Paper presented at the annual Allied Social Science Associations January 3-5, 2015, San Francesco, California.

The Spatial Patterns of E-commerce in China

Xiaobo Zhang
National School of Development, Peking University

Wu Zhu
National School of Development, Peking University

December 31, 2015

Abstract

In the past several years, China has become the largest e-commerce market in the world. Using a unique e-commerce dataset from Alibaba at the county level, we first examine the spatial patterns of online sales and shopping in China. We find that online sellers are more likely to be located in existing industrial clusters, where thousands of firms are concentrated in one place and produce one major product. Online shopping has picked up more rapidly in less populated areas. With e-commerce, people have access to a wider range of commodities which in turn expands the market size of industrial clusters. Ultimately, the growth rate of e-commerce among the top clusters is more rapid than less clustered places—a winner-takes-all phenomenon.

Key words: E-commerce, clusters; China, and winner-takes-all.

China has overtaken the US and become the largest e-commerce market in the world since 2013 (Economist, 2015; see Figure 1). In 2014, online sales accounted for more than ten percent of total sales in China, compared to less than seven percent in the US. From 2013 to 2014, China's online sales grew at 52%, much faster than in the US (17%). Despite its sheer size and rapid growth, little is known about the evolving spatial patterns of China's e-commerce due to a lack of disaggregate date at the regional level.[1] For example, where have online sales and shopping experienced the fastest growth? Has e-commerce strengthened or weakened the existing manufacturing centers?

In this study, we use a unique e-commerce development index at the county level compiled by Alibaba, the largest e-commerce company in the world, in combination with China's firm registry database to answer the above questions. First, we show that online sales are mainly concentrated in the coastal areas where there is a strong presence of industrial clusters; while online purchases are from all over China, in particular from less populated western regions. The wide span of purchases from previously far-reaching areas demonstrates how e-commerce has expanded the market size. Second, we find that the expansion of market has empowered the competitive advantage of existing industrial clusters, creating a winners-take-all phenomenon. As a consequence, China's industrial production has become increasingly more clustered along with e-commerce development.

The paper is organized as follows. Section 2 provides background of China's e-commerce, major data sources, as well as the spatial patterns of e-commerce and cluster development in China. Section 3 explores the relationship between industrial clustering and e-commerce development. Section 4 concludes.

Background and Data

Background

Alibaba is not only the largest e-commerce company in China but also in the world. Alibaba established Taobao, a consumer to consumer platform (C2C), in 2003, which accounts for about 80% of China's C2C market. Tmall, a business to consumer platform (B2C) set up in 2008, dominates more than half of China's B2C online sales. In 2012, Alibaba's online sales were larger than Amazon and eBay combined. Although JD.com, the second largest e-commerce company in China, has grown faster than Alibaba in the past two years, its market share was only 18.6%, compared to 61.4% by Alibaba.[2]

In 2014, Alibaba released an e-commerce development index for the top 100 counties in China. We requested a more detailed online purchase and online sales index at the county level for China as a whole from Alibaba. While the provided data from 2013 is

[1] A few existing studies primarily focus on the mechanism of e-commerce market, such as Fan and Mo (2015).
[2] http://www.nasdaq.com/article/jdcom-grabs-china-online-shopping-share-from-alibaba-cm462542.

complete, the 2014 data is not. Therefore, our main analysis is based on the e-commerce index in 2013. Although we also examine the growth in the e-commerce index from 2013 to 2014 as a robustness check, the limitations of the 2014 data should be kept in mind.

E-commerce index

The e-commerce index includes two components: an online sale index and an online purchase index. The definition of the online sale index at the county level is equal to 0.5*(number of B2B sellers/population + number of B2C sellers/population) + 0.5* total value sold/number of sellers. The online purchase index is defined as 0.5*number of consumers online/population + 0.5*total value bought/number of consumers.

Consumers online refer to those who purchase online at least once a year. Population size is based on the China Population Census 2010. The information on online purchase and sales is calculated by Alibaba using data from its platforms, including Tmall.com and Taobao.com. However, data from other platforms are not included in the index. Considering the dominant share of Alibaba, the omission may not create a too large bias. The overall e-commerce development index is simply the average of the online sale index and online purchase index.

Figure 1 displays the map of the e-commerce index at the county level in 2013. While both the coastal region and western region figure prominently high, while the central region lags behind. The overall e-commerce index masks the difference between online sales and online purchase. This is probably why the spatial pattern shown in the map is not so clear cut. Figure 3 shows the map of online sales. Apparently, the coast regions, where the manufacturing sector is concentrated, score much higher in online sales than the western and central regions.[3] Figure 3 plots the spatial distribution of online purchase. Not only do residents in the eastern part of China shop heavily online, but also do those from the western region. E-commerce enables those in more remote areas to access a large selection of commodities that are not available to them otherwise. In other words, e-commerce has expanded the market to remote areas without traditional bricks and mortars.

Clustering Measure

E-commerce development is closely related to the underlying real economy. In China, most industrial production is cluster-based (Long and Zhang, 2011 and 2012). In order to examine the relationship between e-commerce development and industrial clustering, we also compute the clustering index at the county level by using China's firm registry database and following the methodology outlined in Ruan and Zhang (2015). The China State Administration of Industry and Commerce maintains the records of all the firms' registration data and the database includes firm names as well as their registered capital from the early 1980s to present. By using the number of firms or registered capital as

[3] It is no doubt that some towns in inland rural areas have boomed thanks to e-commerce (Leong, Pan, and Newell, 2015). Overall, the coastal areas have benefited more from online sales than inland areas.

weights, we were able to calculate the clustering index at the county level.

The computation of the clustering index includes four major steps. In the first step, we calculate local quotients for each industry at the county level. The local quotient measures the relative comparative advantage of a particular industry in a county compared to the whole country. A county is considered having a comparative advantage in a certain industry if its corresponding local quotient is larger than 1.

In the second step, we compute the conditional probability that an industry i has a comparative advantage in a county conditional on that another industry j reveals a comparative advantage in the same county measured in local quotients. Likewise, we can define the conditional probability of industry j on i. The smaller value of the two conditional probabilities is defined as the industry proximity measure between industry i and j. These proximity coefficients constitute a proximity matrix.

In the third step, a region's proximity for a given industry can be obtained by summing up its proximity coefficients with all other industries. Finally, we aggregate the above industrial proximity to obtain a location's clustering measure using the relative share of each industry's employment in a region or registered capital as weights.

Figure 5 maps out the clustering measure at the county level, which is calculated based on China Economic Census 2008 with employment as weight. As clearly shown in the figure, industrial clusters are concentrated in the coastal areas, while the western regions are less clustered.

Figure 6 plots the time trend of the clustering index. The clustering index at the county level is computed using the firm registry database. The figure reports the average value of the clustering measure across counties. The upper solid line is weighed by number of firms, while the lower dash line is weighted by registered capital. Both lines reveal a clear upward trend, implying that industrial production has become increasingly clustered over time.

E-commerce and Real Economy

Having shown the spatial patterns of e-commerce and clusters, we will quantitatively examine the relationship between e-commerce development and industrial clusters in this following section.

Table 1 presents the summary statistics of major variables used in the quantitative analyses. Figure 7 plots the average online sale index versus the initial degree of clustering. To filter out noises, we divided counties into 20 identically sized groups according to the initial clustering measure in 2000, which is computed based on the registration database of 2000, and then reported the average online sale index of each group. The figure shows a clear positive correlation between clustering and online sale index. The relationship is particularly strong for the top five groups and online sales dominate in the highly clustered regions.

Figure 7 provides some suggestive evidence on the positive correlation between clustering and e-commerce development. Table 2 reports the ordinary least square (OLS) estimates on the e-commerce index by controlling for more variables. The first regression (column 1) includes only the clustering index in 2000. It is highly significant at the 1% level. E-commerce was not present in China until 2003, when Taobao was established. Therefore, the measure in 2000 represents the degree of industrial clustering prior to the e-commerce era. There is no reverse causality from e-commerce in 2013 on the degree of initial clustering.

China is a large country with drastic regional differences. To account for the regional differences, province fixed effects are added in the second regression. Although the magnitude drops by half, the coefficient for the clustering measure remains highly statistically significant. In the third regression, we further control for local sex ratio among prime-age population, share of urban population, the share of service employment in total employment, share of industrial employment in total employment, average year of schooling, and the share of prime-age population. The coefficient for the clustering variable declines to 2.098, still highly significant. In the fourth regression, we further include two market access variables — minimum distance between county seat and the nearest provincial or national road; minimum distance between county seat and the nearest railway station. Neither coefficient is statistically significant. The coefficient for clustering increases slightly to 2.142. Per capita GDP is further included as a control variable in column (5). The adjusted R^2 is as high as 0.754, suggesting a good fit. With the inclusion of the additional control variable, the coefficient for the clustering measure drops to 1.17.

Despite the difference in magnitude across the five specifications, the initial degree of clustering in 2000 plays a role in shaping the spatial distribution of e-commerce in 2013. Taking the most conservative estimate in the last column as an example, a 10% increase in the mean value of clustering in 2000 would lift the e-commerce index by 0.06% (0.099*1.17/1.77).

Table 3 repeats the above excises by replacing the dependent variable with the online purchase index. Same as in Table 2, the clustering variable is significant in all the five regressions. However, its economic significance (0.02%=0.099*0.556/2.236) is much smaller than in Table 2 if using the most conservative estimate in column (5). The coefficient for the distance to the nearest railway station becomes statistically significant, suggesting that people in remote areas are more likely to shop online.

Table 4 reports the OLS estimates on online sales. The effect of clustering on online sales is much stronger than for online purchase. A 10% increase in initial clustering boosts online sales by at least 5.5% (=0.099*5.52/0.099). Interestingly, in contrast to the previous table, the coefficient for the distance to the nearest railway station is statistically significant and negative, indicating that regions with good transportation infrastructure perform better in online sales. In a word, online sales are more likely to emerge in areas with strong industrial clusters and good transportation infrastructure.

In the previous three tables, we assume that the effect of clustering on e-commerce development is the same across counties. Given the large regional variations in China, there is likely a heterogeneous effect across regions with different degrees of clustering. Table 5 reports the estimates on the heterogeneous effect for the overall e-commerce index (Panel A), online purchase index (Panel B), and online sale index (Panel C). Specifically, we classify the sample into four groups based on the initial value of clustering in 2000 and interact them separately with the clustering measure. Consistently across the three panels, it is in counties with the top quintile in terms of clustering measure that clustering enhances e-commerce development.

The cross-sectional regressions in Tables 2-4 are subject to omitted variable bias. Although we have included a large number of control variables in column 5 in each table, there is still a likelihood that some important factors are missing. To address this concern, in Tables 6-7, we look at the effect of clustering on growth in online sales and online purchase from 2013 to 2014. Unfortunately, there are some missing values for the data in 2014, decreasing the sample size by about one third. Nonetheless, the main results remain hold. The highly clustered counties have experienced a faster growth in online sales than other areas (Table 6). However, they do not score higher than less clustered areas in online purchase as shown in Table 7. Because of the smaller sample size, the results should be read with caution.

Having found that the initial degree of clustering plays a significant role in fostering online sales, we investigate if increasing online sales will further reinforce the existing advantage of clusters or not next. In the first column of Table 8, we regress the clustering measure in 2014 on online sale in 2013 by controlling for the initial clustering in 2012. The clustering measures are in logarithmic form. The coefficient for the online sale index is 1.511, highly significant at a 1% level. In column 2, we add the online purchase index in 2013. The online purchase index in 2013 is negatively associated with the clustering measure in 2014. In column 3, province fixed effects are further included. The coefficient for online purchase index becomes insignificant, while the coefficient for online sales remain highly significant. After including more control variables in columns 4-5, the main results still hold. In areas with higher online sales, the clustering measure increases faster. In counties with greater online purchases, the degree of clustering actually tends to decline.

From 2012 to 2014, the clustering measure increased by 5.4 percent. Based on the point estimation in the last column of Table 8, an increase in one standard deviation of online sale index will boost the clustering measure by 3.4 (=1.417*2.383) percentage points, which amounts to roughly 63% of the overall increase from 2012 to 2014. A rise in one standard deviation of the online purchase index is associated with 2.1 (0.412*5.078) percentage points of decline in clustering measure.

Table 9 further reports the heterogeneous effect of e-commerce development on change in clustering by interacting the online sale index and online purchase index with four dummy variables that indicate the four quintiles of clustering index in 2012. There is a U-shaped relationship between online sales and the clustering formation. The least and most clustered counties see a more rapid growth in clustering than those counties in the middle range of clustering for the same amount of online sales. For regions with the highest quantile of clustering, an increase in one standard deviation of the online sale index pushes up the clustering measure by 5.6 (=1.524*3.64) percentage points.

The effect of online purchase on cluster formation exists only in regions with lower degrees of clustering. The flood of cheaper goods from the more productive places elsewhere may squeeze the profit margin of local business and push them out of business, resulting in lower industrial clustering. Specifically, in counties sitting at the bottom 25% of the clustering measure, an increase in one standard deviation of the online purchase index leads to a decrease of the clustering measure by 3.6 (=0.815*4.39) percentage points.

The expansion or contraction of clusters should be accompanied by the increase or decrease in number of firms. As a robustness check of the findings in Table 8, we further look at the impact of e-commerce development on the growth in the entry of newly established firms in Table 10. The dependent variable in Table 10 is the share of newly established firms in total number of firms in 2014. In all the regressions in Table 10, we control for the initial value of the dependent variable in 2012. In the first column, only the online sale index is included. It is not statistically significant. In the second column, the online purchase index is added. The coefficient for online sales becomes positive and highly significant. Province fixed effects are controlled for in column 3. Population density and market access variables are further included in column 4. The ratio of industrial employment in total employment and local sex ratios are included as additional control variables in column 5. The coefficients for the online sale index and online purchase index are robust to the different specifications. In counties with higher volumes of online sales, the number of newly established firms increases. By contrast, in counties with more online purchases, the number of new entries decline.

Conclusion

E-commerce has taken off rapidly in China, reaching far away to remote markets which have not been served by bricks and mortars. Using a unique e-commerce development index at the county level and universe firm registry database, our paper shows that the expansion of market reinforces the competitive advantage of existing clusters, creating a “winners-take-all” phenomenon in terms of cluster formation. Industrial production has become more spatially concentrated along with e-commerce development.

References

Economist, 2015. “Alibaba, Clicks to Bricks: The Chinese Online Giant Is Looking for New Sources of Growth,” Aug 15th 2015. Available http://www.economist.com/news/business/21660996-chinese-online-giant-looking-new-sources-growth-clicks-bricks?zid=291&ah=906e69ad01d2ee51960100b7fa502595.

Fan, Ying, and Xiao Mo, 2015. “Reputation Premium and Reputation Management: Evidence from the Largest e-Commerce Platform in China,” available from http://www-personal.umich.edu/~yingfan/reputation_FJX.pdf.

Leong, M-L, Pan, S-L and Susan Newell, 2015. “The Emergence of Self-organizing E-commerce Ecosystem in Remote Villages of China: A Tale of Digital Empowerment for Rural Development,” *MIS Quarterly*, forthcoming.

Long, Cheryl, and Xiaobo Zhang, 2011. “Cluster-Based Industrialization in China: Financing and Performance,” *Journal of International Economics*, 84 (1): 112-123.

Long, Cheryl, and Xiaobo Zhang, 2012. Patterns of China's Industrialization: Concentration, Specialization, and Clustering,” *China Economic Review*, 23(3): 593-612.

Ruan, Jianqing and Xiaobo Zhang, 2015. “A Proximity-based Measure of Industrial Clustering.” IFPRI Discussion Paper No. 1468.

Table 1 Summary Statistics of Key Variables

Log of ecommerce development 2013	1.777	1.710	0.418	1930
Log of sale online 2013	0.0992	0.0630	1.012	1926
Log of purchase online 2013	2.336	2.291	0.385	1930
Growth of sale online 2013-2014	13.36	16.48	45.93	1394
Growth of purchase online 2013-2014	7.564	7.263	8.170	1394
100*Log Cluster 2014	-32.991	-26.904	108.619	1830
100*Log Cluster 2013	-34.672	-27.63	109.54	1830
100*Log Cluster 2012	-35.468	-29.243	108.616	1830
100*Log Cluster 2000	-59.906	-61.055	125.263	1777
Cluster 2000	0.991	0.514	1.573	1829
Log cluster index weighted by employee in 2008	-1.295	-1.099	1.596	1917
Clustering index weighted by employee in 2008	0.683	0.332	1.184	1917
Urbanization ratio	32.800	32.066	013.790	1905
Sex ratio	105.794	105.085	5.791	1905
Service ratio	10.865	9.886	5.648	1904
Industrial ratio	21.671	17.557	41.942	1901
Average year of schooling	8.112	8.350	1.124	1905
Prime Ratio	72.409	72.664	4.922	1905
Population density ($10^3/km^2$)	0.299	0.173	1.594	1903
Railroad (m/km^2)	13.594	3.050	19.217	1927
Provincial road density (m/km^2)	21.445	15.038	26.260	1927
National road density (m/km^2)	33.868	16.442	43.928	1927
Highway density (m/km^2)	57.31	28.74	89.90	1927
Market access 1 (log(km))	1.0667	1.188	1.444	1928
Market access 2 (log(km))	2.988	3.238	1.556	1928
Log of GDP 2007 (CNY)	8.93	8.894	0.659	1869

Note: Overall e-commerce development index, online purchase and online sale index are obtained from Alibaba; Sex ratio=ratio of prime-age males to females based on China Population Census 2010; Service ratio=share of employment in the service sector in total employment; Industrial ratio=share of employment in the industry sector in total employment based on China Population Census 2010; Urban ratio=share of population living in city or town based on China Population Census 2010; Market access 1 is the minimum distance between county seat and provincial or national roads and highway; Market access 2 is the minimum distance between county seat and the nearest railway station.

Table 2 The Effect of Clustering on E-commerce Development

	(1)	(2)	(3)	(4)	(5)
VARIABLES	100*Log Ecommerce Development 2013				
Cluster in 2000	13.403***	6.787***	2.098***	2.142***	1.170***
	(1.009)	(0.797)	(0.431)	(0.428)	(0.378)
Sex ratio			0.404***	0.402***	0.235**
			(0.118)	(0.119)	(0.109)
Urbanization			1.047***	1.053***	0.850***
			(0.074)	(0.074)	(0.074)
Service ratio			1.144***	1.140***	1.048***
			(0.188)	(0.188)	(0.175)
Industrial ratio			-0.015	-0.015	-0.016**
			(0.009)	(0.010)	(0.007)
Average year of schooling			6.945***	7.038***	4.774***
			(1.296)	(1.316)	(1.236)
Ratio of primate-age population			0.837**	0.840**	0.504*
			(0.392)	(0.394)	(0.301)
Population density			-0.256***	-0.248***	-0.193***
			(0.062)	(0.062)	(0.064)
Market access 1				-0.309	-0.195
				(0.409)	(0.395)
Market access 2				0.409	0.497
				(0.398)	(0.390)
Log GDP per capita 2005					14.496***
					(1.549)
Province FE	No	Yes	Yes	Yes	Yes
Observations	1,829	1,829	1,800	1,800	1,736
Adjusted R-squared	0.262	0.478	0.735	0.735	0.754
AIC	18240	17633	16118	16121	15424

Note: See Table 1 for the variable definitions. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 3 The Effect of Clustering on Online Purchase

	(1)	(2)	(3)	(4)	(5)
VARIABLES			100*Log (online purchase)		
Cluster in 2000	11.397***	6.065***	1.401***	1.538***	0.556**
	(0.832)	(0.702)	(0.354)	(0.349)	(0.309)
Sex ratio			0.556***	0.546***	0.373***
			(0.109)	(0.110)	(0.099)
Urbanization			1.047***	1.058***	0.841***
			(0.070)	(0.070)	(0.071)
Service ratio			1.138***	1.130***	1.014***
			(0.173)	(0.174)	(0.155)
Industrial ratio			-0.017*	-0.016*	-0.016**
			(0.009)	(0.009)	(0.007)
Average year of schooling			6.192***	6.506***	4.446***
			(1.229)	(1.242)	(1.093)
Ratio of primate-age population			0.959**	0.958**	0.580**
			(0.396)	(0.399)	(0.291)
Population density			-0.344***	-0.329***	-0.267**
			(0.123)	(0.113)	(0.108)
Market access 1				-0.271	-0.190
				(0.357)	(0.338)
Market access 2				0.922***	0.948***
				(0.356)	(0.346)
Log GDP per capita 2005					14.849***
					(1.447)
Province FE	No	Yes	Yes	Yes	Yes
Observations	1,829	1,829	1,800	1,800	1,736
Adjusted R-squared	0.223	0.438	0.744	0.745	0.767
AIC	18027	17462	15756	15753	15019

Note: See Table 1 for the variable definitions. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 4 The Effect of Clustering on Online Sales

	(1)	(2)	(3)	(4)	(5)
VARIABLES			100*Log (online sales)		
Cluster in 2000	29.406***	12.693***	6.960***	6.178***	5.520***
	(2.509)	(1.713)	(1.303)	(1.339)	(1.342)
Sex ratio			-1.031***	-0.947***	-1.075***
			(0.362)	(0.364)	(0.353)
Urbanization			0.763***	0.734***	0.642***
			(0.217)	(0.217)	(0.214)
Service ratio			1.466***	1.487***	1.557***
			(0.486)	(0.489)	(0.489)
Industrial ratio			-0.009	-0.012	-0.013
			(0.010)	(0.010)	(0.008)
Average year of schooling			21.928***	20.040***	16.666***
			(4.144)	(4.166)	(4.389)
Ratio of primate-age population			0.002	0.059	0.020
			(0.524)	(0.524)	(0.513)
Population density			0.714	0.668	0.650
			(0.850)	(0.775)	(0.786)
Market access 1				-1.671	-1.339
				(1.343)	(1.334)
Market access 2				-3.632***	-3.473***
				(1.265)	(1.249)
Log GDP per capita 2005					9.671**
					(4.304)
Province FE	No	Yes	Yes	Yes	Yes
Observations	1,827	1,827	1,798	1,798	1,735
Adjusted R-squared	0.216	0.440	0.510	0.512	0.531
AIC	21545	20957	20378	20370	19549

Note: See Table 1 for the variable definitions. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 5 The Heterogeneous Effect of Clustering on E-commerce Development

VARIABLES	(1)	(2)	(3)	(4)	(5)
Panel A: 100*Log (E-commerce development index)					
Cluster in 2000*bottom 25 percentile	-31.770	-33.999*	-2.028	-2.704	11.141
	(22.760)	(20.085)	(14.202)	(14.291)	(13.545)
Cluster in 2000* 25-50 percentile	13.112	3.184	0.148	0.117	0.832
	(8.112)	(6.597)	(4.469)	(4.461)	(4.325)
Cluster in 2000*50-75 percentile	15.022***	7.704**	0.763	0.793	-0.432
	(3.641)	(3.009)	(1.961)	(1.958)	(1.943)
Cluster in 2000*top 25 percentile	13.053***	6.516***	2.009***	2.048***	1.140***
	(1.158)	(0.850)	(0.450)	(0.446)	(0.392)
Panel B: 100*Log (Online purchase)					
Cluster*bottom 25 percentile	-13.606	-37.452**	-6.151	-8.073	5.885
	(20.864)	(18.340)	(12.211)	(12.251)	(11.459)
Cluster* 25-50 percentile	16.874**	5.267	2.119	2.119	2.685
	(7.375)	(6.171)	(3.998)	(3.979)	(3.852)
Cluster*50-75 percentile	12.916***	6.654**	0.006	0.134	-1.207
	(3.355)	(2.813)	(1.768)	(1.760)	(1.754)
Cluster*top 25 percentile	11.330***	5.821***	1.343***	1.473***	0.553**
	(0.980)	(0.743)	(0.371)	(0.364)	(0.321)
Panel C: 100*Log (Online sales)					
Cluster*bottom 25 percentile	-181.593***	-48.221	-7.977	1.687	11.591
	(58.890)	(49.395)	(46.008)	(46.421)	(46.259)
Cluster* 25-50 percentile	9.109	-7.763	-12.684	-13.108	-11.854
	(21.287)	(16.055)	(14.340)	(14.344)	(14.358)
Cluster*50-75 percentile	50.932***	22.948***	11.463**	10.616*	10.326*
	(9.069)	(6.639)	(5.728)	(5.749)	(5.810)
Cluster*top 25 percentile	27.573***	12.214***	6.663***	5.932***	5.375***
	(2.827)	(1.855)	(1.363)	(1.406)	(1.409)

Note: The sample is divided into four groups according to the clustering index in 2000. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 6 the Effect of Clustering on Growth in Online Sales

	(1)	(2)	(3)	(4)	(5)
VARIABLES	100*Log Sale Online 2014				
100*Log Sale Online 2013	0.787***	0.693***	0.646***	0.645***	0.653***
	(0.022)	(0.024)	(0.025)	(0.025)	(0.024)
Cluster*bottom 25 percentile	-155.468***	-93.726***	-76.910***	-74.395**	-65.706**
	(32.844)	(30.486)	(29.002)	(29.040)	(28.559)
Cluster* 25-50 percentile	-38.139***	-21.752**	-22.887**	-22.956**	-23.477**
	(11.478)	(10.351)	(9.889)	(9.934)	(9.957)
Cluster*50-75 percentile	-8.573*	-5.063	-7.465*	-7.595*	-9.037**
	(4.713)	(4.088)	(3.871)	(3.896)	(3.918)
Cluster*top 25 percentile	7.451***	4.375***	2.382***	2.224***	1.536**
	(0.981)	(0.790)	(0.693)	(0.712)	(0.693)
Sex ratio			-0.034	-0.016	-0.133
			(0.211)	(0.215)	(0.227)
Urbanization			0.373***	0.369***	0.265*
			(0.140)	(0.140)	(0.138)
Service ratio			0.786**	0.788**	0.781**
			(0.315)	(0.316)	(0.315)
Industrial ratio			0.008	0.007	0.007
			(0.011)	(0.011)	(0.009)
Average year of schooling			5.213*	4.864	4.041
			(3.091)	(3.071)	(3.129)
Ratio of primate-age population			0.924**	0.921**	0.618
			(0.462)	(0.466)	(0.464)
Population density			-0.374	-0.383	-0.350
			(0.280)	(0.268)	(0.265)
Market access 1				-0.072	0.022
				(0.846)	(0.856)
Market access 2				-1.077	-0.736
				(0.914)	(0.922)
Log GDP per capita 2005					9.310***
					(3.033)
Province FE	No	Yes	Yes	Yes	Yes
Observations	1,318	1,318	1,316	1,316	1,281
Adjusted R-squared	0.783	0.820	0.832	0.832	0.838
AIC	13818	13597	13496	13498	13106

Note: The sample is divided into four groups according to the clustering index in 2000. Due to missing values in 2014, the number of observation in this table is smaller than the cross-sectional regressions in previous tables. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 7 The Effect of Clustering on the Growth in Online Purchase

	(1)	(2)	(3)	(4)	(5)
VARIABLES	100*Log Purchase Online 2014				
100*Log Purchase Online 2013	1.231***	1.351***	1.043***	1.045***	1.019***
	(0.022)	(0.030)	(0.026)	(0.026)	(0.026)
Cluster*bottom 25 percentile	-11.812	-12.186	3.672	4.186	9.446
	(12.265)	(11.522)	(8.351)	(8.349)	(7.975)
Cluster* 25-50 percentile	-0.800	1.109	0.024	0.009	-0.791
	(4.298)	(3.716)	(2.831)	(2.837)	(2.725)
Cluster*50-75 percentile	0.326	2.315	-0.748	-0.776	-2.043*
	(1.833)	(1.493)	(1.122)	(1.125)	(1.115)
Cluster*top 25 percentile	1.492***	1.455***	0.023	-0.010	-0.565***
	(0.364)	(0.349)	(0.200)	(0.202)	(0.182)
Sex ratio			0.246***	0.250***	0.120*
			(0.079)	(0.079)	(0.071)
Urbanization			0.298***	0.296***	0.180***
			(0.052)	(0.052)	(0.050)
Service ratio			0.314***	0.312***	0.310***
			(0.106)	(0.106)	(0.098)
Industrial ratio			-0.011***	-0.012***	-0.011***
			(0.004)	(0.004)	(0.003)
Average year of schooling			8.341***	8.273***	6.880***
			(0.794)	(0.796)	(0.831)
Ratio of primate-age population			1.089***	1.088***	0.746***
			(0.125)	(0.126)	(0.120)
Population density			0.020	0.019	0.075*
			(0.047)	(0.049)	(0.044)
Market access 1				-0.019	-0.027
				(0.244)	(0.230)
Market access 2				-0.199	-0.019
				(0.241)	(0.230)
Log GDP per capita 2005					9.895***
					(1.076)
Province FE	No	Yes	Yes	Yes	Yes
Observations	1,318	1,318	1,316	1,316	1,281
Adjusted R-squared	0.793	0.857	0.917	0.917	0.926
AIC	11330	10868	10145	10148	9733

Note: The sample is divided into four groups according to the clustering index in 2000. Due to missing values in 2014, the number of observation in this table is smaller than the cross-sectional regressions in previous tables. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 8 The Impact of E-commerce on Clustering

Variables	(1)	(2)	(3)	(4)	(5)
Clustering index in 2012 (log)	0.933***	0.935***	0.942***	0.943***	0.940***
	(0.009)	(0.010)	(0.008)	(0.008)	(0.008)
Online sales in 2013	1.511***	1.826***	1.389***	1.420***	1.417***
	(0.295)	(0.344)	(0.317)	(0.313)	(0.311)
Online purchase in 2013		-0.253**	-0.174	-0.248**	-0.412***
		(0.111)	(0.115)	(0.106)	(0.157)
Population density				0.045	0.067
				(0.100)	(0.084)
Market access 1				0.007	0.153
				(0.300)	(0.299)
Industry ratio					0.001
					(0.002)
Urban ratio					0.036
					(0.052)
Sex ratio					-0.181**
					(0.072)
Average year of schooling					1.482
					(1.223)
Prime age ratio					0.021
					(0.110)
Province FE	No	No	Yes	Yes	Yes
Number of observations	1,819	1,819	1,819	1,794	1,792
Adjusted R-squared	0.971	0.971	0.977	0.978	0.978

Note: The dependent variable is clustering index in 2014 based on China firm registry database and weighted by number of firms following the methodology of Ruan and Zhang (2015). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 9 The Heterogeneous Impact of E-commerce on Clustering

Variables	(1)	(2)	(3)	(4)
Clustering index in 2012 (log)	0.906***	0.920***	0.919***	0.913***
	(0.017)	(0.014)	(0.014)	(0.015)
Online sale*bottom 25 percentile	2.109***	1.579**	1.618**	1.428**
	(0.657)	(0.741)	(0.735)	(0.703)
Online sale *25-50 percentile	0.272	0.311	0.294	0.285
	(0.635)	(0.555)	(0.553)	(0.565)
Online sale *50-75 percentile	1.392**	0.995	0.987	0.953
	(0.690)	(0.611)	(0.623)	(0.656)
Online sale *top 25 percentile	1.464***	1.377***	1.408***	1.524***
	(0.516)	(0.498)	(0.498)	(0.492)
Online purchase*bottom 25%	-0.781***	-0.484**	-0.593***	-0.815***
	(0.215)	(0.202)	(0.185)	(0.235)
Online purchase *25-50 percentile	-0.334**	-0.234*	-0.276**	-0.531***
	(0.144)	(0.139)	(0.133)	(0.178)
Online purchase *50-75 percentile	-0.119	-0.120	-0.191	-0.420***
	(0.154)	(0.138)	(0.131)	(0.162)
Online purchase *top 25 percentile	0.162	0.064	0.012	-0.223
	(0.212)	(0.171)	(0.172)	(0.194)
Population density			0.027	0.052
			(0.089)	(0.070)
Market access 1			-0.049	0.110
			(0.301)	(0.299)
Industry ratio				-0.001
				(0.002)
Urban ratio				0.038
				(0.051)
Sex ratio				-0.175**
				(0.072)
Average year of schooling				2.272*
				(1.242)
Prime age ratio				-0.023
				(0.106)
Province fixed effects	Yes	Yes	Yes	Yes
Observations	1,819	1,819	1,794	1,792
Adjusted R-squared	0.971	0.977	0.978	0.978

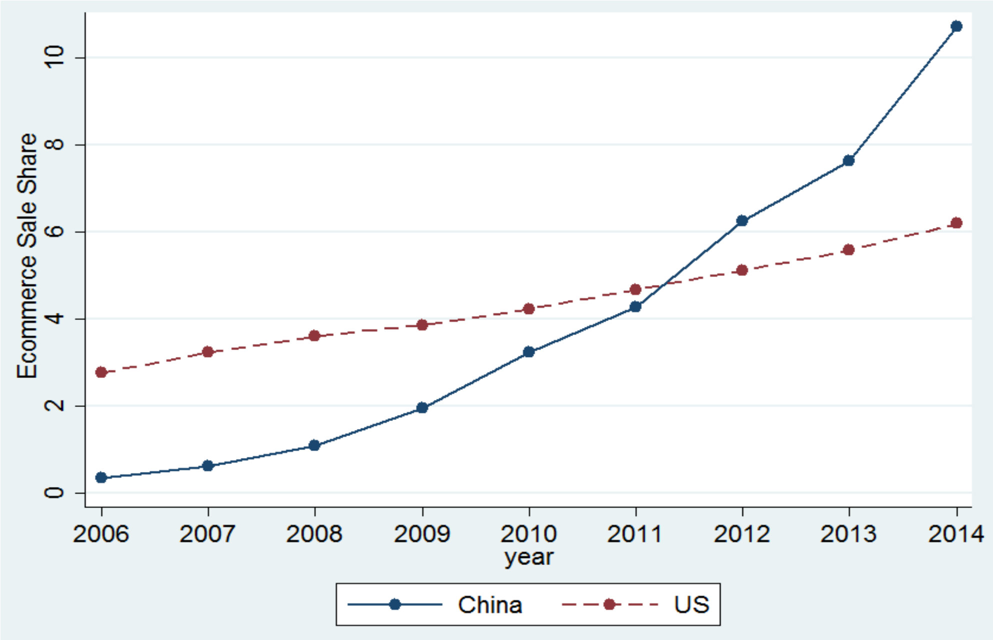
Note: The analysis is at the county level. The sample is divided into four groups based on the value of clustering index in 2012. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 10 The Effect of E-commerce on the Entry of New Manufacturing Firms

	(1)	(2)	(3)	(4)	(5)
Variables					
Firm entry ratio in 2012	0.328***	0.312***	-0.009	-0.010	-0.039
	(0.0692)	(0.0686)	(0.175)	(0.175)	(0.175)
Online sale	-0.025	0.722***	0.640***	0.669***	0.574***
	(0.163)	(0.213)	(0.212)	(0.214)	(0.220)
Online purchase		-0.541***	-0.427***	-0.453***	-0.365***
		(0.083)	(0.103)	(0.107)	(0.101)
Population density				-0.082**	-0.058*
				(0.040)	(0.000)
Market access				0.177	0.290
				(0.227)	(0.220)
Industry ratio					-0.006**
					(0.003)
Sex ratio					-0.240**
					(0.108)
Province FE	No	No	Yes	Yes	Yes
Observations	1,824	1,824	1,824	1,799	1,797
Adjusted R-squared	0.045	0.053	0.189	0.188	0.183

Note: The dependent variable is the entry ratio in 2014. The entry ratio is defined as the share of new firms in manufacturing, energy and water sectors in a particular year in total number of existing firms. Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Figure 1 The Percentage of Ecommerce Sales in Total Sales in China and the US



Data Source: the data of US comes from https://ycharts.com/indicators/ecommerce_sales_as_percent_retail_sales. The data for China is based on a series of reports of *China's E-commerce* by Ministry of Commerce People Republic of China, Press of Tsinghua University.[4]

[4] Please see the website for detail information: http://baike.baidu.com/link?url=RviFNz8JUlt1U_KbUPxdlCsAIrDe3xENPLzZdOfs9Pw1Rdmd6SMGzic3Il y7AgaGf1smmPdegxRMae2f67RquK

Figure 2 Spatial Distribution of E-commerce Development

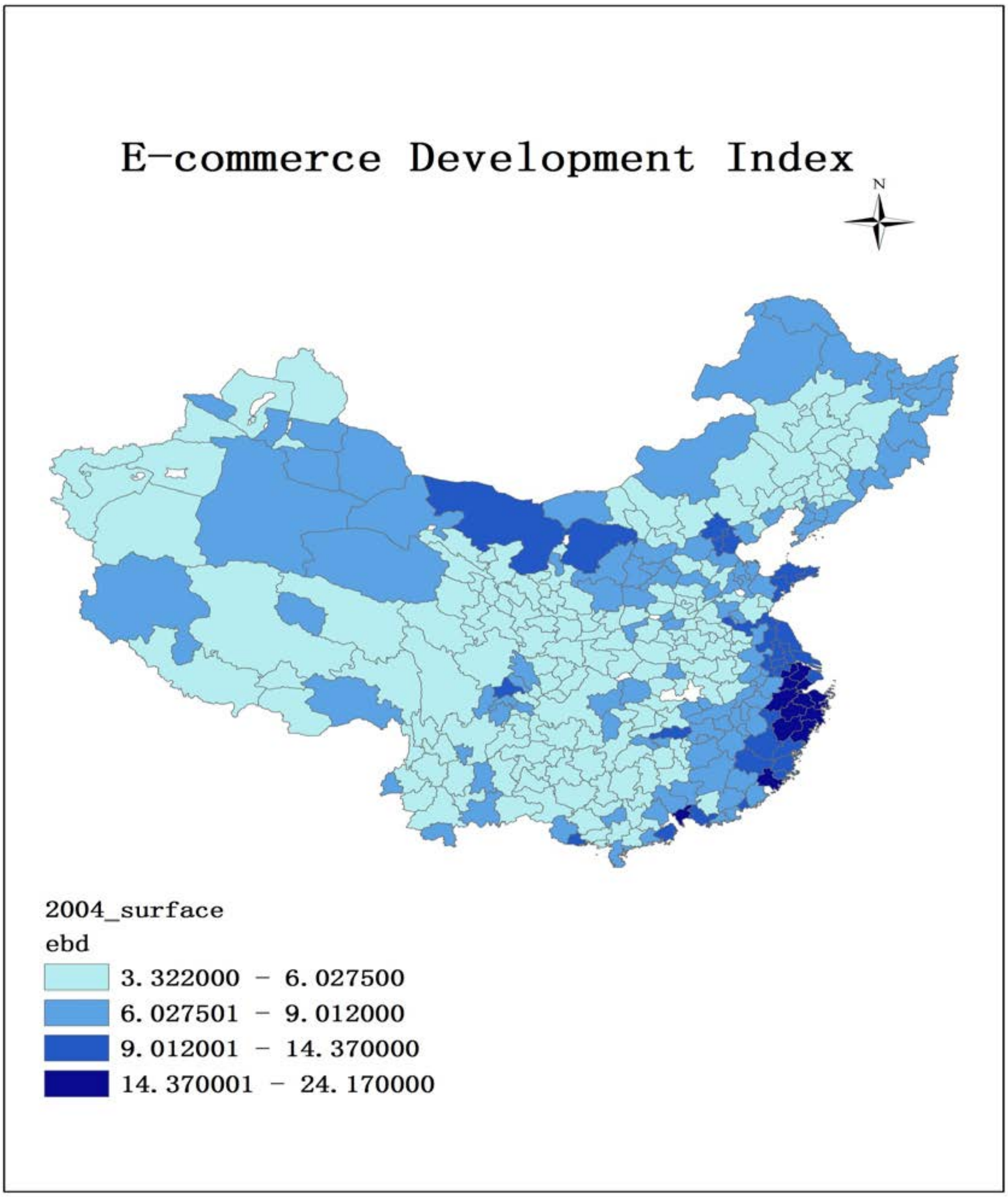


Figure 3 Spatial Distribution of Online Sales

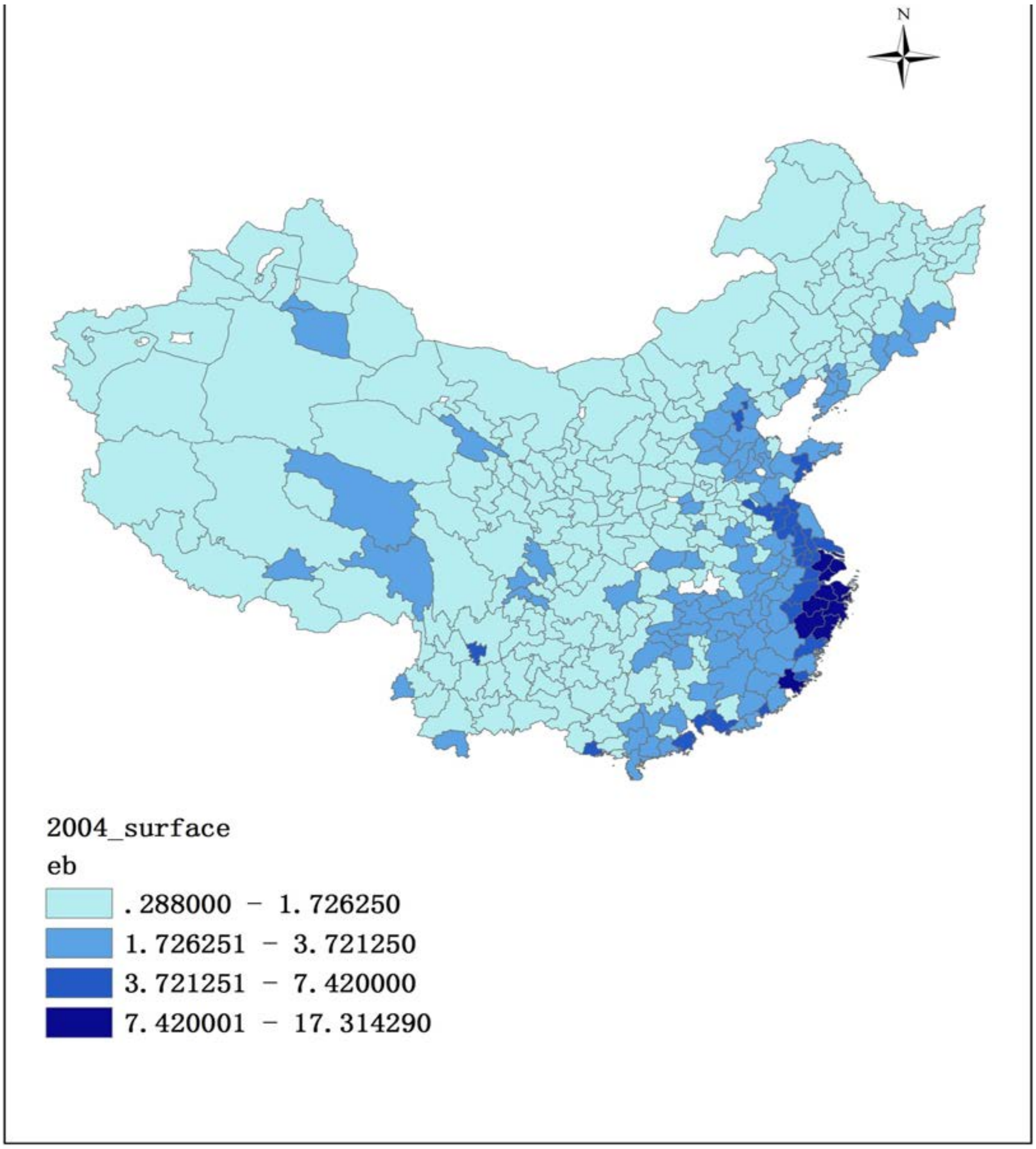


Figure 4 Spatial Distribution of Online Purchase

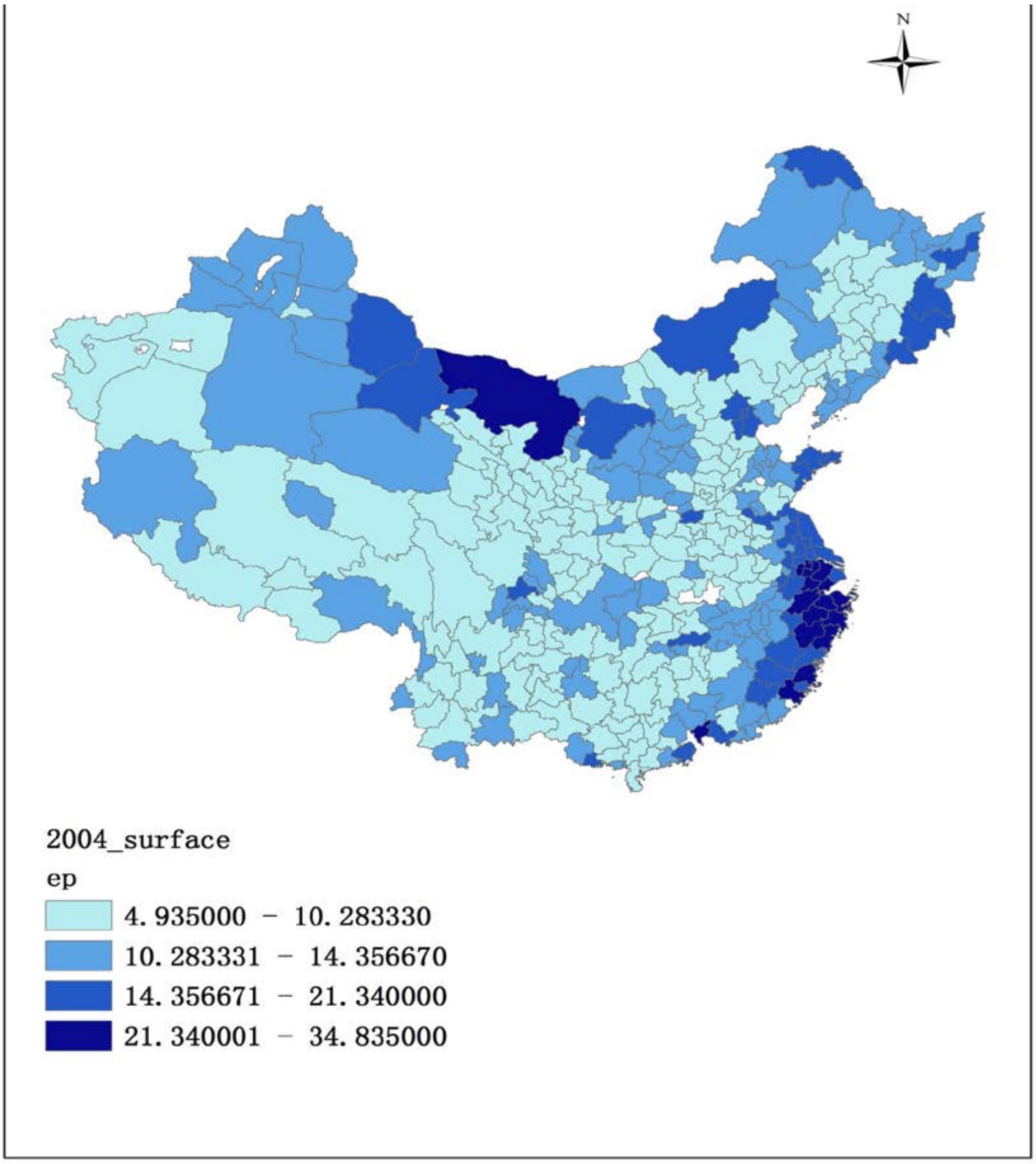
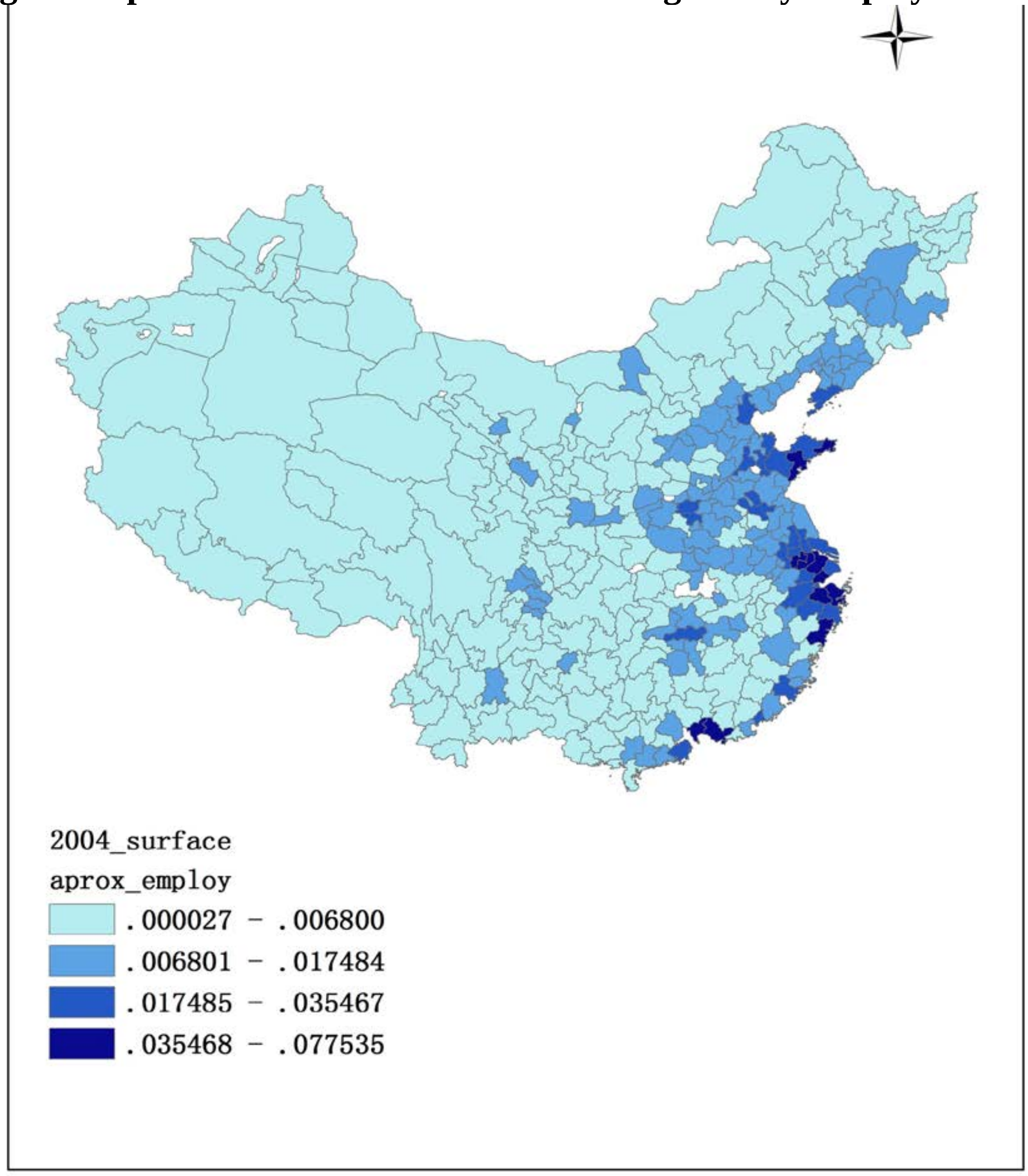
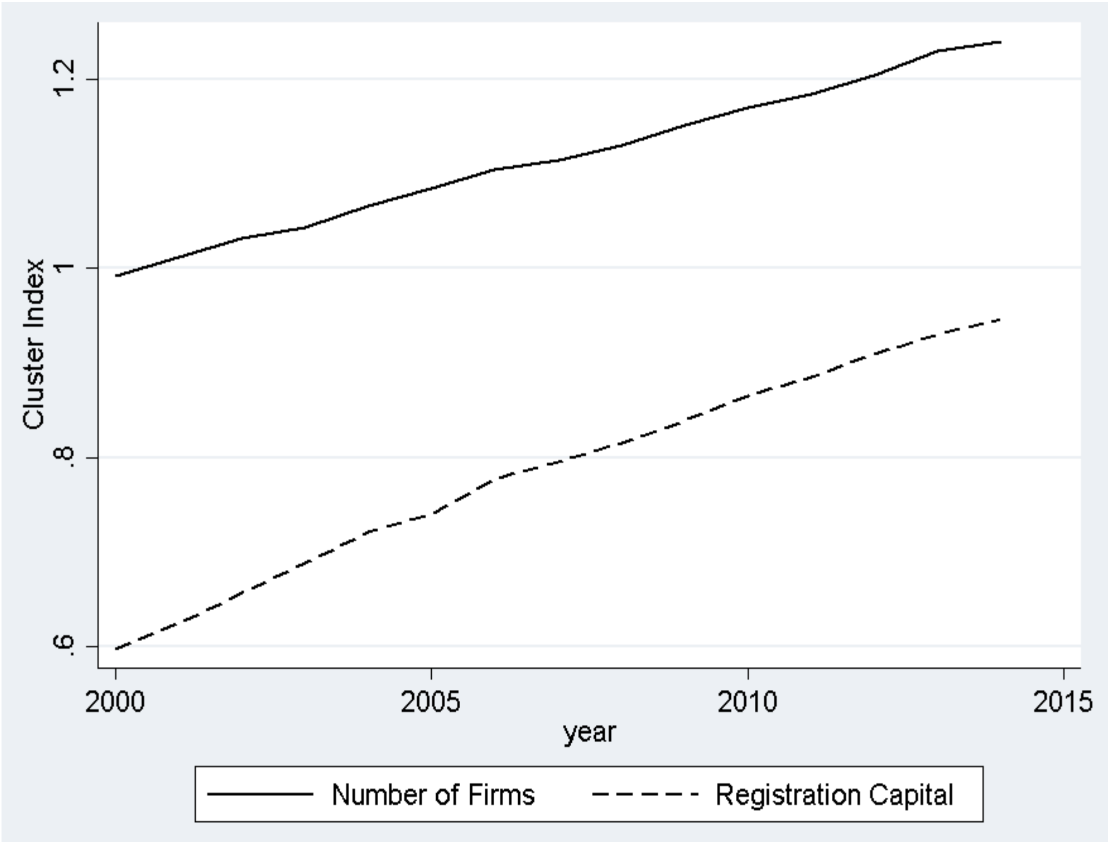


Figure 5 Spatial Distribution of Clusters Weighted by Employment



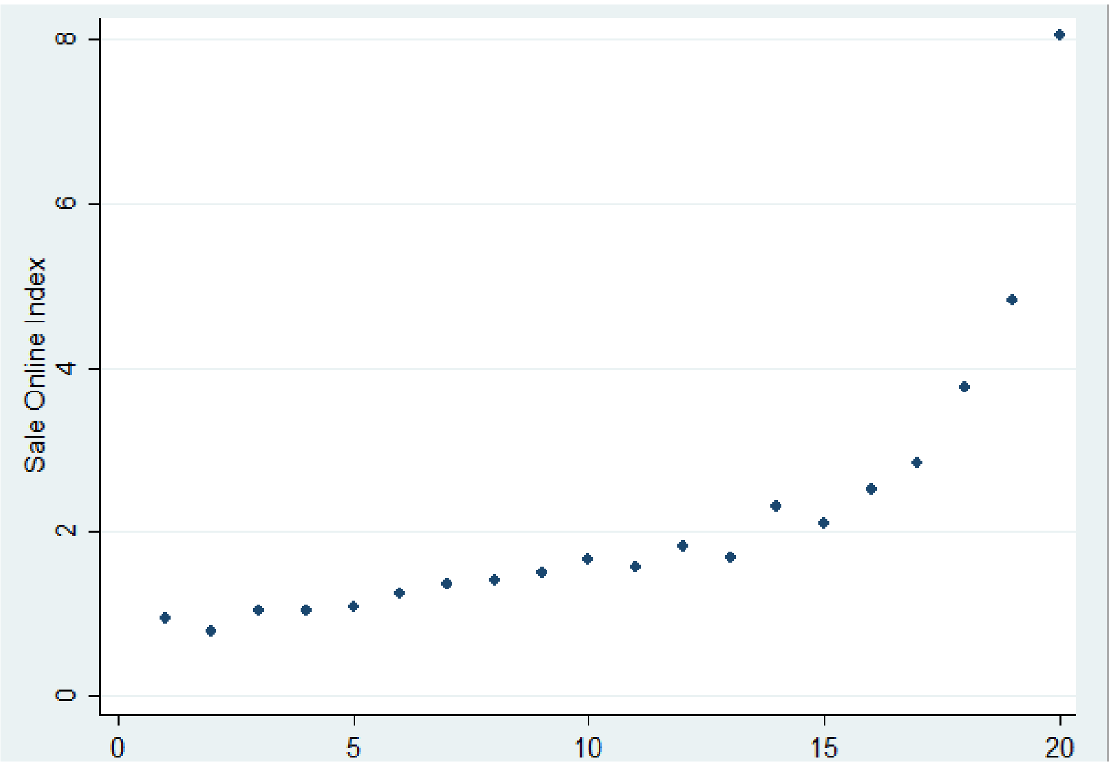
Note: the clustered index is from Ruan and Zhang (2015) which was calculated based on China Economic Census 2008.

Figure 6: Clustering Index by Year



Note: The clustering index is calculated by authors using Chinese firm registry database and following the methodology of Ruan and Zhang (2015). The figure is plotted using the subsample of counties in China which have information of ecommerce development.

Figure 7 Clustering and Online Sales



Note: The clustering measure is computed based on China Economic Census 2004. The sample is divided into 20 groups according to the value of clustering measure. The vertical line stands for the average value of online sale index corresponding to each group of clustering measure.