

Wu, Yiyun; Zhu, Xiwei

Working Paper

Why geographic dispersion before its time: Industrial policy and economic geography in the People's Republic of China

ADB Working Paper, No. 633

Provided in Cooperation with:

Asian Development Bank Institute (ADBI), Tokyo

Suggested Citation: Wu, Yiyun; Zhu, Xiwei (2017) : Why geographic dispersion before its time: Industrial policy and economic geography in the People's Republic of China, ADB Working Paper, No. 633, Asian Development Bank Institute (ADBI), Tokyo

This Version is available at:

<https://hdl.handle.net/10419/163132>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-nc-nd/3.0/igo/>



ADB Working Paper Series

WHY GEOGRAPHIC DISPERSION BEFORE ITS TIME: INDUSTRIAL POLICY AND ECONOMIC GEOGRAPHY IN THE PEOPLE'S REPUBLIC OF CHINA

Yiyun Wu and Xiwei Zhu

No. 633
January 2017

Asian Development Bank Institute

Yiyun Wu is an associate professor in the Policy Simulation Laboratory of the Social Sciences Academy, Zhejiang University. Xiwei Zhu is a professor in the School of Economics, Zhejiang University.

The views expressed in this paper are the views of the author and do not necessarily reflect the views or policies of ADBI, ADB, its Board of Directors, or the governments they represent. ADBI does not guarantee the accuracy of the data included in this paper and accepts no responsibility for any consequences of their use. Terminology used may not necessarily be consistent with ADB official terms.

Working papers are subject to formal revision and correction before they are finalized and considered published.

The Working Paper series is a continuation of the formerly named Discussion Paper series; the numbering of the papers continued without interruption or change. ADBI's working papers reflect initial ideas on a topic and are posted online for discussion. ADBI encourages readers to post their comments on the main page for each working paper (given in the citation below). Some working papers may develop into other forms of publication.

ADB recognizes "China" as the People's Republic of China.

Suggested citation:

Wu, Y., and X. Zhu. 2017. Why Geographic Dispersion Before Its Time: Industrial Policy and Economic Geography in the People's Republic of China. ADBI Working Paper 633. Tokyo: Asian Development Bank Institute. Available: <https://www.adb.org/publications/why-geographic-dispersion-before-its-time-prc>

Please contact the authors for information about this paper.

wyyzju@zju.edu.cn, xwzhu@zju.edu.cn

Asian Development Bank Institute
Kasumigaseki Building, 8th Floor
3-2-5 Kasumigaseki, Chiyoda-ku
Tokyo 100-6008, Japan

Tel: +81-3-3593-5500
Fax: +81-3-3593-5571
URL: www.adbi.org
E-mail: info@adbi.org

© 2017 Asian Development Bank Institute

Abstract

This paper investigates the trends and determinants of geographic concentration and industrial specialization in the People's Republic of China (PRC) using interprovincial panel data for the period from 1999 to 2010. It shows that, after 2005, both geographic concentration and industrial specialization began to decrease, resulting in an increased similarity of provincial industrial structure. Industrial policies of provincial governments cause geographic dispersion and inverse specialization. The result is robust when using instrumental variables to deal with possible reverse causality and omitted variable problems. The mechanism behind this is that central government industrial policy, which tends to last for several years, is an important reference document for each provincial planner. This causes the less-developed regions to deviate from their comparative advantages, resulting in a combination of insufficient geographic concentration and inverse specialization in the PRC.

JEL Classification: L59, L60, R12

Contents

1.	INTRODUCTION	1
2.	DATA AND MEASUREMENT	2
2.1	Data	2
2.2	Measurement	3
3.	THE EVOLUTION OF ECONOMIC GEOGRAPHY IN THE PRC: STYLIZED FACTS	4
3.1	Geographic Concentration	4
3.2	Industrial Specialization and Isomorphism	6
4.	ECONOMIC GEOGRAPHY IN THE PRC: COMPARISON AND HYPOTHESES	9
4.1	Theoretical and Economic Background	9
4.2	Hypotheses	10
5.	INDUSTRIAL POLICY AND ECONOMIC GEOGRAPHY: EMPIRICAL ANALYSIS	12
5.1	Methodology	13
5.2	Results	13
5.3	Panel Regression: Further Test	17
6.	CONCLUSIONS	19
	REFERENCES	21
	APPENDIX 1	24
	APPENDIX 2: TIME SERIES FOR GEOGRAPHIC CONCENTRATION ($CONC_i$) AND THE RELATIVE RATE OF CHANGE	26
	APPENDIX 3: TIME SERIES FOR INDUSTRIAL SPECIALIZATION ($SPEC_i$) AND THE RELATIVE RATE OF CHANGE	28
	APPENDIX 4: DEFINITIONS AND SUMMARY STATISTICS OF KEY VARIABLES	30

1. INTRODUCTION

The People's Republic of China (PRC) experienced convergence of provincial industrial structures with the inverse process of specialization between urban and rural areas at the very beginning of its market-oriented economic reform in the mid-1980s. This phenomenon was termed a "skewed pyramid" to warn against unbalanced development (Ma 1989). Twenty years later, however, geographic concentration and industrial specialization in the PRC, which had increased rapidly and continuously since the late 1980s, began to decline. What caused the PRC's recent geographic manufacturing dispersion? This paper investigates the evolution of the PRC's economic geography and concludes that the spatial dispersion of manufacturing firms and more severe industrial isomorphism among provinces are closely related to the local governments' development policies. The underlying mechanism is quite simple: to apply for special privileges from the central government, local governments have a strong incentive to mimic the former's industrial policies when setting up their own.

There are many papers focusing on whether geographic concentration and industrial specialization in the European Union (EU) and the United States (US) change as per the predictions of New Economic Geography (Krugman 1991a; Kim 1995; Ellison and Glaeser 1997; Amiti 1999; Midelfart-Knarvik et al. 2002). As for the PRC, although Young (2000) previously inspired researchers' interests in the reshaping process of its post-reform economic geography, not much empirical work emerged until high-quality statistical data became available in the past few decades (Wen 2004; Bai et al. 2004; Fan 2004, 2008; Lu and Tao 2006, 2007, 2009; Huang and Li 2006). Compared with the EU's and US's market-oriented evolution, the PRC's evolving geographic manufacturing distribution is triggered not only by agglomeration and dispersion forces suggested in Wen (2004), but also by local protectionism and market segmentation stemming from fiscal decentralization (Young 2000). Bai et al. (2004), Lu and Tao (2007), and Lu and Tao (2009) showed that industries with higher after-tax profit and share of state ownership have less geographic concentration, thus supporting their hypothesis that local protectionism hinders geographic concentration. Jin et al. (2006) also showed that government interventions impede the PRC's industrial agglomeration. In short, local government acts as a troublemaker that distorts the spatial reallocation of industrial activities and causes inefficiency. Meanwhile, large empirical studies also support that market-oriented activities are the main determinants reshaping the PRC's economic geography and its manufacturing industries' agglomeration in the coastal region during 1985–2004 (Wen 2004; Bai et al. 2004; Jin et al. 2006; Lu and Tao 2006, 2007, 2009; Fan 2004, 2008; Luo and Cao 2005; Wang and Wei 2007). Due to data availability, previous studies could only explore the spatial redistribution of manufacturing industries from 1978 to 2005. Lack of further empirical work brings two obvious limitations for understanding the whole story. First, it is impossible to fully evaluate the central government's regional development strategies since 2000 (such as the Western Development Drive, Revitalization of Northeast PRC and Rise of Central PRC). Second, local government is always considered a self-interested troublemaker that undermines spatial efficiency. But local government's reaction is rational under the specifically vested interest pattern; thus, perhaps the central government should also bear some responsibility for the distortion.

Empirical works on the EU and the US (Kim 1995; Miderfart–Knarvik et al. 2000; Aiginger and Davies 2004) indicate that, in the long term, geographic concentration of manufacturing industries experiences a bell curve during the process of regional integration. If this is a common rule, it is important to answer which position on the curve the PRC is on right now. Meanwhile, if geographic concentration of manufacturing industries has ceased its uptrend, it is unfair to blame local governments for the spatial distortion. However, Lu and Tao (2007) revealed that, compared to the US and other economies, the PRC's four-digit-level manufacturing industries are far much less concentrated. Thus, if geographic concentration in the PRC has already started decreasing as Fan and Li (2011) and Wang et al. (2010) showed, it would be interesting to see which forces determine the change: market forces or local government inventions?

Few papers have evaluated the active role local government might play during industrial development and upgrading, which is actually highly stressed by the new structural economics (Lin 2012). In reality, it is common for PRC local governments to leverage development planning and preferential policies to widely affect some specific industries, which is a main influencing factor. Song and Wang (2013) were the first to use local government 5-year plans to study the relation between preferential industrial policy and local productivity. By contrast, this paper pays particular attention to the role of local government 5-year plans in determining provincial industrial structure, and aims to explain the PRC's evolving economic geography.

In this paper, we constructed a panel data of 37 two-digit industries in 31 PRC provinces over the period 1999–2010 based on national and provincial statistical yearbooks, which enables focusing on geographic concentration and industrial specialization over a longer and more recent time period than the relevant literature. The findings show that non-market forces significantly affect the spatial distribution of economic activities and support our hypotheses about the intervention of local governments through preferential policies. Although local governments directly cause lower spatial concentration and more highly assimilated regional industrial structures, the central government should be blamed for inducing them to carry out similar industrial plans. The rest of the paper is organized as follows. In section two, we discuss data source and indices of geographic concentration and industrial specialization. Section 3 provides stylized facts of recent PRC geographic concentration and industrial specialization. Section 4 proposes theoretical hypotheses for testing. Section 5 presents econometric testing of the hypotheses and assesses robustness of the results. Section 6 concludes.

2. DATA AND MEASUREMENT

2.1 Data

The existing literature usually uses two major sets of industrial statistics. The first is those published by National Bureau of Statistics, such as *China Industry Economy Statistical Yearbook*, *China Statistical Yearbook*, or *China Economic Census Yearbook*. Though published statistical yearbooks are easily accessible, the coverage is often limited. The second is unpublished statistics, such as data from the *Annual Survey of Industrial Firms* (ASIF) conducted by the PRC's National Bureau of Statistics. Unpublished statistics sometime provide a wider range of details, such as firm-level data. However, they are hard to access, which prevents most researchers from

performing repeat studies. In addition, common data problems like gaps or typos, or lack of checking methods also undermine research.¹

This paper constructed a panel data of 37 two-digit industries (recycling and mining of other ores are excluded because of missing data) in 31 PRC provinces (including provinces, autonomous regions, and municipalities; hereafter, provinces) for the period from 1999 to 2010.² In general, gross output value, value-added, and employment data are the most popular variables used to construct economic geography indices. We agree with Bai et al. (2004) about redundant personnel where employment data may suffer from the surplus labor problem particularly prevalent in state-owned enterprises and lead to biased measurement. We also agree with Wen's 2004 opinion about gross output value and value-added when constructing spatial Gini coefficients. Besides, as Krugman and Venables (1995) and Venables (1996) argued, the availability of intermediate input also influences firms' selection of location. Taking all these into consideration, this paper uses gross output value to calculate industrial geographic distribution. To ensure comparability over time, we adjust gross output in terms of ex-factory price indices of industrial products by sector from *China Urban Life and Price Yearbook* (2011), using 2003 as the base year.³ All data are from *China Industry Economy Statistical Yearbook*, *China Statistical Yearbook*, *China Economic Census Yearbook*, and provincial statistical yearbooks. The discrepancy between the aggregated gross domestic product (GDP) of provinces and national GDP is often used to question the reliability of PRC statistics; however, it is found that the industrial statistics from provincial and national yearbooks are highly consistent.⁴

2.2 Measurement

There are two aspects of structural change when dealing with economic geography: geographic concentration and industrial specialization. In previous studies, the Herfindahl–Hirschman Index (HHI), Hoover (GINI) coefficient, EG Index, Spatial Dispersion Index, and Entropy Index were mostly used to measure geographic concentration and industrial specialization. Considering the pros and cons of the above indices,⁵ we choose a comprehensive measurement that employs the geographic concentration Entropy index and industrial specialization Entropy index proposed by Aiginger and Davies (2004) as the major measuring indices. Second, the Krugman Index (Krugman 1991a) is introduced to reflect the difference of industry composition among provinces and its shifts. Finally, in line with previous studies and to test the robustness of the study, the Hoover coefficient is also employed to calculate the PRC's geographic concentration and industrial specialization.

¹ For example, *China Statistical Yearbook* (2006) reported that there were 271,835 industrial enterprises above designated size and 251,499 manufacturing enterprises above designated size in 2005. However, Lu and Tao (2009) claimed the data to be 265,739 and 246,379, less by 6,096 and 5,120 respectively. More detailed discussions about the potential problems in ASIF see Nie et al. (2012).

² The selection of the period from 1999 to 2010 is based on two reasons: first, since current studies have fully discussed the evolution of economic geography in the PRC from the 1980s to 2003, this paper focuses on new trends since 2004; second, the statistical criteria saw major revisions in 1998 and 2010, respectively, making it inappropriate to compare data before and around 1998 or 2010. More detailed explanations of data sources and collection can be found in Appendix I.

³ Since *China Urban Life and Price Yearbook* lacks the price index of the agricultural products and byproducts processing industry, printing industry, and ordinary machinery manufacturing industry before 2002, this paper replaces the relevant data by indices of food manufacturing industry, paper and paper products industry, and, for special purposes, the equipment manufacturing industry of that year.

⁴ See Appendix I for details.

⁵ We will not go into details about the pros and cons of different indices. See Cowell (1995) and Palan (2010).

For convenience, x_{ir} represents the output of industry i in province r , where i stands for a given industry ($i = 1, 2, \dots, I$) and r stands for a given province ($r = 1, 2, \dots, R$). $x_{i.}$ represents gross output of industry i , $x_{.,r}$ for gross industry output in province r , and $x_{.,.}$ for gross PRC industrial output. $w_r \equiv x_{.,r}/x_{.,.}$ stands for the share of gross industry output in province r and $v_i \equiv x_{i.}/x_{.,.}$ stands for the share of gross output of industry i .

2.2.1 Entropy Index

As proposed by Aiginger and Davies (2004), the Entropy index of geographic concentration is defined as $CONC_i \equiv -\sum_r (x_{ir}/x_{i.}) \ln(x_{ir}/x_{i.})$. If a given industry i has equal output in all r provinces, $CONC_i = \ln R$. Alternatively, if its output is completely from one province, $CONC_i = 0$. Generally, $CONC_i$ increases the more evenly the industry i spreads its output across provinces; it is therefore an inverse measure of concentration. The Entropy index of industrial specialization is defined as $SPEC_r \equiv -\sum_i (x_{ir}/x_{i.}) \ln(x_{ir}/x_{i.})$. Analogous to $CONC_i$, $SPEC_r$ is an inverse measure of specialization which must lie between $[0, \ln I]$. To capture average levels of specialization and concentration of industries in a country, Aiginger and Davies (2004) further proposed $TYP\text{SPEC}$ and $TYP\text{CONC}$, defined as weighted averages with the weights being, respectively, the province (w_r) and industry (v_i) shares of gross output. Thus, $TYP\text{SPEC} \equiv \sum_r w_r \cdot SPEC_r$ and $TYP\text{CONC} \equiv \sum_i v_i \cdot CONC_i$.

2.2.2 Krugman Index

In Krugman (1991a), the structural difference of industries between two provinces is defined as $k_{r,r'} \equiv \sum_i |x_{ir}/x_{i.} - x_{ir'}/x_{i.}|$, where $k_{r,r'}$ lies between $[0, 2]$. The more specialized the industries in province r and r' , the closer $k_{r,r'}$ is to 2; in contrast, $k_{r,r'} = 0$ if the two provinces are completely symmetric.

2.2.3 Hoover Coefficient

The Hoover coefficient (Hoover 1936) can measure geographic concentration of a given industry i (Hoover coefficient of industry) and specialization of a given province r (Hoover coefficient of localization). Take the Hoover coefficient of industry as an example: first, calculate the location quotient for industry i for all regions, where $LQ_{ir} \equiv (x_{ir}/x_{i.}) / (x_{.,r}/x_{.,.})$. Then rank the location quotients in descending order and calculate the cumulative percentage of output in industry i over regions (y -axis). Finally, calculate the cumulative percentage of output in total over regions (x -axis). The concentration curve of industry i is thus formed. The coefficient of industry is defined as the area between the 45-degree line and the concentration curve divided by the entire triangular area. The localization curve is analogously constructed. By definition, both Hoover coefficients lie between $[0, 1]$. When the coefficient increases, the level of geographic concentration or industry specialization improves.

3. THE EVOLUTION OF ECONOMIC GEOGRAPHY IN THE PRC: STYLIZED FACTS

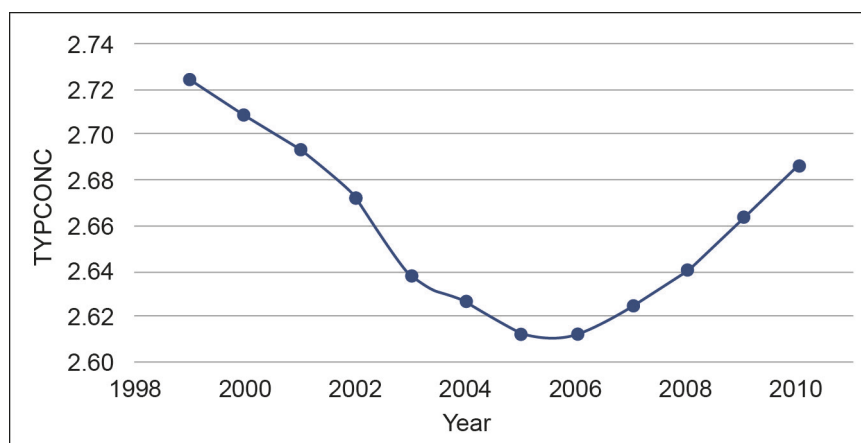
3.1 Geographic Concentration

As described in Section 2, the average geographic concentration ($TYP\text{CONC}$) is calculated using output data for the 37 two-digit industries over the period 1999–2010. The time trend of $TYP\text{CONC}$ is showed in Figure 1 (see Appendix II for $CONC_i$ by industry). From 1999 to 2010, the geographic concentration of industries first increased and then decreased. In 1999, the value of $TYP\text{CONC}$ was 2.7242; from then on, it

went down steadily to 2.6123 in 2005, showing that the concentration level constantly rose during this period, which is consistent with most of the previous studies (e.g., Bai et al. 2004; Wen 2004; Luo and Cao 2005; Lu and Tao 2006). However, after 2005, *TYPCONC* increased steadily to 2.6867 in 2010, which reflects a declining concentration level.

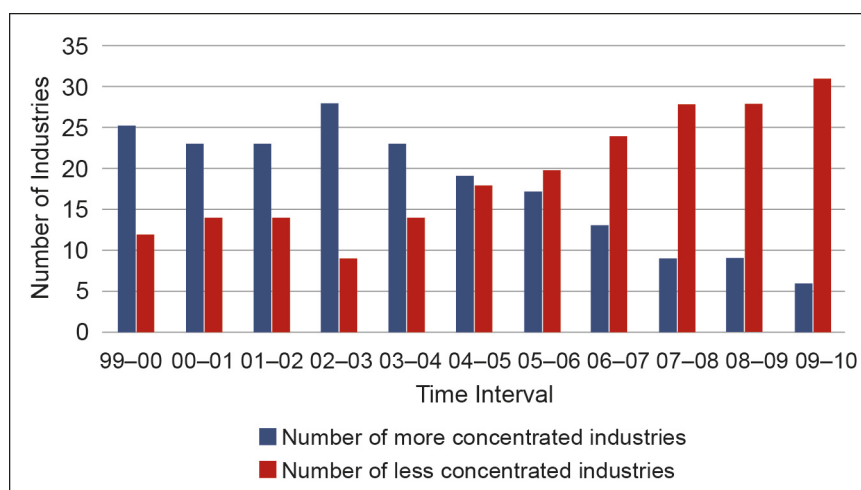
By definition, *TYPCONC* is the weighted average of industrial concentration, with the weights being shares of gross output. When looking into industries, from 1999 to 2005, communication equipment, computers, and other electronics contributed most to the growth of *TYPCONC*; from 2005 to 2010, transport equipment made the largest contribution to its descent. To overcome the influence from those heavily weighted industries, we further probed the change of each industry in neighboring years from 1999 to 2010, as shown in Figure 2. Since 2002, more and more industries started to become dispersed; then, until 2005, more decreasing industries were actually found than increasing ones. In detail, from 1999 to 2005, there were only seven industries whose concentration decreased, while from 2005 to 2010, there were as many as 31 industries that became more dispersed rather than concentrated.

Figure 1: Geographical Concentration over Time



Source: Prepared by the authors.

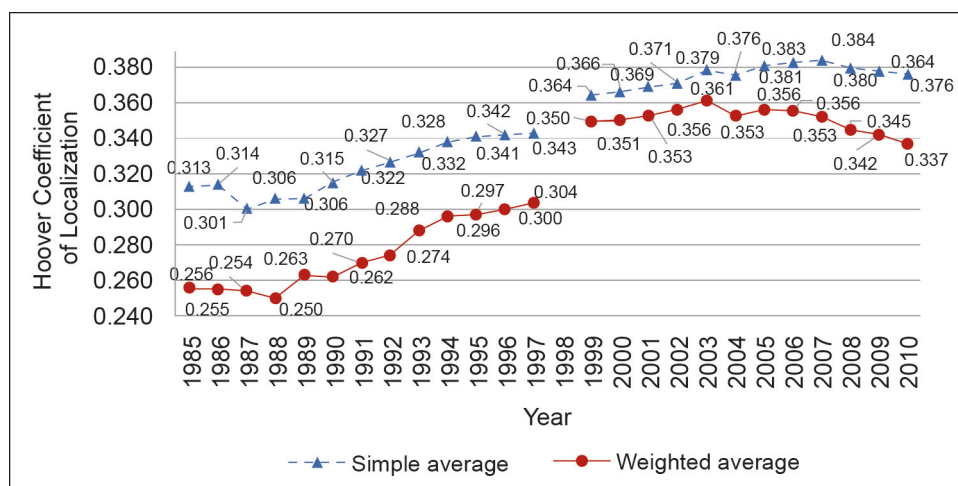
Figure 2: Concentration over Time by Sector



Source: Prepared by the authors.

To test the robustness and compare the results with previous studies, we calculated the Hoover coefficient of localization of 37 industries over the period of 1999–2010.⁶ Similar to the trend of *TYPCONC*, before 2005, the Hoover coefficient of localization did increase by 4.72% (by simple average) or 1.85% (by weighted average), whereas after 2005 it decreased by 1.26% (by simple average) or 5.44% (by weighted average). Furthermore, to trace the time line, we drew the concentration curve of PRC industries from 1985 to 2010 on the basis of Bai et al.'s 2004 data based on 32 industries in 29 provinces from 1985 to 1997 (see Figure 3).⁷ It can be easily seen in the figure that, since the 1980s, the spatial distribution of industries in the PRC initially became more and more concentrated and then subsequently dispersed, which enriches the findings by Wang et al. (2010), and Fan and Li (2011). The results persist even if we focus only on manufacturing.

Figure 3: Time Trend of Average Hoover Coefficient of Localization



Source: Prepared by the authors.

3.2 Industrial Specialization and Isomorphism

Table 1 reports the relative rate of change on industrial specialization from 1999 to 2010, including the Entropy index of the 31 provincial-level divisions ($SPEC_r$), the rate on *TYPESPEC*, and the Hoover coefficient of specialization (by simple average and weighted average). The table shows that the average specialization level also takes a bell-shaped curve within the past 12 years. Twenty-two provinces experienced an increase in specialization from 1999 to 2005, resulting in a decrease in *TYPESPEC* to 2005 of 3.98%; in contrast, after 2005, only 10 provinces had an increase in

⁶ Due to length limit, Appendix II does not report the Hoover coefficient of each industry. Interested readers can contact us for the data. Since the Hoover coefficient of localization and the Entropy index are different in design, the results of the geographic concentration of a given industry are not exactly the same; however, the general trend judged from the two indices is the same.

⁷ Note that data used in Bai et al. (2004) have slightly different statistic criteria from this paper. In addition, they incorporate 32 industries and merge Hainan with Guangdong, and Chongqing with Sichuan, which are also different from here. Though these differences make a gap between data before and after 1998, the general trend of industrial geographic concentration in the PRC remains the same. Similarly, according to Lu and Tao's 2006 results of the Hoover coefficient of localization based on the output of two-digit provincial manufacturing industries, we can have the changing curve of manufacturing geography from 1985 to 2003, which also takes the trend of a bell-shaped concentration curve.

specialization, resulting in an increase of 0.53%. The change of Hoover coefficient confirms the trend shown by *TYPESPEC*, which is also a bell-shaped curve, with 2005 as a turning point (see also Figure 4).⁸

Table 1: Relative Rate of Change on Industrial Specialization (*SPEC_r*) from 1999 to 2010 (%)

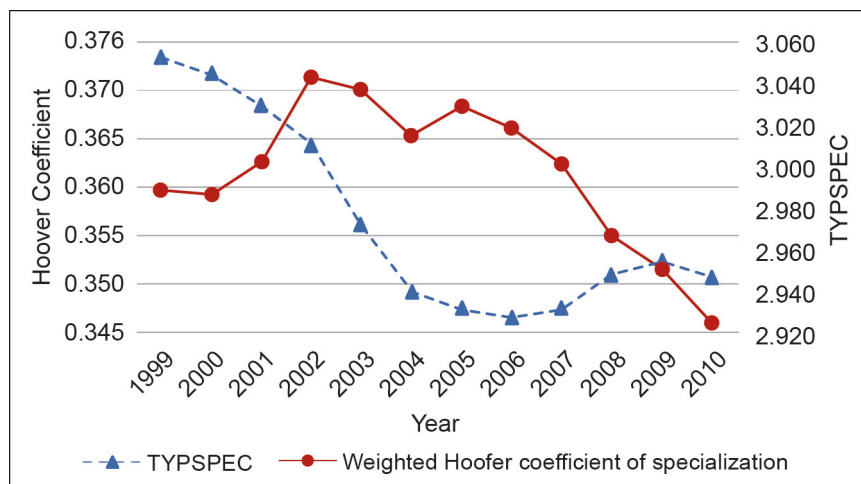
Province	99–10	99–05	05–10	Province	99–10	99–05	05–10
Shanghai	–14.81	–9.30	–6.08	Fujian	–1.06	–3.06	2.06
Guangdong	–12.74	–12.69	–0.06	Guizhou	–0.96	–2.81	1.90
Tibet Autonomous Region	–10.55	–13.68	3.63	Shaanxi	–0.45	–0.43	–0.02
Hainan	–10.26	–5.95	–4.59	Henan	0.21	–0.92	1.15
Beijing	–8.23	–7.93	–0.32	Jiangxi	0.47	0.37	0.09
Jiangsu	–8.12	–6.25	–1.99	Ningxia	0.83	–0.82	1.66
Hebei	–8.04	–9.35	1.44	Hunan	2.44	1.71	0.72
Chongqing	–7.32	–8.87	1.70	Liaoning	2.70	0.44	2.25
Tianjin	–6.95	–10.73	4.24	Sichuan	3.52	1.02	2.48
Hubei	–6.77	–6.08	–0.73	Inner Mongolia	8.14	2.29	5.71
Jilin	–5.23	–8.89	4.01	Yunnan	8.63	3.14	5.32
Shanxi	–4.26	–6.46	2.35	Qinghai	11.65	4.52	6.83
Guangxi	–3.58	–6.49	3.11	Xinjiang	28.64	11.61	15.26
Anhui	–2.83	–2.31	–0.54	Heilongjiang	31.20	20.16	9.19
Shandong	–2.53	–0.97	–1.57	<i>TYPESPEC</i>	–3.47	–3.98	0.53
Zhejiang	–1.68	–1.01	–0.67	HOOVER-a	–0.45	5.19	–5.36
Gansu	–1.45	–7.77	6.86	HOOVER-b	–3.78	2.45	–6.07

Note: HOOVER-a and HOOVER-b refer to simple average and weighted average Hoover coefficient of specialization. For full time series, see Appendix III.

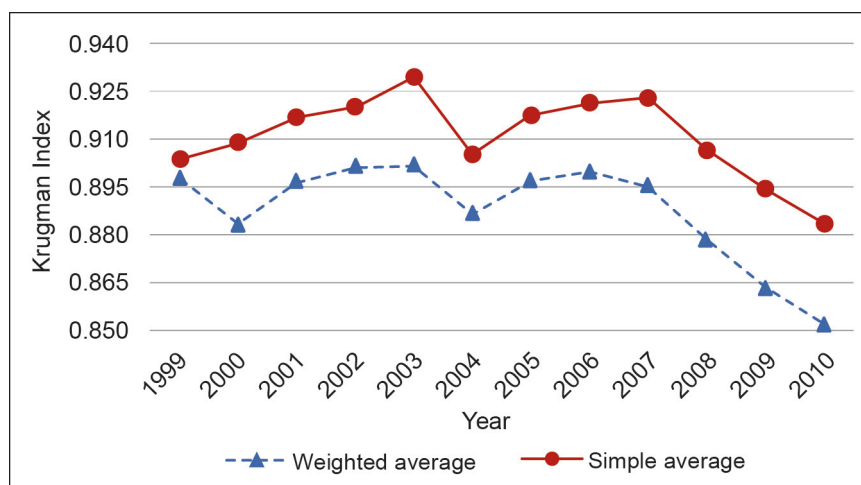
Source: Authors' calculations.

Since *SPEC_r* only measures the change of industry structure within one single province, it is natural to introduce the Krugman Index (KI) to check what happens simultaneously between provinces. As shown in Figure 5, from 1999 to 2010, the interprovincial KI decreased by 2.23% in simple average and by 5.09% in weighted average. Thus, provincial industrial structure in the PRC converged in general from 1999 to 2010. To further confirm the facts, we divided the 31 provinces into four economic regions whose boundary line is set by the National Bureau of Statistics—the East Region (including 11 provinces), the Central Region (including six provinces), the West Region (including 12 provinces), and the Northeast Region (including three provinces). Table 2 reports the results of interprovincial and inter-regional KI. From 1999 to 2010, specialization of the whole country declined sharply, resulting in much more uniform industrial structure across regions. Besides, the similarity of provincial industrial structure is much more severe within the East and the Central Region, than within the West and Northeast Regions. Finally, the similarity of provincial industrial structure within the East and the Central Region is easing, whereas that within and between the Central and the Northeast Region is getting more severe, resulting in a decreasing specialization of the country as a whole.

⁸ Although the results based on the Entropy index and the Hoover coefficient are not exactly the same, their conclusion of the changing trend of industrial specialization in the PRC at the two phases (from 1995 to 2005 and from 2005 to 2010) is the same. Technically speaking, the Entropy index is more sensitive to underdeveloped regions and the Hoover coefficient to developed regions.

Figure 4: Regional Specialization over Time

Source: Prepared by the authors.

Figure 5: Krugman Index over Time

Source: Prepared by the authors.

Table 2: Interprovincial Krugman Index and Its Relative Rate of Change from 1999 to 2010

	E-E	E-C	E-W	E-N	C-C	C-W	C-N	W-W	W-N	N-N	National
1999	0.6256	0.7471	0.9577	0.9280	0.5865	0.7828	0.8424	0.9188	0.9097	0.9405	0.8979
2010	0.6395	0.7280	0.9057	0.8081	0.6034	0.6836	0.7228	0.7874	0.7848	0.8551	0.8521
Change Rate	2.23%	-2.56%	-5.44%	-12.91%	2.87%	-12.67%	-14.20%	-14.30%	-13.73%	-9.09%	-5.09%

Note: E means East Region, C means Central Region, W means West Region and N for Northeast Region; so that E-E refers to the weighted average Krugman Index of provinces from within the East Region; and E-C refers to the weighted average Krugman Index of provinces from the East Region and the Central Region; and it is similar to read the rest notations.

Source: Prepared by the authors.

In summary, two stylized facts of the PRC's economic geography can be concluded as follows: first, 2005 marked the ending of 20 years' increase in geographic concentration

following the economic reform and opening; after 2005, geographic concentration of industries starts to decrease. Second, industrial specialization also adopts a bell-shaped path and the provincial industrial structure is converging as a whole. Based on the above findings, we compare the results with theory and international experiences in the following section to see whether the case of the PRC follows the common evolution path or explores a new way with unique characteristics.

4. ECONOMIC GEOGRAPHY IN THE PRC: COMPARISON AND HYPOTHESES

4.1 Theoretical and Economic Background

The geographic distribution of economic activities is one of the key factors that affect efficiency. Generally speaking, increased industrial specialization is beneficial in terms of exploiting regional advantages and improving productivity. However, the overall benefit of increasing geographic concentration is ambiguous. On the one hand, higher concentration makes it possible to utilize external and scale economies; on the other hand, it enlarges the welfare gap between the core and the periphery. In Krugman's two-region two-sector model (1991a, 1991b), decreasing transportation costs always cause an increase in geographic concentration and industrial specialization. Thus these are sometimes viewed as two sides of a coin. In contrast, when there are more than two regions, Fujita et al. (1999) found that concentration and specialization could change in the opposite direction when transportation costs decrease. More generally, Rossi-Hansberg (2005) and Aiginger and Rossi-Hansberg (2006) concluded that when transportation costs decrease, industrial specialization could increase while geographic concentration would decrease correspondingly. Therefore, in multi-region models, geographic concentration may follow a bell-shaped path and industrial specialization keeps increasing as transportation costs decrease.

In reality, the actual transportation costs of manufactured goods during the 20th century decreased by more than 90% (Glaeser and Kohlhase 2004). Meanwhile, Kim (1995) and Krugman (1991a) showed that the geographic concentration of manufacturing industries in the US reached its peak in the 1930s, after which it started to descend steadily. Midelfart-Knarvik et al. (2000) extended the time period and showed the degree to which US manufacturing geographic concentration continued to decrease in the 1990s. Henley (1994) discovered that in the 1980s the geographic concentration of UK manufacturing industries stopped increasing and industrial activities become more geographically dispersed. Similarly, the *TYPCONC* and Hoover coefficient time line seems to indicate that geographic concentration in the PRC is "copying" the experiences of the US and EU and has already turned from the increasing phase to the decreasing phase. However, it has been verified and widely known that the level of geographic concentration in the PRC is much lower compared with those of the US and European economies (Lu and Tao 2009). So why is the turning point in the PRC arriving before its time? As for the change of industrial specialization in the empirical studies on regional economic integration, Aiginger and Rossi-Hansberg (2006) argued that from 1987 to 1996 the industrial GINI coefficient of 50 states in the US increased by 2.3%, meaning the level of industrial specialization increased prominently when geographic concentration decreased. Both Amiti's 1999 study and Brühlhart's 1998 study on European integration found that the resulting geographic concentration and industrial specialization increased in the 1970s and 1980s. Along with the Single Market Program since 1992, the geographic concentration of EU countries started to show a tendency for decentralization (Brühlhart 2005), while

the industrial specialization continued to increase (Aiginger and Davies 2004; Aiginger and Rossi-Hansberg 2006). Another confusing problem is that, contrary to geographic concentration, the PRC's industrial specialization did not "copy" the changing trend of the US and EU economies. As indicated by the results of *TYPESPEC* and the interprovincial Krugman Index, PRC industrial and interprovincial specialization has tended to decrease in recent years, which contradicts the pattern that every region needs to exploit its comparative advantages to accelerate economic development.

4.2 Hypotheses

As indicated by sections 3 and 4.1, it seems that, from the aspect of economic geography, the PRC takes a road different from developed economies, which are not well predicted by New Economic Geography. Though previous studies have explained the PRC phenomenon from the perspective of local protectionism, which is helpful in understanding the lower level of geographic concentration in industries with higher after-tax profit margins or state-owned shares (e.g., Bai et al. 2004; Lu and Tao 2007; Huang and Li 2006), it still cannot explain why the PRC's geography concentration and industrial specialization presents a downward trend in the process of its entering the World Trade Organization and integrating its domestic market. In fact, provincial governments could affect the speed of local industrial development through "passively defensive" protection; in addition, local officials' promotion competition (Zhou 2004, 2007) gives them strong incentives to achieve better performance via active involvement in industry and economic development through preferential industrial policies. For instance, Xu et al. (2007) observed that an inter-province governor transfer program would increase the speed of economic development for lagging provinces. And economic development is usually achieved by vigorous development of the secondary industry pushed by the government. Song and Wang (2013) further proved that local governments' key industrial policies conceived in 5-year plans indeed increased these industries' productivity.

In the PRC, the *Five-Year Plan for National Economic and Social Development* (henceforth, 5-year plan) is one of the most significant means by which governments could reallocate resources. One of the reasons local governments favor 5-year plans is that officials could make use of the "visible hands" to intervene in the development of local industries, thus completing the task of economic performance needed for their personal promotion. More importantly, when local plans base themselves on the central government's 5-year plan and special plans for industrial development, it is advantageous for local government to apply for affiliated support (e.g., tax privileges for high-tech industry, strategic promotion policy for emerging industry, etc.). In this way, the provinces can acquire extra resources for local development, which also to some extent lead to "similarity" or "homogeneity" of industrial policies across the provinces. To obtain a clearer view, we collated data from the 10th to 11th national and provincial 5-year plans and classified the industries into two categories (at the two-digit industry level according to the National Bureau of Statistics): policy-oriented industries, and other industries. The policy-oriented industries refer to those termed as "pillar industry," "advancing industry," "promising industry," "priority industry," "breakthrough industry," or described with verbs like "to enlarge and strengthen," "to prioritize," "to develop," "to accelerate," "to expand," "to cultivate" when reviewing its 5-year plans. The remaining industries are "other (non-preferred) industries," which are either not mentioned in the plans, or whose prospects are described as "to rationally develop," "to relatively control," "to optimize and adjust," "to transform," "to reduce scale," "to gradually

eliminate,” “to limit,” or “to orderly transfer” in 5-year plans.⁹ Then, for further investigation, we introduce a dummy variable to represent “policy-oriented industries” (=1) and “other industries” (=0) in each 5-year plan. When we look into the 11th 5-year plan, it can be easily seen that those listed in the central government’s 11th 5-year plan as policy-oriented industries on average are also thus chosen by as many as 20 provinces; those that do not appear in the central government’s favorite list would, on average, only appear in nine provinces’ list as policy-oriented industries.¹⁰ To measure policy similarity between the central government and the local governments, we calculate the Jaccard coefficient of similarity. The results are presented in Table 3.¹¹

Table 3: Jaccard Coefficient of Similarity of Preferential Industries in the 10th and 11th Five-Year Plans

Area	10th	11th	Inter-temporal	Area	10th	11th	Area	10th	11th
E-E	0.508	0.492	0.571	E-C	0.475	0.498	E-CG	0.464	0.432
C-C	0.479	0.653	0.570	E-W	0.309	0.346	C-CG	0.389	0.570
W-W	0.357	0.466	0.535	E-N	0.442	0.470	W-CG	0.318	0.481
N-N	0.406	0.620	0.597	C-W	0.345	0.508	N-CG	0.478	0.589
National	0.378	0.464	0.559	C-N	0.419	0.603	LG-CG	0.394	0.493
—	—	—	—	W-N	0.270	0.511	—	—	—

Note: E for East Region, C for Central Region, W for West Region and N for Northeast Region, CG for central government and LG for local government, 10th and 11th for the 10th and 11th 5-year plan; so that E-E refers to the simple average of Jaccard coefficient of similarity among east provinces, and it is similar for the rest notations. Besides, “inter-temporal” refers to the simple average of Jaccard coefficient of similarity between 10th and 11th 5-year plan of the same province.

Source: Prepared by the authors.

From Table 3, we can see first that in the horizontal comparison across provinces, except those within the east region, the Jaccard coefficients of the central, west, and northeast regions, and the inter-provinces between any two different regions all increased, resulting in within-province Jaccard coefficients for the nation as a whole ranging from 0.378 (the 10th 5-year plan) to 0.464 (the 11th 5-year plan). In other

⁹ In some provinces’ 5-year plans, the industries are usually named by collective terms of several industries (such as high-tech industry, information industry), by industrial sectors (such as chemistry industry) or by representative products in the industry. We looked into those provinces’ 5-year plans and unified their terms at two-digit industry levels by referring to standard of classification published by National Bureau of Statistics (NBS) and provincial bureau of statistics, including *Product Classification for Statistics* (China Statistics Press, 2010), *Classification of Strategic Emerging Industry* (2012), *Product Classification of New Materials*, *Classification of High-tech (Manufacturing) Industry* (2013), *Classification of High-tech Industry for Statistics* (NBS [2002] No. 33), *Interim Provisions for Classification of Information Industry for Statistics* and *Classification of Environment Protection*. For details, one may refer to the website of NBS, <http://www.stats.gov.cn/tjsj/tjbz/>. Note that Song and Wang (2013) used a different approach in defining industries: they regarded all the manufacturing mentioned in 30 provinces’ 5-year plans as major industries, following the classification of industry code in National Economy (GB/T4754-2002) and ignored new material and new resource industries. Therefore, their numbers of policy-oriented industries are different from here.

¹⁰ Although Song and Wang (2013) adopted a different approach, their results also indicated that there were many major industries overlapping in provincial and national 5-year plans.

¹¹ Given two objects, A and B, each with n binary attributes (either 1 or 0). Thus, there are four different kinds of combination of attributes for A and B, specified as M11, M01, M10 and M00 with $M11+M01+M10+M00=n$. The Jaccard similarity coefficient, J, is defined as $J = M11/(M01+M10+M11)$. Clearly J lies in [0, 1] and $J = 1$ if A is the same as B. For comparison of two provinces (inter-temporal comparison of single province), we calculate the J directly. For comparison of two regions, e.g., eastern and western area, we first calculate the Jaccard coefficient of any province-pairs (one from the eastern area and the other from the western area), and then use simple average-term to represent similarity between regions.

words, the industrial policies within provinces in the central, west, and northeast regions and across regions become increasingly similar, whereas their policies became much more similar to those of the east provinces. Second, in cross-time comparison, the inter-temporal Jaccard coefficient of all provinces between the 11th 5-year plan and the 10th 5-year plan is 0.559. Only the coefficient of the west provinces is below the average level. This indicates that when all provinces' industrial policies become more similar to that of the east provinces, the policy of the west provinces deviates most from their past policies. Lastly, in terms of the relativity between the industrial policies of the central government and local governments, the Jaccard coefficient in the 10th 5-year plan is 0.394, while in the 11th 5-year plan it raises to 0.493. On the one hand, it indicates a high similarity between the industrial policies of the central government and local governments; on the other hand, it also shows that the level of similarity is even higher in the 11th 5-year plan. Notably, the east provinces are again special among all the four areas. The Jaccard coefficient is lower in the 11th 5-year plan than that in the 10th 5-year plan. That is to say, though all provinces follow the central government in making their own industrial policies, leading to a higher similarity among the industrial policies of all provinces, the east provinces keep themselves relatively independent and are good at adapting to their own conditions.

There is little doubt that once an industry is listed as policy-oriented in the 5-year plan of a province, it receives more support, such as priority in construction land allocation, special provincial development funds, convenience in clearance of imported equipment and parts, deduction of research fees, double amortization in the cost of intangible assets, priority in raw materials and electricity supply, and professional training for certain industries. Therefore, it is speculated in this paper that those measures may allow favorite industries to expand rapidly in the valid period of the 5-year plan, and would consequently influence the spatial distribution of economic activities. On the basis of the previous discussion, we propose two hypotheses:

Hypothesis 1: for a given industry i in province r , if it is listed as policy-oriented in the local 5-year plan, then, during the valid period of the plan, its gross output share r would increase faster.

Hypothesis 2: for a given industry i in province r , if it is listed as policy-oriented in the local 5-year plan, then, during the valid period of the plan, its gross output share i would also increase faster.

If the above two hypotheses can be verified by the empirical study, it means that local government industrial policies do influence the distribution of economic geography. Due to the higher similarity in industrial policy among provinces from the 10th to the 11th 5-year plan, it could explain the changes in the PRC's recent geographic concentration and industrial specialization.

5. INDUSTRIAL POLICY AND ECONOMIC GEOGRAPHY: EMPIRICAL ANALYSIS

To test the above two hypotheses, the national- and provincial-level 10th and 11th 5-year plans are used and a dummy variable $plan_{irn}$ (where 1 represents policy-oriented industries, and 0 others) is introduced to stand for the policy support that industry i faces in the n th 5-year plan of province r . Around 1998, the industrial statistic criteria experienced a major amendment. Thus, to ensure comparability, we only focus on the 10th 5-year plan (valid during 2001 to 2005) and the 11th 5-year plan (valid during 2006 to 2010) to evaluate the effect of industrial policy on economic geography.

5.1 Methodology

First, we will test whether economic geography changes as a result of policy intervention based on the 11th provincial 5-year plan. Let p_{ir} stand for industry i 's output from the gross output of province r , and s_{ir} stand for industry i 's output from the gross output nationwide. The basic regression model is defined as follows:

$$\Delta_{ir} = \alpha + \beta \text{plan}_{ir} + \sum_j \delta_j C_{irj} + \varepsilon_{ir} \quad (1)$$

where dependent variable Δ_{ir} is the change of industry i in the valid period of the 11th 5-year plan of province r . Two alternative variables are used here, i.e., Δp_{ir} , which is the change of p_{ir} , and Δs_{ir} , which is the change of s_{ir} . The variable Δp_{ir} reflects the industrial structure change within province r and Δs_{ir} catches the geographic change within industry i , defined as $\Delta p_{ir} = p_{ir2010} - p_{ir2005}$ and $\Delta s_{ir} = s_{ir2010} - s_{ir2005}$, respectively. The policy explanatory variable plan_{ir} is the paper's core, the coefficient of which is expected to be positive by the hypotheses in section 4. C_{irj} is a group of control variables, measuring the influence of non-policy factors. ε_{ir} is a random error term with constant variance. When Δp_{ir} is the dependent variable, the main control variables include firm_{ir} , that is, the number of firms in the same industry in province r and its quadratic term sfm_{ir} , which are respectively the proxy variables for Marshall Externality (e.g., labor market pooling, specialized supply, and knowledge spillovers) and competitive effect. According to New Economic Geography, the estimated coefficient of Marshall Externality (firm_{ir}) is expected to be positive, while the estimated coefficient of competitive effect (sfm_{ir}) is negative. In addition, p_{ir} is used to control the initial level of industry i in province r before implementing the 11th 5-year plan, and another set of dummy variables ID_i to control industry-specific characteristics. All control variables use lagged value to avoid possible reverse causality. When Δs_{ir} is the dependent variable, we follow Midelfart–Knarvik et al. (2000), in which the logarithmic value of regional population, $\ln \text{pop}_r$, controls the regional specific characteristics, or use the logarithmic value of regional output, $\ln \text{gdp}_r$, as an alternative. The estimated coefficients of these two variables are expected to be positive. Meanwhile, the quadratic term of the total number of firms in industry i , isfm_i , is used to measure the competitive effect, and the improvement of transport infrastructures locally (trans) or regionally (ntrans) to measure the change of transportation costs, the estimated coefficients of which are negative and positive, respectively, according to New Economic Geography. Last, due to data availability, we follow Wooldridge (2004: 300) to deal with the omitted variables problem, where the lagged dependent variables are used in the cross-sectional regression model as a proxy to control the effect from other industrial development factors. For a detailed definition of variables, methodology of construction, and summary statistics, see Appendix IV.

5.2 Results

It needs to be pointed out that though such endogeneity problems as omitted variables and reverse causality have been addressed by lagged dependent and control variables in the model, they may still exist between the policy variable, plan_{ir} , and its dependent variable Δp_{ir} , since there is possibility of self-selection. That is to say, industry i is selected as policy-oriented in province r because the local government would cart its industrial policy to that of the central government's, resulting in a deviation from its comparative advantage and causing correlation between plan_{ir} and ε_{ir} , which makes the OLS result unreliable. To accommodate this, we introduce two other variables: cplan and rplan , representing the central government's industrial policy in the national

11th 5-year plan and its long-term strategies for regional development, respectively. Both are used as instrumental variables in two-stage least squares regression (2SLS). As mentioned above, local governments tend to refer to the central government's 5-year plan in making their own industrial policies, through which they can acquire extra resources for development. Generally speaking, the prioritized industries in the central government's plan are significant references for the local governments to choose their own policy-oriented industries,¹² thus leading to a high similarity in pillar or major industries across provinces. Meanwhile, for those provinces that are included in the central government's regional development program, the development strategies of the central government are also very influential. For example, the document entitled *Some Opinions on Promoting the Further Development of Western China by the State Council* emphasizes that provinces in the west region should cultivate industries with local advantages and establish a major area for national energy and mineral resources. Obviously, those instructive ideas will affect the choice of preferential industries by the provincial governments in the west region. Therefore, local governments' 11th 5-year plan ($plan_{ir}$) highly correlates with the central government's 11th 5-year plan ($cplan$) and its strategies for regional development ($rplan$). It needs to be emphasized that the central government's policies would influence the prioritized industry as a whole, though they cannot determine the development of a specific industry in every province, which means they can influence $\sum_r \Delta p_{ir}$ but not Δp_{ir} . Maximum likelihood estimation (MLE) and generalized method of moments (GMM) estimation are used to accommodate the problem of weak instrumental variables and heteroskedasticity caused by provincial and industrial differences.¹³

Table 4 reports the results with Δp_{ir} as the dependent variable. By regression Eq. (1), the ordinary least squares (OLS) estimation (column 1 of Table 5) shows that the regression coefficient of $plan_{ir}$ is 0.745 with significance level below 1%. This means that local industrial policy does accelerate the development pace of policy-oriented industries. To improve efficiency of the estimates and control for the self-selection problem, the 2SLS model (Column 2 of Table 5) uses $cplan$ and $rplan$ as the instrumental variables for $plan_{ir}$. In the first-stage regression, the robust F-statistic is 81.81, and the p-values of both regression coefficients are less than 0.01, indicating that $cplan$ and $rplan$ can effectively explain $plan_{ir}$. In addition, the p-values of the over-identification test and DWH test are 0.69 and 0.05, respectively, which means the instrumental variables are effective. The estimated regression coefficient of $plan_{ir}$ is positive (1.445) and significant, showing that being listed as a policy-oriented industry does grant extra advantages over other industries in the same province.

Meanwhile, the fact that OLS estimator of $plan_{ir}$ is less than that of 2SLS verifies our conjecture about local government choosing its policy-oriented industry to cart that of the central government's preference, resulting in underestimation by OLS regression. In addition, the results also show that when the number of firms ($firm_{ir}$) in one industry increases, the Marshall's externality benefits its development on one hand; on the other hand, the competition among firms also increases. The competitive effect is confirmed by the negative estimated coefficient of sfm_{ir} as the theory of New Economic Geography predicts. Though it is believed, according to the results of the first stage regression, that there is no serious weak instrument bias, limited information maximum

¹² For instance, when one of the prioritized industries had its name changed from "high and new industry" to "high-tech industry" in the central government's 11th 5-year plan, 21 provinces adopted the new name in their 11th 5-year plan instantly.

¹³ However, when Δs_{ir} is the dependent variable, the local policy variable does not suffer from endogeneity problems. It is further proved by the close parameters estimated by OLS and fixed effects model estimation.

likelihood (LIML) regression is introduced as a precautionary measure. As shown by the results in column 3 of Table 5, the estimators by 2SLS and LIML regression are quite close, which confirms there are no weak instrumental variables. Considering scale differences across the provinces and industries, columns 4 to 6 use GMM estimation, which is more effective in the presence of heteroskedasticity. The results are also close to that of 2SLS estimation. In column 5, the lagged dependent variable Δp_{irL} is replaced with lagged industrial development speed ar_{ir-L6} to catch the influence of other economic trend factors. The results indicate that factors such as techniques and human resources promoting fast development could have a long-term influence on industry. Column 6 further considers the influence of firm scale (scl_{ir04}). However, the results of $plan_{ir}$ are not influenced when introducing firm scale. The estimated coefficient of scl_{ir04} indicates that when the proportion of small business increases by 1%, the share of industrial output in that region would increase by 0.3%. This result to some extent supports Rosenthal and Strange's 2010 conclusion of the positive relationship between small business and industry vitality.

Table 4: Impacts of Preferential Industry Policy on Industrial Specialization
(Δp_{ir} as dependent variable)

	(1) OLS	(2) 2SLS	(3) LIML	(4) GMM	(5) GMM	(6) GMM
$plan_{ir}$	0.745*** (0.105)	1.445*** (0.331)	1.446*** (0.331)	1.477*** (0.321)	1.403*** (0.318)	1.451*** (0.359)
$firm_{ir}$	0.216*** (0.052)	0.167*** (0.048)	0.167*** (0.048)	0.170*** (0.048)	0.192*** (0.051)	0.167*** (0.048)
sfm_{ir}	-0.010*** (0.003)	-0.008** (0.003)	-0.008** (0.003)	-0.009** (0.003)	-0.010*** (0.003)	-0.008 (0.048)
Δp_{irL}	0.202*** (0.054)	0.204*** (0.053)	0.204*** (0.053)	0.201*** (0.052)		0.200*** (0.054)
ar_{ir-L6}					0.229** (0.318)	
p_{ir03}	-0.216*** (0.031)	-0.244*** (0.036)	-0.244*** (0.036)	-0.245*** (0.036)	-0.216*** (0.041)	-0.229*** (0.037)
scl_{ir04}						0.339 (0.224)
Constant	-0.345*** (0.147)	-0.336** (0.149)	-0.336** (0.149)	-0.346** (0.147)	-0.614*** (0.184)	-0.647** (0.263)
Industry dummy	Yes	Yes	Yes	Yes	Yes	Yes
N. of Obs.	1,144	1,144	1,144	1,144	1,107	921
R ²	0.270	0.236	0.236	0.232	0.187	0.258
F-statistic/Wald chi2	18.43	112.65	112.63	115.31	83.08	120.82
Over-identification		0.164 0.6859		0.164 0.6859	0.087 0.7676	0.075 0.7835
DWH		3.928 0.0475				
GMM C statistic chi2				5.116 0.0237	4.05134 0.0441	3.546 0.0597
Robust F-Statistic of the first stage		81.8065 0.0000	81.8065 0.0000	81.8065 0.0000	90.869 0.0000	62.6305 0.0000

OLS = ordinary least squares; 2SLS = two-stage least squares; LIML = limited information maximum likelihood; GMM = generalized method of moments; N. of Obs. = number of observations.

Note: ① Robust standard errors are in parentheses, with the P-value below; ② *, ** and *** stand for significant at 10%, 5%, and 1%; ③ Due to lack of data, regression (6) does not include Beijing, Tianjin, Shanghai, Henan, and Hunan.

Source: Prepared by the authors.

Table 5 reports the results with Δs_{ir} as the dependent variable. Estimated coefficients of $plan_{ir}$ are significant and around 0.4 (columns 1–8 in Table 5). It means that once listed as policy-oriented, the output share of industry i in province r out of the gross output of industry i as a whole would increase. However, when comparing the results in Tables 4 and 5, we can see that provincial plans' influence on industrial structure (around 1.4) is much stronger than on economic geography (around 0.4). This sounds reasonable since provincial industry policy would have larger influence inside a province rather than between provinces. In addition, the regression coefficients of $lnpop_r$ and $lngdp_r$ are both positive and significant as expected, which means local demands have a positive effect on the geographic concentration of the industry. Furthermore, as shown by the estimated coefficient of $rplan$, the central government's regional development strategies accelerate the proportional increase in the middle-west and northeast provinces. Thus regional development strategies have a balancing effect on forerunners and followers. The regression coefficient of independent variables $trans_r$ and $ntrans_r$ measuring the improvement of transportation infrastructures locally or among regions are both positive, which indicates that the decrease of transportation costs is still helpful for the increase of geographic concentration. Last, the lagged dependent variable Δs_{irL} , or the coefficient of lagged development speed ar_{ir-L3} , are positive and, in most cases, significant, which means that historical factors also influence geographic concentration.

Table 5: Impacts of Preferential Industry Policy on Geographic Concentration
(Δs_{ir} as dependent variable)

	(1) OLS	(2) 2SLS	(3) LIML	(4) GMM	(5) GMM	(6) GMM
$plan_{ir}$	0.384*** (0.090)	0.398*** (0.091)	0.370*** (0.085)	0.358*** (0.083)	0.393*** (0.088)	0.392*** (0.089)
Δs_{irL}	0.330** (0.157)			0.329** (0.161)	0.329** (0.161)	0.341** (0.157)
$lnpop_r$	0.280*** (0.041)	0.330*** (0.046)				
$lngdp_r$			0.190*** (0.052)	0.183*** (0.046)	0.183*** (0.045)	0.183*** (0.046)
ar_{ir-L3}		0.093 (0.076)	0.113 (0.074)			
$Rplan$	0.274** (0.108)	0.305*** (0.114)	0.574*** (0.111)	0.442*** (0.118)	0.442*** (0.118)	0.507*** (0.107)
$trans_r$				0.092* (0.055)	0.092* (0.055)	
$ntrans_r$			1.593** (0.738)			1.172* (0.701)
s_{ir03}	-0.115*** (0.024)	-0.118*** (0.024)	-0.117*** (0.025)	-0.115*** (0.024)	-0.115*** (0.024)	-0.116*** (0.024)
$isfm_i$	-0.133 (0.086)	-0.133 (0.089)			-0.136 (0.086)	-0.135 (0.086)
$_cons$	-2.202*** (0.290)	-2.652*** (0.315)	-0.201* (0.103)	-1.833*** (0.376)	-1.785*** (0.378)	-1.882 (0.371)
N. of Obs.	1,147	1,110	1,110	1,147	1,147	1,147
F-Statistic	27.88	27.18	25.32	24.78	22.26	22.84
R-squared	0.2182	0.2035	0.1961	0.2136	0.2150	0.2142

OLS = ordinary least squares; 2SLS = two-stage least squares; LIML = limited information maximum likelihood; GMM = generalized method of moments; N. of Obs. = number of observations.

Note: ① Robust standard errors are in parentheses, with P-value below; ② *, ** and *** stand for significant at 10%, 5% and 1%.

Source: Prepared by the authors.

5.3 Panel Regression: Further Test

Section 5.2 shows that provincial industrial policy explains the economic geography changes during the valid period of the 11th 5-year plan (2006–2010). In this section, we will extend the time period under study to the valid period of both the 10th and 11th 5-year plan (2001–2010) with a panel model. Compared with the former, due to a major amendment to the industrial statistical criteria in 1998, we have fewer variables that could be used to analyze the changes during 2001–2010. However, there are also some advantages. As is known, omitted variables bias is a common problem in empirical studies. When the omitted variables are unobservable and unchanging individual differences, panel data offer another tool for eliminating the bias. Therefore, a two-period panel data constructed in this section could further test the robustness of the results. The panel model is set as follows:

$$\Delta_{im} = \beta_0 + \beta_1 plan_{im} + \sum_j \delta_j C_{imj} + \alpha_i + \gamma_r + \varepsilon_{im} \quad (2)$$

where Δ_{im} is the change of industry i in province r during the valid period of the n -th 5-year plan. Two dependent variables Δp_{im} and Δs_{im} are used as in section 5.1. Here, Δp_{im} is the change of p_{ir} , defined as $\Delta p_{im} = p_{irt} - p_{irt-5}$, and Δs_{im} is the change of s_{ir} , defined as $\Delta s_{im} = s_{irt} - s_{irt-5}$, with t as the last year of the valid period of the n -th 5-year plan in both equations. The dummy variable of industrial policy $plan_{im}$ measures the provincial policy faced by industry i during the valid period of the n -th 5-year plan, with 1 being policy-oriented and 0 the others. As predicted by Hypothesis 1 and Hypothesis 2, the estimated coefficient of $plan_{im}$ is positive, with α_i being a set of time-invariant characteristics of industry i , γ_r being a set of time-invariant characteristics of region r , and ε_{im} being the random error. In addition, C_{imj} is a group of variables controlling the influence of non-policy factors. When Δp_{im} is the dependent variable, the main control variables include the lagged value of industry i 's initial scale x_{imL} in province r and the lagged value of numbers of firms in the industry fm_{imL} . Considering the possible self-selection problem, the central government's n -th 5-year plan $cplan_n$ and its regional development strategies $rplan_n$ are used again as two instrumental variables for local government's preferential industrial policies. When Δs_{im} is the dependent variable, the control variables include s_{imL} , the lagged value of industry i 's initial scale in province r , and $trans_m$, the improvement of transport infrastructures in implementing the n -th 5-year plan.

Table 6 summarizes the results when Δp_{im} is the dependent variable. The Hausman test is in favor of the fixed effects model (FEM); therefore, we first apply the FEM to estimate the regression Eq. (2). As shown in column 1 of Table 6, the estimated coefficient of local policy $plan_{im}$ is positive (0.333) and significant, which means that local industrial policy does promote the growth of the preferential industry's share in local industrial structure. If we further control time-fixed effect, the results remain (column 2 in Table 6). Columns 3 to 6 in Table 6 report two-step GMM estimator or LIML estimation when using $cplan_n$ and $rplan_n$ as the instrumental variables for local policy $plan_{im}$. In the first stage regression, the Angrist–Pischke F-statistic is at 12.8, which exceeds the threshold ($F=10$) suggested by Stock et al. (2002). Thus, it is believed that there is no severe weak instrument bias. Still, columns 4 and 6 use LIML estimation since it is much less sensitive to weak instrumental variables. The

coefficients of LIML estimators are quite close to those of GMM estimators. Judged from the results of columns 3 to 6, the coefficients of control variables are quite stable. Since local policy $plan_{im}$ has a significant and positive effect on the growth of industrial share out of local gross output, it can be concluded that local governments' industrial policies in 5-year plans do significantly influence local industrial structure and specialization.

Table 6: Impacts of Preferential Industry Policy on Industrial Specialization
(Δp_{im} as dependent variable)

	(1) FEM	(2) FEM	(3) IV-FEM (GMM)	(4) IV-FEM (LIML)	(5) IV-FEM (GMM)	(6) IV-FEM (LIML)
$plan_{im}$	0.333** (0.136)	0.268** (0.131)	1.182* (0.666)	1.176* (0.712)	1.292* (0.683)	1.286* (0.732)
x_{imL}					-0.273*** (0.048)	-0.267*** (0.049)
x_{imL2}	-0.340*** (0.061)	-0.350*** (0.067)	-0.351*** (0.063)	-0.343*** (0.062)		
fm_{imL}	0.172*** (0.041)	0.164*** (0.041)	0.177*** (0.043)	0.177*** (0.043)	0.170*** (0.042)	0.170*** (0.420)
obs.	2,254	2,254	2,214	2,214	2,214	2,214
F test	12.22	8.30	10.91	10.99	10.30	10.30
Kleibergen–Paap Chi-sq (2) (Under-identification test)			22.83 0.0000	22.83 0.0000	22.85 0.0000	22.845 0.0000
Hansen J statistic (over-identification test)			1.490 0.2222	1.483 0.2234	1.564 0.2222	1.554 0.2126
Angrist–Pischke F test for first-stage regressions			12.76 0.0000	12.76 0.0000	12.77 0.0000	12.77 0.0000

FEM = fixed effects model; GMM = generalized method of moments; LIML = limited information maximum likelihood; Obs. = observations.

Note: ① Robust standard errors are in parentheses, with P-value below; ② *, ** and *** stand for significant at 10%, 5% and 1%.

Source: Prepared by the authors.

Table 7 summarizes the results when Δs_{im} is the dependent variable. A provincial dummy variable is added to the pooled OLS regression (column 1 in Table 7). As indicated by the results, the estimated coefficient $plan_{im}$ is 0.447, which means that local industrial policy promotes provincial output share growth where this industry is listed as policy-oriented. Columns 2 to 5 in Table 7 apply a fixed effects model to address the time-invariant heterogeneity of unobservable or omitted variables. According to the results in Table 7, the estimated coefficient of $plan_{im}$ by fixed effects model and pooled regression model are relatively close. However, the F test of joint significance of the fixed effects model refuses the null hypothesis ($p=0.00$). Thus the fixed effects model is more appropriate. The estimated coefficient of $plan_{im}$ is positive and significant, indicating that, compared with those provinces that do not list some industry as policy-oriented, local industrial policies do grow its share of nationwide output in the province where this industry is offered with preferential policies. In addition, the estimated coefficient of variable $Intrs_m$, which measures the improvement of local transportation infrastructures, is positive. And again it proves the beneficial effect of decreasing transportation costs on geographic concentration. The coefficients of the lagged dependent variable Δs_{iml} and the lagged development speed ar_{imL} are both positive and significant, showing that historical factors have continuous influence on economic geography.

Table 7: Impacts of Preferential Industry Policy on Geographic Concentration
(Δs_{im} as dependent variable)

Variable	(1) Pooled OLS	(2) FEM	(3) FEM	(4) FEM	(5) FEM
$plan_{im}$	0.447*** (0.071)	0.460*** (0.073)	0.454*** (0.068)	0.438*** (0.078)	0.421*** (0.073)
s_{irL}	-0.112*** (0.022)	-1.044*** (0.057)	-1.080*** (0.056)	-1.047*** (0.056)	-1.083*** (0.056)
ar_{imL}	0.021 (0.021)	0.149** (0.053)		0.152*** (0.055)	
Δs_{imL}			0.502*** (0.118)		0.510*** (0.118)
$Intrs_m$				0.017 (0.014)	0.026** (0.013)
Province Dummy	Yes	None	None	None	None
_cons	-0.176*** (0.040)	3.325*** (0.196)	3.293*** (0.186)	3.324*** (0.216)	3.179*** (0.204)
Obs.	2,211	2,211	2,294	2,211	2,294
F-Statistic	8.89	135.25	152.26	109.91	125.70
R ²	0.1506	0.6273	0.6515	0.6277	0.6525

OLS = ordinary least squares; FEM = fixed effects model; Obs. = observations.

Note: ① Robust standard errors are in parentheses, with P-value below; ② *, ** and *** stand for significant at 10%, 5% and 1%.

Source: Prepared by the authors.

6. CONCLUSIONS

New Economic Geography predicted that market integration would increase industrial specialization and reduce geographic concentration, which has been confirmed in developed countries. However, the evolution of economic geography in the PRC from 1999 to 2010 follows a different path: specialization and concentration started to decrease simultaneously after 2005 and the similarity of industrial structure among provinces began to increase. Though industrial concentration is doomed to change from increase to decrease along with the progress of domestic market integration, it turns around to decrease at a relatively low level of concentration verified by the previous literatures. Meanwhile, the situation of industrial and interprovincial specialization is also deteriorating with decreasing industrial concentration. As shown by the stylized facts, the evolution of economic geography in the PRC is “disturbed” by unique non-market factors other than economic and geographic determinants.

This paper found that local governments’ policy intervention plays an important role in the evolution of PRC economic geography. Through industrial policy, local governments could directly interfere with development. As shown by the empirical results, the interference is so effective that it profoundly changes the industrial structures inside a province and regional specialization among industries, and, as a result, influences the domestic economic landscape. The effectiveness of the instrumental variables used here further proves the guiding role of the central government in provincial industrial policy. The central government’s preference usually reflects more about industrial structures in developed regions, which is itself a function of its long-term goal of pursuing economic efficiency. Since local governments thus have incentives to follow the central government’s lead when drawing up their

own policies, this would lead to a structure deviation of supports, especially for less-developed provinces. Overall, there are pros and cons about the behavior pattern of “local governments following central government.” It is helpful in achieving the central government’s goal of industrial structure adjustment and transformation throughout the country. However, its negative effect is that some provinces sacrifice too much to cater for the central government’s support. Furthermore, the negative effect is much more obvious in underdeveloped provinces, which is also supported by empirical results. Since the early stage of economic reform in the PRC, provinces in the coast region have been striving for the central government’s support, resulting in severe similarity of industrial structures across regions; nowadays, similar trends among underdeveloped provinces in the hinterland are observed. Learning from the experience of the developed regions is to some extent helpful in realizing the potential of underdeveloped provinces, but low concentration and insufficient specialization will definitely bring efficiency loss to such a large economy as the PRC.

REFERENCES

- Aiginger, K., and S. W. Davies. 2004. Industrial Specialisation and Geographic Concentration: Two Sides of the Same Coin? Not for the European Union. *Journal of Applied Economics* 7(2): 231–248.
- Aiginger, K., and E. Rossi–Hansberg. 2006. Specialization and Concentration: A Note on Theory and Evidence. *Empirica* 33(4): 255–266.
- Amiti, M. 1999. Specialization Patterns in Europe. *Weltwirtschaftliches Archiv* 135(4): 573–593.
- Bai, C.–E., Y. Du, Z. Tao, and S. Y. Tong. 2004. Local Protectionism and Regional Specialization: Evidence from China's Industries. *Journal of International Economics* 63(2): 397–417.
- Brühlhart, M. 1998. Economic Geography, Industry Location and Trade: The Evidence. *World Economy* 21(6): 775–801.
- Brühlhart, M., and R. Traeger. 2005. An Account of Geographic Concentration Patterns in Europe. *Regional Science and Urban Economics* 35(6): 597–624.
- Cowell, E. A. 1995. *Measuring Inequality*. London: Prentice Hall.
- Ellison, G., and E. Glaeser. 1997. Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. *Journal of Political Economy* 105(5): 889–927.
- Fujita, M., P. Krugman, and A. Venables. 1999. *The Spatial Economy*. Cambridge, MA: MIT Press.
- Glaeser, E., and J. Kohlhase. 2004. Cities, Regions and the Decline of Transport Costs. *Papers in Regional Science* 83(1): 197–228.
- Henley, A. 1994. Industrial Development in U.K. Manufacturing since 1980. *The Manchester School* 62(1): 40–59.
- Krugman, P. 1991a. *Geography and Trade*. Cambridge, MA: MIT Press.
- . 1991b. Increasing Returns and Economic Geography. *The Journal of Political Economy* 99(3): 483–499.
- Krugman, P., and A. Venables. 1995. Globalization and the Inequality of Nations. *Quarterly Journal of Economics* 110(4): 857–880.
- Kim, S. 1995. Expansion of Markets and the Geographic Distribution of Economic Activities: The Trends in U.S. Regional Manufacturing Structure, 1860–1987. *Quarterly Journal of Economics* 110(4): 881–908.
- Lu, J., and Z. Tao. 2009. Trends and Determinants of China's Industrial Agglomeration. *Journal of Urban Economics* 65(2): 167–180.
- Midelfart–Knarvik, K., H. Overman, P. Lane, and J.–M. Viaene. 2002. Delocation and European Integration: Is Structural Spending Justified? *Economic Policy* 17(35): 321–359.
- Midelfart–Knarvik, K., H. Overman, S. Redding, and A. Venables. 2000. The Location of European Industry. European Commission European Economy—Economic Papers, No.142. Brussels: European Commission.
- Palan, N. 2010. Measurement of Specialization—The Choice of Indices. FIW Working Paper series. Vienna: Austrian Institute of Economic Research.

- Rosenthal, S., and W. Strange. 2010. Small Establishments/Big Effects: Agglomeration, Industrial Organization and Entrepreneurship. In *Agglomeration Economics*, edited by E. Glaeser. Chicago, IL: University of Chicago Press.
- Rossi-Hansberg, E. 2005. A Spatial Theory of Trade. *American Economic Review* 95(5): 1464–1491.
- Stock, J. H., J. H. Wright, and M. Yogo. 2002. A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business and Economic Statistics* 20(4): 518–529.
- Venables, A. 1996. Equilibrium Locations of Vertically Linked Industries. *International Economic Review* 37(2): 341–359.
- Wooldridge, J. M. 2004. *Introductory Econometrics: A Modern Approach* (International Edition). Beijing: Tsinghua University Press.
- Young, A. 2000. The Razor's Edge: Distortions and Incremental Reform in the People's Republic of China. *Quarterly Journal of Economics* 115(4): 1091–1135.

Literature in Chinese

- Fan, J. 2004. Market Integration, Regional Specialization and Tendency of Industrial Agglomeration: An Implication for Regional Disparity. *Social Science in China* 6: 39–51.
- . 2008. Unbalanced Industrial Structure, Spatial Agglomeration and Change of Regional Disparity. *Shanghai Economic Research* 2: 3–13.
- Fan, J., and L. Wenfang. 2011. Effect of Spatial Concentration of Manufacturing in China: A Review. *South China Journal of Economics* 6: 53–66.
- Fan, J., and Y. Jing. 2011. Judgement of China's Industrial Agglomeration Level: Concurrent Studies of Inter-Regional Similarity Trend of Industries. *Jianghai Academic Journal* 5: 89–94.
- Huang, J., and L. Kunwang. 2006. Foreign Trade, Local Protectionism and Industrial Location in China. *China Economic Quarterly* 5(3): 733–760.
- Jin, Y., Z. Chen, and M. Lu. 2006. Industry Agglomeration in China: Economic Geography, New Economic Geography and Policy. *Economic Research* 4: 79–89.
- Lin, J. Y. 2012. *New Structural Economics: A Framework for Rethinking Development and Policy*. Beijing: Beijing University Press.
- Lu, J., and Z. Tao. 2006. Industrial Agglomeration and Co-agglomeration in China's Manufacturing Industries: With International Comparison. *Economic Research* 3: 103–114.
- . 2007. Determinants of Industrial Agglomeration in China: Evidence from Panel Data. *China Economic Quarterly* 6(3): 801–816.
- Luo, Y., and L. Cao. 2005. A Positive Research on Fluctuation Trend of China's Manufacturing Industrial Agglomeration Degree. *Economic Research* 8: 106–115.
- Ma, J. 1989. *The Skewed Pyramid: Unbalancing Industrial Structure in China*. Beijing: Academy Press.

- Nie, H., T. Jiang, and D. Yang. 2012. A Review and Reflection on the Use and Abuse of Chinese Industrial Enterprises Database. *World Economy* 5: 142–158.
- Song, L., and X. Wang. 2013. Industry Policy, Resource Allocation and Productivity. *Management World* 12: 63–77.
- Xu, X., X. Wang, and Y. Shu. 2007. Local Officials and Economic Growth. *Economic Research* 9: 18–31.
- Wang, Y., and H. Wei. 2007. Characteristics of Industries, Competition for Space and the Geographic Concentration in Manufacture, *Management World* 4: 68–77.
- Wang, F., Y. Wang, Y. Tang, and J. Fang. 2010. Whether the Time has Come for Manufacturing Dispersion. *Zhejiang Social Sciences* 9: 2–10.
- Wen, M. 2004. Relocation and Agglomeration of Chinese Industry. *Journal of Development Economics* 73(1): 329–347.
- Zhou, L. 2004. The Incentive and Cooperation of Government Officials in the Political Tournaments: An Interpretation of the Prolonged Local Protectionism and Duplicative Investments in China. *Economic Research* 7: 36–50.
- . 2007. Governing China's Local Officials: An Analysis of Promotion Tournament Model. *Economic Research* 7: 36–50.

APPENDIX 1

1. Data Source and Collection

Data used in the paper are collected from the *China Industry Economy Statistical Yearbook*, the first and the second *China Economic Census Yearbook*, the *China Statistical Yearbook*, the *Almanac of China's Economy*, and provincial statistical yearbooks. First, we used the first and the second *China Economic Census Yearbook* and *China Industry Economy Statistical Yearbook* from 2001 to 2011 to collect the gross product of industrial enterprises above designated size of 37 two-digit industries in the PRC's 31 provinces as follows: 25 industries from 1999 to 2010 (See note 1), 10 industries in 2004 and 2008 (see note 2) and non-metal mining and clothing industries from 2004 to 2010. Second, we collected the gross product of non-metal mining and clothing industries in 2003 and 10 other industries in the remaining years from the provincial statistics yearbooks. Last, due to the inconsistency in some provinces, some data were recalculated or adjusted. The adjustments are as follows: (1) the yearbooks of Henan and Liaoning provinces do not provide the gross product of these 10 industries from 2005 to 2010. Thus, we calculated the gross product of these 10 industries from 2005 to 2007 through the added value and the ratio of added value in *Liaoning Statistical Yearbook*. Because of the lack of a ratio of added value in 2009 and 2010, we calculated the gross product by main operating income and the average ratio between main operating income and gross product in previous years. The missing data in *Henan Statistical Yearbook* was also calculated by the above approach. (2) Since the statistical data reported in 2000 *Shanghai Statistical Yearbook* is at village and above level, the data of plastic product industries in Shanghai in 1999 were adjusted by the ratio between industrial enterprises above designated size and enterprises in villages and above. (3) Zhejiang and other six provinces miss some data or have changed statistical criteria. The missing data of Zhejiang province (from 1999 to 2003) were collected from *60 Years' Collections of Statistics of Zhejiang Province (Zhejiang 60 Nian Tongji Ziliao Huibian)*. The missing data of Guangxi province (2003, from 2005 to 2007) were collected from the relevant years' *Almanac of China's Yearbook*. The missing data of Inner Mongolia (from 2000 to 2001), Qinghai (2002), and Gansu (from 2000 to 2002) provinces were calculated by the average rate of growth of adjacent years. The missing data of Chongqing (from 1999 to 2002) were calculated by the average rate of growth from 2004 to 2010.

2. Data Cleaning

Since the panel data of the 37 industries used in this paper comes from different yearbooks, some of which are even by calculation, it is necessary to check for consistency and accuracy. The verification is conducted at two levels.

First, we adjust the unit of gross product into 100 million with two digits. The 10,354 data from the *China Industry Economy Statistical Yearbook*, the first and the second *China Economic Census Yearbook* are denoted as x_{irt}^{china} , and the corresponding data from provincial yearbooks are marked as x_{irt}^{local} . The difference between the above data sets is defined by $\mu_{irt} \equiv \left| x_{irt}^{china} - x_{irt}^{local} \right| / x_{irt}^{china}$. As shown in Table A1 below, 98.33% of the data are consistent and 99.42% are less than 1% difference. Therefore, we believed the industrial data calculated by the National Bureau of Statistics and local bureaus of statistics are highly consistent. As for the different data, we found that most were caused by the discrepancy between classification of

some industries, and some were caused by the lagged effect of criteria adjustment in local statistics. When difference occurred, we took the data compiled by the NBS.

Table A1: The Comparison of National and Provincial Statistics from 1999 to 2010 (%)

Δ	$\Delta=0$	$0<\Delta\leq 0.5\%$	$0.5\%<\Delta\leq 1\%$	$1\%<\Delta\leq 5\%$	$5\%<\Delta\leq 10\%$	$10\%<\Delta$	Total
%	98.33	0.79	0.29	0.26	0.11	0.20	100

Secondly, we have checked the consistency of data from *China Statistical Yearbook* and data used in this paper from the perspective of sum value. First of all, we add up the output of the 37 industries and from the 31 provinces respectively. The difference between the added-up value by industry (x_{it}^σ) or by province (x_{rt}^σ) and corresponding data from *China Statistical Yearbook* (x_{it}^c and x_{rt}^c respectively) is defined as $\mu_{it} \equiv |x_{it}^c - x_{it}^\sigma|/x_{it}^c$ and $\mu_{rt} \equiv |x_{rt}^c - x_{rt}^\sigma|/x_{rt}^c$, respectively. Due to the data limitation and a major change in the criteria of industry classification in 2003, we select 37 industries (GB/T4754-2002) as the object. Thus, the comparison above should pay attention to the influence of the added-up value from the following industries: (1) from 1999 to 2002, wood and bamboo processing, other minerals processing and weapons and ammunition manufacturing; (2) after 2003, other minerals processing and wasted resources and materials recovery and processing. As shown by the results, the differences of the 372 values by province are all below 0.5%. As for the values by industry, the 434 differences out of 436 are below 0.5%, except for furniture manufacturing (0.846%, 1999) and stationery and sports products manufacturing (0.501%, 2000) (Since *China Statistical Yearbook* lacks relevant data, the added-up of crafts and others manufacturing from 1999 to 2002 were not verified, leaving 436 pairs of data).

By verification at the above two levels, the statistical data this research relies on are highly consistent. In a word, the construction of provincial and industrial database by the statistics from *China Industry Economy Statistical Yearbook*, the census data, and provincial statistical yearbooks is feasible and reliable.

Note 1: These 25 two-digit industries refer to Mining and Washing of Coal, Extraction of Petroleum and Natural Gas, Mining and Processing of Ferrous Metal Ores, Mining and Processing of Non-Ferrous Metal Ores, Processing of Food from Agricultural Products, Processing of Foodstuff, Manufacture of Beverages, Manufacture of Tobacco, Manufacture of Textile, Printing Reproduction of Recording Media, Processing of Petroleum, Coking, Processing of Nuclear Fuel, Manufacture of Raw Chemical Materials and Chemical Products, Manufacture of Medicines, Manufacture of Chemical Fibers, Manufacture of Non-metallic Mineral Products, Smelting and Pressing of Ferrous Metals, Smelting and Pressing of Non-ferrous Metals, Manufacture of Metal Products, Manufacture of General Purpose Machinery, Manufacture of Special Purpose Machinery, Manufacture of Transport Equipment, Manufacture of Electrical Machinery and Equipment, Manufacture of Communication Equipment, Computers and other Electronic Equipment, Manufacture of Measuring Instruments and Machinery for Cultural Activity and Office Work, and Producing and Supply of Electric Power and Heat Power.

Note 2: These 10 two-digit industries refer to Manufacture of leather, Fur, Feather, and Related Products, Processing of Timber, Manufacture of Wood, Bamboo Rattan, Palm and Straw Products, Manufacture of Furniture, Manufacture of Paper and Paper Products, Manufacture of Articles for Culture, Education and Sport Activities, Manufacture of Rubber, Manufacture of Plastics, Manufacture of Artwork and Other Manufacturing, Producing and Supply of Gas, and Producing and Supply of Water.

APPENDIX 2: TIME SERIES FOR GEOGRAPHIC CONCENTRATION ($CONC_i$) AND THE RELATIVE RATE OF CHANGE

Industry Code	1999	2000	2001	2002	2003	2004	2005
06	2.705	2.735	2.690	2.647	2.599	2.576	2.574
07	2.328	2.289	2.326	2.372	2.418	2.417	2.427
08	2.537	2.521	2.503	2.521	2.479	2.673	2.678
09	2.644	2.630	2.666	2.644	2.544	2.593	2.626
10	2.896	2.896	2.902	2.883	2.842	2.809	2.764
13	2.917	2.861	2.828	2.799	2.801	2.777	2.750
14	2.914	2.919	2.912	2.940	2.931	2.947	2.901
15	3.012	3.008	3.002	3.003	2.997	3.027	2.999
16	2.744	2.744	2.782	2.816	2.838	2.846	2.862
17	2.508	2.511	2.450	2.395	2.374	2.286	2.280
18	2.250	2.246	2.235	2.220	2.194	2.208	2.220
19	2.306	2.294	2.286	2.205	2.207	2.246	2.262
20	2.735	2.736	2.734	2.741	2.771	2.732	2.718
21	2.667	2.671	2.578	2.541	2.438	2.384	2.393
22	2.781	2.729	2.713	2.663	2.605	2.561	2.577
23	2.867	2.838	2.819	2.798	2.690	2.665	2.666
24	1.961	1.935	1.956	1.980	1.960	1.930	1.956
25	2.872	2.850	2.855	2.870	2.884	2.901	2.920
26	2.957	2.940	2.898	2.873	2.857	2.840	2.804
27	3.101	3.095	3.086	3.086	3.074	3.061	3.041
28	2.464	2.417	2.374	2.320	2.113	2.035	1.968
29	2.729	2.712	2.631	2.538	2.504	2.497	2.475
30	2.462	2.462	2.489	2.456	2.423	2.435	2.444
31	2.937	2.939	2.929	2.929	2.902	2.871	2.834
32	2.976	2.966	2.972	2.967	2.936	2.914	2.894
33	3.133	3.135	3.129	3.133	3.106	3.064	3.033
34	2.552	2.536	2.528	2.493	2.436	2.446	2.464
35	2.652	2.632	2.618	2.604	2.560	2.580	2.587
36	2.716	2.697	2.698	2.708	2.810	2.786	2.794
37	2.879	2.886	2.889	2.883	2.881	2.910	2.923
39	2.513	2.489	2.462	2.438	2.404	2.389	2.392
40	2.293	2.289	2.227	2.179	2.077	2.024	2.013
41	2.312	2.251	2.296	2.309	2.224	2.230	2.241
42	2.469	2.466	2.448	2.425	2.294	2.371	2.315
44	3.099	3.086	3.098	3.101	3.114	3.089	3.085
45	2.894	2.821	2.875	2.820	2.856	2.820	2.776
46	2.839	2.968	2.977	3.000	2.975	2.949	2.944
TYPCONC	2.724	2.709	2.694	2.672	2.638	2.627	2.612
Hoover1	0.364	0.366	0.369	0.371	0.379	0.376	0.381
Hoover2	0.350	0.351	0.353	0.356	0.361	0.353	0.356
Industry Code	2006	2007	2008	2009	2010	99–05	05–10
06	2.559	2.587	2.658	2.688	2.707	−0.131	0.133
07	2.453	2.474	2.492	2.575	2.542	0.098	0.115
08	2.623	2.658	2.603	2.547	2.611	0.141	−0.067
09	2.738	2.681	2.670	2.619	2.636	−0.018	0.009
10	2.804	2.799	2.847	2.858	2.892	−0.132	0.128

continued on next page

Table A2 *continued*

Industry Code	2006	2007	2008	2009	2010	99–05	05–10
13	2.755	2.769	2.812	2.827	2.875	−0.167	0.124
14	2.897	2.892	2.918	2.919	2.971	−0.012	0.069
15	3.020	3.034	3.048	3.035	3.029	−0.013	0.030
16	2.872	2.876	2.899	2.915	2.910	0.118	0.048
17	2.267	2.277	2.280	2.313	2.354	−0.227	0.074
18	2.223	2.253	2.301	2.353	2.406	−0.030	0.187
19	2.272	2.306	2.322	2.335	2.345	−0.044	0.083
20	2.704	2.699	2.718	2.724	2.761	−0.017	0.043
21	2.427	2.419	2.431	2.486	2.545	−0.274	0.152
22	2.571	2.571	2.600	2.648	2.695	−0.204	0.118
23	2.701	2.720	2.735	2.742	2.761	−0.201	0.095
24	1.962	1.973	1.999	2.019	2.071	−0.006	0.115
25	2.930	2.958	2.971	2.987	3.006	0.048	0.087
26	2.783	2.792	2.800	2.764	2.805	−0.153	0.001
27	3.040	3.029	3.016	3.011	3.014	−0.061	−0.027
28	1.868	1.911	1.853	1.837	1.809	−0.496	−0.159
29	2.489	2.485	2.474	2.453	2.496	−0.254	0.022
30	2.455	2.454	2.524	2.566	2.601	−0.018	0.157
31	2.826	2.826	2.849	2.886	2.930	−0.103	0.096
32	2.891	2.904	2.900	2.888	2.911	−0.082	0.017
33	3.015	3.001	2.971	2.950	2.975	−0.099	−0.058
34	2.458	2.472	2.543	2.591	2.618	−0.088	0.154
35	2.577	2.597	2.620	2.628	2.661	−0.064	0.074
36	2.802	2.814	2.832	2.836	2.831	0.078	0.037
37	2.913	2.905	2.895	2.897	2.902	0.044	−0.021
39	2.406	2.437	2.483	2.535	2.552	−0.122	0.161
40	2.037	2.055	2.038	2.047	2.072	−0.280	0.059
41	2.277	2.290	2.351	2.393	2.387	−0.071	0.146
42	2.312	2.320	2.366	2.425	2.457	−0.154	0.142
44	3.085	3.090	3.121	3.133	3.142	−0.015	0.057
45	2.791	2.791	2.811	2.850	2.887	−0.118	0.111
46	2.844	2.838	2.868	2.905	2.944	0.105	0.000
TYPCONC	2.612	2.625	2.641	2.664	2.687	−0.112	0.074
Hoover1	0.383	0.384	0.380	0.378	0.376	0.017	−0.005
Hoover2	0.356	0.353	0.345	0.342	0.337	0.006	−0.019

Note: Hoover1 and Hoover2 are simple average and weighted average, respectively. Two-digit industry code in the table are explained as follows: Mining and Washing of Coal (06), Extraction of Petroleum and Natural Gas (07), Mining and Processing of Ferrous Metal Ores (08), Mining and Processing of Non-Ferrous Metal Ores (09), Mining and Processing of Nonmetal Ores (10), Processing of Food from Agricultural Products (13), Processing of Foodstuff (14), Manufacture of Beverages (15), Manufacture of Tobacco (16), Manufacture of Textile (17), Manufacture of Textile Wearing Apparel, Footwear, and Caps (18), Manufacture of Leather, Fur, Feather, and Related Products (19), Processing of Timber, Manufacture of Wood, Bamboo Rattan, Palm and Straw Products (20), Manufacture of Furniture (21), Manufacture of Paper and Paper Products (22), Printing and Reproduction of Recording Media (23), Manufacture of Articles for Culture, Education and Sport Activities (24), Processing of Petroleum, Coking, Processing of Nuclear Fuel (25), Manufacture of Raw Chemical Materials and Chemical Products (26), Manufacture of Medicines (27), Manufacture of Chemical Fibers (28), Manufacture of Rubber (29), Manufacture of Plastics (30), Manufacture of Non-metallic Mineral Products (31), Smelting and Pressing of Ferrous Metals (32), Smelting and Pressing of Non-ferrous Metals (33), Manufacture of Metal Products (34), Manufacture of General Purpose Machinery (35), Manufacture of Special Purpose Machinery (36), Manufacture of Transport Equipment (38), Manufacture of Electrical Machinery and Equipment (39), Manufacture of Communication Equipment, Computers and other Electronic Equipment (40), Manufacture of Measuring Instruments and Machinery for Cultural Activity and Office Work (41), Manufacture of Artwork and Other Manufacturing (42), Producing and Supply of Electric Power and Heat Power (44), Producing and Supply of Gas (45), Producing and Supply of Water (46).

APPENDIX 3: TIME SERIES FOR INDUSTRIAL SPECIALIZATION ($SPEC_R$) AND THE RELATIVE RATE OF CHANGE

Province Name	1999	2000	2001	2002	2003	2004	2005
Beijing	2.877	2.736	2.732	2.788	2.775	2.772	2.649
Tianjin	3.056	3.001	2.967	2.934	2.909	2.801	2.728
Hebei	3.122	3.117	3.104	3.073	2.961	2.858	2.830
Shanxi	2.538	2.570	2.530	2.523	2.437	2.379	2.374
Inner Mongolia	2.743	2.742	2.761	2.772	2.800	2.801	2.805
Liaoning	2.974	2.961	2.962	2.955	2.953	2.933	2.987
Jilin	2.633	2.586	2.498	2.344	2.193	2.321	2.399
Heilongjiang	2.229	2.342	2.429	2.507	2.578	2.670	2.679
Shanghai	3.110	3.095	3.038	3.014	2.895	2.889	2.821
Jiangsu	3.138	3.130	3.126	3.106	3.043	2.960	2.942
Zhejiang	3.175	3.186	3.172	3.154	3.165	3.142	3.143
Anhui	3.186	3.198	3.173	3.178	3.149	3.113	3.113
Fujian	3.201	3.211	3.207	3.146	3.126	3.089	3.103
Jiangxi	3.092	3.103	3.093	3.047	3.046	3.069	3.103
Shandong	3.253	3.238	3.244	3.237	3.233	3.237	3.222
Henan	3.208	3.215	3.213	3.198	3.183	3.175	3.178
Hubei	3.138	3.135	3.115	3.053	3.051	2.851	2.947
Hunan	3.122	3.141	3.157	3.176	3.162	3.137	3.175
Guangdong	3.189	3.142	3.068	2.989	2.889	2.830	2.784
Guangxi	3.032	3.023	3.005	2.930	2.869	2.804	2.835
Hainan	2.973	3.033	2.979	2.905	2.780	2.733	2.797
Chongqing	2.825	2.665	2.611	2.581	2.492	2.543	2.575
Sichuan	3.076	3.069	3.061	3.077	3.101	3.118	3.107
Guizhou	2.762	2.802	2.789	2.808	2.837	2.718	2.684
Yunnan	2.485	2.473	2.445	2.447	2.460	2.543	2.563
Tibet Autonomous Region	2.470	2.507	2.435	2.407	2.191	2.087	2.132
Shaanxi	3.009	3.019	2.986	2.990	2.980	3.007	2.996
Gansu	2.794	2.856	2.786	2.834	2.796	2.612	2.577
Qinghai	2.256	2.182	2.294	2.325	2.364	2.277	2.358
Ningxia	2.735	2.759	2.742	2.782	2.762	2.703	2.712
Xinjiang	2.107	2.222	2.313	2.367	2.399	2.392	2.351
TYPESPEC	3.053	3.045	3.030	3.011	2.972	2.940	2.931
Hoover3	0.454	0.458	0.463	0.470	0.475	0.472	0.478
Hoover4	0.360	0.359	0.363	0.371	0.370	0.365	0.368

continued on next page

Table A3 *continued*

Province Name	2006	2007	2008	2009	2010	99–05	05–10
Beijing	2.569	2.544	2.625	2.656	2.641	−0.228	−0.008
Tianjin	2.635	2.656	2.773	2.781	2.843	−0.328	0.116
Hebei	2.826	2.833	2.819	2.829	2.871	−0.292	0.041
Shanxi	2.375	2.350	2.352	2.378	2.430	−0.164	0.056
Inner Mongolia	2.820	2.871	2.884	2.935	2.966	0.063	0.160
Liaoning	3.019	3.053	3.076	3.073	3.054	0.013	0.067
Jilin	2.415	2.431	2.592	2.547	2.495	−0.234	0.096
Heilongjiang	2.735	2.777	2.842	2.862	2.925	0.449	0.246
Shanghai	2.782	2.716	2.721	2.700	2.649	−0.289	−0.171
Jiangsu	2.936	2.914	2.905	2.903	2.884	−0.196	−0.058
Zhejiang	3.140	3.132	3.127	3.123	3.122	−0.032	−0.021
Anhui	3.114	3.112	3.103	3.083	3.096	−0.074	−0.017
Fujian	3.114	3.139	3.155	3.178	3.167	−0.098	0.064
Jiangxi	3.106	3.114	3.105	3.112	3.106	0.011	0.003
Shandong	3.212	3.210	3.203	3.181	3.171	−0.032	−0.051
Henan	3.200	3.160	3.206	3.209	3.215	−0.030	0.036
Hubei	2.950	2.960	2.913	2.959	2.925	−0.191	−0.022
Hunan	3.181	3.185	3.211	3.203	3.198	0.053	0.023
Guangdong	2.769	2.801	2.794	2.802	2.783	−0.405	−0.002
Guangxi	2.843	2.878	2.899	2.876	2.924	−0.197	0.088
Hainan	2.807	2.591	2.671	2.654	2.668	−0.177	−0.128
Chongqing	2.515	2.477	2.588	2.577	2.619	−0.250	0.044
Sichuan	3.129	3.150	3.208	3.200	3.184	0.031	0.077
Guizhou	2.678	2.683	2.718	2.716	2.735	−0.078	0.051
Yunnan	2.591	2.608	2.660	2.700	2.700	0.078	0.136
Tibet Autonomous Region	2.157	2.215	2.263	2.217	2.209	−0.338	0.077
Shaanxi	2.986	2.969	2.988	3.000	2.995	−0.013	−0.001
Gansu	2.593	2.597	2.674	2.725	2.753	−0.217	0.177
Qinghai	2.399	2.446	2.496	2.546	2.519	0.102	0.161
Ningxia	2.689	2.696	2.704	2.767	2.757	−0.023	0.045
Xinjiang	2.430	2.542	2.594	2.651	2.710	0.244	0.359
TYPESPEC	2.928	2.931	2.948	2.954	2.947	−0.122	0.015
Hoover3	0.479	0.476	0.467	0.459	0.452	0.024	−0.026
Hoover4	0.366	0.362	0.355	0.351	0.346	0.009	−0.022

Note: Hoover3 and Hoover4 are simple average and weighted average, respectively.

APPENDIX 4: DEFINITIONS AND SUMMARY STATISTICS OF KEY VARIABLES

Name	Definition	N	Mean	SD	Min	Max
Δp_{ir}	$\Delta p_{ir} = (X_{ir2010}/X_{ir2010} - X_{ir2005}/X_{ir2005}) \cdot 100\%$, change of industry i's output in province r out of the gross output of province r from 2005 to 2010.	1,147	0.000	1.691	-15.579	21.752
Δs_{ir}	$\Delta s_{ir} = (X_{ir2010}/X_{ir2010} - X_{ir2005}/X_{ir2005}) \cdot 100\%$, change of industry i's output in province r out of the gross output of industry i from 2005 to 2010.	1,147	0.000	1.494	-12.384	10.454
$plan_{ir}$	Dummy variable, $plan_{ir}=1$ for industry chosen as priority one in province r's 11th 5-year plan, otherwise $plan_{ir}=0$.	1,147	0.462	0.499	0	1
Δp_{irL}	$\Delta p_{irL} = (X_{ir2004}/X_{ir2004} - X_{ir1999}/X_{ir1999}) \cdot 100\%$, change of industry i's output in province r out of the gross output of province r from 1999 to 2004.	1,147	0.000	2.068	-21.597	14.319
Δs_{irL}	$\Delta s_{irL} = (X_{ir2004}/X_{ir2004} - X_{ir1999}/X_{ir1999}) \cdot 100\%$, change of industry i's output in province r out of the gross output of industry i from 1999 to 2004.	1,147	0.000	0.605	-5.764	4.607
ar_{irL3}	$ar_{irL3} = [(X_{ir2005}/X_{ir2002})^{1/3} - 1] \cdot 100\%$, growth rate of industry i's output in province r from 2002 to 2005.	1,110	1.960	13.008	-1	425.5
ar_{irL6}	$ar_{irL6} = [(X_{ir2004}/X_{ir1999})^{1/5} - 1] \cdot 100\%$, growth rate of industry i's output in province r from 1999 to 2004.	1,110	0.227	0.351	-1	3.659
$Rplan$	Dummy variable, $rplan=1$ for provinces listed in regional development plan during the 11th 5-year plan, otherwise $rplan=0$.	1,147	0.677	0.468	0	1
$Cplan$	Dummy variable, $cplan=1$ for industry chosen as priority one in the central government's 11th 5-year plan, otherwise $cplan=0$.	1,147	0.459	0.499	0	1
p_{ir03}	$p_{ir03} = X_{ir2003}/X_{ir2003} \cdot 100\%$, share of industry i's output in province r out of the gross output of province r in 2003.	1,147	2.703	4.384	0	48.005
s_{ir03}	$s_{ir03} = X_{ir2003}/X_{ir2003} \cdot 100\%$, share of industry i's output in province r out of the gross output of industry i in 2003.	1,147	3.226	5.077	0	37.778
scl_{ir04}	$scl_{ir04} = (X_{ir2004}/X_{ir2004}) \cdot 100\%$, share of output of industry i's small enterprise in province r out of the gross output of industry i in 2004.	924	0.473	0.396	-7.64	1
$firm_{ir}$	Number of province r's firms in industry i in 2000. (100 firms)	1,144	1.371	2.333	0	20.56
sfm_{ir}	quadratic term of $firm_{ir}$	1,144	7.319	30.304	0	422.714
$isfm_i$	Number of firms in industry i in 2000. (10,000 firms)	1,147	0.293	0.418	0	1.894
$lngdp_r$	Logarithm of gross regional product of province r in 2003.	1,147	8.939	1.081	6.027	10.737
$lnpop_r$	Logarithm of population of province r in 2004.	1,147	8.036	0.886	5.598	9.176
$trans_r$	$trans_r = \sum b_k \Delta tr_{rk} / area_r$, Δtr_{rk} stands for change of province r's length of type k (road, rail, river) transportation infrastructure during the implementation period of the 11th 5-year plan, b_k stands for the share of cargo turnover of type k (road, rail, river) transportation infrastructure, $area_r$ stands for area of province r (10,000 km ²).	1,147	1.129	1.039	0.003	4.336
$ntrans_r$	$ntrans_r = \sum a_{ij} \sum b_k \Delta tr_{rk} / \sum a_{ij} \cdot area_r$, Δtr_{rk} , b_k and $area_r$ as above, $a_{ij}=1$, if province r is adjacent to province j, then $a_{ij}=1$, otherwise $a_{ij}=0$	1,147	0.111	0.072	0.000	0.259
$plan_{im}$	Dummy variable, $plan_{im}=1$ for industry chosen as priority one in province r's n-th 5-year plan, otherwise $plan_{im}=0$.	2,294	0.417	0.493	0	1

continued on next page

Table A4 continued

Name	Definition	N	Mean	SD	Min	Max
Δp_{irm}	$\Delta p_{irm} = (x_{ir,t}/x_{ir,t-5} - x_{ir,t-5}/x_{ir,t-5}) \cdot 100\%$, change of industry i's output in province r out of the gross output of province r from year t-5 to t, year t is the last implementation year of the n-th 5-year plan.	2,294	0.000	1.780	-15.677	21.751
Δs_{irm}	$\Delta s_{irm} = (x_{ir,t}/x_{ir,t-5} - x_{ir,t-5}/x_{ir,t-5}) \cdot 100\%$, change of industry i's output in province r out of the gross output of industry i from year t-5 to t, with t as above.	2,294	0.000	1.663	-16.992	23.260
fm_{irmL}	number of province r's firms in industry i in year t-7, t as above. (100 firms)	2,254	1.899	4.116	0	66.47
x_{irmL}	output of industry i in province r in year t-5, t as above (billion yuan)	2,294	1.369	3.89	0	108.477
x_{irmL2}	output of industry i in province r in year t-6, t as above (billion yuan)	2,294	1.154	3.154	0	85.172
ar_{irmL}	$ar_{irmL} = (x_{ir,t-5}/x_{ir,t-5}) - 1$, t as above.	2,211	0.192	0.574	-1	12.5
s_{irmL}	share of industry i's output in province r out of the gross output of industry i in year t-6, t as above.	2,294	3.226	4.937	0	37.431
$Intrs_{rm}$	$Intrs_{rm} = \ln(\sum b_{kn} \cdot \Delta tr_{rkn} / area_r)$, Δtr_{rkn} stands for change of province r's length of type k (road, rail, river) transportation infrastructure during the implementation period of the n-th 5-year plan, b_{kn} stands for the share of cargo turnover of type k (road, rail, river) transportation infrastructure $area_r$ stands for area of province r (10,000 km ²).	2,294	5.357	1.758	0	8.375