China Economic Review 23 (2012) 593–612

Contents lists available at SciVerse ScienceDirect

China Economic Review

Patterns of China's industrialization: Concentration, specialization, and clustering☆

Cheryl LONG [a,b,*], Xiaobo ZHANG [c,d]

[a] Colgate University, United States
[b] WISE, Xiamen University, China
[c] International Food Policy Research Institute, United States
[d] Zhejiang University, China

ARTICLE INFO

Article history:
Received 30 September 2010
Received in revised form 18 September 2011
Accepted 18 September 2011
Available online 24 September 2011

JEL classification:
L1
O1
R1

Keywords:
Industrial clustering
Agglomeration
Geographic concentration
Regional specialization
China

ABSTRACT

This paper presents a few stylized facts on the patterns of China's industrialization by computing a set of multi-dimensional measures on industrial concentration, regional specialization, and clustering based on census data at the firm level in 1995 and 2004. Our results show that China's rapid industrialization is characterized by the following patterns: industries have become more spatially concentrated; regions have become increasingly specialized; and firms have become more interconnected, both within industries and within regions. In addition, the number of firms is growing faster in clustered areas than non-clustered ones. Together these patterns suggest that China's industrialization process is largely cluster-based—a phenomenon in which a large number of highly interconnected firms are located within a well-defined geographic region.

1. Introduction

In the past three decades, China has experienced the same degree of industrialization that took two centuries to occur in Europe (Summers, 2007). The rapid industrialization has been accompanied by the emergence of numerous "specialty cities". Thousands of firms, large and small, each specialized in a finely defined production step, are lumped together in a densely populated region, where some particular manufactured consumer good is churned out in millions (if not billions) annually. Many formerly rural towns in the coastal areas have become so specialized that they proclaim themselves to be the world's Socks City, Sweater City, Kid's Clothing City, Footwear Capital, and so on. Despite numerous popular media reports on this phenomena, few studies have rigorously documented these patterns using data covering a large sample over a long period.[1]

Each of the specialty cities described above fits Porter's concept of an industrial cluster, which is "a geographically proximate group of inter-connected companies (and associated institutions) in a particular field" (Porter, 2000, 6). Is increasing clustering a general pattern of China's industrialization and economic growth over the past three decades? In this paper, we plan to explore

☆ Funding supports from the Faculty Research Council at Colgate University, the International Food Policy Research Institute, the Natural Science Foundation of China (Approval number 70828002), Ministry of Education of China (Approval number 08JJD840206) and Zhejiang Province Modern Agricultural and Rural Development Innovation Fund are gratefully acknowledged.
* Corresponding author at: Colgate University, United States.
E-mail address: cxlong@colgate.edu (C. Long).
[1] For example, see http://www.nytimes.com/2004/12/24/business/worldbusiness/24china.html for a *New York Times* report.

1043-951X/$ – see front matter
doi:10.1016/j.chieco.2011.09.002

this issue by conducting the following tasks. First, we will propose new measures for industrial clustering that capture the degree of interconnectedness among firms within the cluster. Next, we will document the patterns of clustering in China by constructing three sets of multi-dimensional measures: (1) the conventional concentration measure for geographic concentration of industries, (2) an extension of the concentration measure to gage the degree of regional specialization, and (3) the clustering measure proposed above. Finally, we will evaluate the drivers behind clustering in China.

Geographic concentration of industries (or industrial concentration) and regional specialization are sometimes used interchangeably in the literature (for example, Bai, Duan, Tao, & Tong, 2004). In our view, however, regional specialization is better reserved for describing how much a region chooses to focus on production in a limited number of industries. This is quite distinct from an industry's total production being concentrated in a small number of regions, which is the defining feature of industrial concentration.

In contrast to the two concepts discussed above, agglomeration is often used in the literature of spatial economics. It not only emphasizes how specialized a region is, but it also relates to how close (and consequently interconnected) the firms are within a region. However, because the degree of agglomeration is usually measured by the regional concentration of industries, these two concepts are often used interchangeably in the literature. To highlight the importance of interconnectedness among firms and industries, we will reserve the usage of clustering throughout this paper for the case where firms (and other related institutions) within a well-defined geographic region maintain a high level of interaction among themselves. In contrast, we will use *agglomeration* in a more general sense (i.e., any place witnessing high industrial concentration, also referred as *spatial agglomeration*), and will treat *clustering* as a special case of agglomeration where the interaction among firms is an integral feature. For a graphic illustration for how these concepts relate to one another, see Diagram 1.

The definition of clustering thus suggests that the phenomenon is usually accompanied by regional specialization, a relationship that follows naturally from the high interconnectedness among firms in the same region of clustering. In addition, increased geographic concentration of industries may also occur in successful clusters thanks to the increased productivity of firms in the cluster and their enhanced competitiveness on the national market. Consequently, multi-dimensional variables are needed to fully describe the phenomenon of clustering–including the conventional ones measuring regional specialization and industrial concentration–as well as those documenting interactions among firms and industries. The aforementioned is the approach taken in the current study.

The structure of this paper is as follows. Section 2 reviews the literature on the measurement of these related phenomena and that on China's industrialization. Section 3 describes the data and the specific procedures for constructing the various measures, while Section 4 presents the patterns of China's industrialization based on these measures. Section 5 provides a preliminary analysis of why China has seen increased clustering during 1995–2004, and a short conclusion is offered in Section 6.

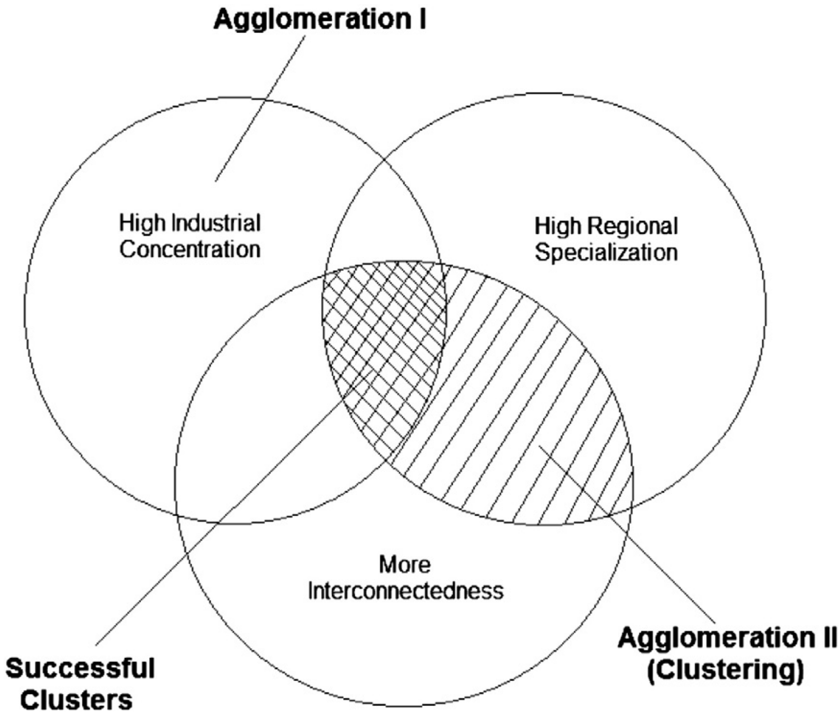

Diagram 1. Regional Specialization within Industry (1995 v. 2004): CR3. Note: The left figure presents the distributions of the Krugman Gini coefficient in 1995 versus 2004 computed at the county level, while the right figure presents the corresponding distributions for the proximity measures. All the data is computed by authors based on China National Bureau of Statistics, 1995 and China National Bureau of Statistics, 2004.

2. Literature review

The first strand of literature relating to our paper is on industrial districts. As highlighted by Marshall (1920), industrial districts generate several positive externalities for firms: better access to the market and suppliers, labor pooling, and easy flow of technological know-how. Along this line of argument, Fujita, Krugman, and Venables (2001) stress that spatial concentration is a key feature of industrialization; indeed, industrialization is often accompanied by the spatial agglomeration of industrial activities.

The industrialization process in Italy, Japan, and other East Asian countries and regions closely follow the Marshallian industrial district model (Becattini, 1990). In this development model, a large number of small and medium enterprises (SMEs) often cluster together with comprehensive vertical division of labor. The Marshallian industrial district model was popular in the early era of European industrialization. One noted example is the putting-out system in the U.K. prior to its industrialization, in which a merchant obtained market orders and subcontracted the production to nearby farmers or skilled workers who usually finished the work in their homes or family workshops (Hounshell, 1984). Industrial districts where different workshops and factories cluster together were ubiquitous in France and Italy until the mid-twentieth century and are still viable in some regions of Italy (Piore & Sabel, 1984; Porter, 1998). The putting out system was also widely observed in nineteenth-century Japan (Nakabayashi, 2006). Outsourcing (or subcontracting), the modern variant of the traditional putting-out system, remains a major feature of industrial production organization in contemporary Japan and Taiwan (Sonobe & Otsuka, 2006a, 2006b).

However, the patterns of industrialization in the later period in the US and UK seem to reject the Marshallian industrial district model. In the U.K., the decentralized production system scattered in family workshops was replaced by a large integrated factory system during the Industrial Revolution (Landes, 1998). The trend was similar and more evident in the U.S. during its industrialization in the 19th century (Chandler, 1977). The auto industry, for example, became highly concentrated in the Detroit metropolitan area with several dominant firms. Markusen (1996) proposed three additional types of industrial districts: a hub-and-spoke industrial district, centering on a few dominant externally oriented firms; a satellite platform—a congregation of branch facilities of externally based multi-plant firms; and a state-anchored district, which rely on one or more public institutions. The Detroit auto cluster fits into the hub-and-spoke model. In the US, the hub-and-spoke model is the most popular mode of clustering. In this paper, we will focus our discussions on Marshallian and hub-and-spoke models, as the two types of clusters are less prevalent in China.

The key difference between the Marshallian and hub-and-spoke clustering models is that the first encompasses a finer vertical division of labor in the production process. By dividing a production process into incremental stages, a large lump-sum investment can be divided into many small amounts, therefore lowering the capital entry barriers (Long & Zhang, 2011; Ruan & Zhang, 2009; Schmitz, 1995). Therefore, this mode of industrial organization is more widely observed in countries or regions with scarce capital, abundant labor, and a less developed financial sector.[2]

Unfortunately, the existing literature has not provided much insight on how to measure and distinguish between the two types of industrial districts described above. Instead, most studies rely on various measures of the geographic concentration of industries. For example, the market share in output of the top, say, three regions in an industry is often used as a concentration measure. The advantage of this measure is that it is easy to calculate and interpret, but when the distribution of firms is relatively spread out, it may miss those regions below the cut-off lines. To overcome this problem, the Gini coefficient is often used to calculate the regional variation of output shares for all of the firms in an industry. Krugman (1991) further modifies the Gini coefficient by accounting for the discrepancy between a region's share of output in an industry and its share in all industries in calculating the Gini coefficient.

Obviously these concentration measures do not distinguish between the two kinds of industrial districts: one where a small number of large firms with minimal inter-firm connections are located, versus the other where a large number of small firms congregate and interact closely with one another, which we refer to as clusters. While the first type characterizes cities such as Detroit, the second type of industrial district, or clustering, is the one that better fits the patterns observed in coastal China (see Diagram 1). The first contribution of the current paper is thus to propose and construct measures that capture the feature of interconnectedness in clusters.

The second strand of literature relevant to the current study is the debate on China's industrialization, for which previous studies seem to have provided mixed conclusions related to two questions. First, have Chinese regions become more or less specialized? Second, have Chinese industries become more or less geographically concentrated?

On the issue on regional specialization, Young (2000) found that China's regions had become less specialized up to the early 1990s among the broadly defined sectors of agriculture, industry, construction, transportation, and commerce (or at the even more aggregated primary, secondary, and tertiary sectors). These results are interpreted as evidence for high inter-regional trade barriers in China, which had obstructed regional specialization among industries and hindered the division of labor among regions, thus leading to more market fragmentation, at least till the early 1990s. Yet several subsequent studies produced findings opposite to those in Young (2000) (Huang, Rozelle, & Chang, 2004; Wei & Fan, 2006; Zhang & Tan, 2007). In particular, using data similar to Young (2000) but with a longer and more recent period, Zhang and Tan (2007) showed that the regional distribution of industries initially became more dispersed before growing more concentrated after 1991. Thus, the increasing market fragmentation up to the early 1990s documented in Young (2000) may have been reversed later on.

Related to the issue of agglomeration (or geographic concentration of industries), using data from the Second and the Third national industrial censuses, Wen (2004) showed that Chinese manufacturing industries have become increasingly geographically concentrated up to 1995. Based on a panel data set of 32 industries in 29 provinces, Bai et al. (2004) similarly found that the geographic concentration of

[2] This does not mean that Marshallian clusters are absent in developed countries. The Marshallian industrial districts can still be observed in developed countries. For example, the high-tech clusters in the biotech sector in the US are largely dominated by small firms.

industrial production has become increased for the period of 1984–1997. Although they do not directly address the same question as Young (2000), the findings from both studies suggest a more integrated market in China, which is in contrast to Young's results.[3]

The mixed findings from previous studies may be due to their different levels of aggregation and time periods covered. In addition, these studies all use data from the earlier time period of the 1980s or the 1990s. Thus an additional contribution that the current paper can make to the literature is to extend the previous analysis to a more recent period and shed more light on the debate described above, given our access to firm-level data nationwide for a longer time period including more recent years, which can be aggregated at different levels.

Several in-depth case studies and popular media reports suggest that China has followed the cluster-based industrialization path (Huang, Zhang, & Zhu, 2008; Ruan & Zhang, 2009; Sonobe, Hu, & Otsuka, 2002, 2004). However, there is a large degree of diversity among clusters in different places. For example, in Guangdong province, there are hundreds of specialized towns (Bellandi & Di Tommaso, 2005; Enright, Scott, & Chang, 2005). Some of them are structured around foreign firms or big joint ventures with multinational firms, while others are dominated by SMEs, some of which were converted from town-village enterprises. In Zhejiang province, however, most clusters originated from local entrepreneurs (Marukawa, 2006) and the share of the output values from FDI accounts for less than 5% in 2007 (Wang & Mei, 2009). Given China's vast size and heterogeneity across regions, it is unclear whether clusters in China generally adhere more closely to the hub-and-spoke model or the Marshallian industrial district model.

However, to our knowledge, there have been no systematic quantitative studies based upon large samples to empirically test which type of agglomeration China has followed during its course of rapid industrialization. [4]

For example, Sonobe et al. (2002) studied how a garment cluster formed in a rural town in Zhejiang Province starting with small-scale production. Fleisher, Hu, McGuire, and Zhang (2010) conducted a follow up survey on the same cluster to study the quality upgrade process of the cluster. Huang et al. (2008) detailed how the footwear cluster in Wenzhou helped overcome financial, institutional, and technological barriers. Ruan and Zhang (2009) in particular demonstrated that clustering lowered capital entry barriers and enabled more entrepreneurs to participate in the production process. These studies provide insight into both the emergence of industrial clusters in China and how these clusters have worked to facilitate further growth in their regions. However, it is not clear whether these case studies can be generalized and thus help shed light on the evolution of China's industrialization over time.

To fill in these gaps in the literature, we will first propose measures that better capture the feature of clustering, i.e., the interconnectedness among firms in the same cluster. Furthermore, we will compare these clustering measures with the more conventional measures on geographic concentration and regional specialization, based on a comprehensive data set comprising of all industrial firms in China in 1995 and 2004. Thus, we are able to document the general patterns of China's industrialization and better describe the various features of clustering in China using more wide-ranging measures.

3. Data and measures of clustering

We begin by describing the data, and then move on to propose new measures for interconnectedness: industrial proximity (based on Hausmann and Klinger's work), and the ratio between value added and output. The rest of the section discusses the construction of the new and conventional measures for clustering.

3.1. Data description

We utilize firm-level data from the China National Bureau of Statistics, 1995 and the China National Bureau of Statistics, 2004 for analysis in this paper. Compared to datasets used in previous studies on China's industrialization patterns (Bai et al., 2004; Lu & Tao, 2009; Wen, 2004; Young, 2000; Zhang & Tan, 2007), our datasets have more comprehensive coverage with respect to both time and the number of firms—spanning a time period of nine years and including industrial firms of all sizes (not only those above a certain scale). Table 1 and Table 2 present summary statistics of gross industrial output from the censuses by industry and by region, respectively. Table 3 compares the sample of our data sets with the published national aggregate statistics for China in 1995 and 2004. As shown in the table, our data sets capture the whole universe of Chinese industrial firms in these two years.

Since the data is at the firm level, we can aggregate them at any level of our choice, such as county, prefecture, or province, for regional aggregation, and 2, 3, or 4-digit industry level for sectoral aggregation. For the main part of the analysis, we choose province and 2-digit CIC (China Industry Code) as the levels of aggregation. But for robustness tests, we also use prefecture and county levels for geographic aggregation, and 3-digit and 4-digit CIC for industrial aggregation.

Because China modified its industry coding system in 2002 (switching from GB1994 to GB2002), we match industry codes that have changed from 1994 to 2002 as follows: for industry codes that have become more disaggregated, we use the 1994 codes as the standard; for those that have become more aggregated, we use the 2002 codes as the standard. In other words, we use the more aggregate codes to group and compare industries between 1995 and 2004. During the period between the two censuses

[3] Bai et al. (2004) use the terms “regional specialization” and “geographic concentration of industries” interchangeably (see page 400 in Bai et al., 2004), which is in contrast to our distinct treatments of the terms.

[4] In spirit, our study is similar to Lu and Tao (2009) which is based on a sample of large firms. They found that the extent of industrial agglomeration in China has increased steadily from 1998 to 2005. However, their sample excludes the small and medium firms and focuses only on regional concentration measures.

Table 1
Summary statistics of firm gross industrial output by 2-digit industry.

Industry	1995			2004		
	Mean	sd	No. of firms	Mean	sd	No. of firms
Coal mining and dressing	9664	96,605	11,953	17,643	224,515	26,822
Petroleum and natural gas extraction	1,066,011	4,254,463	134	962,613	5,758,893	481
Ferrous metals mining and dressing	5228	25,019	2141	9554	46,737	10,256
Nonferrous metals mining and dressing	8554	36,978	3766	14,919	104,298	6075
Non-metallic mineral mining and dressing	3087	11,114	11,820	3293	19,150	34,945
Other minerals mining and dressing	2515	4604	149	3948	23,920	263
Foodstuff processing industry	10,042	40,142	30,962	13,719	105,609	69,521
Foodstuff manufacturing industry	5856	34,955	16,313	11,026	87,351	29,811
Beverage manufacturing industry	7852	43,376	14,719	10,740	96,640	25,485
Tobacco processing	237,406	902,942	423	885,794	2,275,005	281
Spinning industry	18,002	53,656	24,459	14,029	100,348	83,011
Manufacturers of clothes and other fiber products	7453	25,219	18,937	9671	58,140	48,250
Leather, fur, feather and other products	9308	30,236	10,468	13,816	59,022	22,677
Timber processing and bamboo, cane, palm, straw products	2620	12,020	15,480	5072	32,941	37,028
Furniture	2580	9564	8760	6255	39,320	23,892
Paper makers and paper products	7304	27,232	13,890	10,005	83,103	39,669
Printing and record medium reproduction	2553	9898	16,763	4234	21,719	44,070
Teaching and sport products for daily use	8575	31,219	5356	9702	42,732	14,711
Oil processing and refining	73,925	716,204	2744	126,789	1,373,127	7146
Chemical material and products	13,750	116,761	26,872	19,175	239,355	69,120
Pharmaceutical and medicine manufacturing	16,527	61,057	6051	29,861	148,194	11,271
Chemical fibers	76,277	474,156	1034	59,128	329,805	3372
Rubber products	13,294	74,614	4663	13,490	128,384	15,178
Plastic products	5856	20,254	19,255	7573	44,341	69,729
Non-metallic mineral products	4926	18,678	61,278	6306	32,082	157,734
Smelting and pressing of ferrous metals	44,108	479,425	8429	84,284	961,069	20,494
Smelting and pressing of non-ferrous metals	29,697	143,612	4621	41,174	259,277	15,162
Metal products	5534	24,407	26,744	7849	52,709	80,976
Common machines	7719	41,075	31,474	9032	80,815	113,691
Special equipment	10,805	69,347	18,391	10,556	84,887	55,095
Traffic equipment	17,009	216,986	19,522	27,664	475,230	51,844
Manufacture of electrical machinery and apparatus	13,206	74,917	18,928	20,680	249,967	54,979
Electrical machines and equipment	34,343	250,416	5489	74,793	1,048,297	15,211
Electronic and communication equipment	12,552	75,699	9735	40,516	532,205	35,203
Instruments, culture and office devices	4576	17,211	12,127	6966	38,783	26,627
Recycling of material waste and scrap	2799	10,227	4440	4491	28,720	6156
Electricity, steam, thermal power production and supply	19,369	107,370	12,600	60,653	1,052,128	24,568
Coal gas production and supply	20,474	82,663	372	30,310	130,000	1445
Tap water production and supply	3545	29,942	5147	5058	34,173	11,035
Total	10,763	134,410	506,409	16,198	310,686	1,363,284

Calculated by authors based on China National Bureau of Statistics, 1995 and China National Bureau of Statistics, 2004. The gross industrial output is reported in 1000 s of RMB at current price.
*** Statistically significant at 1% level.

(1995–2004), some counties have also been elevated to cities and have changed their names. We have carefully tracked these changes to match the counties throughout the time period.

3.2. A new measure of interconnectedness: industry proximity

As discussed before, the conventional measures used in the agglomeration literature do not adequately describe the type of agglomeration observed in China, which we refer to as clustering. Although the clustering concept is intuitive and easy to understand, the measurement of interconnectedness seems more elusive. This is mainly because the ideal information at the firm level needed for measuring interconnectedness is largely absent in reality, as such information will require detailed accounts on who and how often the firms interact with on a regular basis.

Lacking the first-best information, we read Porter's concept of clustering more carefully to explore alternative ways of measuring interconnectedness among firms. When delineating the main actors within a cluster, Porter states, “They include, for example, *suppliers of specialized inputs such as components, machinery, and services as well as providers of specialized infrastructure*. Clusters also often extend downstream to channels or customers and laterally to *manufacturers of complementary products or companies related by skills, technologies, or common inputs*” (Porter, 2000, 16–17, italics added by authors). In addition, Porter emphasizes that one main benefit derived from geographically concentrated clusters is that industries in the same cluster share common technologies, skills, knowledge, inputs, and institutions.

The statements cited above suggest one way to measure interconnectedness as envisioned in the cluster concept by Porter. If industries and firms produce similar goods, then they are more likely to use similar combinations of inputs in their production

Table 2
Summary statistics of firm gross industrial output by province.

	1995			2004		
Province	Mean	sd	Number of firms	Mean	sd	Number of firms
Beijing	15,067	256,247	9623	18,926	418,528	31,364
Tianjin	13,638	155,565	10,735	23,949	500,129	25,432
Hebei	9216	80,343	23,592	15,789	201,284	64,062
Shanxi	8538	91,723	11,416	14,490	216,240	28,641
Neimeng	5559	83,740	9432	19,689	223,747	11,759
Liaoning	10,184	171,331	29,435	16,844	365,899	54,115
Jilin	7751	167,292	13,100	22,085	542,689	16,037
Heilongjiang	8524	308,216	18,745	19,613	738,370	20,101
Shanghai	23,260	273,997	16,690	26,263	499,886	55,315
Jiangsu	15,815	106,861	41,582	15,618	262,800	187,212
Zhejiang	10,363	58,130	32,725	11,236	173,580	187,588
Anhui	6912	64,940	23,474	10,808	189,114	38,827
Fujian	8080	47,063	19,038	15,126	225,394	49,532
Jiangxi	4528	45,443	18,253	9331	146,981	29,144
Shandong	17,466	179,590	26,980	20,477	303,314	119,699
Henan	9703	75,318	23,119	12,065	164,347	76,292
Hubei	10,432	139,562	20,881	18,191	359,619	28,937
Hunan	5738	66,417	23,720	9668	145,047	43,529
Gongdong	17,715	114,052	34,536	22,969	473,908	136,606
Gongxi	7719	42,786	12,312	11,870	155,918	18,753
Hainan	9932	52,781	1278	21,086	198,547	2025
Chongqing	6676	82,141	11,456	12,677	149,313	20,359
Sichuan	6675	74,866	26,380	12,137	168,971	43,325
Guizhou	5500	48,962	7450	13,831	178,497	10,996
Yunnan	13,970	223,904	6267	16,157	239,845	14,271
Tibet	2343	6554	295	7004	22,096	354
Shaanxi	6182	60,000	12,950	12,251	209,434	25,573
Gansu	8260	109,848	7140	14,648	305,597	11,549
Qinghai	8193	77,821	1446	17,524	218,482	2168
Ningxia	9011	56,606	1706	15,132	151,483	3984
Xinjiang	9990	187,435	5077	28,813	441,176	5735
Total	10,725	134,909	500,833	16,198	310,686	1,363,284

Calculated by authors based on China National Bureau of Statistics, 1995 and China National Bureau of Statistics, 2004. The gross industrial output is reported in 1000 s of RMB at current price.

processes, and more likely to rely on the same set of suppliers and clients, and thus are more likely to be interconnected through skills, technologies, and other common inputs. The similarity among products of industries can thus be used as a measure for clustering, as defined by Porter.

The proximity measures proposed by Hausmann and Klinger (2007) allow us to implement the above measure of interconnectedness among industries (and participating firms) in a cluster. They constructed a proximity matrix for all four-digit SITC products, in which the proximity between any two goods captures their similarity in the following sense: If the two goods need the same combination of inputs (or endowments and capabilities) to produce, then there is a higher probability that a country has a comparative advantage in both, and the two products are more likely to be both exported. In other words, the proximity between each pair of goods can be computed as the

Table 3
Comparing the aggregated output values from the two censuses with published figures.

	Gross Industrial Output (Trillions of RMB, at current price)		
	Sample (1)	Statistical Yearbook (2)	(1)/(2)*100%
1995	5.495	5.528	99.438
2004	20.174	18.722	107.754
	Industrial value added (Trillions of RMB, at current price)		
	Sample (3)	Statistical yearbook (4)	(3)/(4)*100%
1995	1.536	1.545	99.445
2004	5.678	5.481	103.604

The figures under the headings of Sample (2) and (3) are calculated by authors based on China National Bureau of Statistics, 1995 and China National Bureau of Statistics, 2004. The official figures for gross industrial output and industrial value added under the headings of Statistical Yearbooks (2) and (4) are from *China Statistical Yearbook* in 1996 and 2005. However, the official figures in the China National Bureau of statistics, 2005 do not include non-state owned small enterprises below a certain scale. Therefore, the ratio of the tabulated to official figures exceeds one in 2004.

probability that a country has net exports in both, averaged over all countries in the world (See the Appendix for a more detailed discussion on how the Hausmann–Klinger proximity matrix is constructed and what advantages it has in measuring industry clustering).

It follows then that firms and industries that produce products with a higher proximity are more likely to interact with one another in various ways, including dependence on similar inputs (be they raw materials, labor, or machinery), reliance on similar technologies and research and development, and even dependence on the same supply or marketing facilities. Thus, those industries producing commodities that are more proximate in the Hausmann–Klinger space are likely to be more interconnected in the Porter sense. As a result, this proximity measure can be used to provide a gage for how closely interconnected industries and their participating firms are within a specific region.

To implement the idea of measuring interconnectedness among firms using product proximity, we begin with the Hausmann and Klinger proximity matrix. Because the matrix is computed for products at the four-digit SITC level, we have made a concerted effort to convert the CICs first to ISICs and then to SITCs based on the manuals obtained from China's National Bureau of Statistics as well as correspondence tables from Eurostat and the United Nations. Given that both CICs and ISICs are industry-level codes, while SITCs are based on products, there are many cases in which we have one industry corresponding to more than one product. In such cases, we assign equal weights (that sum up to 1) to all the products made in the industry.

We then follow the procedures below to construct the interconnectedness measure, which we will refer to as the proximity measure henceforth: (1) aggregate firm level data to the cell level, where the cell is defined as a combination of province (or county) and a 4-digit CIC industry; (2) for each industry in a region, calculate its average proximity to all industries located in the same region, using the Hausmann–Klinger product proximity matrix, where the average proximity (for each industry) is computed as a weighted average using the size of the other industry in each pair as the weight; (3) finally, the average industry proximity for each region is computed as the average of the (average) proximities of all the industries in that region, weighted by the size of each industry.

The proximity measure can be based on output, employment, or assets, depending on the weights used to adjust for the size of each industry. We could potentially use all of these measures, as they provide different angles of clustering and interconnectedness. Among the three advantages outlined by Marshall (1920), output-weighted proximity is probably more conducive to technological spillovers, since the output can be used as input in the production of other industries in the same region, while employment-weighted proximity implies more labor-market pooling, and asset-weighted proximity more specialized supplies, especially in capital goods. All of these effects of agglomeration will lead to higher productivity at the firm level, thus revealing the positive spillover effects of clustering.[5]

3.3. Other measures of clustering

While the proximity measure described above provides information on how industries within a certain region are technologically close to one another, the value added/output ratio measures how close firms are to one another technologically within an industry. As the ratio increases, the number of transactions among firms within industries drops, implying a lower level of interconnectedness. Thus there is an inverse relationship between the average value added/output ratio among firms in the same industry and the level of interaction in that industry.

To better distinguish the Detroit style versus the Chinese style of clustering, we also compute one more set of measures—the per region number of firms and the regional average firm size for each industry, as well as their geographic distributions. In addition, we compute the conventional concentration measures to study both the geographic concentration of industries and regional specialization, which also permit comparison with the existing findings.

In particular, the concentration indices we use include the total share of gross industrial output (asset or employment) contributed by the top three industries or regions (referred to as CR3 henceforth), as well as the corresponding Gini coefficient and Krugman–Gini coefficient. When these measures are computed for each industry (thus with regions as producers), they are referred to as industrial concentration (or geographic concentration of industry) indices. Similar measures become indicators for regional specialization when they are computed for each region (i.e., with industries as producers).

In summary, we have three sets of measures describing the geographic concentration of Chinese industries (CR3, Gini, and Krugman–Gini of regional output for each industry), the distribution of their numbers and sizes, and interaction of firms within the same industry (the average value added/output ratio for each industry). We also have two measures describing the regional specialization in China (CR3, Gini, and Krugman–Gini of sectoral output for each region) and the interaction of firms within the same region (industry proximity for each region). Next we will discuss the patterns of China's industrialization based on these measures.

4. Patterns of China's industrialization

The above measures reveal several patterns regarding China's industrialization process between 1995 and 2004. We begin with patterns along the sectoral dimension (4.1–4.3) and then present those along the regional dimension (4.4 and 4.5). Although the discussion below is based on measures using output, results based on asset values are very similar. Measures using employment data tend to give different patterns, but the employment-based results are less dependable, as the state-owned enterprises

[5] See Long and Zhang (2011) for more detailed discussions.

Table 4
Industrial concentration.

Industry	CR3_2004	CR3_1995	yr2004_gini	yr1995_gini	yr2004_Kgini	yr1995_Kgini
Coal mining and dressing	0.495	0.390	0.610	0.574	0.556	0.567
Petroleum and natural gas extraction	0.486	0.570	0.632	0.651	0.712	0.655
Ferrous metals mining and dressing	0.486	0.422	0.607	0.574	0.612	0.772
Nonferrous metals mining and dressing	0.488	0.369	0.638	0.582	0.542	0.579
Non-metallic mineral mining and dressing	0.348	0.322	0.524	0.540	0.382	0.483
Other minerals mining and dressing	0.612	0.488	0.746	0.598	0.750	0.592
Foodstuff processing industry	0.419	0.302	0.574	0.470	0.326	0.244
Foodstuff manufacturing industry	0.341	0.307	0.530	0.512	0.337	0.336
Beverage manufacturing industry	0.319	0.305	0.487	0.483	0.352	0.259
Tobacco processing	0.367	0.457	0.525	0.614	0.626	0.688
Spinning industry	0.629	0.434	0.748	0.641	0.438	0.306
Manufacturers of clothes and other fiber products	0.591	0.511	0.770	0.716	0.595	0.465
Leather, fur, feather and other products	0.593	0.431	0.771	0.665	0.596	0.441
Timber processing and bamboo, cane, palm, straw products	0.414	0.290	0.617	0.526	0.441	0.531
Furniture	0.500	0.373	0.693	0.559	0.395	0.292
Paper makers and paper products	0.486	0.297	0.666	0.481	0.365	0.226
Printing and record medium reproduction	0.443	0.352	0.609	0.504	0.323	0.308
Teaching and sport products for daily use	0.659	0.559	0.803	0.693	0.658	0.445
Oil processing and refining	0.338	0.457	0.529	0.632	0.517	0.493
Chemical material and products	0.420	0.349	0.570	0.512	0.192	0.215
Pharmaceutical and medicine manufacturing	0.308	0.326	0.472	0.503	0.406	0.276
Chemical fibers	0.710	0.598	0.777	0.706	0.555	0.462
Rubber products	0.493	0.432	0.678	0.580	0.483	0.397
Plastic products	0.544	0.448	0.706	0.636	0.374	0.299
Non-metallic mineral products	0.371	0.320	0.561	0.501	0.318	0.233
Smelting and pressing of ferrous metals	0.378	0.391	0.527	0.558	0.321	0.357
Smelting and pressing of non-ferrous metals	0.293	0.285	0.465	0.418	0.493	0.494
Metal products	0.524	0.443	0.705	0.612	0.339	0.241
Common machines	0.493	0.592	0.668	0.683	0.350	0.330
Special equipment	0.407	0.359	0.607	0.532	0.300	0.338
Traffic equipment	0.290	0.389	0.541	0.570	0.524	0.418
Manufacture of electrical machinery and apparatus	0.553	0.440	0.718	0.654	0.392	0.371
Electrical machines and equipment	0.720	0.555	0.813	0.727	0.658	0.588
Electronic and communication equipment	0.638	0.588	0.792	0.723	0.634	0.497
Instruments, culture and office devices	0.610	0.406	0.774	0.658	0.559	0.428
Recycling of material waste and scrap	0.348	0.453	0.569	0.645	0.545	0.337
Electricity, steam, thermal power production and supply	0.306	0.267	0.446	0.419	0.250	0.258
Coal gas production and supply	0.404	0.499	0.545	0.633	0.408	0.418
Tap water production and supply	0.391	0.381	0.508	0.521	0.357	0.390
Weighted sample mean	0.460	0.387	0.622	0.563	0.424	0.370
Difference	0.073 (0.015)***		0.060 (0.010)***		0.054 (0.013)***	

The concentration measures for each of the 2-digit CIC industry across provinces are computed by authors based on China National Bureau of Statistics, 1995 and China National Bureau of Statistics, 2004.
*** Statistically significant at 1% level.

shed millions of jobs in the restructuring process during this time period, rendering the employment figures less reliable. The section finishes with a discussion on how these patterns relate to one another. Although mutually consistent to a large degree, there are also some important differences in how well the various measures capture the patterns of clustering in China.

4.1. Increasing industrial concentration

Table 4 compares three concentration measures between 1995 and 2004 at the more aggregate level: the total output share of the top three producing provinces, the Gini coefficient of output across all provinces, and the Krugman–Gini coefficient of output across all provinces, all computed for each of the 2-digit CIC industries. It is clear that between 1995 and 2004, most industries in China have become more concentrated geographically, as all the three measures yield consistent results of increased industrial concentration. The last two rows list the weighted sample means for each of the measures as well as the 1995–2004 differences and their *t*-statistics, showing that the differences are statistically significant.[6]

Table 5 further lists the top three provinces ranked by output for each of the industries. A detailed analysis of the determinants of such location patterns is beyond the scope of this paper, but explanations offered in Wen (2004) for the geographic patterns of industries in 1995 seem to largely apply for the time period of 1995–2004 as well. Briefly, affinity to natural resources, availability of labor and infrastructure, as well as policy initiatives, could all explain the output rankings of Chinese regions in different

[6] The few industries that experienced declines in their geographic concentration mainly include the tobacco industry, the oil and utility industries, and the transportation equipment industry, where the decline in concentration may be explained by continued local protectionism (Bai et al., 2004).

Table 5
Top producing provinces for each 2-digit industry.

Industry	Prov1 _04	Prov2 _04	Prov3 _04	Prov1 _95	Prov2 _95	Prov3 _95
Coal mining and dressing	SX	SD	HEN	SX	SD	HEB
Petroleum and natural gas extraction	HLJ	SD	XJ	HLJ	SD	XJ
Ferrous metals mining and dressing	HEB	LN	SX	HEB	LN	AH
Nonferrous metals mining and dressing	SD	HEN	HUN	SD	JX	HEN
Non-metallic mineral mining and dressing	SD	HEN	ZJ	SD	JS	AH
Other minerals mining and dressing	HLJ	HEN	SD	GX	ZJ	HEN
Foodstuff processing industry	SD	HEN	JS	SD	JS	GD
Foodstuff manufacturing industry	SD	GD	HEN	SH	GD	SD
Beverage manufacturing industry	SD	GD	SC	SD	GD	JS
Tobacco processing	YN	SH	JS	YN	HUN	SH
Spinning industry	JS	ZJ	SD	JS	GD	SH
Manufacturers of clothes and other fiber products	JS	GD	ZJ	SH	JS	GD
Leather, fur, feather and other products	ZJ	GD	FJ	GD	FJ	ZJ
Timber processing and bamboo, cane, palm, straw products	JS	SD	ZJ	JS	GD	HLJ
Furniture	GD	ZJ	SD	GD	SD	JS
Paper makers and paper products	SD	GD	ZJ	SD	GD	ZJ
Printing and record medium reproduction	GD	ZJ	JS	SH	GD	JS
Teaching and sport products for daily use	GD	ZJ	JS	GD	SH	JS
Oil processing and refining	LN	SD	GD	ZJ	LN	SD
Chemical material and products	JS	SD	GD	JS	SH	GD
Pharmaceutical and medicine manufacturing	JS	SD	ZJ	SH	JS	GD
Chemical fibers	ZJ	JS	SD	SH	JS	ZJ
Rubber products	SD	JS	ZJ	SH	SD	JS
Plastic products	GD	ZJ	JS	GD	JS	ZJ
Non-metallic mineral products	SD	HEN	GD	JS	SD	GD
Smelting and pressing of ferrous metals	HEB	JS	LN	SH	LN	HEB
Smelting and pressing of non-ferrous metals	JS	ZJ	HEN	JS	SH	LN
Metal products	GD	JS	ZJ	SH	JS	GD
Common machines	JS	ZJ	SD	SD	JS	SH
Special equipment	SD	JS	GD	JS	SD	SH
Traffic equipment	SH	JL	GD	SH	JS	HUB
Manufacture of electrical machinery and apparatus	GD	JS	ZJ	SH	JS	HEN
Electrical machines and equipment	GD	SH	JS	GD	SH	JS
Electronic and communication equipment	GD	JS	TJ	GD	SH	JS
Instruments, culture and office devices	GD	ZJ	SD	SD	GD	JS
Recycling of material waste and scrap	SAX	JS	SD	GD	SH	JS
Electricity, steam, thermal power production and supply	GD	ZJ	JS	GD	JS	SD
Coal gas production and supply	GD	JS	SC	SH	GD	LN
Tap water production and supply	GD	JS	ZJ	GD	SH	JS

Calculated by authors based on China National Bureau of Statistics, 1995 and China National Bureau of Statistics, 2004. Prov1, Prov2 and Prov3 stand for the top three provinces in terms of output concentration for each 2-digit industry. See Table 2 for the abbreviations of provincial names.

industries. In addition, as the trend of industrial concentration has intensified between 1995 and 2004, the initial conditions listed above seem to have become more important in determining the geographic allocation of industries over time. It is also apparent from the table that several coastal provinces, such as Guangdong, Zhejiang, Jiangsu, and Shandong, have become dominant in a greater number of industries.

Although the tables above provide information that can be easily interpreted, industrial clusters are more appropriately defined at a more disaggregate level, such as county or township level.[7] Thus we now present additional evidence showing the evolution of clustering at the county level. Fig. 1-1 illustrates that patterns similar to those found above are observed when county level data are used to compute the Krugman Gini for each of the 4-digit CIC industries. The density distributions for the Krugman Gini are shown for 1995 and 2004, with the distribution in the later year substantially to the right of that in the earlier year, indicating greater concentration among Chinese counties in 2004 than in 1995 even for 4-digit CIC industries.

4.2. Greater concentration of more equally sized firms

As discussed previously, the industrial concentration measures do not distinguish between the Detroit-style and the Chinese-style industrial districts, since the concentration measure is an aggregate measure concealing most firm-level details. To explore the distinction between these two types, we study additional measures, with the first ones being the number of firms and the average firm size in a certain region, constructed for each industry.

Table 6 lists the per province number of firms and the provincial average firm size for each industry. As shown in the table, for each 2-digit industry, the number of firms in a province has on average increased from 543 to 1550 from 1995 to 2004, a very

[7] Because the data available to us do not have location identifiers at the township level, we can only calculate the cluster measures at the more aggregate provincial and county level but not at the township level.

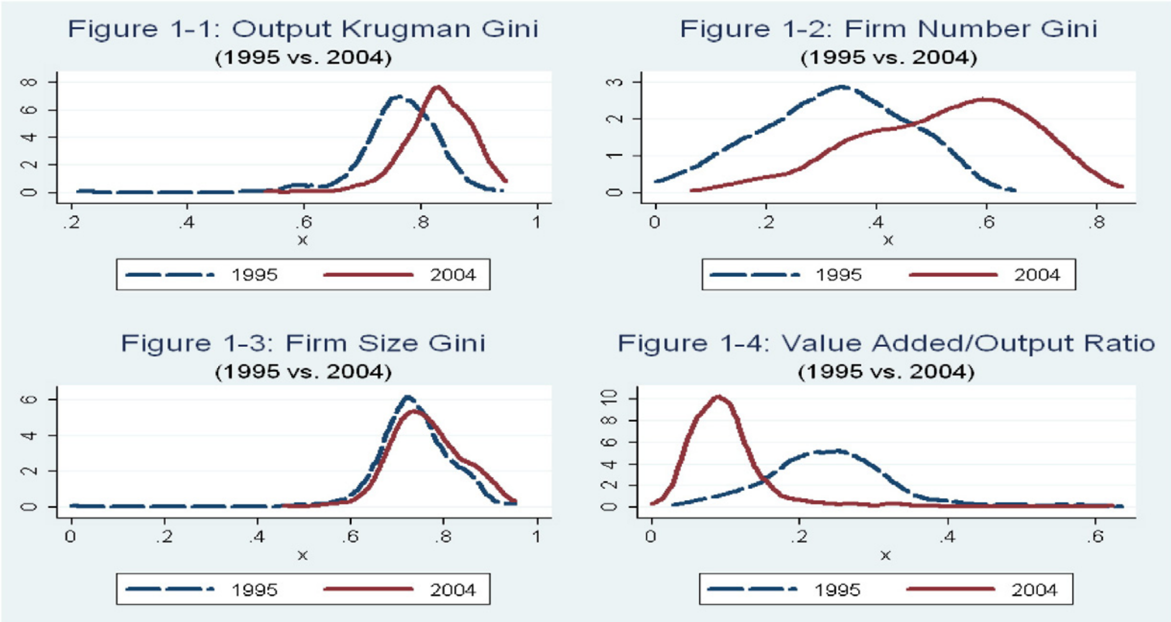

Fig.1. Industrial concentration, regional specialization, interconnectness and clustering. Note: Drawn by authors.

large and significant change; while the average gross output value of firms has increased from RMB 47.3 million to RMB 69.0 million, a change that is not statistically significant. These measures suggest that the higher geographic concentration of industries observed above has been accompanied by a larger number of similar-sized firms.

Table 7 studies the distributions of number of firms and firm size, presenting the Gini coefficients for both the number of firms and the average size of firms. Clearly, the number of firms has become significantly less evenly distributed, while the average size has become slightly more evenly distributed. In other words, while the average firm size has become more equal across regions, the number of firms has become more concentrated across regions, with the top producing regions in each sector having a larger share of the firms in that sector in 2004 than in 1995. The results in Tables 6 and 7 suggest that the increased industrial concentration across regions is mainly driven by the larger number of small or medium-sized firms located in the top production regions.

Again, these patterns are robust to different levels of aggregation. Fig. 1-2 and 1-3 present the density distributions of the Gini-coefficients in 1995 versus 2004 for the average number of firms and the average firm size per county at the 4-digit CIC industry level. While the number of firms has become significantly more concentrated during this period, the average firm size has not shown significant change.

4.3. Increasing interaction among firms

The greater concentration of more equally sized firms that we observed in 2004 suggests a trend of clustering in Chinese industries that differs from the Detroit-style agglomeration. But is there indeed much interaction among firms located in the same region and industry? To answer this question, Table 8 presents the average value-added/output ratio for the 2-digit CIC industries in 1995 and 2004, where the ratio is computed by dividing the total value added by the total gross industrial output value (both at the provincial level) in the same 2-digit industry. As shown in the table, the value-added/output ratio has generally decreased from 1995 to 2004 and the change is statistically significant.

Using 3-digit or 4-digit industries gives similar results. Fig. 1-4 uses 4-digit CIC industries (aggregated from county level data) to compare the density distribution of the value added/output ratio in 1995 with that in 2004. The lower value-added/output ratios in the later year indicate that Chinese firms in the same industry are increasingly interacting with one another through finer division of labor.

4.4. Increasing regional specialization

The findings presented above show that Chinese industries have become more concentrated geographically and several provinces on the coast have appeared as the top producers in multiple industries (see Table 5). Have regions become less or more specialized while industries have become more concentrated?

To answer this question, we use measures of regional specialization. They are similar to measures for industrial concentration, with the only difference being that a region instead of an industry is used as an observational unit. Table 9 compares the following measures between 1995 and 2004: the total output share of the top three CIC-2 producing industries, the Gini coefficient of

Table 6
Number of firms and average firm size per province.

Industry	n_04	n_95	size_04	size_95
Coal mining and dressing	925	386	8360	9664
Petroleum and natural gas extraction	21	6	456,127	1,066,011
Ferrous metals mining and dressing	342	71	4527	5228
Nonferrous metals mining and dressing	203	126	7069	8554
Non-metallic mineral mining and dressing	1165	369	1560	3087
Other minerals mining and dressing	9	7	1871	2515
Foodstuff processing industry	2243	968	6501	10,042
Foodstuff manufacturing industry	962	510	5225	5856
Beverage manufacturing industry	822	460	5089	7852
Tobacco processing	10	14	419,726	237,406
Spinning industry	2678	764	6648	18,002
Manufacturers of clothes and other fiber products	1556	592	4583	7453
Leather, fur, feather and other products	732	327	6547	9308
Timber processing and bamboo, cane, palm, straw products	1194	484	2403	2620
Furniture	771	274	2964	2580
Paper makers and paper products	1280	448	4741	7303
Printing and record medium reproduction	1422	524	2006	2553
Teaching and sport products for daily use	525	173	4597	8575
Oil processing and refining	238	91	60,078	73,925
Chemical material and products	2230	840	9086	13,750
Pharmaceutical and medicine manufacturing	364	189	14,149	16,527
Chemical fibers	116	34	28,017	76,277
Rubber products	506	150	6392	13,294
Plastic products	2249	602	3588	5856
Non-metallic mineral products	5088	1915	2988	4926
Smelting and pressing of ferrous metals	683	263	39,937	44,108
Smelting and pressing of non-ferrous metals	489	149	19,510	29,697
Metal products	2612	836	3719	5534
Common machines	3667	1015	4280	7719
Special equipment	1777	575	5002	10,805
Traffic equipment	1672	610	13,108	17,009
Manufacture of electrical machinery and apparatus	1833	592	9799	13,206
Electrical machines and equipment	507	177	35,440	34,343
Electronic and communication equipment	1173	314	19,198	12,552
Instruments, culture and office devices	859	379	3301	4576
Recycling of material waste and scrap	199	143	2128	2799
Electricity, steam, thermal power production and supply	793	394	28,740	19,369
Coal gas production and supply	48	12	14,362	20,474
Tap water production and supply	356	161	2397	3545
Difference	659 (110)***		21,730 (16,882)	

Calculated by authors based on China National Bureau of Statistics, 1995 and China National Bureau of Statistics, 2004. Firm size is measured as the average output, with 2004 data deflated to 1995 RMB amount.
*** Statistically significant at 1% level.

output across all industries, and the Krugman–Gini coefficient of output across all industries, all computed for each of the Chinese provinces.

These different measures yield largely consistent patterns: Within each region, the top industries account for a significantly larger share of the total regional output in 2004 than in 1995. In other words, Chinese regions have also become more specialized during this time period. Measures at the county level reveal the same trend, as shown in Fig. 2-1, which compares the density distribution of the Krugman Gini coefficient for output in 1995 and that in 2004. Our results based on more updated data thus support the view that Chinese regions have become increasingly specialized since the mid-1990s (Zhang & Tan, 2007). In other words, the product market has become more integrated over time.

4.5. Increasing product proximity within each region

The results presented above provide supporting evidence for the pattern that Chinese regions have become more specialized. In other words, they have increasingly allocated more resources to the top few industries. But how are these industries chosen? Do they tend to be closely related or diversely distributed industries? We use the measure of product proximity to answer this question. Table 10 reports the industry proximity measure (weighted by output) for each of the Chinese provinces in 1995 and 2004. The measures constructed at the county and the prefecture levels give the same pattern of higher industry proximity in each region. Fig. 2-2 supports the pattern using measures computed at the county level, as the density distribution of the proximity measure in 2004 is located substantially to the right of that in 1995.

Table 7
Distributions of number of firms and average firm size.

Industry	n_04	n_95	size_04	size_95
Coal mining and dressing	0.580	0.531	0.668	0.526
Petroleum and natural gas extraction	0.533	0.680	0.524	0.643
Ferrous metals mining and dressing	0.660	0.541	0.524	0.776
Nonferrous metals mining and dressing	0.541	0.475	0.718	0.418
Non-metallic mineral mining and dressing	0.426	0.404	0.305	0.381
Other minerals mining and dressing	0.512	0.435	0.648	0.414
Foodstuff processing industry	0.512	0.433	0.316	0.364
Foodstuff manufacturing industry	0.444	0.321	0.341	0.408
Beverage manufacturing industry	0.432	0.427	0.308	0.419
Tobacco processing	0.411	0.432	0.548	0.461
Spinning industry	0.728	0.526	0.318	0.233
Manufacturers of clothes and other fiber products	0.700	0.507	0.358	0.430
Leather, fur, feather and other products	0.732	0.446	0.350	0.412
Timber processing and bamboo, cane, palm, straw products	0.569	0.492	0.281	0.402
Furniture	0.551	0.364	0.395	0.410
Paper makers and paper products	0.566	0.373	0.311	0.253
Printing and record medium reproduction	0.514	0.341	0.261	0.316
Teaching and sport products for daily use	0.738	0.515	0.459	0.373
Oil processing and refining	0.507	0.486	0.455	0.528
Chemical material and products	0.524	0.394	0.230	0.264
Pharmaceutical and medicine manufacturing	0.391	0.366	0.194	0.255
Chemical fibers	0.740	0.556	0.391	0.405
Rubber products	0.610	0.428	0.476	0.423
Plastic products	0.650	0.468	0.256	0.302
Non-metallic mineral products	0.477	0.405	0.317	0.308
Smelting and pressing of ferrous metals	0.474	0.403	0.321	0.508
Smelting and pressing of non-ferrous metals	0.552	0.395	0.415	0.365
Metal products	0.635	0.401	0.286	0.347
Common machines	0.643	0.446	0.233	0.285
Special equipment	0.630	0.439	0.271	0.323
Traffic equipment	0.548	0.385	0.437	0.380
Manufacture of electrical machinery and apparatus	0.673	0.469	0.275	0.349
Electrical machines and equipment	0.711	0.579	0.493	0.457
Electronic and communication equipment	0.698	0.562	0.420	0.295
Instruments, culture and office devices	0.701	0.461	0.345	0.418
Recycling of material waste and scrap	0.517	0.378	0.467	0.432
Electricity, steam, thermal power production and supply	0.583	0.521	0.619	0.595
Coal gas production and supply	0.472	0.447	0.495	0.751
Tap water production and supply	0.531	0.485	0.508	0.585
Weighted sample mean	0.586	0.461	0.375	0.391
Difference	0.125 (0.014)***		−0.016 (0.020)	

The distribution is measured as Gini coefficient in number of firms (n_95 and n_04) and firm size (size_95 and size_04) by industry in 1995 and 2004. Calculated by authors based on China National Bureau of Statistics, 1995 and China National Bureau of Statistics, 2004.
*** Statistically significant at 1% level.

4.6. Comparing various measures of clustering

As we have used multiple measures to examine the patterns of industrialization in China between 1995 and 2004, we now compare the various measures to see how they relate to one another in portraying the industrialization patterns in China.

Tables 11 and 12 present the correlations among different measures in industry concentration and regional specialization. At the industry level, Table 11 shows that the various concentration measures are highly correlated. Furthermore, the more geographically concentrated their gross outputs are, the finer the division of labor among firms in 2004, as the correlation coefficient between the value added/output ratio is negative with the concentration measures whenever the correlation is significant. But such a pattern is missing in 1995, where the correlation coefficient between the value added/output ratio and the concentration measure was never significant. This suggests that between 1995 and 2004 industrial agglomeration in China has exhibited more interactions, whereby industries that are more geographically concentrated also enjoy finer vertical division of labor among firms. Table 12 shows that for a particular region, the more specialized it is in choosing the industrial allocation, the greater proximity there is among industries in 2004. Yet interestingly, the correlation between the proximity measure and the other measures is insignificant or significant in the wrong direction in 1995. In other words, regions with a higher degree of specialization are also more specialized in a range of highly interconnected industries in 2004, but more specialized regions were equally or even less interconnected than other regions in 1995. Thus Chinese regions not only have become more specialized over time, but they have also better at specializing. By focusing on a range of highly interconnected activities, they are more capable to take advantage of the positive spillovers from clustering in later years (Marshall, 1920; Long & Zhang, 2011).

Table 8
Value added/Output ratio (2004 v. 1995).

Industry	2004	1995
Coal mining and dressing	0.353	0.508
Petroleum and natural gas extraction	0.617	0.555
Ferrous metals mining and dressing	0.336	0.394
Nonferrous metals mining and dressing	0.292	0.348
Non-metallic mineral mining and dressing	0.202	0.385
Other minerals mining and dressing	0.053	0.400
Foodstuff processing industry	0.207	0.165
Foodstuff manufacturing industry	0.260	0.216
Beverage manufacturing industry	0.347	0.312
Tobacco processing	0.683	0.533
Spinning industry	0.238	0.184
Manufacturers of clothes and other fiber products	0.205	0.237
Leather, fur, feather and other products	0.192	0.198
Timber processing and bamboo, cane, palm, straw products	0.176	0.257
Furniture	0.160	0.264
Paper makers and paper products	0.217	0.234
Printing and record medium reproduction	0.229	0.310
Teaching and sport products for daily use	0.160	0.185
Oil processing and refining	0.194	0.257
Chemical material and products	0.276	0.260
Pharmaceutical and medicine manufacturing	0.373	0.293
Chemical fibers	0.189	0.189
Rubber products	0.205	0.221
Plastic products	0.185	0.189
Non-metallic mineral products	0.242	0.318
Smelting and pressing of ferrous metals	0.264	0.266
Smelting and pressing of non-ferrous metals	0.218	0.221
Metal products	0.194	0.249
Common machines	0.241	0.292
Special equipment	0.226	0.216
Traffic equipment	0.230	0.252
Manufacture of electrical machinery and apparatus	0.233	0.243
Electrical machines and equipment	0.237	0.228
Electronic and communication equipment	0.281	0.277
Instruments, culture and office devices	0.200	0.276
Recycling of material waste and scrap	0.164	0.219
Electricity, steam, thermal power production and supply	0.356	0.530
Coal gas production and supply	0.308	0.081
Tap water production and supply	0.426	0.478
Weighted sample mean	0.262	0.282
Difference	−0.021 (0.013)*	

Calculated by authors based on China National Bureau of Statistics, 1995 and China National Bureau of Statistics, 2004.
* Statistically significant at 10% level.

These results support the argument that China's “Reform and Opening-up” policies since the 1980s have helped the industrial clustering process. The lack of correlation between regional specialization and industry proximity in 1995 indicates the existence of local protectionism, as the planned economy legacy implies that industries had developed in many regions not based on their comparative advantages (Young, 2000). Largely in response to the then widespread market fragmentation, China has implemented a series of policy reforms since the mid-1990s to integrate the product market (Zhang & Tan, 2007). Reassuringly, our industry proximity measure became significantly correlated with the CR3 and Gini measures by 2004. In other words, regions that are more specialized in their industry allocations are also more likely to have closer linkages among upstream and downstream industries, a key feature of industrial clusters.

5. Explaining greater proximity in China

A more detailed study is required to fully understand the mechanisms through which Chinese regions and industries have become more clustered between 1995 and 2004, but we test a few conjectures here. First of all, different industries have different tendencies for forming agglomerations as they may benefit differently from locating in clusters. Following Marshall (1920), clustering should be preferred by industries for which access to market and suppliers is important, for which labor skills are important, and for which benefits from technological spillovers are more important.

In transition economies such as China, some additional factors may also play important roles in explaining agglomeration. As financial access is often limited, Chinese industries that rely more on external finances may have a greater need for clustering as closer interactions may facilitate inter-firm financing practices such as trade credit. Similarly, clustering will be more desired by firms in regions with lower levels of financial development. In addition, the ownership structure of Chinese firms may determine

Table 9
Regional specialization.

Province	CR3_04	CR3_95	gini_04	gini_95	Kgini_04	Kgini_95
Beijing	0.357	0.351	0.664	0.627	0.483	0.414
Tianjin	0.402	0.304	0.656	0.560	0.477	0.366
Hebei	0.400	0.307	0.625	0.565	0.402	0.359
Shanxi	0.590	0.485	0.789	0.713	0.689	0.548
Neimeng	0.394	0.337	0.679	0.614	0.570	0.542
Liaoning	0.386	0.318	0.640	0.599	0.325	0.520
Jilin	0.571	0.437	0.757	0.647	0.534	0.519
Heilongjiang	0.499	0.432	0.719	0.639	0.699	0.578
Shanghai	0.322	0.320	0.589	0.553	0.289	0.315
Jiangsu	0.293	0.309	0.633	0.564	0.310	0.280
Zhejiang	0.280	0.284	0.552	0.551	0.403	0.312
Anhui	0.259	0.269	0.559	0.529	0.284	0.322
Fujian	0.244	0.205	0.600	0.489	0.357	0.555
Jiangxi	0.286	0.231	0.562	0.525	0.407	0.534
Shandong	0.239	0.235	0.531	0.506	0.274	0.276
Henan	0.294	0.231	0.551	0.541	0.361	0.335
Hubei	0.428	0.334	0.666	0.597	0.400	0.279
Hunan	0.255	0.238	0.550	0.544	0.378	0.329
Gongdong	0.385	0.234	0.676	0.537	0.381	0.367
Gongxi	0.383	0.318	0.661	0.573	0.431	0.496
Hainan	0.443	0.371	0.670	0.626	0.577	0.608
Chongqing	0.451	0.406	0.683	0.665	0.467	0.476
Sichuan	0.279	0.283	0.560	0.583	0.342	0.309
Guizhou	0.426	0.308	0.698	0.644	0.601	0.503
Yunnan	0.478	0.510	0.728	0.708	0.709	0.662
Tiebet	0.614	0.453	0.734	0.635	0.697	0.740
Shaanxi	0.308	0.252	0.584	0.563	0.551	0.383
Gansu	0.490	0.374	0.733	0.660	0.587	0.500
Qinghai	0.612	0.497	0.802	0.711	0.705	0.614
Ningxia	0.389	0.331	0.716	0.666	0.591	0.599
Xinjiang	0.610	0.549	0.782	0.742	0.690	0.588
Weighted sample average	0.340	0.293	0.620	0.563	0.386	0.368
Difference	0.047 (0.019)**		0.057 (0.014)***		0.018 (0.015)	

CR3, Gini and Kgini measure the total output share of the top three CIC-2 producing industries, the Gini coefficient of output across all industries, and the Krugman-Gini coefficient of output across all industries by province. Calculated by authors based on China National Bureau of Statistics, 1995 and China National Bureau of Statistics, 2004.
*** Statistically significant at 1% level.
** Statistically significant at 5% level.

the development of industrial clusters for two reasons: firms of different ownership types may have different demands for clustering due to their different abilities to benefit from agglomeration, and the access to financing may also vary substantially depending on the firm's ownership type. Finally, various governments in China often have a large presence in determining the development of industrial clusters.

As shown in Long and Zhang (2011), the newly constructed proximity measure best fits the pattern of industrial clustering observed in China. We thus focus on how various factors determine the different degree of proximity across Chinese industries and regions. Various factors may determine the degree of proximity. At the industry level, initial conditions such as reliance on external finances, reliance on contracting, capital intensity, human intensity, as well as export intensity may all affect the potential benefits from locating within an industrial cluster and thus the incentive for forming one. We rely on Gao (2007) for data on Chinese industries' export intensity and Ciccone and Papaioannou (2009) for information on the other industry characteristics. Note that all these variables were evaluated before the early 1990s; thus they are appropriate to use as the initial values in 1995. Along the regional dimension, the level of local financial development, ownership structure of local industries, and whether the region is along the coast or in the inland also potentially influence the level of proximity. To obtain the regional level information, we use the 1995 Census data, aggregating firm level data to obtain regional averages. For ownership structure, we compute the average percentage of privately owned firm shares and that of foreign-owned shares in the county (both weighted by sales). To measure the degree of financial constraint faced by firms located in a certain region, we computed the coefficient of variation for the marginal product of capital for the region (as measured by the log of the valued added/capital ratio); thus a higher value for this variable indicates a lower level of local financial development.[8]

To explore the effects of both sectoral and regional features, we conduct two sets of estimations using proximity measures computed at the 4-digit CIC industry-county level. In the first set of estimations, we explain proximity by various industry

[8] The variable is also used in Long and Zhang (2011).

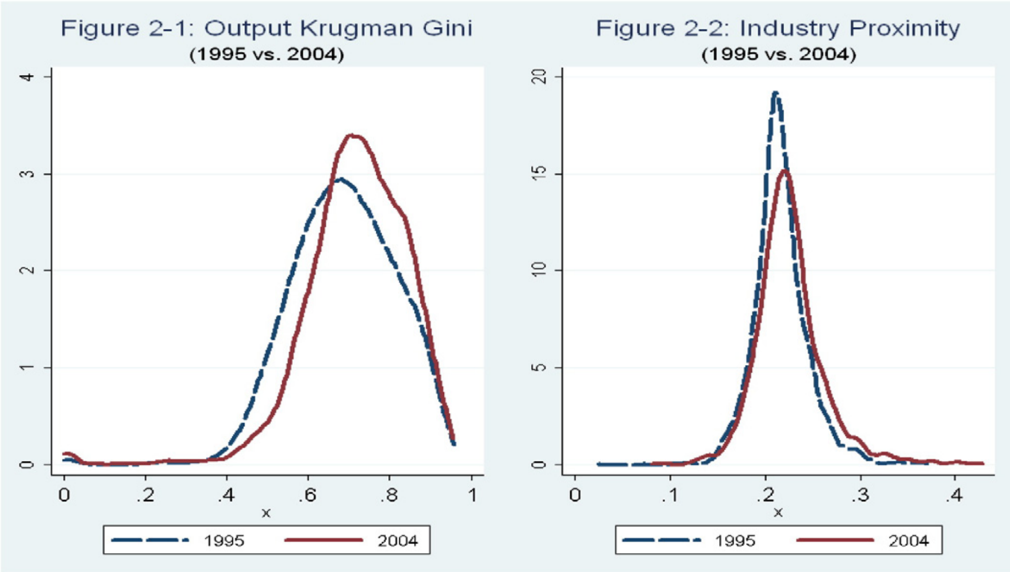

Fig. 2. Industrial concentration (1995 v. 2004). Note: Fig. 1-1 presents the distribution of the Krugman Gini Coefficients as the concentration index in 1995 versus 2004 for each 4-digit CIC industry; Fig. 1-2 presents the distribution of Gini coefficients in 1995 versus 2004 in firm number per county for each 4-digit CIC industry; Fig. 1-3 presents the distribution of Gini coefficients in 1995 versus 2004 in average firm size per county for each 4-digit CIC industry; while Fig. 1-4 presents the distribution of the average value added/output ratio in 1995 versus 2004 for each 4-digit CIC industry. All the data used in graphing the figures is computed by authors based on China National Bureau of Statistics, 1995 and China National Bureau of Statistics, 2004.

features. To further allow the effects to vary across different ownership types, we also include the interaction terms between the industry's external finance reliance and the region's ownership measures, as well as the interactions between the industry's export intensity and the region's ownership structure. To abstract from the effects of county level characteristics, we control for county fixed effects in the estimations, with standard errors clustered at the 4-digit CIC level.

Columns 1–3 in Table 12 indicate that the proximity measure in 2004 is significantly correlated with its value in 1995, no matter whether the weight used to construct the proximity measure is sales, employment, or assets. More interestingly, sectors with higher capital intensity tend to show higher degrees of proximity, regardless of which weight is used. These results are consistent with the agglomeration effects based on knowledge and technology. As spillover effects of technology (as captured by capital) are important sources of benefit for firms within a conglomeration, there is a greater incentive for technology-intensive industries to form clusters.

But the patterns observed above are also consistent with a less sanguine interpretation. Local governments in China have been playing prominent roles in promoting industrial clusters and they are often drawn to industries with higher capital intensity because these industries can quickly bring about higher local GDP growth. The findings from Long and Zhang (2011), however, show that compared to other clusters, smaller benefits are obtained from those involving industries with higher capital intensity. The positive correlation between industry capital intensity and the degree of proximity observed above can then be interpreted as evidence that the government's intervention might have led to the over-clustering of capital intensive industries in China.

There is also some evidence that industries with higher reliance on contracting are more likely to have higher degrees of proximity (see Column 2). This suggests that industries that have the need for greater interaction with their downstream and upstream industries have somewhat succeeded to locate more closely to their related industries. Although neither the industry's reliance on external finance nor its export intensity has a significant effect on the degree of proximity, the estimates for the interaction terms tell a more subtle story. Industries that rely more heavily on external finances tend to have significantly higher levels of proximity when located in regions with a substantial presence of either private ownership or foreign ownership. These results are consistent with the argument that a main benefit from clustering in China is the access to finances, especially for private firms and foreign firms, which are the outsiders of a financial system dominated by state banks that prefer other SOEs.

Similar results are obtained regarding export intensity, with industries that have comparative advantage enjoying higher degrees of proximity when they are located in regions with a substantial presence of private ownership. This is consistent with Marshall's arguments that enhanced access to market and knowledge spillovers are both important sources of benefit for firms within industrial clusters. As information regarding overseas markets is more easily obtained within a cluster, more firms are drawn to agglomeration in industries that have comparative advantages in exporting. But note that these effects are only observed when the presence of private ownership or foreign ownership is substantial. Marshall's predictions do not apply when the regions are dominated by Chinese SOEs, as their operations are still not fully profit driven.

Columns 4–6 in Table 13 present results from the second set of estimations, where we explain proximity by regional characteristics. Similar to the first set of estimations above, we also include interaction terms between ownership structure and

Table 10
Product proximity.

Province	2004	1995
Beijing	0.220	0.206
Tianjin	0.208	0.194
Hebei	0.219	0.212
Shanxi	0.208	0.207
Neimeng	0.214	0.198
Liaoning	0.205	0.204
Jilin	0.220	0.206
Heilongjiang	0.197	0.186
Shanghai	0.219	0.222
Jiangsu	0.210	0.210
Zhejiang	0.220	0.211
Anhui	0.211	0.204
Fujian	0.202	0.208
Jiangxi	0.206	0.200
Shandong	0.205	0.200
Henan	0.209	0.201
Hubei	0.216	0.207
Hunan	0.210	0.201
Gongdong	0.215	0.209
Gongxi	0.214	0.208
Hainan	0.207	0.201
Chongqing	0.197	0.206
Sichuan	0.202	0.198
Guizhou	0.196	0.188
Yunnan	0.197	0.187
Tibet	0.238	0.223
Shaanxi	0.192	0.191
Gansu	0.205	0.199
Qinghai	0.217	0.197
Ningxia	0.220	0.215
Xinjiang	0.199	0.190
Weighted sample average	0.211	0.206
Difference	0.005 (0.001)***	

The industry proximity measure (weighted by output) for each of the Chinese provinces in 1995 and 2004 are calculated by authors based on China National Bureau of Statistics, 1995 and China National Bureau of Statistics, 2004.
*** Statistically significant at 1% level.

Table 11
Correlation matrix for industry concentration and firm interaction measures.

	CR3	Gini_output	K-Gini_output	Firm number	Firm size	Value added/output
2004						
Gini_output	0.3896	1				
	(0.000)					
K-Gini_output	0.4179	0.5204	1			
	(0.000)	(0.000)				
Firm number	−0.3783	0.2707	−0.1028	1		
	(0.000)	(0.000)	(0.027)			
Firm size	0.3996	0.6104	0.5205	−0.2815	1	
	(0.000)	(0.000)	(0.001)	(0.000)		
Value added/output	0.0454	−0.0848	−0.0965	−0.0718	0.0127	1
	(0.325)	(0.069)	(0.038)	(0.124)	(0.785)	
1995						
Gini_output	0.1595	1				
	(0.000)					
K-Gini_output	0.1891	0.5524	1			
	(0.000)	(0.000)				
Firm number	−0.4953	0.3192	0.0192	1		
	(0.000)	(0.000)	(0.676)			
Firm size	0.3008	0.8553	0.5586	0.0318	1	
	(0.108)	(0.639)	(0.000)	(0.489)		
Value added/output	−0.0043	0.0405	0.0296	0.0678	0.0271	1
	(0.925)	(0.378)	(0.521)	(0.140)	(0.555)	

Calculated using county level data by authors based on China National Bureau of Statistics, 1995 and China National Bureau of Statistics, 2004. The figures in parentheses are p-values of the correlation coefficients among the concentration and firm interaction measures.

Table 12
Correlation matrix for regional specialization and product proximity measures.

	2004			1995		
	CR3	Gini	K-Gini	CR3	Gini	K-Gini
Gini	0.5568	1		0.6900	1	
	(0.000)			(0.000)		
K-Gini	0.4495	0.8019	1	0.5182	0.7218	1
	(0.000)	(0.000)		(0.000)	(0.000)	
Proximity	0.1147	0.0232	0.0025	−0.0108	−0.0299	−0.1320
	(0.001)	(0.224)	(0.894)	(0.571)	(0.117)	(0.000)

Calculated by authors based on China National Bureau of Statistics, 1995 and China National Bureau of Statistics, 2004, using county level data. The figures in parentheses are *p*-values of the correlation coefficients among regional specialization and product proximity measures.

industry's reliance on external finance as well as ownership's interactions with industry's export intensity. There is evidence in two out of three columns that greater proximity may have occurred due to greater difficulty in accessing finances (see Columns 4 and 5), as regions with lower levels of financial development tend to have higher degrees of proximity. Finally, coastal regions tend to see more industrial clustering (see Columns 4 and 5)—this is consistent with other empirical findings (see, for instance, Li & Fung, 2006). Finally, coefficient estimates for the interaction terms tell stories very similar to those found in Columns 1–3. A higher degree of proximity is more likely to emerge when private ownership (or foreign ownership) is combined with a greater reliance on external finance (or export comparative advantage).

Table 13
Explaining proximity using industry and regional characteristics.

	(1)	(2)	(3)	(4)	(5)	(6)
Proximity(output)	0.865***			0.736***		
	(0.0442)			(0.00946)		
Proximity(employment)		0.729***			0.570***	
		(0.0578)			(0.00900)	
Proximity(asset)			0.867***			0.737***
			(0.0491)			(0.00916)
Comparative advantage	0.00186	0.00551	0.00346			
	(0.00423)	(0.00622)	(0.00458)			
Reliance on external finance	−0.00509	−0.0107	−0.00391			
	(0.00384)	(0.00684)	(0.00418)			
Reliance on contracting	0.00731	0.0254*	0.00682			
	(0.00597)	(0.0125)	(0.00636)			
Capital intensity	0.00695**	0.0125***	0.00740**			
	(0.00246)	(0.00411)	(0.00263)			
Human capital intensity	0.00173	−0.000378	0.00231			
	(0.00204)	(0.00281)	(0.00223)			
Private share				−0.0129	0.00155	−0.0221
				(0.0120)	(0.0145)	(0.0140)
Foreign share				−0.00257	0.00918**	0.000655
				(0.00431)	(0.00409)	(0.00454)
Financial constraint				0.000920*	0.00159***	0.000208
				(0.000550)	(0.000567)	(0.000638)
Coast dummy				0.00107*	0.00302***	0.000379
				(0.000643)	(0.000639)	(0.000645)
Comparative advantage	0.0401***	0.0548***	0.0367**	0.0321**	0.0369**	0.0273*
* private share	(0.0129)	(0.0114)	(0.0152)	(0.0162)	(0.0158)	(0.0146)
Comparative advantage	0.0229***	0.0281*	0.0140**	0.0168***	0.0197***	0.00939**
* foreign share	(0.00671)	(0.0141)	(0.00653)	(0.00435)	(0.00447)	(0.00444)
Ext. finance reliance	0.0703***	0.0445**	0.0745**	0.0559**	0.0102	0.0579**
* private share	(0.0216)	(0.0183)	(0.0346)	(0.0242)	(0.0262)	(0.0243)
Ext. finance reliance	0.0589***	0.0781***	0.0518***	0.0479***	0.0481***	0.0383***
* foreign share	(0.00940)	(0.0224)	(0.0122)	(0.00957)	(0.00759)	(0.00933)
Constant	−0.00501	0.0307	−0.0121	0.0570***	0.0910***	0.0580***
	(0.0253)	(0.0356)	(0.0271)	(0.00205)	(0.00201)	(0.00206)
Observations	33,941	33,893	33,945	33,784	33,724	33,792
Adjusted R-squared	0.746	0.699	0.757	0.657	0.631	0.673

Dependent variables are the proximity measures at the industry-county level, computed by the authors using different weights (output in Columns 1 and 4, employment in Columns 2 and 5, and asset in Columns 3 and 6). Chinese industries' comparative advantage data is from Gao (2007), while other industry level data is from Ciccone and Papaioannou (2009). County level information is computed by the authors based on the 1995 census data.
*** Statistically significant at 1% level.
** Statistically significant at 5% level.
* Statistically significant at 10% level.

To summarize, we have found evidence that both regional features and industry features play significant roles in explaining the increase in proximity during 1995 and 2004 in China. While the regional features mainly work through the channel of financial development and ownership structure, the effects of industry characteristics illustrate the importance of intrinsic technological requirements in influencing how clusters form. Finally, the interaction effects of ownership structure and technological features highlight how technological factors and institutional arrangements can impact each other's influence on economic outcomes.

6. Conclusion

By computing various measures of clustering using firm-level census data for 1995 and 2004, we have shown in this paper that China's industrialization has been accompanied by greater spatial concentration, increasing regional specialization, and more interactions among firms within industries and within regions. The increased spatial concentration resembles the industrialization path in other countries. But our results also indicate that the number of firms is growing faster and firm size is not significantly larger in clustered areas than in non-clustered regions; at the same time there is finer division of labor and closer technological affinity among firms. This pattern mirrors the East Asian cluster-based industrialization led by SMEs but differs from the U.S. experience, where industrial districts were dominated by large firms.

These patterns put together suggest that China's industrialization process has been accompanied by industrial clusters as defined by Porter, where a large number of highly inter-connected firms are located within a well-defined geographic region. This cluster-based industrialization may have fit well with China's comparative advantage at the onset of its reform that was marked with limited capital and abundant labor, as the cluster-based business model makes use of more entrepreneurs and labor and less capital compared to the non-clustered large factories. This may explain why clusters have emerged as the organizational choice of Chinese firms over time. A recent paper by Long and Zhang (2011) has provided empirical evidence that clustering has indeed helped Chinese firms overcome financial constraints and further improve their productivity and export performance.

Although the patterns are clear and robust, many important issues remain to be explored. For example, is there a natural tendency toward clustering in the process of industrialization? Why is the scale of clustering in China much higher than many parts of the world? Is it because of China's large domestic market or the more active intervention of local governments? What factors tend to promote the development of industrial clusters? Is the cluster-based industrialization model observed in China applicable to other developing countries, such as Sub-Saharan African countries?

Furthermore, what are the mechanisms through which clustering affects firm growth and economic development? Two crucial problems facing Chinese firms are credit constraints and weak protection of property and contracting rights. Does clustering play any role in helping to resolve these difficulties? Some characteristics of clustering suggest a potentially important role of industrial clusters in promoting economic growth. Close proximity and intense competition among firms within a cluster may reduce the temptation of cheating, and finer division of labor and frequent usage of trade credit among firms within a cluster may reduce the need for external finances.

As firms cluster in a narrowly defined geographic region, how do they cope with natural and market shocks at the regional level? Given that the cluster-based model is labor intensive, how will the clusters respond to the rising labor cost in China (Wang & Mei, 2009)? Finally, innovation is the engine for growth in the long run. But what role does clustering play in upgrading the local economy's structure and bringing about technological, managerial, and institutional innovations? In other words, will clustering become a long-term feature of China's economic growth or will it serve as a temporary arrangement waiting to be replaced by some other more efficient system?

The empirical findings in this paper are just the first step in identifying the regularities of China's rapid industrialization. A better understanding of these issues will help shed light on whether the lessons drawn from China's experience in industrialization can be applied to other developing regions in the world.

Appendix. More on the Hausmann–Klinger proximity matrix

Specifically, Hausmann and Klinger propose the proximity between two products i and j to be computed as follows: $P_{i,j} = \min\{P(x_i|x_j), P(x_j|x_i)\}$, where $x_i = 1$ if a country has the revealed comparative advantage in product i (or if $RCA_i > 1$), and 0 otherwise, while the conditional probabilities $P(x_i|x_j)$, $P(x_j|x_i)$ are computed using trade information on all countries.

To get to the intuition of the formula, consider the pair of goods of ostrich meat (good i) and metal ores (good j). Some countries such as Australia export both goods. The formula implies that the probability of exporting metal ores given that a country exports ostrich meat is large, but the probability that a country exports ostrich meat given that it exports metal ores is very low, since although Australia exports both, Chile, Peru, and Zambia do not export ostrich meat but do export metals. The proximity between ostrich meat and metal ores will thus be low, because the formula requires the minimum of the two conditional probabilities, with $P(x_i|x_j)$ being low, despite a high $P(x_j|x_i)$. Thus, the formula is superior to a simple conditional probability $P(x_i|x_j)$or$P(x_j|x_i)$.

The proximity measure also isolates the degree of similarity between the two goods from how prevalent they are in different countries. An alternative measure for proximity is the joint probability $P(x_i \cap x_j)$, which Hausmann and Klinger rejected for the following reason. Consider ostrich meat and ostrich eggs, two goods with extremely high similarity, because every single country that exports ostrich eggs also exports ostrich meat. But if only three countries in the world export these two goods, "then the joint probability for any single country exporting the two would be small, instead of large." (Hausmann & Klinger, 2007, page 10) The problem with using the joint probability to measure proximity is that it combines the degree of similarity between the two goods and their prevalence in different parts of the world.

Some additional nice features of the proximity measure as an indicator of industry interconnectedness include the following: (1) as a characteristic of the production technology based on export/import information from all countries, it applies to all countries, be they open or closed; (2) computed as the minimum between two conditional probabilities, it is a symmetric measure; (3) by focusing on countries with a revealed comparative advantage in product *i* (i.e., $x_i=1$ if a country has the revealed comparative advantage in product *i* [or if $RCA_i>1$] and 0 otherwise), the measure captures all the significant exports but leaves aside the noise; and, (4) finally, by computing the average proximities between 1998 and 2000, the Hausmann–Klinger proximity matrix integrates some stability over time.

The proximity measure can be based on assets, employment, or output, depending on the weights used to adjust for the size of each industry. We use all these measures, as they may provide different angles of clustering. An illustration follows. Consider a region with three industries: steel, automobiles, and rubber. Intuitively, the automobile industry has a high proximity to both the steel and rubber industries, while the proximity between the other two is low. Now suppose that the region has experienced faster growth in the auto industry than in the other industries. Following the procedures describe above, we see that the average proximity of the auto industry has not changed, since the relative weights of the other two industries have not changed. But the average industry proximity of the whole region has increased, because the industry that is closer to the others, the auto industry in this case, has grown faster. Now consider the role of the weight. If the growth of the auto industry is in its output relative to those of the other industries, then the greater interconnectedness among industries in the region will be reflected in a greater proximity using output as the weight. Proximity measures weighted by asset or employment can be understood accordingly.

These three proximities may therefore measure different kinds of interconnectedness, which in turn imply different kinds of cluster effects. Marshall (1920) outlined three types of advantages from agglomeration (which includes clusters as a special case): labor market pooling, specialized supplies, and technological spillovers. Large populations of skilled laborers enter the area and are able to exchange knowledge, ideas, and information. In addition, there is increased access to the specialized goods and services provided for the clustering firms, which provides increasing returns to scale for each of the firms located within that area because of the proximity to the available sources needed for production. Finally, clustering in specific fields leads to quicker diffusion and adoption of ideas.

Although likely to contribute to all three of these advantages, output-weighted proximity is probably more conducive to technological spillovers, since the output can be used as input in the production of other industries in the same region, while employment-weighted proximity implies more labor-market pooling, and asset-weighted proximity implies more specialized supplies, especially in capital goods. All of these effects of agglomeration will lead to higher productivity at the firm level, thus revealing the positive spillover effects of clustering.

References

Bai, Chong-En, Duan, Yingjuan, Tao, Zhigang, & Tong, Sarah T. (2004). Local protectionism and regional specialization: Evidence from China's industries. *Journal of International Economics*, *63*(2), 397–417.

Becattini, Giàcomo (1990). The marshallian industrial district as socio-economic notion. In F. Pyke, G. Becattini, & W. Sengenberger (Eds.), *Industrial Districts and Inter-firm Cooperation in Italy*.

Bellandi, Marco, & Di Tommaso, Marco R. (2005). The case of specialized towns in Guangdong, China. *European Planning Studies*, *13*(5), 707–729.

Chandler, Alfred, Jr. (1977). *The visible hand: The managerial revolution in American business.* Cambridge: Harvard University Press.

China National Bureau of Statistics, various years. *China Statistical Yearbooks*. Beijing: China Statistical Publishing House.

China National Bureau of Statistics (1995). *China industrial census.*

China National Bureau of Statistics (2004). *China economic census.*

Ciccone, Antonio, & Papaioannou, Elias (2009). Human capital, the structure of production, and growth. *Review of Economics and Statistics*, *91*(1), 66–82.

Enright, Michael J., Scott, Edith E., & Chang, Ka-mun (2005). *Regional powerhouse: The greater Pearl River Delta and the rise of China*: John Wiley and Sons (Asia) Inc.

Fleisher, Belton, Hu, Dinghuan, McGuire, William, & Zhang, Xiaobo (2010). The evolution of an industrial cluster in China. (Belton Fleisher, Dinghuan Hu, William McGuire, and Xiaobo Zhang). *China Economic Review*, *21*(3), 456–469.

Fujita, Masahisa, Krugman, Paul, & Venables, Anthony J. (2001). *The spatial economy: Cities, regions, and international trade.* Massachusetts: MIT Press.

Gao, Yuan (2007). Dynamic comparative advantage: A comparison of China and India. University of Montreal Department of Economics Working Paper.

Hausmann, Ricardo, & Klinger, Bailey (2007). *The structure of the product space and the evolution of comparative advantage.* Center for International Development Working Paper No. 146, Harvard University.

Hounshell, David A. (1984). *From the American system to mass production, 1800–1932.* Baltimore: The Johns Hopkins University Press.

Huang, Jikun, Rozelle, Scott, & Chang, Ming (2004). Tracking distortions in agriculture: China and its accession to the World Trade Organization. *World Bank Economic Review*, *18*(1), 59–84.

Huang, Zuhui, Zhang, Xiaobo, & Zhu, Yunwei (2008). The role of clustering in rural industrialization: A case study of Wenzhou's footwear industry. *China Economic Review*, *19*, 409–420.

Krugman, Paul (1991). *Geography and trade.* Cambridge, MA: MIT Press.

Landes, David S. (1998). *The wealth and poverty of nations.* New York: W.W. Norton and Co.

Li and Fung Research Center. 2006, Industrial Cluster Series, May Issue 1.

Long, Cheryl, & Zhang, Xiaobo (2011). Cluster-based industrialization in China: Financing and performance. *Journal of International Economics*, *84*(1), 112–123.

Lu, Jianyong, & Tao, Zhigang (2009). Trends and determinants of China's industrial agglomeration. *Journal of Urban Economics*, *65*, 167–180.

Markusen, Ann (1996). Sticky places in slippery space: A typology of industrial districts. *Economic Geography*, *72*(3), 293–313.

Marshall, Alfred (1920). *Principles of Economics.* London: Macmillan and Co., Ltd.

Marukawa, Tomoo (2006). The geography and history of industrial clusters in Zhejiang Province, China. http://www.iss.u-tokyo.ac.jp/~marukawa/zhejiangcluster.pdf

Nakabayashi, Masaki (2006). Flexibility and diversity: The putting-out system in the silk fabric industry of Kiryu, Japan. *Discussion Paper in Economic and Business No. 06–10, Osaka University*.

Piore, Michael J., & Sabel, Charles F. (1984). *The second industrial divide: Possibilities for prosperity*: Basic Books.

Porter, E. Michael (1998). Clusters and the new economics of competition. *Harvard Business Review*, *76*(6), 77–90.

Porter, E. Michael (2000). Location, competition, and economic development: Local clusters in a global economy. *Economic Development Quarterly*, *14*(1), 15–34.

Ruan, Jianqing, & Zhang, Xiaobo (2009). Finance and cluster-based industrial development in China. *Economic Devleopment and Cultural Change*, *58*, 143–164.

Schmitz, H. (1995). Collective efficiency: Growth path for small-scale industry. *Journal of Development Studies*, *31*(4), 529–566.
Sonobe, Tetsushi, Hu, Dinghua, & Otsuka, Keijiro (2002). Process of cluster formation in China: A case study of a garment town. *Journal of Development Studies*, *39*(1), 118–139.
Sonobe, Tetsushi, Hu, Dinghuan, & Otsuka, Keijiro (2004). From inferior to superior products: An inquiry into the Wenzhou model of industrial development in China. *Journal of Comparative Economics*, *32*, 542–563.
Sonobe, Tetsushi, & Otsuka, Keijiro (2006a). The division of labor and the formation of industrial clusters in Taiwan. *Review of Development Economics*, *10*(1), 71–86.
Sonobe, Tetsushi, & Otsuka, Keijiro (2006b). *Cluster-based industrial development: An East Asia model.* New York: Palgrave MacMillan.
Summers, Lawrence (2007). The rise of Asia and the global economy. *Research Monitor (the bi-annual newsletter of the Global Development Network), Special Issue: 4–5.*
Wang, Jici, & Mei, Lixia (2009). Dynamics of labor-intensive clusters in China: Relying on low labor costs of cultivating innovation. *International Institute for Labor Studies Discussion Paper Series No. 195.*
Wei, Xiangdong, & Fan, C. Simon (2006). Converge to the law of one price in China. *The Review of Economics and Statistics*, *88*(4), 682–697.
Wen, Mei (2004). Relocation and agglomeration of Chinese industry. *Journal of Development Economics*, *73*(1), 329–347.
Young, Alwyn (2000). The razor's edge: Distortions and incremental reform in the People's Republic of China. *Quarterly Journal of Economics*, *115*(4), 1091-1135.
Zhang, Xiaobo, & Tan, Kong-Yam (2007). Incremental reform and distortions in China's product and factor markets. *World Bank Economic Review*, *21*(2), 279–299.