For Online Publication:

Internet Appendix for

Business Group Spillovers

Naaraayanan and Wolfenzon

Forthcoming in the Review of Financial Studies

Appendix A Matching datasets

In this section, we discuss the matching between firm-bank pairs in Prowess to firm-bank pairs in the loan-level dataset from the credit registry at the MCA. We start with the sample of firms in Prowess and match them to firms in the loan-level dataset scrape between May 2021 to December 2021. We match them using the company identifier (CIN) provided by the MCA, yielding a match rate of 95%.

In this matched dataset, we keep loans starting from 1980 until 2016. We drop short maturity loans, those that are less than 3 years, and keep loan amounts larger than 10 million INR. Subsequently, we carry out cleaning and standardize names to merge with information on banks and financial institutions from Prowess. Note that this standardization on names is performed on both datasets i.e., Prowess and the loans. In the next step, we perform a fuzzy match of the names across the two datasets, yielding a match rate of 57% (out of 1,175 names in Prowess). This was a non-trivial task due to the fact that the loan-level sample included various financial institutions, including non-banks and private entities for which we had no information. As a result, we decided to only focus on banks and financial institutions reported in Prowess, resulting in a sample that covers larger firm-bank pairs.

In the last step, we merge the dataset to the sample of standalone firms that lay along the GQ road network. This merge yields 2,430 loans to 302 unique firms and 140 banks. Further, in our empirical specification, we require that standalone firms: (i) borrow from multiple banks, and (ii) borrow from multiple banks both before and after the GQ upgrades. Such a restriction leads the final sample to consist of 163 loans to 17 unique firms to 15 unique banks. Appendix Table I reports the descriptive statistics for firms and loan level characteristics for the matched sample.

Appendix B Discussion of identifying assumptions

In this section, we discuss in detail the identifying assumptions for interpreting β in Equation 2 as the causal impact of GQ upgrades. The identifying assumption is that the timing of the road construction is orthogonal to the investment opportunities for firms located along the GQ road network. We confirm the validity of this assumption by (i) ruling out pre-trends in investments and (ii) showing that observable differences in firm characteristics cannot explain the timing of GQ road upgrade.

We begin by assessing pre-trends in investments. Results are presented in event-time in panel A of Appendix Figure I. This figure confirms that treated firms indeed did not show any pre-trends, thus ruling out concerns that the increase in investment would have occurred regardless of the road upgrade.

Next, we investigate whether observable firm characteristics can predict the timing of road construction. To do so, we run Cox hazard rate regressions of the time to road upgrade. The explanatory variables include city-level averages (calculated over different periods) of investments, market concentration (based on sales), market share (based on sales), fixed assets, investment growth, sales growth, cash holdings, and profitability.

Appendix Table II presents the results. Panel A reports results using city-level averages computed using all firms while panel B reports results using city-level averages computed using only group-affiliated firms. All specifications include state fixed effects. The results suggest that observable differences in average firm characteristics cannot explain the timing of GQ road upgrade. While it is impossible to test whether the timing of the road upgrade is orthogonal to the (unobserved) investment opportunities, the results in this table provide comfort that at least the timing is orthogonal to a broad set of observable firm characteristics that are likely correlated with investments.

Lastly, in panel B of Appendix Figure I, we show robustness of event study results to

recent concerns about the two-way fixed effect estimator providing biased estimates when treatment is staggered and in the presence of treatment effect heterogeneity (Goodman-Bacon 2021; Roth, Sant'Anna, Bilinski, and Poe 2023).

The imputation estimator and other recent econometric advances in the staggered treatment adoption literature propose using the "never treated" group as the control to obtain unbiased estimates of the treatment effect. As our baseline sample consists of all firms that are eventually treated, we identify the control group by expanding our sample to include firms that did not lay along the GQ road network. Further, as business group shares and organizational form are not randomly allocated, we choose to match each treatment firm in our sample to a control firm based on the following pre-GQ characteristics, one year before : (i) the level of business group share in the city, and (ii) the same organizational form (group-affiliated vs. standalone). Lastly, we also require that the control firms have the same (i) the level of investment, (ii) the level of leverage, and (iii) operate in the same state and two-digit industry, as the treatment firm. The figure confirms that (i) there are no pre-trends prior to the GQ upgrades, and (ii) the increase in investments start around two years into the GQ upgrades and are persistent up to five years after the GQ road upgrades.

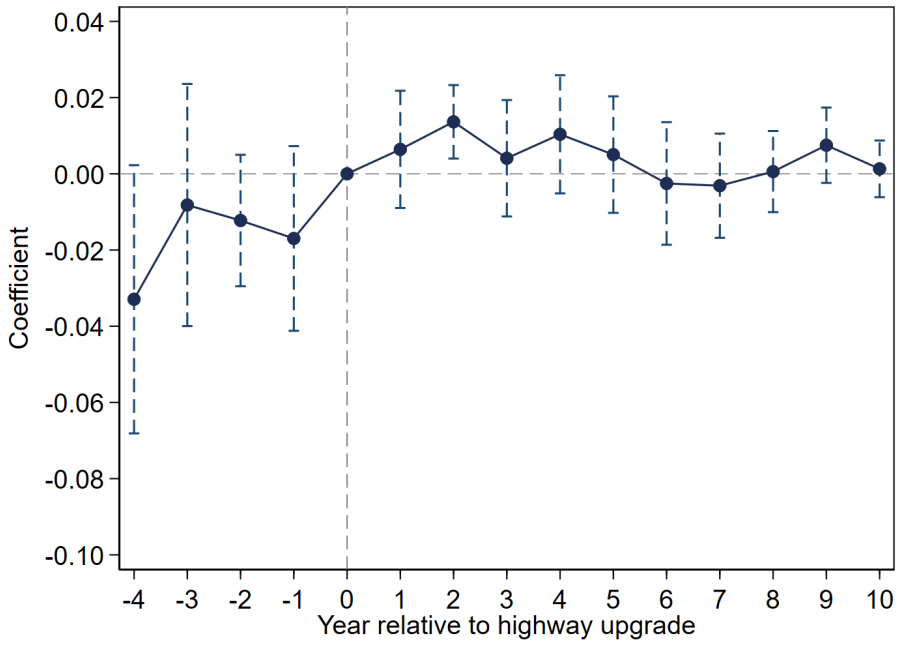
Appendix C Robustness to fixed effects estimator

In this section, we examine robustness of our baseline results to recent concerns about the two-way fixed effects estimator providing biased estimates when treatment is staggered and in the presence of treatment effect heterogeneity (Goodman-Bacon 2021; Roth, Sant'Anna, Bilinski, and Poe 2023).

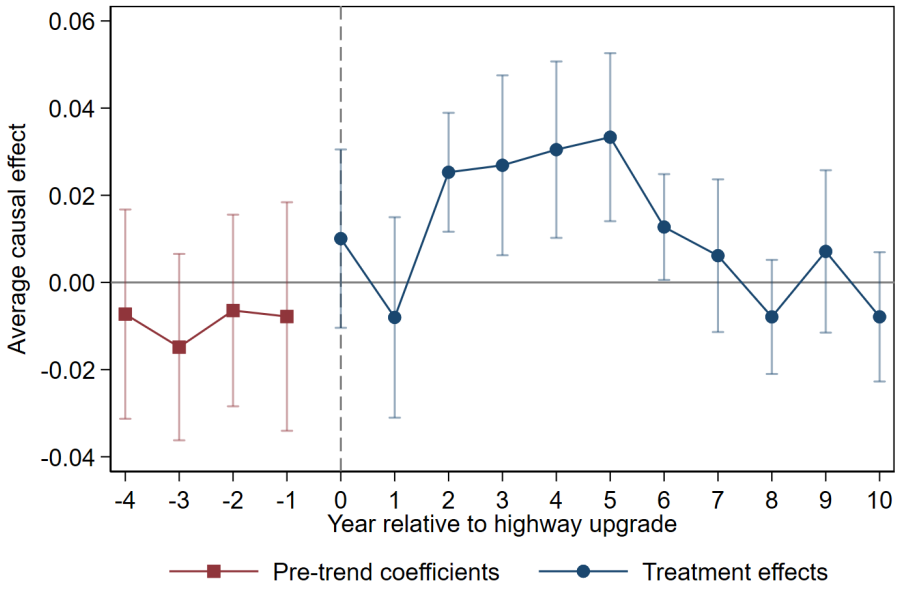
The imputation estimator and other recent econometric advances in the staggered treatment adoption literature propose using the "never treated" group as the control to obtain unbiased estimates of the treatment effect. As our baseline sample consists of all firms that are eventually treated, we identify the control group by expanding our sample to include firms that did not lay along the GQ road network. Further, as business group shares and organizational form are not randomly allocated, we choose to match each treatment firm in our sample to a control firm based on the following pre-GQ characteristics, one year before : (i) the level of business group share in the city, and (ii) the same organizational form (i.e., standalone). Lastly, we also require that the control firms have the same (i) the level of investment, (ii) the level of leverage, (iii) operate in the same state and two-digit industry, and (iv) operate in *High or Low BGS* regions, as the treatment firm.

In Appendix Table III, we present the results from two separate regressions for firms in *High BGS* and *Low BGS* regions. Each regression is akin to a difference-in-differences (DiD) estimation. The table confirms that there are no pre-trends for firms in either regions prior to the GQ upgrades. More importantly, there is a lower investment sensitivity to GQ upgrades for firms in *High BGS* regions relative to control firms (column 1) but for firms in *Low BGS* regions relative to control firms, we find a higher investment sensitivity to the GQ upgrades (column 2). These results are consistent with our baseline estimates from Figure III. Moreover, we note that the lower investment sensitivity in *High BGS*

regions start in the immediate year around the GQ upgrades and are persistent up to ten years after the GQ road upgrades while the higher investment sensitivity in *Low BGS* regions are temporary.



(a) Fixed effects estimator



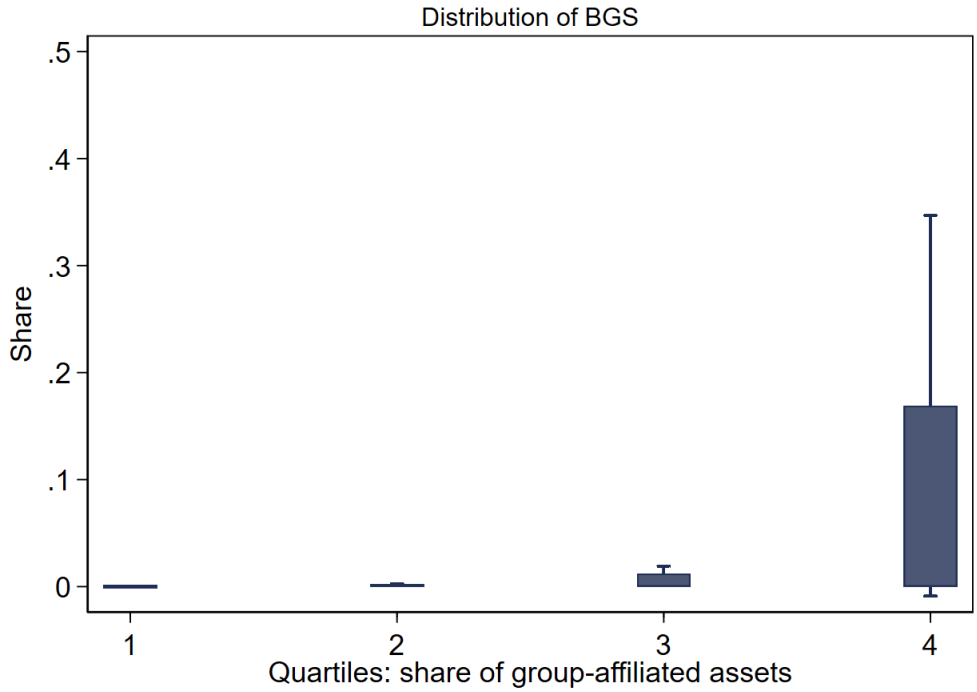
(b) Imputation estimator

Appendix Figure I. Firm investment around GQ upgrades

Panel A of the figure displays the dynamic coefficients (μ_k) and their corresponding 90% confidence intervals of investment by firms around the upgrade of the GQ road network. We use the fixed effects estimator to estimate a fully dynamic specification that allows us to capture the dynamics of firm investment relative to the year of commencement of GQ upgrade. Specifically, we estimate the following equation:

$$\text{Investment}_{ijcst} = \alpha_i + \sum_{k=-1}^{-4} \mu_k + \sum_{k=1}^{10} \mu_k + \theta_{jt} + \epsilon_{ijcst}$$

All coefficients are plotted relative to investment at k=0, which is normalized to zero. Panel B of the figure plots coefficients from the imputation estimator suggested in **?** and includes matched pair × year fixed effects. Details are in Appendix B.



Appendix Figure II. Distribution of BGS

This figure illustrates the variation the share of group-affiliated assets split by quartiles. The bars display the mean within each group while the lines present the standard deviation within each group.

Appendix Table I. Summary statistics, loan-level dataset

This table reports the descriptive statistics of firms matched to the Prowess sample. Panel A reports the loan-level characteristics while panel B reports the descriptive statistics for firms that borrow from multiple banks five years around the GQ upgrades. Following **?**, we aggregate multiple loans of a firm from the same bank and collapse the data at the firm-bank pair level with two observations (pre and post) for each pair. All variables are defined in Appendix Table XIV. Data source: CMIE Prowess and Ministry of Corporate Affairs.

	Panel A: Loan-level characteristics					
	N	**Mean**	**SD**	**P25**	**P50**	**P75**
Loan amount (INR millions)	163	147.2	383.4	26.0	70.0	136.6
Change in log lending	163	0.2	1.0	-0.4	0.2	0.8
Group exposure amount (INR billions)	163	55.68	64.51	8.47	39.25	126.19

	Panel B: Firm-level characteristics					
	N	**Mean**	**SD**	**P25**	**P50**	**P75**
Total assets	163	3,999	7,150	716	1,458	4,441
Firm age (years)	163	31.33	16.60	19.00	26.00	38.00
Cash flow	163	0.08	0.10	0.03	0.07	0.13
Profitability	163	0.13	0.08	0.10	0.13	0.16
Listed	163	0.60	0.49	0.00	1.00	1.00
Investment	163	0.46	0.20	0.31	0.45	0.60
Debt	163	0.29	0.18	0.18	0.28	0.37
Total factor productivity	163	2.98	2.17	1.87	2.29	3.30

Appendix Table II. Timing of GQ and pre-existing firm characteristics: Survival analysis

A Cox proportional hazards model is fitted to investigate the predictability of placement of GQ segments based on pre-existing firm characteristics. The explanatory variables include city-level averages (calculated over different periods) of investments, market share (based on sales), firm assets, investment growth, sales growth, cash holdings, and profitability. Panel A reports results using city-level averages computed using all firms while panel B reports results using city-level averages computed using only group firms. All specifications include state fixed effects. Standard errors are corrected for heteroscedasticity and autocorrelation and clustered at the city level. Standard errors are reported in parentheses. *,**, and *** indicate significance at 10%, 5%, and 1% respectively. All variables are defined in Appendix Table XIV. Data source: CMIE Prowess.

	Panel A. All firms			
	1994-97	1992-97	1990-97	1988-97
	(1)	(2)	(3)	(4)
Investment (City avg.)	-39.324	3.266	3.599	15.796
	(27.251)	(15.239)	(18.431)	(21.039)
Market share (City avg.)	-1.305	-0.689	-0.835	-0.697
	(3.277)	(2.172)	(2.163)	(2.130)
Sales growth (City avg.)	2.994	-0.768	-0.885	-0.592
	(2.857)	(1.517)	(1.482)	(1.506)
Investment growth (City avg.)	4.197	0.697	0.877	0.519
	(3.851)	(2.881)	(2.788)	(2.762)
Profitability (City avg.)	-11.911	-5.321	-5.411	-5.750
	(8.407)	(3.502)	(3.542)	(3.502)
Firm Size (City avg.)	-0.670	-0.052	-0.040	-0.050
	(0.742)	(0.336)	(0.315)	(0.306)
State fixed effects	Yes	Yes	Yes	Yes
χ^2 statistic	53.856	46.665	45.915	47.201
Log Pseudo likelihood	-75.639	-101.905	-101.856	-101.695
Observations	430	430	430	430

	Panel B. Group firms			
	1994-97	1992-97	1990-97	1988-97
	(1)	(2)	(3)	(4)
Investment (City avg.)	-2.262	-5.438	-5.236	17.734
	(20.788)	(25.234)	(30.136)	(41.819)
Market share (City avg.)	-1.775	-1.194	-1.379	-1.816
	(2.339)	(1.250)	(1.363)	(1.606)
Sales growth (City avg.)	1.891	0.213	-0.268	0.253
	(1.613)	(1.631)	(1.743)	(1.876)
Investment growth (City avg.)	-0.690	-0.999	-1.446	-1.757
	(1.427)	(1.060)	(1.232)	(1.516)
Profitability (City avg.)	7.711	6.770	9.246	9.189
	(12.036)	(6.473)	(7.509)	(7.431)
Firm Size (City avg.)	0.307	0.345	0.325	0.301
	(0.374)	(0.418)	(0.442)	(0.437)
State fixed effects	Yes	Yes	Yes	Yes
χ^2 statistic	72.366	51.023	65.776	78.016
Log Pseudo likelihood	-69.759	-70.971	-70.827	-70.784
Observations	364	364	364	364

Appendix Table III. Robustness to fixed effects estimator

This table presents estimates from the imputation estimator suggested in Borusyak, Jaravel, and Spiess (2022). Column 1 (column 2) presents coefficients for firms located in *High BGS (Low BGS)* regions compared to matched control firms from the "never treated" cities that did not lay along the GQ road network. *Before*^{k} *(After*^{k}*)* is an indicator variable taking value 1 for all years leading upto (after) the GQ upgrade in the city. *High BGS* is an indicator variable set to one if the share of assets of group-affiliated firms from that city is in the top quartile in the year before the announcement of the GQ road network upgrades. All regressions include matched pair × year fixed effects. Details are in Appendix C. Standard errors are corrected for heteroscedasticity and autocorrelation and clustered at the city level. Standard errors are reported in parentheses.*,**, and *** indicate significance at 10%, 5%, and 1% respectively. All variables are defined in Appendix Table XIV. Data source: CMIE Prowess.

Dependent variable	Investment	
	High BGS	Low BGS
	(1)	(2)
Before^{-4}	-0.058	0.049
	(0.054)	(0.0369)
Before^{-3}	-0.045	0.059
	(0.041)	(0.044)
Before^{-2}	-0.009	0.071
	(0.020)	(0.0296)
Before^{-1}	0.004	0.154
	(0.019)	(0.0502)
After^0	-0.020***	0.096
	(0.009)	(0.0422)
After^1	-0.063***	-0.024
	(0.015)	(0.0244)
After^2	0.006	0.183**
	(0.015)	(0.0262)
After^3	-0.027*	0.143*
	(0.015)	(0.0361)
After^4	0.013	-0.049
	(0.011)	(0.0287)
After^5	0.009	-0.011
	(0.0121)	(0.047)
After^6	-0.085***	-0.016
	(0.0128)	(0.0348)
After^7	-0.017	-0.059**
	(0.0119)	(0.0393)
After^8	-0.061*	-0.07
	(0.0129)	(0.0473)
After^9	0.013	-0.011
	(0.0137)	(0.0476)
After^{10}	-0.033**	-0.017
	(0.012)	(0.0452)
Fixed effects:		
Matched pair × year	Yes	Yes
Observations	8,465	5,601
Sample: Standalone firms	Yes	Yes

Appendix Table IV. Horse race regressions, robustness to firm size

This table presents estimates from horse race regressions relating the effect of business group prevalence on standalone firms' investment. We consider the following covariates: *Listed share*, *Firm age*, and *TFP*. For each covariate, we define an indicator variable which is set to one if the specific characteristic of firms from that city is in the top quartile in the year before the announcement of the GQ road network upgrades. Panel A defines the indicator based on all firms while Panel B defines the indicator using only standalone firms. *PostGQ* is an indicator variable taking value 1 for all years including and after the GQ upgrade in the city. *High BGS* is an indicator variable set to one if the share of assets of group-affiliated firms from that city is in the top quartile in the year before the announcement of the GQ road network upgrades. All regressions include firm fixed effects, High BGS × state × year, and High BGS × industry × year fixed effects. As TFP is estimated for manufacturing firms, columns 4 and 5 restrict the sample to these industries. Standard errors are corrected for heteroscedasticity and autocorrelation and clustered at the city level. Standard errors are reported in parentheses.*,**, and *** indicate significance at 10%, 5%, and 1% respectively. All variables are defined in Appendix Table XIV. Data source: CMIE Prowess.

Dependent variable	**Panel A: Definition using all firms**				
	Investment				
	Baseline (1)	Listed share (2)	Firm size (3)	TFP (4)	All (5)
PostGQ	0.039	0.046*	0.039	0.002	-0.028
	(0.028)	(0.027)	(0.029)	(0.035)	(0.043)
PostGQ × High BGS	-0.038***	-0.084***	-0.094	-0.105**	-0.170***
	(0.012)	(0.029)	(0.092)	(0.041)	(0.058)
PostGQ × High Listed Share		-0.216**			-0.262***
		(0.095)			(0.083)
PostGQ × High Firm Size			0.016		0.003
			(0.090)		(0.044)
PostGQ × High Firm TFP				0.006	0.012
				(0.025)	(0.025)
Fixed effects:					
Firm	Yes	Yes	Yes	Yes	Yes
High BGS × industry × year	Yes	Yes	Yes	Yes	Yes
High BGS × state × year	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.70	0.70	0.70	0.66	0.66
Observations	15,842	15,842	15,842	11,520	11,520
Sample: Standalone firms	Yes	Yes	Yes	Yes	Yes

Continued...

Dependent variable	**Panel B: Definition using standalone firms**				
	Investment				
	Baseline (1)	Listed share (2)	Firm size (3)	TFP (4)	All (5)
PostGQ	0.039	0.061**	0.101*	0.007	0.105*
	(0.028)	(0.030)	(0.053)	(0.036)	(0.052)
PostGQ × High BGS	-0.038***	-0.099***	-0.064**	-0.110**	-0.116**
	(0.012)	(0.032)	(0.029)	(0.041)	(0.044)
PostGQ × High Listed Share (standalones)		-0.065**			-0.009
		(0.025)			(0.055)
PostGQ × High Firm Size (standalones)			-0.075*		-0.168***
			(0.042)		(0.061)
PostGQ × High Firm TFP (standalones)				-0.026	-0.060
				(0.031)	(0.036)
Fixed effects:					
Firm	Yes	Yes	Yes	Yes	Yes
High BGS × industry × year	Yes	Yes	Yes	Yes	Yes
High BGS × state × year	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.70	0.70	0.70	0.66	0.66
Observations	15,842	15,842	15,842	11,520	11,520
Sample: Standalone firms	Yes	Yes	Yes	Yes	Yes

Appendix Table V. Stock price reactions to new plant announcements by standalone firms

This table shows stock price reactions to announcement of new plants by standalone firms around GQ upgrade as a function of business group share. Panel A presents results without controlling for project size while Panel B includes size decile fixed effects. Across both panels, the dependent variables are cumulative abnormal returns (CARs) and cumulative excess returns (CERs), and we use several event windows starting from one day before to one day after the announcement of a new plant. To calculate the abnormal returns, we assume a single-factor model, where beta is estimated using the data from the pre-event window. *Abnormal returns* are estimated as the difference between the return on a firm's stock and the return predicted by the capital asset pricing model (CAPM) with the S&P Nifty as the benchmark market portfolio. *Excess returns* are measured as the difference between the return on a firm's stock and the return on the benchmark S&P Nifty index. *PostGQ* is an indicator variable taking value 1 for all years including and after the GQ upgrade in the city. *High BGS* is an indicator variable set to one if the share of assets of group-affiliated firms from that city is in the top quartile in the year before the GQ road network upgrades. All regressions include firm and industry x year fixed effects. Standard errors are corrected for heteroscedasticity and autocorrelation and clustered at the city level. Standard errors are reported in parentheses.*,**, and *** indicate significance at 10%, 5%, and 1% respectively. Data sources: CMIE Prowess and CapEx database.

	Panel A: Without controlling for project size							
	Cumulative abnormal returns				Cumulative excess returns			
Event window	(-1,1)	(-2,2)	(-3,3)	(-5,5)	(-1,1)	(-2,2)	(-3,3)	(-5,5)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PostGQ	0.090***	0.120***	0.184***	0.218***	0.081***	0.102***	0.143***	0.133***
	(0.005)	(0.012)	(0.014)	(0.019)	(0.004)	(0.011)	(0.014)	(0.014)
PostGQ × High BGS	-0.005	0.032	0.021	0.030	0.015	0.052	0.035	0.073
	(0.027)	(0.035)	(0.023)	(0.060)	(0.033)	(0.038)	(0.026)	(0.080)
Fixed effects:								
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry × year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.26	0.24	0.19	0.21	0.23	0.21	0.16	0.16
Observations	1,759	1,759	1,759	1,759	1,759	1,759	1,759	1,759
Sample: Standalone firms	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

	Panel B: Controlling for project size							
	Cumulative abnormal returns				Cumulative excess returns			
Event window	(-1,1)	(-2,2)	(-3,3)	(-5,5)	(-1,1)	(-2,2)	(-3,3)	(-5,5)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PostGQ	0.117***	0.161***	0.226***	0.254***	0.092***	0.125***	0.161***	0.141***
	(0.015)	(0.020)	(0.024)	(0.032)	(0.016)	(0.021)	(0.026)	(0.033)
PostGQ × High BGS	-0.026	0.026	0.013	0.024	-0.015	0.051	0.030	0.064
	(0.019)	(0.026)	(0.028)	(0.065)	(0.025)	(0.030)	(0.026)	(0.074)
Fixed effects:								
Size decile	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry × year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.33	0.31	0.26	0.27	0.31	0.30	0.27	0.29
Observations	1,078	1,078	1,078	1,078	1,078	1,078	1,078	1,078
Sample: Standalone firms	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Appendix Table VI. Standalone firms' inventory efficiency around GQ upgrades

This table presents estimates comparing days sales of inventory for standalone firms around the investment opportunity shock and as a function of business group prevalence in the local area. *PostGQ* is an indicator variable taking value 1 for all years including and after the GQ upgrade in the city. *High BGS* is an indicator variable set to one if the share of assets of group-affiliated firms from that city is in the top quartile in the year before the announcement of the GQ road network upgrades. All regressions include firm fixed effects, High BGS × state × year, and High BGS × industry × year fixed effects. We restrict the sample to firms with days sales of inventory between 5 and 150 days to mitigate the effect of outliers. Standard errors are corrected for heteroscedasticity and autocorrelation and clustered at the city level. Standard errors are reported in parentheses.*,**, and *** indicate significance at 10%, 5%, and 1% respectively. All variables are defined in Appendix Table XIV. Data source: CMIE Prowess.

Dependent variable	Days sales of inventory	
	(1)	(2)
PostGQ	-10.479**	-23.713**
	(4.621)	(9.332)
PostGQ × High BGS		-3.397
		(3.096)
Fixed effects:		
Firm	Yes	Yes
High BGS × industry × year	Yes	Yes
High BGS × state × year	Yes	Yes
Adjusted-R^2	0.63	0.63
Observations	10,845	10,845
Sample: Standalone firms	Yes	Yes

Appendix Table VII. Mechanism: Political connections in infrastructure-related industries

This table presents estimates from regressions examining political connections as a plausible mechanism. Panel A focuses on the baseline measure of business group prevalence while panel B focuses on a measure of business group prevalence based on the largest 25 business groups to proxy for political influence. In both panels, column 1 restricts the sample to group-affiliated firms while column 2 restricts the sample to standalone firms. *PostGQ* is an indicator variable taking value 1 for all years including and after the GQ upgrade in the city. *High BGS* is an indicator variable set to one if the share of assets of group-affiliated firms from that city is in the top quartile in the year before the GQ road network upgrades. *High BGS (Largest 25)* is an indicator variable set to one, if the share of assets of group-affiliated firms from that city that belongs to the 25 largest (by size) business groups, is in the top quartile in the year before the GQ road network upgrades. The sample includes firms operating in "other manufacturing industries" (NIC code: 321-329), "coke and refined petroleum products" (NIC code: 191-199), and "construction firms" (NIC code: 420-439). All regressions include firm and year fixed effects. Due to small number of clusters in these tests, the standard errors are corrected for heteroscedasticity and are reported in parentheses.*,**, and *** indicate significance at 10%, 5%, and 1% respectively. All variables are defined in Appendix Table XIV. Data source: CMIE Prowess.

	Panel A: Baseline	
Dependent variable	Investment	
	(1)	(2)
PostGQ	0.111	-0.095***
	(0.118)	(0.032)
PostGQ × High BGS	0.100	0.007
	(0.050)	(0.048)
Fixed effects:		
Firm	Yes	Yes
Industry × year	Yes	Yes
Adjusted-R^2	0.65	0.74
Observations	524	1,477
Sample	Group-affiliated firms	Standalone firms

	Panel B: Largest 25 groups	
Dependent variable	Investment	
	(1)	(2)
PostGQ	0.111	-0.023
	(0.118)	(0.071)
PostGQ × High BGS (Largest 25)	0.100	-0.080*
	(0.050)	(0.039)
Fixed effects:		
Firm	Yes	Yes
Industry × year	Yes	Yes
Adjusted-R^2	0.65	0.74
Observations	524	1,477
Sample	Group-affiliated firms	Standalone firms

Appendix Table VIII. Robustness: Controlling for pre-GQ firm characteristics

This table reports robustness to controlling for preGQ firm characteristics. *PostGQ* is an indicator variable taking value 1 for all years including and after the GQ upgrade in the city. *High BGS* is an indicator variable set to one if the share of assets of group-affiliated firms from that city is in the top quartile in the year before the GQ road network upgrades. In column 2, we interact *PostGQ* with whether the firm is *Listed* in the year before the GQ road network upgrades. In columns 3 and 4, we interact *PostGQ* with the median total factor productivity and natural logarithm of firm age before the GQ road network upgrades, respectively. All regressions include firm fixed effects, High BGS x industry x year fixed effects, and High BGS x state x year. Standard errors are corrected for heteroscedasticity and autocorrelation and clustered at the city level. Standard errors are reported in parentheses.*,**, and *** indicate significance at 10%, 5%, and 1% respectively. All variables are defined in Appendix Table XIV. Data source: CMIE Prowess.

Dependent variable	Investment				
	Baseline	Listed	TFP	Firm age	All
	(1)	(2)	(3)	(4)	(5)
PostGQ	0.039	-0.051	0.003	-0.048	-0.213
	(0.028)	(0.064)	(0.050)	(0.079)	(0.143)
PostGQ × High BGS	-0.038***	-0.130**	-0.102**	-0.122*	-0.327**
	(0.012)	(0.063)	(0.042)	(0.067)	(0.145)
PostGQ × Listed$_i$		0.097			0.153**
		(0.065)			(0.062)
PostGQ × TFP$_i$			0.005		0.014
			(0.012)		(0.013)
PostGQ × Firm age$_i$				0.033	0.018
				(0.024)	(0.034)
Fixed effects:					
Firm	Yes	Yes	Yes	Yes	Yes
High BGS × industry × year	Yes	Yes	Yes	Yes	Yes
High BGS × state × year	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.70	0.65	0.62	0.65	0.62
Observations	15,842	7,491	5,274	7,491	5,274
Sample: Standalone firms	Yes	Yes	Yes	Yes	Yes

Appendix Table IX. Robustness: High Business Group Share definition

This table reports robustness for the business group share measure we use in our estimations. Column 1 reproduces the coefficients from baseline estimations in Table III. Note that in the baseline estimations *High BGS* is defined as an indicator variable set to one if the share of assets of group-affiliated firms from that city is in the top quartile in the year before the GQ road network upgrades. Column 2 repeats the same estimation, however, without the High BGS interaction with Industry × year and State × year fixed effects. Column 3 presents the interaction with the continuous measure while column 4 defines *High BGS* using Hirschman Herfindahl Index (HHI) based on group-affiliated firms' sales at each location. Column 5 presents outlines the quartile specification and lastly, column 6 presents the interaction with terciles of business group share. *PostGQ* is an indicator variable taking value 1 for all years including and after the GQ upgrade in the city. Standard errors are corrected for heteroscedasticity and autocorrelation and clustered at the city level. Standard errors are reported in parentheses.*,**, and *** indicate significance at 10%, 5%, and 1% respectively. All variables are defined in Appendix Table XIV. Data source: CMIE Prowess.

Dependent variable	Investment					
	Baseline	No interaction with High BGS	Continuous	HHI	Quartile	Tercile
	(1)	(2)	(3)	(4)	(5)	(6)
PostGQ	0.039	0.022	2.719	0.017		
	(0.028)	(0.030)	(2.803)	(0.025)		
PostGQ × High BGS	-0.038***	-0.036**				
	(0.012)	(0.015)				
PostGQ × BGS (continuous)			-0.931**			
			(0.378)			
PostGQ × High BGS (HHI)				-0.072**		
				(0.030)		
PostGQ × Quartile$_1$					-0.036	
					(0.041)	
PostGQ × Quartile$_2$					0.177***	
					(0.041)	
PostGQ × Quartile$_3$					-0.008	
					(0.029)	
PostGQ × Quartile$_4$					-0.036**	
					(0.029)	
PostGQ × Tercile$_1$						-0.025
						(0.033)
PostGQ × Tercile$_2$						0.079***
						(0.026)
PostGQ × Tercile$_3$						-0.026*
						(0.014)
Fixed effects:						
Firm	Yes	Yes	Yes	Yes	Yes	Yes
Industry × year	Yes	Yes	Yes	Yes	Yes	Yes
State × year	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted-R^2	0.70	0.70	0.70	0.71	0.70	0.70
Observations	15,842	15,842	15,842	13,097	15,660	15,842
Sample: Standalone firms	Yes	Yes	Yes	Yes	Yes	Yes

Appendix Table X. Robustness: Drop exiting firms

This table examines the robustness to dropping firms that exit the sample. Column 1 repeats the baseline estimation (Table III) while column 2 repeats the estimation on the subsample of firms exiting the sample if they stop filing annual reports or have been legally struck-off from the business register, identified using data from the MCA. *PostGQ* is an indicator variable taking value 1 for all years including and after the GQ upgrade in the city. *High BGS* is an indicator variable set to one if the share of assets of group-affiliated firms from that city is in the top quartile in the year before the GQ road network upgrades. All regressions include firm fixed effects, High BGS × industry × year fixed effects, and High BGS × state × year. Standard errors are corrected for heteroscedasticity and autocorrelation and clustered at the city level. Standard errors are reported in parentheses.*,**, and *** indicate significance at 10%, 5%, and 1% respectively. All variables are defined in Appendix Table XIV. Data source: CMIE Prowess and Ministry of Corporate Affairs.

Dependent variable	Investment	
	Baseline	Drop exiting firms
	(1)	(2)
PostGQ	0.039	0.039
	(0.028)	(0.028)
PostGQ × High BGS	-0.033**	-0.027**
	(0.012)	(0.012)
Fixed effects:		
Firm	Yes	Yes
High BGS × industry × year	Yes	Yes
High BGS × state × year	Yes	Yes
Adjusted-R^2	0.70	0.70
Observations	15,827	15,718
Sample : Standalone firms	Yes	Yes

Appendix Table XI. Robustness: Drop firms with extreme negative sales growth

This table examines the robustness to dropping firms that experience extreme negative sales growth. Column 1 repeats the baseline estimation (Table III) while other columns drop firms in the right tail of the sales growth distribution (10%ile in column 2, 5%ile in column 3, and 1% ile in column 4, respectively). *PostGQ* is an indicator variable taking value 1 for all years including and after the GQ upgrade in the city. *High BGS* is an indicator variable set to one if the share of assets of group-affiliated firms from that city is in the top quartile in the year before the GQ road network upgrades. All regressions include firm fixed effects, High BGS × industry × year fixed effects, and High BGS × state × year. Standard errors are corrected for heteroscedasticity and autocorrelation and clustered at the city level. Standard errors are reported in parentheses.*,**, and *** indicate significance at 10%, 5%, and 1% respectively. All variables are defined in Appendix Table XIV. Data source: CMIE Prowess.

Dependent variable	Investment			
	Baseline	Bottom 10 %ile	Bottom 5 %ile	Bottom 1 %ile
	(1)	(2)	(3)	(4)
PostGQ	0.039	0.029	0.029	0.041
	(0.028)	(0.031)	(0.032)	(0.032)
PostGQ × High BGS	-0.038***	-0.105***	-0.099***	-0.039**
	(0.012)	(0.030)	(0.019)	(0.016)
Fixed effects:				
Firm	Yes	Yes	Yes	Yes
High BGS × industry × year	Yes	Yes	Yes	Yes
High BGS × state × year	Yes	Yes	Yes	Yes
Adjusted-R^2	0.70	0.75	0.73	0.71
Observations	15,842	9,774	12,892	15,210
Sample : Standalone firms	Yes	Yes	Yes	Yes

Appendix Table XII. Mergers and acquisitions of standalone firms

This table changes in merger and acquisition activity whereby standalone firms are target, as function of business group prevalence around GQ upgrades. Column 1 focuses on all years in the sample while column 2 focuses on acquisitions on or after the financial year 2000 due to limited data on transactions in the prior period. *PostGQ* is an indicator variable taking value 1 for all years including and after the GQ upgrade in the city. *High BGS* is an indicator variable set to one if the share of assets of group-affiliated firms from that city is in the top quartile in the year before the GQ road network upgrades. All regressions include firm fixed effects, High BGS × industry × year fixed effects, and High BGS × state × year. Standard errors are corrected for heteroscedasticity and autocorrelation and clustered at the city level. Standard errors are reported in parentheses.*,**, and *** indicate significance at 10%, 5%, and 1% respectively. All variables are defined in Appendix Table XIV. Data source: CMIE Prowess.

Dependent variable	Target_{SA}	
	Full sample	≥2000
	(1)	(2)
PostGQ	0.041	0.051
	(0.038)	(0.045)
PostGQ × High BGS	-0.009	-0.020
	(0.039)	(0.046)
Fixed effects:		
Firm	Yes	Yes
High BGS × industry × year	Yes	Yes
High BGS × state × year	Yes	Yes
Adjusted-R^2	0.09	0.09
Observations	14,970	14,795
Sample : Standalone firms	Yes	Yes

Appendix Table XIII. Standalone entry at regional-level around GQ upgrades

This table reports examines entry by standalone firms around GQ road network upgrades as a function of business group share. The dependent variable in column 1 is the natural logarithm of new business incorporations at the city-level each year while the dependent variable in column 2 is the inverse hyperbolic sine of new business incorporations at the city-level each year. *PostGQ* is an indicator variable taking value 1 for all years including and after the GQ upgrade in the city. *High BGS* is an indicator variable set to one if the share of assets of group-affiliated firms from that city is in the top quartile in the year before the GQ road network upgrades. All regressions include city fixed effects, High BGS × industry × year fixed effects, and High BGS × state × year. Standard errors are corrected for heteroscedasticity and autocorrelation and clustered at the city level. Standard errors are reported in parentheses.*,**, and *** indicate significance at 10%, 5%, and 1% respectively. All variables are defined in Appendix Table XIV. Data source: CMIE Prowess.

Dependent variable	Log(incorporations)	IHS(incorporations)
	(1)	(2)
PostGQ	0.475**	0.397**
	(0.171)	(0.164)
PostGQ × High BGS	0.273	0.131
	(0.290)	(0.270)
Fixed effects:		
City	Yes	Yes
High BGS × industry × year	Yes	Yes
High BGS × state × year	Yes	Yes
Adjusted-R^2	0.10	0.10
Observations	967	967

Appendix Table XIV. Variable definitions

Variable	Definition	Data source
A. Firm characteristics		
Firm age	Firm i's age since incorporation.	CMIE Prowess
Cash flow	Ratio of cash flow from operations relative to book value of assets.	CMIE Prowess
Days sales of inventory	Ratio of ending inventory to cost of good solds multiplied by 365.	
Debt	Total outstanding debt from bank and financial institutions relative to book value of assets.	CMIE Prowess
High BGS	Indicator variable set to one if the share of group-affiliated firms' assets from that city is in the top quartile in the year before the GQ road network upgrades.	CMIE Prowess
Investment	Ratio of net fixed assets relative to book value of assets.	CMIE Prowess
Listed	Indicator variable set to 1 if the firm is listed on either the National Stock Exchange (NSE) or the Bombay Stock Exchange (BSE) by the financial year.	CMIE Prowess
Profitability	Earnings before interest, depreciation, taxes, and amortization relative to book value of assets.	CMIE Prowess
Return on assets	Profit after tax relative to book value of assets.	CMIE Prowess
Sales growth	Measured as the annual growth rate of sales.	CMIE Prowess
Size	Measured as the log of book value of assets.	CMIE Prowess
Total factor productivity	Estimation methodology as in Levinsohn and Petrin (2003). Details in Appendix D.	CMIE Prowess
B. Regional characteristics		
Average time to fill skilled worker (manager) vacancy	Average time in weeks to fill vacancy of a manager or a technician.	World Bank Enterprise Survey
Bank branches	Total number of bank branches scaled by the city population as recorded in the Population Census of 2001.	Reserve Bank of India
Bad roadways	Indicator set to 1 if the firm gives a rating of 1 or 2 (1 being worse and 10 being excellent) on the availability of road transport.	World Bank Enterprise Survey
Fraction of listed firms	Share of firms from that city that are listed on either the National Stock Exchange (NSE) or the Bombay Stock Exchange (BSE) by the year before the GQ road network upgrades.	CMIE Prowess
Labor constraint in contracting	Indicator set to 1 if the firm reports constraints in contracting labor.	World Bank Enterprise Survey
Loan amount	The natural logarithm of loan amount in Rs. million.	Ministry of Corporate Affairs
Obstacle to growth (transport)	Indicator set to 1 if the firm gives a rating of 2 ("Moderate obstacle"), 3 ("Major obstacle"), or 4 ("Very severe obstacle") on whether transportation is a problem for the operation and growth of their business.	World Bank Enterprise Survey
Obstacle to growth (labor)	Indicator set to 1 if the firm gives a rating of 2 ("Moderate obstacle"), 3 ("Major obstacle"), or 4 ("Very severe obstacle") on whether availability of skilled and educated Workers is a problem for the operation and growth of their business..	World Bank Enterprise Survey

Appendix D Total factor productivity: Estimation procedure

This section outlines the variables and their definitions that we use for estimating total factor productivity. All numbers are deflated using industry deflators to reflect real values. In estimating firm-level total factor productivity, we include firm size as a control variable.

Output: Value of total sales that includes income earned by the company from the sale of industrial goods as well as their raw materials, byproducts, stores and waste.

Capital: Gross fixed assets of a firm that includes both tangible assets, such as land, building, plant, and machinery, and intangible assets, such as goodwill assets, software, etc.

Labor: Compensation to employees that includes all cash and payments in kind made by a company to its employees.

Intermediate inputs: Combined value of raw materials, power and fuel consumptions. Raw materials are the sum of expenses on raw materials, stores, spares and tools used up by firms in the production process. Power and fuel include expenses made by the firms on power, fuel and water. The sum of these three variables is used as the proxy in the estimation of the production function.