Journal of Rural Studies 80 (2020) 403–417

Contents lists available at ScienceDirect

Journal of Rural Studies

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

The spatial aggregation of rural e-commerce in China: An empirical investigation into Taobao Villages

Min Liu [a], Qian Zhang [b], Song Gao [a], Jikun Huang [c,d,*]

[a] College of Pastoral Agricultural Science and Technology, Lanzhou University, China
[b] College of Land Science and Technology, China Agricultural University, China
[c] China Centre for Agricultural Policy, School of Advanced Agricultural Sciences, Peking University, China
[d] College of Economics and Management, Jiangxi Agricultural University, China

ARTICLE INFO

JEL classification:
O13
O33
P25

Keywords:
Rural e-commerce
Spatial aggregation
Taobao village
Web crawler
Nighttime lights

ABSTRACT

China's rural e-commerce has been developing quickly over the last decade, and it has shown significant spatial aggregation in some areas. This paper examines this development and investigates factors that impact its spatial aggregation. The development of Taobao Villages is a typical example that reflects the fast development and spatial aggregation of rural e-commerce in villages. Thus, all the Taobao Villages that existed in China by 2017 are used as our research sample for the empirical analysis. Considering that village-level data involving long-term and large-scale observations are lacking in China, we innovatively combine report data from Alibaba, spatial data from Geography Information Systems technology, company data from web crawler technology and nighttime light data from remote sensing technology to quantify the factors of interest for each sample village. Ultimately, the consistent results of a random-effects probit model based on data about 2266 villages across six years and a negative binominal model based on cross-section data of 2092 villages demonstrates that the spatial aggregation at village-level with regard to rural e-commerce is significantly driven by the local industrial base and neighborhood effects. Interestingly, the socioeconomic conditions of the surrounding regions that villages locate in impact the spatial aggregation of rural e-commerce significantly and it presents non-linear relationship between them, but the local socioeconomic conditions of villages per se do not present significant impacts on it. This paper concludes with the policy implication for the promotion of spatial aggregation of rural e-commerce.

1. Introduction

The past few years have seen an explosion of attention paid to how information and communication technologies (ICTs) influence the global economic landscape (Wei and Liefner, 2012). One particularly thriving application is in electronic commerce (e-commerce) (Goldstein & O'connor, 2000; OCED, 2000). Using e-commerce changes the way in which goods, services, information, and knowledge are exchanged (Lal, 2004). Compared with countries such as the US and the UK, China started developing e-commerce late: in April 1998 the first online transaction in China was carried out (CINIC, 2017). China has been experiencing a rapid development of e-commerce since the 2010s (Lele and Goswami, 2017). For example, the total value of China's online sales in 2013 amounted to 1885 billion RMB (1 US$ = 6.19 RMB), which has surpassed that of the online sales of the US since then. And it continuously increased to 10.63 trillion RMB (1 US$ = 6.90 RMB) in 2019 (NBSC , 2020). China has become the world's largest e-commerce market (CECRC, 2016).

Scholars as well as policy-makers have been concerned with developing e-commerce in rural areas since it would significantly help rural residents to break the geographic boundaries of the traditional market to buy and sell goods and services (Liu et al., 2015; Aker et al., 2016; Deichmann et al., 2016). During recent years as a way not only to support the economic development of rural areas but also to decrease the rural-urban economic divide, the Chinese national government has announced a policy priority of bolstering rural e-commerce (Couture et al., 2018). The International Telecommunication Union suggested in 1984 developing telecommunications infrastructure as a way to eliminate poverty. Developed countries started implementing rural e-commerce during the 1990s, but some scholars have noted that statistical evidence on farmers' e-commerce participation is rare (e.g., Morehart and Hopkins, 2000; Mueller, 2001). Rural e-commerce in China started

* Corresponding author: Jikun Huang
E-mail addresses: minliu@lzu.edu.cn (M. Liu), qian.zhang@cau.edu.cn (Q. Zhang), gaos12@lzu.edu.cn (S. Gao), jkhuang.ccap@pku.edu.cn (J. Huang).

https://doi.org/10.1016/j.jrurstud.2020.10.016
Received 4 May 2020; Received in revised form 11 September 2020; Accepted 14 October 2020
Available online 22 October 2020
0743-0167/

booming in the 2010s, although it did not begin until the late 2000s. Remarkably, its development has shown considerable spatial aggregation in some areas (Ling, 2015; Zhu et al., 2016). To be more specific, many e-commercial companies and home-based online stores that sell the similar kinds of products have aggregated in the same areas of rural China. The aggregation of e-commercial companies and stores has further prompted the growth of local related industries like manufacturing, logistics, processing, e-commercial services and so forth. As such, some scholars have suggested that the spatial aggregation of rural e-commerce forms e-commerce clusters, which results in the integrated development of primary, secondary and tertiary industries in rural China (Liu et al., 2015; Qi et al., 2019).

In the existing literature, we found that rural e-commerce development became a hot topic discussed by scholars in developed countries around 2000 following the popularization of Information Technology in the 1990s (Baily and Lawrence, 2001; Mueller, 2001; Williams, 2001). In the 2010s, Chinese scholars have published an increasing number of research papers on rural e-commerce (e.g., Liu et al., 2015; Lin et al., 2016). Meanwhile, many scholars from other developing countries such as India have discussed the significance of developing rural e-commerce in their countries (e.g., Lele and Goswami, 2017). Not only the research time but also the research content on rural e-commerce is different when comparing developed with developing countries. How agricultural companies and agri-food supply chains adopt e-commerce for business has been the primary focal point of academic research on rural e-commerce in developed countries, but research has barely studied the adoption by individual farmers as e-commerce was rarely used by individual farmers in developed countries (e.g., Salin, 2000; Mueller, 2001; Lal, 2004). On the contrary, several studies have investigated the factors that drive rural households to adopt e-commerce as the use of e-commerce has been more and more popular among farmers to sell their products in China (Zeng and Guo, 2016; Liu et al., 2020).

Other developing countries such as India and international NGOs like the World Bank and the Asian Development Bank, have been closely monitoring (rapid) rise of adopting e-commerce by rural households in China because it is embraced as a promising mode of improving rural economy in developing countries (Lin et al., 2016; World Bank Group, 2016; Lele and Goswami, 2017); however, there is a lack of reliable references on how to expand rural e-commerce to other developing areas or countries (Zhu et al., 2016; Xiong et al., 2017). Moreover, some scholars have used case studies and qualitative observations to investigate the spatial aggregation of rural e-commerce in China (e.g., Ling, 2015). A few scholars from geographical sciences have also presented the spatial aggregation of rural e-commerce on maps; however, the existing studies have failed to apply quantitative models to analyze the causes of spatial aggregation rigorously and deeply (e.g., Zhu et al., 2016; Shan and Luo, 2017).

Therefore, this paper examines the factors that drive the development and spatial aggregation of rural e-commerce in China based on quantitative analysis, and further explores the possible approaches to expand rural e-commerce to other areas or countries. In this regard, all Taobao Villages in China are employed as our research sample since they have been experiencing typical aggregated development in regards to rural e-commerce. Thereby the specific research question is which factors drive Taobao Villages appeared earlier in some areas than in the others. It is hoped to provide scientific references on how to promote the development of rural e-commerce based on the case of the development of Taobao Villages in China. It is worth noting that this paper is innovative in its combination of report data from Alibaba, spatial data from Geography Information Systems technology, company data from web crawler technology and nighttime light data from remote sensing technology to conduct an empirical analysis and quantify the considerable factors at the village level.

In the following sections, we discuss rural e-commerce development within the context of China and introduce the Taobao Villages in section 2. Section 3 discusses the potential determinants of spatial aggregation of rural e-commerce according to the existing literature and fieldwork experience. Section 4 presents our novel methods for data collection. Following that, we present the empirical models to examine the determinants of spatial aggregation of rural e-commerce in section 5. In the subsequent section, we describe the model results and share further discussion of them. Finally, we conclude this paper with our research findings and study limitations in section 7.

2. Background

2.1. Rural e-commerce development in China

Rural e-commerce development in China has two features, which are as follows: it has developed rapidly during the recent decade; and it has shown significant spatial aggregation. In 2017, the value of retail sales of rural e-commerce was 1244.9 billion RMB (1 US$ = 6.74 RMB), an increase of 39.2% compared with the 894.5 billion RMB (1 US$ = 6.66 RMB) in sales in 2016 (CIECC, 2017, 2018). Although rural e-commerce in China only explains 15.5% of the total value of online sales in 2018, it increased faster than the average growth of China's online retail sales (CIECC, 2018). Collectively, the value of the agricultural products sold by online retailers was 158.9 billion RMB in 2016, which increased to 243.7 billion RMB in 2017 (CINIC, 2017). In particular, fruits, tea and nuts have been the top three agricultural products in online retail sales, and aquatic products, vegetables and dairy products have taken the first three places in terms of the magnitude of increase (MCC, 2018). Rural online stores are mainly located in the e-commercial platforms of Taobao, Jingdong, Pinduoduo, Yunji, Youzan, Ganjie, and so forth. Remarkably, rural e-commerce development has mainly aggregated in Eastern China and has also shown significant regional disparity. For instance, rural e-commerce's share of the retail market in Eastern China accounted for 78.7%, 12.3% in Central China, 6.8% in Western China, and 2.2% in Northeast China in 2018 (CIECC, 2018).

2.2. Policy review

The Chinese government has expended great efforts to promote e-commerce development since the 2000s. On the one hand, it improved China's Internet infrastructure through the establishment of new policies and projects. On the other hand, many policies and official documents have been issued to facilitate the advancement and supervision of e-commerce development. We learned that the first official document by the State Council of China that included the development of e-commerce was released in 1998. And the first document for developing rural e-commerce was released in 2001. In addition, we used the keywords "Internet," "Internet in rural China," "e-commerce" and "rural e-commerce" to identify relevant policies and official documents that are published on the State Council of China's website (www.gov.cn). Fig. 1 presents the number of those policy documents released by the national government in each year between 2006 and 2017. It reveals that the

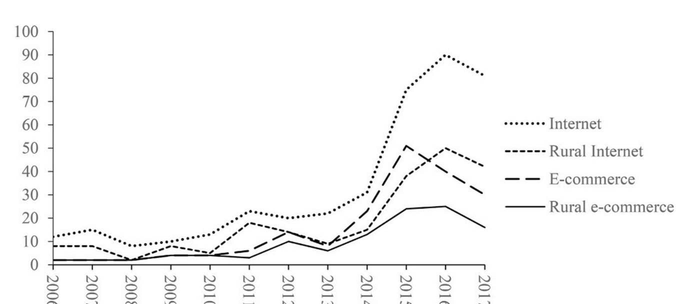

Fig. 1. The number of national policy documents on Internet services, rural Internet, e-commerce and rural e-commerce, 2006–2017. Source: Authors' compilation based on the information from the State Council of China.

national government paid increasing attention to the development of Internet services, rural Internet, e-commerce, and rural e-commerce between 2006 and 2017. The biggest shift was between 2014 and 2015.

According to the official documents released online by the Chinese national government, we reviewed the policies and projects that concern rural e-commerce development. Fig. 2 shows the yearly main concerns of the Chinese government with regard to the promotion of rural e-commerce between 2006 and 2017.

Source: Authors' compilation based on the online public database of the State Council of China, China's Ministry of Agriculture, China's Ministry of Commerce and China's Ministry of Industry and Information Technology.

Briefly, these policies and official documents are for the advancement of infrastructures and the establishment of a soft environment appropriate for rural e-commerce development. 2013 was marked with the release of the "Strategy of Broadband" by China's State Council. The focus point of this strategy lays on updating the broadband Internet and increasing Internet access throughout China, particularly in rural areas. In the following year, it formed the "poverty alleviation through e-commerce" strategy. Since then, the majority of local governments have focused on improving Internet conditions and other infrastructures that pertain to trade agricultural products, such as storing, processing, shipping, and marketing. However, to set up a soft environment is to prioritize offering training and platforms for developing e-commerce to local people and companies. For instance, local governments encourage migrant workers who previously worked in big cities to move back to their hometowns and found e-commerce businesses through offering these workers trainings or financial supports or both, which are mostly based on the e-commercial support projects or anti-poverty projects issued by the national or local governments (General Office of the State Council, 2015; Chen et al., 2019). E-commerce parks that are used to aggregate local e-commerce entrepreneurs have been developing as well. These parks are used to help the new companies obtain additional financial and technological supports for e-commerce operations.

Some national projects are intended to encourage the spatial aggregation of rural e-commerce in a particular area. For instance, "the project of One Village One Product" of China has been ongoing since 2007. The project goal is for every village to have successfully developed local distinguishable products (Qin et al., 2009). Relying on Internet to sell local products from villages has been developed during recent years, which further promoted not only the development of One Village One Product but also the spatial aggregation of rural e-commerce in some areas. Moreover, the national project of "the pilot counties of introducing e-commerce to rural areas" that focuses on funding poverty-stricken areas to develop rural e-commerce was started in 2014. All counties participating in this project has been funded with 20 million RMB to develop local e-commerce based on their local products, which had included 756 counties by 2017.

In short, rural e-commerce development can be considered to have two stages, as shown in Fig. 2. Prior to 2014, the rural e-commerce market of China was in its beginning stage. Many e-commercial companies had planned and had begun to delve into the rural e-commercial market, yet few made any exceptional investments. Local governments had in comparison fewer documents and practical support available for developing rural e-commerce. Since 2014, rural e-commerce has been in the developing stage. Both state-owned and private enterprises started to inject considerable capital into the rural e-commercial market. The Chinese national and local governments are tremendously and increasingly invested in developing rural e-commerce, especially with regard to improving local transportation, Internet and logistics conditions, building e-commercial platforms and organizing e-commercial trainings.

2.3. Taobao Villages

A Taobao Village, which aggregates e-commercial companies and stores that sell goods on the e-commercial "Taobao.com" platform in a village, is a typical outcome of the spatial aggregation of rural e-commerce at the village level. There were only 3 Taobao Villages in 2009. By 2017 there were 2118. The value of retail sales by Taobao Villages exceeded 220 billion RMB and amounted to more than 10% of the total value of retail sales for overall rural e-commerce of China in 2018 (Aliresearch, 2018).

2.3.1. The definition of Taobao Villages

As reported by Alibaba, a Taobao Village is a rurally situated village, has a total yearly turnover greater than 10 million RMB on the Taobao platform, and has either greater than 100 live online stores or more than

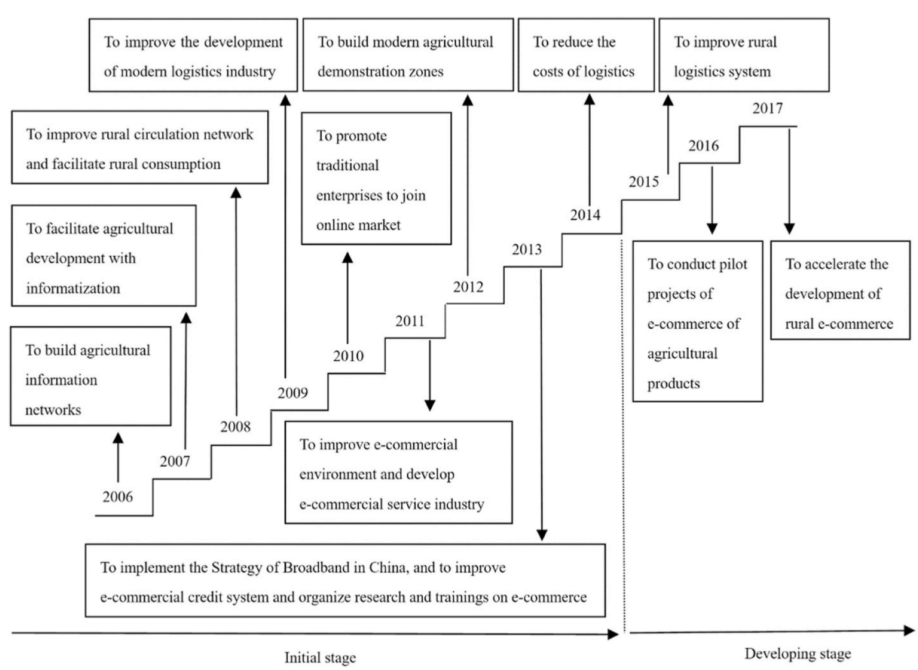

Fig. 2. The main content of the policies and official documents of China concerning rural e-commerce development, 2006–2017.

10% of its local households operating online stores (Aliresearch, 2016; Qi et al., 2019). Thus, a Taobao Village is an aggregation in a rural area of e-commerce vendors that collectively pass a certain threshold in regard to the number of online sales and stores. Based on this definition, 2009 was the year that Alibaba officially approved the first Taobao Village.

There are two types of Taobao Villages, which are distinguished by their products: agricultural and non-agricultural Taobao Villages. The former mostly sell agricultural and processed agricultural goods, while the latter mostly sell industrial products, e.g., clothes and shoes. Table 1 displays the number of existing agricultural and non-agricultural Taobao Villages during the years 2012–2017. Although the number of Taobao Villages is increasing rapidly, this number is miniscule compared with the number of villages in China. There were 554,218 administrative villages in China in 2017, but only 2118 (0.38%) of them were Taobao Villages. The number of agricultural Taobao Villages is much less, at one in 2012 and 93 in 2017. The number of agricultural Taobao Villages to Taobao Villages was a mere 4% in 2017. Moreover, there are some Taobao Villages that have dropped from the list because they no longer qualified. For example, 126 Taobao Villages left the list in 2017.

2.3.2. The formation of Taobao Villages

According to the information from our survey and existing news and reports, we found that there are two main ways to form a Taobao Village. The first way is for non-locals to migrate their e-business from an urban to a rural area. Specifically, the first wave of Taobao vendors were exclusively located in urban areas because of the availability of Internet access and convenient transportation. As the Internet penetration rate rose in rural China, Taobao vendors relocated to rural areas, because in rural areas the costs of labor and land for packing and storage are much lower than in urban areas. For instance, as the formation of the Taobao Village of Qinyanliu in Yiwu city, some Taobao sellers first moved to Qinyanliu village from Yiwu city to sell petty commodities because the cost for renting storage was much lower in the village. Their arrival prompted more Taobao sellers aggregate there, and eventually China's first Taobao Village was formed. The second way is for a few local firms and individuals to introduce e-commerce to sell their locally produced products at the beginning, and their success to stimulate their neighbors, relatives and other locals to follow them to conduct e-commerce, which was the case with the formation of the Taobao Village of Bainiu in Linan city. With the growth in the number of local e-commerce sellers, shipping and logistics industries correspondingly broadened their networks in order to include these localities, which in turn promoted more e-commerce sellers to aggregate. Thereby, a Taobao Village is gradually formed and eventually presents as an aggregation with e-commerce sellers who sell local products online and the related industries.

2.3.3. The distribution of Taobao Villages

E-commerce sellers have demonstrated spatial aggregation at the village-level. To a greater extent, rural e-commerce development has raised regional disparity. We depict the spatial distribution of all Taobao Villages and agricultural Taobao Villages in Fig. 3. Panel A demonstrates that the expansion of Taobao Villages moved very rapidly from 2012 to 2017 and reflects the regional disparity of the distribution of Taobao Villages. Most Taobao Villages aggregate in Eastern China. In 2012, only the seven provinces of Zhejiang, Guangdong, Jiangsu, Shandong, Fujian, Hebei and Jiangxi had Taobao Villages, whereas as of 2017 this group consisted of 24 provinces and municipalities. Nevertheless, there were still some provinces in Western China without Taobao Villages in 2017. For some provinces, the growth rate regarding the number of Taobao Villages is quite low and even negative. For example, in the case of the Yunnan province the number of Taobao Villages diminished from two to one between 2015 and 2016. In 2017, it remained at one.

In Panel B, we can see that there was only one agricultural Taobao Village in 2012, which was in Zhejiang province. The number of agricultural Taobao Villages also shows an obvious increase from 2012 to 2017, and most of them are in Eastern China.

More than 90% of Taobao Villages were located in Eastern China from 2012 to 2017 (see Table 2). Central China has the second highest number of Taobao Villages, and Northeast China has the least. The first Taobao Village appeared in Western China in 2014 and in Northeastern China in 2015.

Table 1
Changes in the number of Taobao Villages, 2012–2017.

Year	Taobao Villages	Agricultural Taobao Villages	Dropped Taobao Villages
2012	16	1	0
2013	20	3	0
2014	212	8	2
2015	780	40	1
2016	1311	62	37
2017	2118	93	126

Source: Authors' compilation based on the annual report by Alibaba and the China statistical yearbooks.

3. Variables

Rural e-commerce in China has formed spatial aggregations, which is new compared with e-commerce development in any other country. Taking into account our field research and the existing literature, we are aware that the local industrial base, neighborhood effects, geographic location and socioeconomic conditions have been discussed to drive the spatial aggregation of rural e-commerce in China (e.g., Zeng and Guo, 2016; Zhu et al., 2016; Diao et al., 2017; Xu et al., 2017; Zhou et al., 2017).

3.1. Local industrial base

It has been proved that companies and farmers in some rural areas of China have formed an obvious aggregation to produce the same products, which has become the local industrial base (Huang et al., 2008; Zhang and Hu, 2014; Min et al., 2017). For example, most of the farmers in Qixia county of Shandong province have planted apple trees since the 1980s, and the apple industry has been its local industrial base. Some local companies and individual households started to sell local apples through Internet around 2008. With the increase in local online stores, the aggregation of e-commerce has been gradually formed in Qixia county. It is understandable that the local industrial base indicates the experience and ability in producing and processing any agricultural or industrial products by the locals, which ensures an ample product supply for developing e-commercial industry. It is suggested that the local industrial base is crucial for the aggregation of e-commerce locally (Ling, 2015; Zeng and Guo, 2016).

3.2. Neighborhood effects

Neighborhood effects have interested many scholars. Foster and Rosenzweig (1995), for example, found that in India a household adopting new seeds is partly dependent on their neighbor's experience utilizing them. Similarly, Conley and Udry (2010) highlighted that in Ghana information from neighbors holds significant weight when it comes to the cultivation decision of pineapple farmers. It is clear that in the developing world neighborhood effects are significant when considering the livelihood and development prospects of village economies, especially in facilitating technology diffusion (Sampson et al., 2002; Yamauchi, 2007; Jackson, 2010; Magnan et al., 2015; Di Falco et al., 2018). Villagers in China maintain strong "acquaintance networks," in particular "kinship networks." As such, neighbors' and relatives' experiences have a significant impact on a household's decision to

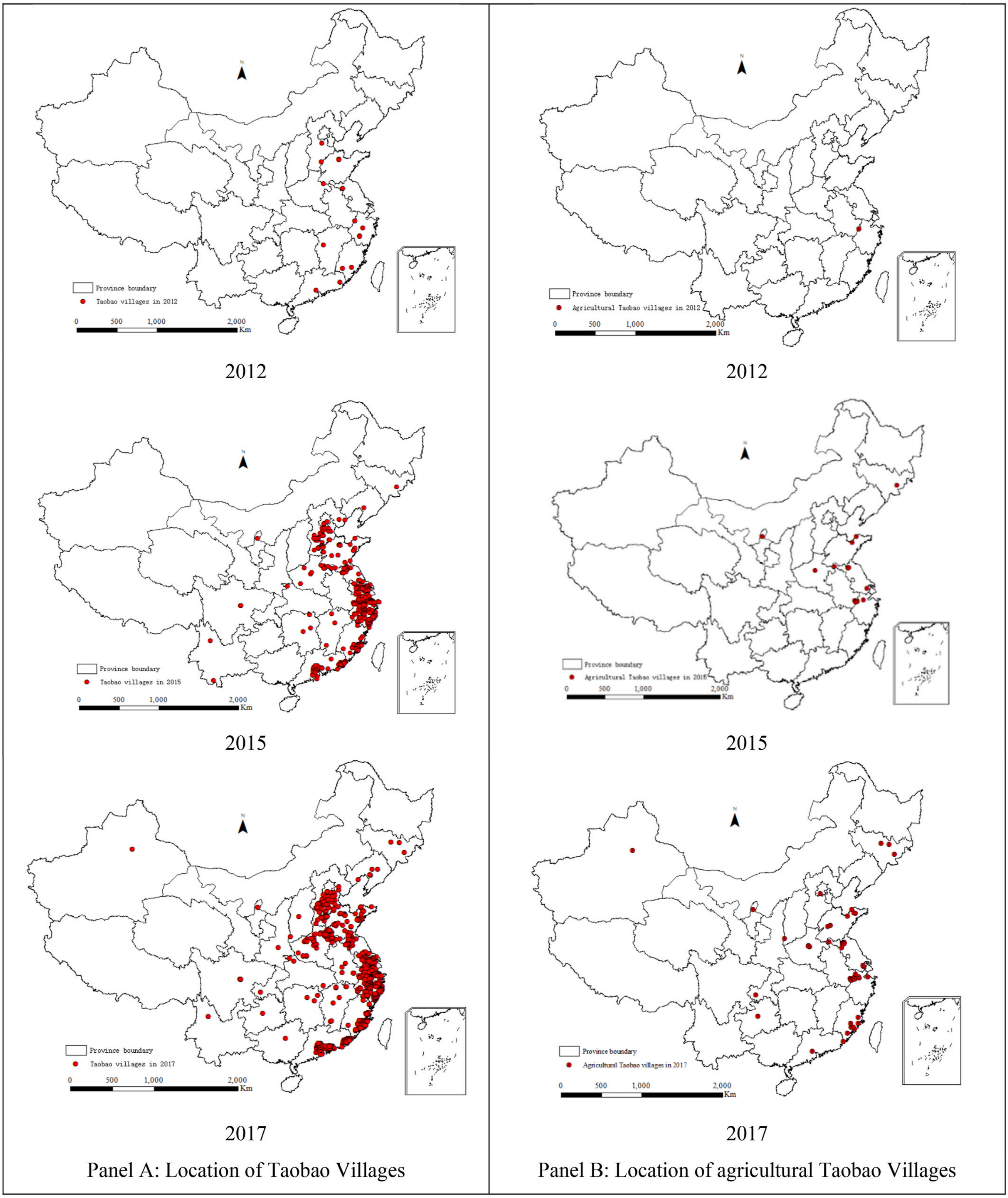

Fig. 3. The spatial distribution of Taobao Villages and agricultural Taobao Villages in 2012, 2015 and 2017. Source: Authors' compilation based on report data from Alibaba and the spatial analysis by GIS.

Table 2
The share of Taobao Villages in each region of China (%), 2012–2017.

	Eastern	Central	Western	Northeast
2012	93.8	6.2	0.0	0.0
2013	95.0	5.0	0.0	0.0
2014	98.2	0.9	0.9	0.0
2015	97.7	1.4	0.6	0.3
2016	97.7	1.5	0.4	0.4
2017	96.3	2.7	0.5	0.5

Source: Authors' compilation based on report data from Alibaba.

adopt the practice of an e-commerce vendor (Chen et al., 2014; Lin et al., 2016). In addition, Zhu et al. (2016) and Xu et al. (2017) observed that neighborhood effects on the aggregated development of e-commerce are evident among villages.

3.3. Geographic location

The geographic location of villages is a significant determinant of the aggregation of rural e-commerce because geographical advantages bring better Internet and transportation access (Xu et al., 2017). Conversely, poor geographic location has become in some cases a serious bottleneck for promoting the aggregation of rural e-commerce because of poor Internet and transportation access (Ling, 2015). It has been proved that the available Internet access is the precondition to conducting e-commerce. Logistics, which depends on local transportation access, significantly impacts the cost of developing e-commerce in rural areas (Diao et al., 2017). Hence, geographic location is considered to affect the spatial aggregation of rural e-commerce.

3.4. Socioeconomic condition

The socioeconomic condition indicates the population density, the extent of economic activity and the status of other socioeconomic indicators. Some scholars noted that there is a higher likelihood that the spatial aggregation of rural e-commerce appears in areas with larger population density and more ideal economic conditions as the consumption demand is higher there (e.g., Xu et al., 2017). Others argued that rural e-commerce aggregates in the areas that have lower costs for labor and land because the aggregation of e-commerce has large demands for both (e.g., Zeng and Guo, 2016; Diao et al., 2017). These controversial opinions will be further investigated in this study.

In brief, based on the existed observation and qualitative discussion by scholars, it is suggested that the local industrial base, neighborhood effects, and geographic location may affect the aggregation of rural e-commerce significantly and positively, but the impact of socioeconomic condition is still vague. However, the impacts of these possible determinants on the spatial aggregation of rural e-commerce have not yet been rigorously analyzed, we therefore conduct an empirical analysis for the case of the Taobao Villages.

4. Data

4.1. Data sources

All the Taobao Villages by the year of 2017 are used as our research sample. We employ Geography Information Systems (GIS) data, remote sensing technology and web crawler technology to quantify the related variables of each village, considering that long-term and large-scale data at village level is not available from the existing statistical datasets. The details on the data collection are presented below.

First, based on the reports released by Alibaba from 2012 to 2017, we collected data, including name and address, on every Taobao Village. There are 2281 existing Taobao Villages from 2012 to 2017, which includes 166 Taobao Villages that have been dropped from the list, and three villages that were Taobao Villages at the beginning, were subsequently dropped, and then reclaimed their Taobao Village designation by 2017.

Second, the GIS and remote sensing technology are used to collect spatial data. The address of each Taobao Village was used to identify its longitude and latitude coordinates. Afterward, GIS is used to measure the number of Taobao Villages within a 10-km distance from each Taobao village. We use this measurement to present the factor of neighborhood effects. And the shortest driving distance between the sample village and its township is used to present the actual geographic location of each village, which is based on the dataset of a Baidu map (similar to a Google map). The remote sensing technology is employed to measure the intensity of nighttime lights to present the socioeconomic condition.

With respect to the intensity of nighttime lights, satellites detect luminosity from outdoor and indoor use of light, fires, gas flares, and the aurora, and remote sensing technology is employed to measure the strength of light emitted at night captured by satellites (Henderson et al., 2012). It is suggested that nighttime light is the best available proxy measure for subnational economic growth (Keola et al., 2015; Bunte et al., 2018). Remotely sensed nighttime light imagery, for example, has been used to estimate socioeconomic parameters like population density, economic activity, electricity usage, as well as urbanization level since the late 1990s, and multitemporal nighttime light data can be used to present changes over time (Elvidge et al., 1997; Henderson et al., 2012; Bennett and Smith, 2017). In our case, the data of nighttime lights is based on the database of radiance measurements taken of Earth provided by the Visible Infrared Imaging Radiometer Suite sensor from NASA/NOAA's Suomi National Polar-orbiting Partnership satellite. Nighttime light intensity in areas within a 2-km, 20-km, 50-km and 100-km radii around village committees between 2012 and 2016 are measured at the 1 km by 1 km grid cell level. The lights from fires, gas flares, and the aurora are eliminated. The intensity of nighttime lights within a 2-km radius is used to present the local socioeconomic conditions of our sample villages considering the average area of villages in China, and the others are used to demonstrate the socioeconomic conditions of the different regions that each sample village locates in (Zhu et al., 2016; Diao et al., 2017).

Third, web crawler technology is used to extract information about the companies located in each sample village. We used this data to depict the local industrial base. For example, Xiaoling village in Jing county of Anhui province had 58 registered companies by 2017, and 55 of them are paper-making factories. Thus, it follows that paper-making has become the industrial base of this village. Web crawler technology is an effective tool to amass all visible information that is published on the Internet (Thelwall, 2001). This program scours websites, inspects the retrieved webpages and outputs the requested information (Sundaresan and Yi, 2000; Korfiatis et al., 2006). In this study, a web crawling program written in Python was used to assess the Baidu Enterprise Credit System. It presents the basic information of all the registered companies, including the years of their establishment, addresses, business scopes, shareholders, operating status and so forth. Based on matching addresses between our sample villages and the companies, we obtain the data on the number of companies in each village for every year. Ultimately, 2266 sample villages remained for our empirical model after 15 out of 2281 Taobao Villages were dropped as their address information was unclear or incorrect.

4.2. Data description

The statistical description on the local industrial base, neighborhood effects, geographic location and regional socioeconomic condition of our sample villages are described below.

The vertical box plots in Fig. 4 demonstrate the number of local companies in the sample villages from 2012 to 2017, which shows the level of the local industrial base for each village. The lower hinge, median, upper hinge and upper adjacent values increase year by year,

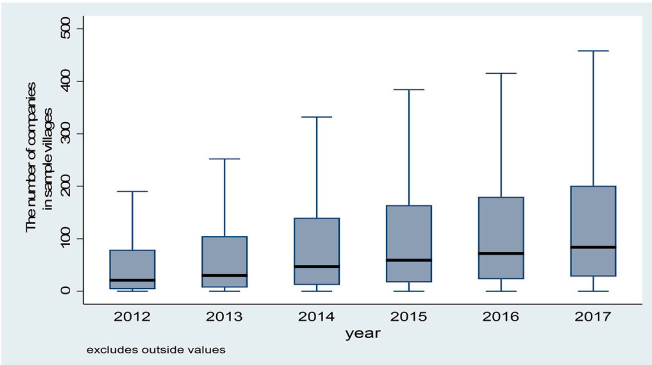

Fig. 4. The number of local companies in sample villages, 2012–2017. Source: Authors' compilation based on company data obtained by the web crawler from Baidu Enterprise Credit System.

which suggests that the number of local companies is generally increasing. The median shows that 50% of sample villages have more than 21 companies in 2012 and more than 84 in 2017.

Fig. 5 illustrates the relationship between new Taobao Villages and already existing Taobao Villages in the neighboring area, which indicates the neighborhood effects. The y-axis indicates the increased number of Taobao Villages in the areas within a 10-km radius around sample villages from year *t-1* to year *t*, and the x-axis presents the number of Taobao Villages in the same area in year *t-1*. The red line shows the general trend that more Taobao Villages would arise provided that in the previous year more Taobao Villages in the neighboring area are existed. It seemingly indicates that the neighborhood effects encourage Taobao Village development. Moreover, our data show that the average growth rate of new Taobao Villages reached 80% in the areas within a 10-km radius around the existed villages from 2016 to 2017, but the total number of Taobao Villages only grew by 62% in this period. It indicates that new Taobao Villages appeared more in the areas that have existed Taobao Villages.

Fig. 6 presents the shortest driving distances from villages to their township seats, which indicates the geographic location of our sample villages. It demonstrates that the most Taobao Villages are approximately 2.4 km away from their townships. The number of Taobao Villages is decreasing with the increase in the distances between village committees and their townships after the distance exceeds 2.4 km. Fifty percent of Taobao Villages are less than 3.5 km away from their townships, and 95% are less than 9.4 km. The farthest distance in our case is 25.7 km.

Fig. 7 presents the intensity of nighttime lights with 20-km, 50-km and 100-km radii around sample villages during 2012–2016, which demonstrates the socioeconomic conditions of the surrounding regions

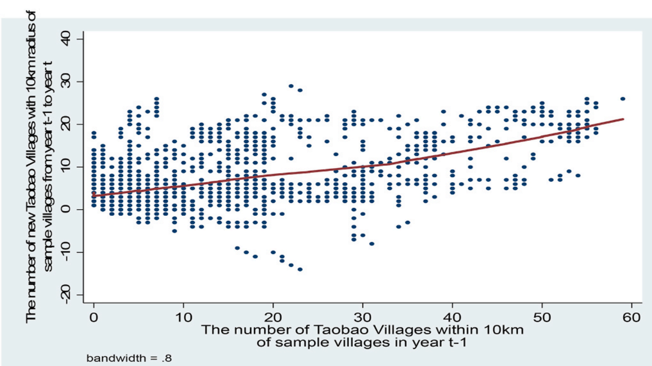

Fig. 5. The number of nearby Taobao Villages. Source: Authors' compilation based on report data from Alibaba and the spatial analysis by GIS.

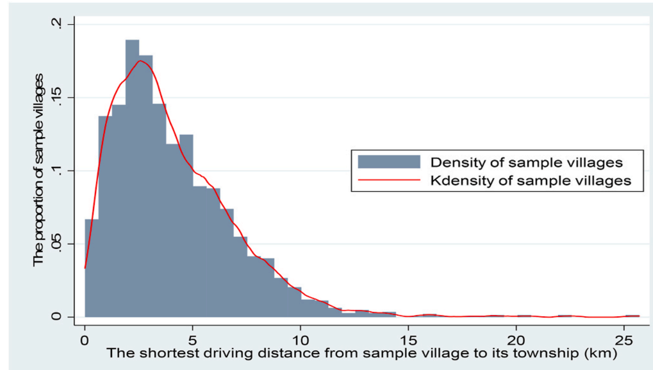

Fig. 6. The distances between sample villages and their townships. Source: Authors' compilation based on data from the Baidu map.

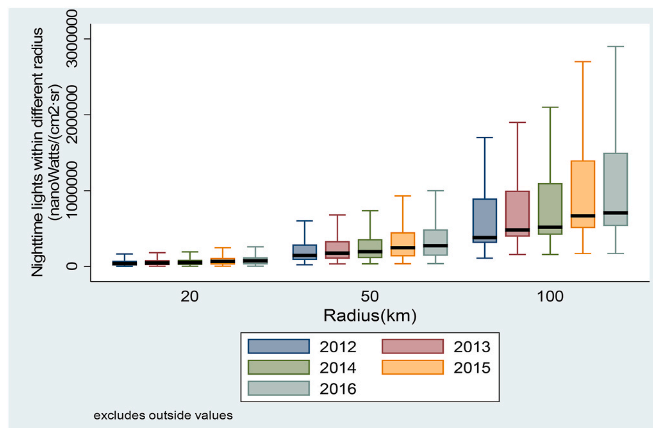

Fig. 7. The intensity of nighttime lights around sample villages, 2012–2016. Source: Authors' compilation based on the nighttime lights data collected via remote sensing technology.

that sample villages locate in. The intensity of nighttime lights within 2-km radius is too small to present in Fig. 7. The box plots show that the intensity of nighttime lights is increasing over time. This indicates that the regional socioeconomic conditions of the sample villages at different ranges have gradually improved between 2012 and 2016.

5. Empirical model

Two econometric models are employed to investigate the specific impacts of potential determinants on Taobao Village development. The first model uses a random-effects probit model that was developed on the basis of the 6-year panel data. The second uses a Poisson regression model (negative binomial model) that is based on the cross-section data. The former employs a dummy variable to present the development of Taobao Villages, and the later uses a count variable to present that. These two different models based on different types of datasets and dependent variables are used in order to examine the robustness of the model results from different perspectives.

5.1. Random-effects panel probit model

Given that the dependent variable (D_{it}) is whether village i is a Taobao Village in year t, fixed-effects and random-effects panel logit/probit models are considered (Butler and Moffitt, 1982). The fixed-effects probit model, however, presents with the practical difficulty that it raises thousands of dummy variable coefficients in the estimation (Greene et al., 2002). More importantly, the fixed-effects probit model has a methodological issue, the incidental parameters problem, which results in biased and inconsistent estimators (Greene et al., 2002; You et al., 2016). Thus, a random-effects panel probit model is employed in this study, and its validity will be further tested later. The specification of the random-effects panel probit model is as below (Guilkey and Murphy, 1993):

$$D_{it}^{*} = \lambda_i \overline{X}_{it-1} + \delta_{it} + \omega_i, \quad D_{it} = \begin{cases} 1 & if \quad D_{it}^{*} > 0 \\ 0 & if \quad D_{it}^{*} \leq 0 \end{cases} \tag{1}$$

D_{it}^{*} is a latent variable that represents the probability that village i is a Taobao Village in year t, which is determined by the observed binary variable D_{it}, which takes the value of 1 for being a Taobao Village in year t and 0 otherwise; $\overline{X}_{it-1}$ is a vector of explanatory variables. Among them are the number of local companies in year *t-1* (c_{it-1}), the intensity of local nighttime lights in year *t-1* (l_{it-1}), the number of nearby Taobao Villages in year *t-1* (n_{it-1}), the distances from villages to the seats of their townships (d_i), and the intensity of nighttime lights in year *t-1* in the surrounding regions of the villages (including 20-km($l20_{it-1}$), 50-km ($l50_{it-1}$), and 100-km radius ($l100_{it-1}$)). Moreover, the interaction term of the number of local companies (c_{it-1}) and the intensity of local nighttime lights (l_{it-1}) is considered to show the combined effects of the local industrial base and local socioeconomic conditions. The quadratic terms of ($l20_{it-1}$), ($l50_{it-1}$) and ($l100_{it-1}$) are included to present whether the impacts of regional socioeconomic conditions in the surrounding areas are non-linear. The variables for year and province that the villages affiliate with are controlled. It is noted that the major variations in government policies and development of rural economy often occurred among provinces, and especially the regional disparity of the development of rural e-commerce is most obvious at the province-level, thus we control for each province, rather than cities or counties.λ_i Is a vector of estimated parameters; δ_{it} is the individual- and time-variant error term; ω_i is the unobserved random variable that presents individual heterogeneity; and ($\delta_{it} + \omega_i$) constitute the composite error term. All variables for the random-effects panel probit model and their definitions can be seen in Table 3.

Table 3
Variables and their definitions for the random-effects panel probit model.

Variables	Definition of Variable	Obs.	Mean	Std. Dev.
D_{it}	= 1 if village *i* is a Taobao Village in year *t*, = 0 otherwise	13,596	0.33	0.47
c_{it-1}	The logarithm of the number of companies in village *i* in year *t-1*	11,330	3.60	1.78
l_{it-1}	The logarithm of the intensity of nighttime lights within a 2-km radius around the committee of village *i* in year *t-1*	11,330	6.31	1.47
c_l_{it-1}	The interaction term of c_{it-1} and l_{it-1}	11,330	23.69	14.02
n_{it-1}	The number of Taobao Villages within a 10-km radius of village *i* in year *t-1*	11,330	4.54	8.99
d_i	The shortest driving distance from village *i* to the seat of its town government	13,596	4.13	2.91
$l20_{it-1}$	The logarithm of the intensity of nighttime lights within a 20-km radius around the committee of village *i* in year *t-1*	11,330	10.71	1.00
$l20_{it-1}^2$	The quadratic term of $l20_{it-1}$	11,330	115.68	21.04
$l50_{it-1}$	The logarithm of the intensity of nighttime lights within a 50-km radius around the committee of village *i* in year *t-1*	11,330	12.28	0.80
$l50_{it-1}^2$	The quadratic term of $l50_{it-1}$	11,330	151.42	19.81
$l100_{it-1}$	The logarithm of the intensity of nighttime lights within a 100-km radius around the committee of village *i* in year *t-1*	11,330	13.44	0.67
$l100_{it-1}^2$	The quadratic term of $l100_{it-1}$	11,330	181.21	18.06
y_t	Year *t*, including 6 years from 2012 to 2017	13,596		
p_n	Province *n*, including 24 provinces	13,596		

Notes: 13,596 observations include 2266 villages and cross 6 years from 2012 to 2017. Some variables only have 11,330 observations because they were taken with a one-year lag, and thus 2266 observations were dropped for them.
Source: authors' calculations.

5.2. Poisson regression model

In addition to using a dummy variable to present the development of Taobao Villages, we employ the number of years that village *i* has been a Taobao Village by 2017 (Y_i) as the dependent variable. Y_i is a count variable and always is a positive integer. And the distribution of variable Y_i is strongly skewed to the left (see Fig. A.1 in Appendix A). The Poisson regression model is the preferred choice with respect to modeling count data (Michener and Tighe, 1992; Famoye, 1993; Cameron and Trivedi, 1998; Saphores et al., 2009). If Y_i has a Poisson distribution, the probability that $Y_i = y \geq 0$ is given by Eqn 2:

$$Pr_i(y) = \frac{exp(-\mu_i)\mu_i^y}{y!} \tag{2}$$

where μ_i is the Poisson parameter for village *i*. We assume that the natural logarithm of μ_i is a combination of various explanatory variables that are denoted by $\overline{Z}_i^{'}$ and a vector β of unknown coefficients (as shown in Eqn 3):

$$\ln(\mu_i) = \overline{Z}^{'}_i \beta. \tag{3}$$

All the related variables and their definitions are presented in Table 4.

However, the Poisson model is restrictive because the model requires the assumption that the conditional mean equals the conditional variance (hypothesis of equi-dispersion). In this regard, a Poisson model is a special case of a negative binomial model. In other words, a negative binomial model becomes a Poisson model when this particular condition is met (Agresti and Kateri, 2011). Therefore, whether a Poisson model fits the data will be further tested, otherwise a negative binomial model will be used.

6. Results

6.1. Empirical results

Table 5 shows the outcomes of the random-effects panel probit model that is estimated using the standard maximum likelihood procedure, including the regression results and their marginal effects. Three regressions including $l20_{it-1}$, $l50_{it-1}$ and $l100_{it-1}$ are alternatively conducted, which are presented as Models 1, 2 and 3, respectively. The consistent results of these three regressions, especially that the significance of coefficients of the common variables (such as c_{it-1}, l_{it-1}, c_l_{it-1}, n_{it-1} and d_i) are the same and the value is approximate, which verifies the robustness of our models (Garikipati, 2008; Liu et al., 2014). The rho of these three regressions is not near zero, which indicates that the unobserved individual heterogeneity (ω_i shown in Eqn 1) should be considered. The chibar2 (01) further presents that the null hypothesis, which states that there are no individual random effects, is rejected. This result confirms that the random-effects model is suitable. Furthermore, the Wald x^2 tests for these three regressions are significantly different from zero. In other words, all corresponding models are statistically valid.

Table 5 presents in what way the determinants impact the development of a Taobao Village based on the random-effects panel probit model. In models 1 and 2, the coefficients for the variable of the number of local companies (c_{it-1}) are significant and positive. Based on this

Table 4
Variables and their definitions for the Poisson regression model.

Variables	Variable definition	Obs.	Mean	Std. Dev.
Y_i	The number of years that village *i* has been a Taobao Village by 2017	2092	1.96	1.01
c_i	The logarithm of the number of companies at village *i* in 2012	2092	3.05	1.81
l_i	The logarithm of the intensity of nighttime lights in the area within a 2-km radius around village *i* in 2012	2092	6.00	1.52
c_l_i	The interaction term of c_i and l_i	2092	19.30	13.32
n_i	The number of Taobao Villages within a 10-km radius of village *i* in 2012	2092	0.14	0.42
d_i	The shortest driving distance from village *i* to the seat of its town government	2092	4.10	2.84
$l20_i$	The logarithm of the intensity of nighttime lights in the area within a 20-km radius around village *i* in 2012	2092	10.43	1.04
$l20_i^2$	The quadratic term of $l20_i$	2092	109.88	21.37
$l50_i$	The logarithm of the intensity of nighttime lights in the area within a 50-km radius around village *i* in 2012	2092	12.00	0.82
$l50_i^2$	The quadratic term of $l50_i$	2092	144.66	19.97
$l100_i$	The logarithm of the intensity of nighttime lights in the area within a 100-km radius around village *i* in 2012	2092	13.15	0.68
$l100_i^2$	The quadratic term of $l100_i$	2092	173.44	18.20
$\triangle dc_i$	The difference in c_i between 2012 and 2016	2092	1.11	0.90
$\triangle dl_i$	The difference in l_i between 2012 and 2016	2092	0.62	0.42
$\triangle dn_i$	The difference in n_i between 2012 and 2016	2092	13.06	14.26
$\triangle dl20_i$	The difference in $l20_i$ between 2012 and 2016	2092	0.54	0.16
$\triangle dl50_i$	The difference in $l50_i$ between 2012 and 2016	2092	0.53	0.12
$\triangle dl100_i$	The difference in $l100_i$ between 2012 and 2016	2092	0.53	0.12
p_n	Province *n*, including 24 provinces	2092		

result, we infer that the local industrial base facilitates the development of Taobao Villages, mirrors the findings of Zhu et al. (2016). Furthermore, the significant and positive coefficients of n_{it-1} demonstrate that existing Taobao Villages significantly promotes Taobao Village development in their neighboring areas. This result proves the existence of neighborhood effects in relation to Taobao Village development. Contrastingly, the coefficients of l_{it-1} are not significant, which indicates that the local socioeconomic conditions of villages do not impact Taobao Village development. However, the coefficients of $l20_{it-1}$, $l50_{it-1}$, $l100_{it-1}$ and their quadratic terms are significant, which indicates that Taobao Village development and the socioeconomic condition of the surrounding regions to which the villages belong have a non-linear relationship, as shown in Fig. 8. It is noted that we have tried a range of radii as well, from 10-km to 100-km with every 10 km interval. The model results are stable and consistent.

Fig. 8, which is drawn based on our results displayed in Table 5, presents the marginal effects of being Taobao Villages decreasing with the increase in the intensity of nighttime lights first, and then slightly increasing with the growth in the intensity of nighttime lights. This indicates that Taobao Villages appeared earlier in the areas with relatively lower nighttime light intensity that have lower population density, less economic activity and lower urbanization levels (Henderson et al., 2012). The turning points demonstrate, however, that the marginal effects of being Taobao Villages increase with the growing intensity of nighttime lights to some extent. To test the robustness of the estimation by random-effects panel probit model, we conducted the probit model on the pooled data. The results are displayed in Table A.1 of Appendix B, which are in line with the results shown in Table 5.

Since there are no non-Taobao Villages in our research sample to compare with the Taobao Villages because of data limitation, the failed Taobao villages are used for the comparison. Table 6 displays the marginal effects the determinants have on the development of a single Taobao Village, which are estimated by the random-effects panel probit model based on two subsamples. The research sample is divided into the successful Taobao-village group and the failed Taobao-village group. The failed Taobao-village group includes the samples that the villages had been Taobao Villages before 2017, but they have been dropped from the list by 2017. And the successful Taobao-village group includes all the remaining samples which were still Taobao Villages in 2017. In Table 6, the significance of the coefficients for all variables under models 4, 5 and 6 is the same with that in the regression results under models 1, 2 and 3 in Table 5, but the magnitude of the corresponding marginal effects is larger in comparison. In contrast, the coefficients for all variables under models 7, 8 and 9 are not significant. It indicates that the concerned explanatory variables present stronger impacts in the successful Taobao-village group, but they do not show any impacts in the failed Taobao-village group.

Considering the mean of $Ytbvil_i$ does not equal to its variance, the negative binomial model seems preferable than the Poisson model (Agresti and Kateri, 2011). Based on the results of likelihood-ratio test of alpha estimated by negative binomial model, the null hypothesis that alpha equals 0 is significantly rejected, which indicates that it has overdispersion, and therefore it proves that the negative binomial model is more appropriate than the Poisson regression model in our case (Famoye et al., 2004).

Table 7 shows the average marginal effects estimated by the negative binomial model with robust standard errors. We note that 174 villages are dropped from the total 2266 sample villages because their value on Y_i (the number of years that village *i* has been a Taobao Village by 2017) is not available. Specifically, they were either off the Taobao Village list between 2012 and 2017 or they became Taobao Villages before or during 2012. Therefore, we have 2092 observations remaining in the cross-section data. Six regressions are conducted, alternatively including the variable of nighttime lights within 20-km, 50-km, 100-km radii, with or without the differences of the linear variables between the years of 2012 and 2016, seen as Models 10–15. To test the robustness of the estimation by negative binominal model, we conducted the ordinary least squares (OLS) regression model on the cross-section data. The results are shown in Table A.2 of Appendix B. It presents that the model results of OLS regression are consistent with that estimated by negative binominal model.

In Table 7 it is shown that the coefficients of the variables of the number of local companies (c_i) and the Taobao Villages in the neighboring area (n_i) are significant and positive in all six regressions. Thus, we infer that the local industrial base and neighborhood effects enable Taobao Village development. The insignificant coefficients of l_i indicate that local socioeconomic conditions of villages do not impact Taobao Village development. Interestingly, the coefficients of the interactive term (c_l_i) are significantly negative, which indicates that the impact of the local industrial base is reduced with the increase in the local socioeconomic conditions of the village. The significant coefficients of the variables of the distance to the township (d_i) presents that Taobao Villages appeared earlier in the areas farther away from townships. The significant coefficients of $l20_i$, $l50_i$, $l100_i$ and their quadratic terms shows that the development of Taobao Villages does have a nonlinear relationship with the regional socioeconomic condition of villages. Moreover, the coefficients of $\triangle dc_i$ and $\triangle dn_i$ in models 13, 14 and 15 are significant and positive, which shows that the growth of the number of local companies and neighboring Taobao Villages supports Taobao Village development. In general, the results are consistent with that

Table 5
Determinants of the development of Taobao Villages: random-effects panel probit model estimation.

	Model 1		Model 2		Model 3	
	Regression results	Marginal effects	Regression results	Marginal effects	Regression results	Marginal effects
c_{it-1}	0.169*	0.028*	0.180**	0.029**	0.157	0.026
	(1.73)	(1.72)	(2.01)	(1.98)	(1.36)	(1.34)
l_{it-1}	−0.047	−0.008	−0.063	−0.010	−0.094	−0.015
	(-1.04)	(1.02)	(-1.34)	(-1.33)	(-1.30)	(-1.31)
c_l_{it-1}	−0.015	−0.002	−0.017	−0.003	−0.014	−0.002
	(-1.14)	(-1.13)	(-1.40)	(-1.38)	(-0.88)	(-0.87)
n_{it-1}	0.021***	0.003***	0.022***	0.004***	0.022***	0.004***
	(9.24)	(9.63)	(8.89)	(9.55)	(8.52)	(8.49)
d_i	0.018	0.003	0.015	0.002	0.015	0.003
	(1.58)	(1.62)	(1.29)	(1.31)	(1.44)	(1.46)
$l20_{it-1}$	−1.821*	−0.297*				
	(-1.70)	(-1.75)				
$l20^2_{it-1}$	0.081*	0.013*				
	(1.67)	(1.71)				
$l50_{it-1}$			−3.837**	−0.627**		
			(-1.96)	(-2.03)		
$l50^2_{it-1}$			0.152**	0.025**		
			(2.00)	(2.12)		
$l100_{it-1}$					−4.389***	−0.718***
					(-3.40)	(-3.69)
$l100^2_{it-1}$					0.160***	0.026***
					(3.51)	(3.82)
y_t	Yes	Yes	Yes	Yes	Yes	Yes
p_n	Yes	Yes	Yes	Yes	Yes	Yes
lnsig2u	−0.396**		−0.401**		−0.393**	
sigma_u	0.820		0.818		0.821	
Rho	0.402		0.401		0.403	
chibar2 (01)	382.97***		379.01***		384.24***	
Wald x^2	2281.53***		2278.29***		2281.63***	
Obs.	11,330		11,330		11,330	

Notes: z-statistics in parentheses. *, ** and *** denote $p < 0.10$, $p < 0.05$ and $p < 0.010$ respectively.
Source: authors' calculations.

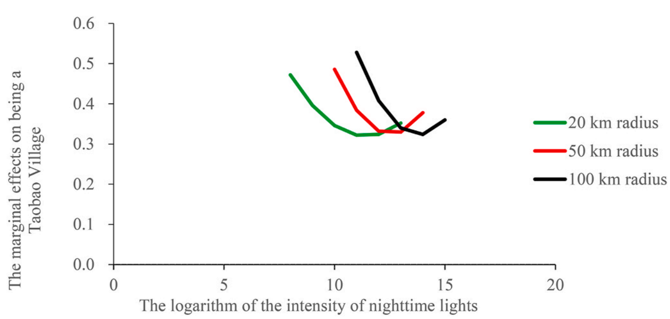

Fig. 8. The relationship between Taobao Village development and the socio-economic condition of the surrounding regions to which the villages belong. Source: authors' calculations.

shown in Table 5.

6.2. Further discussion

Based on the case of Taobao Villages in China, our empirical analysis proves that the development of rural e-commerce in China is significantly and positively driven by the level of local industrial base and neighborhood effects. The regional socioeconomic conditions around villages impact the spatial aggregation of rural e-commerce significantly, but the local socioeconomic conditions of villages do not. The spatial aggregation of rural e-commerce appeared earlier in the areas relatively farther away from townships and with less-developed regional socioeconomic conditions. This is explained by the fact that the development of rural e-commerce is partially driven by the demand in land for production and storage, and the cost of land is much lower farther from townships (Diao et al., 2017). To be more specific, the products sold by Taobao Villages mainly include agricultural products, industrial products and handcrafts, which are produced locally (Zhou et al., 2017) and require massive inputs of land and labor. Agricultural products in particular, such as nursery plants in Shuyang and pecans in Linan, have high demands for land for planting, and those Taobao Villages are not close to their townships.

On the other hand, the less-developed areas with lower labor costs and more available space for production and storage easily attract aggregations of e-commerce sellers. The areas that have high population density and good economic conditions may have other more profitable businesses to focus on, rather than developing e-commerce. It has been found that Taobao Villages have not developed in economically developed areas (Diao et al., 2017; Shan and Luo, 2017). This phenomenon is in line with the development of traditional business clusters in developing countries, such as the shoe industry in Sinos Valley, Brazil (Schmitz, 1995), the cotton knitwear industry in Tiruppur, India (Cawthorne, 1995), and the footwear industry in Wenzhou, China (Huang et al., 2008), in which clusters appear and produce their goods in less-developed and low-cost countries or areas (Qi et al., 2019).

Nevertheless, 90% of our Taobao Villages are aggregated in Eastern China, where the population density and GDP are much higher than in other regions (Cai et al., 2002). In this case, our research findings actually indicate that rural e-commerce has mainly aggregated in the relatively less-developed areas of the developed region in China. We speculate that a developed region such as Eastern China that has better Internet and transportation infrastructure supports the rapid development of rural e-commerce, but the aggregation of rural e-commerce survives in the less-developed areas that have lower costs of labor and land for production and storage.

Table 6
Determinants of Taobao Village development: random-effects panel probit model estimation on two subsamples.

	Successful Taobao-village group			Failed Taobao-village group		
	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
c_{it-1}	0.033*	0.035**	0.029	−0.005	0.004	0.003
	(1.92)	(2.23)	(1.41)	(-0.11)	(0.09)	(0.06)
l_{it-1}	−0.006	−0.010	−0.017	−0.017	−0.014	−0.165
	(-0.61)	(-0.81)	(-1.22)	(-0.64)	(-0.52)	(-0.64)
c_l_{it-1}	−0.004	−0.004*	−0.003	0.003	0.002	0.002
	(-1.54)	(-1.93)	(-1.27)	(0.45)	(0.25)	(0.30)
n_{it-1}	0.005***	0.005***	0.005**	0.003	0.002	0.002
	(2.67)	(2.57)	(2.41)	(0.97)	(0.72)	(0.57)
d_i	0.004*	0.004	0.004	0.002	0.002	0.002
	(1.69)	(1.26)	(1.36)	(0.49)	(0.54)	(0.54)
$l20_{it-1}$	−0.447**			0.266		
	(-2.14)			(0.76)		
$l20^2_{it-1}$	0.020**			−0.015		
	(2.09)			(-0.87)		
$l50_{it-1}$		−0.842**			0.176	
		(-2.77)			(0.37)	
$l50^2_{it-1}$		0.034**			−0.009	
		(2.87)			(-0.46)	
$l100_{it-1}$			−0.850**			−0.524
			(-2.21)			(-0.56)
$l100^2_{it-1}$			0.031**			0.017
			(2.11)			(0.50)
Province FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Wald x^2	234.66***	237.59***	236.61***	205.23***	205.29***	202.76***
Observations	8412	8412	8412	624	624	624

Notes: z-statistics in parentheses. ***, ** and * denote $p < 0.01$, $p < 0.05$ and $p < 0.10$ respectively. All observations from 2017 have been dropped automatically in the regressions of models 4, 5, and 6 because all of the samples in 2017 are Taobao Villages for the successful Taobao-village group and their dependent variables equal to 1. Similarly, all the observations in 2017 have been dropped automatically in models 7, 8, and 9.
Source: authors' calculations.

7. Conclusion

This paper reviewed the development of rural e-commerce in China and the national policies for developing e-commerce during the last decade. The development of rural e-commerce has been rapid and presents obvious spatial aggregation. The development of Taobao Villages is a typical example showing the spatial aggregation of rural e-commerce at the village level, and it is therefore used to investigate the determinants that promote the development and aggregation of rural e-commerce in China. All China's Taobao Villages are employed as sample villages. GIS data, remote sensing technology and web crawler technology were used to collect data from 2266 sample villages between 2012 and 2017, considering that village-level data, including long-term and large-scale observations, are lacking.

The results of a random-effects panel probit model and Poisson regression model consistently show that the number of companies in a village and the existing Taobao Villages in its neighboring area positively and significantly impact the development of Taobao Villages. This finding indicates that the local industrial base and neighborhood effects significantly drive the development of rural e-commerce. In addition, the model results show that Taobao Villages appeared earlier in the areas farther away from townships and with less-developed regional socioeconomic conditions, but in fact, more than 90% of Taobao Villages are located in developed Eastern China. This indicates that rural e-commerce has aggregated in the less-developed areas of China's most developed region. In this regard, the policy implications suggest that the priority on developing rural e-commerce, especially with regard to financial support, should focus on Central China, Western China and the Northeast considering the regional disparity in the development of rural e-commerce. The relatively less-developed areas that have a local industrial base should be involved in promoting the spatial aggregation of local rural e-commerce, and utilizing the neighborhood effects among villages is an excellent way to promote the aggregation of rural e-commerce.

In practice, the national government of China has allocated a large amount of funds to poverty-stricken counties for developing rural e-commerce since 2014, but some local governments have used the funds for developing local e-commerce unwisely. For instance, most local governments considered Internet and transportation as the primary condition for developing their e-commerce, and thus the areas near the townships and with higher population density and better economic activity were funded first to develop local e-commerce. Our research findings provide them with some scientific references on how to select optimal areas for developing local e-commerce and forming aggregation. We suggest the local governments focus on funding their less-developed areas that may be remote or exhibit poor socioeconomic conditions, but that do have or had local business or industry. In other words, local governments should consider whether the areas have existing local business or industry first, and then improve Internet access and transportation.

The ongoing poverty alleviation through developing rural e-commerce in China and the development of Taobao Villages have been followed by many other developing countries and intentional NGOs such as the World Bank (World Bank Group, 2016; Xiong et al., 2017). It is believed that the development of rural e-commerce provides opportunities for economic advancement for rural areas and can contribute to

Table 7
Determinants of the development of Taobao Villages: the marginal effects estimated by negative binominal model based on the cross-section data.

	Model 10	Model 11	Model 12	Model 13	Model 14	Model 15
c_i	0.080***	0.079***	0.074**	0.110***	0.113***	0.102***
	(2.67)	(2.66)	(2.52)	(3.63)	(3.74)	(3.41)
l_i	−0.014	−0.023	−0.034**	−0.006	−0.026	−0.038**
	(-0.76)	(-1.28)	(-1.96)	(-0.30)	(-1.43)	(-2.08)
c_l_i	−0.009*	−0.009*	−0.008*	−0.012**	−0.013***	−0.011**
	(-1.78)	(-1.84)	(-1.74)	(-2.40)	(-2.62)	(-2.38)
n_i	0.162***	0.156***	0.167***	0.077***	0.078**	0.082**
	(6.86)	(6.50)	(6.97)	(2.74)	(2.57)	(2.57)
d_i	0.009**	0.007*	0.006	0.010***	0.007*	0.005
	(2.15)	(1.68)	(1.41)	(2.59)	(1.70)	(1.42)
$l20_i$	−0.788***			−1.085***		
	(-3.45)			(-4.20)		
$l20_i^2$	0.037***			0.051***		
	(3.33)			(4.04)		
$l50_i$		−1.610***			−2.593***	
		(-4.22)			(-6.07)	
$l50_i^2$		0.067***			0.110***	
		(4.25)			(6.20)	
$l100_i$			−3.201***			−4.009***
			(-3.91)			(-4.65)
$l100_i^2$			0.122***			0.155***
			(3.97)			(4.79)
$\triangle dc_i$				0.068***	0.073***	0.069***
				(3.68)	(3.92)	(3.79)
$\triangle dl_i$				0.018	0.003	0.007
				(0.61)	(0.11)	(0.26)
$\triangle dn_i$				0.005***	0.005***	0.006***
				(5.49)	(5.23)	(6.13)
$\triangle dl20_i$				−0.00003		
				(-0.00)		
$\triangle dl50_i$					0.361**	
					(2.24)	
$\triangle dl100_i$						0.156
						(0.89)
Province FE	YES	YES	YES	YES	YES	YES
Obs.	2092	2092	2092	2092	2092	2092

Notes: z-statistics in parentheses. *, ** and *** denote $p < 0.10$, $p < 0.05$ and $p < 0.010$ respectively.
Source: authors' calculations.

upgrading the structure of the local economy in underdeveloped areas (Xiong et al., 2017; Zhu et al., 2016). Based on our research findings, we suggest targeting the areas that have local industrial bases first when developing rural e-commerce. The relatively less-developed areas in a developed region are one potential area for developing rural e-commerce. Moreover, the neighborhood effects can be used to promote the aggregation of rural e-commerce.

This paper makes several contributions. First, as a study on the development of rural e-commerce in China, we investigated its spatial aggregation based on empirical analysis and took the development of all Taobao Villages as an example, which is newly developing. Second, it is novel to combine report data and data from GIS, remote sensing technology and web crawler technology to quantify indicators at village level, which solved the problem of the lack of data involving large-scale areas and long-term observations at the village level in China. This is a methodological contribution. In particular, the Python web crawler used to mine the data for every village represents the cutting-edge in the field of agricultural economics (Zilberman, 2019). Third, the consistent results from the random-effects probit model on panel data and a Poisson regression model on cross-section data proved the robustness of our model results, which provide reliable references for understanding the development of rural e-commerce of China.

Ultimately, we would like to mention several limitations of our analysis that present challenges for future research. First, there are no non-Taobao Villages included in the research sample, because we have no information on non-Taobao Villages and selecting the corresponding non-Taobao Villages to match with our Taobao Villages limits our focus to areas that have already developed rural e-commerce. Second, we did not answer the question regarding what level of socioeconomic conditions (such as the extent of population density and GDP) are advantageous for developing rural e-commerce, because the specific numerical relationship between the intensity of nighttime lights and socioeconomic conditions is not investigated in this paper. Third, we used the number of companies to reflect the local industrial base generally, but information about companies' scales and their main products is lacking to interpret the impacts from different sizes and types of local industries on the development of rural e-commerce. Fourth, the impacts of policies on developing rural e-commerce are not specified because we lack the data on local policies, which are assumed to be controlled by the dummy variables of each year and each province.

CRediT authorship contribution statement

Min Liu: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Writing - original draft. **Qian Zhang:** Data curation. **Song Gao:** Data curation. **Jikun Huang:** Conceptualization, Writing - review & editing, Project administration.

Acknowledgements

The authors would like to thank Dr. Shi Shu, Dr. Jie Zhou, Yueying Bai, Xiya Zhao, Yaqi Hu and Yuling Li for their great help in data collection, and Dr. Wanglin Ma, Dr. Hongdong Guo, Dr. Fu (Jeff) Jia, Dr. Qihui Chen, Dr. Shi Min, Dr. Hongmei Yi, Dr. Xiaomeng Cui and Dr. Jiaqi Qi for their constructive suggestions and comments during this manuscript preparation stage. The authors also appreciate the committee of 2018 CAER-IFPRI Annual International Conference for selecting this manuscript as the best conference paper.

Appendices.

Appendix A

Fig. A.1. Fig. A.1The distribution of the dependent variable (Y_i). Source: authors' calculations.

Appendix B

Table A.1
Determinants of the development of Taobao Villages: Probit model estimation based on the pooled data

	Model A		Model B		Model C	
	Regression results	Marginal effects	Regression results	Marginal effects	Regression results	Marginal effects
c_{it-1}	0.158***	0.033***	0.169***	0.035***	0.152***	0.032***
	(3.46)	(3.46)	(3.74)	(3.75)	(3.39)	(3.39)
l_{it-1}	−0.023	−0.005	−0.038	−0.008	−0.062**	−0.013**
	(-0.79)	(-0.79)	(-1.32)	(-1.32)	(-2.24)	(-2.24)
c_l_{it-1}	−0.015**	−0.003**	−0.017**	−0.004**	−0.015**	−0.003**
	(-2.16)	(-2.16)	(-2.50)	(-2.50)	(-2.17)	(-2.17)
n_{it-1}	0.025***	0.005***	0.026***	0.005***	0.026***	0.005***
	(9.96)	(10.04)	(10.15)	(10.23)	(9.94)	(10.02)
d_i	0.015***	0.003***	0.012**	0.003**	0.012**	0.003**
	(2.75)	(2.76)	(2.16)	(2.16)	(2.17)	(2.17)
$l20_{it-1}$	−1.453***	−0.304***				
	(-3.85)	(-3.85)				
$l20^2_{it-1}$	0.064***	0.013***				
	(3.60)	(3.60)				
$l50_{it-1}$			−3.333***	−0.697***		
			(-5.48)	(-5.50)		
$l50^2_{it-1}$			0.133***	0.028***		
			(5.42)	(5.44)		
$l100_{it-1}$					−4.538***	−0.950***
					(-3.80)	(-3.81)
$l100^2_{it-1}$					0.166***	0.035***
					(3.80)	(3.81)
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	11,330		11,330		11,330	

Notes: z-statistics in parentheses. *, ** and *** denote $p < 0.10$, $p < 0.05$ and $p < 0.010$ respectively.
Source: authors' calculations.

Table A.2
Determinants of the development of Taobao Villages: OLS estimation based on the cross-section data

	Model a	Model b	Model c	Model d	Model e	Model f
c_i	0.168***	0.168***	0.158***	0.226***	0.234***	0.215***
	(2.84)	(2.85)	(2.71)	(3.77)	(3.96)	(3.66)
l_i	−0.026	−0.042	−0.064*	−0.010	−0.045	−0.065*
	(-0.75)	(-1.25)	(-1.96)	(-0.27)	(-1.31)	(-1.96)
c_l_i	−0.019**	−0.019**	−0.018**	−0.025***	−0.027***	−0.025***
	(-1.98)	(-2.06)	(-1.96)	(-2.65)	(-2.91)	(-2.70)
n_i	0.352***	0.343***	0.363***	0.181***	0.182***	0.183***
	(6.59)	(6.43)	(6.79)	(2.98)	(2.92)	(2.85)
d_i	0.017**	0.014*	0.012	0.020**	0.013*	0.011
	(2.21)	(1.73)	(1.48)	(2.58)	(1.71)	(1.46)
$l20_i$	−1.543***			−2.023***		
	(-3.22)			(-4.04)		
$l20_i^2$	0.073***			0.096***		
	(3.12)			(3.90)		
$l50_i$		−3.394***			−5.065***	
		(-4.23)			(-6.02)	
$l50_i^2$		0.141***			0.214***	
		(4.25)			(6.12)	
$l100_i$			−6.395***			−7.720***
			(-3.87)			(-4.69)
$l100_i^2$			0.243***			0.298***
			(3.92)			(4.80)
Δdc_i				0.126***	0.134***	0.129***
				(3.81)	(4.05)	(3.92)
Δdl_i				0.035	0.008	0.019
				(0.59)	(0.14)	(0.33)
Δdn_i				0.010***	0.010***	0.011***
				(5.11)	(5.02)	(5.81)
$\Delta dl20_i$				−0.007		
				(-0.04)		
$\Delta dl50_i$					0.623**	
					(2.13)	
$\Delta dl100_i$						0.217
						(0.65)
Province FE	YES	YES	YES	YES	YES	YES
Obs.	2092	2092	2092	2092	2092	2092

Notes: t-statistics in parentheses. *, ** and *** denote $p < 0.10$, $p < 0.05$ and $p < 0.010$ respectively.
Source: authors' calculations.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jrurstud.2020.10.016.

Funding

This work was supported by the Chinese Postdoctoral committee [grant number 8201400849]; and the Asian Development Bank [grant number 2017 × 127. CCA].

References

Agresti, A., Kateri, M., 2011. Categorical Data Analysis. Springer Berlin Heidelberg, pp. 206–208.
Aker, J.C., Ghosh, I., Burrell, J., 2016. The promise (and pitfalls) of ICT for agriculture initiatives. Agric. Econ. 47 (S1), 35–48.
Aliresearch, 2016. The Research Report on Taobao Village of China in 2016.
Aliresearch, 2018. The Research Report on Taobao Villages of China in 2018.
Baily, M.N., Lawrence, R.Z., 2001. Do we have a new e-conomy? Am. Econ. Rev. 91 (2), 308–312.
Bennett, M.M., Smith, L.C., 2017. Advances in using multitemporal night-time lights satellite imagery to detect, estimate, and monitor socioeconomic dynamics. Rem. Sens. Environ. 192, 176–197.
Bunte, J.B., Desai, H., Gbala, K., Parks, B., Runfola, D.M., 2018. Natural resource sector FDI, government policy, and economic growth: quasi-experimental evidence from Liberia. World Dev. 107, 151–162.
Butler, J.S., Moffitt, R., 1982. A computationally efficient quadrature procedure for the one-factor multinomial probit model. Econometrica 50 (3), 761–764.
Cai, F., Wang, D., Du, Y., 2002. Regional disparity and economic growth in China: the impact of labor market distortions. China Econ. Rev. 13 (2), 197–212.
Cameron, A.C., Trivedi, P.K., 1998. Regression Analysis of Count Data. Cambridge University Press, New York, NY.
Cawthorne, P.M., 1995. Of networks and markets: the rise and rise of a South Indian town, the example of Tiruppur's cotton knitwear industry. World Dev. 23 (1), 43–56.
CECRC, 2016. The annual monitoring report of online retail sales of China. Retrieved from. http://www.360doc.com/content/17/0604/10/43746527_659751979.shtml, 8th of May, 2019.
Chen, X.X., Wang, X.Y., Liu, J.X., 2019. The impact of E-business poverty alleviation on decision-making and development of rural enterprises in the perspective of game theory. On Econ. Prob. 1, 84–91.
Chen, X., Zhang, L., Li, W., 2014. A network evolution model for Chinese traditional acquaintance networks. IEEE Intell. Syst. 29 (5), 5–13.
CIECC, 2017. The report on the development of rural e-commerce of China (2016-2017). Retrieved from. http://www.100ec.cn/detail–6419058.html, 8th of May, 2019.
CIECC, 2018. The report on the development of rural e-commerce of China (2017-2018). Retrieved from. https://www.sohu.com/a/259675536_99947052, 8th of May, 2019.
CINIC, 2017. The 40th China statistical report on internet development. Retrieved from. http://www.cac.gov.cn/2017-08/04/c_1121427672.htm, 8th of May, 2019.
Conley, T.G., Udry, C.R., 2010. Learning about a new technology: pineapple in Ghana. Am. Econ. Rev. 100 (1), 35–69.
Couture, V., Faber, B., Gu, Y., Liu, L., 2018. E-commerce integration and economic development: evidence from China (No. w24384). Natl. Bur. Econ. Res. Retrieved from http://www.nber.org/papers/w24384.pdf 8th of May, 2019.
Deichmann, U., Goyal, A., Mishra, D., 2016. Will digital technologies transform agriculture in developing countries? Agric. Econ. 47 (S1), 21–33.
Di Falco, S., Feri, F., Pin, P., Vollenweider, X., 2018. Ties that bind: network redistributive pressure and economic decisions in village economies. J. Dev. Econ. 131, 123–131.

Diao, B., Chen, K., Ding, L., Zeng, K., 2017. Spatial pattern and influence factors of Taobao Village. Trop. Geogr. 37, 56–65.
Elvidge, C.D., Baugh, K.E., Kihn, E.A., Kroehl, H.W., Davis, E.R., Davis, C.W., 1997. Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. Int. J. Rem. Sens. 18 (6), 1373–1379.
Famoye, F., 1993. Restricted generalized Poisson regression model. Commun. Stat. Theor. Methods 22 (5), 1335–1354.
Famoye, F., Wulu, J.T., Singh, K.P., 2004. On the generalized Poisson regression model with an application to accident data. J. Data Sci. 2 (2004), 287–295.
Foster, A.D., Rosenzweig, M.R., 1995. Learning by doing and learning from others: human capital and technical change in agriculture. J. Polit. Econ. 103 (6), 1176–1209.
Garikipati, S., 2008. The impact of lending to women on household vulnerability and women's empowerment: evidence from India. World Dev. 36 (12), 2620–2642.
General Office of the State Council, 2015. Opinions by the general office of the State Council on supporting the migrant workers to return their hometowns to start businesses. Retrieved from. http://www.gov.cn/zhengce/content/2015-06/21/content_9960.htm, 8th of May, 2019.
Goldstein, A., O'connor, D., 2000. E-commerce for development: prospects and policy issues. Working paper. OECD Development centre, Paris, 164.
Greene, W., Han, C., Schmidt, P., 2002. The bias of the fixed effects estimator in nonlinear models. In: Unpublished Manuscript, vol. 29. New York University, Stern School of Business, New York.
Guilkey, D.K., Murphy, J.L., 1993. Estimation and testing in the random effects probit model. J. Econom. 59 (3), 301–317.
Henderson, J.V., Storeygard, A., Weil, D.N., 2012. Measuring economic growth from outer space. Am. Econ. Rev. 102 (2), 994–1028.
Huang, Z., Zhang, X., Zhu, Y., 2008. The role of clustering in rural industrialization: a case study of the footwear industry in Wenzhou. China Econ. Rev. 19 (3), 409–420.
Jackson, M.O., 2010. Social and Economic Networks. Princeton University Press, Princeton, NJ.
Keola, S., Andersson, M., Hall, O., 2015. Monitoring economic development from space: using nighttime light and land cover data to measure economic growth. World Dev. 66, 322–334.
Korfiatis, N.T., Poulos, M., Bokos, G., 2006. Evaluating authoritative sources using social networks: an insight from Wikipedia. Online Inf. Rev. 30 (3), 252–262.
Lal, K., 2004. E-business and export behavior: evidence from Indian firms. World Dev. 32 (3), 505–517.
Lele, U., Goswami, S., 2017. The fourth industrial revolution, agricultural and rural innovation, and implications for public policy and investments: a case of India. Agric. Econ. 48 (S1), 87–100.
Lin, G., Xie, X., Lv, Z., 2016. Taobao practices, everyday life and emerging hybrid rurality in contemporary China. J. Rural Stud. 47, 514–523.
Ling, S.X., 2015. Research on forming and evolving mechanism of E-commerce industrial cluster in China's rural Area. Commer. Res. 453 (1), 104–109.
Liu, C., Li, J., Liu, J., 2015. Rural E-commerce and new model of rural development in China: a comparative study of" Taobao Village" in jiangsu province. Asian Agric. Res. 7 (11), 35–46.
Liu, M., Shi, P.F., Wang, J.X., Wang, H.M., Huang, J.K., 2020. Do farmers earn more from selling their agricultural products through e-commerce? (Working paper).
Liu, Q., Lu, R., Zhang, C., 2014. Entrepreneurship and spillovers from multinationals: evidence from Chinese private firms. China Econ. Rev. 29, 95–106.
Magnan, N., Spielman, D.J., Lybbert, T.J., Gulati, K., 2015. Leveling with friends: social networks and Indian farmers' demand for a technology with heterogeneous benefits. J. Dev. Econ. 116, 223–251.
MCC, 2018. The report on China's e-commerce in 2017. Retrieved from. http://baijiahao.baidu.com/s?id=1602348460992344469&wfr=spider&for=pc.
Michener, R., Tighe, C., 1992. A Poisson regression model of highway fatalities. Am. Econ. Rev. 82 (2), 452–456.
Min, S., Huang, J., Bai, J., Waibel, H., 2017. Adoption of intercropping among smallholder rubber farmers in Xishuangbanna, China. Int. J. Agric. Sustain. 15 (3), 223–237.
Morehart, M., Hopkins, J., 2000. On the upswing: online buying & selling of crop inputs & livestock. US Department of Agriculture. ERS Agricultural Outlook, Washington, DC, pp. 1–4.
Mueller, R.A.E., 2001. E-commerce and entrepreneurship in agricultural markets. Am. J. Agric. Econ. 83 (5), 1243–1249.
NBSC (National Bureau Of Statistics of China), 2020. China Statistical Yearbook.
OCED, 2000. Information Technology Outlook 2000: ICTs, E-Commerce and the Information Economy. Organisation for Economic Co-operation and Development, Paris.
Qi, J., Zheng, X., Guo, H., 2019. the formation of Taobao villages in China. China Econ. Rev. 53, 106–127.
Qin, F., Zhong, Y., Zhang, M., Wang, Q., 2009. Thinking of the development of "one village one product" in China. Iss. Agric. Econ. 8, 4–8.
Salin, V., 2000. Information technology and cattle-beef supply chains. Am. J. Agric. Econ. 82 (5), 1105–1111.
Sampson, R.J., Morenoff, J.D., Gannon-Rowley, T., 2002. Assessing "neighborhood effects": social processes and new directions in research. Annu. Rev. Sociol. 28, 443–478.
Saphores, J.-D.M., Nixon, H., Ogunseitan, O.A., Shapiro, A.A., 2009. How much e-waste is there in US basements and attics? Results from a national survey. J. Environ. Manag. 90 (11), 3322–3331.
Schmitz, H., 1995. Collective efficiency: growth path for small-scale industry. J. Dev. Stud. 31 (4), 529–566.
Shan, J.S., Luo, Z.D., 2017. Agglomeration and fission: spatial distribution characteristics and evolution trends of Taobao Villages and towns. Shanghai Urban Plan. Rev. 2, 98–104.
Sundaresan, N., Yi, J., 2000. Mining the web for relations. Comput. Network. 33 (1), 699–711.
Thelwall, M., 2001. A web crawler design for data mining. J. Inf. Sci. 27 (5), 319–325.
Wei, Y.D., Liefner, I., 2012. Globalization, industrial restructuring, and regional development in China. Appl. Geogr. 32 (1), 102–105.
Williams, J., 2001. E-commerce and the lessons from nineteenth century exchanges. Am. J. Agric. Econ. 83 (5), 1250–1257.
World Bank Group, 2016. World Development Report 2016: Digital Dividends. World Bank Publications, Washington, DC.
Xiong, X., Nie, F., Bi, J., Waqar, M., 2017. The research on the path of poverty alleviation of e-commerce: a case study of Jing Dong. J. Simulat. 5 (2), 73.
Xu, Z., Wang, Z., Zhou, L., Wang, H., 2017. The spatial distribution characteristics and driving factors of "Taobao Village" in China. Econ. Geogr. 37, 107–114.
Yamauchi, F., 2007. Social learning, neighborhood effects, and investment in human capital: evidence from green-revolution India. J. Dev. Econ. 83 (1), 37–62.
You, J., Imai, K.S., Gaiha, R., 2016. Declining nutrient intake in a growing China: does household heterogeneity matter? World Dev. 77, 171–191.
Zeng, Y., Guo, H., 2016. The formation mechanism of agro-Taobao Village: a multiple-case study. Iss. Agric. Econ. 37, 39–48.
Zhang, X., Hu, D., 2014. Overcoming successive bottlenecks: the evolution of a potato cluster in China. World Dev. 63, 102–112.
Zhou, J., Yang, Z., Gao, W., 2017. The development characteristics and mechanism analysis of Taobao Villages in jiangsu province under e-commerce economy. Urban Dev. Stud. 24, 9–14.
Zhu, B., Song, Y., Li, G., Yu, T., 2016. Spatial aggregation pattern and influencing factors of "Taobao Village" in China. Econ. Geogr. 36, 92–98.
Zilberman, D., 2019. Agricultural economics as a poster child of applied economics: big data & big issues. Am. J. Agric. Econ. 101 (2), 353–364.