Do Expressways Affect Local Development Differentially? Evidence from China and a Demand-side Explanation

Guojun He Yang Xie Bing Zhang

HKUST UC Riverside Nanjing U

June 2, 2019

All in one slide

All in one slide

No expressway in China in 1983; 111,000km in 2015

How did expressway connection affect the local economies?

All in one slide

No expressway in China in 1983; 111,000km in 2015

How did expressway connection affect the local economies?

Our approach

- System designed to link key cities; link peripheral counties less intentionally

All in one slide

No expressway in China in 1983; 111,000km in 2015

How did expressway connection affect the local economies?

Our approach

- System designed to link key cities; link peripheral counties less intentionally
- DiD, 1600 peripheral counties, 2000–2012
- Matching by initial income for potential heterogeneity

All in one slide

No expressway in China in 1983; 111,000km in 2015

How did expressway connection affect the local economies?

Our approach

- System designed to link key cities; link peripheral counties less intentionally
- DiD, 1600 peripheral counties, 2000–2012
- Matching by initial income for potential heterogeneity

Minor average effect, significant heterogeneity

- GDP $\uparrow$ in initially poor counties, $\downarrow$ in initially rich counties
- Further empirics reveals driven by the demand side

All in one slide

No expressway in China in 1983; 111,000km in 2015

How did expressway connection affect the local economies?

Our approach

- System designed to link key cities; link peripheral counties less intentionally
- DiD, 1600 peripheral counties, 2000–2012
- Matching by initial income for potential heterogeneity

Minor average effect, significant heterogeneity

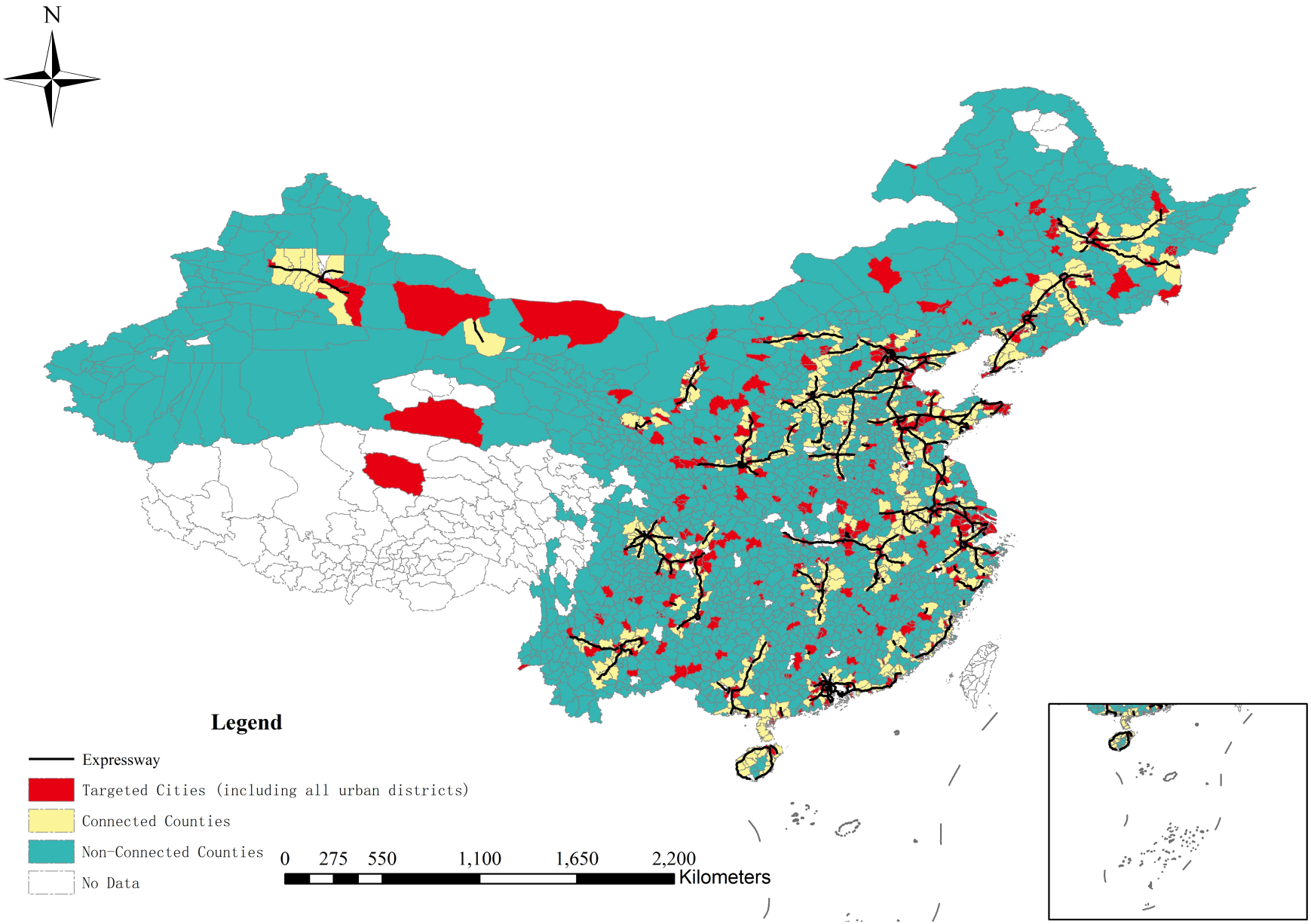
- GDP $\uparrow$ in initially poor counties, $\downarrow$ in initially rich counties
- Further empirics reveals driven by the demand side

An explanation through the consumption–environment preferences

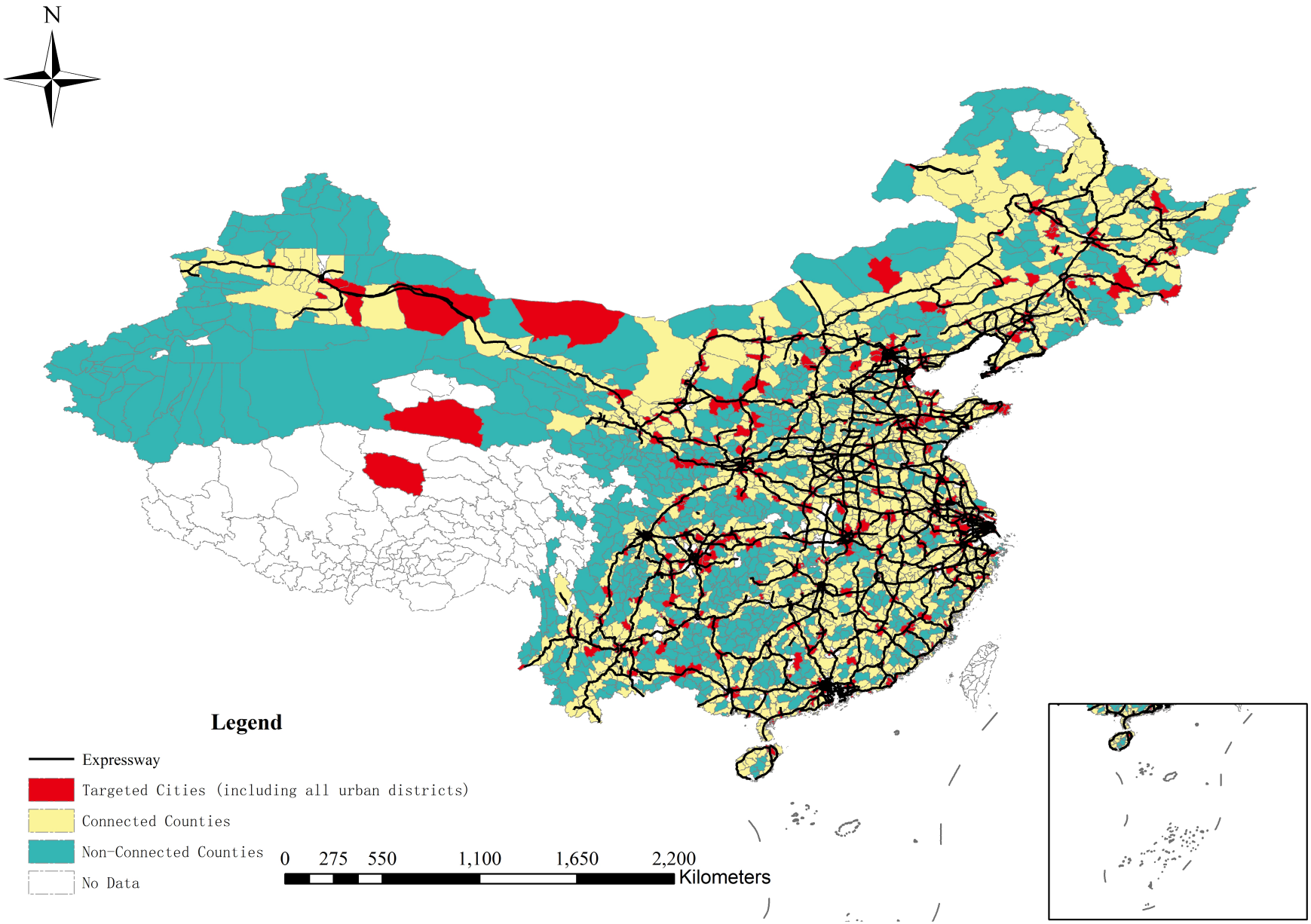
- Consistent pattern in data of emissions and industrial structure, etc.

Literature

From 2000 . . .


N
Legend
Expressway
Targeted Cities (including all urban districts)
Connected Counties
Non-Connected Counties
No Data
0 275 550 1,100 1,650 2,200
Kilometers

. . . to 2012


N
Legend
Expressway
Targeted Cities (including all urban districts)
Connected Counties
Non-Connected Counties
No Data
0 275 550 1,100 1,650 2,200
Kilometers

Econometric model

Generalized DiD with two-way fixed effects:

$$\ln \underbrace{y_{i,t}}_{\substack{\text{Real GDP or} \\ \text{GDP per capita} \\ \text{of county } i \text{ in year } t}} = \alpha + \beta \cdot \underbrace{Connect_{i,t}}_{\substack{\mathbf{I}(\text{Connected})}} + \underbrace{\rho_t}_{\substack{\text{Year-fixed} \\ \text{effect}}} + \underbrace{\mu_i}_{\substack{\text{County-fixed} \\ \text{effect}}} + \epsilon_{i,t}$$

Econometric model

Generalized DiD with two-way fixed effects:

$$\underbrace{\ln y_{i,t}}_{\substack{\text{Real GDP or} \\ \text{GDP per capita} \\ \text{of county } i \text{ in year } t}} = \alpha + \beta \cdot \underbrace{Connect_{i,t}}_{\mathbf{I}(\text{Connected})} + \underbrace{\rho_t}_{\substack{\text{Year-fixed} \\ \text{effect}}} + \underbrace{\mu_i}_{\substack{\text{County-fixed} \\ \text{effect}}} + \epsilon_{i,t}$$

Potential heterogeneity:

$$\ln y_{i,t} = \alpha + \beta \cdot Connect_{i,t}$$

Econometric model

Generalized DiD with two-way fixed effects:

$$\ln \underbrace{y_{i,t}}_{\substack{\text{Real GDP or} \\ \text{GDP per capita} \\ \text{of county } i \text{ in year } t}} = \alpha + \beta \cdot \underbrace{Connect_{i,t}}_{\substack{\mathbf{I}(\text{Connected})}} + \underbrace{\rho_t}_{\substack{\text{Year-fixed} \\ \text{effect}}} + \underbrace{\mu_i}_{\substack{\text{County-fixed} \\ \text{effect}}} + \epsilon_{i,t}$$

Potential heterogeneity:

$$\ln y_{i,t} = \alpha + \beta \cdot Connect_{i,t} + \gamma \cdot Connect_{i,t} \cdot \ln \underbrace{X_{i,2000}}_{\substack{\text{Real GDP per capita} \\ \text{in 2000}}}$$

Econometric model

Generalized DiD with two-way fixed effects:

$$\ln \underbrace{y_{i,t}}_{\substack{\text{Real GDP or} \\ \text{GDP per capita} \\ \text{of county } i \text{ in year } t}} = \alpha + \beta \cdot \underbrace{Connect_{i,t}}_{\substack{\mathbf{I}(\text{Connected})}} + \underbrace{\rho_t}_{\substack{\text{Year-fixed} \\ \text{effect}}} + \underbrace{\mu_i}_{\substack{\text{County-fixed} \\ \text{effect}}} + \epsilon_{i,t}$$

Potential heterogeneity:

$$\ln y_{i,t} = \alpha + \beta \cdot Connect_{i,t} + \gamma \cdot Connect_{i,t} \cdot \ln \underbrace{X_{i,2000}}_{\substack{\text{Real GDP per capita} \\ \text{in 2000}}} + \rho_t + \mu_i + \epsilon_{i,t}$$

Concerns on identification

We hope connection were random, but it is definitely not

Concerns on identification

We hope connection were random, but it is definitely not

Connection correlated unobservable factors that affect the outcomes?

- Connection driven by time-invariant county characteristics, e.g., geography?

  We control for the county-fixed effects

- Connection driven by the same shock to all counties, e.g., national policies?

  We control for the year-fixed effects

Concerns on identification

We hope connection were random, but it is definitely not

Connection correlated unobservable factors that affect the outcomes?

- Connection driven by time-invariant county characteristics, e.g., geography?

  We control for the county-fixed effects

- Connection driven by the same shock to all counties, e.g., national policies?

  We control for the year-fixed effects

- Given these, connection driven by local shocks?

  Plan announced well in advance, route never changed

Concerns on identification

We hope connection were random, but it is definitely not

Connection correlated unobservable factors that affect the outcomes?

- Connection driven by time-invariant county characteristics, e.g., geography?

  We control for the county-fixed effects

- Connection driven by the same shock to all counties, e.g., national policies?

  We control for the year-fixed effects

- Given these, connection driven by local shocks?

  Plan announced well in advance, route never changed

- Policymakers knowing more than us, compound treatments, etc.?

  Classic concern for DiD – IV and event studies could help to some extent

Concerns on identification

We hope connection were random, but it is definitely not

Connection correlated unobservable factors that affect the outcomes?

- Connection driven by time-invariant county characteristics, e.g., geography?

  We control for the county-fixed effects

- Connection driven by the same shock to all counties, e.g., national policies?

  We control for the year-fixed effects

- Given these, connection driven by local shocks?

  Plan announced well in advance, route never changed

- Policymakers knowing more than us, compound treatments, etc.?

  Classic concern for DiD – IV and event studies could help to some extent

Parallel pre-trends confirmed by data Parallel

Panel-data across Chinese counties over 2000–2012

Expressway expansion data from historical GIS data

Socio-economic data from yearbooks

Emission data from the Environmental Survey and Reporting database

- Ministry of Environmental Protection of China
- All major polluting sources at the plant level
- Primary measure: Chemical Oxygen Demand (COD) in water Why COD?
- Alternative measures: NH_3N, SO_2, NO_x

Summary Match

Average treatment effect

Table 2. The Average Treatment Effects of Expressway Connection on GDP

	GDP (million yuan, log)			Per capita GDP (yuan, log)		
	(1)	(2)	(3)	(4)	(5)	(6)
Expressway	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
County FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	N	Y	Y	N
Provincial Trends	N	Y	N	N	Y	N
Province-Year FE	N	N	Y	N	N	Y
Obs.	13,440	13,440	13,440	13,347	13,347	13,347
R^2	0.87	0.90	0.91	0.86	0.89	0.90

Notes: This table estimates the impacts of expressway connection on GDP measures using a variety of specifications. GDP data are deflated, where Beijing-2000 is the base province-year. We probe the robustness of estimates accuracy by clustering the standard errors at three different levels: county level, province level and county and province-year level (multi-way clustering suggested by Cameron, Gelbach, and Miller (2011)). These standard errors are respectively reported in the parentheses below the estimated coefficients. Our preferred specification clusters standard errors at the county level. *** p<0.01, ** p<0.05, * p<0.1.

On average, connection hardly changed GDP

- Robust w.r.t. different levels at which the standard errors are clustered
- Robust w.r.t. controlling for provincial trends or province-year fixed effects

Heterogenous treatment effects

Table 3. Heterogeneous Treatment Effect with respect to Initial Income

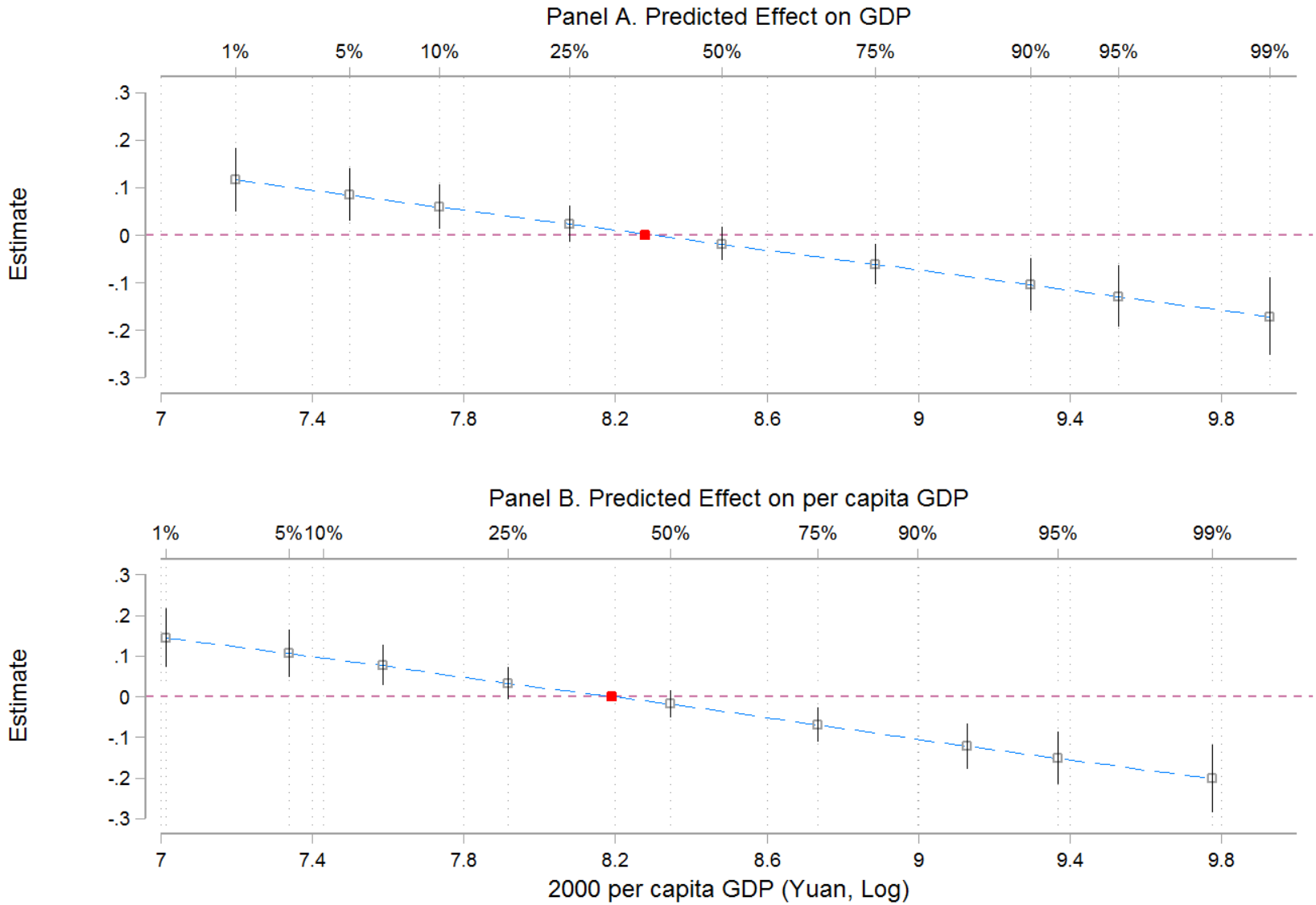
	GDP (million yuan, log)			Per capita GDP (yuan, log)		
	(1)	(2)	(3)	(4)	(5)	(6)
Expressway	0.87***	0.75***	0.80***	1.06***	0.92***	0.98***
	(0.21)	(0.18)	(0.18)	(0.22)	(0.19)	(0.19)
	(0.41)	(0.29)	(0.29)	(0.44)	(0.29)	(0.29)
	(0.27)	(0.22)	(0.24)	(0.29)	(0.23)	(0.24)
Expressway*GDP pc	-0.11***	-0.09***	-0.10***	-0.13***	-0.11***	-0.12***
(yuan, log, Year 2000)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
	(0.05)	(0.03)	(0.03)	(0.05)	(0.03)	(0.03)
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
County FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	N	Y	Y	N
Provincial Trends	N	Y	N	N	Y	N
Province-Year FE	N	N	Y	N	N	Y
Obs.	13,440	13,440	13,440	13,347	13,347	13,347
R^2	0.87	0.90	0.91	0.86	0.89	0.90

Notes: This table estimates the heterogeneous impacts of expressway connection on GDP measures using a variety of specifications. GDP data are deflated, where Beijing-2000 is the base province-year. We probe the robustness of estimates accuracy by clustering the standard errors at three different levels: county level, province level and county and province-year level (multi-way clustering suggested by Cameron, Gelbach, and Miller (2011)). These standard errors are respectively reported in the parentheses below the estimated coefficients. Our preferred specification clusters standard errors at the county level. *** p<0.01, ** p<0.05, * p<0.1.

Initially richer counties saw more negative impacts on GDP

- Robust w.r.t. different levels at which the standard errors are clustered
- Robust w.r.t. controlling for provincial trends or province-year fixed effects

Predicted treatment effects



Concerns and robustness checks

Nonlinear specification? Driven by intrinsically converging growth? Betty

Parallel pre-trends/the same pre-levels? Within subgroups? Event Subgroups

Unmatched sample Unmatched

IV approach IV

Potential spill-overs Spill-overs

Why does the heterogeneity emerge?

Initial income correlated with many things

- A story about supply, e.g., endowment, comparative advantage, etc.?
- About demand, e.g., consumer and governmental preferences?
- About initial access to nearby market, e.g., initial trade cost?

Why does the heterogeneity emerge?

Initial income correlated with many things

- A story about supply, e.g., endowment, comparative advantage, etc.?
- About demand, e.g., consumer and governmental preferences?
- About initial access to nearby market, e.g., initial trade cost?

Some specific stories

- Driven by the comparative advantage? No, otherwise impact always positive
- By the home market effect? No, otherwise impact always negative
- A combination of these two?

Why does the heterogeneity emerge?

Initial income correlated with many things

- A story about supply, e.g., endowment, comparative advantage, etc.?
- About demand, e.g., consumer and governmental preferences?
- About initial access to nearby market, e.g., initial trade cost?

Some specific stories

- Driven by the comparative advantage? No, otherwise impact always positive
- By the home market effect? No, otherwise impact always negative
- A combination of these two?

Can factors related to the two stories dilute the main result?

Impact heterogeneity across initial income not diluted

Table 5. Explore Heterogeneity Patterns

	Panel A. GDP (million yuan, log)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Expressway	0.87***	1.61***	0.44	0.71***	0.86***	0.89***	0.89***	0.91***	2.00***
	(0.21)	(0.24)	(0.31)	(0.22)	(0.25)	(0.21)	(0.29)	(0.21)	(0.34)
Expressway*GDP pc	-0.11***	-0.11***	-0.10***	-0.10***	-0.10***	-0.12***	-0.11***	-0.11***	-0.23***
(Year 2000)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)
Expressway*X	0.00	-0.11***	0.05**	0.07***	-0.00	0.01	-0.00	0.00*	/
(Year 2000)	(0.00)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.00)	(0.00)	/
	Panel B. GDP per capita (yuan, log)								
Expressway	1.06***	1.76***	0.57*	0.84***	1.03***	1.07***	1.08***	1.10***	2.15***
	(0.22)	(0.25)	(0.33)	(0.24)	(0.26)	(0.22)	(0.31)	(0.22)	(0.35)
Expressway*GDP pc	-0.13***	-0.13***	-0.11***	-0.11***	-0.13***	-0.14***	-0.13***	-0.14***	-0.24***
(Year 2000)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)
Expressway*X	-0.00	-0.11***	0.05**	0.07***	0.00	0.01	-0.00	0.00**	/
(Year 2000)	(0.00)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.00)	(0.00)	/
X Indicator	Distance (km)	Population (log)	Land Area (log)	Land per capita (log)	# Industrial firms (log)	Output Value (log)	Agriculture (%)	Manufacturing (%)	All
County FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Obs.	13,347	12,245	12,236	12,236	12,210	13,264	13,347	13,347	12,144
R^2	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.87

Notes: This table estimates the heterogeneous impacts of expressway connection on GDP measures. GDP data are deflated, where Beijing-2000 is the base province-year. Standard errors are clustered at county level and reported in the parentheses below the estimated coefficients. *** p<0.01, ** p<0.05, * p<0.1.

Similar results with the income group specification

A demand side explanation

A demand side explanation

Expressway connection $\Rightarrow$ trade cost $\downarrow$

A demand side explanation

Expressway connection $\Rightarrow$ trade cost $\downarrow$

$\Rightarrow$ economy's ability to transform emissions into consumption $\uparrow$

A demand side explanation

Expressway connection $\Rightarrow$ trade cost $\downarrow$

$\Rightarrow$ economy's ability to transform emissions into consumption $\uparrow$

$\Rightarrow$ rebalance between consumption and the environment

A demand side explanation

Expressway connection $\Rightarrow$ trade cost $\downarrow$

$\Rightarrow$ economy's ability to transform emissions into consumption $\uparrow$

$\Rightarrow$ rebalance between consumption and the environment

Depends on the preference over the two

A demand side explanation

Expressway connection $\Rightarrow$ trade cost $\downarrow$

$\Rightarrow$ economy's ability to transform emissions into consumption $\uparrow$

$\Rightarrow$ rebalance between consumption and the environment

Depends on the preference over the two

Poor regions care more about consumption, rich primarily the environment

- Generally true; Chinese context:

  Scientific Outlook of Development (2003)

  Council of State Directive on cadre evaluation (2005)

  CCCPC Organizational Department Directive on cadre evaluation (2013)

  Evidence in mayor promotion (e.g., Sun et al., 2014; Zheng et al., 2014)

A demand side explanation

Expressway connection $\Rightarrow$ trade cost $\downarrow$

$\Rightarrow$ economy's ability to transform emissions into consumption $\uparrow$

$\Rightarrow$ rebalance between consumption and the environment

Depends on the preference over the two

Poor regions care more about consumption, rich primarily the environment

- Generally true; Chinese context:

  Scientific Outlook of Development (2003)

  Council of State Directive on cadre evaluation (2005)

  CCCPC Organizational Department Directive on cadre evaluation (2013)

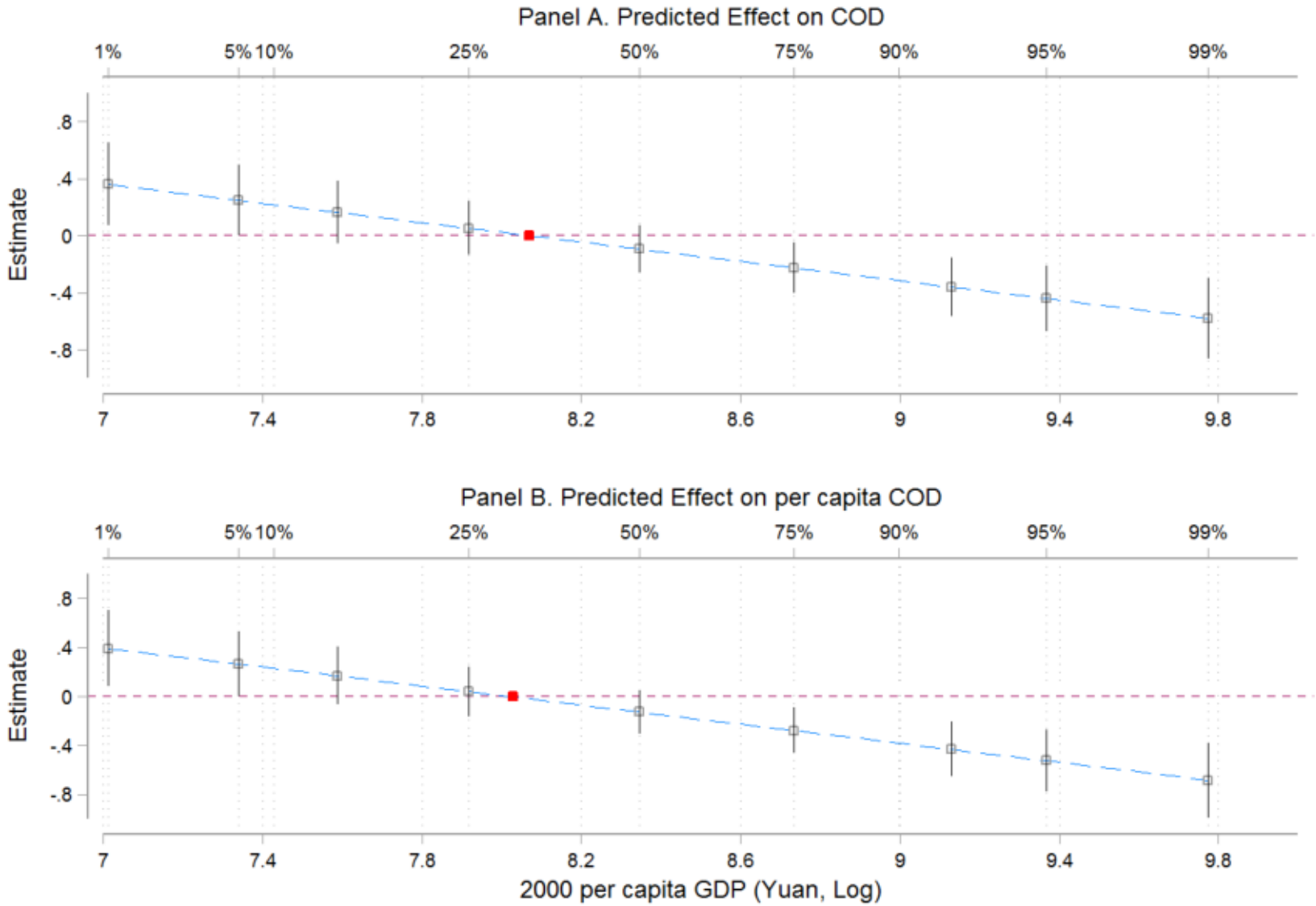
  Evidence in mayor promotion (e.g., Sun et al., 2014; Zheng et al., 2014)

Expressway gives the poor an opportunity to develop,

the rich to sacrifice more consumption for the environment

Model

Expressway impacts on emissions



Consistent result; other emission measure results as expected

Emission

How did the rebalance happen?

Table 8. Expressway Connection: Channels

	COD Emission Intensity (ton per RMB-value of output, log)		Number of Key Polluting Firms (log)		Output Value of Key Polluting Firms		Share of the Secondary Industry (%, log)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Expressway	-0.04	2.77***	-0.07	1.06**	0.04	1.17***	-0.00	0.63***
	(0.08)	(0.81)	(0.05)	(0.42)	(0.03)	(0.27)	(0.01)	(0.14)
Expressway*GDP pc		-0.33***		-0.13***		-0.13***		-0.07***
(yuan, log, Year 2000)		(0.09)		(0.05)		(0.03)		(0.02)
County FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Obs.	14,531	14,531	14,711	14,711	14,713	14,713	8,051	8,051
R^2	0.11	0.11	0.54	0.54	0.39	0.39	0.21	0.21

Notes: This table estimates the heterogeneous impacts of expressway connection environmental and economic outcomes. GDP data are deflated, where Beijing-2000 is the base province-year. Standard errors are clustered at county level and reported in the parentheses below the estimated coefficients. *** p<0.01, ** p<0.05, * p<0.1.

Technology changed, firms entered/exited/moved, industry shifted

Concluding remarks

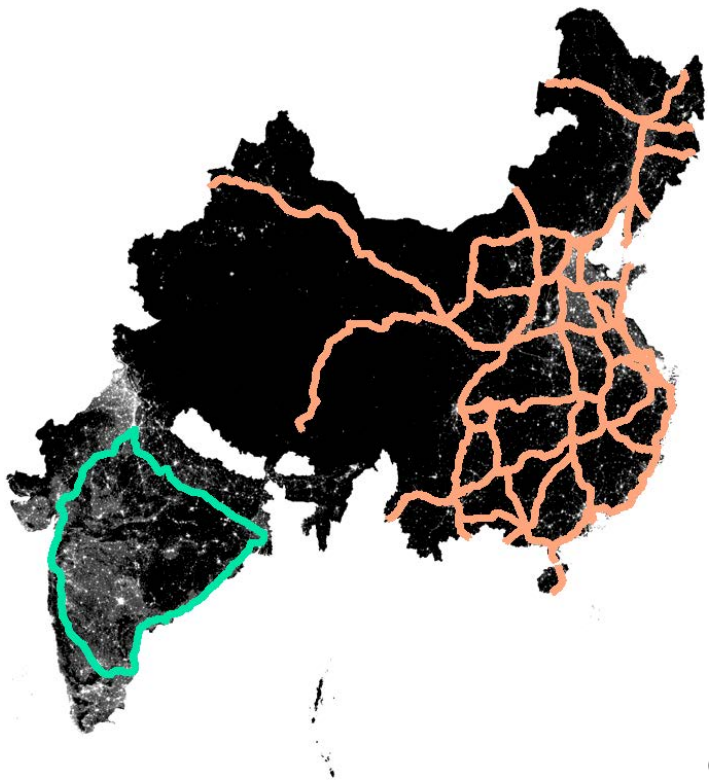
In China, expressways alleviate inequality among peripheral areas

- At an environmental cost; GDP–environment preference looks important

Concluding remarks

In China, expressways alleviate inequality among peripheral areas

- At an environmental cost; GDP–environment preference looks important

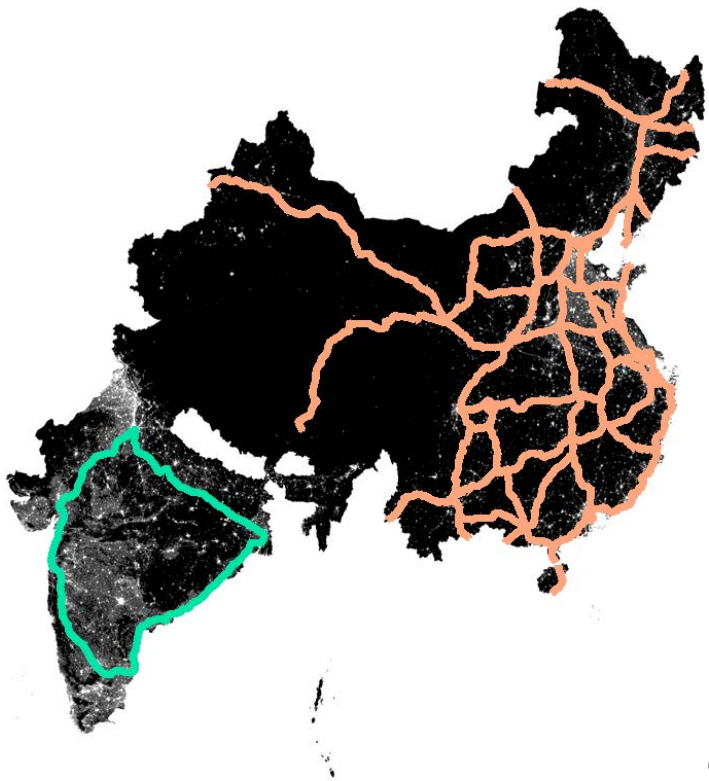


Golden Quadrilateral vs. National Expressway Network
Source: Alder (2019)

Concluding remarks

In China, expressways alleviate inequality among peripheral areas

- At an environmental cost; GDP–environment preference looks important



Golden Quadrilateral vs. National Expressway Network
Source: Alder (2019)

Alder (2019): Efficiency *and* distribution gains in India from "Chinese" roads

Acknowledgement

Hunt Allcott, Michael Anderson, Richard Arnott, Chong-En Bai, Michael Bates, Nathaniel Baum-Snow, Cyndi Berck, Peter Berck, Judd Boomhower, Jimmy Chan, Gordon Dahl, Alain de Janvry, Ozkan Eren, Thibault Fally, Jingting Fan, Fred Finan, Shihe Fu, Joshua Graff Zivin, Michael Greenstone, Jie He, Steven Helfand, Sarojini Hirshleifer, Wei Huang, Ruixue Jia, Larry Karp, Bree Lang, Matt Lang, Bryan Leonard, Weijia Li, Jeremy Magruder, Aprajit Mahajan, Daniel McMillen, Helene Ollivier, Albert Park, Martino Pelli, Jeffrey Perloff, Obie Porteous, Han Qi, David Rapson, Elisabeth Sadoulet, Ruoyao Shi, Leo Simon, Michael Song, Kenneth Small, Qu Tang, Itai Trilnick, Reed Walker, Shaoda Wang, Brian Wright, Yiqing Xu, Jia Yan ...

CU Boulder, CUHK, Fudan, HKBU, HKUST, PKU, SJTU, SMU, THU, UC Berkeley, UC Riverside, Universit de Sherbrooke, USC and 2017 AERE, CAERE, CES, EAERE, ESAM, Fudan–UC ...

Xiaoxiao Shen, Yuhang Pan, Chun Wai Cheung, Ziteng Lei, Tingjun Man, Jing Yang

HKUST IEMS research grant

Economic impact of transport infrastructure investment

Economic impact of transport infrastructure investment

A long list of empirical literature, in particular about China

Economic impact of transport infrastructure investment

A long list of empirical literature, in particular about China

Prefecture or city level analysis

Economic impact of transport infrastructure investment

A long list of empirical literature, in particular about China

Prefecture or city level analysis

- Either reduced-formed or structural, but endogeneity problem in general

Economic impact of transport infrastructure investment

A long list of empirical literature, in particular about China

Prefecture or city level analysis

- Either reduced-formed or structural, but endogeneity problem in general
- IV – exclusion criteria often challenged, usually lack of variation over time

Economic impact of transport infrastructure investment

A long list of empirical literature, in particular about China

Prefecture or city level analysis

- Either reduced-formed or structural, but endogeneity problem in general
- IV – exclusion criteria often challenged, usually lack of variation over time

Faber (2014 *RES*)

- County level analysis – the connected and unconnected more comparable

Economic impact of transport infrastructure investment

A long list of empirical literature, in particular about China

Prefecture or city level analysis

- Either reduced-formed or structural, but endogeneity problem in general
- IV – exclusion criteria often challenged, usually lack of variation over time

Faber (2014 *RES*)

- County level analysis – the connected and unconnected more comparable
- DiD (1997 *and* 2006) with IV – former concerns about IV still apply

Economic impact of transport infrastructure investment

A long list of empirical literature, in particular about China

Prefecture or city level analysis

- Either reduced-formed or structural, but endogeneity problem in general
- IV – exclusion criteria often challenged, usually lack of variation over time

Faber (2014 *RES*)

- County level analysis – the connected and unconnected more comparable
- DiD (1997 *and* 2006) with IV – former concerns about IV still apply
- Negative impact on average, interpreted as the home market effect

Economic impact of transport infrastructure investment

A long list of empirical literature, in particular about China

Prefecture or city level analysis

- Either reduced-formed or structural, but endogeneity problem in general
- IV – exclusion criteria often challenged, usually lack of variation over time

Faber (2014 *RES*)

- County level analysis – the connected and unconnected more comparable
- DiD (1997 *and* 2006) with IV – former concerns about IV still apply
- Negative impact on average, interpreted as the home market effect

This paper

Economic impact of transport infrastructure investment

A long list of empirical literature, in particular about China

Prefecture or city level analysis

- Either reduced-formed or structural, but endogeneity problem in general
- IV – exclusion criteria often challenged, usually lack of variation over time

Faber (2014 *RES*)

- County level analysis – the connected and unconnected more comparable
- DiD (1997 *and* 2006) with IV – former concerns about IV still apply
- Negative impact on average, interpreted as the home market effect

This paper

- Panel data of 13 years, 1600 counties – longest, largest, finest in the literature

Economic impact of transport infrastructure investment

A long list of empirical literature, in particular about China

Prefecture or city level analysis

- Either reduced-formed or structural, but endogeneity problem in general
- IV – exclusion criteria often challenged, usually lack of variation over time

Faber (2014 *RES*)

- County level analysis – the connected and unconnected more comparable
- DiD (1997 *and* 2006) with IV – former concerns about IV still apply
- Negative impact on average, interpreted as the home market effect

This paper

- Panel data of 13 years, 1600 counties – longest, largest, finest in the literature
- DiD with matching by initial income – better identification for relative effect

Economic impact of transport infrastructure investment

A long list of empirical literature, in particular about China

Prefecture or city level analysis

- Either reduced-formed or structural, but endogeneity problem in general
- IV – exclusion criteria often challenged, usually lack of variation over time

Faber (2014 *RES*)

- County level analysis – the connected and unconnected more comparable
- DiD (1997 *and* 2006) with IV – former concerns about IV still apply
- Negative impact on average, interpreted as the home market effect

This paper

- Panel data of 13 years, 1600 counties – longest, largest, finest in the literature
- DiD with matching by initial income – better identification for relative effect
  - Parallel pre-trends/the same pre-levels; results robust when IV used

Economic impact of transport infrastructure investment

A long list of empirical literature, in particular about China

Prefecture or city level analysis

- Either reduced-formed or structural, but endogeneity problem in general
- IV – exclusion criteria often challenged, usually lack of variation over time

Faber (2014 *RES*)

- County level analysis – the connected and unconnected more comparable
- DiD (1997 *and* 2006) with IV – former concerns about IV still apply
- Negative impact on average, interpreted as the home market effect

This paper

- Panel data of 13 years, 1600 counties – longest, largest, finest in the literature
- DiD with matching by initial income – better identification for relative effect
  - Parallel pre-trends/the same pre-levels; results robust when IV used
- Heterogeneity emphasized, hinting demand-side stories

Economic impact of transport infrastructure investment

A long list of empirical literature, in particular about China

Prefecture or city level analysis

- Either reduced-formed or structural, but endogeneity problem in general
- IV – exclusion criteria often challenged, usually lack of variation over time

Faber (2014 *RES*)

- County level analysis – the connected and unconnected more comparable
- DiD (1997 *and* 2006) with IV – former concerns about IV still apply
- Negative impact on average, interpreted as the home market effect

This paper

- Panel data of 13 years, 1600 counties – longest, largest, finest in the literature
- DiD with matching by initial income – better identification for relative effect
  - Parallel pre-trends/the same pre-levels; results robust when IV used
- Heterogeneity emphasized, hinting demand-side stories

PHH EKC Back to setting

Environmental impact of trade

E.g., pollution haven hypothesis

Openness to trade seldom exogenous

Our empirical setting more credible

The emission results consistent with the hypothesis

The GDP results inconsistent with its underlying mechanism

Back to setting

Environmental Kuznets curve

Empirical effort following Krueger and Grossman (1995), evidence mixed

Critiques from Arrow et al. (1995) and Stern (2004)

Copeland and Taylor (2004): Source of economic growth important

This paper

- Results inconsistent with the environmental Kuznets curve
- The same kind of shock move income and emissions together
- But the direction can differ across income levels
- Need 3-, not 2-dimensional analysis

Back to setting

Why COD?

Primary indicator for water pollution

- Oxygen required to oxidize (organic) pollutants in water

Produced and reported by almost all key pollution sources and industries

Why COD?

Primary indicator for water pollution

- Oxygen required to oxidize (organic) pollutants in water

Produced and reported by almost all key pollution sources and industries

NH_3N: Poor data quality

SO_2: Emission concentrated in certain industries and areas

NO_x: Data for 2000–2005 unavailable

Back to data

Eventually connected vs. never connected

Table 1. Summary Statistics of Sampled Counties

Variable	Un-Matched Sample			Matched Sample		
	Overall	Connected	Un-Connected	Overall	Connected	Un-Connected
GDP (million yuan, 2000)	2,978	3,416	2,455	3,067	3,265	2,868
	(6,910)	(3,241)	(9,588)	(4,616)	(2,942)	(5,827)
GDP (million yuan, 2012)	12,925	14,881	10,570	13,249	14,098	12,400
	(27,918)	(15,408)	(37,723)	(19,604)	(14,031)	(23,911)
GDP per capita (yuan, 2000)	5,710	6,231	5,081	6,084	6,087	6,082
	(5,114)	(4,567)	(5,646)	(4,287)	(3,946)	(4,608)
GDP per capita (yuan, 2012)	26,654	28,052	24,961	28,516	27,472	29,559
	(29,377)	(27,713)	(31,210)	(30,823)	(28,432)	(33,051)
# of Counties	1,646	897	749	1,614	807	807

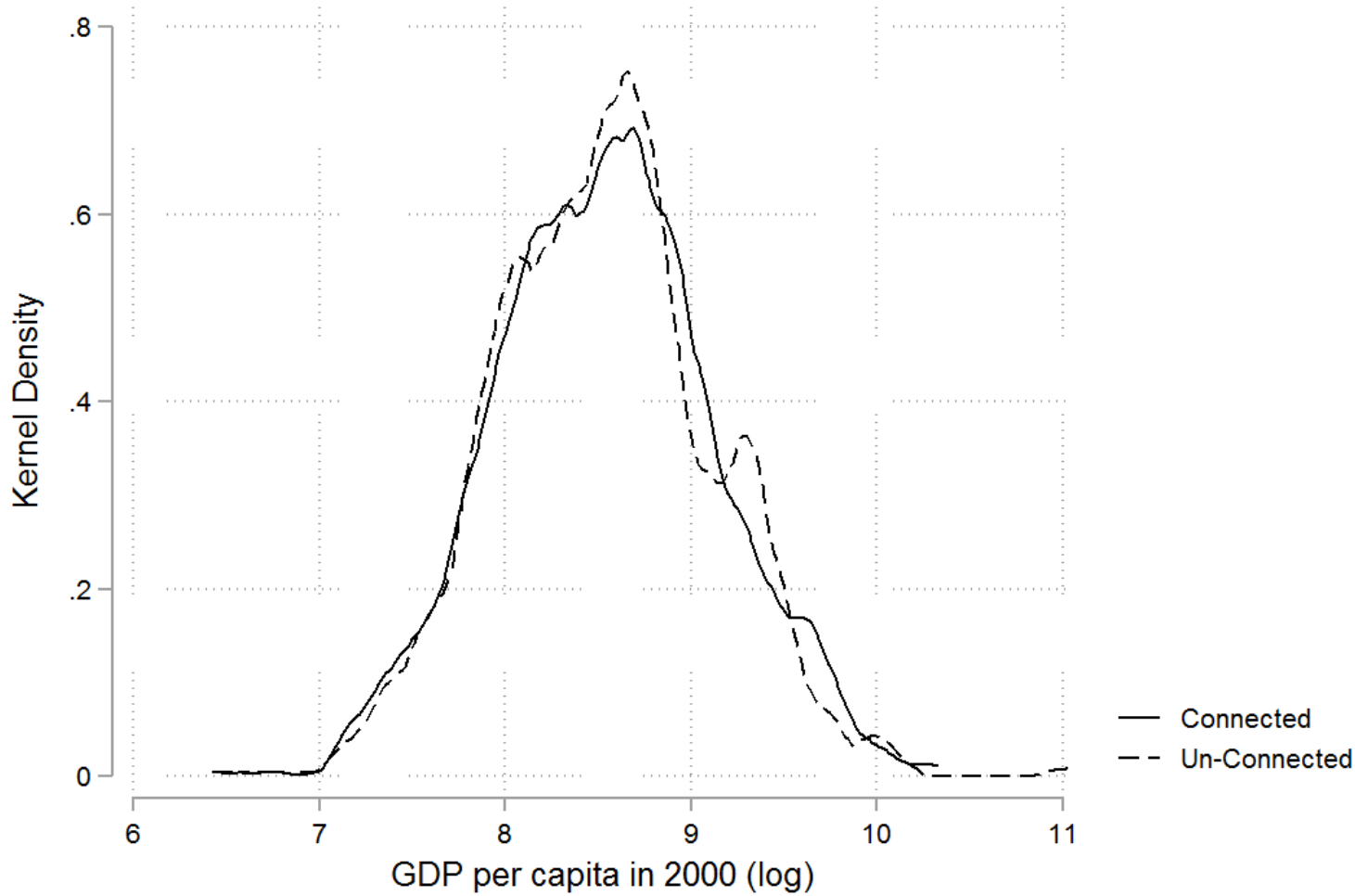
Notes: County-level nominal GDP and population data are collected from provincial statistical yearbooks, China City Statistical Yearbooks, China County Statistical Yearbooks and China Economic Database from CEIC (www.ceicdata.com). GDP data are deflated overtime and across regions using provincial CPI and cross-provincial CPI from the National Bureau of Statistics of China (Yu, 2006), taking Beijing-2000 as the base province-year. Standard deviations are reported in the parentheses below the means.

- Connected counties: 11% in 2000, more than 50% in 2012
- Average income changed a lot over time

  Sample mean of real GDP per capita grew 13.7% every year on average

Back to data

Matching by initial income within the province



Back to data

Parallel pre-trends?

An event-study approach

- D_{it}^k for the kth year after connection, baseline $k = -1$

$$\ln y_{it} = \underbrace{\sum_{k \leq -2} \delta_k \cdot D_{it}^k}_{\text{Pre-trends}} + \underbrace{\sum_{k \geq 0} \delta_k \cdot D_{it}^k}_{\text{Impact dynamics}} + \rho_t + \mu_i + \epsilon_{it}$$

Parallel pre-trends?

An event-study approach

- D_{it}^k for the kth year after connection, baseline $k = -1$

$$\ln y_{it} = \underbrace{\sum_{k \leq -2} \delta_k \cdot D_{it}^k}_{\text{Pre-trends}} + \underbrace{\sum_{k \geq 0} \delta_k \cdot D_{it}^k}_{\text{Impact dynamics}} + \rho_t + \mu_i + \epsilon_{it}$$

Matching by initial income within the province

- Primarily for better investigation on income heterogeneity
- The connected and unconnected comparable in income level

Parallel pre-trends?

An event-study approach

- D_{it}^k for the kth year after connection, baseline $k = -1$

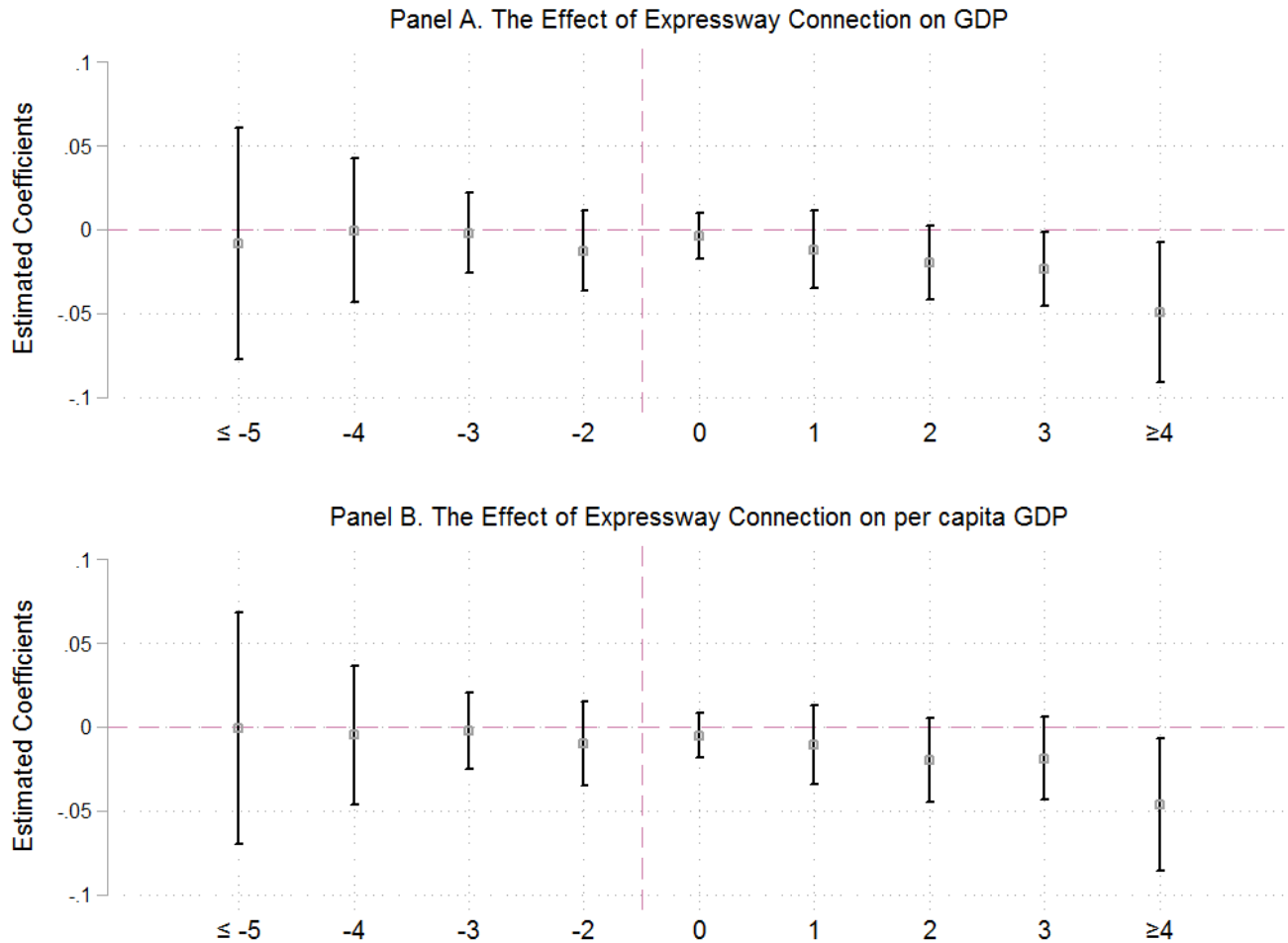
$$\ln y_{it} = \underbrace{\sum_{k \leq -2} \delta_k \cdot D_{it}^k}_{\text{Pre-trends}} + \underbrace{\sum_{k \geq 0} \delta_k \cdot D_{it}^k}_{\text{Impact dynamics}} + \rho_t + \mu_i + \epsilon_{it}$$

Matching by initial income within the province

- Primarily for better investigation on income heterogeneity
- The connected and unconnected comparable in income level
- Parallel pre-trends would imply the same pre-levels – even better identification

Back to data

Event study



Parallel pre-trends/the same pre-levels; negligible impact within 3 years

Back to concerns

Two major concerns

Table 4. Heterogeneous Treatment Effect with respect to Different Initial Income Groups

	GDP (million yuan, log)				Per capita GDP (yuan, log)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Expressway	0.10***	0.11***	0.06**	0.06**	0.11***	0.12***	0.06**	0.07**
	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)
High Income*Expressway	-0.16***	-0.16***	-0.10***	-0.10***	-0.17***	-0.18***	-0.10***	-0.11***
	(0.02)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)
County FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Provincial Trends	Y	N	N	N	Y	N	N	N
Province-Year FE	N	Y	N	N	N	Y	N	N
Income Group Trends	N	N	Y	N	N	N	Y	N
Income Group * Year FE	N	N	N	Y	N	N	N	Y
Obs.	13,440	13,440	13,440	13,440	13,347	13,347	13,347	13,347
R^2	0.90	0.91	0.87	0.87	0.89	0.91	0.87	0.87

Notes: This table estimates the heterogeneous impacts of expressway connection on GDP. GDP data are deflated, where Beijing-2000 is the base province-year. Standard errors are clustered at county level and reported in the parentheses below the estimated coefficients. *** p<0.01, ** p<0.05, * p<0.1.

- Robust w.r.t. nonlinear specification, e.g., two income-groups
- Robust w.r.t. controlling for group-specific dynamics

Back to concerns

Event study for each income-group

Table S2. Parallel Trends Tests Separately for Different Income Groups

	Overall		Low Income Group		High Income Group	
	GDP (log)	Per capita GDP (log)	GDP (log)	Per capita GDP (log)	GDP (log)	Per capita GDP (log)
	(1)	(2)	(3)	(4)	(5)	(6)
>= 5Years Before	-0.01	-0.00	-0.09	-0.08	0.01	0.01
	(0.04)	(0.04)	(0.08)	(0.08)	(0.03)	(0.03)
4 Years Before	-0.00	-0.00	-0.06	-0.06	0.01	0.01
	(0.02)	(0.02)	(0.04)	(0.04)	(0.02)	(0.02)
3 Years Before	-0.00	-0.00	-0.01	-0.00	-0.00	-0.00
	(0.01)	(0.01)	(0.03)	(0.02)	(0.01)	(0.01)
2 Years Before	-0.01	-0.01	-0.01	0.00	-0.01	-0.01
	(0.01)	(0.01)	(0.02)	(0.03)	(0.01)	(0.01)
Year of Connection	-0.00	-0.01	0.02	0.03*	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
1 Year Later	-0.01	-0.01	0.02	0.04	-0.02	-0.02
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
2 Years Later	-0.02*	-0.02	0.03	0.05*	-0.03***	-0.03***
	(0.01)	(0.01)	(0.02)	(0.03)	(0.01)	(0.01)
3 Years Later	-0.02**	-0.02	0.03	0.05*	-0.04***	-0.04**
	(0.01)	(0.01)	(0.02)	(0.03)	(0.01)	(0.01)
>=4 Years Later	-0.05**	-0.05**	0.02	0.03	-0.06***	-0.06***
	(0.02)	(0.02)	(0.04)	(0.03)	(0.02)	(0.02)
County FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Obs.	13,440	13,347	2,976	2,951	10,464	10,396
R^2	0.87	0.86	0.91	0.90	0.87	0.86

Notes: We conduct an event study by including leads and lags of the first expressway connection dummy in the regressions. The dummy indicating one-year prior treatment status is omitted from the regression. GDP data are deflated, where Beijing-2000 is the base province-year. Standard errors are clustered at the county level and reported in the parentheses. ** p<0.01, ** p<0.05, * p<0.1.

Back to concerns

Unmatched sample

Table S4. The Effects of Expressway Connection on GDP: Results from Un-matched Sample

	GDP (million yuan, log)				Per capita GDP (yuan, log)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Expressway	-0.01	2.80***	1.74**	2.03***	-0.01	3.27***	1.94**	2.25***
	(0.01)	(0.74)	(0.77)	(0.77)	(0.01)	(0.80)	(0.84)	(0.85)
	(0.02)	(0.87)	(0.72)	(0.70)	(0.02)	(0.97)	(0.83)	(0.80)
	(0.01)	(0.81)	(0.79)	(0.86)	(0.01)	(0.89)	(0.87)	(0.94)
Expressway*GDP pc		-0.35***	-0.22**	-0.26***		-0.41***	-0.25***	-0.29***
(yuan, log, Year 2000)		(0.09)	(0.09)	(0.09)		(0.09)	(0.10)	(0.10)
		(0.10)	(0.09)	(0.08)		(0.11)	(0.10)	(0.09)
		(0.09)	(0.09)	(0.10)		(0.10)	(0.10)	(0.11)
County FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	N	Y	Y	Y	N
Provincial Trends	N	N	Y	N	N	N	Y	N
Province-Year FE	N	N	N	Y	N	N	N	Y
Obs.	19,835	18,179	18,179	18,179	19,472	18,007	18,007	18,007
R^2	0.87	0.08	0.12	0.16	0.86	0.08	0.13	0.17

Notes: This table estimates the impacts of expressway connection on GDP measures using the sample before matching. GDP data are deflated, where Beijing-2000 is the base province-year. We probe the robustness of estimates accuracy by clustering the standard errors at three different levels: county level, province level and county and province-year level (multi-way clustering suggested by Cameron, Gelbach, and Miller (2011)). These standard errors are respectively reported in the parentheses below the estimated coefficients. Our preferred specification clusters standard errors at the county level. *** p<0.01, ** p<0.05, * p<0.1.

Back to concerns

Straight-line IV

Table S3. Estimates Using Straight-Line IV

	1st Stage: Expressway	2SLS: Δ log (GDP) between 2000 and 2012		2SLS:Δ log (per capita GDP) between 2000 and 2012	
	(1)	(2)	(3)	(4)	(5)
Straight-Line IV	0.34***				
	(0.03)				
Expressway		-0.05	1.79***	0.03	2.19***
		(0.06)	(0.56)	(0.08)	(0.52)
Expressway*GDP pc (yuan, log, Year 2000)			-0.22***		-0.26***
			(0.06)		(0.06)
Specification	1st Stage	2SLS	2SLS	2SLS	2SLS
Province FE	Y	Y	Y	Y	Y
Observations	1,684	1,586	1,564	1,547	1,547
R^2	0.23	0.25	0.27	0.29	0.32

Notes: Each column in the table represents a separate regression. The instrumental variable (IV) is constructed using straight lines that connect pairs of target cities. If a county is located on the straight line between two target cities, the IV equals to 1, and otherwise 0. For columns 3 and 5, the straight-line IV interacted with per capita GDP in 2000 is used to instrument the expressway connection interacted with per capita GDP in 2000. GDP data are deflated, where Beijing-2000 is the base province-year. Standard errors are clustered at the province level. * significant at 10% ** significant at 5% *** significant at 1%.

Back to concerns

Impact of neighbouring connections

Table S6. Heterogeneous Effect of Neighbor County Expressway Connection

	GDP (million yuan, log)			Per capita GDP (yuan, log)		
	(1)	(2)	(3)	(4)	(5)	(6)
Neighbor Expressway	0.52*	0.55**	0.50**	0.78***	0.65***	0.63**
	(0.29)	(0.22)	(0.25)	(0.29)	(0.22)	(0.25)
N-Expressway*GDP pc (yuan, log, Year 2000)	-0.06	-0.06**	-0.06*	-0.09**	-0.07***	-0.07**
	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)
County FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	N	Y	Y	N
Provincial Trends	N	Y	N	N	Y	N
Province-Year FE	N	N	Y	N	N	Y
Obs.	8,732	8,732	8,732	8,688	8,688	8,688
R^2	0.87	0.89	0.89	0.86	0.88	0.89

Notes: This table estimates the heterogeneous impacts of expressway connection on GDP measures using a variety of specifications. GDP data are deflated, where Beijing-2000 is the base province-year. Standard errors clustered at the county level are reported in the parentheses below the estimated coefficients. *** p<0.01, ** p<0.05, * p<0.1.

The same pattern as the main result – spill-overs works *against* it

Coefficients smaller – neighbouring connection weaker than own connection

Back to concerns

Model (in appendix)

2 sectors producing consumption goods with different emission intensity

Labor input immobile across boundaries

Local emissions damage local environmental quality

Model (in appendix)

2 sectors producing consumption goods with different emission intensity

Labor input immobile across boundaries

Local emissions damage local environmental quality; no dynamics

No substitution between consumption goods – fixed composition

Consumers enjoy consumption composite and local environmental quality

Small open economy, world prices of consumption goods given

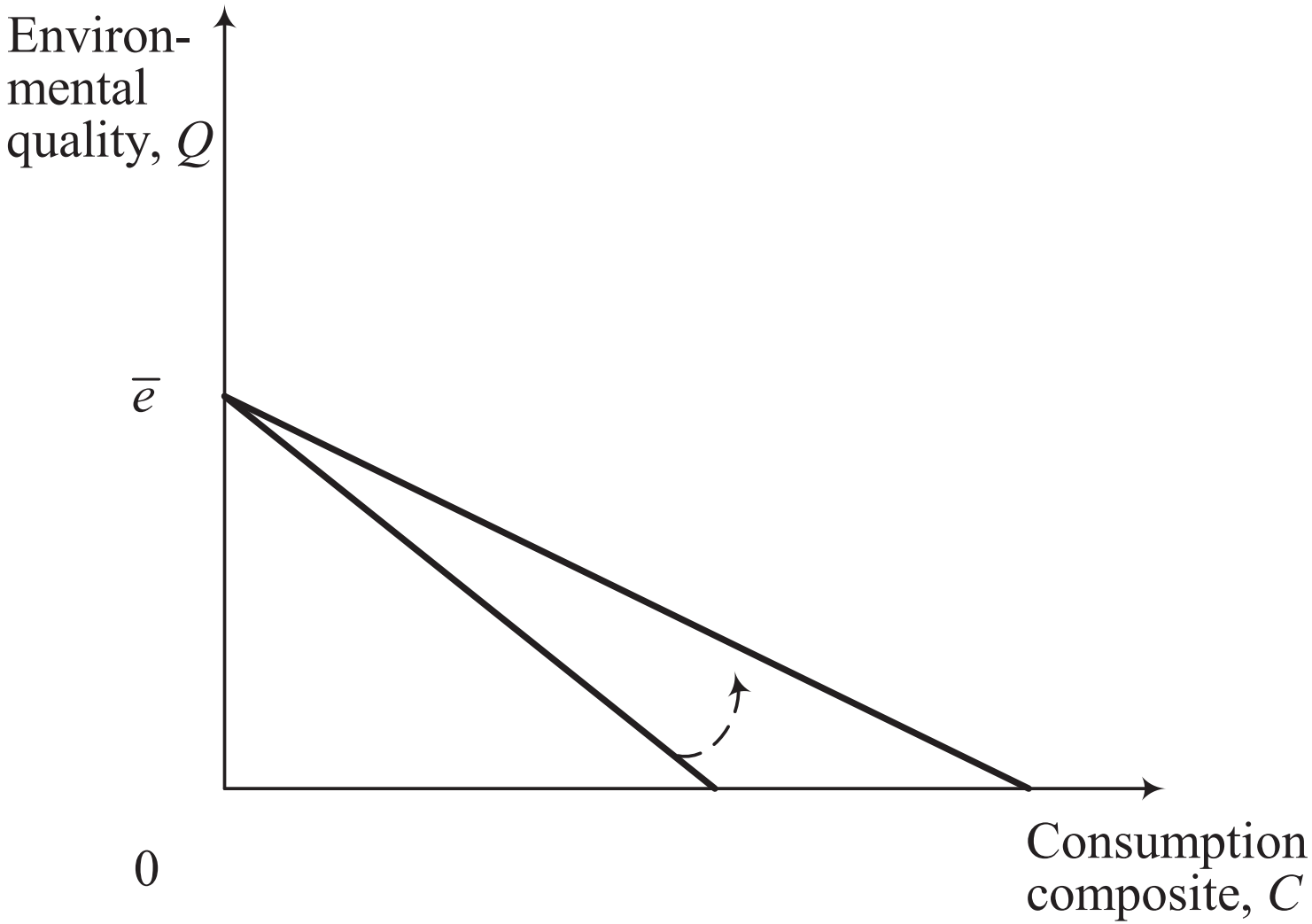
Iceberg trade cost for consumption goods

Lower trade cost – expressway connection; consumption composite – GDP

Planner – local government

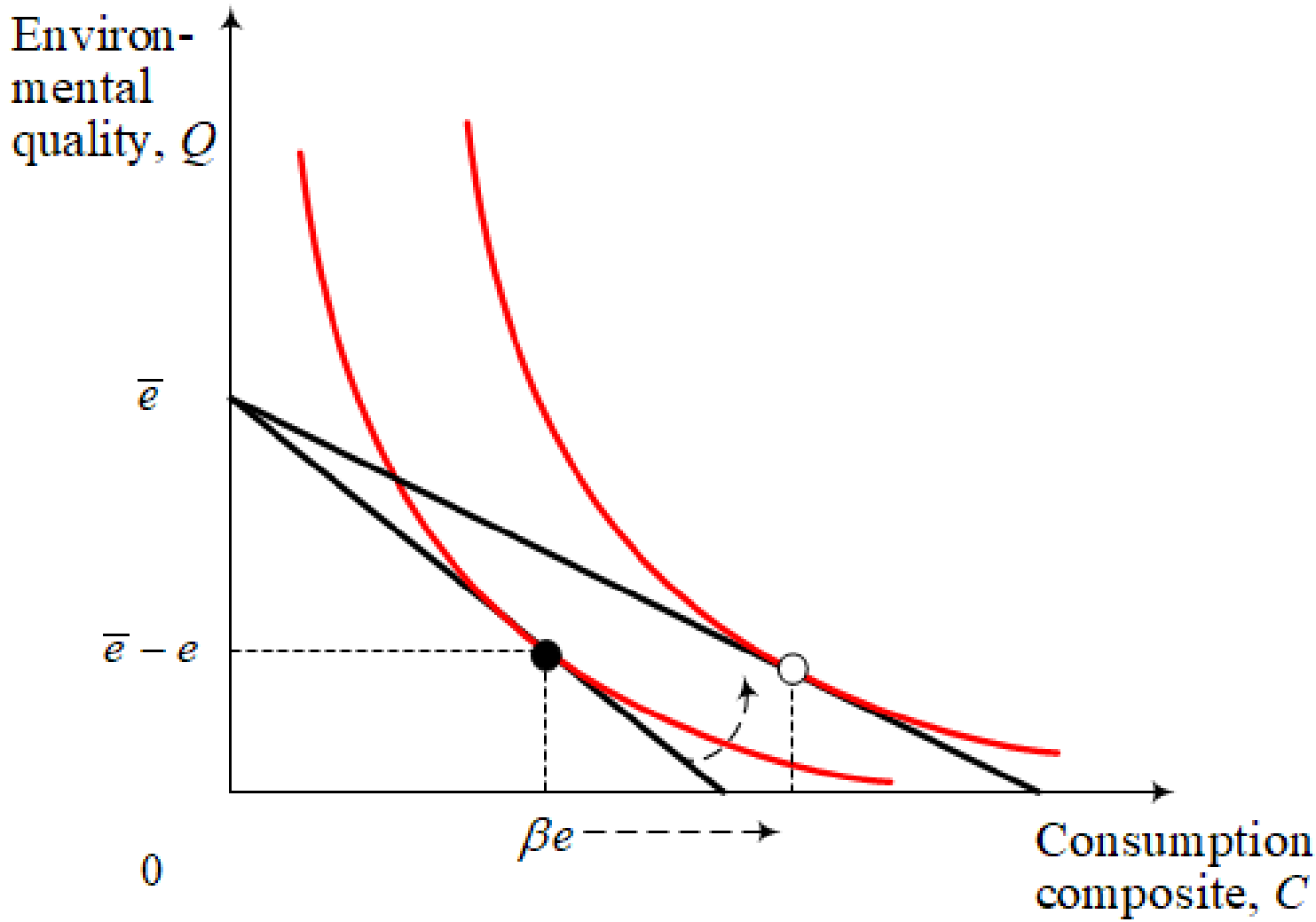
Back to emission graph

Lower trade cost rotates endogenous possibility frontier



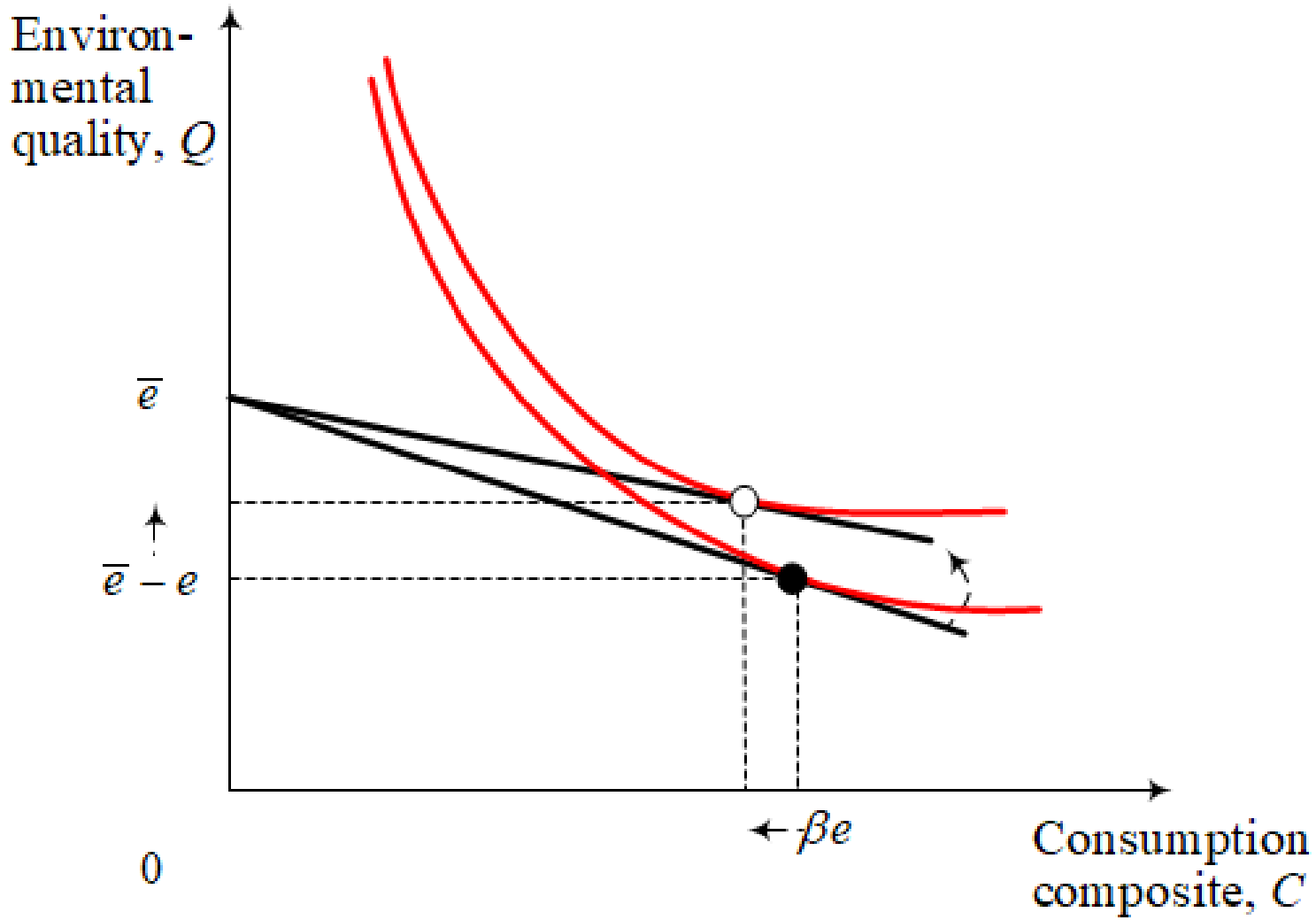
Back to emission graph

Comparative statics at low income



Back to emission graph

Comparative statics at high income



Back to emission graph

Other emission measures

Table 7. The Effects of Expressway Connection on Other Emission Measures

	COD Emissions from Key Polluting Sites (ton, log)		Per capita COD Emissions from Key Polluting Sites (kg, log)		NH3-N Emissions (ton, log)		Per capita NH3-N Emissions (kg, log)		SO2 Emissions (ton, log)		Per capita SO2 Emissions (kg, log)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Expressway	-0.05	3.55***	-0.09	4.03***	-0.08	3.64***	-0.13	4.31***	-0.07	-0.35	-0.08	-0.17
	(0.09)	(0.78)	(0.10)	(0.85)	(0.11)	(1.19)	(0.13)	(1.34)	(0.06)	(0.56)	(0.06)	(0.56)
Expressway*GDP pc		-0.43***		-0.49***		-0.44***		-0.52***		0.03		0.01
(yuan, log, Year 2000)		(0.09)		(0.10)		(0.14)		(0.16)		(0.07)		(0.07)
County FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Obs.	14,551	14,551	14,408	14,408	11,380	11,380	11,249	11,249	13,548	13,548	13,413	13,413
R^2	0.07	0.08	0.07	0.07	0.14	0.14	0.16	0.16	0.13	0.13	0.12	0.12

Notes: This table estimates the heterogeneous impacts of expressway connection on other emission measures. GDP data are deflated, where Beijing-2000 is the base province-year. Standard errors are clustered at county level and reported in the parentheses below the estimated coefficients. *** p<0.01, ** p<0.05, * p<0.1.

As expected, SO_2 performs less well

Back to Channel