Online Appendix (Not for Publication)

A1. The Determination of M

Let $r_d(\varphi) = p_N(\varphi)q_N(\varphi)$ and $r_{ROW}(\varphi) = p_{ROW}q_{ROW}$ denote firm's revenue from selling locally and non-locally. Plug in the optimal pricing rule $p_N = \frac{\sigma}{\sigma-1}\frac{w}{\varphi}$, the profit $\pi_d(\varphi)$ can be rewritten as

$$\pi_d(\varphi) = p_N(\varphi)q_N(\varphi) - \frac{w}{\varphi}q_N(\varphi) = \frac{1}{\sigma-1}\frac{w}{\varphi}q_N(\varphi) = \frac{1}{\sigma}r_d(\varphi)$$

Similarly, we find $\pi_{ROW} = \frac{1}{\sigma}r_{ROW}(\varphi)$. And the zero cutoff profit conditions in Eq.(8) give

$$\pi_d(\varphi^*) = \frac{1}{\sigma}r_d(\varphi^*) = wf \qquad \pi_{ROW}(\varphi^*_{ROW}) = \frac{1}{\sigma}r_{ROW}(\varphi^*_{ROW}) = wf_{ROW} \tag{A1}$$

Plug the demands in Eq.(3) and (4) into revenue functions, we obtain

$$r_d(\varphi) = p_N(\varphi)^{1-\sigma}\frac{X}{P^{\sigma+1}} = \varphi^{\sigma-1}\left(\frac{\sigma w}{\sigma-1}\right)^{1-\sigma}\frac{X}{P^{\sigma+1}} \tag{A2}$$

$$r_{ROW}(\varphi) = Bp_{ROW}(\varphi)^{1-\sigma} = \tau^{1-\sigma}\varphi^{\sigma-1}\left(\frac{\sigma w}{\sigma-1}\right)^{1-\sigma}B \tag{A3}$$

Let $\tilde{\varphi}(\varphi)$ and $\tilde{\varphi}_{ROW}(\varphi)$ denote the average productivity of successful firms and exporting firms, respectively. Given the demands in equations (3) and (4), we have

$$\frac{r_d(\tilde{\varphi}(\varphi))}{r_d(\varphi^*)} = \left(\frac{\tilde{\varphi}(\varphi)}{\varphi^*}\right)^{\sigma-1} \qquad \frac{r_{ROW}\tilde{\varphi}_{ROW}(\varphi))}{r_{ROW}(\varphi^*_{ROW})} = \left(\frac{\tilde{\varphi}_{ROW}(\varphi)}{\varphi^*_{ROW}}\right)^{\sigma-1} \tag{A4}$$

With the Pareto distribution, the ratio of the average and cutoff productivity is a constant:

$$\tilde{\varphi} = \left(\frac{\kappa}{\kappa-(\sigma-1)}\right)^{\frac{1}{\sigma-1}}\varphi^* \tag{A5}$$

so the average revenue of a firm that produces is given by

$$\begin{aligned}\bar{r} &= r_d(\tilde{\varphi}(\varphi)) + Prob_{ROW} r_{ROW}(\tilde{\varphi}_{ROW}(\varphi)) = \sigma\left(\bar{\pi} + wf + Prob_{ROW} w f_{ROW}\right) \\ &= \frac{\sigma\kappa w}{\kappa - (\sigma - 1)}(f + Prob_{ROW} f_{ROW})\end{aligned}$$

Use equations (A2) and (A3),

$$\frac{r_{ROW}(\varphi^*_{ROW})}{r_d(\varphi^*)} = \tau^{1-\sigma}\left(\frac{\varphi^*_{ROW}}{\varphi^*}\right)^{\sigma-1}\frac{BP^{\sigma+1}}{X} = \frac{f_{ROW}}{f}$$

so

$$Prob_{ROW} = \left(\frac{\varphi^*_{ROW}}{\varphi^*}\right)^{-\kappa} = \tau^{-\kappa}\left(\frac{X}{BP^{\sigma+1}}\frac{f_{ROW}}{f}\right)^{-\frac{\kappa}{\sigma-1}}$$

As in Melitz (2003), the aggregate revenue R equals to the wage times the size of the labor in the non-agriculture sector: $R = wL_N$. The mass of firms therefore can be pin down by:

$$M = \frac{R}{\bar{r}} = \frac{wL_N}{\frac{\sigma\kappa w}{\kappa-(\sigma-1)}(f + Prob_{ROW} f_{ROW})} = \frac{(\kappa - (\sigma-1))L_N}{\sigma\kappa\left(f + \tau^{-\kappa}(\frac{X}{BP^{\sigma+1}}\frac{f_{ROW}}{f})^{-\frac{\kappa}{\sigma-1}} f_{ROW}\right)} \tag{A6}$$

A2. Proof of Proposition 1

Take derivative of M w.r.t. τ,

$$\begin{aligned}\frac{\partial M}{\partial \tau} = &\frac{(\kappa - (\sigma-1))\bar{L}}{\sigma\kappa(f + \tau^{-\kappa}(\frac{X}{BP^{\sigma+1}}\frac{f_{ROW}}{f})^{-\frac{\kappa}{\sigma-1}} f_{ROW})^2} \\ &\left\{-\frac{\gamma\tau^{-\gamma}\frac{\bar{a}}{T}(\frac{1}{p_{A_{ROW}}})^{\gamma-1}}{\left(1 + (\frac{1}{\tau}\frac{1}{p_{A_{ROW}}})^{\gamma-1}\right)^{\frac{2\gamma-1}{\gamma-1}}}\left(f + \tau^{-\kappa}\left(\frac{f_{ROW}}{f}\frac{X}{BP^{\sigma-1}}\right)^{-\frac{\kappa}{\sigma-1}} f_{ROW}\right)\right. \\ &\left. + \kappa\tau^{-\kappa-1}\left(\frac{f_{ROW}}{f}\frac{X}{BP^{\sigma-1}}\right)^{-\frac{\kappa}{\sigma-1}} f_{ROW}\left(1 - \frac{\bar{a}}{T\left(1 + (\frac{1}{\tau}\frac{1}{p_{A_{ROW}}})^{\gamma-1}\right)^{\frac{\gamma}{\gamma-1}}}\right)\right\}\end{aligned} \tag{A7}$$

Assertion 1: This follows trivially from a derivative of equation (9) with respect to either argument, and the fact that that $L_N = \bar{L} - L_A$.

Assertion 2: Fix L_N, then the derivative from equation (A6) yields the above expression without the term in the second row. The terms in the first row is positive, and the term in the third row is positive given assumption (1), so the number of firms decreases as trade costs fall.

Assertion 3: Inspecting equation (A6), $\partial M/\partial L_N > 0$ trivially since $\kappa > \sigma - 1$ given the Chaney (2008) assumption, i.e., the necessary assumption for $M > 0$ in the same equation.

Assertion 4: When $T \to \underline{T}$, the term in the third row of the above expression is zero and $\frac{\partial M}{\partial \tau} < 0$. As $T \to \infty$ the first term is zero and $\frac{\partial M}{\partial \tau} > 0$. In addition, the derivative of $\frac{\partial M}{\partial \tau}$ with respect T:

$$\frac{\partial^2 M}{\partial \tau \partial T} \propto \frac{1}{T^2}\left(\frac{\gamma \bar{a} \tau^{-\gamma} \left(\frac{1}{p_{A_{ROW}}}\right)^{\gamma-1}}{\left(1+\left(\frac{1}{\tau}\frac{1}{p_{A_{ROW}}}\right)^{\gamma-1}\right)^{\frac{2\gamma-1}{\gamma-1}}} \left(f + \tau^{-\kappa}\left(\frac{f_{ROW}}{f}\frac{X}{BP^{\sigma-1}}\right)^{-\frac{\kappa}{\sigma-1}} f_{ROW}\right)\right.$$
$$\left. + \frac{\kappa \bar{a} \tau^{-\kappa-1}\left(\frac{f_{ROW}}{f}\frac{X}{BP^{\sigma-1}}\right)^{-\frac{\kappa}{\sigma-1}} f_{ROW}}{T\left(1+(\frac{1}{\tau}\frac{1}{p_{A_{ROW}}})^{\gamma-1}\right)^{\frac{\gamma}{\gamma-1}}}\right) > 0$$

that is, $\frac{\partial M}{\partial \tau}$ increases monotonically with T. Given the limiting endpoints and the continuity of $\frac{\partial M}{\partial \tau}$, the existence of a crossing T^* (and corresponding agricultural employment) comes from the Intermediate Value Theorem, while its uniqueness comes from the monotonicity of $\frac{\partial M}{\partial \tau}$. Therefore, $\frac{\partial M}{\partial \tau} < 0$ when $T < T^*$ and $\frac{\partial M}{\partial \tau} > 0$ when $T > T^*$.

A3. The Profit-Wage Ratio

The average profit of a firm is given by

$$\bar{\pi} = (\pi_d(\tilde{\varphi}) - wf) + Prob_{ROW}\left(\pi_{ROW}(\tilde{\varphi}_{ROW}) - wf_{ROW}\right)$$

Combine with Eq. (A1), (A4), and (A5),

$$\begin{aligned}
\bar{\pi} &= \left(\frac{1}{\sigma} r_d(\tilde{\varphi}) - wf\right) + Prob_{ROW}\left(\frac{1}{\sigma} r_{ROW}(\tilde{\varphi}_{ROW}) - wf_{ROW}\right) \\
&= \left(\left(\frac{\tilde{\varphi}}{\varphi^*}\right)^{\sigma-1} - 1\right) wf + Prob_{ROW}\left(\left(\frac{\tilde{\varphi}_{ROW}}{\varphi^*_{ROW}}\right)^{\sigma-1} - 1\right) wf_{ROW} \\
&= \frac{w(\sigma-1)}{\kappa - (\sigma-1)}\left(f + Prob_{ROW} f_{ROW}\right)
\end{aligned}$$

Plug in Eq. (A6), the total profit of all firms equals to

$$\Pi = M\bar{\pi} = \frac{w(\sigma-1)}{\sigma\kappa} L_N$$

So the ratio between profit and wage is

$$\frac{\Pi}{w\bar{L}} = \frac{\sigma - 1}{\sigma\kappa}\frac{L_N}{\bar{L}} = \frac{\sigma - 1}{\sigma\kappa}\left(1 - \frac{\frac{\bar{a}}{T}}{\left(1 + (\frac{1}{\tau}\frac{1}{p_{A_{ROW}}})^{\gamma-1}\right)^{\frac{\gamma}{\gamma-1}}}\right)$$

Take derivative w.r.t. τ,

$$d\frac{\Pi}{w\bar{L}}/d\tau = -\frac{1}{\tau}(\gamma - 1)\frac{\sigma - 1}{\sigma\kappa}\frac{\frac{\bar{a}}{T}(\frac{1}{\tau}\frac{1}{p_{A_{ROW}}})^{\gamma-1}}{\left(1 + (\frac{1}{\tau}\frac{1}{p_{A_{ROW}}})^{\gamma-1}\right)^{\frac{2\gamma}{\gamma-1}}} < 0$$

B. Additional Figures

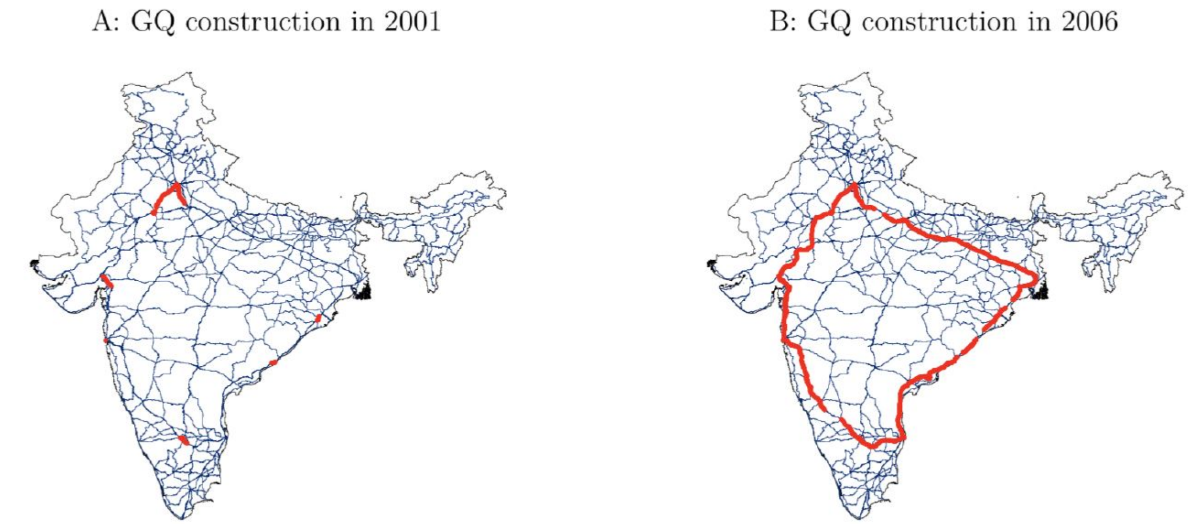

Figure B.1: The Golden Quadrilateral: Asturias, Garcìa-Santana, and Ramos (2019)

(a) NTHS construction in 2000

(b) NTHS construction in 2011

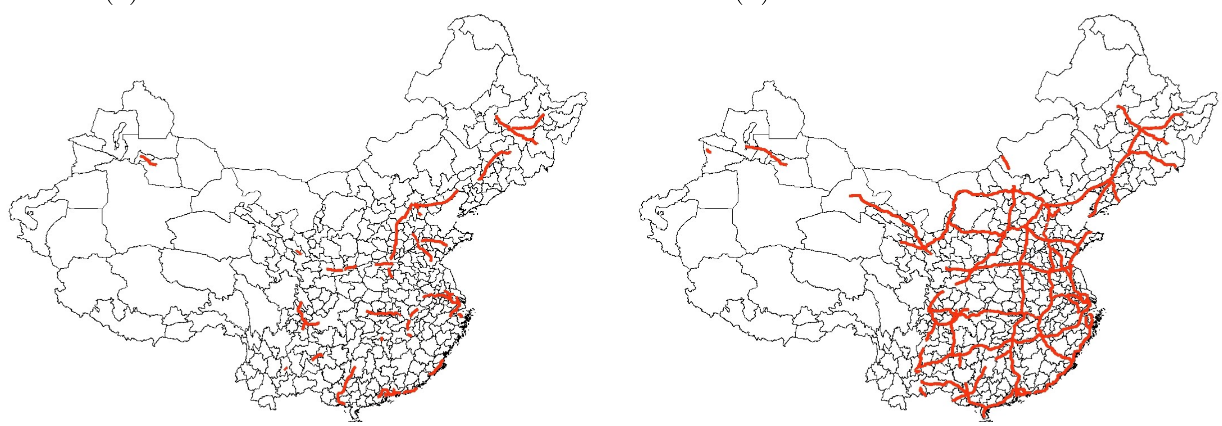

Figure B.2: China's NTHS Completion: Lu (2023)

C. Heterogeneous Exit

In this subsection, we show that the reduction in the number of establishments in India after highway connection is driven by small firms. Table H.5 shows the effect of highway connection on the number of establishments by establishment size. The dependent variable is the log of the number of establishments in each size category in each location at a given year. An establishment is classified as small if the number of employees is less than 25, and an establishment is considered a large establishment if the number of employees is greater than or equal to 500. We find that the number of small establishments decreased and the number of large establishments increased in connected areas relative to unconnected areas after highway connection.

Table C.1: Effect of Highway on the Number of Establishments by Firm Size

	(1) Small	(2) Median	(3) Large	(4) Small	(5) Median	(6) Large
$\chi_{\text{tercile}=1,2} {}^{*} \chi_{t>t_c}$	-0.179***	-0.173***	0.096	-0.169***	-0.111**	0.114*
	(0.064)	(0.066)	(0.074)	(0.058)	(0.055)	(0.064)
log(output)				0.196***	0.388***	0.347***
				(0.021)	(0.027)	(0.035)
log(population)				0.009	0.135*	0.079
				(0.072)	(0.080)	(0.063)
Location FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,153	5,293	3,653	5,104	5,241	3,631
First stage F-stat	749.6	821.3	448.6	756.5	832	457.2

Note: Standard errors clustered at the district level are in parentheses. *** significant at 1%. ** significant at 5%. * significant at 10%.

D. OLS results for China

In this section, we present the OLS regression results for specification 12 for China's NTHS. Table D.1 presents the effect of highway on the number of manufacturing establishments China. Column 1 shows that relative to the unconnected districts in the third tercile, connected districts in the first and second tercile experienced over 6.4% increase in the number of establishments after connection. The estimated effect of connecting to NTHS stays positive and statistically significant with the inclusion of additional controls.

Table D.1: Impact of Highway on the Number of Establishments

Dependent variable = log(number of establishments)			
	(1)	(2)	(3)
$\chi_{\text{tercile}=1,2}{}^{*}\chi_{t>t_c}$	0.119***	0.122***	0.122***
	(0.0378)	(0.0367)	(0.0365)
log(output)		0.418***	0.428***
		(0.0888)	(0.0866)
log(population)			-0.482
			(0.437)
Location FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	2,925	2,925	2,925
Adjusted R-squared	0.943	0.946	0.946

Note: Standard errors clustered at the district level are in parentheses. *** significant at 1%. ** significant at 5%. * significant at 10%.

The results in Table D.2 illustrate the impact of highways on the number of employees in an establishment. Opposite to what was found with the IV regressions, the OLS regression indicates that the impact of highway on average employment in China is positive and significant, the two first terciles increased the average number of employees by about 4%. It suggests the actual highway was placed where employment is expected to grow.

Table D.2: Impact of Highway on the Average Number of Employees

Dependent variable = log(average number of employees)			
	(1)	(2)	(3)
$\chi_{\text{tercile}=1,2}{}^{*}\chi_{t>t_c}$	0.0441**	0.0444**	0.0440**
	(0.0182)	(0.0181)	(0.0180)
log(output)		0.0444	0.0357
		(0.0417)	(0.0442)
log(population)			0.411
			(0.283)
Location FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	2,919	2,919	2,919
Adjusted R-squared	0.961	0.961	0.961

Note: Standard errors clustered at the district level are in parentheses. *** significant at 1%. ** significant at 5%. * significant at 10%.

We also test the effect of highways on initial agricultural employment with the OLS regression and the results are presented in Table D.3. In China, agricultural employment decreases significantly, by roughly 11 log points in locations connected by the highway.

Table D.3: Impact of Highway on Agricultural Employment

Dependent variable = log(agricultural employment)			
	(1)	(2)	(3)
$\chi_{\text{tercile}=1,2}{}^{*}\chi_{t>t_c}$	-0.117*	-0.118*	-0.117*
	(0.0656)	(0.0660)	(0.0655)
log(output)		-0.109	-0.0925
		(0.219)	(0.215)
log(population)			-0.800
			(0.844)
Location FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	2,909	2,909	2,909
Adjusted R-squared	0.826	0.826	0.826

Note: Standard errors clustered at the district level are in parentheses. *** significant at 1%. ** significant at 5%. * significant at 10%.

Table D.4 shows the effect of highway on the number of establishments using terciles based on the share of people working in agriculture in 2000. Columns 1-3 show that in China, the number of establishment increases in cities where the initial agricultural employment is higher.

Table D.4: Number of Establishments and Agricultural Employment Share

Dependent variable = log(number of establishments)			
	(1)	(2)	(3)
	Low agr	Med agr	Large agr
$\chi_{\text{tercile}=1,2} {}^* \chi_{t>t_c}$	-0.00501	0.0282	0.207***
	(0.0635)	(0.0454)	(0.0596)
log(output)	0.516***	0.352***	0.465***
	(0.172)	(0.131)	(0.160)
log(population)	0.304	-0.996*	-2.352**
	(0.347)	(0.502)	(0.973)
Location FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	899	899	885
Adjusted R-squared	0.949	0.951	0.940

Note: Standard errors clustered at the district level are in parentheses. *** significant at 1%. ** significant at 5%. * significant at 10%.

E. Event study for China

This section shows the results from an event study analysis. Since the timing of the highway connection is varying and continuous, we use the imputation proposed by Borusyak et al. (2024) to address the potential of unrestricted treatment effect heterogeneity, Each graph shows the estimated coefficients for a given year relative to the connection year in the sample period and their 95% confidence intervals. The vertical line at 0 indicates the year in which at least 80% of NTHS in the province is completed. We omitted the year 0 in the regression, so the coefficient for the connection year shows up as zero in the graphs.

Figure E.1 shows the impact of NTHS on the number of establishments over time. We find no evidence of a pre-trend in the number of establishments before the NTHS is completed. The estimated coefficients become significantly negative after the connection year. The results suggest that the NTHS construction increased the number of establishments in the connected city-prefectures relative to unconnected city-prefectures. Since the connection status is continuous, we see an upward treatment effect.

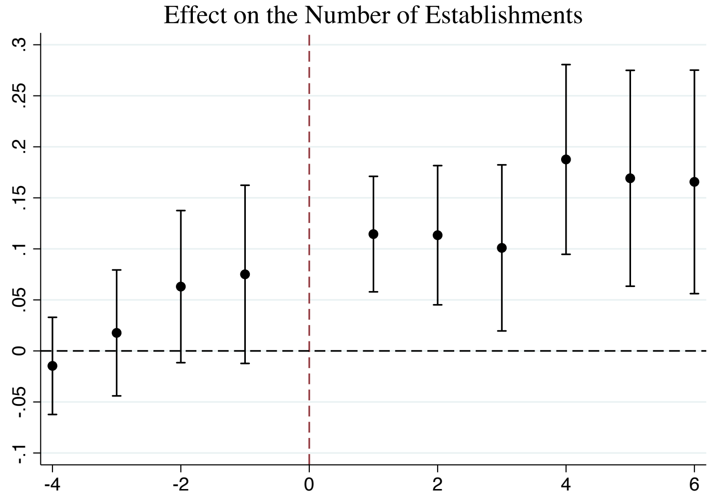

Figure E.1: Impact of highway proximity on the number of establishments over time. This graph plots the estimates of the effect of highway over time, along with their 95% confidence intervals. The vertical line at 0 indicates the year in which at least 80% of NTHS in the province is completed.

Figure E.2 shows the impact of NTHS on the average employment of of establishments over time. Prior to year 0, the estimated coefficients are not statistically different from zero. There is admittedly an upward pretrend, but this is not surprising since the road construction is continuous and only the threshold of 80% completion gives us a discrete treatment of "connection". The estimated coefficients become positive after connection, but it is significant in year 1 suggesting a brief positive impact of NTHS on average employment in connected areas.

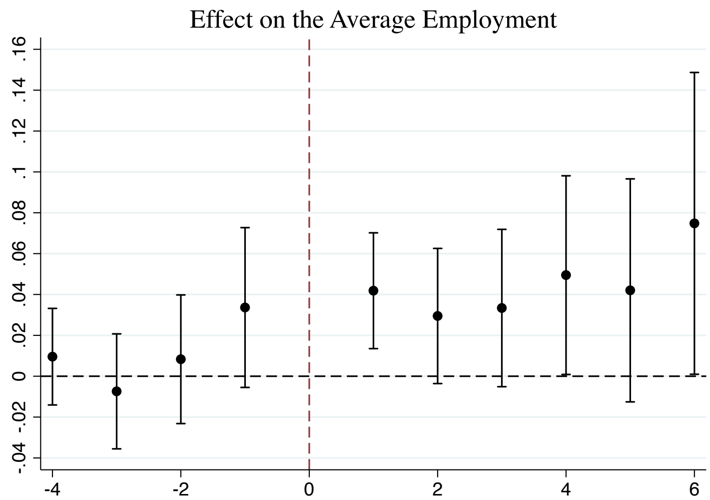

Figure E.2: Impact of Highway proximity on the average employees over time. This graph plots the estimates of the effect of highway over time, along with their 95% confidence intervals. The vertical line at 0 indicates the year in which at least 80% of NTHS in the province is completed.

Figure E.3 shows the impact of NTHS on the agricultural employment in the city-prefecture. The estimated coefficients over time are negative but not statistically significant.

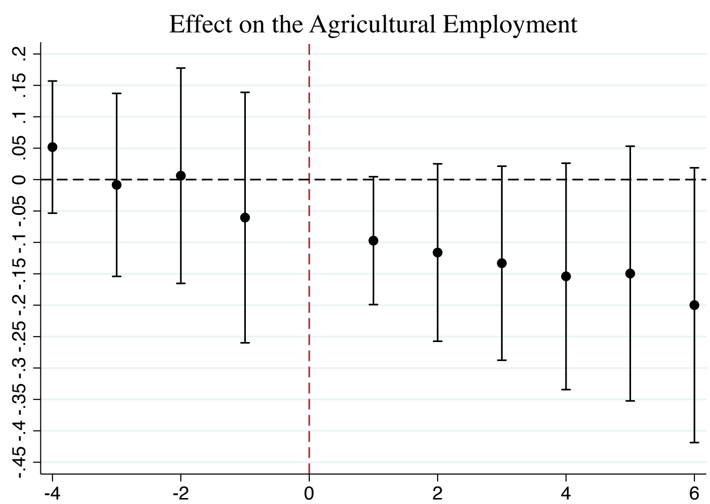

Figure E.3: Impact of Highway proximity on the agricultural employment over time. This graph plots the estimates of the effect of highway over time, along with their 95% confidence intervals. The vertical line at 0 indicates the year in which at least 80% of NTHS in the province is completed.

F. OLS results for India

In this section, we present the OLS regression results for specification 12 for India's GQ. Table F.1 shows the effect of GQ construction on the number of establishments in each district. The results are qualitatively the same as the results using an IV specification. The connected districts saw a significant decline in the number of establishments relative to the unconnected districts after GQ construction.

Table F.1: Impact of Highway on the Number of Establishments

Dependent variable = log(number of establishments)			
	(1)	(2)	(3)
$\chi_{\text{tercile}=1,2}{}^{*}\chi_{t>t_c}$	-0.095*	-0.088**	-0.084**
	(0.050)	(0.041)	(0.041)
log(output)		0.298***	0.300***
		(0.019)	(0.019)
log(population)			0.022
			(0.061)
Location FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	5,676	5,676	5,611
Adjusted Adjusted R-squared	0.909	0.928	0.929

Note: Standard errors clustered at the district level are in parentheses. *** significant at 1%. ** significant at 5%. * significant at 10%.

Table F.2 shows the impact of GQ on the average number of employees per establishment in a district. The estimated coefficient on the interaction term is positive, suggesting that the number of employees increased in connected districts relative to unconnected districts after GQ construction.

Table F.2: Impact of Highway on the Average Number of Employees

Dependent variable = log(average number of employees)			
	(1)	(2)	(3)
$\chi_{\text{tercile}=1,2}{}^{*}\chi_{t>t_c}$	0.057	0.062*	0.064*
	(0.040)	(0.037)	(0.038)
log(output)		0.202***	0.199**
		(0.018)	(0.020)
log(population)			0.060
			(0.051)
Location FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	5,669	5,669	5,605
Adjusted Adjusted R-squared	0.729	0.761	0.762

Note: Standard errors clustered at the district level are in parentheses. *** significant at 1%. ** significant at 5%. * significant at 10%.

Table F.3 shows the impact of GQ on the number of agricultural employees in the district. We find no significant change in the log of agricultural employment in the connective district relative to unconnected district after GQ construction.

Table F.3: Impact of Highway on Agricultural Employment

Dependent variable = log(agricultural employment)			
	(1)	(2)	(3)
$\chi_{\text{tercile}=1,2}{}^{*}\chi_{t>t_c}$	-0.016	-0.016	-0.006
	(0.037)	(0.037)	(0.029)
log(output)		-0.004	-0.005
		(0.008)	(0.006)
log(population)			0.905***
			(0.046)
Location FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	5,601	5,601	5,601
Adjusted Adjusted R-squared	0.915	0.915	0.934

Note: Standard errors clustered at the district level are in parentheses. *** significant at 1%. ** significant at 5%. * significant at 10%.

Table F.4 shows the effect of GQ on the number of establishments using terciles based on the share of people working in agriculture in 2000. The estimated coefficient on the interaction term is negative and statistically significant for the bottom tercile with low share of agricultural employment in 2000. The results suggest that the negative effect of GQ on the number of establishment is concentrated in the districts which have initially low agricultural employment.

Table F.4: Number of Establishments and Agricultural Employment Share

Dependent variable = log(number of establishments)			
	(1)	(2)	(3)
	Low agr	Med agr	Large agr
$\chi_{\text{tercile}=1,2}{}^{*}\chi_{t>t_c}$	-0.176***	-0.103	0.0244
	(0.063)	(0.068)	(0.082)
log(output)	0.305***	0.303***	0.294***
	(0.036)	(0.028)	(0.034)
log(population)	0.082	-0.057	0.051
	(0.099)	(0.116)	(0.099)
Location FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	1,918	1,855	1,838
Adjusted Adjusted R-squared	0.945	0.917	0.906

Note: Standard errors clustered at the district level are in parentheses. *** significant at 1%. ** significant at 5%. * significant at 10%.

G. Event study for India

This section shows the results from an event study analysis. Each graph shows the estimated coefficients for a given year in the sample period and their 95% confidence intervals. The vertical line at 2004 indicates the year in which at least 80% of GQ is completed. We omitted the year 2004 in the regression, so the coefficient for 2004 shows up as zero in the graphs. We also highlight the year 2006 using a different vertical line in which more than 90% of GQ is completed. In the literature, it is more common to use 2006 as the completion year for GQ.[11] In this paper, we choose 2004 as the completion cutoff so that the analysis of India becomes consistent with that of China. Our results are robust to changing the GQ completion year to 2006.

Figure G.2 shows the impact of GQ on the number of establishments over time. We find no evidence of a pre-trend in the number of establishments before the GQ is completed. The estimated coefficients become significantly negative after 2006. The results suggest that the GQ construction reduces the number of establishments in the connected districts relative to unconnected districts.

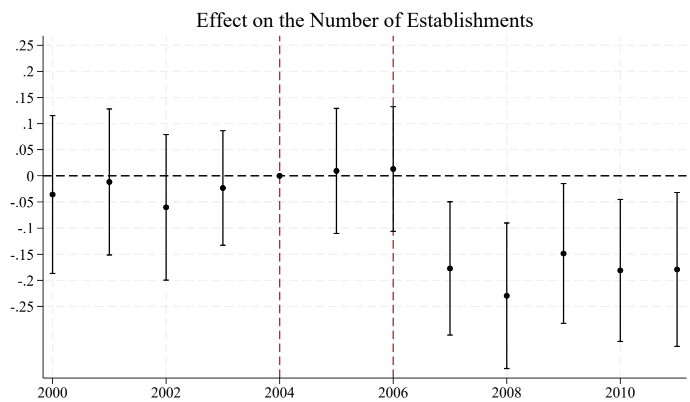

Figure G.1: Impact of highway proximity on the number of establishments over time. This graph plots the estimates of the effect of highway over time, along with their 95% confidence intervals. The vertical line at 2004(2006) indicates the year in which at least 80%(90%) of GQ is completed.

Figure G.2 shows the impact of GQ on the average employment of establishments over time. Prior to 2006, the estimated coefficients are not statistically different from zero. The estimated coefficients become significantly positive after 2006, suggesting a positive impact of GQ on average employment in connected areas.

[11] See Ghani et al. (2016), Asturias et al. (2019), and Brooks et al. (2021).

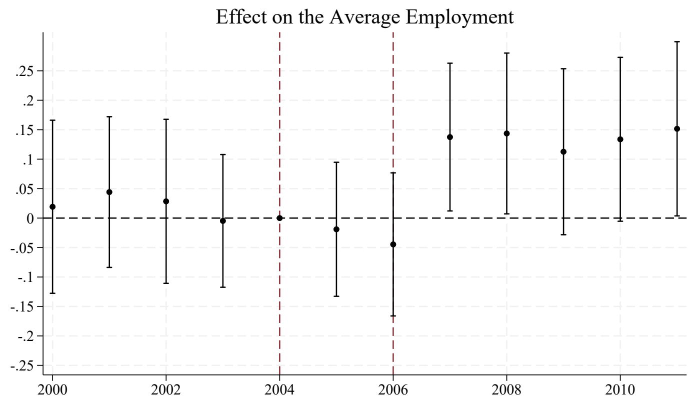

Figure G.2: Impact of Highway proximity on the average employees over time. This graph plots the estimates of the effect of highway over time, along with their 95% confidence intervals. The vertical line at 2004(2006) indicates the year in which at least 80%(90%) of GQ is completed.

Figure G.3 shows the impact of GQ on the agricultural employment in the district. The estimated coefficients over time are not statistically different from zero.

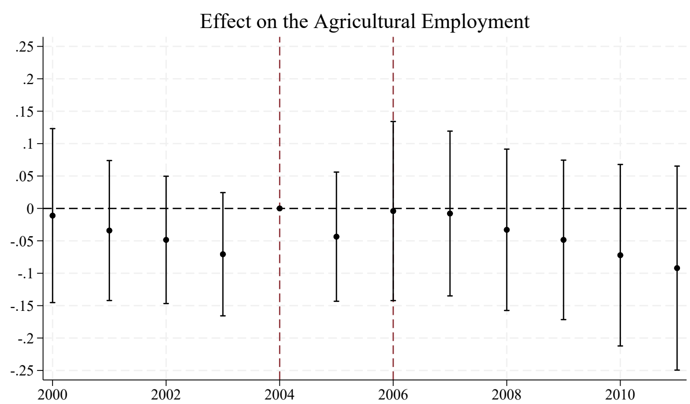

Figure G.3: Impact of Highway proximity on the agricultural employment over time. This graph plots the estimates of the effect of highway over time, along with their 95% confidence intervals. The vertical line at 2004(2006) indicates the year in which at least 80%(90%) of GQ is completed.

H. Robustness Checks for India

In this subsection, we present the results about the impact of GQ in India using a subset of firms that satisfy a comparable threshold as the one imposed in the Chinese data. The threshold is 5 million RMB, and we use the exchange rate 1Rupee = 0.18RMB to convert it to rupees.

The results are qualitatively the same as the baseline, indicating that the qualitative difference in results between India and China are not the result of left-censoring of small firms in the Chinese data. We continue to find a significant decline in the number of establishments in the connected districts after connection compared to unconnected districts. The connected districts did not experience a significant change in the agricultural employment after highway connection as compared to unconnected districts. The decline in the number of districts is concentrated in the districts that have high initial agricultural employment.

Table H.1: Impact of Highway on the Number of Establishments

	Dependent variable = log(number of establishments)		
	(1)	(2)	(3)
$\chi_{\text{tercile}=1,2} * \chi_{t>t_c}$	-0.152**	-0.132**	-0.129**
	(0.068)	(0.050)	(0.050)
log(output)		0.559***	0.559***
		(0.031)	(0.033)
log(population)			0.056
			(0.071)
location FE	Yes	Yes	Yes
year FE	Yes	Yes	Yes
Observations	5,144	5,144	5,096
		First stage	
$\chi_{\text{tercile}=1,2} * \chi_{t>t_c}$	0.847***	0.847***	0.848***
	(0.029)	(0.029)	(0.029)
log(output)		0.007	0.007
		(0.005)	(0.005)
log(population)			0.029
			(0.023)
location FE	Yes	Yes	Yes
year FE	Yes	Yes	Yes
Observations	5,144	5,144	5,096
Adj. R^2	0.834	0.835	0.834
First Stage F-Stat	867.7	871.3	878.2

Note: Standard errors clustered at the district level are in parentheses. *** significant at 1%. ** significant at 5%. * significant at 10%.

Table H.2: Impact of Highway on the Average Number of Employees

	Dependent variable = log(average number of employees)		
	(1)	(2)	(3)
$\chi_{\text{tercile}=1,2}{}^{*}\chi_{t>t_c}$	0.066	0.068	0.064
	(0.058)	(0.058)	(0.057)
log(output)		0.053	0.043
		(0.034)	(0.036)
log(population)			0.037
			(0.070)
location FE	Yes	Yes	Yes
year FE	Yes	Yes	Yes
Observations	5,138	5,138	5,091
	First stage		
$\chi_{\text{tercile}=1,2}{}^{*}\chi_{t>t_c}$	0.847***	0.847***	0.848***
	(0.029)	(0.029)	(0.029)
log(output)		0.007	0.007
		(0.005)	(0.005)
log(population)			0.029
			(0.023)
location FE	Yes	Yes	Yes
year FE	Yes	Yes	Yes
Observations	5,138	5,138	5,091
Adj. R^2	0.834	0.835	0.833
First Stage F-Stat	866.6	870.1	877.7

Note: Standard errors clustered at the district level are in parentheses. *** significant at 1%. ** significant at 5%. * significant at 10%.

Table H.3: Impact of Highway on Agricultural Employment

	Dependent variable = log(agricultural employment)		
	(1)	(2)	(3)
$\chi_{\text{tercile}=1,2}{}^{*}\chi_{t>t_c}$	-0.057	-0.057	-0.031
	(0.046)	(0.046)	(0.038)
log(output)		0.001	0.001
		(0.009)	(0.008)
log(population)			0.880***
			(0.049)
location FE	Yes	Yes	Yes
year FE	Yes	Yes	Yes
Observations	5,090	5,090	5,090
	First stage		
$\chi_{\text{tercile}=1,2}{}^{*}\chi_{t>t_c}$	0.847***	0.847***	0.848***
	(0.029)	(0.029)	(0.029)
log(output)		0.007	0.007
		(0.005)	(0.005)
log(population)			0.029
			(0.023)
location FE	Yes	Yes	Yes
year FE	Yes	Yes	Yes
Observations	5,090	5,090	5,090
Adj. R^2	0.832	0.833	0.833
First Stage F-Stat	856.7	861.9	870.3

Note: Standard errors clustered at the district level are in parentheses. *** significant at 1%. ** significant at 5%. * significant at 10%.

Table H.4: Number of Establishments and Agricultural Employment Share

	Dependent variable = log(number of establishments)		
	(1)	(2)	(3)
	Low agr	Med agr	Large agr
$\chi_{\text{tercile}=1,2}{}^{*}\chi_{t>t_c}$	-0.213***	-0.0839	-0.0790
	(0.070)	(0.108)	(0.090)
log(output)	0.654***	0.467***	0.583***
	(0.057)	(0.048)	(0.055)
log(population)	0.144	0.0396	-0.0225
	(0.092)	(0.140)	(0.117)
location FE	Yes	Yes	Yes
year FE	Yes	Yes	Yes
Observations	1,751	1,704	1,641
		First stage	
$\chi_{\text{tercile}=1,2}{}^{*}\chi_{t>t_c}$	0.916***	0.766***	0.841***
	(0.035)	(0.066)	(0.049)
log(output)	0.007	0.004	0.007
	(0.008)	(0.007)	(0.010)
log(population)	0.037	-0.008	0.060
	(0.038)	(0.035)	(0.055)
location FE	Yes	Yes	Yes
year FE	Yes	Yes	Yes
Observations	1,918	1,855	1,838
Adj. R^2	0.898	0.816	0.793
First Stage F-Stat	678.1	136.5	297.3

Note: Standard errors clustered at the district level are in parentheses. *** significant at 1%. ** significant at 5%. * significant at 10%.

In the following table, we test the effect of the GQ on the number of establishments using different size cutoffs. The dependent variable is the log of the number of establishments in each size category in each location at a given year. An establishment is classified as small if the number of employees is less than 25, and an establishment is considered a large establishment if the number of employees is greater than or equal to 500. The results suggest that the decline in the number of establishments in the connected districts after highway construction is driven primarily by the small firms.

Table H.5: Effect of Highway on the Number of Establishments by Firm Size

	(1) Small	(2) Median	(3) Large	(4) Small	(5) Median	(6) Large
$\chi_{\text{tercile}=1,2}{}^{*}\chi_{t>t_c}$	-0.201*	-0.141*	0.065	-0.165*	-0.118**	0.089
	(0.105)	(0.072)	(0.073)	(0.098)	(0.056)	(0.063)
log(output)				0.376***	0.533***	0.328***
				(0.048)	(0.034)	(0.034)
log(population)				0.096	0.067	0.065
				(0.095)	(0.077)	(0.063)
Location FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,962	4,887	3,623	2,943	4,844	3,601
First stage F-stat	401.3	793.8	448	410.9	800.3	457.5

Note: Standard errors clustered at the district level are in parentheses. *** significant at 1%. ** significant at 5%. * significant at 10%.