1) District 0-10 km from GQ	0.129	0.182	0.368	0.286	0.155	0.111	-0.219	0.087	0.154	0.026	-0.159	-0.388++	-0.191	0.154	-0.025
Median value and below	(0.342)	(0.211)	(0.319)	(0.224)	(0.151)	(0.301)	(0.185)	(0.276)	(0.171)	(0.125)	(0.295)	(0.182)	(0.246)	(0.172)	(0.102)
C. Panel A split regressions using median of industry external finance dependence															
(0,1) District 0-10 km from GQ	1.117++	0.668+++	0.745++	0.022	0.076	0.496	0.149	-0.127	-0.110	-0.038	0.529	0.327	0.186	-0.003	0.067
Above median industry	(0.440)	(0.251)	(0.348)	(0.147)	(0.140)	(0.381)	(0.249)	(0.278)	(0.161)	(0.106)	(0.384)	(0.239)	(0.287)	(0.147)	(0.105)
(0,1) District 0-10 km from GQ	0.335	0.283	0.424	0.242	0.064	0.396	-0.033	0.199	0.142	-0.008	0.032	-0.136	-0.098	-0.013	-0.017
Median value and below	(0.338)	(0.210)	(0.286)	(0.215)	(0.134)	(0.326)	(0.194)	(0.263)	(0.147)	(0.105)	(0.318)	(0.199)	(0.249)	(0.177)	(0.096)
D. Panel A split regressions using median of industry contractual intensity															
(0,1) District 0-10 km from GQ	0.452	0.279	0.475	0.193	0.114	0.671	0.149	0.425	0.022	0.034	0.419	0.207	0.317	0.089	-0.182
Above median	(0.742)	(0.381)	(0.508)	(0.328)	(0.163)	(0.627)	(0.393)	(0.500)	(0.188)	(0.137)	(0.658)	(0.399)	(0.453)	(0.208)	(0.190)
(0,1) District 0-10 km from GQ	0.708++	0.383+	0.744++	0.433++	0.315++	0.329	-0.155	0.040	0.035	-0.026	0.299	-0.003	0.002	0.164	0.062
Median value and below	(0.317)	(0.196)	(0.300)	(0.194)	(0.128)	(0.327)	(0.185)	(0.263)	(0.166)	(0.099)	(0.262)	(0.165)	(0.212)	(0.149)	(0.086)
E. Panel Asplit regressions using median of industry relationship stickiness															
(0,1) District 0-10 km from GQ	0.543	0.161	0.453	0.425++	0.287++	0.367	-0.167	0.101	0.180	-0.007	0.166	-0.253	0.008	0.237	-0.019
Above median	(0.353)	(0.198)	(0.306)	(0.208)	(0.120)	(0.333)	(0.182)	(0.267)	(0.161)	(0.102)	(0.293)	(0.163)	(0.216)	(0.150)	(0.087)
(0,1) District 0-10 km from GQ	1.015++	0.866+++	1.077+++	0.100	0.165	0.931++	0.423	0.385	-0.044	0.028	1.198++	0.578++	0.517	0.105	0.097
Median value and below	(0.469)	(0.325)	(0.395)	(0.205)	(0.145)	(0.427)	(0.304)	(0.368)	(0.146)	(0.098)	(0.571)	(0.291)	(0.414)	(0.146)	(0.104)

V. Plant Panel Analysis

Using plant panel data, this section explores two additional elements of the adjustment process to greater connectivity. *First*, we ask how much of the observed response is due to adjustments on the extensive (exit) versus the intensive margin (changes within firms). *Second*, we explore how the characteristics of incumbent plants, this time measured at the plant level, may affect their process of adjustment. Our baseline specification takes the form of difference-in-difference estimation where the *plant-level* outcome variable in district *d* at time *t,* $Y_{p,d,t}$ is estimated as:

$$Y_{p,d,t} = \sum_{d \in D} \beta_d * GQ\ Distance\ band_{i,d} * GQ\ segment\ completed_{d,t} + \Gamma_p + \Omega_t + \varepsilon_{p,d,t} \quad (3)$$

$Y_{p,d,t}$ is the outcome measure for plant *p* in district *d* in year *t*, Γ_p captures the vector of time-invariant plant specific traits (e.g. location, industry, ownership and other attributes of the plant) and Ω_t captures year-specific factors as well as secular trends. The estimate of β_1 is the focal parameter, which captures the plant-level difference-in-differences effect of the nearest GQ highway segment opening among those plants located in non-nodal districts within 10km of the GQ network, relative to plants more than 50km away (the reference group).[16]

Analysis presented in Table 2 reveals that the sample focusing more on smaller plants (sample scheme) yielded the most provocative results, while those for the larger sample (census scheme) are presented in Annex Table A.5. We summarize the latter, when germane. Because the Annual Survey of Industries is structured as a rotating panel for the majority of the plant size distribution, we limit the sample to the set of plants observed in the data in the five years prior to their nearest GQ segment being completed, thereby halving the number of observations. For changes at the intensive margin, our sample contains two observations per plant, corresponding to records observed just before and just after the opening of the nearest GQ segment. Although the resulting subsamples are not directly mapped to district-level findings, the results demonstrate remarkable consistency.

i. Extensive margin

Exit, defined as the failure to observe the same plant five years after the completion of the nearest GQ segment, falls significantly with proximity, consistent with the rising district plant count in figure 1. This is true across firm age but is particularly so among younger firms where exit falls 14% (table 1, column 1)

[16] All variables with nominal values are deflated to 2005 prices, and districts are concorded to be consistently defined across the timespan analyzed (2000 to 2009). We further inspect the plant panel data for outliers and make a small number of values missing that imply implausibly large within-plant year-to-year changes. We calculate TFP-R under the assumption of constant-returns-to-scale Cobb-Douglas production with structural parameters set as the median input cost share by three-digit industry.

compared to roughly 10% for mature and older firms. There is no significant effect in the larger firm sample (Annex Table A.5) for any age cohort, and the point estimates are a third to a half smaller, suggesting that aggregate plant count effects noted in Table 1 and Figure 1 are being largely driven by increased survival of smaller and medium size firms.

Table 4: Plant-level effects of GQ access: estimation of plant exit/closure and performance outcomes

	Sample scheme				
DV - Logged values of	**Exit[0/1]**	**Output**	**Avg. plant size (emp.)**	**Labor productivity**	**TFPR**
	(1)	(2)	(3)	(4)	(5)
			A. All		
(0,1) District 0-10 km from GQ	-0.114++	0.123+++	0.016	0.113++	0.015
	(0.054)	(0.040)	(0.021)	(0.045)	(0.020)
N	18321	10213	10211	10172	9853
R^2	0.022	0.947	0.926	0.912	0.907
			B. Young		
(0,1) District 0-10 km from GQ	-0.140++	0.145++	0.071	0.050	0.013
	(0.057)	(0.063)	(0.046)	(0.061)	(0.033)
N	6255	3486	3486	3479	3373
R^2	0.026	0.944	0.905	0.910	0.912
			C. Mature		
(0,1) District 0-10 km from GQ	-0.100+	0.111+	-0.014	0.145++	0.017
	(0.056)	(0.058)	(0.025)	(0.057)	(0.029)
N	9459	5245	5243	5222	5051
R^2	0.019	0.950	0.936	0.911	0.909
			D. Old		
(0,1) District 0-10 km from GQ	-0.106++	0.114+	-0.019	0.164++	0.015
	(0.050)	(0.061)	(0.044)	(0.064)	(0.052)
N	2607	1482	1482	1471	1429
R^2	0.024	0.939	0.937	0.915	0.878

Note: The table limits the sample to plants from the sample scheme observed in the data in the five before and after the nearest GQ segment being completed. "Exit" is an indicator variable measured as the failure to observe the same plant in the post-GQ period. Table reports OLS coefficient estimates of the effect on plant outcomes due to proximity to the GQ segment. All specifications include an unreported constant term, regressors for completion interacted with an indicator for nodal district and district 10 - 50 km from the nearest GQ segment, and plant and year fixed effects. Sample is restricted to comprise the same set of districts available in the repeated cross-section analysis above. Standard errors are clustered by district. Significance levels are indicated by + < :1, ++ < :05, + + + < :01.

ii. Intensive margin:

Also consistent with the district-level findings, average output rises significantly across all plant sizes with proximity to the GQ, again with a larger impact on young plants (14.5%) relative to the mature or older ones (11%). The impacts on employment are largely insignificant although the point estimates, again, suggest a similar gradient with young firms growing 7% and old firms shrinking -1.9%. This combination

is consistent with the observed significant increase of labor productivity by 11.3% in the aggregate and increasing with age: young plants show an insignificant 5% while mature plants show a significant 14.5% and older plants increase labor productivity by 16.4%.[17] Older plants may be growing or gaining higher rents à la de Loecker et al., while shedding some labor, relative to younger plants that hire as they expand.

iii. ***The role of initial plant capabilities/characteristics***

Table 3 splits the regressions presented in Table 2 by plant position above or below median values of four pre-GQ levels of plant capabilities: skill intensity defined as the average plant wage per employee; capital intensity measured as the ratio of the value of tangible assets to total revenue; labor productivity; and TFPR.

The impact of proximity on exit follows the same decreasing gradient with age found previously although surprisingly, there appears little relationship with initial plant capabilities. Except for capital intensity for young plants, in general more able plants do not seem to survive better with proximity than those less so.

However, there are important differential effects on output and plant size that are consistent with findings in the aggregate data, that again suggest two types of adjustments for young and old plants. These adjustments indicate that exit may be a longer-term process as noted in Figure 1. With GQ proximity, young plants with above median skill intensity show a 10% increase for output compared to those below which show a negative but insignificant response. Older plants show the same pattern but where skill now mitigates contraction: output in below median skills plants fell by 11.8% and those with above median skills showing no significant effects.

A similar pattern emerges with capital intensity which is consistent with the industry characteristic effects discussed earlier: output in young plants with above median capital intensity increases by 13%, while those below the median contract by 9%; older plants with above median capital intensity show no effect while those below show a contraction of 8.5%. More sophisticated young plants grow with proximity to the highway network, while sophisticated older plants, contract less. Plant size (employment) also follows the same pattern, albeit with less sensitivity to plant age.[18]

[17] The only significant coefficient in the census scheme suggests this gradient as well with young firms, again an insignificant 4.3% and older firms with a significant 14.4% (Appendix table A.5).

[18] Young firms with above median skills are 7.4% larger and those below are 6% (insignificant), very similar to older firms. Those above average in labor productivity show increases of 9.2% and those below, 6.9%. A similar pattern appears, although again, more mutedly, for mature firms for skills, capital intensity and labor productivity. Similar effects appear for labor productivity for small firms (capital intensity), mature firms (skills) although signs flip for above and below average labor productivity suggest perhaps mechanical mean reversion effects dominating. Similarly, TFPR for small and mature firms shows the same pattern for capital intensity differences, and perhaps mean reversion effects for labor productivity.

Table 5: Plant heterogeneity and the impact of market access: Sample Scheme

DV - Logged values of	Young					Mature					Old				
	Exit[0/1]	**Output**	**plant size (emp.)**	**Labor productivity**	**TFPR**	**Exit[0/1]**	**Output**	**plant size (emp.)**	**Labor productivity**	**TFPR**	**Exit[0/1]**	**Output**	**plant size (emp.)**	**Labor productivity**	**TFPR**
		(1)	(2)	(3)	(4)		(5)	(6)	(7)	(8)		(9)	(10)	(11)	(12)
A. Sample scheme interactions with initial skill intensity															
(0,1) District 0-10 km from GQ	-0.138++	0.098++	0.074++	-0.012	0.024	-0.107+	0.004	0.034	-0.076++	0.016	-0.104+	0.017	0.098+	-0.05	0.002
*above median	(0.066)	(0.043)	(0.031)	(0.043)	(0.022)	(0.061)	(0.037)	(0.022)	(0.037)	(0.017)	(0.054)	(0.043)	(0.056)	(0.041)	(0.027)
(0,1) District 0-10 km from GQ	-0.139+++	-0.045	-0.06	0.041	-0.017	-0.095+	0.011	-0.044++	0.096++	-0.005	-0.110++	-0.118++	-0.066+	-0.001	-0.019
*below median	(0.053)	(0.042)	(0.039)	(0.035)	(0.022)	(0.054)	(0.039)	(0.021)	(0.037)	(0.016)	(0.053)	(0.053)	(0.040)	(0.045)	(0.036)
B. Sample scheme interactions with initial capital intensity															
(0,1) District 0-10 km from GQ	-0.170+++	0.132+++	0.036	0.103+++	0.059+++	-0.100+	0.081++	0.038+	0.025	).055+++	-0.098+	-0.007	0.05	-0.024	0.024
*above median	(0.063)	(0.035)	(0.027)	(0.036)	(0.022)	(0.055)	(0.034)	(0.020)	(0.033)	(0.018)	(0.056)	(0.051)	(0.056)	(0.044)	(0.041)
(0,1) District 0-10 km from GQ	-0.112++	-0.091++	-0.04	-0.064	-0.061+++	-0.101+	-0.069+	-0.041+	-0.02	0.044+++	-0.110++	-0.085+	-0.017	-0.013	-0.038
*below median	(0.056)	(0.041)	(0.030)	(0.041)	(0.021)	(0.059)	(0.037)	(0.021)	(0.033)	(0.016)	(0.051)	(0.045)	(0.028)	(0.038)	(0.033)
C. Sample scheme interactions with initial TFPR															
(0,1) District 0-10 km from GQ	-0.148++	0.027	0.044	-0.013	-0.012	-0.092	0.005	0.02	-0.029	-0.016	-0.099++	-0.023	0.066	-0.006	-0.053
*above median	(0.058)	(0.040)	(0.035)	(0.037)	(0.025)	(0.059)	(0.030)	(0.019)	(0.032)	(0.018)	(0.049)	(0.036)	(0.055)	(0.045)	(0.036)
(0,1) District 0-10 km from GQ	-0.147++	0.028	-0.044	0.065	0.019	-0.114++	0.015	-0.017	0.034	0.03	-0.110++	-0.087	-0.033	-0.045	0.051
*below median	(0.061)	(0.041)	(0.032)	(0.043)	(0.021)	(0.055)	(0.043)	(0.019)	(0.036)	(0.020)	(0.055)	(0.062)	(0.038)	(0.048)	(0.036)
D. Sample scheme interactions with initial labor productivity															
(0,1) District 0-10 km from GQ	-0.126++	0.028	0.092+++	-0.087+	-0.03	-0.104+	-0.018	0.039+	-0.097+++	-0.011	-0.093+	-0.032	0.062	-0.039	0.008
*above median	(0.061)	(0.048)	(0.032)	(0.047)	(0.025)	(0.060)	(0.038)	(0.021)	(0.033)	(0.015)	(0.049)	(0.050)	(0.054)	(0.050)	(0.025)
(0,1) District 0-10 km from GQ	-0.152+++	0.029	-0.069++	0.108+++	0.038+	-0.096+	0.039	-0.047++	0.120+++	0.026	-0.110+	-0.043	-0.034	0.003	-0.017
*below median	(0.057)	(0.040)	(0.035)	(0.038)	(0.023)	(0.053)	(0.037)	(0.021)	(0.033)	(0.018)	(0.058)	(0.051)	(0.035)	(0.047)	(0.045)

Notes: See Table 4. Estimations in the base effect presented in Table 4 is interacted with indicator variables for above or below median values for indicated plant traits.

In the large firm sample of the census scheme there are virtually no significant coefficients on any outcome variables in the aggregate (Annex Table A.5), although breaking apart by initial conditions offers some evidence of similar patterns – higher skill and capital intensity are correlated with high output and plant size responsiveness for young and mature plants, and with higher TFPR for mature and old plants (Annex Table A.6).

These results support the expected heightened effect of increased competition and the opportunity tapped by more capable firms. For younger plants, the effect is to expand size, while for the older ones, it is to defend the existing size. The less capable older plants are eventually likely to exit. The relative rise in labor productivity of older plants, despite lower output gains suggests some relative downsizing, but also in line with district-level pseudo panel results, a possible de Loecker et al. (2016) effect.

V. Conclusion

The natural experiment offered by the GQ upgrades allows examining the adjustment process across plants of different age cohorts and characteristics to an increase in competition and access to markets. Young plants drive the previously documented aggregate positive effects of proximity to the GQ network, while older plants show insignificant or negative effects. Younger plants tend to be more capital intensive and engaged in value chains, suggesting an important domestic "offshoring" dynamic along the GQ. Though in the short-run, exit falls for all plant cohorts in districts along the GQ network, the diverging responsiveness over time between young and older plants suggests that in the long-run older plants will contract and possibly exit. Strikingly, the evidence suggests that this process may take decades.

It is also clear that building roads is not sufficient for local industrial development. The responsiveness of young plants is blunted where factor markets are incomplete, or the business climate is distorted. With the exception of the capabilities of plants and the work force, older plants are less sensitive to misallocation in input and output markets, suggesting that over time they have devised ways around them. The fact that older plants respond better to proximity to the GQ in more distorted districts suggests a de Loecker et al effect where they protect their output market while benefiting from cheaper inputs.

Finally, plant and workforce capabilities are critical complements to physical infrastructure, allowing escape from new competition and taking advantage of new markets: young plants with high human capital or capabilities drive the aggregate positive response to connectivity among these plants, while older plants with low human capital or capabilities drive their contractionary tendencies.

References

Adalet McGowan, M., D. Andrews, and V. Millot. (2017) "The Walking Dead? Zombie Firms and Productivity Performance in OECD Countries," OECD Economics Department Working Papers #1372.

Aghion, P., N. Bloom, R. Blundell, R. Griffith, and P. Howitt. (2005) "Competition and Innovation: An Inverted-U Relationship," *The Quarterly Journal of Economics*, 120(2):701–728

Aghion, P., R. Blundell, R. Griffith, P. Howitt, and S. Prantl (2009). "The effects of entry on incumbent innovation and productivity." *The Review of Economics and Statistics*, 91(1): 20‑32.

Aghion, P., S. Bechtold, L. Cassar, and H. Herz (2014). "The causal effects of competition on innovation: Experimental evidence." NBER Working Paper. National Bureau of Economic Research.

Ahlfeldt, G. M., and A. Feddersen (2010). "From periphery to core: Economic adjustments to high-speed rail." *Economics Letters*, 107(2): 172–174.

Akcigit, U., and Q. Liu. (2016) "The Role of Information in Innovation And Competition," *Journal of the European Economic Association*, 14(4): 828-870,

Akcigit, U., J. Grigsby, and T. Nicholas. (2017). "The Rise of American Ingenuity: Innovation and Inventors of the Golden Age." NBER Working Paper 23047, National Bureau of Economic Research

Anyadike-Danes, M., and M. Hart (2017), "High performing firms and job creation: a longitudinal analysis (1998-2013)", Enterprise Research Centre Insight Paper, April 2017.

Asturias, J., M. Garćıa-Santana, and R. Ramos (2019). "Competition and the welfare gains from transportation infrastructure: Evidence from the Golden Quadrilateral in India." *Journal of the European Economic Association*, 17(6): 1881–1928.

Banerjee, A., Duflo, E.and Qian, N. (2012). 'On the road: access to transportation infrastructure and economic growth in China', NBER Working Paper No. 17897, National Bureau of Economic Research.

Baum-Snow, N. (2007). "Did highways cause suburbanization?" *The Quarterly Journal of Economics*, 122(2): 775–805.

Baum-Snow, N. (2010). "Changes in transportation infrastructure and commuting patterns in US metropolitan areas, 1960–2000." *American Economic Review: Papers & Proceedings*, 100(2): 378–382.

Baum-Snow, N., L. Brandt, J. V. Henderson, M. A. Turner, and Q. Zhang (2016a). "Roads, railroads, and decentralization of Chinese cities." *The Review of Economics and Statistics*, 99(3): 435–448.

Bernard, A.B., Redding, S.J. and Schott, P.K. (2007). "Comparative advantage and heterogeneous firms." *The Review of Economic Studies*, 74(1): 31-66.

Bloom, N., & Van Reenen, J. (2007). "Measuring and explaining management practices across firms and countries." *The quarterly journal of Economics*, 122(4): 1351-1408.

Bloom, N., E. Brynjolfsson, L. Foster, R. Jarmin, M. Patnaik, I. Saporta-Eksten, J. Van Reenen. (2019) "What Drives Differences in Management Practices?" *American Economic Review*, 109(5):1648-83,

Bloom, N., Iacovone, L., Pereira-Lopez, M., & Van Reenen, J. (2022). "Management and misallocation in Mexico", NBER working paper No. w29717. National Bureau of Economic Research.

Barzin, S., D'Costa, S., Graham, D.J., 2018. "A pseudo-panel approach to estimating dynamic effects of road infrastructure on firm performance in a developing country context." *Regional Science and Urban Economics,* 70: 20-34.

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.

Chor, Davin and Kalina Manova. (2012). "Off the cliff and back? Credit conditions and international trade during the global financial crisis". *Journal of International Economics*, 87(1): 117-133,

Cosar, A. K., and P. D. Fajgelbaum (2016). "Internal geography, international trade, and regional specialization." *American Economic Journal: Microeconomics*, 8(1): 24–56.

Crozet, Matthieu and Demir, Banu and Smarzynska Javorcik, Beata. (2021). International Trade and Letters of Credit: A Double-Edged Sword in Times of Crises (October 2021). CEPR Discussion Paper No. DP16630

Cusolito, A. P., A. Garcia-Marin, A., and W. Maloney. (2022). "Proximity to the Frontier, Markups, and the Response of Innovation to Foreign Competition.", *American Economic Review Insights*

Datta, S. (2011). 'The impact of improved highways on Indian firms', *Journal of Development Economics*, 99(1): 46–57

Das, A., E. Ghani, A. Grover, W. Kerr, and R. Nanda (2023). "Infrastructure and finance: Evidence from India's GQ highway network." *Journal of Urban Economics*, 142: 103593.

De Borger, B., Mulalic, I., Rouwendal, J., 2025. "Productivity and wage effects of an exogenous improvement in transport infrastructure: Accessibility and the Great Belt Bridge." *Regional Science and Urban Economics*, 114: 104133.

De Loecker, J., P. K. Goldberg, A. K. Khandelwal , and N. Pavcnik. (2016) "Prices, Markups, and Trade Reform," *Econometrica,* 84(2): 445-510.

Donaldson, D., and R. Hornbeck (2016). "Railroads and American economic growth: A 'market access' approach." *The Quarterly Journal of Economics*, 131(2): 799–858.

Donaldson, D. (2018). "Railroads of the Raj: Estimating the impact of transportation infrastructure." *American Economic Review*, 108(4–5): 899–934.

Duranton, G., and M. A. Turner (2011). "The fundamental law of road congestion: Evidence from US cities." *American Economic Review*, 101(6): 2616–2652.

Duranton, G., M. A. Turner, and M. E. Morrow (2014). "Roads and trade: Evidence from the US." *The Review of Economic Studies*, 81(2): 681–724.

Duranton, G.; E. Ghani, Ejaz; A. Grover; W. Kerr. (2015a). "The Misallocation of Land and Other Factors of Production in India," Policy Research Working Paper; No. 7221. World Bank Group, Washington, DC.

Duranton, G.; E. Ghani, Ejaz; A. Grover; W. Kerr. (2015b) "A Detailed Anatomy of Factor Misallocation in India," World Bank Policy Research Working Paper No. 7547. World Bank Group, Washington, DC.

Duranton, G., and M. A. Turner (2012). "Urban growth and transportation." *The Review of Economic Studies*, 79(4): 1407–1440.

Eslava, M., & Haltiwanger, J. (2017). The Life-cycle Growth of Plants in Colombia: Fundamentals vs. Distortions," CAF Working paper Number 2017/18, Buenos Aires: CAF.

Faber, B. (2014). "Trade integration, market size, and industrialization: Evidence from China's National Trunk Highway System." *The Review of Economic Studies*, 81(3): 1046–1070.

Fisman, R. and I. Love (2003). "Trade credit, financial intermediary development, and industry growth". Journal of Finance, 58: 353-374

Foster, L., J. Haltiwanger, and C. Syverson. (2016). "The slow growth of new plants: learning about demand? *Economica*, 83(329): 91-129.

Foster, V., Gorgulu, N., Jain, D., Straub, S., Vagliasindi, M., 2023. "The impact of infrastructure on development outcomes: A meta-analysis." *World Bank Policy Research Working Papers*, 10350

Fujita, M., P. Krugman, and A. Venables, (1999), The Spatial Economy: Cities, Regions, and International Trade, The MIT Press, Cambridge, MA.

García-López, M.-Á., A. Holl, and E. Viladecans-Marsal (2015). "Suburbanization and highways in Spain." *Journal of Urban Economics*, 85: 52–67.

Ghani, E., A. Grover, and W.R. Kerr. (2016a). "Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing." *Economic Journal*, 126(591): 317-357.

Ghani, E., A. Grover, and W.R. Kerr. (2016b). "Highways and Spatial Location within Cities: Evidence from India," *World Bank Economic Review*, World Bank Group, 30: 97-108.

Ghani, E., A. Grover, and W. Kerr. (2012). 'Is India's manufacturing sector moving away from cities?', NBER Working Paper No. 17992, National Bureau of Economic Research.

Gibbons, S., T. Lyytikäinen, H. G. Overman, and R. Sanchis-Guarner (2019). "New road infrastructure: The effects on firms." *Journal of Urban Economics*, 110: 35–50.

Goksoy, A., Vayvay, O., & Ergeneli, N. (2013). "Gaining competitive advantage through innovation strategies: an application in warehouse management processes." *American Journal of Business and Management*, 2(4): 304-321.

Grover, A., and V. Karplus. (2021). "Coping with COVID-19: Does Management Make Firms More Resilient?." Policy Research Working Paper No. 9514. World Bank, Washington, DC.

Grover, A., D. Medvedev and E. Olafsen. (2019). High-Growth Firms: Facts, Fiction, and Policy Options for Emerging Economies (English). Washington, D.C.: World Bank Group. http://documents.worldbank.org/curated/en/391161542221639630/High-Growth-Firms-Facts-Fiction-and-Policy-Options-for-Emerging-Economies

Grover, A., L. Iacovone and P. Chakraborty. (2019). "Management Practices in Croatia: Drivers and Consequences for Firm Performance." Policy Research Working Paper; No. 9067. World Bank, Washington, DC.

Haltiwanger, J., R. S. Jarmin, and J. Miranda. (2013) "Who Creates Jobs? Small versus Large versus Young," *The Review of Economics and Statistics*, 95(2): 347–361

Holl, A. (2004a). "Transport infrastructure, agglomeration economies, and firm birth: Empirical evidence from Portugal." *Journal of Regional Science*, 44(4): 693–712.

Hornbeck, R., & Rotenberg, M. (2019). "Railroads, reallocation, and the rise of American manufacturing" NBER working paper No. w26594. National Bureau of Economic Research.

Hornbeck, R., & Rotenberg, M. (2021). "Railroads, Market Access, and Aggregate Productivity Growth." University of Chicago Booth School of Business, mimeo.

Hsieh, C.-T., and P. J. Klenow. (2009) "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 124(4):1403–1448, https://doi.org/10.1162/qjec.2009.124.4.1403

Hsieh, C. and P. Klenow. (2014). "The Life Cycle of Plants in India And Mexico," *The Quarterly Journal of Economics*, 129(3): 1035-1084.

Jedwab, R., and A. Moradi (2016). "The permanent effects of transportation revolutions in poor countries: Evidence from Africa." *The Review of Economics and Statistics*, 98(2): 268–284.

Li, Y., and M. Rama. (2015) "Firm Dynamics, Productivity Growth, and Job Creation in Developing Countries: The Role of Micro- and Small Enterprises," *World Bank Research Observer,* 30(1): 3 – 38

Martin, L., S. Nataraj, S., and A. Harrison. (2017). "In With the Big, Out With the Small: Removing Small-Scale Reservations in India". *American Economic Review*, 107(2): 354-386.

Martin, Julien, Isabelle Mejean, Mathieu Parenti. (2023). "Relationship Stickiness, International Trade, and Economic Uncertainty". *The Review of Economics and Statistics* 1-45; doi: https://doi.org/10.1162/rest_a_01396

Melitz, M.J., 2003. "The impact of trade on intra-industry reallocations and aggregate industry productivity." *Econometrica*, 71(6):1695-1725.

Melo, P. C., D. J. Graham, and R. B. Noland (2010). "Impact of transport infrastructure on firm formation: Evidence from the English regions." *Regional Science and Urban Economics*, 40(6): 451–464.

Melo, P.C., Graham, D.J., Brage-Ardao, R., 2013. "The productivity of transport infrastructure investment: a meta-analysis of empirical evidence." *Regional Science and Urban Economics*, 43(5): 695-706.

Michaels, G. (2008). "The effect of trade on the demand for skill: Evidence from the interstate highway system." *The Review of Economics and Statistics*, 90(4): 683–701.

Olley, S., & Pakes, A. (1996). "Market share, market value and innovation in a panel of British manufacturing firms." *Econometrica*, 64(6): 1263-1297.

Rajan, R. and L. Zingales. (1998). "Financial dependence and growth". The American Economic Review, 88: 559-586

Sedláček, P. and V. Sterk. 2017. "The Growth Potential of Startups over the Business Cycle." *American Economic Review*, 107(10): 3182-3210.

Sivadasan, Jagadeesh. (2009) "Barriers to competition and productivity: Evidence from India" *BE Journal of Economic Analysis & Policy*, 9(1.42):1–64.

Tybout, James, R. 2000. "Manufacturing Firms in Developing Countries: How Well Do They Do, and Why?" *Journal of Economic Literature*, 38 (1): 11-44.

World Bank. 2024. World Development Report 2024: The Middle-Income Trap. World Bank, Washington DC. http://hdl.handle.net/10986/41919

Online Annex Tables and Figures

Table A.1: Descriptive statistics

DV - Change in logged values of manufacturing trait listed in column header:	Count of districts	A. Total levels of activity in 2000, combining districts within spatial range: Output (MM Rs.)	Employment	Plant count	Avg. plant size (output, MM Rs.)	Avg. plant size (emp.)	Labor productivity	B. Total Levels of activity in 2009 combining districts within spatial: Output (MM Rs.)	Employment	Plant count	Avg. plant size (output, MM Rs.)	Avg. plant size (emp.)	Labor productivity
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(9)	(10)	(11)	(12)	(13)	(14)
Young cohort of plants (0-5 years)													
Total	311												
Nodal district for GQ	9	1.07E+10	1.17E+05	1,927	5.57E+06	61	92,001	2.14E+10	1.79E+05	2,123	1.01E+07	84	119,487
District 0-10 km from GQ	69	3.05E+10	3.37E+05	6,310	4.83E+06	53	90,377	6.55E+10	5.57E+05	6,361	1.03E+07	88	117,487
District 10-50 km from GQ	37	1.27E+10	8.44E+04	1,871	6.79E+06	45	150,498	2.10E+10	1.12E+05	1,710	1.23E+07	66	186,666
District over 50 km from GQ	196	3.73E+10	4.09E+05	9,610	3.88E+06	43	91,247	6.81E+10	6.19E+05	9,453	7.20E+06	65	109,954
Nodal district for NS-EW	9	7.48E+09	1.01E+05	1,661	4.51E+06	61	74,030	2.42E+10	2.41E+05	1,899	1.27E+07	127	100,088
District 0-10 km from NS-EW	75	1.40E+10	1.64E+05	3,918	3.57E+06	42	85,370	3.87E+10	3.50E+05	5,321	7.27E+06	66	110,355
District 10-50 km from NS-EW	54	1.23E+10	1.47E+05	2,901	4.24E+06	51	83,531	1.73E+10	1.50E+05	2,465	7.02E+06	61	115,356
District over 50 km from NS-EW	173	5.75E+10	5.35E+05	11,238	5.11E+06	48	107,391	9.57E+10	7.26E+05	9,962	9.61E+06	73	131,929
Mature cohort of plants (6-24 years)													
Total	311												
Nodal district for GQ	9	2.49E+10	2.68E+05	5,578	4.46E+06	48	92,788	2.14E+10	1.79E+05	2,123	1.01E+07	84	119,487
District 0-10 km from GQ	69	7.71E+10	1.04E+06	14,890	5.18E+06	70	74,054	6.55E+10	5.57E+05	6,361	1.03E+07	88	117,487
District 10-50 km from GQ	37	2.00E+10	1.98E+05	3,651	5.49E+06	54	100,979	2.10E+10	1.12E+05	1,710	1.23E+07	66	186,666
District over 50 km from GQ	196	9.95E+10	1.23E+06	21,615	4.60E+06	57	80,868	6.81E+10	6.19E+05	9,453	7.20E+06	65	109,954
Nodal district for NS-EW	9	1.94E+10	2.48E+05	4,207	4.62E+06	59	78,241	3.49E+10	4.31E+05	3,126	1.12E+07	138	80,810
District 0-10 km from NS-EW	75	4.19E+10	4.98E+05	9,756	4.29E+06	51	84,033	7.62E+10	7.11E+05	8,594	8.86E+06	83	107,052
District 10-50 km from NS-EW	54	3.08E+10	4.25E+05	6,817	4.52E+06	62	72,373	3.38E+10	3.18E+05	3,647	9.26E+06	87	106,129
District over 50 km from NS-EW	173	1.29E+11	1.57E+06	24,954	5.19E+06	63	82,646	1.62E+11	1.30E+06	17,048	9.50E+06	76	124,898
Old cohort of plants (25 and above years)													
Total	311												
Nodal district for GQ	9	2.67E+10	2.61E+05	3,021	8.84E+06	86	102,339	6.83E+10	5.31E+05	6,377	1.07E+07	83	128,736
District 0-10 km from GQ	69	5.00E+10	9.33E+05	4,991	1.00E+07	187	53,603	1.72E+11	1.47E+06	15,854	1.08E+07	93	116,618
District 10-50 km from GQ	37	6.70E+09	1.03E+05	1,053	6.36E+06	98	65,195	2.91E+10	2.07E+05	3,023	9.64E+06	68	141,052
District over 50 km from GQ	196	5.86E+10	8.97E+05	8,193	7.16E+06	110	65,348	1.99E+11	1.92E+06	22,682	8.77E+06	85	103,553
Nodal district for NS-EW	9	1.10E+10	1.28E+05	1,608	6.82E+06	79	85,940	4.97E+10	4.99E+05	4,916	1.01E+07	102	99,479
District 0-10 km from NS-EW	75	2.44E+10	3.41E+05	3,670	6.66E+06	93	71,637	8.83E+10	9.98E+05	11,447	7.72E+06	87	88,501
District 10-50 km from NS-EW	54	1.83E+10	3.02E+05	2,617	7.01E+06	115	60,753	3.99E+10	3.69E+05	5,581	7.16E+06	66	108,181
District over 50 km from NS-EW	173	8.83E+10	1.42E+06	9,363	9.43E+06	152	62,041	2.90E+11	2.26E+06	25,992	1.12E+07	87	128,166

Notes: Descriptive statistics calculated from Annual Survey of Industries. 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 for the organized sector. Nodal districts include Delhi, Mumbai, Kolkata, and Chennai and their contiguous suburbs (Gurgaon, Faridabad, Ghaziabad, and NOIDA for Delhi; Thane for Mumbai). The indicator variable for District 0-10 km from GQ takes a unit value for non-nodal districts that have minimum straight-line distance from the GQ network of less than 10 km; other distance-related indicator variables are defined analogously. For the NS-EW network, we define Delhi, Chandigarh, NOIDA, Gurgaon, Faridabad, Ghaziabad, Hyderabad, and Bangalore to be the nodal districts. Labor productivity is total output per employee.

Table A.2: Long-differenced estimations of the impact of GQ improvements, comparing 2009 to 2000 (Expanded spatial bands)

DV - Change in logged values of manufacturing trait listed in column header:	**Output**	**Employment**	**Plant count**	**Avg. plant size (emp.)**	**Avg. plant size (output)**	**Labor productivity**	**TFPR**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				A. Young plants (0-5 years)			
(0,1) District 0-10 km from GQ	0.931++	0.410	0.273	0.450+	0.849++	0.405++	0.202
	(0.368)	(0.314)	(0.264)	(0.245)	(0.332)	(0.202)	(0.135)
(0,1) District 10-50 km from GQ	0.297	0.051	0.147	-0.057	0.171	0.231	0.081
	(0.535)	(0.394)	(0.284)	(0.321)	(0.462)	(0.302)	(0.138)
(0,1) District 50-125 km from GQ	0.406	0.265	0.376	0.307	0.393	0.133	-0.063
	(0.399)	(0.311)	(0.251)	(0.266)	(0.353)	(0.231)	(0.120)
(0,1) District 125-200 km from GQ	0.281	0.156	0.189	0.117	0.222	-0.170	-0.051
	(0.408)	(0.331)	(0.255)	(0.291)	(0.372)	(0.226)	(0.125)
				B. Mature plants (6-24 years)			
(0,1) District 0-10 km from GQ	0.804+	0.434	0.566++	0.013	0.359	0.225	0.078
	(0.420)	(0.288)	(0.233)	(0.231)	(0.322)	(0.211)	(0.108)
(0,1) District 10-50 km from GQ	0.400	0.116	-0.018	0.026	0.410	0.344	0.085
	(0.495)	(0.312)	(0.252)	(0.292)	(0.441)	(0.282)	(0.142)
(0,1) District 50-125 km from GQ	0.716+	0.380	0.246	0.334	0.541+	0.299	0.075
	(0.382)	(0.246)	(0.218)	(0.209)	(0.306)	(0.233)	(0.126)
(0,1) District 125-200 km from GQ	0.634+	0.279	0.210	0.127	0.469	0.284	0.195
	(0.367)	(0.257)	(0.242)	(0.223)	(0.304)	(0.219)	(0.122)
				C. Old plants (25 years and above)			
(0,1) District 0-10 km from GQ	0.368	0.068	0.203	-0.041	0.173	0.360++	0.062
	(0.328)	(0.256)	(0.223)	(0.183)	(0.235)	(0.181)	(0.110)
(0,1) District 10-50 km from GQ	-0.509	-0.688++	-0.386	-0.532++	-0.208	0.363	-0.019
	(0.465)	(0.333)	(0.294)	(0.254)	(0.347)	(0.220)	(0.132)
(0,1) District 50-125 km from GQ	-0.062	-0.254	-0.302	0.051	0.239	0.259	-0.028
	(0.342)	(0.253)	(0.217)	(0.196)	(0.258)	(0.199)	(0.136)
(0,1) District 125-200 km from GQ	0.164	-0.141	-0.125	-0.134	0.255	0.376++	0.077
	(0.339)	(0.267)	(0.233)	(0.193)	(0.262)	(0.187)	(0.129)

Notes: See Table 1

Table A.3a: Long-differenced estimations comparing the impact of GQ improvements for young plants only

DV - Change in logged values of manufacturing trait listed in column header:	**Output**	**Employment**	**Plant count**	**Avg. plant size (emp.)**	**Avg. plant size (output)**	**Labor productivity**	**TFPR**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	A. Within GQ improvements: New construction versus upgrades						
(0,1) District 0-10 km from GQ*	0.464	0.194	0.078	0.139	0.230	0.267	0.099
(0,1) New construction district	(0.407)	(0.312)	(0.230)	(0.254)	(0.344)	(0.194)	(0.117)
(0,1) District 0-10 km from GQ*	0.894++	0.332	0.095	0.454++	0.990+++	0.518++	0.346++
(0,1) Road upgrade district	(0.353)	(0.306)	(0.281)	(0.220)	(0.331)	(0.221)	(0.155)
	B. GQ improvements versus the planned but not upgraded NS-EW network						
	Panel B.1: Effects for districts based upon distance from the GQ network:						
(0,1) District 0-10 km from GQ	0.710++	0.263	0.079	0.290	0.652++	0.453++	0.242+
	(0.290)	(0.233)	(0.206)	(0.189)	(0.283)	(0.185)	(0.127)
(0,1) District 10-50 km from GQ	0.139	-0.043	-0.016	-0.171	0.005	0.241	0.125
	(0.460)	(0.318)	(0.227)	(0.263)	(0.395)	(0.273)	(0.120)
	Panel B.2: Effects for districts based upon distance from the NS-EW network:						
(0,1) District 0-10 km from NS-EW	0.378	0.184	0.105	0.095	0.250	0.201	0.029
	(0.263)	(0.209)	(0.176)	(0.173)	(0.230)	(0.145)	(0.095)
(0,1) District 10-50 km from NS-EW	-0.249	-0.249	-0.185	-0.196	-0.135	0.184	0.012
	(0.352)	(0.255)	(0.214)	(0.221)	(0.320)	(0.199)	(0.114)

Notes: See Table 1. Panel B long-differenced estimations compare results from proximity to the GQ network to the NS-EW highway network that was planned for the upgrade at the same time as the GQ project but was then delayed. The regressions control for the initial district conditions listed in Table 2 and state fixed effects.

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

DV - Change in logged values of manufacturing trait listed in column header:	Output	Employment	Plant count	Avg. plant size (emp.)	Avg. plant size (output)	Labor productivity	TFPR
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Young plants (0-5 years)							
Panel A: Effects for districts based upon distance from the GQ network:							
(0,1) Nodal district	0.236	0.243	0.271	-0.052	0.114	-0.017	0.260+
	(0.775)	(0.681)	(0.548)	(0.348)	(0.388)	(0.296)	(0.136)
(0,1) District 0-10 km from GQ	0.700++	0.255	0.075	0.276	0.643++	0.456++	0.236+
	(0.290)	(0.231)	(0.205)	(0.190)	(0.284)	(0.186)	(0.127)
(0,1) District 10-50 km from GQ	0.101	-0.082	-0.032	-0.210	-0.021	0.252	0.107
	(0.469)	(0.321)	(0.230)	(0.268)	(0.404)	(0.276)	(0.119)
Panel B: Effects for districts based upon distance from the NS-EW network:							
(0,1) Nodal district for NS-EW	2.204++	1.777++	1.081+	1.409+++	1.193+++	0.133	0.195+
	(0.879)	(0.773)	(0.618)	(0.336)	(0.432)	(0.354)	(0.117)
(0,1) District 0-10 km from NS-EW section scheduled for Phase I	0.588+	0.402	0.191	0.304	0.392	0.157	0.100
	(0.323)	(0.261)	(0.224)	(0.197)	(0.267)	(0.171)	(0.102)
(0,1) District 0-10 km from NS-EW section scheduled for Phase II	0.110	-0.093	-0.004	-0.167	0.070	0.269	-0.081
	(0.428)	(0.327)	(0.265)	(0.252)	(0.356)	(0.232)	(0.157)
(0,1) District 10-50 km from NS-EW	-0.235	-0.235	-0.180	-0.181	-0.124	0.179	0.020
	(0.354)	(0.257)	(0.215)	(0.222)	(0.322)	(0.200)	(0.113)
B. Mature plants (6-24 years)							
Panel A: Effects for districts based upon distance from the GQ network:							
(0,1) Nodal district	0.975	0.825	0.574	0.218	0.422	-0.004	0.126
	(0.774)	(0.576)	(0.506)	(0.353)	(0.427)	(0.316)	(0.119)
(0,1) District 0-10 km from GQ	0.411	0.216	0.405++	-0.112	0.101	0.086	-0.009
	(0.333)	(0.222)	(0.175)	(0.172)	(0.255)	(0.153)	(0.091)
(0,1) District 10-50 km from GQ	-0.009	-0.097	-0.162	-0.115	0.114	0.170	-0.012
	(0.413)	(0.254)	(0.217)	(0.243)	(0.382)	(0.229)	(0.123)
Panel B: Effects for districts based upon distance from the NS-EW network:							
(0,1) Nodal district for NS-EW	1.769++	1.290++	0.553	1.008+++	1.190++	0.078	-0.021
	(0.879)	(0.625)	(0.528)	(0.370)	(0.524)	(0.463)	(0.130)
(0,1) District 0-10 km from NS-EW section scheduled for Phase I	0.825++	0.441++	0.085	0.462+++	0.750+++	0.252	0.015
	(0.324)	(0.215)	(0.200)	(0.174)	(0.255)	(0.158)	(0.108)
(0,1) District 0-10 km from NS-EW section scheduled for Phase II	0.231	-0.018	0.032	0.072	0.240	0.259	-0.024
	(0.393)	(0.282)	(0.221)	(0.223)	(0.310)	(0.193)	(0.123)
(0,1) District 10-50 km from NS-EW	0.144	-0.024	-0.141	0.124	0.232	0.095	0.072
	(0.314)	(0.207)	(0.188)	(0.185)	(0.272)	(0.194)	(0.107)
C. Old plants (25 years and above)							
Panel A: Effects for districts based upon distance from the GQ network:							
(0,1) Nodal district	0.366	0.258	0.809	-0.464	-0.333	0.087	0.278++
	(0.747)	(0.672)	(0.618)	(0.345)	(0.374)	(0.191)	(0.136)
(0,1) District 0-10 km from GQ	0.274	0.133	0.295	-0.028	-0.004	0.156	0.067
	(0.285)	(0.210)	(0.181)	(0.163)	(0.211)	(0.155)	(0.089)
(0,1) District 10-50 km from GQ	-0.501	-0.535+	-0.253	-0.465++	-0.319	0.169	-0.026
	(0.425)	(0.294)	(0.263)	(0.233)	(0.316)	(0.196)	(0.101)
Panel B: Effects for districts based upon distance from the NS-EW network:							
(0,1) Nodal district for NS-EW	1.377+	1.562++	0.811	0.780++	0.561	-0.240	0.076
	(0.749)	(0.678)	(0.628)	(0.312)	(0.348)	(0.204)	(0.128)
(0,1) District 0-10 km from NS-EW section scheduled for Phase I	-0.284	-0.153	-0.006	-0.133	-0.288	-0.098	0.045
	(0.293)	(0.225)	(0.202)	(0.173)	(0.230)	(0.156)	(0.104)
(0,1) District 0-10 km from NS-EW section scheduled for Phase II	0.105	-0.059	-0.184	0.200	0.315	0.127	0.063
	(0.316)	(0.229)	(0.226)	(0.199)	(0.278)	(0.243)	(0.134)
(0,1) District 10-50 km from NS-EW	-0.409	-0.332	-0.184	-0.207	-0.265	-0.117	0.106
	(0.322)	(0.227)	(0.194)	(0.173)	(0.229)	(0.151)	(0.103)

Notes: See Table 1. 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 1 and state fixed effects.

Annex Table 3c: IV estimations of the impact of GQ improvements, comparing 2009 to 2000

DV - Change in logged values of manufacturing trait listed in column header:	**Output**	**Employment**	**Plant count**	**Avg. plant size (emp.)**	**Avg. plant size (output)**	**Labor productivity**	**TFPR**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				A. Young plants (0-5 years)			
(0,1) District 0-10 km from GQ	1.024+++	0.494++	0.261	0.432++	0.778+++	0.417+++	0.236++
OLS, with no district controls	(0.272)	(0.229)	(0.204)	(0.169)	(0.220)	(0.132)	(0.098)
(0,1) District 0-10 km from line ROUTE 1	1.977++	1.135	1.083+	0.614	1.168+	0.594	0.478++
	(0.967)	(0.725)	(0.595)	(0.483)	(0.655)	(0.422)	(0.216)
Exogeneity test p-value	0.244	0.285	0.0902	0.713	0.522	0.622	0.198
(0,1) District 0-10 km from line ROUTE 2	1.786+++	1.076++	0.805+	0.726++	1.021++	0.460	0.379++
	(0.678)	(0.527)	(0.428)	(0.337)	(0.454)	(0.296)	(0.151)
Exogeneity test p-value	0.149	0.157	0.0974	0.374	0.533	0.782	0.253
				B. Mature plants (6-24 years)			
(0,1) District 0-10 km from GQ	0.578+	0.320	0.589+++	-0.128	0.078	0.123	-0.029
OLS, with no district controls	(0.316)	(0.225)	(0.190)	(0.160)	(0.214)	(0.126)	(0.079)
(0,1) District 0-10 km from line ROUTE 1	1.699++	0.628	1.234++	0.0514	0.722	0.802++	0.133
	(0.858)	(0.629)	(0.548)	(0.406)	(0.575)	(0.376)	(0.215)
Exogeneity test p-value	0.0998	0.502	0.163	0.630	0.162	0.0340	0.369
(0,1) District 0-10 km from line ROUTE 2	1.371++	0.624	0.922++	0.123	0.646	0.445+	0.115
	(0.613)	(0.458)	(0.402)	(0.281)	(0.397)	(0.235)	(0.144)
Exogeneity test p-value	0.0854	0.345	0.261	0.291	0.0580	0.0998	0.195
				C. Old plants (25 years and above)			
(0,1) District 0-10 km from GQ	0.723+++	0.480++	0.597+++	0.009	0.131	0.241++	0.020
OLS, with no district controls	(0.250)	(0.201)	(0.172)	(0.133)	(0.169)	(0.118)	(0.073)
(0,1) District 0-10 km from line ROUTE 1	1.339++	0.860	1.141++	0.0470	0.160	0.326	0.167
	(0.679)	(0.576)	(0.547)	(0.339)	(0.417)	(0.327)	(0.198)
Exogeneity test p-value	0.357	0.494	0.244	0.863	0.951	0.706	0.427
(0,1) District 0-10 km from line ROUTE 2	1.041++	0.811+	0.994++	0.0508	0.0434	0.0939	0.173
	(0.511)	(0.427)	(0.392)	(0.242)	(0.311)	(0.227)	(0.136)
Exogeneity test p-value	0.473	0.383	0.195	0.820	0.644	0.669	0.204

See Table 1. The table reports IV estimation that instruments being within 10 kilometres from the GQ network with being within 10 kilometres 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 A.4: Share of young plants by industry attributes

Industry category	Plant Share		Employment Share		Output Share	
	2000	*2009*	2000	2009	2000	2009
	A. Overall					
	24%	22%	15%	22%	19%	22%
	B. GVC intensity					
Above median	22%	20%	16%	19%	18%	20%
Below median	27%	19%	12%	18%	22%	18%
	C. External finance dependence					
Above median	21%	20%	14%	17%	17%	19%
Below median	27%	19%	15%	21%	23%	20%
	D. contractual intensity					
Above median	22%	19%	14%	19%	22%	22%
Below median	24%	19%	15%	19%	19%	19%
	E. Relationship stickiness					
Above median	25%	19%	13%	18%	19%	19%
Below median	23%	20%	17%	19%	20%	20%

Notes: The table presents the share of young plants above and below median of the given industry trait.

Table A.5: Plant-level effects of GQ access: estimation of plant exit/closure and performance outcomes

	Census scheme				
DV - Logged values of	**Exit[0/1]**	**Output**	**Avg. plant size (emp.)**	**Labor productivity**	**TFPR**
	(1)	(2)	(3)	(4)	(5)
			A. All		
(0,1) District 0-10 km from GQ	-0.055	0.041	-0.012	0.057	-0.009
	(0.053)	(0.039)	(0.025)	(0.037)	(0.022)
N	7096	9437	9431	9419	9196
R^2	0.01	0.964	0.949	0.942	0.903
			B. Young		
(0,1) District 0-10 km from GQ	-0.029	0.068	0.038	0.043	0.039
	(0.061)	(0.071)	(0.056)	(0.077)	(0.041)
N	1802	2292	2288	2286	2228
R^2	0.012	0.961	0.932	0.939	0.891
			C. Mature		
(0,1) District 0-10 km from GQ	-0.059	0.03	0.006	0.011	-0.027
	(0.058)	(0.062)	(0.042)	(0.065)	(0.031)
N	3552	4727	4726	4720	4609
R^2	0.009	0.963	0.949	0.938	0.907
			D. Old		
(0,1) District 0-10 km from GQ	-0.071	0.039	-0.08	0.144++	-0.014
	(0.052)	(0.061)	(0.050)	(0.059)	(0.051)
N	1742	2418	2417	2413	2359
R^2	0.011	0.969	0.96	0.95	0.901

Notes: See Table 4. This table replicates the analysis from Table 4 for the larger plants covered in the census scheme of ASI.

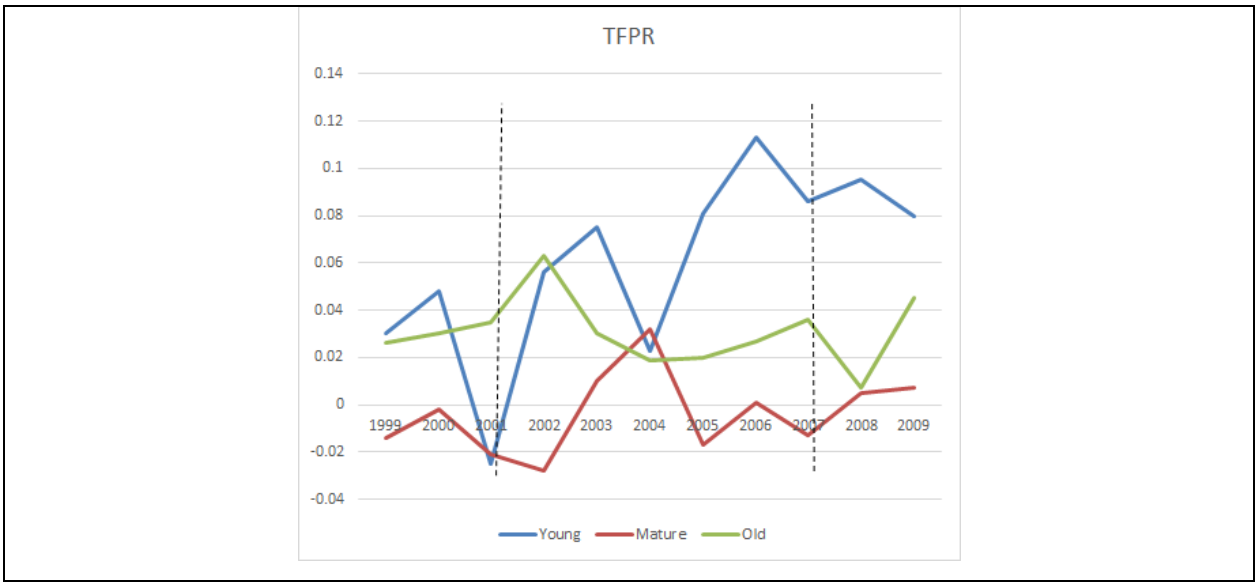
Table A.6: Plant heterogeneity and the impact of market access: Census Scheme

DV - Logged values of	Young					Mature					Old				
	Exit[0/1]	**Output**	**plant size (emp.)**	**Labor productivity**	**TFPR**	**Exit[0/1]**	**Output**	**plant size (emp.)**	**Labor productivity**	**TFPR**	**Exit[0/1]**	**Output**	**plant size (emp.)**	**Labor productivity**	**TFPR**
		(1)	(2)	(3)	(4)		(5)	(6)	(7)	(8)		(9)	(10)	(11)	(12)
A. Census scheme interactions with initial skill intensity															
(0,1) District 0-10 km from GQ	-0.042	0.075+	0.04	0.027	0.018	-0.066	0.041	0.035	-0.014	0	-0.075	0.047	0.045+	0.014	-0.034
*above median	(0.064)	(0.043)	(0.031)	(0.038)	(0.026)	(0.061)	(0.035)	(0.028)	(0.030)	(0.012)	(0.057)	(0.032)	(0.025)	(0.033)	(0.022)
(0,1) District 0-10 km from GQ	-0.032	-0.06	-0.086++	0.036	-0.027	-0.055	-0.033	-0.028	-0.003	-0.012	-0.055	-0.061+	-0.070+++	0.019	-0.076++
*below median	(0.066)	(0.047)	(0.036)	(0.037)	(0.019)	(0.064)	(0.028)	(0.024)	(0.030)	(0.019)	(0.056)	(0.033)	(0.025)	(0.031)	(0.024)
B. Census scheme interactions with initial capital intensity															
(0,1) District 0-10 km from GQ	-0.086	0.065+	0.041	0.016	0.006	-0.076	0.072+++	0.048++	0.011	0.028++	-0.068	0.032	0.032	0.014	-0.033+
*above median	(0.060)	(0.038)	(0.026)	(0.035)	(0.018)	(0.058)	(0.027)	(0.022)	(0.029)	(0.013)	(0.051)	(0.035)	(0.022)	(0.027)	(0.020)
(0,1) District 0-10 km from GQ	-0.003	-0.07	-0.106+++	0.049	-0.018	-0.047	-0.073++	-0.045	-0.033	-0.047+++	-0.053	-0.045	-0.056+	0.02	-0.077++
*below median	(0.070)	(0.047)	(0.035)	(0.037)	(0.027)	(0.060)	(0.030)	(0.029)	(0.028)	(0.017)	(0.061)	(0.032)	(0.029)	(0.033)	(0.022)
C. Census scheme interactions with initial TFPR															
(0,1) District 0-10 km from GQ	-0.042	-0.001	-0.024	0.022	-0.011	-0.03	-0.01	-0.001	-0.018	-0.016	-0.075	-0.004	-0.008	0.013	-0.060++
*above median	(0.072)	(0.036)	(0.023)	(0.028)	(0.015)	(0.064)	(0.024)	(0.022)	(0.021)	(0.011)	(0.052)	(0.024)	(0.021)	(0.021)	(0.017)
(0,1) District 0-10 km from GQ	-0.048	0.084	-0.023	0.128	0.076	-0.084	0.126	0.087++	0.024	0.092++	-0.043	-0.079	-0.022	-0.004	-0.014
*below median	(0.059)	(0.102)	(0.057)	(0.085)	(0.063)	(0.057)	(0.088)	(0.040)	(0.078)	(0.045)	(0.063)	(0.120)	(0.064)	(0.093)	(0.057)
D. Census scheme interactions with initial labor productivity															
(0,1) District 0-10 km from GQ	-0.053	0.024	0.029	-0.004	0.004	-0.066	0.019	0.066+++	-0.064++	-0.02	-0.048	0.014	0.015	0.006	-0.036+
*above median	(0.059)	(0.033)	(0.026)	(0.031)	(0.019)	(0.052)	(0.026)	(0.021)	(0.026)	(0.013)	(0.062)	(0.029)	(0.024)	(0.035)	(0.020)
(0,1) District 0-10 km from GQ	-0.017	-0.009	-0.081++	0.073+	-0.013	-0.05	-0.013	-0.076+++	0.064+	0.012	-0.077	-0.027	-0.036	0.027	-0.075++
*below median	(0.072)	(0.052)	(0.032)	(0.037)	(0.024)	(0.069)	(0.033)	(0.026)	(0.034)	(0.017)	(0.055)	(0.045)	(0.030)	(0.032)	(0.025)

Notes: See Table 5.

Figure A.1: Cumulative effects of GQ completion





Notes: To estimate the cumulative systemic effect of GQ upgrades for each year starting from 1999 to 2009, relative to the year 1994 the specification is:

$$Y_{i,t} = \sum_{1999}^{2009} \beta_t * (0,1)year_t * (0,1)GQDist_{i,d<10km} + \emptyset_i + \vartheta_t + \varepsilon_{i,t}$$

Both estimations include a vector of district and state fixed effects $\emptyset_i$ and a vector of year fixed effects ϑ_t which control for the main effects of distance from the GQ network, and of the post GQ-upgrades period. Standard errors are clustered by district and we weight by districts' log population in 2001. The sample comprises 1,188 observations as the cross of 12 years with the 99 non-nodal districts within 50 kilometers of the GQ network for which we can always observe their activity. For parsimony, we drop the lagged dependent variable which has little effect on the coefficients, and we use the 10-50 category as the base case since it behaves similarly to the 50+, dropping the latter from the sample. Figure A.1 presents the cumulative system time path generated by estimating the above equation. The vertical lines mark when the GQ upgrades began and when they reached the 80% completion mark in 2005.