Infrastructure and Output: An Earthquake IV Analysis

Charles Feng, Dean Reuben, Aditya Mahajan

April 2025

1   Introduction

In David Alan Aschauer's 1990 paper, '*Why is Infrastructure Important?',* he argues that public investment in 'core' infrastructure drives productivity growth. Core infrastructure, defined as the 'public stock of social and economic overhead capital' (Aschauer, 1990, p. 22), such as utilities and transportation routes, drives growth through two vectors: first, it smooths transactions; goods can be shipped on predictable schedules, allowing for credible promise of payment and delivery with minimal transaction cost. Second, it reduces operating costs: a dollar spent repairing a key highway is several dollars a transportation company does not have to spend on repairing trucks damaged by poor roads or circumventing the roads altogether. David Aschauer (1990) argues that the US productivity slowdown of the 1970s can be largely attributed to a slowdown in public capital investment.

We test this hypothesis in the context of another large emerging economy: China. In recent years, productivity growth in the country has slowed despite record levels of infrastructure investment at all levels of society. We conducted an IV analysis of the impact of highway construction on output per worker with recorded damage-causing earthquakes as the instrument, exogenous highway construction as our independent variable, and output per worker as our dependent variable.

Our question is: What is the effect of *exogenous* core infrastructure growth on regional productivity, denominated by the output per worker, in the 31 provinces and special administrative cities of China from 2008 to 2021?

2   Independent Variable

Core infrastructure is defined as roads, bridges, power plants, or power networks. The myriad economic benefits of transport networks are well-studied: Donaldson (2009) famously established that British-built railroad networks in Colonial India lowered transportation costs and brought down inter-regional price differences. We believe we can extend this analysis one step further- that infrastructure is responsible for growth in productivity itself.
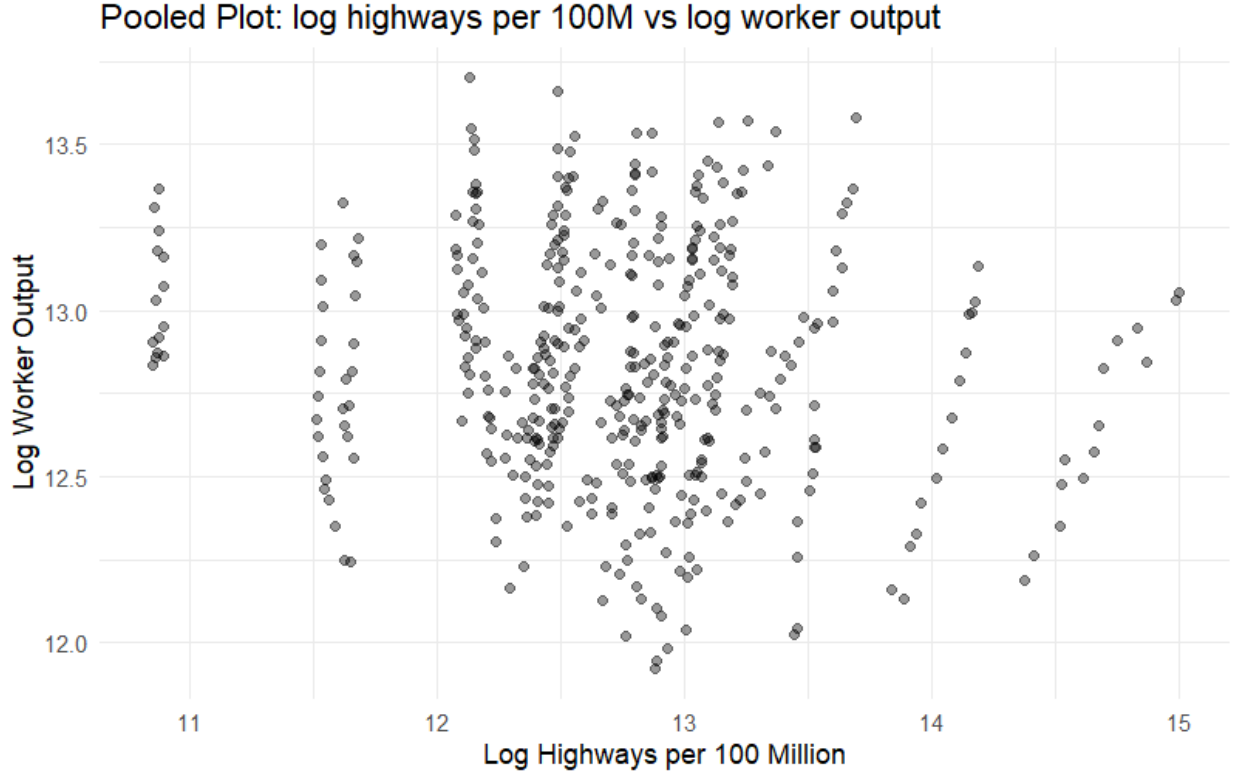
Figure 1: log Output vs. log Highways

We plotted natural logged Highways per 100 million against natural logged output per worker, observing that it's highly clustered by province. A slight positive relationship can be seen in some province clusters, while others show no clear relationship.

3 Controls

We control for the rival capital investments of private firms, which we define as the available equipment, facilities, and inventory for worker use, and the non-infrastructure components of productivity.

For private capital stock, We use the variable *Private, Non-current Assets per 100,000 Workers*, as Chinese national capital stock estimates do not distinguish between private capital and national infrastructure. This metric is derived from the balance sheets of fully private Chinese enterprises. We believe that this is an adequate representative measure of private capital stock, as non-current balance sheet assets have been used to proxy for capital stock in past corporate studies, famously such as McConnell and Servaes (1990) .

The Solow Residual, the macroeconomic measure of total factor productivity (TFP), has been studied extensively. We borrow the ideas of Romer (1990), who endogenized science and technology within the Solow Residual, as well as the work of Lucas (1988), who found evidence of an education dividend to productivity. To represent human capital, we introduce the variable *log(college enrollments per 100,000 working age people*. The Chinese statistics board defines working age as 25-64. We also introduce the variable of *log(patents per 100 million)* to model the trend and variation in scientific discovery across the provinces, relative to their total populations.

Table 1: Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
log(RMB Output per Worker)	n=31 t=14	12.83029	0.1833	12.4578	13.2383
log(km Highways per 100 million people)	n=31 t=14	12.7294	0.7118	10.8716	14.6778
log(RMB Non-current Assets per 100 thousand workers)	n=31 t=14	12.8189	0.8543	10.4276	14.0984
log(College enrollments per 100,000 people aged 25-65)	n=31 t=14	11.5937	0.4361	10.3215	12.8851
log(Patents per 100 million people)	n=31 t=14	11.9793	1.0296	10.0031	13.9658

Notes: units are adjusted such that all variables are at similar magnitudes. We take this measure to prevent unusually-sized coefficients from distorting our analysis

We display the summary statistics of the logged variables we use for our regression, observing that after the fitting and scaling measures we've taken, the magnitude and variances of each variable are sufficiently similar for analysis. The most volatile series is highways- this makes sense, as China's provinces vary greatly in size. We take note of this, and introduce individual fixed effects dummies to control for provincial characteristicsm, and time effects dummies to filter out wider unrelated Chinese trends.

3.1 Theory

We based our model of production on the simple Cobb-Douglas Production Model:

$$Y = AK^\alpha L^{1-\alpha}$$

Y is output, K, L are capital stock and labor supply, respectively, and A is total factor productivity, the variable we're interested in. Per worker, the formula is:

$$y = Ak^\alpha$$

Where y, k are output per worker and capital stock per worker, respectively. The key features of this model are the diminishing returns on capital per worker, $0 < \alpha < 1$, and the Solow Residual, representing the 'force multipliers' of an Economy, such as a strong scientific research regime, health, or human capital accumulation. The properties of logarithms enable us to depict this relationship in linear form:

$$log(y) = \alpha + \beta log(k)$$

To reflect the fact that infrastructure is a component of capital stock, we specify the model:

$$log(y) = \alpha + \beta(PrivateCapital + Infrastructure)$$

We define *PrivateCapital* as the stock of physical private, rival capital not available for use by the general populace, such as equipment, machinery, and private facilities. The reasoning for this model specification is our assumption that a comparable gain in private capital and infrastructure will result in a similar percentage increase in output per worker, regardless of the source of capital stock growth.

The above model reflects the *null* hypothesis, where infrastructure is a component of capital stock, and is a mere input, not a *force multiplier* of the economy, a component of the Solow Residual. In the alternative case that *Infrastructure* drives α, then the model is:

3

$$log(y) = \delta + \gamma log(Infrastructure) + \beta(log(PrivateCapital) + log(Infrastructure))$$
$$log(y) = \delta + \theta log(Infrastructure) + \beta(PrivateCapital)$$
$$\theta = \gamma + \beta > \beta$$

We focus our natural experiment on showing that there is a greater return on infrastructure investment than standard rival private capital. We can demonstrate this as such- in situations where the overall capital stock decreased, if we can identify a significant, large positive partial elasticity of output per worker w.r.t. infrastructure, we take this as evidence of a productivity dividend.

3.2   Experimental Design

We want to isolate the exogenous effects of infrastructure from the correlated effects of private capital, as well as other predictors of productivity, such as human capital and scientific research. We need a *natural experiment*- a special situation in which infrastructure *diverges* from these correlated productivity factors, allowing us to estimate an infrastructure effect free from potential positive biases.

We look to a unique characteristic of Chinese infrastructure- its rapid recovery following natural disasters. Chen (2022) finds that significant investments in infrastructure were instrumental in the economic recovery following the 2008 Wenchuang earthquake, making the important distinction between infrastructure recovery and economic recovery. While high investments in roadways characterized Sichuan's recovery, key indicators of economic health did not follow. Supporting this, Fayazi et al. (2019), meanwhile, found evidence of increased inequality among the Chinese rural poor.

We see an opportunity to exploit the positive predictive power of earthquakes towards infrastructure, as well as the divergence between infrastructure and other key economic metrics to obtain relatively unbiased estimators of the effects of infrastructure on output. We will run a natural experiment in the form of a two-stage instrumental variable regression. We seek to establish the relevance of the Earthquake instrument at predicting for expected increases in provincial infrastructure spending, as well as its independence of our controls. We will then show that this exogenous earthquake infrastructure has positive predictive power in regards to output per worker.

4   Experiment: Earthquake IV Analysis of effects of Infrastructure on Output per Worker

We estimate the probit model:

$$P(Earthquake_{it} = 1|i,t) = \Phi(\theta_0 + \theta_1 log(Highways_{it}) + \theta_2 log(assets_{it}) + \theta_3 controls_{it}) + \epsilon_{it}$$

This allows us to model the probability of a damage-causing earthquake in a province and year, as predicted the values of the independent predictors in that year. Individual and time effects are controlled for with fixed effects dummies. The significance levels of our coefficients are as follows:

Table 2: Probit CLT-enabled z-test Results

Variable	Coefficient	Std. Error	z-Value	Pr($> |z|$)
lHighway	3.08370	1.62934	1.8926	0.05974
lassets	-0.84773	0.50654	-1.6736	0.09566
lCollege	-0.40497	0.19369	-2.0908	0.03771
lPatent	0.13299	0.39760	0.3345	0.73834

We observe that our education proxy is not statistically significant. That being said, all other estimators are significant, minimum at the 10 % level. We're especially interested in the *sign* of these significant coefficients- our human capital and private capital predictors were both *negatively* correlated with earthquake probability, while the density of highways relative to population *rose.* In province years where a damage-causing earthquake took place, our model suggests more infrastructure, and *less* assets, college, and research.

We believe this effect is worth studying. While it's not accurate to say that our instrument is *strictly* exogenous, our instrument presents a unique situation where infrastructure and other drivers of productivity are likely to *diverge,* enabling us to more precisely study the effects of infrastructure, free of the potential *positive* bias of private capital, education, etc. We estimate a first-stage regression, using a two-way fixed effects model:

$$\log(\text{Highway}_{it}) = \Lambda_1 log(Earthquake_{it}) + \Lambda_2 log(Assets_{it}) + \Lambda_3 log(Patent_{it}) + \Lambda_4 log(College_{it})$$
$$+ a_i + d_t + \gamma_{it}$$

We test the significance of the coefficients, once again using four-year lagged Newey-West HAC-SEs:

Table 3: Results of t and F tests of first-stage regression

Variable	Coefficient	Std. Error	t-Value	Pr($> |t|$)
Quake	0.0245	0.0114	2.1543	0.0318
lassets	0.0406	0.0172	2.361	0.0187
lPatent	0.0640	0.0060	10.6031	1.993e-13
lCollege	0.0281	0.0087	3.2162	0.0014
Joint F-statistic	165.881 (p = 2.22e-16)			

All coefficients are significant at the 5% level or higher. We can see that the partial elasticity of highways is positive with respect to earthquakes, college rate, the patent rate, and the accumulation of private non-current assets. We recall that in the above regression, highways were positively predictive of an earthquake, while assets, college, and patents tended to move in the opposite direction in earthquake years. Because the bias of coefficients relative to each other moves in the direction of covariance, we argue that the coefficient on quake is a compelling *lower* bound estimate, and that the real, bias-free number is likely to be higher, free of the other, *negative* earthquake effects of damage to assets, human capital, and physical private capital. A correlation matrix shows that the estimator on *Quake* is *not* meaningfully correlated with other estimators, with correlations no higher than 0.15, and is suited for instrumental variable analysis:

5

Table 4: Correlation Matrix

	Quake	lCollege	lPatent	lassets
Quake	1.000	-.063	-.106	.154
lCollege	-.063	1.000	-.243	-.078
lPatent	-.106	-.243	1.000	-.748
lassets	.154	-.078	-.748	1.000

Finally, to interpret the coefficient on *Quake,* it tells us that earthquake-afflicted areas saw a statistically *significant* expected .0245 % increase in highways per million people over areas without earthquakes. To answer whether this 'earthquake effect' actually explained a positive portion of the observed variation in output, we'll estimate the second-stage regression:

$$\log(\text{workeroutput}_{it}) = \phi_1 log(\widehat{Highway}_{it}) + \phi_2 log(Patent_{it}) + \phi_3 log(College_{it}) + \phi_4 log(Assets_{it}) + a_i + d_t + \nu_{it}$$

$Highway_{it}$ is the logged highway construction that the model exclusively attributes to earthquakes. A coefficient joint significance test using 4-lagged Newey-West standard errors yields:

Table 5: Results of t and F tests of second-stage regression

Variable	Coefficient	Std. Error	t-Value	$\Pr(> |t|)$
$l\widehat{Highway}$	1.0850	0.5295	2.0491	0.0411
lCollege	0.0234	0.0104	2.2573	0.0245
lPatent	0.0320	0.0215	1.4870	0.1378
lassets	0.1281	0.0246	5.1981	3.271e-7
Joint F-statistic	18.471 (p = 6.69e-14)			

The fitted increase in infrastructure resulting exclusively from earthquakes that we established in the first-stage regression has predictive power at the 5% significance level: a 1% increase in infrastructure due to an earthquake drives an approximate 1.09% expected higher level of output per worker *over and above* the effects of human capital, worker capital, and research and technology levels. Having established the lack of positive bias in our earthquake year infrastructure estimator, and having shown that there is a stiatistically significant predicted positive elasticity of output per worker with regards to earthquake-predicted infrastructure, we argue that this is compelling evidence of infrastructure's contribution to the Solow Residual. We argue that our diverging probit model shows that the decreasing series of lCollege, lPatent, and lassets cannot account for the positive expected effects of infrastructure in earthquake-stricken provinces.

4.1   Findings

We set out to show that the "infrastructure effect" of earthquakes in Chinese provinces drove gains to output per worker. Using a two-stage IV regression with an 'earthquake' dummy as the instrument, we showed that our relevant and exogenous instrument of damage-causing earthquakes brought about a statistically significant expected % increase in infrastructure levels. Observing the effects of this fitted exogenous infrastructure on output per worker, we show that a 1% increase in infrastructure as a result of earthquakes increased output per worker by at least 1.09%.

We argue that the benefit of increasing public, nonrival capital stock is intuitive. Roads are, to an extent, nonrival public goods. A new, more efficient route to a paying customer allows a truck or train to return multiples of output to its operator at no added expense, and all trucks can take advantage of this dividend (free of congestion), thereby scaling returns on labor and private capital. A filled pothole may be a thousand prevented tire repairs. Both our econometrics-driven experiment and public goods theory substantiate our claim that infrastructure is a *force multiplier* of labor and rival private capital.

We'll address potential shortcomings in our experimental methods and model with robustness checks. One such check is a Hausman-Wu test of instrument exogeneity.

We estimate an alternative OLS model 6.

Table 6: Alternative OLS Model – Coefficient Estimates

Variable	Estimate	Std. Error	t-value	$\Pr(> |t|)$
lHighway	0.3271	0.0579	5.646	3.18×10^{-8} ***
lCollege	0.0152	0.0099	1.532	0.1262
lPatent	0.0216	0.0160	1.347	0.1786
lassets	0.1168	0.0154	7.575	2.67×10^{-13} ***
Signif. codes: *** 0.001, ** 0.01, * 0.05, . 0.1				
joint F-statistic	26.5979 (p = 2.22e-16)			

We run a Hausman-Wu test with the null hypothesis that this OLS model is consistent and efficient, and an alternative hypothesis that this model is *inconsistent,* and the IV model is *efficient.* We show that OLS has almost certainly failed to arrive at an unbiased estimator, as reported in Table **??**.

Table 7: Hausman–Wu Test

Statistic	Value
Test statistic (χ^2)	43.246
Degrees of freedom	3
p-value	2.182×10^{-9}
Null: this OLS model is consistent.	
Alternative: one model is inconsistent.	

Another concern of ours is the significance of inflation- it's feasible that an increase in price level driven by a devastating earthquake in an important province might create a 1% rise in nominal output. We included time effects categorical variables that control for variation across time but not individual. Inflation is a *time effect*, and is controlled for because it would affect all Chinese provinces.

Lastly, we highlight that earthquakes, while not being affected by output, *do* hit the provinces at unequal frequencies. In particular, Xinjiang and Sichuan have been affected by almost yearly earthquakes. It's feasible that these provinces have internalized some *unrelated* long-run effects, such as more resilient private capital or architecture, *mitigating* earthquake-driven negative effects, that our model wrongly sees as a return to infrastructure. We can rerun our models *without* these two exceptionally earthquake-prone regions and rerun our tests:

Table 8: Results t and F tests of second stage IV, NO HIGHQUAKE

Variable	Coefficient	Std. Error	t-Value	$\mathbf{Pr(> |t|)}$
$\widehat{lHighway}$	1.0850	0.5295	2.0491	0.0411
lCollege	0.0234	0.0104	2.2573	0.0245
lPatent	0.0320	0.0215	1.4870	0.1378
lassets	0.1281	0.0246	5.1981	3.271e-7
Joint F-statistic	18.471 (p = 6.69e-14)			

The coefficient, p-values, and t-scores aren't meaningfully affected after we removed the two most earthquake-prone provinces in our data.

While we observed positive correlation between independent controls, this multicollinearity is isolated to the controls. As earthquake likelihood is negatively correlated with each and every one of our controls, earthquake-fitted infrastructure estimates won't be positively inflated by this effect.

5   Discussion and Literature Review

Aschauer (1990) uses national-level time series and percentage changes to show a potential causal relationship, but his exclusive focus on the United States and limited controls left the rigor of his findings vulnerable to endogeneity and bias.. We saw an opportunity in Chinese earthquake data to undertake an event-driven study of the effects of an infrastructure and science/human capital *divergence*. We believe that our paper contributes to the pool of economic knowledge by testing the infrastructure-productivity link with more statistically rigorous methods.

Our study was made possible by the implications of Chinese earthquake recovery studies. That being said, we cannot uncritically accept their results as gospel. We highlight the risks of research done in authoritarian countries with a vested interest in "cooking the books" to portray the regime's effectiveness- subsequent independent study of the efficacy of Chinese earthquake recovery efforts should be done to test the validity of our natural experiment.

Qin X., Wu H, and Shan T., (2022) found that rural Chinese infrastructure investments significantly decreased poverty rates by improving market access for rural households, reducing transportation and transaction costs, and facilitating capital accumulation. The potential benefits of infrastructure may be more than current literature estimates- we remain hopeful that an infrastructure breakthrough may inform sensible policy for decades to come.

References

Aschauer, D. A. (1989). Is public expenditure productive? *Journal of Monetary Economics*, 23(2):177–200.

Chen, P. (2022). Analysis of the post-earthquake economic recovery of the most severely affected areas in the 2008 wenchuan earthquake. *Natural Hazards*, 114(3):2633–2655.

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

Fayazi, M., Yeh, E. T., and Li, F. (2019). Development and divergent post-disaster trajectories in a mountain village: Temporal dynamics of differentiation after the 2008 wenchuan earthquake. *World Development*, 124:104663.

McConnell, J. J. and Servaes, H. (1990). Additional evidence on equity ownership and corporate value. *Journal of Financial Economics*, 27(2):595–612.

National Bureau of Statistics of China (2025). National data. https://data.stats.gov.cn/english/index.htm?isLogin=0. Accessed: 29 April 2025.

Qin, X., Wu, H., and Shan, T. (2022). Rural infrastructure and poverty in china. *PLOS ONE*, 17(6):e0266528.

Robert E. Lucas, J. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, 22(1):3–42.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, 98(5, Part 2):S71–S102.