INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
IFPRI

IFPRI Discussion Paper 01468

October 2015

A Proximity-Based Measure of Industrial Clustering

Jianqing Ruan

Xiaobo Zhang

Development Strategy and Governance Division

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

The International Food Policy Research Institute (IFPRI), established in 1975, provides evidence-based policy solutions to sustainably end hunger and malnutrition and reduce poverty. The Institute conducts research, communicates results, optimizes partnerships, and builds capacity to ensure sustainable food production, promote healthy food systems, improve markets and trade, transform agriculture, build resilience, and strengthen institutions and governance. Gender is considered in all of the Institute's work. IFPRI collaborates with partners around the world, including development implementers, public institutions, the private sector, and farmers' organizations, to ensure that local, national, regional, and global food policies are based on evidence. IFPRI is a member of the CGIAR Consortium.

AUTHORS

Jianqing Ruan is an associate professor at the School of Management at Zhejiang University, Hangzhou, China.

Xiaobo Zhang is a senior research fellow in the Development Strategy and Governance Division of the International Food Policy Research Institute, Washington, DC and a distinguished professor of economics at the National School of Development at Peking University.

Notices

[1.] IFPRI Discussion Papers contain preliminary material and research results and are circulated in order to stimulate discussion and critical comment. They have not been subject to a formal external review via IFPRI's Publications Review Committee. Any opinions stated herein are those of the author(s) and are not necessarily representative of or endorsed by the International Food Policy Research Institute.

[2.] The boundaries and names shown and the designations used on the map(s) herein do not imply official endorsement or acceptance by the International Food Policy Research Institute (IFPRI) or its partners and contributors.

Contents

Abstract	v
1. Introduction	vi
2. Literature Review	2
3. A Proximity-Based Measure of Industrial Clustering	5
4. An Application to China	8
5. Which Clustering Measure Can Most Accurately Predict China’s Top 100 Clusters?	15
6. Conclusion	17
References	18

Tables

4.3 Top and bottom 10 counties according to proximity-based clustering measure 13

4.4 Top 10 provinces according to different clustering measures 14

5.1 Comparing the values of clustering measures between the top 100 clusters and elsewhere 16

5.2 Comparing the predictions of clustering measures on the top 100 clusters 16

Figures

4.1 Network of industrial proximity with the strongest links 10

4.2 Spatial distribution of industrial clustering 12

ABSTRACT

An industrial cluster is a locality with a high concentration of firms in related businesses. Although relatedness and concentration are the two defining features of an industrial cluster, the commonly used measures of clustering often fail to simultaneously capture both dimensions. Based on the product space literature, we first compute the degree of relatedness based on the concept of industrial proximity. Next, we develop a clustering index that takes into account both proximity and concentration. Finally, we calculate this new proximity-based index using the 1995 China Industrial Census and the 2004 and 2008 China Economic Census. The new index predicts China's top 100 industrial clusters much more accurately than existing cluster measures.

Keywords: industrial clusters, proximity, concentration, China

JEL: L10, L50, and L60

ACKNOWLEDGMENTS

This work was undertaken as part of, and funded by, the CGIAR Research Program on Policies, Institutions, and Markets (PIM) led by the International Food Policy Research Institute (IFPRI). Additional financial support from the Natural Science Foundation of China (Approval numbers 71441008, 71341002, and 71350002) is also gratefully acknowledged. This paper has not gone through IFPRI's standard peer-review procedure. The opinions expressed here belong to the authors, and do not necessarily reflect those of PIM, IFPRI, or CGIAR.

1. INTRODUCTION

Clusters are ubiquitous in both developed and developing countries (Piore and Sabel 1984; Sonobe and Otuska 2006; Otuska and Sonobe 2011). Based on observations during the period of British industrialization, Marshall (1920) highlighted three positive externalities—technology spillover, proximity to market, and labor pooling—in industrial districts, a term preceding the currently popular concept of the cluster. Scholars have subsequently identified additional advantages of clusters. For example, clusters are instrumental in creating collective efficiency (Schmitz and Nadvi 1999), such as sharing tools. In the early stage of a business's development, financial constraints may limit entrepreneurship. In a cluster, by dividing a production process into many incremental steps, the capital requirement of starting and running a business is greatly reduced, enabling more potential entrepreneurs to realize their dreams (Huang, Zhang, and Zhu 2008; Ruan and Zhang 2009; Long and Zhang 2011). Most firms in clusters in developing countries are labor intensive, thus making the cluster-based model particularly viable in the early stage of development, when labor is more abundant than capital. Not surprisingly, the United Nations Industrial Development Organization has advocated cluster-based development as an instrument for poverty reduction (Nadvi and Barrientos 2004).

To better understand the drivers and consequences of cluster-based development, one must be able to correctly measure the degree of clustering. People immediately know that a place is a cluster when seeing it, but a great challenge is to quantifiably measure it. In practice, the concepts of regional specialization, industrial concentration, and clustering are often interchangeably used. According to Porter (1990), a cluster is a place with highly concentrated firms that are closely related. However, the commonly used concentration or specialization indexes, such as the Hirschman-Herfindahl index and the Krugman index, focus on industrial concentration or regional specialization, ignoring the aspect of relatedness.

In this paper, we develop a clustering measure that takes into account both concentration and relatedness. First, we compute the degree of industrial relatedness within a location based on the idea of product space (Hidalgo et al. 2007). Next, we combine relatedness and relative scale in the calculation of the clustering index and apply it to the Chinese data. Finally, we compare the accuracy of the predictions of our index with other popular indexes of China's top clusters. Among six measures, our index performs the best.

Our paper is arranged as follows. Section 2 reviews the literature. Section 3 details the method of constructing our clustering index. A set of cluster measures is calculated using China's industrial and economic censuses in Section 4. Section 5 offers a comparison of the effectiveness of alternative measures of predicting China's top 100 clusters. Section 6 concludes.

2. LITERATURE REVIEW

Although clustering has been widely used as a concept, no specific measures exist for it. Instead, researchers have used several popular indexes, originally designed for measuring regional specialization or industrial concentration, as proxies. Regional specialization offers a territorial perspective on the distribution of industries within a region. If a region is dominated by one industry, then the region is highly specialized. Industrial concentration is the other side of the same coin. It measures the spatial distribution of an industry in a country. Mathematically, the measures for regional specialization and industrial concentration are the same if one swaps industries with regions in the notation. Since clusters are largely territory based, we refer only to the measures of regional specialization in this paper.

The concentration index (CRn), the Hirschman-Herfindahl index, the Krugman index, the Ellison and Glaeser index, and the local quotient are the most popular measures of regional specialization. CRn is defined as the share of the top *n* industries in total output (or employment) in a location. The concept is easy to understand: a larger value of the CRn index means a region is more specialized. However, the measure's problem lies in the fact that the index hinges upon the value of *n* and therefore cannot be compared across regions if *n* varies territorially. In the literature, *n* is often set to 3. In this case, CRn becomes CR3.

The Hirschman-Herfindahl index can largely address that problem by summing the squared shares of output or employment for all the industries in a region. The Gini coefficient also makes use of the output or employment shares of all the industries within a region. By measuring the inequality in shares of employment or output across industries in a location, the Gini coefficient provides another measure of regional specialization. The larger the Gini coefficient, the more specialized a region's economy.

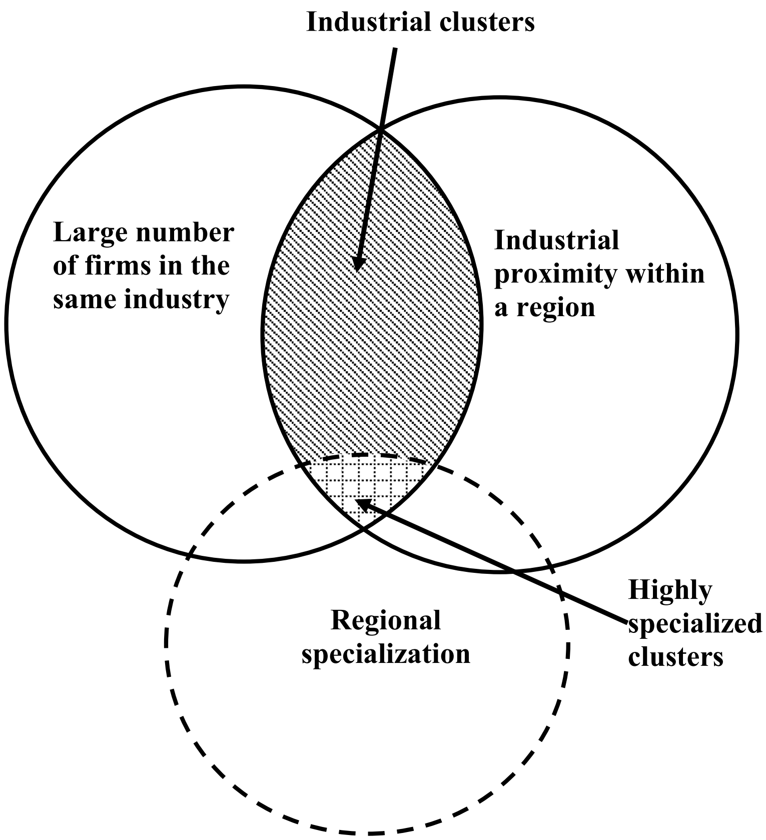
However, the preceding three indexes only measure absolute specialization. Whether a region has a comparative advantage in a certain industry also depends upon other regions. Krugman put forward a relative specialization measure by comparing the differences between a region's employment shares and corresponding national employment shares across industries. A larger value indicates that a region's industry structure deviates greatly from the national average and reveals more comparative advantage relative to other regions.

The aforementioned measures, however, are not scale free. A small region tends to have relatively fewer industries than a large area, thus becoming more likely to have seemingly large values of the measure of regional specialization. To address that concern, Ellison and Glaeser (1997) developed a new index that takes into account differences in the size distribution of plants and of geographic areas.[1] The Ellison–Glaeser index is more comparable across space and time than the aforementioned measures. Although the Ellison–Glaeser index takes relative scale into account, it does not consider firm relatedness in a location, an important feature of industrial clusters.

Most empirical works (Bai et al. 2004; Delgado, Porter, and Stern 2010) use measures of regional specialization as a proxy for clustering. Yet regional specialization is not identical to clustering. Most of the regional specialization measures capture only the dimension of specialization, ignoring two other key aspects of clustering: large scales and close firm relatedness. Figure 2.1 graphs the distinction between the concepts of regional specialization, production scale, and firm relatedness. A cluster must be a place with a large number of closely related firms. Yet the existing measures of regional specialization fail to capture relative scale, with the exception of the Ellison–Glaeser index, and firm relatedness, embedded in clusters.

[1] In the original paper, the index is developed to measure industrial concentration. By switching the notation of industries and regions, the index can also be used to measure regional specialization.

Figure 2.1 A diagram on clustering and regional specialization



Source: Drawn by authors.

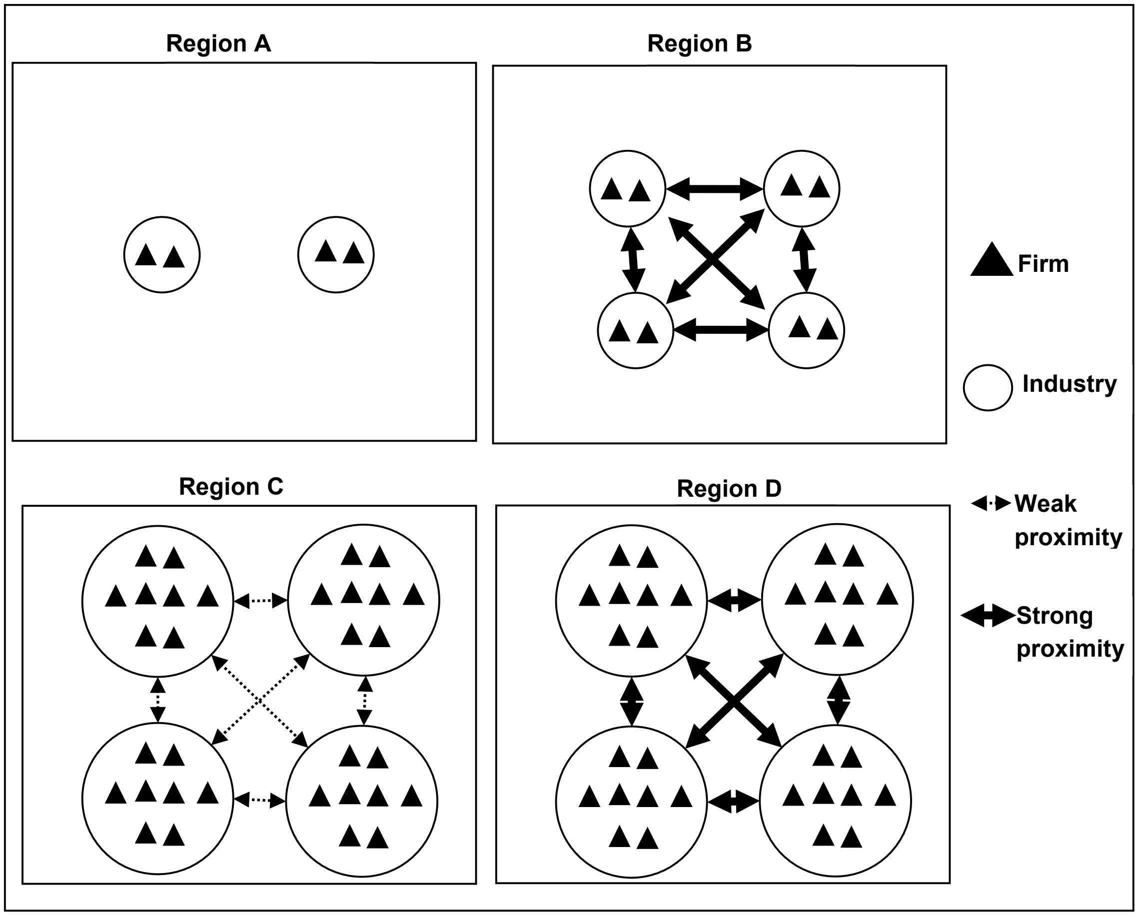
When using the traditional measures of regional specialization, many western counties in China, which often consist of only a few industries, would score higher than coastal counties, which usually have a large number of firms spanning different closely related industries. Because of this problem, many empirical papers (Rui and Swann 1998; Henderson 2003; Wen 2004) on clusters opt to use some simple yet crude measures, such as a region's output or employment in a particular industry as a share of national total output or employment. These measures do capture the relative scale of a region's industry; however, they totally ignore firm relatedness within the region.

Let's use Figure 2.2 to illustrate the distinction between clustering and regional specialization using four cases. We use black triangles to stand for firms, circles for industries, and double arrows for industrial relatedness. Thicker arrow lines indicate stronger ties among industries. For simplicity, we assume that the number of firms represents the production scale.[2] Region A includes only two industries with two firms in each industry. It is apparent that this area is not a cluster. In comparison, Region D is composed of four industries. There are eight firms in each industry, indicating a large scale. In addition, the firms in the four industries are closely related. Intuitively, Region D is much more clustered than Region A. However, according to the

[2] Of course, there are other ways to measure production scale, such as output, assets, and number of workers. We use number of firms purely for the purpose of illustration.

traditional measures of regional specialization, Region A would score higher than Region D, suggesting the limitation of using the regional specialization measure as a proxy for clustering.

Figure 2.2 An illustration of the clustering concept



Source: Drawn by authors.

Region B is similar to Region D except for its smaller production scale. Each industry in Region B includes only two firms, compared with eight firms in Region D. So an ideal clustering measure should indicate a higher degree of clustering in Region D than in Region B.

Region C differs from Region D in that the ties among the four industries in Region C are weaker. The conventional regional specialization measures would yield the same value for both regions because these measures do not take firm relatedness into account.

In sum, the commonly used regional specialization measures or crude measures of relative scale are not suitable for measuring the degree of clustering.

3. A PROXIMITY-BASED MEASURE OF INDUSTRIAL CLUSTERING

To overcome the limitations of the regional specialization measures, we develop a new clustering measure by taking both industrial scale and relatedness into account. We first compute industrial relatedness before constructing our clustering index.

Proximity

It is not an easy task to measure the degree of relatedness across industries. Firms are related in many ways, such as an input–output relationship, technology spillovers, and information sharing. The difficulty in measuring such various relationships is likely a key reason why the concept of relatedness has been rarely used in the literature on regional specialization and clustering.

The seminal work by Hidalgo et al. (2007, shortened as HKBH hereafter) provides a way to measure the relatedness of product space. The key insight is that if two products share similar inputs and production technologies, they are likely to have similar comparative advantages in terms of export. If one product is exported, the chance of the other product also being exported should be very high. Based on this insight, the proximity between each pair of goods can be computed as the probability that a country has exported both products (averaged over all countries in the world). Specifically, HKBH use the following algorithm to compute product relatedness: first, they compute the average probability of product B with a revealed comparative advantage (RCA > 1) across countries given that product A has a revealed comparative advantage; next, they compute product A's conditional revealed comparative advantage probability if product B has a revealed comparative advantage (RCA > 1); the smaller value of the preceding two conditional probabilities is defined as the measure of relatedness between products A and B. Using this method, HKBH derive a matrix of proximity among all export products based on trade data.

Based on the proximity matrix, Long and Zhang (2011 and 2012) develop a new clustering index that takes relatedness into account. Despite the improvement, their method still suffers a few shortcomings. First, the world trade database covers only tradable products. Nontradable products or little-exported products cannot be fully captured. Second, the industrial classifications are not totally consistent with those used in China. As such, converting the matrix of proximity from the international standard industrial classification (SIC) to the Chinese SIC may generate some errors. Third, exports are influenced not only by production technologies but also by other factors, such as the tariff rate and exchange rate. Thus, equalizing export proximity to product proximity may bring about some biases. Finally, their index does not consider the relative size of a region's output or employment.

Building on the method of Long and Zhang (2011) and following in the spirit of the HKBH proximity measure, this paper directly uses firm-level data from the China Industrial Census and China Economic Census to calculate relatedness across industries instead of using world trade data. The HKBH approach measures product proximity based on the revealed comparative advantage (RCA) of export products. Since our focus is on the comparison of clustering across regions within a country, we employ an industry's local quotient as a measure of its comparative advantage. One key advantage of using the local quotient is that more firm information, such as employment, assets, and output, is readily available, whereas RCA exclusively relies on export data.

We use the following algorithm to calculate industry proximity: first, we compute the local quotient of all the industries in a location. If an industry's local quotient is greater than 1, then that industry demonstrates comparative advantage in the location. Second, by looking at all the regions, we calculate the conditional probability of industry j demonstrating a comparative advantage when industry i has a comparative advantage in a location. Third, we repeat the preceding step to compute the conditional probability that industry i displays a comparative advantage when industry j shows a comparative advantage. Fourth, we select the smaller value from the above two conditional probabilities as the measure of industry proximity.

Next we use Chinese data to illustrate the procedure in more detail. We treat counties and districts as a territory unit. The classification of industries follows SIC3. The number of workers is used as an outcome variable in the calculation of the local quotient.

First, we compute each county/district's local quotient as follows:

$$LQ_{rj} = \frac{E_{rj}/E_r}{E_{cj}/E_c}, \quad (1)$$

where r stands for region; j refers to industry; E_{rj} is the total employment in industry j in region r; E_r amounts to total employment in region r; E_{cj} represents national total employment in industry j; and E_c is defined as the total employment at the country level. When LQ_{rj} is greater than 1, industry j in region r reveals a comparative advantage relative to the national average.

If LQ_{rj} is greater than 1, we can calculate the conditional probability that industry i in region r also reveals a comparative advantage as follows:

$$P(LQ_{ri} > 1 | LQ_{rj} > 1) = \frac{P(LQ_{ri}>1 \cap LQ_{rj}>1)}{P(LQ_{rj}>1)}. \quad (2)$$

Likewise, we can compute the conditional probability of industry j also having a comparative advantage when industry i shows a comparative advantage, $P(LQ_{rj} > 1 | LQ_{ri} > 1)$. Following Hidalgo et al. (2007), we define the smaller value of the two conditional probabilities as the industry proximity measure:

$$\emptyset_{ije} = \min\{P(LQ_{ri} > 1 | LQ_{rj} > 1), P(LQ_{rj} > 1 | LQ_{ri} > 1)\}. \quad (3)$$

In the preceding calculation, we have used number of workers as an outcome variable. Since employment represents only one aspect of firm performance, we can also use output, assets, and number of firms as alternative outcome variables. To obtain a more comprehensive view of industry relatedness in this paper, we also calculate three industry proximity measures based on output, assets, and number of firms, $\emptyset_{ijo}$, $\emptyset_{ijk}$, and $\emptyset_{ijf}$. We define the industry proximity as the simple average of the four measures:

$$\emptyset_{ij} = {(\emptyset_{ije} + \emptyset_{ijo} + \emptyset_{ijk} + \emptyset_{ijf})}/{4}. \quad (4)$$

With this information, we can create a matrix of industrial proximity.

A Proximity-Based Clustering Measure

With the matrix of industrial proximity, we are now in a position to compute the regional clustering index. First, we calculate an industry's overall proximity with all other industries in a location. Suppose a region has n industries. Using the total number of workers as a weight, industry i's proximity with other industries can be obtained as follows:

$$\emptyset_{ri} = \sum_{j\neq i}^{n}(\emptyset_{ij}\frac{E_{rj}}{\sum_{j\neq i}^{n} E_{rj}}), \quad (5)$$

where E_{rj} represents employment in industry j in region r; n is the total number of industries in the region; and $\emptyset_{ij}$ stands for the proximity coefficient between industry i and industry j available from the matrix of industrial proximity.

Next, we aggregate the above industrial proximity using the relative share of each industry's employment or output in a region as weights. In particular, we adopt the following method to compute the clustering index in region *r*:

$$\mathrm{Cluster_r} = \sum_{i=1}^{n} (\emptyset_{ri} \frac{E_{ri}}{E_{ci}}), \tag{6}$$

where E_{ri} and E_{ci} represent industry *i*'s total employment in region *r* and the country as a whole, respectively; and $\emptyset_{ri}$ is industry *i*'s proximity with other industries in region *r* obtained from (5). Of course, alternative weights, such as output, assets, and number of firms can also be used as weights in (5) and (6). This clustering index captures not only the relative size of a region's industries but also the relatedness among industries within the region. In principle, it accords more closely with our common sense appraisal of industrial clusters.

4. AN APPLICATION TO CHINA

Data

Most previous studies (Li and Lu 2009; Lu and Tao 2009) on China's clusters make use of the Annual Survey of Industrial Enterprises in China (ASIEC). However, the database covers all the state-owned enterprises (SOEs) and above-scale private firms with sales exceeding 5 million yuan but excludes many domestic private small and medium-size enterprises (SMEs). Since clusters are often composed of numerous SMEs, a clustering measure based on a dataset that excludes SMEs will likely differ from reality. For example, whereas ASIEC includes about 0.3 million firms in 2004, the total number of enterprises covered in the China Economics Census 2004 approximates 1 million. To cover the sample bias of ASIEC, we opt to use the firm-level data from the China Industrial Census 1995, China Economic Census 2004, and China Economic Census 2008 in calculating the proximity index and clustering index. Since the focus of the paper is on industrial clusters, we exclude the agriculture, construction, and service sectors. At the SIC3 level, our remaining dataset includes 191 industries.

Industrial Proximity in 1995, 2004, and 2008

Based on the China Industrial Census 1995, China Economic Census 2004, and China Economic Census 2008, we compute four sets of industrial proximity in 1995, 2004, and 2008 with employment, output, assets, and number of firms as weights, respectively. Table 4.1 reports the summary statistics of industrial proximity in the three years. The four measures largely reveal the same pattern. Except for the asset-based measure, all the measures inch up in proximity from 1995 to 2004. All four measures reveal a unanimous noticeable increase in proximity from 2004 to 2008.

Table 4.1 Summary statistics of industrial proximity

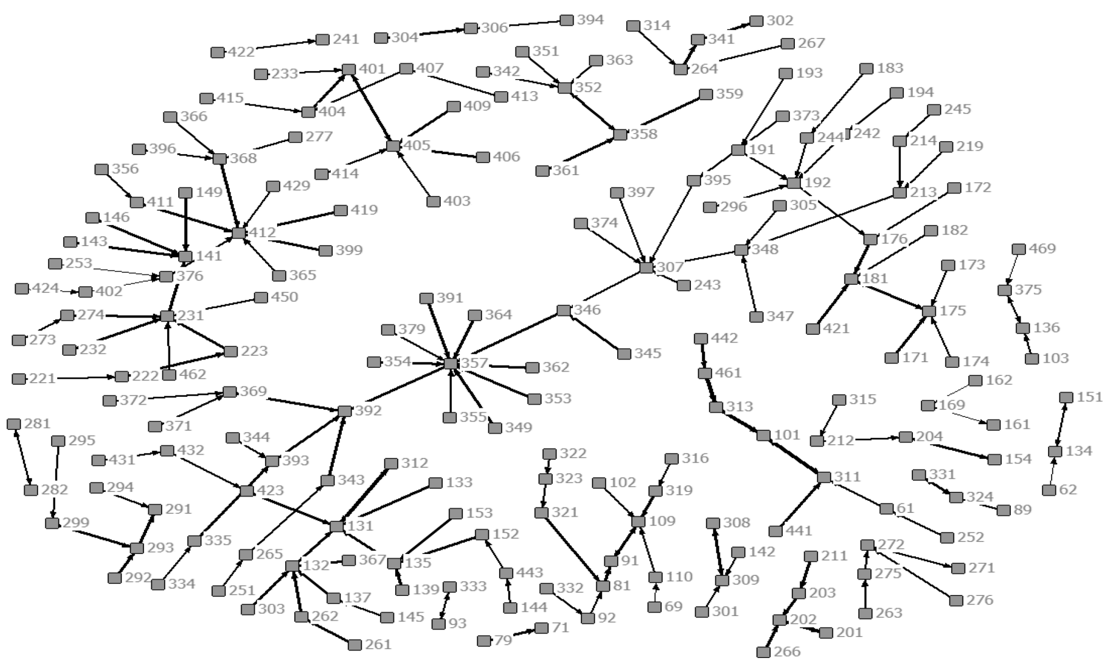
Measures based on the following variable	1995					2004					2008				
	Mean	Med	Min	Max	Std.	Mean	Med	Min	Max	Std.	Mean	Med	Min	Max	Std.
Employment	0.118	0.106	0.000	0.630	0.082	0.114	0.102	0.000	0.617	0.076	0.122	0.110	0.000	0.640	0.079
Output	0.099	0.087	0.000	0.586	0.071	0.100	0.088	0.000	0.562	0.069	0.108	0.096	0.000	0.598	0.073
Assets	0.106	0.091	0.000	0.674	0.080	0.105	0.093	0.000	0.547	0.072	0.116	0.105	0.000	0.607	0.079
No. of firms	0.159	0.153	0.000	0.563	0.106	0.161	0.149	0.000	0.618	0.099	0.167	0.154	0.000	0.697	0.099
Mean	0.121	0.111	0.000	0.613	0.082	0.120	0.111	0.000	0.569	0.075	0.128	0.118	0.000	0.616	0.079

Source: The pairwise industrial proximity is computed based on the China Industrial Census (1995), the China Economic Census (2004), and the China Economic Census (2008). It is based either on employment, output, assets, or number of firms.

Note: Med = median; Min = minimum; Max = maximum; Std. = standard error.

The industrial proximity matrix is a 191 × 191 matrix, too large to be presented in a table. Instead, we plot the matrix in 2008 using a social network graph in Figure 4.1. For simplicity, the figure reports only links of paired industries with the strongest relatedness. It is apparent from the figure that large variations in proximity exist across industries. Some industries occupy key nodes in the network with strong links with multiple industries, while a large number of industries possess only one strong link.

Figure 4.1 Network of industrial proximity with the strongest links



Source: Drawn by authors based on the China Economic Census (2008).

Clustering Index

Based on industrial proximity, we can calculate the clustering index at the county/district level in 1995, 2004, and 2008. Because the share of a county's employment or output in the national total is generally small, we transform the shares into a percentage by multiplying by 100. Table 4.2 summarizes the clustering index in the three years based on four different weights. The measures vary greatly from zero to 112. Regardless of the weight, all four measures indicate that Chinese counties/districts have become increasingly clustered from 1995 to 2008, in accordance with the findings in Long and Zhang (2012).

Table 4.2 Summary statistics of proximity-based clustering measure

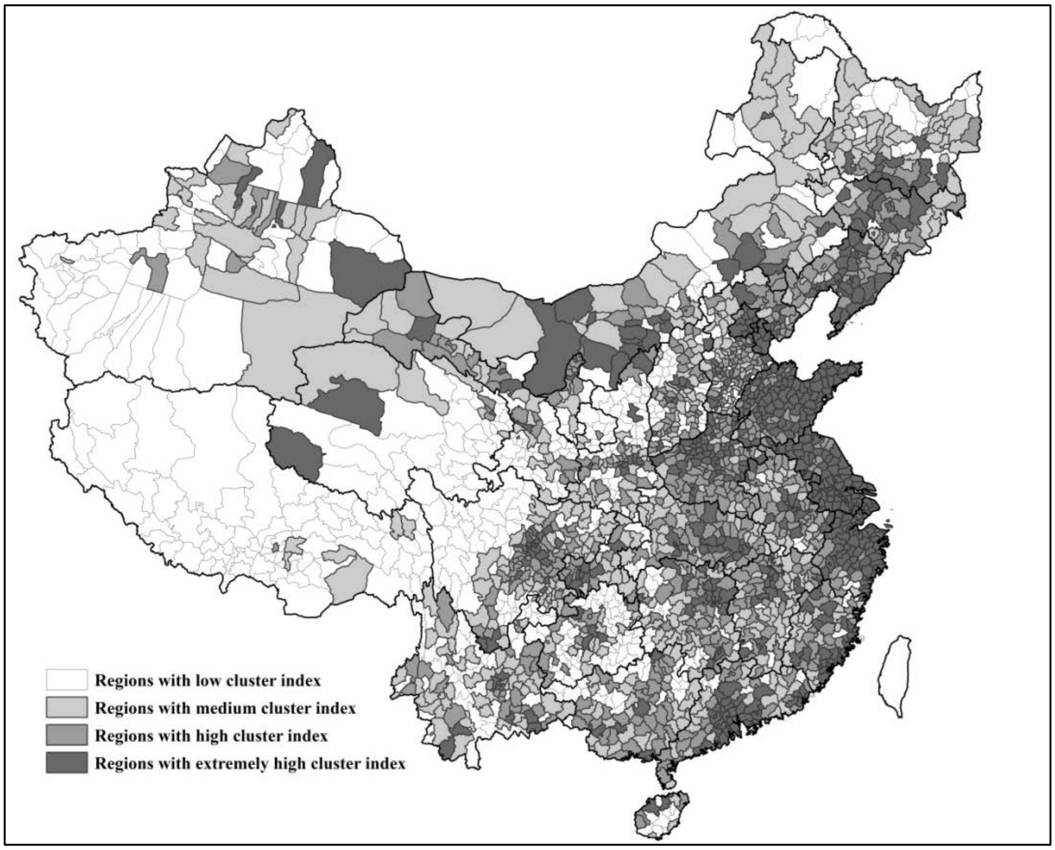
Measures based on the following weight	1995					2004					2008				
	Mean	Med	Min	Max	Std.	Mean	Med	Min	Max	Std.	Mean	Med	Min	Max	Std.
Employment	1.02	0.51	0.00	46.04	1.77	1.06	0.47	0.00	105.73	2.84	1.12	0.52	0.00	111.94	2.99
Output	1.00	0.33	0.00	27.68	2.08	1.02	0.33	0.00	63.36	2.60	1.08	0.40	0.00	58.76	2.47
Assets	0.96	0.36	0.00	34.05	1.94	1.00	0.33	0.00	63.55	2.56	1.07	0.40	0.00	62.41	2.54
No. of firms	1.06	0.61	0.00	30.22	1.57	1.08	0.51	0.00	52.04	2.19	1.16	0.57	0.00	47.49	2.22

Source: The clustering measures are computed based on the China Industrial Census (1995), the China Economic Census (2004), and the China Economic Census (2008).

Note: Med = median; Min = minimum; Max = maximum; Std. = standard error. The clustering index is calculated at the county/district level using employment, output, assets, and number of firms as weights, respectively.

Since there are more than 2,000 counties and districts in China, we cannot report all of them in this paper. Instead, we plot the clustering index at the county/district level in 2008 in a map (Figure 4.2). The coastal area figures most prominently in clustering, in particular, with clusters most highly concentrated in Shandong, Jiangsu, Zhejiang, and Guangdong provinces. In inland China, some clusters are evident in Henan, Hubei, Hunan, Sichuan, and Chongqing provinces.

Figure 4.2 Spatial distribution of industrial clustering



Source: Drawn by authors based the China Economic Census (2008).

Note: The clustering measure at the county level is based on the China Economic Census 2008 using employment as weight. The figures are similar when using output, assets, and number of firms as weights. Taiwan is excluded due to lack of data.

Table 4.3 further lists the names and the clustering index of the top and bottom 10 counties in 2008. Among the top clusters, Guangdong province accounts for five of the top 10. Dongguan in Guangdong province ranks at the top in four clustering measures. It is followed by three other cities/districts in Guangdong, Bao’an, Zhongshan, and Shunde. A few counties in Shanghai, Jiangsu, and Zhejiang also rank in the top 10. Nine out of the 10 bottom counties are located in Tibet, and the remaining one is in Qinghai. Both Tibet and Qinghai have very few industrial clusters. The results match closely with observed reality.

Table 4.3 Top and bottom 10 counties according to proximity-based clustering measure

			Clustering measure			
	County/district	Province	Employment	Output	Assets	Number of firms
Top 10	Dongguan	Guangdong	111.9444	58.764	62.4103	47.4902
	Bao`an district,	Guangdong	56.6861	38.234	38.0668	34.6448
	Zhongshan	Guangdong	35.9553	29.733	25.3073	22.4534
	Shendu district,	Guangdong	21.7857	23.398	15.8657	21.1430
	Wujin district,	Jiangsu	13.6392	12.299	13.2887	20.8477
	Nanhai district,	Guangdong	20.0250	23.668	12.8790	20.6230
	Kunshan	Jiangsu	21.2187	22.401	26.3461	18.7268
	Songjiang	Shanghai	15.7119	16.570	20.7157	18.1683
	Cixi	Zhejiang	12.5522	9.0147	12.6856	18.0854
	Jiading	Shanghai	16.0711	18.944	20.0453	17.7861
Bottom 10	Kangma	Xizang	0.0014	0.0008	0.0016	0.0026
	Qiongjie	Xizang	0.0012	0.0002	0.0010	0.0022
	Xietongmen	Xizang	0.0019	0.0019	0.0016	0.0021
	Bailang	Xizang	0.0025	0.0005	0.0011	0.0020
	Luolong	Xizang	0.0012	0.0001	0.0003	0.0017
	Zaduo	Qinghai	0.0010	0.0001	0.0008	0.0015
	Dingqing	Xizang	0.0010	0.0003	0.0003	0.0015
	Gongjue	Xizang	0.0006	0.0000	0.0002	0.0015
	Zuogong	Xizang	0.0005	0.0000	0.0001	0.0014
	Gongga	Xizang	0.0005	0.0000	0.0001	0.0006

Source: The clustering measures are computed based on the China Economic Census (2008).
Note: The ranking is based on our clustering measure with number of firms as weight.

Based on the county-level clustering index, we can obtain the average value of the clustering index at the province level. To compare the performance of our proximity-based cluster index with other popularly used indexes at the provincial level, we also separately compute the average values of five other clustering measures at the provincial level: the Long and Zhang clustering index, the Ellison–Glaeser index, the CR3 index, the Gini coefficient, and the Krugman index. Table 4.4 lists the top 10 provinces according to the six different clustering measures. The proximity-based index reveals that coastal provinces dominate the top 10, with only Henan, an inland province, as an exception. According to the Long and Zhang index, only three out of the top 10 provinces are not from coastal areas. In comparison, according to the other four popular indexes, inland provinces almost exclusively occupy the top 10 list. Hebei and Jiangsu are the only coastal provinces among the top 10 based on the CR3 index. These four measures' rankings contradict the fact that most industrial clusters are located in the coastal regions.

Table 4.4 Top 10 provinces according to different clustering measures

Rank	Proximity-based	Long and Zhang	Ellison and Glaeser	CR3	Gini	Krugman
1	Shanghai	Hainan	Shanxi	Xizang	Shanxi	Shanxi
2	Guangdong	Xizang	Guizhou	Qinghai	Ningxia	Xizang
3	Beijing	Beijing	Gansu	Shanxi	Shaanxi	Qinghai
4	Tianjin	Shanghai	Ningxia	Hebei	Qinghai	Heilongjiang
5	Zhejiang	Hebei	Shaanxi	Jiangsu	Guizhou	Xinjiang
6	Jiangsu	Zhejiang	Chongqing	Ningxia	Gansu	Guizhou
7	Shandong	Qinghai	Inner Mongolia	Hunan	Hebei	Inner Mongolia
8	Fujian	Heilongjiang	Sichuan	Heilongjiang	Heilongjiang	Shaanxi
9	Henan	Guangdong	Heilongjiang	Xinjiang	Liaoning	Gansu
10	Liaoning	Tianjin	Hunan	Shaanxi	Chongqing	Sichuan

Source: The clustering measures are computed based on the China Economic Census (2004).

Note: The measure at the provincial level takes the average value of clustering measures at the county level, which use employment as weight based on the China Economic Census 2004.

5. WHICH CLUSTERING MEASURE CAN MOST ACCURATELY PREDICT CHINA'S TOP 100 CLUSTERS?

In 2007, the Institute of Industrial Economy in the Chinese Academy of Social Science published China's top 100 clusters based on a rich set of indicators, including opinions of leading informants in different industries, from more than 1,000 clusters. In this section, we would like to test which clustering measures offer the best prediction of the top 100 clusters.

Most industrial clusters are found at the county/district level. However, they do not perfectly match with territorial lines. In some cases, a county or district may have more than one cluster. For example, the Lucheng district of Wenzhou city includes two famous clusters: footwear and lighters. A few clusters span more than one county/district, such as the motorbike cluster in Jiangmen city. In this case, we treat all the counties in the cluster as a top cluster county. Since our clustering measures are at the county/district level, we map the top 100 clusters with corresponding counties/districts. We end up with a list of 109 clusters at the county/district level.

Based on the China Economic Census 2004, we compute our newly developed clustering measure as well as several alternative measures—Ellison–Glaeser, CR3, Gini, the Krugman index, and the Long and Zhang index. Table 5.1 presents the average clustering measures in the top cluster counties/districts and other counties as well as the *t*-test for the difference between the two mean values. Apparently, our clustering measure scores much higher in the top cluster counties/districts than elsewhere, and the difference is statistically significant. The result is robust with respect to the choice of alternative weights. The Long and Zhang index also indicates a higher value in the areas of top clusters, but the difference is statistically significant only when employment is used as a weight. In contrast, the other cluster measures tend to yield greater values in regions outside the top clusters.

We also rank counties/districts according to the six different clustering measures and count how many of the top 109 clusters overlap with the published top clusters in 2007. Table 5.2 reports the prediction results. When using employment as weight, our clustering measure predicts 53 of 109 top clusters with a success rate of 48.62 percent. By comparison, the other five measures can at most successfully predict two top clusters. The patterns remain the same when using assets, output value, or number of firms as weights. In summary, our newly developed clustering measure offers a great improvement in the prediction of clusters over other commonly used measures.

Table 5.1 Comparing the values of clustering measures between the top 100 clusters and elsewhere

	Employment			Output			Assets			Number of workers		
	Top 100 clusters?		*p*-value of *t*-test	Top 100 clusters?		*p*-value of *t*-test	Top 100 clusters?		*p*-value of *t*-test	Top 100 clusters?		*p*-value of *t*-test
	Yes	No		Yes	No		Yes	No		Yes	No	
Proximity-based	6.862	0.830	0.000	7.151	0.777	0.000	6.624	0.778	0.000	6.597	0.866	0.000
Long and Zhang	0.229	0.219	0.001	0.229	0.226	0.180	0.229	0.226	0.158	—	—	—
Ellison and Glaeser	0.042	0.066	0.990	0.041	0.065	0.955	0.039	0.035	0.459	0.031	0.060	1.000
CR3	0.913	0.906	0.130	0.925	0.927	0.615	0.917	0.919	0.627	0.920	0.897	0.000
Gini	0.388	0.400	0.808	0.420	0.451	0.985	0.408	0.439	0.985	0.390	0.364	0.013
Krugman	0.467	0.590	1.000	0.588	0.735	1.000	0.596	0.738	1.000	0.343	0.460	1.000

Source: The clustering measures are computed based on the China Economic Census (2004).

Note: The national top 100 clusters cover 109 districts and counties. All the clustering measures are based on the China Economic Census 2004.

Table 5.2 Comparing the predictions of clustering measures on the top 100 clusters

	Employment		Output		Assets		Number of workers	
	Number	%	Number	%	Number	%	Number	%
Proximity-based measure	53	48.62	53	48.62	48	44.04	53	48.62
Long and Zhang	1	0.92	3	2.75	1	0.92	—	—
Ellison and Glaeser	0	0.00	1	0.92	2	1.83	0	0.00
CR3	0	0.00	0	0.00	0	0.00	0	0.00
Gini	2	1.83	2	1.83	1	0.92	8	7.34
Krugman	1	0.92	2	1.83	1	0.92	0	0.00

Source: The clustering measures are computed based on the China Economic Census 2004.

Note: The percentage column equals the number of correctly predicted counties in the top 100 clusters divided by 109.

6. CONCLUSION

Clusters are widespread in the real world and play an important role in local economic development—but we still lack a good quantification of the clustering phenomenon. Previous measures of regional specialization or industrial concentration are often used as a proxy; unfortunately, more often than not such measures fail to reflect the actual situation of clustering. One main root cause is that those measures do not take into account industrial proximity, a key feature inherent in clusters. To remedy this deficiency, our paper develops a new clustering index based on the idea of product space. Both the relative size of a region's industries and the proximity across industries within a region are considered in the construction of our new index. Using firm-level census data in China, we illustrate how to construct a clustering index, then compare its performance with other commonly used measures in terms of predicting the observed top industrial clusters in China. Our proximity-based clustering index undoubtedly offers the most accurate predication and can serve as a stepping-stone toward further studying the consequences and causes of industrial clusters.

REFERENCES

Bai, C.-E., Y. Duan, Z. Tao, and S. T. Tong. 2004. "Local Protectionism and Regional Specialization: Evidence from China's Industries." *Journal of International Economics* 63 (2): 397–417.

Delgado, M., M. E. Porter, and Scott Stern. 2010. "Clusters and Entrepreneurship." *Journal of Economic Geography* 10: 495–518.

Ellison, G., and E. L. Glaeser. 1997. "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach." *Journal of Political Economy* 105 (5): 889–927.

Henderson, J. V. 2003. "Marshall's Scale Economies." *Journal of Urban Economics* 53: 1–28.

Hidalgo, C. A., B. Klinger, A.-L. Barabási, and R. Hausmann. 2007. "The Product Space Conditions the Development of Nations." *Science* 317: 482–487.

Huang, Z., X. Zhang, and Y. Zhu. 2008. "The Role of Clustering in Rural Industrialization: A Case Study of Wenzhou's Footwear Industry." *China Economic Review* 19: 409–420.

Li, B., and Y. Lu. 2009. "Geographic Concentration and Vertical Disintegration: Evidence from China." *Journal of Urban Economics* 65: 294–304.

Long, C., and X. Zhang. 2011. "Cluster-Based Industrialization in China: Financing and Performance." *Journal of International Economics* 84 (1): 112–123.

Long, C., and X. Zhang. 2012. "Patterns of China's Industrialization: Concentration, Specialization, and Clustering." *China Economic Review* 23 (3): 593–612.

Lu, J., and Z. Tao. 2009. "Trends and Determinants of China's Industrial Agglomeration." *Journal of Urban Economics* 65 (2): 167–180.

Marshall, A. 1920. *Principles of Economics.* London: Macmillan.

Nadvi, K., and S. Barrientos. 2004. *Industrial Clusters and Poverty Reduction.* Vienna: United Nations Industrial Development Organization.

Otsuka, K., and T. Sonobe. 2011. *A Cluster-Based Industrial Development Policy for Low-Income Countries.* Policy Research Working Paper 5703. Washington, DC: World Bank.

Piore, M. J., and C. F. Sabel. 1984. *The Second Industrial Divide.* New York: Basic Books.

Porter, M. E. 1990. *The Competitive Advantage of Nations.* New York: Free Press.

Ruan, J., and X. Zhang. 2009. "Finance and Cluster-Based Industrial Development in China." *Economic Development and Cultural Change* 58: 143–164.

Rui, B., and P. Swann. 1998. "Do Firms in Clusters Innovate More?" *Research Policy* 27: 525–540.

Schmitz, H., and K. Nadvi. 1999. "Clustering and Industrialization: Introduction." *World Development* 27: 1503–1514.

Sonobe, T., and K. Otsuka. 2006. *Cluster-Based Industrial Development: An East Asian Model.* United Kingdom: Palgrave Macmillan.

Wen, M. 2004. "Relocation and Agglomeration of Chinese Industry." *Journal of Development Economics* 73: 329–347.

RECENT IFPRI DISCUSSION PAPERS

For earlier discussion papers, please go to www.ifpri.org/pubs/pubs.htm#dp.
All discussion papers can be downloaded free of charge.

1467. *.Aspirations and the role of social protection: Evidence from economic shocks in rural Pakistan*. Katrina Kosec and Cecilia Hyunjung Mo, 2015.

1466. *Addressing human capital development in public agriculture extension in southern Africa: Assessing Mozambique's experience.* Hélder R. Gêmo and Kristin E. Davis, 2015.

1465. *Microinsurance decisions: Gendered evidence from rural Bangladesh*. Daniel J. Clarke and Neha Kumar, 2015.

1464. *Institutions and market integration: The case of coffee in the Ethiopian Commodity Exchange.* Manuel A. Hernandez, Shahidur Rashid, Solomon Lemma, and Tadesse Kuma, 2015.

1463. *Happiness in the air: How does a dirty sky affect subjective well-being?* Xin Zhang, Xiaobo Zhang, and Xi Chen, 2015.

1462. *The impact of the use of new technologies on farmers' wheat yield in Ethiopia: Evidence from a randomized controlled trial.* Gashaw Tadesse Abate, Alan de Brauw, Nicholas Minot, and Tanguy Bernard, 2015.

1461. *Living like there's no tomorrow: Saving and spending following the Sichuan earthquake*. Mateusz Filipski, Ling Jin, Xiaobo Zhang, and Kevin Chen, 2015.

1460. *Agricultural value chain development in practice: Private sector–led smallholder Development.* Shashidhara Kolavalli, Akwasi Mensah-Bonsu, and Saima Zaman, 2015.

1459. *Cropping Intensity gaps: The potential for expanding global harvest areas.* Wenbin Wu, Liangzhi You, and Kevin Chen, 2015.

1458. *Migration, gender, and farming systems in Asia: Evidence, data, and knowledge gaps*. Valerie Mueller, Chiara Kovarik, Kathryn Sproule, and Agnes Quisumbing, 2015.

1457. *Los efectos de la roya en las economías Centroamericanas.* Valeria Piñeiro, Samuel Morley, and Pablo Elverdin, 2015.

1456. *Agriculture, gendered time use, and nutritional outcomes: A systematic review.* Deborah Johnston, Sara Stevano, Hazel Malapit, Elizabeth Hull, and Suneetha Kadiyala, 2015.

1455. *Public benefits of private technology adoption: the localized spatial externalities of water conservation in eastern Uttar Pradesh.* Anil K. Bhargava, Travis J. Lybbert, and David J. Spielman, 2015.

1454. *Supply-side dynamics of chickpeas and pigeon peas in India.* Kalimuthu Inbasekar, Devesh Roy, and P. K. Joshi, 2015.

1453. *Measuring women's decisionmaking: Indicator choice and survey design experiments from cash and food transfer evaluations in Ecuador, Uganda, and Yemen.* Amber Peterman, Benjamin Schwab, Shalini Roy, Melissa Hidrobo, and Daniel Gilligan, 2015.

1452. *The potential of farm-level technologies and practices to contribute to reducing consumer exposure to aflatoxins: A theory of change analysis.* Nancy Johnson, Christine Atherstone, and Delia Grace, 2015.

1451. *How will training traders contribute to improved food safety in informal markets for meat and milk? A theory of change analysis.* Nancy Johnson, John Mayne, Delia Grace, and Amanda Wyatt, 2015.

1450. *Communication and coordination: Experimental evidence from farmer groups in Senegal*. Fo Kodjo Dzinyefa Aflagah, Tanguy Bernard, and Angelino Viceisza, 2015.

1449. *The impact of household health shocks on female time allocation and agricultural labor participation in rural Pakistan.* Gissele Gajate-Garrido, 2015.

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

www.ifpri.org

IFPRI HEADQUARTERS

2033 K Street, NW
Washington, DC 20006-1002 USA
Tel.: +1-202-862-5600
Fax: +1-202-467-4439
Email: ifpri@cgiar.org