China Economic Review 93 (2025) 102460

Contents lists available at ScienceDirect

China Economic Review

journal homepage: www.elsevier.com/locate/chieco

The role of access to expressways on industrial clustering development in China☆

Yunfei Zhang [a], Kevin Chen [a,b,*]

[a] China Academy for Rural Development, Zhejiang University, Hangzhou 310058, China
[b] International Food Policy Research Institute, East and Central Asia Office, Beijing 100081, China

ARTICLE INFO

JEL codes:
D80
L60
O10

Keywords:
Expressway
Industrial clustering
Regional development
China

ABSTRACT

Infrastructure development is considered a key driver of industrial growth. In this study, we used the difference-in-differences method to examine the effect of access to expressways on industrial clustering development. Our findings highlight that access to expressways was associated with a 6.5 % increase in the number of firms in specific industries. The results remained robust after several endogeneity tests. Furthermore, our mechanism analysis demonstrated that the effect was driven by the reduction of trade costs and alleviation of information friction.

1. Introduction

Over the past few decades, China's industrial growth has accelerated significantly, aided by numerous market-oriented reforms that facilitated its transformation into the world's factory (Naughton, 2007). By 2015, China accounted for 18.5 % of global manufactured exports, a dramatic surge from 2.8 % in 1990 (Hanson, 2020). China's rapid industrial expansion has spurred the rise of numerous "specialty cities," many of which are rural towns in coastal areas. The emergence of these specialty cities indicate that China's industrial growth has been largely driven by localized industrial clusters (Long & Zhang, 2012).

According to cluster theory, firms in a geographically defined cluster accrue advantages from clustering, fostering economic growth in the region (Porter, 1990). Extensive empirical evidence has confirmed the positive agglomeration externalities associated with industrial clusters (Glaeser et al., 2010; Guo et al., 2020; Ruan & Zhang, 2009). However, creating a cluster from scratch is a significant challenge for governments (Shang et al., 2025; van der Linde, 2003), making research on the determinants of cluster formation crucial.

Infrastructure investment is considered a key driver of economic development (see Faber (2014) for a review). When transportation infrastructure connects metropolitan cities to peripheral regions, the literature suggests two opposing outcomes: diffusion (Baum-Snow et al., 2017; He et al., 2020; Wu et al., 2023) and concentration (Faber, 2014) of industrial activities. Peripheral regions could benefit from diffusion but might suffer adverse effects from concentration, contingent on empirical findings.

In this study, we examined the role of expressways in industrial clustering development. Industrial clustering is measured by the firm number in a specific industry, given the density of firms in an industry at the county level is an important feature of entrepreneurial clustering in China. We initially employed a difference-in-differences (DID) method to examine the effect of access to

☆ This article is part of a Special issue entitled: 'Cities Inno Dev in China' published in China Economic Review.

* Corresponding author at: China Academy for Rural Development, Zhejiang University, Hangzhou 310058, China.
E-mail address: k.chen@cgiar.org (K. Chen).

https://doi.org/10.1016/j.chieco.2025.102460
Received 27 May 2024; Received in revised form 23 May 2025; Accepted 1 June 2025
Available online 3 June 2025

expressways. Our findings indicate that expressway access can substantially enhance the development of industrial clustering. More specifically, after gaining access to expressways, firms in a specific industry at the county level increased by 6.5 %.

Several studies (Faber, 2014; He et al., 2020; Wu et al., 2023) suggest that expressway routing choices may be affected by unobservable factors, leading to potential endogeneity issues. To address the concern, we employed three methods. First, we used a matching approach to ensure comparability between treatment and control groups. Second, we conducted robustness checks to account for the endogeneity of expressway access over time. Third, we applied Faber's (2014) least cost path spanning tree networks approach to construct an instrumental variable, which we then used in long-difference regression. Our baseline results remained robust after these examinations.

Furthermore, our mechanism analysis demonstrated that the effect was driven by the reduction of trade costs and alleviation of information friction. First, we found expressways improved logistics to reduce trade costs, using road freight volume as a proxy. Second, the results show that expressways reduced information friction by facilitating migrant flows. Importantly, migrant flow is predetermined to ensure exogeneity. Information friction was measured by the surname distance between counties.

This study contributes to the existing literature from three perspectives. First, it contributes to our understanding of determinants of industrial clustering and the agglomeration economy. Prior literature focused on the role of entrepreneurship (Zhu et al., 2019), clan culture (Fan et al., 2023), universities (Lee, 2021), and institutional regulation (Lu & Tao, 2009). This study examined the role of transportation infrastructure on industrial clustering, providing new insights regarding this strand of literature.

Second, this study speaks to the literature on the role of transportation infrastructure. As mentioned, when transportation infrastructure links metropolitan cities to peripheral regions, the literature has proposed two contrasting outcomes for industrial activities: diffusion (He et al., 2020; Wu et al., 2023) and concentration (Faber, 2014). Our empirical results support the diffusion effect. After access to expressways, resource endowments flow more efficiently. Peripheral regions, with lower land and labor costs, are better suited for the development of labor-intensive industrial clusters. Our research provides empirical evidence for understanding the role of infrastructure in promoting balanced regional development.

Third, our study contributes to the literature regarding the role of migrant flows. Several studies found that migrant flow can facilitate regional affinity, promoting economic communication between regions (Burchardi et al., 2019; Parsons & Vézina, 2018). Our results demonstrated that expressways reduced information friction by facilitating migrant flows. Therefore, in inland regions of China, these outflow migrants can act as linkages to developed coastal areas. Our study also offers new insights into this strand of literature.

The remainder of the study is organized as follows. Section 2 introduces the institutional background and related literature. Data are discussed in Section 3. Section 4 shows the empirical specification and baseline results. Robustness checks are implemented in Section 5. In Section 6, we examine several potential mechanisms of access to expressways. The study's conclusions are presented in Section 7.

2. Background and related literature

2.1. Industrial clustering

The concept of industrial clusters originates from Marshall's (1920) *Principles of Economics*, in which he defined the phenomenon of large-scale agglomeration of similar product producers in specific spaces, termed specialized production zones. Porter (1990) later formalized this concept, defining industrial clusters as constellations of geographically proximate companies and related institutions interacting in a particular sector. These clusters are united by shared characteristics and mutual dependencies. Drawing on the foundational theories of Marshall (1920), Arrow (1962), and Romer (1986), Glaeser et al. (1992) consolidated and advanced the understanding of industrial clusters, presenting the MAR model, which emphasizes three primary characteristics of clusters: (a) knowledge spillover in the region, (b) shared intermediate inputs, and (c) a region-specific common labor pool.

In the literature, indicators characterizing the degree of regional specialization are commonly used, such as HHI (Herfindahl-Hirschman Index), Gini coefficient, LQ (location quotient), and Krugman index (Delgado et al., 2014; Ellison et al., 2010; Ellison & Glaeser, 1997). These indicators share the common characteristic of emphasizing specialization. Guo et al. (2020) argued that measuring regional specialization may not be the most suitable method to capture the regional patterns of industrial clustering in China. This is primarily because many heavy industries are state-owned enterprises, and their locations were determined by administrative decisions (Fan & Zou, 2021). These firms are typically large, dominate their region, and have high specialization scores.

Hence, Guo et al. (2020) adopted a different approach to measure industrial clusters, using density-based index (DBI). In this DBI measure, a county is defined as having an industrial cluster if it ranks in the highest 5 % percentile of all counties in terms of firm density for a given industry sector.

The DBI measure for industrial clusters aligns well with the characteristics of clusters in China. Industrial clusters in China often take the form of specialty towns, typically characterized by a concentration of numerous small and medium-size enterprises and family-owned workshops. Multiple case studies of Chinese industrial clusters have highlighted this pattern. For example, Ruan and Zhang (2009) analyzed data from Puyuan Township in 2005, identifying approximately 10,000 firms and family workshops producing more than 500 million cashmere sweaters annually. Similarly, Fleisher et al. (2010) found that Zhili Township had over five thousand workshops engaged in children's garment production in 2005. Therefore, the density of enterprises serves as a key indicator of the presence and strength of these industrial clusters.

However, the DBI method has some shortcomings. First, the 5 % threshold limits cross-year comparisons, because it reflects a relative level of development. An industrial cluster may be redefined as not clustered in the following year if its firm growth is

Table 1
Number of firms in the top five percentile of two-digit industries in 1998.

	Two-digit industry	Five percent threshold
13	Food processing	59
14	Food manufacturing	76
15	Beverage manufacturing	70
16	Tobacco processing	2
17	Textile	56
18	Garment	161
19	Leather, fur, and feathers products	35
20	Timber and wood	75
21	Furniture	32
22	Paper products	59
23	Printing	81
24	Culture, education and sports articles	86
25	Petroleum and coking	19
26	Raw chemical materials	62
27	Medicine	62
28	Chemical fibers	14
29	Rubber and Plastics	128
30	Nonmetallic products	126
31	Ferrous metals	64
32	Nonferrous metals	15
33	Metal products	101
34	General machinery	171
35	Special machinery	134
36	Automobile	58
37	Transport equipment	43
38	Electrical Machinery and Equipment	83
39	Electronic equipment	77
40	Instruments and Meters	60
41	Other manufacturing	22
42	Comprehensive Utilization of Waste	45
43	Repair industry	2

Notes: The number is calculated by authors based on China's firm registration database.

relatively slow, even though the industry may still be significant. Therefore, this measure can pose issues regarding longitudinal comparisons. Second, this method overlooks the heterogeneity of industries. Different industries may have different thresholds. Finally, the definition of the 5 % threshold is also controversial, given it is an arbitrary designation. In the empirical analysis section of this study, we address these three issues when applying the DBI method.

2.2. National Expressway Network Plan

Since the late 1970s, China has undergone significant economic reforms and adopted opening-up policies, leading to rapid and sustained economic growth. However, infrastructure investment, particularly in road networks, initially lagged behind. Prior to the 1990s, China's freight transportation relied heavily on railways and rivers, with road freight accounting for less than 5 % (Baum-Snow et al., 2017; Li et al., 2017). In response to these challenges, the Chinese Ministry of Transportation proposed the "7–5 network" in the late 1980s as part of the National Expressway Network Plan.

The network was designed with seven horizontal and five vertical axes, with the goal of creating a national expressway system featuring approximately 30,000 km of high-grade roads primarily for automobiles by the end of 2007. This plan received approval from the State Council in 1992. According to the 1990 Yearbook of China Transportation & Communications, the 7–5 network aimed to connect 203 of 567 cities, covering around 70 % of the urban population.

The construction process involved two phases: a kickoff phase from 1992 to 1997, followed by rapid development from 1998 to 2007 (Faber, 2014). By the end of 1997, 10 % of the planned mileage had been completed. In 1998, the State Council decided to accelerate construction. By 2003, 91 % of the planned mileage had been completed, and by 2007, all targets were met. This study concentrates on the period from 2000 to 2007, which corresponds to the rapid development of expressway construction.

2.3. Internal migration in China

Massive interregional migration has been a prominent phenomenon in China. In 1958, the household registration system (*hukou*) was formally implemented in China. Every Chinese resident receives a hukou under this system, which can be classed as either agricultural or nonagricultural and connected to a particular region. Prior to 1978, hukou served as an individual's authorization to reside and conduct business in the specified locality. People were forbidden to leave their hukou residence (Kinnan et al., 2018). Since the reform and opening up in the late 1970s, restrictions on internal migration in China have been substantially relaxed. Therefore, the country has witnessed mass migration during the last four decades. Migrants move from rural areas to urban cities and towns and from

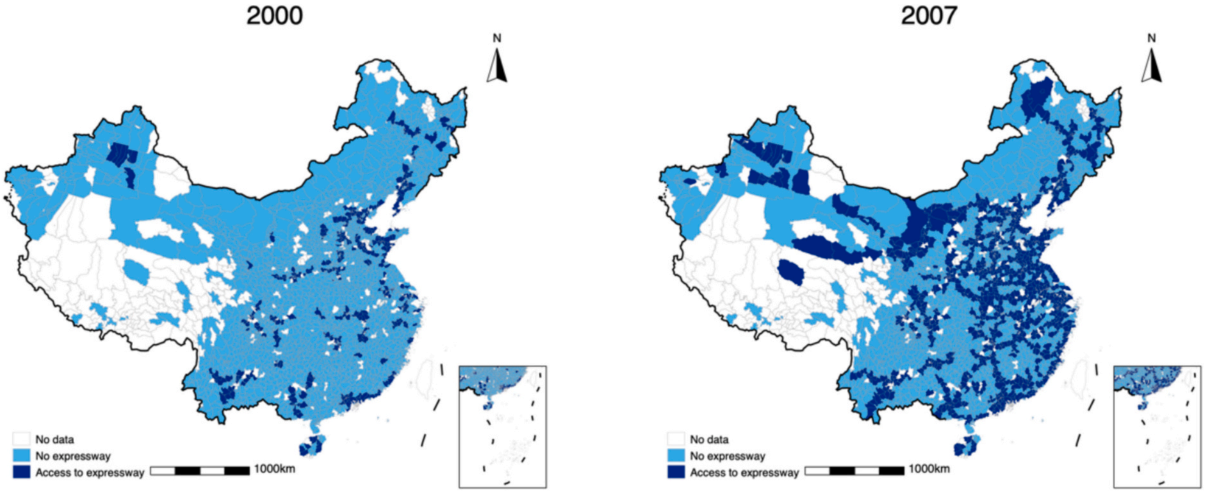

Fig. 1. China's National Expressway Network in 2000 and 2007.
Notes: Fig. 1 shows the geographical distribution of connected counties based on China's National Expressway Network in 2000 and 2007.

inland regions to coastal regions to seek nonagricultural jobs.

However, these migrants still maintain close ties with their hometowns (Zhao, 2002). Migrant workers struggle to obtain urban *hukou* in most Chinese cities. Without this local hukou, they face restricted access to public services, including public housing, social security, and health insurance. Additionally, they experience labor market discrimination (An et al., 2024). Because institutional barriers to permanent settlement in urban areas have not been completely removed, many migrants in China are temporary migrants (Murphy, 1999; Zhao, 2002). These temporary migrants sustain intimate connections with their hometown, frequently shuttling back and forth between their hometown and place of residence multiple times a year. After the access to expressways, the phenomenon may have become more pronounced.

3. Data

3.1. Firm registration database

China's firm registration database was founded by the State Administration for Market Regulation in China. It covers all enterprises registered in China since the reform and opening up in 1978. The database provides indicators including the year of establishment, enterprise location, value of registered capital, industry code, legal representatives, and shareholders. The database includes firm registration information from 1990 to 2020. For consistency with the coverage period of other datasets, we used data from 2000 to 2007 in our study.

China's firm registration database was used to measure the development of industrial clustering—namely, the DBI measurement developed by Guo et al. (2020). To capture an industry's annual firm density, we calculated the firm number at the county-industry-year level based on the year of establishment and location information obtained from the dataset.

Compared to other commonly used enterprise databases (e.g., ASIFP Data), the advantage of China's firm registration database is that the coverage is broad enough and still well represented in the county-industry-year dimension. Given our use of the DBI measure for industrial clusters, we acknowledge this measure's shortcomings and addressed them as follows. First, the 5 % threshold poses limitations for cross-year comparisons. Therefore, we calculated the critical value of 5 % based on the data in 1998 in Table 1.[1] Second, as Table 1 illustrates, some industries have a small number of firms, such as tobacco manufacturing and maintenance. Therefore, these two industries are not covered in the subsequent discussion. Finally, the definition of 5 % is controversial, so we also used the firm number in each two-digit industry as a proxy variable for cluster development.

3.2. China's National Expressway Network

Data on China's National Expressway Network was collected from the China Administrative Spatio-Temporal Expressway Database. The database consists of information on China's expressway network in 2000, 2002, 2003, 2005, and 2007. Because the Spatio-Temporal Expressway Database has gaps, we collected the data from the *Yearbook of China Transportation & Communications (1998–2007)* by hand to fill the gaps. In Fig. 1, we present the geographical distribution of connected counties based on China's National Expressway Network in 2000 and 2007. Fig. 1 shows that in 2000, relatively few counties had access to expressways and they

[1] Because our empirical setting covered 2000 to 2007, we selected 1998 as a preestablished baseline year.

Table 2
Summary statistics of main variables.

Variable	Cell	Obs	Mean	SD
Firm	County-industry-year	234,320	19.58	60.03
Dummy_cluster	County-industry-year	234,320	0.05	0.22
Expressway	County-year	234,320	0.26	0.44
Nightlight	County	234,320	573.36	835.78
Manufacture	County	234,320	452.51	725.76

Notes: "Firm" denotes the firm number for the specific industry in a county. "Dummy_cluster" represents $\mathbf{1}[Cluster_{cit} > 0]$, which equals one if the firm density ranks in the top five percentile among all counties. "Expressway" was a dummy variable and equaled one when the county had access to the expressway network. "Nightlight" denotes the intensity of nightlight at the county level in 1998. "Manufacture" represents the firm number in the manufacturing sector at the county level in 1998.

Table 3
Baseline results based on matched groups.

	Log (Firm+1)		IHS (Firm)	
	(1)	(2)	(3)	(4)
Expressway	0.065***	0.065***	0.076***	0.076***
	(0.011)	(0.010)	(0.013)	(0.012)
County FEs	√	√	√	√
Year-Industry FEs	√	√	√	√
Control*Year		√		√
Observations	234,320	234,320	234,320	234,320
Adjusted R^2	0.650	0.651	0.648	0.650

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. IHS (Firm) is calculated by inverse hyperbolic sine transformation, which equals $log\left(Firm + \sqrt{1 + Firm^2}\right)$. "Firm" denotes the firm number in county c in industry i in year t. "Expressway" represents the access to expressway network in county c in year t. Control variables include the firm number in the manufacture sector and the intensity of night lights in county c in 1998. Standard errors are clustered at the county level.

were not connected in a network. In comparison, the 7–5 network had become clearly established by 2007.

3.3. Other datasets

First, we exploit the Prolonged Artificial Nighttime-Light Dataset of China (1984–2020) to construct nighttime lighting intensity, serving as a proxy for county-level economic performance. Second, geographic distances between counties were computed using geographic information system software. Third, we used China population census 2000 and the 1 % national population sample survey in 2005 (mini-census) to calculate interregional migrants. The calculation method is based on Imbert et al. (2022). Additionally, because the mini-census in 2005 contained individual surname information, we employed it to calculate surname distance based on the method used by Bai and Kung (2022). Fourth, we collected other control variables from the *China City Statistical Yearbook* and the *China County Statistical Yearbook*.

4. Empirical setting and baseline results

We first estimated the following standard DID specification to investigate the effect of access to expressways:

$$Y_{cit} = \beta_0 + \beta_1 Expressway_{ct} + \beta_2 X_c \times Year + \mu_c + \lambda_{it} + \varepsilon_{cit}, \tag{1}$$

where Y_{cit} is the dependent variable measuring the development of industrial clustering. We had two measurements for the dependent variable in county c and industry i in year t: the logarithmic value of the number of firms and a dummy variable, $\mathbf{1}[Cluster_{cit} > 0]$, indicating whether an industry was clustered. $Expressway_{ct}$ represents access to expressways in county c in year t. X_c is a series of control variables, including firm number in the manufacturing sector and the intensity of nighttime lighting in county c in 1998. Subscript c denotes the county, i denotes the two-digit industry, and t denotes the year. ε_{cit} is the error term. Standard errors were clustered at the county level.

However, expressway routing choices may be affected by unobservable factors, leading to potential endogeneity issues. To address this problem, we used three approaches. First, we applied a matching method to ensure comparability between treatment and control groups, thereby better satisfying the parallel trends assumption in the empirical framework. Additionally, we used the matched sample as our baseline. Second, we conducted robustness checks to account for the endogeneity of access to expressways over time. Third, following Faber (2014), we constructed an instrumental variable and employed it in a long-difference regression.

Table 4
Robustness check for timing assumption of expressway completion.

	Log (Firm+1)		IHS (Firm)	
	(1)	(2)	(3)	(4)
Expressway2004	0.084***	0.082***	0.097***	0.095***
	(0.014)	(0.014)	(0.017)	(0.016)
County FEs	√	√	√	√
Year-Industry FEs	√	√	√	√
Control*Year		√		√
Observations	234,320	234,320	234,320	234,320
Adjusted R^2	0.650	0.651	0.648	0.650

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. IHS (Firm) is calculated by inverse hyperbolic sine transformation, which equals $log\left(Firm + \sqrt{1 + Firm^2}\right)$. "Firm" denotes the firm number in county c in industry i in year t. "Expressway2004" represents the access to expressway network in county c in year t and the timing of access for all counties in the treatment group is 2004. Control variables include the firm number in the manufacture sector and the intensity of night lights in county c in 1998. Standard errors are clustered at the county level.

First, we exploited the matching method to make the treatment and control groups comparable. For counties with access to expressways as of 2000, we followed Callaway and Sant'Anna (2021) and excluded them from the sample. The matching covariate was the number of manufacturing firms at the county level in 1998. This covariate had two advantages. First, the variable was predetermined, making it more exogenous. Second, it controlled for the overall industrial development level.

This study calculated descriptive statistics based on the matched sample. Because large cities are the target of expressway connectivity, we followed the literature (He et al., 2020) and dropped the sample of municipalities and provincial capitals. Table 2 presents detailed descriptive statistics for the primary variables.

Industrial clustering was measured using two variables: "firm," which indicated the firm number in a specific industry in a county, and "dummy_cluster," represented as $\mathbf{1}[Cluster_{cit} > 0]$, which equaled 1 if the firm density ranked in the top 5 % among all counties. The average value of "firm" was 19.58, indicating that on average, there were 19.58 enterprises in a specific industry for a county per year. Furthermore, the mean value of "dummy_cluster" was 0.05, which approximately matches the cutoff threshold value.

"Expressway" was a dummy variable and equaled one when the county had access to the expressway network. Moreover, the intensity of nighttime light and the firm number in the manufacturing sector at the county level in 1998 were controlled in the following regression specification. Summary statistics of additional variables used in robustness checks and further discussion are shown in Table A1.

Table 3 reports the results of the baseline regression based on the matched sample (see the results of the unmatched sample in Table A2). As mentioned, the density of firms in an industry at the county level is an important feature of entrepreneurial clustering in China. Therefore, we used the logarithmic value of the firm number plus one as the key dependent variable, as shown in Table 3. Logarithmic transformation can lead to issues with the estimated coefficients (Cohn et al., 2022). Consequently, following Burbidge et al. (1988), we used the inverse hyperbolic sine transformation as an alternative specification for our outcome variable, as shown in Columns 3 and 4. This transformation is defined as $log\left(Firm + \sqrt{1 + Firm^2}\right)$. Additionally, Columns 1 and 3 only incorporate county fixed effects and year-industry fixed effects, whereas Columns 2 and 4 also introduce the interaction of control variables and year dummy variables.

Table 3 demonstrates that various methods of transforming the dependent variable yielded consistent results. The results demonstrate that access to expressways significantly enhanced the development of industrial clustering. Additionally, the coefficients in the first and second columns are nearly identical, suggesting that the matched sample was already highly comparable even without the inclusion of control variables. Specifically, we used the coefficient in Column 2 as the reference point. After expressway access, the firm survival number for the specific industry increased by 6.5 %.

Second, the timing of expressway completion may also involve endogeneity, so we conducted a robustness check. By the end of 2003, completed expressway mileage accounted for 91 % of the goal, indicating that the network was nearly complete. Counties with earlier access to expressways also needed the network's completion to gain greater transportation benefits. Therefore, we set the access to expressways for all counties to 2004. Table 4 shows the robustness check for this timing assumption. The coefficients are positive and significant.

Third, we adopted the least cost path spanning tree networks approach from Faber (2014) to construct an instrumental variable. Specifically, using a minimum-spanning tree algorithm, we generated the expressway network with the smallest length that connected all major cities in the network by 2007. Counties along this path were defined as the straight-line instrumental variable. Table A3 shows that the results for the instrumental variable were also significant.

Finally, to test the impact dynamically, we employed the event study method to estimate the difference between the ex-ante and ex-post treatment groups and the control group flexibly:

$$Y_{cit} = \sum_{\tau=-4,\tau\neq-1}^{+4} \beta_{1\tau} Expressway_{ct}^{\tau} + \beta_2 X_c \times Year + \mu_c + \lambda_{it} + \varepsilon_{cit}. \quad (2)$$

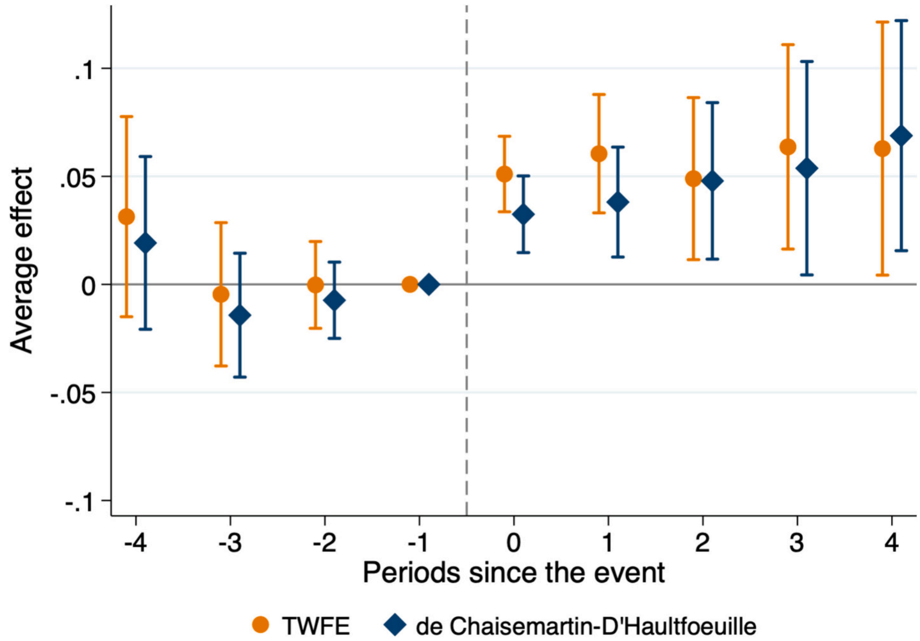

Fig. 2. Common trend test.
Notes: Common trend test of Eq. (2). The figure plots the event study coefficients with corresponding 95 % confidence intervals.

Table 5
Robustness check of alternative specifications.

	(1)	(2)	(3)	(4)
	PPML	Spillover	Add controls	Add controls
Expressway	0.065**	0.064***	0.063***	0.066***
	(0.028)	(0.010)	(0.010)	(0.011)
Distance (0-10 km)		−0.023		
		(0.031)		
Distance (10-20 km)		0.031		
		(0.024)		
County FEs	√	√	√	√
Year-Industry FEs	√	√	√	√
Control*Year	√	√	√	√
Economic Controls			√	
Infrastructure Controls				√
Observations	234,320	234,320	234,320	213,150
R^2	0.629	0.651	0.651	0.645

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. "Expressway" represents the access to expressway network in county c in year t. "Distance (0-10 km)" represents the counties located 10 km from the treatment areas. Control variables include the firm number in the manufacture sector and the intensity of night lights in county c in 1998. Economic Controls include bank deposit balances, government expenditures, GDP per capita, and the industrial share. Infrastructure Controls include the presence of high-speed rail, the presence of an airport, and railway freight volume. Standard errors are clustered at the county level.

In Eq. (2), we replaced the core coefficient β_1 with the leading and lagging terms in event studies $\beta_{1\tau}$. If $\tau < 0$, it represents the $-\tau$ th period before access to expressways, and if $\tau \geq 0$, it represents the τ th period after access to expressways. To improve the precision of the estimation, we combined all periods before four periods and after four periods of treatment.

The recent econometric literature has pointed out that the DID estimator may suffer from the problem of negative weights when treatment effects are allowed to be heterogeneous. Some studies found that traditional two-way fixed-effect estimators can cause bias (Callaway & Sant'Anna, 2021; de Chaisemartin & D'Haultfœuille, 2020; Goodman-Bacon, 2021). To avoid the potential negative weighting problem in this paper, we used the heterogeneity treatment effect robust estimator to conduct further robustness checks in the subsequent analysis.

Fig. 2 illustrates the estimated coefficients and their 95 % confidence intervals of Eq. (2). The results show that the estimated value of β_{1t} was not significantly different from 0, indicating no significant difference between treatment and control groups before the policy

Table 6
Robustness check on different cut-off thresholds of clustering.

	7 % cut-off		5 % cut-off		3 % cut-off	
	(1)	(2)	(3)	(4)	(5)	(6)
Expressway	0.006***	0.006***	0.007***	0.006***	0.002	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
County FEs	√	√	√	√	√	√
Year-Industry FEs	√	√	√	√	√	√
Control*Year		√		√		√
Observations	234,320	234,320	234,320	234,320	234,320	234,320
Adjusted R^2	0.351	0.351	0.321	0.321	0.268	0.268

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Seven/five/three percent cut-off denotes a county was considered to have a particular industry's cluster, if it was in the top seven/five/three percentile of all counties in terms of that industry's firm density. "Expressway" represents the access to expressway network in county *c* in year *t*. Control variables include the firm number in manufacture sector and the intensity of night lights in county *c* in 1998. Standard errors are clustered at the county level.

was implemented. We also referred to de Chaisemartin and D'Haultfœuille (2020) to demonstrate that the results were robust to treatment effect heterogeneity.

5. Robustness check

This section describes several robustness checks used to validate the reliability of our results. First, we used Poisson regression to address potential issues with the dependent variable. Second, we tested for spillover effects to ensure that our identification strategy did not violate the stable unit treatment value assumption (SUTVA). Third, we controlled for omitted variables by adding relevant control variables. Finally, we experimented with different definitions of the dependent variable to assess the sensitivity of our findings to the cutoff threshold. The results from these robustness checks provide further confidence in the consistency and validity of our conclusions.

5.1. *Poisson regression*

Using the log of the dependent variable can be problematic, especially due to the high prevalence of zeros in the data. To address this, we employed inverse hyperbolic sine transformations. Another alternative specification we employed was Poisson pseudo-maximum likelihood (PPML), serving as an additional robustness check for the dependent variable specification. These results are presented in Column 1 of Table 5. The coefficients were both significant and of similar magnitude to those in the baseline regressions. This suggests that different transformations of the dependent variable did not lead to biased estimates.

5.2. *Spillover effect*

The DID method requires the stable unit treatment value assumption to hold. However, transportation infrastructure may have diffusion effects, meaning that control groups are likely to be affected by the policy. This could lead to biased estimates. To address this, we tested for spillover effects, shown in Table 5. In Column 2, we explicitly controlled for spillover effects, following the literature (Huang & You, 2025). We included neighboring areas in the control group as a hypothetical treatment group, using 2004 as the assumed policy shock year. Therefore, the coefficient for expressway access captures the policy's impact after accounting for spillover effects. We found no significant spillover effects in counties 20 km from the treatment areas. This suggests that our identification strategy did not violate SUTVA.

5.3. *Omitted variables*

The baseline specification may suffer from omitted variable bias. To address this, we added a set of control variables. The first group of control variables consisted of county-level socioeconomic indicators, including bank deposit balances, government expenditures, GDP per capita, and industrial share. These indicators are important determinants of industrial cluster development. In Column 3 of Table 5, we further controlled for these variables. The results remained robust after including these controls.

The second group of control variables included other transportation infrastructure. We compiled data on high-speed rail openings, airport openings, and railway freight volume from 2000 to 2007. Prior to 2008, there was no true high-speed rail, except for the Qinshen Passenger Dedicated Line, which began operation in October 2003 and reached a maximum operating speed of 200 km/h. We considered counties through which the Qinshen line passed as the treatment group affected by high-speed rail. Information on airport openings was sourced from publicly available online resources, and all counties in cities with newly opened airports were regarded as the treatment group. Data on railway freight volume were taken from the statistical yearbooks of prefecture-level cities. We included the presence of high-speed rail, presence of an airport, and railway freight volume as control variables in the regression. The results,

Table 7
The effect of expressways on freight volume at the prefecture level.

	(1)	(2)
Expressway_num	0.033*** (0.010)	
Expressway_ratio		0.206*** (0.072)
Prefecture FEs	√	√
Year FEs	√	√
Control	√	√
Observations	1047	1047
Adjusted R^2	0.879	0.878

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the log value of freight volume at the prefecture level. "Expressway_num" represents the number of counties with access to expressways within the prefecture. "Expressway_ratio" represents the proportion of counties with access to expressways within the prefecture. Control variables include GDP per capita, fiscal expenditures, and the industrial share at the prefecture level. Standard errors are clustered at the prefecture level.

Table 8
Heterogeneity check by county location.

	(1)	(2)
Expressway	0.118*** (0.019)	0.275*** (0.030)
Expressway*Distance_city	−0.042*** (0.013)	
Expressway*Distance_industry		−0.020*** (0.003)
County FEs	√	√
Year-Industry FEs	√	√
Control*Year	√	√
Observations	234,320	234,320
Adjusted R^2	0.651	0.663

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the log value of the firm survival number. "Expressway" represents the access to expressway network in county *c* in year *t*. "Distance_city" represents the distance from each county to the nearest provincial capital. "Distance_industry" represents the average distance between a county and pre-2000 cluster counties. Control variables include the firm number in the manufacture sector and the intensity of night lights in county *c* in 1998. Standard errors are clustered at the county level.

shown in Column 4, reveal that the coefficients are statistically significant, with a value of 0.066, which is very close to the baseline coefficient of 0.065.

5.4. Definition of the dependent variable

In this section, we altered the definition of the dependent variable to the dummy variable definition, $\mathbf{1}[Cluster_{cit} > 0]$. A potential concern regarding the definition of the dummy variable is the arbitrary nature of the cutoff threshold (5 %). Different cutoff thresholds may yield inconsistent results. Therefore, we adjusted the threshold to 7 % and 3 %, and the results are presented in Table 6. The coefficients for these cutoffs were positive and significant, although the coefficients for the 3 % cutoff became nonsignificant. As the cutoff threshold decreased, the size threshold of the clusters increased. The findings in Table 6 suggest that the impact of expressways is more pronounced for smaller industry clusters.

6. Further discussion of underlying mechanisms

In this section, we explored how expressway access affects clustering development. First, we analyzed how expressways improve logistics to reduce trade costs, using road freight volume as a proxy. Our results show that expressway access significantly increased road freight volume, especially in prefectures with more counties connected to expressways. We also examined how expressway development interacted with geographic and industrial proximity to reduce trade costs. Second, we investigated how expressways reduced information friction by facilitating migrant flows. The findings indicate that expressway access had a stronger effect in regions with higher migrant outflows and greater surname distance.

Table 9
The effect of expressways on migration flow at the prefecture level.

	(1)	(2)
	Outflow	Inflow
Expressway_num	0.034**	0.046**
	(0.015)	(0.023)
Prefecture FEs	√	√
Year FEs	√	√
Control	√	√
Observations	758	758
Adjusted R^2	0.933	0.813

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the log value of outflow and inflow migrants at the prefecture level. "Expressway_num" represents the number of counties with access to expressways within the prefecture. Control variables include GDP per capita, fiscal expenditures, and the industrial share at the prefecture level. Standard errors are clustered at the prefecture level.

6.1. *Mechanism 1: reducing trade costs*

In this section, we examined the first mechanism of how expressways promote industrial cluster development by reducing trade costs. We first used road freight volume as the dependent variable to examine how expressways affected the mobility of production factors and products. However, due to data limitations, we could only obtain this variable at the prefecture level. Therefore, the independent variables needed to be aggregated to the prefecture level. We defined two explanatory variables. "Expressway_num" represented the number of counties with access to expressways in the prefecture. "Expressway_ratio" represented the proportion of counties with access to expressways in the prefecture.

Table 7 shows that both explanatory variables had a positive and significant effect on road freight volume. The control variables included GDP per capita, fiscal expenditures, and industrial share at the prefecture level. This result supports that one of the key mechanisms is by improving logistics and transportation.

Additionally, we calculated two types of distance as indirect evidence for reducing trade costs. First, we computed the distance from each county to the nearest provincial capital, referred to as "Distance_city." Next, we aimed to capture the distance between each county and counties that had industrial clusters prior to 2000. This reflected the distance between each county and the industrial frontier, represented by "Distance_industry." The variable measured the average distance between a county and pre-2000 cluster counties.

We constructed interaction terms between these two types of distances and expressway development to explore their impact on industrial cluster development. The results are presented in Table 8. First, we found a negative interaction term between "Distance_city" and expressway, indicating that the farther a county is from the provincial capital, the smaller the positive effect of the expressway. Specifically, the coefficient of the interaction term suggests that for every 100 km increase in distance from the nearest provincial capital, the impact of the expressway on the firm number decreased by 4.2 %. This decline represented approximately two thirds of the expressway's effect in the baseline regression.

The interaction between "Distance_industry" and expressway yielded a similar result. As expected, counties closer to the original clusters experienced greater effects. Because industries tend to spread outward from initial clusters, areas closer to these original clusters likely benefited from lower trade costs once they gained access to expressways.

6.2. *Mechanism 2: reducing information friction*

Another mechanism we examined is that access to expressways can boost the interregional labor mobility, thereby reducing interregional information friction. In the section, we first provided direct evidence that access to expressways facilitate interregional migration flow. We referred to Imbert et al. (2022) to construct the migrant flow dataset. They used the Mini-Census in 2005 to create a migration flow matrix between all Chinese prefectures for each year from 2000 to 2005. Based on the method, we constructed the variable of outflow and inflow migrants at the prefecture level. Table 9 shows that expressways promote both outflow and inflow migration. For each additional county with expressway access, outflow migrants increase by 3.4 %, and inflow migrants rise by 4.6 %.

Additionally, we used a triple difference regression model to estimate the impact of expressways on clusters through migrant flow.

$$Y_{cit} = \alpha Expressway_{ct} + \beta Expressway_{ct} \times Migrant_{ci} + X\gamma + \varepsilon_{cit} \tag{3}$$

Migration flow may be an endogenous variable. Therefore, we measured migration flow using the outflow migrant number from 2000. This ensures that migration flow is pre-determined within our analysis period, helping to mitigate concerns about endogeneity. Moreover, we referred to Barwick et al. (2024) for the discussion of endogeneity. They argued that two assumptions must be met to obtain consistent estimates of β in Eq. (3). The first assumption is $\varepsilon_{cit} \perp Expressway_{ct} \mid X$ and the second is $Migrant_{ci} \perp Expressway_{ct} \mid X$. The first assumption is the standard exogeneity assumption, which we have discussed in Section 4. We followed Barwick et al. (2024) to test the second assumption. The results in Table A4 shows that the population outflow in 2000 has no effect on the timing of expressway access, which support the second assumption.

Table 10
The effect of expressways and migrant flow on clustering.

	(1)	(2)	(3)
Expressway	−0.010	1.819***	−0.017
	(0.012)	(0.031)	(0.034)
Expressway*Migrant	0.097***	−0.147	0.179***
	(0.010)	(0.201)	(0.028)
Expressway* Distance_industry		−0.262***	
		(0.034)	
Expressway*Migrant* Distance_industry		0.032	
		(0.029)	
Expressway* Distance_surname			−0.011
			(0.024)
Expressway*Migrant* Distance_surname			0.066***
			(0.018)
County FEs	√	√	√
Year-Industry FEs	√	√	√
Control*Year	√	√	√
Observations	234,320	234,320	234,320
Adjusted R^2	0.652	0.664	0.651

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the log value of the firm survival number. “Expressway” represents the access to expressway network in county *c* in year *t*. “Migrants” denotes the log value outflow migrant number in county c in industry i in 2000. “Distance_city” represents the distance from each county to the nearest provincial capital. “Distance_industry” represents the average distance between a county and pre-2000 cluster counties. “Distance_surname” represents the average surname distance between a county and pre-2000 cluster counties. Control variables include the firm number in the manufacture sector and the intensity of night lights in county *c* in 1998. Standard errors are clustered at the county level.

For the estimation of Eq. (3), we expected the impact of expressway access to be more pronounced in regions with higher populations of outflow migrants. As shown in Table 10, we found that the interaction term between migrant and expressway was significantly positive. Moreover, for every 1 % increase in migrant outflow, the effect of expressway access increased by 1.49 %.[2]

Finally, we provided evidence that expressway access and migrant flow can reduce information friction. However, we lacked direct data to measure information friction. Instead, we used surname distance as a proxy for information friction. Surname distance measures potential biological, cultural, and institutional differences in population characteristics (Bai & Kung, 2022). Differences in human traits and characteristics may act as barriers to industrial transfer between different regions.

The variable “Distance_surname” was calculated as follows. First, we employed the Mini-Census in 2005 to calculate surname distance at the county-pair level based on the method in Bai and Kung (2022). Second, we constructed a new variable, “Distance_surname,” representing the average surname distance between a county and pre-2000 cluster counties.

As shown in Table 10, we introduced a new interaction term, denoted as “Expressway*Migrant*Distance_surname” in the regression model. The results in Table 10 suggest that the effect of access to expressways and migrant flow was more pronounced in counties and industries that were farther apart in terms of surname distance. However, the effect of geographic distance was nonsignificant. This implies that migrant flow may reduce interregional informational friction.

7. Conclusion

In this study, we examined the effect of access to expressways on industrial clustering development in China. We found that access to expressways can stimulate the development of industrial clustering. The empirical findings show that expressways were related to a 6.5 % increase in number of firms at the industry level. Moreover, we demonstrated that the effect was driven by the reduction of trade costs and alleviation of information friction.

Our research supports that investment in transportation infrastructure is crucial for economic growth. Improved transportation enhances the spatial reallocation of economic resources and activities, reshaping the structure and distribution of the regional economy. At the same time, we highlight the importance of social ties. Migrants serve as bridges between regions, driving the development of their hometowns.

Owing to the agglomeration effect of industrial clustering, cluster-based economic development policies have gained popularity among policymakers during the past two decades (Wolman & Hincapie, 2015). This study shed light on how better to implement cluster-based economic development policies by emphasizing the importance of investing in transportation infrastructure. These strategic interventions are pivotal in fostering the development of industrial clusters.

The limitations of this study could be addressed in future research. First, we have not yet validated the classic theory of industrial

[2] This figure was derived from the following calculation: A 100 % increase in migrant outflow led to a change of 0.097 in the effect of the expressway, which was 149 % of the baseline regression value of 0.065. Therefore, for every 1 % increase in migrant outflow, the effect of expressway access increased by 1.49 %.

clusters. Future studies could explore this in three areas: input sharing, labor pooling, and knowledge spillovers. Second, a long-term panel dataset could be used to examine the long-term effects.

Financial support

This study was funded by the Humanities and Social Science Fund of Ministry of Education of China (22JJD790076), the National Social Science Fund of China (No. 21&ZD091; No.22&ZD080), the Major Program of National Fund of Philosophy and Social Science of China (23ZDA035) and the Agricultural Science and Technology Innovation Program (10-IAED-05-2024; 10-IAED-RY-01-2024).

Declaration of competing interest

The authors have no relevant financial or non-financial interests to disclose.

Acknowledgements

We thank the Qiyan Data Technology (Hangzhou) Company for the kind help with the use of China's firm registration database.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chieco.2025.102460.

Data availability

The data that has been used is confidential.

References

An, L., Qin, Y., Wu, J., & You, W. (2024). The local labor market effect of relaxing internal migration restrictions: Evidence from China. *Journal of Labor Economics, 42*(1), 161–200. https://doi.org/10.1086/722620

Arrow, K. J. (1962). The economic implications of learning by doing. *The Review of Economic Studies, 29*(3), 155–173. https://doi.org/10.2307/2295952

Bai, Y., & Kung, J. K. (2022). Surname distance and technology diffusion: The case of the adoption of maize in late imperial China. *Journal of Economic Growth, 27*(4), 569–607. https://doi.org/10.1007/s10887-022-09207-3

Barwick, P. J., Li, S., Lin, L., & Zou, E. Y. (2024). From fog to smog: The value of pollution information. *American Economic Review, 114*(5), 1338–1381. https://doi.org/10.1257/aer.20200956

Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A., & Zhang, Q. (2017). Roads, railroads, and decentralization of Chinese cities. *The Review of Economics and Statistics, 99*(3), 435–448. https://doi.org/10.1162/REST_a_00660

Burbidge, J. B., Magee, L., & Robb, A. L. (1988). Alternative transformations to handle extreme values of the dependent variable. *Journal of the American Statistical Association, 83*(401), 123–127. https://doi.org/10.2307/2288929

Burchardi, K. B., Chaney, T., & Hassan, T. A. (2019). Migrants, ancestors, and foreign investments. *The Review of Economic Studies, 86*(4), 1448–1486. https://doi.org/10.1093/restud/rdy044

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics, 225*(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

de Chaisemartin, C., & D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review, 110*(9), 2964–2996. https://doi.org/10.1257/aer.20181169

Cohn, J. B., Liu, Z., & Wardlaw, M. I. (2022). Count (and count-like) data in finance. *Journal of Financial Economics, 146*(2), 529–551. https://doi.org/10.1016/j.jfineco.2022.08.004

Delgado, M., Porter, M. E., & Stern, S. (2014). Clusters, convergence, and economic performance. *Research Policy, 43*(10), 1785–1799. https://doi.org/10.1016/j.respol.2014.05.007

Ellison, G., & Glaeser, E. L. (1997). Geographic concentration in U.S. manufacturing industries: A dartboard approach. *Journal of Political Economy, 105*(5), 889–927. https://doi.org/10.1086/262098

Ellison, G., Glaeser, E. L., & Kerr, W. R. (2010). What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review, 100*(3), 1195–1213. https://doi.org/10.1257/aer.100.3.1195

Faber, B. (2014). Trade integration, market size, and industrialization: Evidence from China's national trunk highway system. *The Review of Economic Studies, 81*(3), 1046–1070. https://doi.org/10.1093/restud/rdu010

Fan, H., Li, C., Xue, C., & Yu, M. (2023). Clan culture and patterns of industrial specialization in China. *Journal of Economic Behavior & Organization, 207*, 457–478. https://doi.org/10.1016/j.jebo.2023.01.026

Fan, J., & Zou, B. (2021). Industrialization from scratch: The "construction of third front" and local economic development in China's hinterland. *Journal of Development Economics, 152*, Article 102698. https://doi.org/10.1016/j.jdeveco.2021.102698

Fleisher, B., Hu, D., McGuire, W., & Zhang, X. (2010). The evolution of an industrial cluster in China. *China Economic Review, 21*(3), 456–469.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., & Shleifer, A. (1992). Growth in cities. *Journal of Political Economy, 100*(6), 1126–1152.

Glaeser, E. L., Kerr, W. R., & Ponzetto, G. A. M. (2010). Clusters of entrepreneurship. *Journal of Urban Economics, 67*(1), 150–168. https://doi.org/10.1016/j.jue.2009.09.008

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics, 225*(2), 254–277. https://doi.org/10.1016/j.jeconom.2021.03.014

Guo, D., Jiang, K., Xu, C., & Yang, X. (2020). Clustering, growth and inequality in China. *Journal of Economic Geography, 20*(5), 1207–1239. https://doi.org/10.1093/jeg/lbz038

Hanson, G. (2020). Who will fill China's shoes? The global evolution of labor-intensive manufacturing. *East Asian Economic Review, 24*(4), 313–336. https://doi.org/10.11644/KIEP.EAER.2020.24.4.382

He, G., Xie, Y., & Zhang, B. (2020). Expressways, GDP, and the environment: The case of China. *Journal of Development Economics, 145*, Article 102485. https://doi.org/10.1016/j.jdeveco.2020.102485

Huang, K., & You, Y. (2025). Evaluating the impact of the fourth round of China's poverty alleviation program. *American Journal of Agricultural Economics, 107*(2), 583–610.

Imbert, C., Seror, M., Zhang, Y., & Zylberberg, Y. (2022). Migrants and firms: Evidence from China. *American Economic Review, 112*(6), 1885–1914. https://doi.org/10.1257/aer.20191234

Kinnan, C., Wang, S.-Y., & Wang, Y. (2018). Access to migration for rural households. *American Economic Journal: Applied Economics, 10*(4), 79–119. https://doi.org/10.1257/app.20160395

Lee, J. (2021). The role of a University in Cluster Formation: Evidence from a National Institute of Science and Technology in Korea. *Regional Science and Urban Economics, 86*, Article 103617. https://doi.org/10.1016/j.regsciurbeco.2020.103617

Li, Z., Wu, M., & Chen, B. R. (2017). Is road infrastructure investment in China excessive? Evidence from productivity of firms. *Regional Science and Urban Economics, 65*, 116–126. https://doi.org/10.1016/j.regsciurbeco.2017.05.001

van der Linde, C. (2003). The demography of clusters—Findings from the cluster meta-study. In J. Bröcker, D. Dohse, & R. Soltwedel (Eds.), *Innovation clusters and interregional competition* (pp. 130–149). Springer. https://doi.org/10.1007/978-3-540-24760-9_7.

Long, C., & Zhang, X. (2012). Patterns of China's industrialization: Concentration, specialization, and clustering. *China Economic Review, 23*(3), 593–612. https://doi.org/10.1016/j.chieco.2011.09.002

Lu, J., & Tao, Z. (2009). Trends and determinants of China's industrial agglomeration. *Journal of Urban Economics, 65*(2), 167–180. https://doi.org/10.1016/j.jue.2008.10.003

Marshall, A. (1920). *Principles of economics*. MacMillan.

Murphy, R. (1999). Return migrant entrepreneurs and economic diversification in two counties in South Jiangxi, China. *Journal of International Development, 11*(4), 661–672. https://doi.org/10.1002/(SICI)1099-1328(199906)11:4<661::AID-JID609>3.0.CO;2-I

Naughton, B. (2007). *The Chinese economy: Transitions and growth*. MIT Press.

Parsons, C., & Vézina, P.-L. (2018). Migrant networks and trade: The Vietnamese boat people as a natural experiment. *The Economic Journal, 128*(612), F210–F234. https://doi.org/10.1111/ecoj.12457

Porter, M. E. (1990). *The competitive advantage of nations*. Free Press.

Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of Political Economy, 94*(5), 1002–1037.

Ruan, J., & Zhang, X. (2009). Finance and cluster-based industrial development in China. *Economic Development and Cultural Change, 58*(1), 143–164. https://doi.org/10.1086/605208

Shang, S., Zhang, X., & Zhang, Y. (2025). The role of social networks and institutional support in cluster formation. *China Economic Review, 91*, Article 102403. https://doi.org/10.1016/j.chieco.2025.102403

Wolman, H.(H.), & Hincapie, D. (2015). Clusters and cluster-based development policy. *Economic Development Quarterly, 29*(2), 135–149. https://doi.org/10.1177/0891242413517136

Wu, M., Yu, L., & Zhang, J. (2023). Road expansion, allocative efficiency, and pro-competitive effect of transport infrastructure: Evidence from China. *Journal of Development Economics, 162*, Article 103050. https://doi.org/10.1016/j.jdeveco.2023.103050

Zhao, Y. (2002). Causes and consequences of return migration: Recent evidence from China. *Journal of Comparative Economics, 30*(2), 376–394. https://doi.org/10.1006/jcec.2002.1781

Zhu, X., Liu, Y., He, M., Luo, D., & Wu, Y. (2019). Entrepreneurship and industrial clusters: Evidence from China industrial census. *Small Business Economics, 52*(3), 595–616. https://doi.org/10.1007/s11187-017-9974-3