ABSTRACT

QI, JIAQI. The Formation and Effects of Taobao Villages. (Under the direction of Dr. Xiaoyong Zheng).

This dissertation aims to provide valuable information on the development and effects of Taobao villages, which are China's unique rural e-commerce clusters. Specifically, I study how Taobao villages are formed and how Taobao villages affect villagers' livelihood and economic well-being. In Chapter 1, I conduct the first econometric analysis to examine the determinants of the formation of Taobao villages. I examine the roles of various factors such as location, education, transportation and telecommunication infrastructure, local economic development levels, the structure of local economy, urban-rural income difference and spillover from nearby Taobao villages in the formation of Taobao villages. It is found that education plays an important and positive role in the formation of Taobao villages, while credit support for other businesses and existing Taobao villages nearby make the formation of new Taobao villages less likely.

Chapter 2 examines whether originating from an Agro-Taobao village influences a farmer's migration decision, using self-collected household survey data from three Agro-Taobao villages, which are clusters of e-commerce agribusinesses recently appeared in China, and three nearby non-Agro-Taobao villages. Results show that individuals from Agro-Taobao villages are 18% less likely to migrate than their counterparts from non-Agro-Taobao villages, after controlling for various factors that are likely to influence the migration decisions such as demographic characteristics, migration experience, educational background, and health status of the individuals and their spouses. Policy implications of the findings are also discussed.

Chapter 3 further examines the effects of Taobao villages as well as various economic indexes such as fiscal policy, financial development, and rural infrastructure on real rural net

income and urban-rural income inequality. It is found that the development of Taobao villages can be an effective method for raising incomes for rural residents, and to narrow the income gap between urban and rural China.

The Formation and Effects of Taobao Villages

by
Jiaqi Qi

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Economics

Raleigh, North Carolina

2018

APPROVED BY:

Xiaoyong Zheng Committee Chair	Mitch Renkow
Edward Kick	Zheng Li

BIOGRAPHY

Prior to completing the Ph.D. in economics at North Carolina State University, Jiaqi Qi received an MS in economics from the North Carolina State University in 2012. Before starting the graduate study, he earned a BS also in economics at Zhejiang University of Finance and Economics.

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Xiaoyong Zheng for his guidance on my dissertation and the patience on me, as well as the help from my committee, specifically, Dr. Edward Kick, Dr. Mitch Renkow and Dr. Zheng Li. Also, I would like to thank Dr. Liqun Zhu and Peng Cao from Nanjing Agricultural University, for the help on conducting the survey and collecting the data.

TABLE OF CONTENTS

LIST OF TABLES vi

LIST OF FIGURES vii

Introduction 1

REFERENCES 5

Chapter 1: The Formation of Taobao Villages in China 7

Introduction 7

Background 11

Data 17

Empirical Strategy 19

Poisson Panel Data Models 20

Negative Binomial Models 23

Standard Errors 25

Variables 25

Results 32

Conclusions 35

REFERENCES 38

Chapter 2: The Effect of Agribusiness Clusters on Farmers' Migration Decisions in China 53

Introduction 53

Survey and Data 60

Variables 62

Summary Statistics 66

Estimation and Results 68

Conclusions 71

REFERENCES 72

Chapter 3: Estimating the Effects of Taobao Villages on Rural Net Income and Urban-Rural Income Gap 80

Introduction 80

Survey and Data ... 86
Empirical Strategy ... 88
The Rural Income ... 88
The Intra-County Rural-Urban Income Gap ... 89
Explanatory Variables ... 90
Empirical Results ... 95
Conclusions ... 100
REFERENCES ... 102
Conclusions ... 112
APPENDIX ... 115
Appendix A The proof of partial coverage for Chapter 1 ... 116

LIST OF TABLES

Table 1.1 Districts Dropped from the Sample ... 43

Table 1.2 Summary Statistics ... 44

Table 1.3 Estimation Results from Poisson Count Models ... 45

Table 1.4 Estimation Results from Negative Binomial Count Models ... 47

Table 1.5 Estimation Results from Linear Models ... 50

Table 2.1 Definition of Variables ... 76

Table 2.2 Summary Statistics ... 77

Table 2.3 Summary Statistics by Migration Status ... 78

Table 2.4 Regression Results ... 79

Table 3.1 Districts Dropped from the Sample ... 106

Table 3.2 Summary Statistics and Definition of Variables ... 108

Table 3.3 Estimation of Rural Income ... 110

Table 3.4 Estimation of Urban-rural Income Gap ... 111

LIST OF FIGURES

Figure 1.1 Distribution of Taobao villages in China of 2013 51
Figure 1.2 Distribution of Taobao villages in China of 2014 51
Figure 1.3 Distribution of Taobao villages in China of 2015 52
Figure 1.4 Distribution of Taobao villages in China of 2016 52

Introduction

This dissertation aims to provide valuable information on the development and effects of Taobao villages, which are China's unique rural e-commerce clusters. Taobao village are defined as a cluster of rural e-retailers within an administrative village where: 1) Residents got started in e-commerce spontaneously primarily with the use of Taobao Marketplace; 2) Total annual e-commerce transaction volume is at least RMB 10 million ($1.6 million); and 3) At least 10% of the village households actively engage in e-commerce or at least 100 active online shops have been started by villagers (Gao et al., 2014). This dissertation studies how Taobao villages are formed and how Taobao villages affect village residents' livelihood and economic well being. Specifically, I study: 1) the determinants of the formation of Taobao villages, 2) how is the migration decision of the villagers influenced by whether the home village has become a Taobao village and 3) the effects of Taobao villages on rural net income and urban-rural income inequality.

Chinese regional policy has undergone several decades of trials, adjustments, and readjustments (Fan, 1997). During the first three decades since the establishment of the People's Republic of China in 1949, "Maoist development strategy" attempted to divert investments from more developed coastal urban areas to poorer interior rural areas, not only to correct the inherent uneven growth, but also to tap into inland raw materials and resources (Yang, 1990). However, these policies were found to lack economic efficiency and failed to accelerate national economic growth, since support of enterprises in the developed coastal region was much more efficient.

In the early 1980s, Deng Xiaoping's post-Mao policies have reversed the Maoist model of regional development strategy and priorities. As a result, several coastal regions were selected to be developed first. Compared with other areas, those open zones are better endowed with raw materials, energy, and natural resources. As a result, China’s gross domestic product (GDP) between 1983 and 2013 saw an average annual growth rate of 10.12%, making China's economy the second-largest in the world (Bajpai, 2013).

However, Deng’s policy has also created substantial income inequality both between provinces as well as between urban and rural areas. The per capita disposable income in urban areas was, on average, 2.63 times that of rural areas between 1978 and 2010, which is very high by international standards. In China, the large rural-urban income gap lead to a sectoral reallocation of labor from rural to urban areas. The massive internal rural-to-urban migration which started around mid-1980s has played a vital role in China’s rapid urbanization and industrialization process, and the number of migrants increased from about 30 million in the late 1980s to 130 million in 2006, and to over 220 million in 2010. However, the large number of migrant workers could potentially cause various social and economic problems including: 1) left-behind children and empty-nest elders; 2) rural human capital drain, as migrants tend to be more talented and productive (Gao, Zheng and Bu, 2014); and 3) it will be more difficult for the left-behind poor in rural areas to escape relative poverty (Su and Heshmati, 2013). Given these problems caused by income gap and internal migration, it will be important to examine and identify factors that can contribute to an increase in rural income or a reduction in the urban-rural income gap and can give migrant workers incentives to return to their home villages.

Studies find E-commerce to be an effective method to reduce poverty and raise income in many countries. Therefore, we expect rural e-commerce to be one of the solutions for raising rural income, reduce rural-urban income inequality, and attract migrants to return. E-commerce, and rural e-commerce in particular, has developed rapidly in China recently, and the growth has mostly been driven by the largest online trading platform, Taobao.com, founded in 2003 by Alibaba. In 2013, China's total online sales surpassed that of U.S. and became the largest market for e-business in the world (Tan, 2015), and there were 1.05 million online shops registered in villages and towns, where most of these online sellers are located. Taobao and e-commerce diffuses quickly in rural China since individuals can learn how to operate online stores easily from their families and friends, and the village becomes a Taobao village once its online businesses reach a certain scale. Taobao villages are expected to affect income in rural China in two ways: by helping to increase the income of e-commerce employers through Taobao businesses, and by increasing employees' income through job opportunities that Taobao and related businesses create in Taobao villages. Moreover, migrating villagers from Taobao villages have several reasons to return: 1) the e-commerce business environment or culture of Taobao villages significantly lowers the difficulty for migrants to start their own e-commerce businesses upon returning. 2) e-commerce businesses in Taobao villages create many job opportunities, which allow return migrants to earn an income while taking care of dependents and other family members at the same time.

As mentioned before, Taobao villages could bring huge economic and social benefits to the local residents. The first Taobao village appeared in 2009, and the number increased to 20 in 2013, 211 in 2014, and 780 in 2015 (Chen et al., 2016). There are thousands of villages in rural areas of China, yet only several hundred have become Taobao villages. Given the enormous benefits brought by Taobao villages, it is of interest to researchers and policymakers to understand what factors contributed to the formation of existing Taobao villages.

To date, there have been only a few case studies on the topic of Taobao villages (Guo, Liang and Luo, 2014; Zeng and Guo, 2015; Zeng, Qiu and Guo, 2015), and return migration decisions (Connelly, Roberts and Zheng, 2012; Demurger and Xu, 2011; Demurger and Xu, 2015; Chunyu, Liang, and Wu, 2013). This dissertation contributes to the existing literature of Taobao villages by: 1) conducting the first ever econometric analysis of the determinants of the formation of Taobao villages and in particular examining the roles of various factors such as location, education, transportation and telecommunication infrastructure, level of economic development, the structure of local economy, urban-rural income difference and spillover from nearby Taobao villages in the formation of Taobao villages; 2) conducting the first econometric analysis to examine whether a migrant's home village has become an Agro-Taobao village is a factor for the migrant workers to return; and 3) trying to explain the composition of intra-county income gap by rural e-commerce growth, using China's county/district level panel data. It is also the first study to provide an econometric analysis of the determinants of China's urban-rural income gap at the county/district level.

The rest of the dissertation is organized as follows. Chapters 1, 2 and 3 present the analyses on the factors which determine the formation of Taobao villages, the effect of the home village having become an Agro-Taobao village on the villagers' migration decisions, and the effects of the number of Taobao villages on the rural income and urban-rural income gap in China, respectively. Finally, the conclusion summarizes the results and limitations of the study, and discusses possible future research topics.

REFERENCES

Bajpai, P. (2013). "China's GDP Examined: A Service-Sector Surge," available at: https://www.investopedia.com/articles/investing/103114/chinas-gdp-examined-servicesector-surge.asp [Accessed 21 Feb. 2018].

Chen, L., Cui, L., Jiang, Z., Li, Z, Luo, Z., Sheng, Z., Wang, J. and Zeng, Y. (2016). "China Taobao Village Research Report of 2016," available at: http://i.aliresearch.com/img/20170414/20170414113512.pdf [Accessed 5 Jun. 2017].

Chunyu, M., Liang, Z. and Wu, Y. (2013). "Interprovincial Return Migration in China: Individual and Contextual Determinants in Sichuan Province in the 1990s," *Environment And Planning A, 45*, 2939 - 2958.

Connelly, R., Roberts, K. and Zheng, Z. (2012). "The Role of Children in the Migration Decisions of Rural Chinese Women," *Journal of Contemporary China, 21,* 93-111.

Demurger, S. and Xu, H. (2011). "Return Migrants: The Rise of New Entrepreneurs in Rural China," *World Development, 39*, 1847-1861.

Demurger, S. and Xu, H. (2015). "Left-Behind Children and Return Migration in China," *Journal of Migration, 4*, 1-21.

Fan, C. (1997). "Uneven development and beyond: Regional development theory in post-Mao China," *International Journal Of Urban And Regional Research, 21*, 620- 639.

Gao, H., Wang, X., Chen, L., Sheng, Z., Zhang, R. and Jiang, Z. (2014). "China Taobao Village Research Report of 2014," available at: http://www.jianshu.com/p/f2c28a4f53a0 [Accessed 5 Jun. 2017]. (in Chinese).

Gao, Y., Zheng, J. and Bu, M. (2014). "Rural-urban Income Gap and Agricultural Growth in China ," *China Agricultural Economic Review, 6*, 92-107.

Guo, G., Liang, Q. and Luo, G. (2014). "Effects of Clusters on China's E-Commerce: Evidence from the Junpu Taobao Village," *IJBM International Journal of Business and Management, 9*, 51-72.

Su, B. and Heshmati, A. (2013). "Analysis of the Determinants of Income and Income Gap between Urban and Rural China," *China Economic Policy Review, 2*, 1-29.

Tan, J. (2015). "Alibaba: The Giant With 81.5% Market Share In China's Online Shopping Market," available at: http.//www.sharesinv.com/articles/2015/11/09/ alibaba-the-giant-with-81-5-market-share-in-chinas-online-shopping-market [accessed 05 September 2017].

Yang, D. (1990). "Patterns of China's Regional Development Strategy," *The China Quarterly, 122*, 230-257.

Zeng, Y. and Guo, H. (2015). "The Formation Mechanism of Agro-Taobao Village: A Multiple-Case Study," *Agricultural Economy, 4*, 39-48. (in Chinese).

Zeng, Y.,Qiu, D. and Guo, H. (2015). "Study on the Formation of Taobao Village: Taking Dongfeng Village and Junpu Village as Examples," *Economics Geography, 35,* 90-97. (in Chinese).

CHAPTER 1 The Formation of Taobao Villages in China

Introduction

China is now the largest market for e-commerce in the world. The growth in e-commerce in China has been largely driven by Taobao.com, a website similar to Amazon.com in the US. Taobao.com not only provides consumers an easier way to purchase a larger variety of goods, but also provides sellers a place to sell goods without a physical store or large capital. The first generation of Taobao sellers were urban residents because only they had access to internet. As internet penetrated the rural areas, more and more Taobao sellers emerged in rural villages as they took advantage of the lower labor, capital and living costs there. Very often, many sellers from the same village sell the same or very similar products. Then quickly, stores of related businesses along the supply chain of the main products the village sells and other support businesses such as web page design and maintenance, shipping, financial services, photography, video and image processing, brand and management consulting and legal services were opened in the same village. These villages are called Taobao villages by Alibaba, the parent company of Taobao.com. They are essentially a new generation of low-cost manufacturing or factor endowment clusters that are common in a wide range of countries and sectors (Nadvi and Schmitz, 1994; Altenburg and Meyer-Stamer, 1999). Similar to the old generation of business clusters in developing countries such as the cotton knitwear industry in Tiruppur, India (Cawthorne, 1995) and the shoe industry in Sinos Valley, Brazil (Schmitz, 1995), the clusters produce their goods in less-developed and low-cost countries or areas and then sell goods to more developed ones. In the case of Taobao villages here, clusters appear in the less developed rural areas of China and their markets are mainly the more prosperous urban areas of China. Unlike the old generation of clusters, businesses in Taobao villages rely on the internet and e-commerce platforms to market and sell their products.

Firms in clusters enjoy the economies of agglomeration (Marshall, 1920). Marshall noted that the agglomeration of firms engaged in similar or related activities generated a range of localized external economies that lowered costs for clustered producers. Such advantages include a pool of specialization workers, easy access to suppliers of specialized inputs and services and the quick dissemination of new knowledge. As pointed out by Schmitz and Nadvi (1999), such economies of agglomeration are especially beneficial to small enterprises as clustering facilitates greater specialization and division of labor and effective investment in small steps rather than taking on large and wild risks. Taobao villages form and grow to exploit economies of agglomeration and sellers in Taobao villages become more competitive on e-commerce platforms.

Taobao villages also bring economic and social benefits to the local residents. First, the rural e-commerce helps narrow the income gap between urban and rural areas (Zheng, 2007). In recent years, income for people in underdeveloped rural areas in China has become more and more difficult to increase. The main reason is farming is the main source of income for people in these areas and growth in agricultural productivity has become very slow. On the other hand, urban areas have enjoyed relatively fast growth for the past thirty years. Participating in e-commerce or related businesses provides a potential source of alternative income for people in rural areas. On average, holding other factors constant, engaging in e-commerce raises a rural household's annual income by 20.5 thousand RMB (Chen et al., 2016). Studies of clusters in other industries and countries have also shown that clusters benefit local residents by helping them earning more (Guo, Liang and Luo, 2014; Fleisher et al., 2010; Martin et al., 2015; Rivera, Gligor and Sheffi, 2016).

Second, clusters like Taobao villages attract a lot of businesses and create a lot of job opportunities (Guo, Liang and Luo, 2014; Fleisher et al., 2010; Rivera, Gligor and Sheffi, 2016; Visser, E., Távara, J. and Villaran, 2015), enlarging the tax base for local governments significantly. With more tax revenue, governments can invest more in infrastructure improvement and provide more public services, thus raising the living standards for local residents.

Third, China has developed unequally during the last three decades. The enormous growth in eastern provinces due to the Open Door Policy and Coastal Development Strategy in the 1980s, especially in urban areas, has created substantial inequality between provinces and between urban and rural areas (Zhao and Tong, 2000; Sicular et al., 2007). This economic inequality has triggered massive migration from rural to urban areas and from western to eastern provinces since the late 1980s as people look for more job opportunities and better pay (Knight and Song, 1999; Knight, 2008). However, many cities do not provide public services such as education for migrant children and therefore their kids stay in the rural areas, taken care of by their grandparents instead. Parents' migration usually have negative effect on physical health (de Brauw and Mu, 2011), mental health (He et al., 2012) and academic performance (Amuedo-Dorantes and Pozo, 2006; Kandel and Kao, 2001) of left- behind children. Also, some senior citizens have difficulty getting needed help as their adult children worked in cities that are far away from their villages (Lin, Yin and Loubere, 2014; Giles and Mu, 2007). This creates all sorts of social problems. Taobao villages provide local residents more job opportunities with decent income. More importantly, people from these areas do not have to leave their hometown for work so that they can take better care of their kids and old parents.

There are hundreds of thousands of villages in rural areas of China, yet only several hundred have become Taobao villages. Given the benefits brought by Taobao villages, it is of interest to researchers and policymakers to understand what factors contributed to the formation of existing Taobao villages. The latest statistics show that China's internet penetration rate was 49.3 percent in 2014, compared to 87.4 percent in the US, indicating a large room for future growth (Lee, 2017). In particular, the internet penetration rate in rural areas was only 28.8 percent in 2014 (Tan, 2015). Therefore, identifying the main factors that contributed to the formation of existing Taobao villages can help policymakers design corresponding policies to maximize the chance that new Taobao villages form and develop.

In this paper, we conduct the first ever econometric analysis of the determinants of the formation of Taobao villages. To date, there have been only a few case studies on this topic (Guo, Liang and Luo, 2014; Zeng and Guo, 2015; Zeng, Qiu and Guo, 2015). We examine the roles of various factors such as location, education, transportation and telecommunication infrastructure, level of economic development, the structure of local economy, urban-rural income difference and spillover from nearby Taobao villages in the formation of Taobao villages. We find that education plays an important and positive role in the formation of Taobao villages, while credit support for other businesses and existing Taobao villages nearby make the formation of new Taobao villages less likely.

The rest of the paper is organized as follows. The next Section provides background information on the development of e-commerce and the formation of Taobao villages in China.

Section three describes our data. Our empirical strategy is presented in Section four. Section five presents and discusses the results while the Section six concludes.

Background

China has a late start in the development of e-commerce. The first online transaction was completed in April 1998. The growth was relatively slow during the following decade (Cao and Zhang, 2009). In 2008, online sales totaled 120 billion RMB, which accounted for only 1.15% of the total retail sales in China in that year. This number was much lower than those of the U.S. (3.72%) and the U.K. (4.5%) in the same year (Zhang and Li, 2008). However, e-commerce developed rapidly after that. In 2010, total online sales increased to 523 billion RMB and its share as a percentage of total retail sales increased to 3.3%. This share had further increased to 8% in 2013, 10.7% percent in 2014, and 12.3% in 2015. In 2013, China's total online sales surpassed that of U.S. and China is now the largest market for e-business in the world (Tan, 2015).

The growth of e-commerce in China has been mainly driven by Taobao.com thus far. Taobao.com was founded in 2003 by Alibaba,[1] the predominant firm in China's online market. In 2014, Taobao and its associated website Tmall occupied as much as 81.5% of the online shopping market in China, in term of gross merchandise volume (Tan, 2015). In comparison, the largest e-commerce company in the US, Amazon, only accounted for 27.1% of the U.S. retail market in 2012 (Yang, 2014). By the end of 2014, the average number of daily active users of

[1] Alibaba, which was founded in 1999, is the parent company of Taobao.com and has many other business lines including business to business (B2B) trading platforms, online retailing, search engines for shopping, cloud computing services and third-party payment services (Alipay). In September 2014, Alibaba's IPO (BABA) on NYSE raised $21.8 billion and it was the biggest IPO in history at that time.

Taobao had increased to over 120 million (Gao et al., 2014).

Like online shopping portals in other countries, Taobao.com provides consumers a new and more convenient way to purchase goods without going out, and probably more importantly, it offers consumers a significantly more variety of goods than what they can find in local grocery stores and supermarkets. One important reason for its huge success is that Taobao solves the asymmetric information problems for online transactions using several tools. First, it requires all transactions conducted on its website to go through the third-party payment system Alipay.[2] Once Alipay receives the payment for an item from the buyer, it will give the seller a notification to ship the good. After that, the buyer has 10 days to receive the good (shipping usually takes 1-5 days depending on distance), examine it and request for refund or return if the buyer is not satisfied with the good received. Alipay will only send the payment to the seller if the buyer confirms he has received the good in good condition or does not take any action within the 10-day window.

Second, Taobao employs a comprehensive rating system to rate sellers and displays the ratings on its website so that consumers can see them (Yu, 2016). When a transaction is completed, Taobao gives the seller one trading credit. Therefore, sellers with more credits are more experienced and reliable and are also more likely to be listed at or near the top of the list when consumers search for products. In addition, when a transaction is completed, the buyer will be asked to leave comments and ratings for both the seller and the product he bought. The buyer can rate the seller in three aspects: whether the item received is the same as the one described, whether the seller is easy to work with, and the shipping speed. A score from one to five is assigned by the buyer for each category. These are called detailed seller ratings. For each

[2] Alipay, also owned by Alibaba, is very similar to PayPal. Beyond what PayPal can provide, Alipay can also be used as an interest-bearing savings account.

category, the average score the seller got from his past transactions is displayed. In addition, Taobao compares the seller's rating with his peers, that is, sellers who sell products in the same major category of goods (such as garment, household products, digital goods and 13 others), and lists how this seller ranks in terms of percentiles among its peers for each category. Taobao also provides a seller's rating in terms of selling specific products. The product specific rating is simple and can either be positive, neutral or negative. Buyers can see the percentages of positive, neutral and negative ratings the seller got from selling this product in particular in the past. This prevents opportunistic sellers from building a good reputation by selling one product first and then cash in through cheating when selling another product.

The fast growth of China's e-commerce and Taobao.com not only provides customers an easier way to purchase the goods they want, but also provides new opportunities for individual entrepreneurs and small- and medium-size firms. Through Taobao, sellers can sell their products without a physical store (and hence there is no rent to pay) and with very few or no employees than themselves. Also, Taobao doesn't charge the seller a fee to open a store and doesn't take a cut from the transactions completed on its website. This relatively low capital requirement feature significantly reduces the entry barrier facing many entrepreneurs when starting their own businesses, especially those with limited resources. Also, Taobao provides sellers access to the national market rather than a local or regional market. For some products, demand at the local or regional level may be too low for the sellers to break even, but the aggregate demand at the national level is high enough for a seller to start producing and selling. Therefore, Taobao provides some sellers the opportunity to sell certain products they otherwise would not produce and sell.

Selling on Taobao requires the seller to have access to internet. Therefore, it is not surprising that at the early stage, a large percentage of Taobao sellers lived and operated their businesses in cities. Taobao.com was founded in 2003 and in 2006, the network coverage in China's rural areas was only 3.1%, compared with 20.2% in cities (China Internet Network Information Center, 2016). This means very few people in rural areas had access to the internet. However, as time went by and the internet penetrated rural areas. In December 2015, the network coverage in China's rural area was 31.6%, compared with over 65.8% in cities (China Internet Network Information Center, 2016). There are more and more Taobao sellers moved their factories or workshops to rural areas and more people in rural areas started new Taobao stores[3] while people in the cities closed their Taobao stores.[4] The main reason behind this geographic shift is the fact that business operating costs such as renting land and buildings for factories and storage and wages for employees are much lower in rural areas than in the cities. Also at the same time, huge improvements in infrastructure such as internet, telecommunication and highway were made in rural areas.[5] The shipping and logistics industries also grew very fast and expanded their networks to rural areas because of the rapid growth of rural e-commerce. All these developments made it feasible and actually more competitive for a seller to run his Taobao business from a rural area. On the other hand, the developing internet-related infrastructure and logistics system enable more villagers to purchase goods online. China's Ministry of Commerce reported online purchases from rural areas were 180 billion RMB in 2014, 353 billion RMB in 2015, 894.5 billion RMB in 2016 and 537.6 billion RMB in the first six months of 2017.

[3] In November 2013, there were 2.039 million "Taobao" shops registered in rural areas, an increase of 24.9% compared with the number at the end of 2012 (Liu, Li and Liu, 2015).
[4] China e-Business Research Center estimated the total number of private consumer-to-consumer sellers in China had decreased 17.8 percent to 11.22 million from 2012 to 2013 and predicted a further decrease to 9.18 million by the end of 2014, and Taobao is estimated to account for four-fifths of this market (Clover, 2014).
[5] Since 2010, approximately one-third of the rural investment have been spent on basic infrastructure including water, power grids, roads and natural gas each year. China built more than 0.38 million kilometers of new roads in rural areas from 2010 to 2012, and by the end of 2012, the total mileage of rural roads in China had reached 3.63 million kilometers (National Development and Reform Commission, 2013).

Once Taobao hit rural areas, it diffused very quickly. Unlikely the cities, China's rural villages are often regarded as the "acquaintances societies," which have their own widely accepted rules and the information is relatively public (Chen, Zhang and Li, 2014). Once an individual starts operating a Taobao store online in a village and his business does well, his successful experience will be learned by his relatives and neighbors very quickly.[6] Because of the low entry barrier to starting a Taobao store, many of his relatives and neighbors in the same village will soon start their own Taobao businesses. The successes of some of them will then attract even more people to join the business. At some villages, this process continues and the villages eventually become rural e-commerce clusters. Take Junpu village[7] as an example (Zeng, Qiu and Guo, 2015). In July 2012, Mr. Haijing Huang and his friends moved their online clothing business from Guangzhou, the capital city of Guangdong province to Junpu village. Their e-shops operated well and they shared their experiences with their relatives and friends and provided them with information and help. The number of online shops had increased to over 50 by the end of that year, and further increased to nearly 300 by the end of 2013. The success of e-commerce attracted a lot of migrant workers and college graduates who left Junpu to return to the village to open and operate e-shops. By the end of 2014, 348 of the 490 households and nearly 2000 villagers in Junpu had been engaged in e-commerce or related businesses. They opened more than 3,000 Taobao shops which had an annual sales of 1.68 billion RMB and net profit of 15 million RMB, and their products included garments, plastic products, leather goods, electronics, hardware and toys. Alibaba defines a Taobao village as a village where the number

[6] This is similar to the knowledge spillover phenomenon studied in the trade and development literatures. For example, Koenig, Mayneris, and Poncet (2010) found exporting decisions by firms in France are influenced by the same decisions by their neighbors and Foster and Rosenzweig (1995) found Indian households' decisions of adopting high-yielding seed varieties (HYVs) are influenced by their neighbors' experience with HYVs.
[7] Junpu village is located in Jieyang, Guangdong, with a total area of 0.53 square kilometers, 490 households, and a population of 2,695 in 2014. It is one of the first Taobao villages.

of active online shops exceeds 100, or 10% of the number of households and annual e-commerce sales exceed 10 million RMB (Gao et al., 2014). The first Taobao village appeared in 2009 and the number increased to 20 in 2013, 211 in 2014 and 780 in 2015 (Chen et al., 2016). Figure 1.1 shows the distribution of Taobao villages across different provinces between 2013 and 2016 and how the number of Taobao villages grew with the rural internet penetration rate.

The designation as Taobao villages will not bring direct benefits to the villages, either from Alibaba or from local governments. However, clustering will bring both benefits and costs to firms in the cluster (Huang, Zhang and Zhu, 2008). First, as discussed above, clustering allows firms to enjoy economies of agglomeration. This is especially beneficial to small enterprises as clustering facilitates greater specialization and division of labor and effective investment in small steps rather than taking on large and wild risks (Schmitz and Nadvi, 1999). Through specialization, the firms become more efficient, innovative and productive and hence more competitive on the Taobao platform. Second, clustering in one industry attracts businesses in related industries. Again, take Junpu village as an example, from the first Taobao seller in July 2012 to the end of 2014, 15 shipping companies had set up offices in Junpu. There are also other professionals who moved from cities providing services including financing, photography, video and image processing, branding and management consulting, legal services and even e-shop operator training (Zeng, Qiu and Guo, 2015). These businesses and services provide convenience and reduce the production and marketing costs for Taobao sellers in the village and make them more competitive on the Taobao platform. Third, as more businesses come to the village, more farmers are able to find off-farm work and their incomes are higher and more diversified. In Taobao villages, villagers participating in e-business are usually return migrant workers, college or high school graduates and children of farmers who didn't farm before starting running e-shops on Taobao. However, there are also farmers who learn from their families or friends to run a Taobao online shop, and when that happens, they usually let their farmland idle or lease it to other farmers.

Clustering also has its downside. As an e-commerce cluster, the Taobao village is a result of diffusion achieved through imitation and replication. Therefore, most sellers in the same village are selling similar or even identical goods. This creates intense competition and it drives down the price for the goods and bids up wages and other input costs. As a result, the entire industry becomes less profitable (Fleisher et al., 2010). For example, in a Taobao village called Dongfeng, rapid growth of online shops triggered a price war and the profit margin dropped from 75% in 2008 to 9% in 2014 (Zeng, Qiu and Guo, 2015). With a lower profit margin, some sellers may lower the quality of the products offered to cut costs, which could then lead some consumers to stop buying from these sellers. This is a vicious cycle of lower profit margin, and then lower quality, and then lower demand and then even lower profit margin. For example, in a Taobao village called Qinghe, after a period of quick expansion, lack of brand recognition and dominant firms was identified as the main challenge faced by the e-commerce industry in the village (Wu and Huang, 2014).

Data

To examine the determinants of the formation of Taobao villages, a county-level data set was collected from the statistical yearbooks of Zhejiang and Jiangsu provinces for the years 2012 to 2015. We focus on this time period because 2012 is the year Alibaba began to count the number and publish the list of Taobao villages and data for the year 2016 or after have not become available yet. The reasons for choosing these two provinces is: 1) they are the first two provinces that have Taobao villages. 2) Each year, Zhejiang, Guangdong and Jiangsu are the three provinces with the most numbers of Taobao villages, from year 2012 to 2016. We excluded Guangdong provinces from our study because the provincial yearbook of Guangdong did not provide some important indicators at county and district level, such as the number of internet user, total mileage of roads and the number of private vehicles owned. The statistical yearbooks are edited and published by local bureaus of statistics and report statistics on population, economy, employment, investment in fixed assets, energy, transportation, finance, foreign economic relations, wages and other income, household characteristics as well as other social economic variables at the province, city and county level.

The administrative divisions of China are different from most of other countries. There are a total of 34 provinces in China. Each province consists of 10 to 30 prefecture-level cities, each of which in turn consists of 5 to 15 districts and counties. Within a prefecture-level city, districts are usually located in urban areas while counties are usually located in rural areas, although some newly formed districts still cover large rural areas. Each district or county consists of 5 to 15 townships, each of which then consists of 10 to 50 villages. Therefore, on average, there are about 300 to 500 villages in a county or district. We excluded the districts that cover no or little rural areas because our study focuses on the Taobao villages, which by definition, are located in rural areas. We calculated the population density for each district and dropped those with a density higher than one thousand people per squared kilometers as urban areas usually have much higher population densities. Table 1.1 lists all the districts that were dropped.

Ideally, we would like to use the online transaction revenue from each village as our dependent variable, but we are not allowed to have access to this data. Therefore, we use the number of Taobao villages in each county instead as our key variable of interest. It was constructed using the lists of Taobao villages, published annually by Alibaba since 2012 (Gao et al., 2014; Gao et al., 2015; Chen et al., 2016). To examine the spillover and learning effects from neighboring Taobao villages, for each county in each year, we also counted the number of Taobao villages in surrounding counties in the previous year. We defined surrounding counties as those counties that share a common border with the county of interest as long as the border is not Yangtze River[8], which is the longest river in China and is about 2 kilometers wide in Jiangsu, or Taihu Lake, which has an area of more than 2,000 square kilometers.

Empirical Strategy

As we have a panel dataset and our dependent variable, the number of Taobao villages, is a count variable, we estimate various panel data count models for our empirical analysis. We start with Poisson count models, the most widely used count models, followed by Negative Binomial count models. We also estimate the standard linear model for robustness check.

Poisson Panel Data Models

Random Effects Model

The first model we estimate is the Poisson random effects model, first proposed by Hausman, Hall and Griliches (1984). This model assumes that,

(1) $y_{it} | x_i, c_i \sim Poisson\,[c_i \, \exp(x_{it}\beta)]$,

[8] In Jiangsu, there are 12 bridges on Yangtze Rivers now in 2018, and five of them are located in Nanjing. Thus, we can say the transportation between two counties on two sides of Yangtze Rivers is relatively insufficient, and that is the reason we excluded those counties from neighboring counties.

where y_{it} is the number of Taobao villages in county/district i in year t, x_{it} is a $1*k$ vector of observed explanatory variables that vary across both counties and years, c_i is the time-invariant county specific unobservable effect that only varies across counties, β is a $k*1$ vector of parameters, $x_i \equiv [x_{i1}' x_{i2}' ... x_{iT_i}']'$, and $Poisson\,[c_i \exp(x_{it}\beta)]$ denotes a Poisson density with mean $c_i \exp(x_{it}\beta)$. The model further assumes that y_{it} and y_{ir} are independent conditional on x_i and c_i for any $t \neq r$. In addition, it is assumed that c_i is independent of x_i and follows a $Gamma(\delta, \frac{1}{\delta})$ density and thus $E(c_i) = 1$ and $Var(c_i) = 1/\delta$.

With these assumptions, we can integrate out the time-invariant unobserved effect c_i and get the joint likelihood function for the dependent variables for county i across different years conditional only on the observed explanatory variables,

$$(2)\ \Pr(Y_{i1} = y_{i1}, ..., Y_{iT_i} = y_{iT_i} | x_i) = \frac{\prod_{t=1}^{T} \lambda_{it}^{y_{it}} \Gamma(\delta + \sum_{t=1}^{T_i} y_{it})}{\prod_{t=1}^{T} y_{it}! \Gamma(\delta) (\sum_{t=1}^{T_i} \lambda_{it})^{\sum_{t=1}^{T_i} y_{it}}} w_i^{\delta} (1 - w_i)^{\sum_{t=1}^{T_i} y_{it}},$$

where T_i is the number of years observed for county i, $\lambda_{it} = \exp(x_{it}\beta)$, $w_i = \dfrac{\delta}{\delta + \sum_{t=1}^{T_i} \lambda_{it}}$, and the log likelihood function can be written as,

$$(3)\ \begin{aligned} \ell_i &= \log\Gamma(\delta + \textstyle\sum_{t=1}^{T_i} y_{it}) - \log\Gamma(\delta) + \delta \log w_i + \log(1 - w_i) \sum_{t=1}^{T_i} y_{it} + \sum_{t=1}^{T_i} y_{it}(x_{it}\beta) \\ &- (\textstyle\sum_{t=1}^{T_i} y_{it}) \log(\sum_{t=1}^{T_i} \lambda_{it}) \end{aligned} .$$

The random effect estimator $\hat{\beta}_{RE}$ maximizes the log likelihood $L = \sum_{i=1}^{N} \ell_i$, where N is the number of counties in the dataset and ℓ_i is defined in (3).

Random Effects Model with Weaker Assumptions

The random effects model above assumes that the dependent variable follows the Poisson distribution (1) above. This assumption might be too strong in some cases. Alternative models with weaker assumptions are available. Suppose we only assume that

(4) $E(y_{it}|x_i, c_i) = c_i \exp(x_{it}\beta)$ and

(5) $E(c_i|x_i) = E(c_i) = 1$.

Note (1) implies (4) but not vice versa. Also, (5) means c_i is uncorrelated with x_{it} for all t. In this case, the random effect c_i can be removed using the law of iterated expectations, yielding,

(6) $E[E(y_{it}|x_i, c_i)|x_i] = E[c_i \exp(x_{it}\beta)|x_i] = \exp(x_{it}\beta)E[c_i|x_i] = \exp(x_{it}\beta)$.

With (6), we can use the Poisson Quasi-Maximum Likelihood Estimation (QMLE) method and treat the panel data as a pooled cross-sectional data to obtain a consistent estimator for β (Wooldridge, 1997). In particular, the random effects QMLE estimator $\hat{\beta}_{QMLE}$ maximizes $\sum_{i=1}^{\mathrm{N}} \sum_{t=1}^{T_i} l_{it}(\beta)$ where l_{it} is the log likelihood function for y_{it} when it follows the Poisson density with mean $\exp(x_{it}\beta)$.

Fixed Effects Model

The two random effects models above assume that the unobserved county-specific effect c_i is uncorrelated or independent of the observed explanatory variables x_{it}. This assumption is often regarded as too strong because time-invariant county-specific effect is often correlated with other time-varying county characteristics. The fixed effects model, again, first proposed by Hausman, Hall and Griliches (1984), relaxes this assumption.

In particular, the fixed effects model assumes (1) but allows arbitrary correlation between c_i and x_{it}. It also assumes that y_{it} and y_{ir} are independent conditional on x_i and c_i for any $t \neq r$. With these assumptions, it can be shown that y_i, the T_i *1 vector for the dependent variables y_{it} for county i from all years, follows,

(7) $y_i | K_i, x_i, c_i \sim \text{Multinomial}\{K_i, p_1(x_i, \beta), ..., p_{T_i}(x_i, \beta)\}$,

where $K_i \equiv \sum_{t=1}^{T_i} y_{it}$, $p_t(x_i, \beta) \equiv \exp(x_{it}\beta) / [\sum_{r=1}^{T_i} \exp(x_{ir}\beta)]$, and the Multinomial $\{K_i, p_1(x_i, \beta), ..., p_{T_i}(x_i, \beta)\}$ denotes the multinomial density with parameters $K_i, p_1(x_i, \beta), ..., p_{T_i}(x_i, \beta)$. Therefore, we can estimate the parameters of interest by maximizing the multinomial log likelihood function based on (7). In particular, the log likelihood for county i is,

(8) $\ell_i(\beta) = \sum_{t=1}^{T_i} y_{it} \log[p_t(x_i, \beta)]$.

The fixed effects estimator $\hat{\beta}_{FE}$ maximizes $\sum_{i=1}^{N} \ell_i(\beta)$, where $\ell_i(\beta)$ is defined in (8).

Negative Binomial Models

In addition to the Poisson count data models above, we also estimate several negative binomial count data models. One restriction imposed by the Poisson models is that they assume the conditional mean of the dependent variable equals exactly the conditional variance. However, the variance is often larger than the mean empirically, implying "overdispersion" in the data (Hausman, Hall and Griliches, 1984). Negative binomial models accommodates this important data feature.

Pooled Cross-sectional Model

We first ignore the panel feature of the data and estimate a pooled cross-sectional negative binomial model. The model assumes that,

(9) $y_{it}|x_{it},c_{it} \sim Poisson[c_{it}\exp(x_{it}\beta)]$,

where c_{it} is an unobserved factor that varies across both year and county. Other variables are defined above. It is further assumed that c_{it} is independent of x_{it} and follows a $Gamma(\delta, \frac{1}{\delta})$ distribution.

With these assumptions, we can integrate out the unobserved effect c_{it} and get the likelihood function for the dependent variable for county i in year t conditional only on the observed explanatory variables,

(10) $\Pr(y_{it}|x_{it}) = \frac{\Gamma(y_{it}+\delta)}{\Gamma(y_{it}+1)\Gamma(\delta)}\left[\frac{\delta}{\delta+\exp(x_{it}\beta)}\right]^{\delta}\left[\frac{\exp(x_{it}\beta)}{\delta+\exp(x_{it}\beta)}\right]^{y_{it}}$

and the pooled cross-sectional negative binomial estimator maximizes $\sum_{i=1}^{N}\sum_{t=1}^{T_i}\log\Pr(y_{it}|x_{it})$.

Random Effects Model

Both the random effects and fixed effects negative binomial models are from Hausman, Hall and Griliches (1984). In the random effects model, we assume

(11) $y_{it}|\lambda_{it} \sim Poisson[\lambda_{it}]$,

where λ_{it} follows the $Gamma(\exp(x_{it}\beta), \frac{1}{\delta_i})$ distribution and $\frac{\delta_i}{\delta_i+1}$ is a random variable (hence the random effects model) following the $Beta(r,s)$ distribution and independent of x_i. With these assumptions, the joint likelihood function for the dependent variables for county i across different years conditional only on the observed explanatory variables is,

$$(12)\ \Pr(Y_{i1}=y_{i1},\ldots,Y_{iT_i}=y_{iT_i}\,|\,x_i)=\frac{\Gamma(r+s)\Gamma(r+\sum_{t=1}^{T_i}\lambda_{it})\Gamma(s+\sum_{t=1}^{T_i}y_{it})}{\Gamma(r)\Gamma(s)\Gamma(r+s+\sum_{t=1}^{T_i}\lambda_{it}+\sum_{t=1}^{T_i}y_{it})}\prod_{t=1}^{T_i}\frac{\Gamma(\lambda_{it}+y_{it})}{\Gamma(\lambda_{it})\Gamma(y_{it}+1)}$$

where $\lambda_{it}=\exp(x_{it}\beta)$. The random effect estimator maximizes the log likelihood $L = \sum_{i=1}^{N} logPr_i$ where Pr_i is defined in (12).

Fixed Effects Model

The fixed effects negative binomial model has exactly the same assumptions as those of the random effects negative binomial model with the only differences that δ_i can be correlated with the explanatory variables x_i and we do not impose assumption regarding the distribution of δ_i. Hausman, Hall and Griliches (1984) show in this case, the sum of the total counts across years is a sufficient statistic and the joint probability of the counts for county i across different years conditional on the total counts as well as the observed explanatory variables x_i is,

$$(13)\ Pr_i=\Pr(Y_{i1}=y_{i1},\ldots,Y_{iT_i}=y_{iT_i}\,|\,x_i,\sum_{t=1}^{T_i}Y_{it}=\sum_{t=1}^{T_i}y_{it})=\frac{\Gamma(\sum_{t=1}^{T_i}\lambda_{it})\Gamma(\sum_{t=1}^{T_i}y_{it}+1)}{\Gamma(\sum_{t=1}^{T_i}\lambda_{it}+\sum_{t=1}^{T_i}y_{it})}$$

$$\prod_{t=1}^{T_i}\frac{\Gamma(\sum_{t=1}^{T_i}\lambda_{it})\Gamma(\sum_{t=1}^{T_i}y_{it}+1)}{\Gamma(\lambda_{it})\Gamma(y_{it}+1)}.$$

The fixed effects estimator maximizes the log likelihood $L = \sum_{i=1}^{N} logPr_i$ where Pr_i is defined in (13).

Standard Errors

As all the panel data count models are estimated using the maximum likelihood estimation method, for each estimation, we obtain robust standard errors using the Huber-White sandwich formula (Huber, 1967; White, 1982).

Variables

In this section, we discuss the explanatory variables we use in the estimation and their expected effects on the number of Taobao villages. First, it is important to control for some of the most basic differences among counties. As our data come from two provinces, one dummy variable for Jiangsu province is included to control for any factors at the provincial level that determine the formation of Taobao villages. Also, some counties are large while others are small. Some have more people while others have less. Larger counties or counties with more people have more villages and hence are naturally more likely to have more Taobao villages. Therefore, we include the variable $area$, defined as the total area of the county in squared kilometers, and the variable pop, defined as number of people in the county (in thousands), in the estimation. In fixed effects models, any time-invariant variables are subsumed in the county-specific effects so the area and the province dummy variables are only included in the random effects estimations.

Among the 780 Taobao villages in 2015, over 90% of them are located in five provinces[9] on the east coast. This clustering phenomena could reflect some innate advantages these five provinces possess that are amenable to the development of Taobao villages, but it could also reflect that the knowledge spillover effect is at play. The knowledge spillover effect has been

[9] Zhejiang, Jiangsu, Guangdong, Fujian and Shandong.

studied in the trade and development literatures. For example, Fernandes and Tang (2014) show that Chinese firms are more likely to become exporters if neighboring firms are successful in the foreign market. If there is a similar knowledge spillover effect in e-commerce, then villages close to existing Taobao villages are more likely to become Taobao villages themselves as farmers in these villages can quickly learn from their neighbors which can be consider as the complementary effect of neignboring Taobao villages. Therefore, we include the variable nei_taobao, defined as the number of Taobao villages in surrounding countries in the previous year, in the estimation. Here, surrounding counties are defined as counties that share a common border with the county of interest. However, neighboring Taobao villages could have a second effect on the county of interest, that is, the crowding out effect or competition effect. Pulles and Schiele (2013) pointed out that from the resource point of view, the competitive advantage of regional clusters is explained by the resources that cluster firms possess, and the social interactions between cluster firms and preferential resource allocation enable them to obtain better resources than firms located outside of the cluster. Consider an existing Taobao village as a regional cluster of e-commerce, firms (Taobao sellers) in the regional cluster have better access to resources such as cheap raw materials, low transportation and service fees, and experienced workers than Taobao sellers outside the cluster. These competitive advantages of existing Taobao villages will make nearby counties less likely to have new Taobao villages. Therefore, the effect of nei_taobao on the number of Taobao villages is ambiguous.

E-commerce is conducted through the internet and transactions are completed by shipping goods from sellers to buyers. Therefore, the development of Taobao villages is impossible without the development of transportation, telecommunication and logistic infrastructure. For this reason, we include the number of broadband internet users per thousand people in the county,

$internet$, and the revenue for the shipping industry per county resident in thousand RMB in the county, $post$, in the estimation. Furthermore, Agrawal, Galasso and Oettla (2016) find that a 10% increase in a region's highway mileage causes a 1.7% increase in regional innovation growth over a five-year period, suggesting that the variable $roads$, which is defined as the length of road per capita in kilometers, may be correlated with regional growth. The higher these three variables are, the better the county's infrastructure is, and thus, these three variables are expected to have positive effects on the formation of Taobao villages. One concern here, is that these variables are likely to be endogenous in the regression due to reverse causality. Counties with more Taobao villages have more internet users as well as higher revenue for the shipping industry by construction. Also, the emergence of Taobao villages will also encourage local governments to invest more on roads in order to improve the infrastructure. Therefore, we lag these three variables to lessen the concern from endogeneity.

Running an e-commerce business requires the use of a computer to operate the shop online and very often a vehicle for transporting inputs and outputs. Since the 1980s, computer and programming courses have been part of the high school education in China to prepare students for widespread use of computers in modern society (Chen, 1988). As a result, farmers with high school education are more likely to have the necessary knowledge and skills to learn how to operate a shop online than those haven't started or completed high school. Ideally, we should use the percentage of people that have high school education to represent the regional level of education, but such variable is not available at county level. Therefore, we include $high$, defined as the number of high school students per thousand people in the county as our education variable, and also include $vehicle$, defined as the number of privately owned vehicles per thousand people in the county into our regression. Both variables are expected to have positive effects on the formation of Taobao villages. Again, the $high$ variable is likely to suffer from the reverse causality problem. This is because high school education is not mandatory in China. It is possible for some students in Taobao villages to drop out of high school to help with family business. Therefore, we lag the $high$ variable in the estimation.

Business owners employ themselves and how many people are interested in becoming self-employed depends on other jobs available. For this reason, we include $uempl$, the number of people employed by non-private companies or government agencies per one thousand people in urban areas of the county, in our estimation. Non-private jobs are usually more stable and people are less likely to become self-employed if more of this type of jobs are available. In China, the ratio of the average salary of central government employees and the average wage for private-sector workers is 2.22, which is higher than many other countries in the world,[10] and the ratio of the average wage of local government employees and the average wage for private-sector workers is 1.13, which is also higher than many other major countries in the world. In addition, public sector employees also enjoy various kinds of benefits including low-prices apartments, employer paid pension contributions and healthcare subsidy (Lang, 2016). Hence, this variable is expected to have a negative impact on the number of Taobao villages. Again, we lag it to avoid the endogeneity concern caused by reverse causality as Taobao villages may also attract some people to leave such jobs to start their own e-commerce businesses.

Urban-rural income difference, ur_inc, is another variable that influences the number of Taobao villages in a county. Higher income in the urban areas of the county will attract more farmers to seek employment in the urban areas (e.g. Mundlak and Strauss, 1978; Barkley, 1990) and hence are less likely to stay in villages to start their own e-commerce businesses. On the

[10] The ratios are 2.05, 1.7, 1.79, 1.74, 1.49, 1.36 and 1.23 in Canada, Japan, UK, US, Germany and France, respectively.

other hand, for those who stay, it is more likely for them to try something new such as starting an online business to try to close the income gap. Therefore, the effect of this variable is ambiguous. Again, we lag this variable to avoid the endogeneity concern caused by reverse causality as Taobao villages often help close the urban-rural income difference.

Next, we include several county level economic indicators in the estimation to control for differences among counties in levels of economic development. The first variable is the per capita GDP of the county in thousand RMB, GDP. Counties with higher GDP tend to have more firms and factories. As some of the online shops simply sell locally produced goods online especially in Agro-Taobao villages[11] (Guo, Liang and Luo, 2014; Zeng and Guo, 2015; Wu and Huang, 2014), counties with more firms and factories are more likely to have Taobao villages. On the other hand, more firms and factories mean more traditional jobs for the county's residents and then they have less incentives to start their own businesses. Therefore, the effect of the GDP variable is ambiguous. Second, the fixed asset investment share of the GDP, $asset$, is included as a control variable. Fixed asset investment refers to investments in infrastructure and real estate. This variable has an ambiguous effect on the number of Taobao villages. On one hand, better transportation and telecommunication infrastructure is amenable to the formation of Taobao villages. On the other hand, real estate construction employs a lot of people, who often are migrant workers from the rural areas (Shi, 2008; Swider, 2015), making it less likely for people to stay in rural areas to start their own e-commerce businesses. Similarly, the ratio of fiscal revenue to GDP, $fiscal$, is included as a control. Higher fiscal revenue indicates the local tax base is larger. When the government has more money, it is more likely to invest in public infrastructure, which will facilitate the development of Taobao villages. On the other hand,

[11] Taobao villages which specialize in selling local produced agricultural products.

higher fiscal revenue also allows government to have more benefits and welfare for low-income people like farmers in rural areas, giving them less incentives to start their own online businesses. Therefore, the expected effect of this variable is ambiguous. Fourth, the tertiary industry share of the GDP, $tert$, is included as a control variable. This variable represents the importance of the service industries in the local economy. Services such as advertising, shipping and website design and maintenance are critical for successful e-commerce businesses and hence we expect this variable to have a positive effect. Fifth, we control for the value of output per worker in large manufacturing firms in million RMB, $output$. On average, the value of output per worker in heavy industries is more than twice of that of light industries, so a lower value of this variable means the light industries have a larger share of the local economy. A large portion of the goods sold on Taobao are manufactured by the light industries[12] so this variable is expected to have a negative effect on the number of Taobao villages. Sixth, the total amount of loans lent by local banks as a share of GDP, $loan$, is included as another control variable. A higher value of this variable indicates credit is easier to access. In China, 4.16% of the outstanding loans were agriculture-related in 2014, and the number changed to 3.74% in 2015. Therefore we can say outstanding loans are mostly non- agricultural. As an industry with low capital and technology entry barriers, the decision of being a Taobao seller is not affected too much by the difficulty of borrowing money from banks and other financial institutions (Guo, Liang and Luo, 2014). However, other businesses are more likely to get loans, make investments and hire more people, giving people less incentives to start their own online businesses on Taobao. Therefore, the effect of this variable is expected to be negative. Seventh, we control for the per capita total retail sales of consumer goods including e-commerce in thousand RMB, $retail$. Retail sales are purchases

[12] The top 10 selling categories of goods on Taobao in 2014 were: women's clothing, men’s clothing, cellphone, beauty products, computer accessories, leather goods, furniture, shoes, auto parts, and toys (Gao et al., 2014).

of finished goods and services by consumers and measure demand for consumer goods in a certain area (Amadeo, 2017). Therefore, this variable represents the purchasing power of the local residents and higher purchasing power gives potential sellers confidence to start their e-commerce businesses. As a result, this variable is expected to have a positive effect on the number of Taobao villages. Obviously, the development of e-commerce and the formation of Taobao villages will affect all the economic indicators discussed here so all of them are lagged to avoid the potential endogeneity concern due to reverse causality.

Finally, we include year dummies to control for unobserved macro-level factors that affect all counties in the same year. And as discussed in the previous section, county-specific and time-invariant unobserved determinants of the number of Taobao villages are controlled by the c_i term in the models. Table 1.2 lists the definitions of all the variables used in estimation and also reports their summary statistics.

Results

The estimation results are reported in Tables 1.3, 1.4 and 1.5. By construction, counties with zero Taobao villages in all three years cannot be included in the fixed effects estimations of count data model and negative binomial model, and there are 44 such counties in our sample. As a result, the numbers of observations for fixed effects estimations are about one third less than those of other estimations. Also, those counties will not be excluded in the linear model, thus the robustness check will focus on comparing the coefficients of random effects models.

We first discuss the Poisson estimation results from Table 1.3. Four results are robust to the estimation method used. Results show that counties in Zhejiang province have more Taobao villages. This is not surprising as Zhejiang is the birthplace of Alibaba and Taobao. Also, Zhejiang has the long tradition of small businesses as it has the most number of light manufacturing industrial clusters in China (Zhu, 2015) including: world's largest wholesale center Yiwu, a county-level city with 70,000 wholesale booths; "China's capital of hardware and door" Yongkang, a county-level city with several important industrial clusters: automobile components, door industry, cup industry, electronic tools, electronic appliances and cookware; "Capital of Packaging in China" Longgang; "China's capital of low-voltage electric appliances" Yueqing; "China's capital of mold and art ware" Huangyan; "Capital of Socks in China" Zhuji; "Capital of Umbrella in China" Shangyu and hundreds of other clusters of accessories and hardware. Second, more Taobao villages appeared in 2014 and 2015 than in 2013. This simply reflects the trend that more and more people in rural areas have started their e-commerce businesses through Taobao.com in recent years after learning the successful stories from others. Third, counties with a greater fraction of high school graduates tend to have more Taobao villages. Counties with more high school graduates have more college graduates as well and some of them return to the villages after graduation because of the higher living cost in big cities and the fierce competition in urban job markets, especially after the expansion of higher education in China. The unemployment rate of college graduates was more than 10% from 2007 to 2010, and nearly 100,000 new college graduates could not find jobs in 2011 (Mai and He, 2015). These people have the necessary computer, communication and other skills required to run a successful business online. The fixed effects estimation results show that when the fraction of high school graduates increases by one per thousand local residents (the sample average is 42 per thousand local residents in the dataset), the expected number of Taobao villages in the county will increase by 0.19. This effect is both economically and statistically significant. Fourth, easier access to credit crowds out Taobao villages. In China, most banks are state-owned (Dong et al., 2014) and they like to lend more to large firms for two reasons. First, many of the large firms are also state-owned and they are obliged to support their investments. Second, banks regard loans made to large firms as less risky as many of them are state-owned and hence are backed implicitly by the government (Yeung, 2009; Chi and Li, 2017). As a result, small businesses such as online shops on Taobao are not likely to get loans from banks. Our results here imply that as other businesses get loans, make investments and hire more people, less people choose to start their own businesses online.

There are also some interesting results that are not robust across different estimation methods, especially between random effects models and the fixed effects model. The differences between the random and fixed effects models can come from two sources. First, random effects models assume the unobserved county level effects are uncorrelated with other explanatory variables while the fixed effects model does not reply on this assumption. Second, the sample for the fixed effects model is smaller as counties with zero Taobao villages in all three years cannot be included in the fixed effects estimation. Several results are worth mentioning. First, there is weak evidence that the competition effect of Taobao villages in nearby counties dominates the complementary effect of those Taobao villages. The estimated coefficients for the nei_taobao variable are not statistically significant in the random effects models, but are negative and statistically significant at 10% level in the fixed effects model. When surrounding counties have one more Taobao village in the previous year, there will be 0.13 less Taobao villages in the county this year. Second, the two random effects models show that more Taobao villages appear in more populous counties as we expected. However, this effect is statistically insignificant in the fixed effects model.[13] Third, as we expected, more non-private sector jobs in urban areas lead to less number of Taobao villages according to the random effects models. However, this effect is positive according to the fixed effects model. One potential explanation for the latter result is that many shops on Taobao were started by people who returned from urban areas (Zeng, Qiu and Guo, 2015). These farmers were exposed to e-commerce when they worked in urban areas and started their own when they returned. Counties with more urban jobs employ more people from rural areas and when they return, they are more likely to start their own e-commerce businesses.

The negative binomial results reported in Table 1.4 are very similar to those of the Poisson models. The province dummy, year dummies, number of high school graduates and the easiness to access credit variables continue to have the same robust results across different specifications. Two new results from the fixed effects model are worth mentioning. Higher share of the service industries (the $tert$ variable) and higher ratio of the fiscal revenue to GDP (the $fiscal$ variable), lead to more Taobao villages. Both results conform to our expectations. Finally, Table 1.5 collects the results of linear models for robustness check. Comparing the linear random effects model with the other two random effects models, we find that the coefficients for the province dummy and year dummy of 2014 from the linear model are statistically insignificant. They are not consistent with the results of other two models which means under the assumption of this particular linear model, Zhejiang's number of Taobao villages is not significantly different from that of Jiangsu's, and there is no evidence that time trend played an important role in increasing the number of Taobao villages in 2014. The estimation coefficients for the year dummy of 2015, population (the pop variable), education (the $high$ variable), and urban employment (the $uempl$ variable) are robust across three random effects models in terms of both the sign and magnitude.

[13] If we replace the area and population variables with the population density variable in our estimation, the sign and significance of poisson and negative binomial model did not change a lot.

Besides, other significant coefficients in linear model are robust to either Poisson count data model or negative binomial count model.

Conclusions

In this paper, we use various count regression models to examine the determinants of the formation of Taobao villages. We find that education plays an important and positive role in the formation of Taobao villages, while credit support for other businesses and existing Taobao villages nearby make the formation of new Taobao villages less likely.

Our results have important policy implications. In order to maintain the high growth rate of GDP in China, especially since the subprime mortgage crisis in 2008, China tried to transit from investment oriented growth model to consumption based model. In 2012, the contribution of consumption to GDP growth was 51.8%, which surpassed the contribution of investment for the first time since 2000. The number kept increasing each year and finally reached 64.5% in 2017 which means the consumption is now playing more and more important role in China's economy. On the other hand, online retail accounts for 18.54% of the total retail sales in China and the number is expected to rise in the future. Therefore, policies that support the development of online businesses in rural area with e-commerce potential are necessary for local governments. To help new Taobao villages to form and develop, local governments should focus on promoting high school attendance and give less credit support to other businesses. Also, support for areas that are far away from existing Taobao villages are more likely to be successful.

Due to data availability, our econometric analysis is conducted at the county level and we are only able to include a limited set of variables as the potential determinants of the formation of Taobao villages. Also, data on Taobao transaction volumes are not available to us so we can only examine how various socioeconomic factors affect the number of Taobao villages in a county, not the transaction volume. If data at the town or village level or data on Taobao transaction volume become available in the future, a more detailed study can be conducted to further examine the determinants of the formation of Taobao villages or e-commerce clusters in general.

Having examined the determinants of the formation of Taobao villages, naturally the next question to ask is what are the benefits of the Taobao villages bring to the local residents and the society in general. Does participating in e-commerce increase income in rural areas of China? Does e-commerce help solve other social problems such as left behind children? Does e-commerce replace small vendors in cities, who are also from rural areas? These are the questions left for future research.

REFERENCES

Agrawal, A., Galasso, A. and Oettl, A. (2016). "Roads and Innovation," *Review of Economics and Statistics, 99,* 417-434

Altenburg, T. and Meyer-Stamer, J. (1999). "How to Promote Clusters: Policy Experiences from Latin America," *World Development*, *27*, 1693-1713.

Amadeo, K. (2017). "What Is Retail Sales? What Are Its Components?." available at: https://www.thebalance.com/what-is-retail-sales-3305722 [Accessed 17 Sep. 2017].

Amuedo-Dorantes, C. and Pozo, S. (2006). "International Migration, Remittances and the Education of Children: The Dominican Case," Working Paper, Department of Economics, Western Michigan University.

Barkley, A. P. (1990). "The Determinants of the Migration of Labor out of Agriculture in the United States, 1940-1985," *American Journal of Agricultural Economics*, *72,* 567-573.

Cao, L. and Zhang, Z. (2009). "the Research Report of the Twelve Years of China's E-Commerce: 1997-2009," available at: http://tech.qq.com/2009921e/ebaogao.doc [Accessed 5 Jun. 2017]. (in Chinese).

Cawthorne, P. (1995). "Of Networks and Markets: The Rise and Rise of a South Indian Town, the Example of Tiruppur's Cotton Knitwear Industry," *World Development*, *23*, 43-56.

Chen, F. (2016). "Research on Logistics Development of Taobao Village in China." available at: http://www.xzbu.com/3/view-8138714.htm [Accessed 17 Sep. 2017]. (in Chinese).

Chen, L., Cui, L., Jiang, Z., Li, Z, Luo, Z., Sheng, Z., Wang, J. and Zeng, Y. (2016). "China Taobao Village Research Report of 2016," available at: http://i.aliresearch.com/img/20170414/20170414113512.pdf [Accessed 5 Jun. 2017].

Chen, Q. (1988). "Computer Education in Secondary Schools in the People's Republic of China." *Journal of Research of Science Teaching, 25*, 493-500

Chen, X., Zhang, L. and Li, W. (2014). "A Network Evolution Model for Chinese Traditional Acquaintance Networks," *IEEE Intelligent Systems, 29,* 5-13.

Chi, Q. and Li, W. (2017). "Economic Policy Uncertainty, Credit Risks and Banks' Lending Decisions: Evidence from Chinese Commercial Banks," *China Journal of Accounting Research, 10,* 33-50.

China Internet Network Information Center. (2016). "Research Report on China's Rural Internet Development, " available at: http./www.199it.com/archives/512565.html [Accessed 6 Jun. 2017]. (in Chinese).

Clover, C. (2014). "Alibaba's Sellers Rethink Business Models," available at: http.//www.ft.com/cms/s/0/e7939ece-fc8f-11e3-81f5-00144feab7de.html?ft_site=falcon#axzz4jF5NtDWS [Accessed 7 Jun. 2017].

de Brauw, A. and Mu, R. (2011). "Migration and the Overweight and Underweight Status of Children in Rural China," *Food Policy, 36*, 88-100.

Dong, Y., Meng, C., Firth, M. and Hou, W. (2014). "Ownership Structure and Risk-Taking: Comparative Evidence from Private and State-Controlled Banks in China," *International Review of Financial Analysis, 36*, 120-130.

Fernandes, A. P. and Tang, H. (2014). "Learning to Export from Neighbors," *Journal of International Economics, 94*, 67-84.

Fleisher, B., Hu, D., Mcguire, W. and Zhang, X. (2010). "The Evolution of an Industrial Cluster in China," *China Economic Review, 21*, 456-469.

Foster, A. D. and Rosenzweig, M. R. (1995). "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture," *Journal of Political Economy, 103*, 1176-1209.

Gao, H., Wang, X., Chen, L., Sheng, Z., Zhang, R. and Jiang, Z. (2014). "China Taobao Village Research Report of 2014," available at: http://www.jianshu.com/p/f2c28a4f53a0 [Accessed 5 Jun. 2017]. (in Chinese).

Gao, H., Chen, L., Sheng, Z., Zhang, R. and Jiang, Z. (2015). "China Taobao Village Research Report of 2015," available at: http://www.jianshu.com/p/6a24c126d8f8 [Accessed 5 Jun. 2017]. (in Chinese).

Giles, J. and Mu, R. (2007). "Elderly Parent Health and the Migration Decisions of Adult Children: Evidence from Rural China," *Demography*, 44, 265-288.

Golley, J. (2010). "China's prospects for diminishing regional disparities," available at: http://www.eastasiaforum.org/2010/08/31/chinas-prospects-for-diminishing- regional-disparities/ [Accessed 29 Aug. 2017].

Guo, G., Liang, Q. and Luo, G. (2014). "Effects of Clusters on China's E-Commerce: Evidence from the Junpu Taobao Village," *IJBM International Journal of Business and Management, 9*, 51-72.

Hausman, J., Hall, B. H. and Griliches, Z. (1984). "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," *Econometrica*, *52*, 909-938.

He, B., Fan, J., Li, H., Wang, Y., Williams, J. And Wong, K. (2012). "Depression Risk of 'Left-Behind Children' in Rural China," *Psychiatry Research, 200*, 306-312.

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

Huber, P. J. (1967). "The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions," in *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics*, Volume 1. Berkeley. University of California Press, 221-233.

Kandel, W. and Kao, G. (2001). "The Impact of Temporary Labor Migration on Mexican Children's Educational Aspirations and Performance," *International Migration Review, 35*, 1205-1231.

Knight, J. (2008). "Reform, Growth, and Inequality in China," *Asian Economic Policy Review* 3. 140-158.

Knight, J., Song, L. (1999). "The Urban-Rural Divide: Economic Disparities and Interactions in China," Oxford University Press, New York.

Lang, X. (2016). "China Civil Servants' Wages Won Two 'World First'." available at: http://www.aboluowang.com/2016/0224/697400.html [accessed 17 Sep. 2017].

Lee, T. B. (2017). "Internet Users by Country (2016)," available at: http.//www.internetlivestats.com/internet-users-by-country/ [accessed 05 June 2017].

Li, R. (1988). "The Growth Pole Theory in China's Regional Development and the Study of Policies," *Economic Management*, 4, 63-70. (in Chinese).

Lin, K., Yin, P. and Loubere, N. (2014). "Social Support and the 'Left Behind' Elderly in Rural China: A Case Study from Jiangxi Province," *Journal of Community Health, 39*, 674-681.

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 Agricultural Research, 7*, 35-37.

Mai, Y. and He, Y. (2015). "Higher Education Expansion in China and the 'Ant Tribe' Problem," *Higher Education Policy, 28*, 333-352.

Martin, R., Florida, R., Pogue, M. and Mellander, C. (2015). "Creativity, Clusters and the Competitive Advantage of Cities," *Competitiveness Review, 25*, 482-496.

Mundlak, Y. and Strauss J. (1978). "Occupational Migration Out of Agriculture in Japan," *Journal of Development Economics*, *5*, 55-71.

Nadvi, K. and Schmitz H. (1994). "Industrial Clusters in Less Developed Countries: Review of Experiences and Research Agenda," Discussion Paper, No. 339. Institute of Development Studies, University of Sussex, Brighton.

National Development and Reform Commission. (2013). "Report of Rural Infrastructure Construction Development," available at: https.//njs.ndrc.gov.cn/tzzn/201504/W020150401347572595490.doc [Accessed 7 Jun. 2017]. (in Chinese).

Pulles, N. J. and Schiele H. (2013). "Social Capital Determinants of Preferential Resource Allocation in Regional Clusters," *Management Revue, 24*, 96-113.

Rivera, L., Gligor, D. and Sheffi Y. (2016). "The Benefits of Logistics Clustering," *International Journal of Physical Distribution & Logistics Management, 46*, 242-268.

Schmitz, H. (1995). "Collective Efficiency: Growth Path for Small-Scale Industry," *Journal of Development Studies*, *31*, 529-566.

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

Shi, L. (2008). "Rural Migrant Workers in China: Scenario, Challenges and Public Policy," Working Paper, Policy Integration and Statistics Dept., International Labor Office, Geneva.

Sicular, T., Yue, X., Gustafsson, B. and Li, S. (2007). "The Urban-Rural Income Gap and Inequality in China," *Review of Income and Wealth, 53*, 93–126.

Swider, S. (2015). "Building China: Precarious Employment among Migrant Construction Workers," *Work, Employment & Society, 29*, 41-59.

Tan, J. (2015). "Alibaba: The Giant With 81.5% Market Share In China's Online Shopping Market,", available at: http://www.sharesinv.com/articles/2015/11/09/ alibaba-the-giant-with-81-5-market-share-in-chinas-online-shopping-market [accessed 01 September 2016].

Visser, E., Távara, J. and Villaran, F. (2015). "Growing but not Developing: Long-Term Effects of Clustering in the Peruvian Clothing Industry," *Tijdschrift voor economische en sociale geografie, 106*, 78-93.

White, H. (1982). "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, *50,* 1-26.

Wooldridge, J. M. (1997): "Quasi-Likelihood Methods for Count Data," in *Handbook of Applied Econometrics*, vol 2, ed. M. H. Pesaran and P. Schmidt. Oxford: Blackwell, 352-406.

Wu, X. and Huang, C. (2014). "Research on the Development of Chinese Agriculture-Related E-Commerce," Working Paper, Zhejiang University College of Management, Zhejiang University, Hangzhou. (in Chinese).

Yang, G. S. (2014). "Potentials and Perils of E-Business in China," In I. Lee, ed. *Encyclopedia of e-commerce development, implementation, and management,* Hershey, PA, USA. Business Science Reference pp. 1250-1262.

Yeung, G. (2009). "How Banks in China make Lending Decisions," *Journal of Contemporary China, 18*, 285-302.

Yu, H. (2016). "Essays on Reputation Mechanisms in Online Marketplaces," Ph.D. Dissertation, North Carolina State University.

Zeng, Y. and Guo, H. (2015). "The Formation Mechanism of Agro-Taobao Village: A Multiple-Case Study," *Agricultural Economy, 4*, 39-48. (in Chinese).

Zeng, Y.,Qiu, D. and Guo, H. (2015). "Study on the Formation of Taobao Village: Taking Dongfeng Village and Junpu Village as Examples," *Economics Geography, 35,* 90-97. (in Chinese).

Zhang, Y., Li, M. (2008). "China's Online Shopping Market is Warming Rapidly," *People 's Daily Overseas Edition*, 15 November, p. 1.

Zhao, X. and Tong, S. (2000). "Unequal Economic Development in China: Spatial Disparities and Regional Policy Reconsideration, 1985–1995," *Regional Studies 34.* 549–561.

Zheng, Y. (2007). "Principal Component Cluster Analysis of the Development of Rural E-Commerce Regional Infrastructure,"*Forum on Science and Technology in China, 1*, 119-122. (in Chinese)

Zhu, J. (2015). “Industrial Clusters in Zhejiang Province, China,” available at. https://jingsourcing.com/industrial-clusters-in-zhejiang-province-china/ [Accessed 5 Jun. 2017].

Table 1.1 Districts Dropped from the Sample

Fuzhou City districts	Gulou, Taijiang, Cangshan, Jinan, Mawei
Xiameng City districts	Siming, Huli, Jimei, Haicang, Tongan, Xiang’an
Putian City districts	Chengxiang, Hanjiang, Licheng, Xiuyu
Sanming City districts	Meilie, Sanyuan
Quanzhou City districts	Licheng, Fengze, Luojiang, Quangang
Zhangzhou City districts	Xiangcheng, Longwen
Hangzhou City districts	Gongshu, Shangcheng, Xiacheng, Jianggan, Xihu, Binjiang
Ningbo City districts	Haishu, Jiangdong, Jiangbei, Beilun, Zhenghai
Wenzhou City districts	Lucheng, Longwan, Ouhai, Dongtou
Jiaxing City districts	Nanhu, Xiuzhou
Huzhou City districts	Wuxing, Nanxun
Shaoxing City districts	Yuecheng
Jinhua City districts	Wucheng, Jindong
Quzhou City districts	Kecheng, Qujiang
Zhoushan City districts	Dinghai, Putuo
Taizhou City districts	Jiaojiang, Huangyan, Luqiao
Lishui City districts	Liandu
Nanjing City districts	Xuanwu, Gulou, Qinhuai, Jianye, Yuhuatai, Xixia, Jiangning, Pukou, Liuhe, Lishui, Gaochun
Wuxi City districts	Liangxi, Xishan, Huishan, Binhu, Xinwu
Xuzhou City districts	Gulou, Yunlong, Jiawang, Quanshan, Tongshan
Changzhou City districts	Tianning, Zhonglou, Xinbei, Wujin
Suzhou City districts	Gusu, Huqiu, Wuzhong, Xiangcheng, Wujiang
Nantong City districts	Chongchuan, Gangzha
Lianyungang City districts	Lianyun, Haizhou, Ganyu
Huaian City districts	Huaian, Huaiyin, Qingjiangpu, Hongze
Yancheng City districts	Tinghu, Yandu, Dafeng
Yangzhou City districts	Guangling, Hanjiang, Jiangdu
Zhengjiang City districts	Jingkou, Runzhou, Dantu
Taizhou City districts	Hailing, Gaogang, Jiangyan
Suqian City districts	Sucheng, Suyu

Table 1.2. Summary Statistics

Variables	Obs	Mean	Std. Dev.	Min	Max
N_taobao	333	1.348	3.809	0	37
Province	333	0.405	0.492	0	1
Area	333	1.529	0.701	0.097	4.452
Pop	331	770.431	381.626	74.1	1651.2
Nei_taobao	333	0.961	1.882	0	12
Internet	333	200.718	76.682	55.62	493.92
Posts	333	1.050	0.522	0.35	5.15
High	332	42.490	9.307	20.4	72.39
Vehicle	328	126.216	62.265	35.79	402.4
Uempl	333	135.614	84.617	34.24	500.76
Ur_inc	330	15.699	4.402	6.4	28.22
Gdp	333	59.256	29.760	13.46	181.85
Asset	333	59.573	15.737	24.7	125.9
Fiscal	333	8.137	1.859	4.73	13.55
Tert	333	40.691	5.561	28.95	59.99
Aoutput	316	0.993	0.378	0.298	2.279
Loan	332	97.015	39.659	25.2	208.8
Retail	333	20.020	8.103	6.13	42.19

N_taobao: number of Taobao villages in the county;
Province: Whether the county is in Jiangsu province;
Area: area of the county in thousand square kilometers;
Pop: the population of the country in thousands;
Nei_taobao: lagged number of Taobao villages in all bordering counties;
Internet: lagged number of broadband internet users per one thousand residents;
Posts: lagged posts revenue per person in thousand RMB;
High: lagged number of high school students per thousand residents;
Vehicle: number of private vehicles owned by per thousand residents;
Uempl: lagged number of people employed in non-private sector per one thousand residents;
Ur_inc: lagged difference in average salary between urban and rural areas in thousand RMB;
GDP: lagged per capita GDP of the county;
Asset: lagged fixed asset investment to GDP ratio;
Fiscal: lagged local fiscal revenue to GDP ratio;
Tert: lagged tertiary industry share of the GDP;
Aoutput: lagged output per worker in large industrial enterprises in millions RMB;
Loan: lagged loan to GDP ratio;
Retail: lagged total per capita retail sales of consumer goods in thousand RMB.

Table 1.3 Estimation Results from Poisson Count Models

Variables	Random effects	Random Effects with Weaker Assumptions	Fixed effects
	(1)	(2)	(3)
D2014	2.460***	2.483***	2.162**
	(6.32)	(5.32)	(2.56)
D2015	4.139***	4.022***	3.275***
	(9.68)	(7.21)	(2.69)
Province	-1.330**	-0.676*	
	(-2.04)	(-1.69)	
Area	-0.073	0.099	
	(-0.32)	(0.37)	
Pop	0.002***	0.002***	0.01
	(4.75)	(5.44)	(0.75)
Nei_taobao	-0.027	0.033	-0.128*
	(-0.60)	(0.88)	(-1.66)
Internet	-0.001	0.002	-0.008
	(-0.38)	(0.65)	(-1.30)
Posts	0.203	0.052	0.084
	(1.08)	(0.29)	(0.23)
High	0.049**	0.049**	0.186*
	(2.43)	(2.37)	(1.93)
Vehicle	0.013**	0.005	0.025
	(2.23)	(1.37)	(1.23)
Uempl	-0.005***	-0.004***	0.071***
	(-3.36)	(-3.79)	(-4.14)
Ur_inc	-0.003	-0.032	0.026
	(-0.04)	(-0.63)	(0.30)
GDP	0.004	0.003	0.087
	(0.51)	(0.5)	(0.63)
Asset	-0.012	-0.018	0.020
	(-1.07)	(-1.20)	(0.44)
Fiscal	-0.118	-0.134*	0.728
	(-1.38)	(-1.84)	(1.41)
Tert	0.020	0.024	0.147
	(0.55)	(0.94)	(1.12)
Aoutput	-0.527	-0.710*	0.258
	(-1.22)	(-1.88)	(0.08)
Loan	-0.015**	-0.007	-0.065***

Table 1.3 (continued)

	(-1.98)	(-1.36)	(-2.73)
Retail	-0.010	0.002	0.041
	(-0.28)	(0.07)	(0.24)
Constant	-4.353**	-4.332***	
	(-2.04)	(-1.46)	
County FE			Included
Observations	309	309	182
Number of county	111	111	63
Log-Likelihood	-284.586	-370.582	-59.177

Notes: T-statistics in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.4 Estimation Results from Negative Binomial Count Models

Variables	Pooled Cross-sectional	Random Effects	Fixed effects
	(1)	(2)	(3)
D2014	2.512***	2.459***	1.820***
	(5.32)	(5.7)	(3.09)
D2015	4.069***	4.144***	2.772***
	(8.19)	(8.89)	(2.84)
Province	-1.314***	-1.298**	
	(-2.66)	(-2.00)	
Area	0.009	-0.078	
	(0.05)	(-0.37)	
Pop	0.002***	0.002***	-0.009***
	(5.16)	(3.77)	(-2.94)
Nei_taobao	0.021	-0.026	-0.090
	(0.45)	(-0.58)	(-1.13)
Internet	-0.001	-0.001	-0.004
	(-0.19)	(-0.27)	(-0.68)
Posts	0.125	0.209	0.228
	(0.54)	(1.07)	(0.63)
High	0.044***	0.049***	0.115**
	(2.92)	(2.61)	(2.07)
Vehicle	0.011**	0.012**	0.022
	(2.33)	(2.21)	(1.36)
Uempl	-0.006***	-0.005**	0.044**
	(-4.20)	(-2.53)	(2.31)
Ur_inc	-0.017	-0.005	0.008
	(-0.29)	(-0.08)	(0.09)
GDP	0.007	0.005	
	(0.95)	(0.42)	
Asset	-0.015	-0.012	-0.022
	(-1.45)	(-0.93)	(-0.44)
Fiscal	-0.119*	-0.120	1.067**
	(-1.81)	(-1.15)	(2.25)
Tert	0.023	0.020	0.343**
	(0.9)	(0.66)	(2.38)
Aoutput	-0.708**	-0.521	
	(-2.01)	(-1.01)	
Loan	-0.014***	-0.015*	-0.060***

Table 1.4 (continued)

	(-2.59)	(-1.91)	(-2.78)
Retail	-0.014	-0.009	0.115
	(-0.46)	(-0.29)	(0.99)
Constant	-3.514**	-0.212	-19.180**
	(-2.33)	(-0.03)	(-2.48)
County FE			Included
Observations	309	309	192
Number of county	111	111	64
Log-Likelihood	-304.428	-284.571	-64.919

Notes: T-statistics are in parentheses.
* p < 0.10, ** p < 0.05, *** p < 0.01

Table 1.5 Estimation Results from Linear Models

Variables	OLS	Random Effects	Fixed effects
	(1)	(2)	(3)
D2014	0.315	0.315	0.514
	(0.71)	(0.78)	(0.35)
D2015	2.271***	2.257***	2.325
	(2.91)	(2.84)	(0.88)
Province	-0.459	-0.464	
	(-0.54)	(-0.54)	
Area	-0.195	-0.196	
	(-0.66)	(-0.53)	
Pop	0.00222***	0.00222***	-0.0188
	(3.86)	(3.76)	(-1.42)
Nei_taobao	0.181	0.186	0.242
	(1.01)	(0.97)	(0.87)
Internet	0.0133	0.0134	0.023
	(0.93)	(0.95)	(1.07)
Posts	-1.224	-1.26	-2.32
	(-1.23)	(-1.17)	(-1.61)
High	0.0492*	0.0491*	-0.00164
	(1.80)	(1.81)	(-0.01)
Vehicle	0.0169*	0.0170*	0.0341
	(1.94)	(1.82)	(1.09)
Uempl	-0.00773***	-0.00774***	-0.00639
	(-2.95)	(-3.71)	(-0.72)
Ur_inc	0.00861	0.00834	0.0231
	(0.10)	(0.09)	(0.08)
GDP	-0.0139	-0.0138	-0.19
	(-0.83)	(-1.01)	(-1.49)
Asset	-0.00303	-0.00277	-0.0763*
	(-0.18)	(-0.15)	(-1.89)
Fiscal	-0.219**	-0.218**	0.177
	(-2.33)	(-2.12)	(0.21)
Tert	0.112*	0.112**	-0.0497
	(1.84)	(2.03)	(-0.20)
Aoutput	-0.986**	-1.002*	-2.218*
	(-2.02)	(-1.77)	(-1.85)
Loan	-0.0247**	-0.0249**	-0.0851

Table 1.5 (continued)

	(-2.34)	(-2.21)	(-1.63)
Retail	0.0382	0.0395	0.537**
	(0.59)	(0.58)	(2.34)
Constant	-4.817	-4.79	24.71
	(-1.51)	(-1.62)	(1.64)
County FE			Included
Observations	309	309	309
Number of county	111	111	111
R-squared (within)	0.3616	0.3310	0.3900

Notes: T-statistics in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

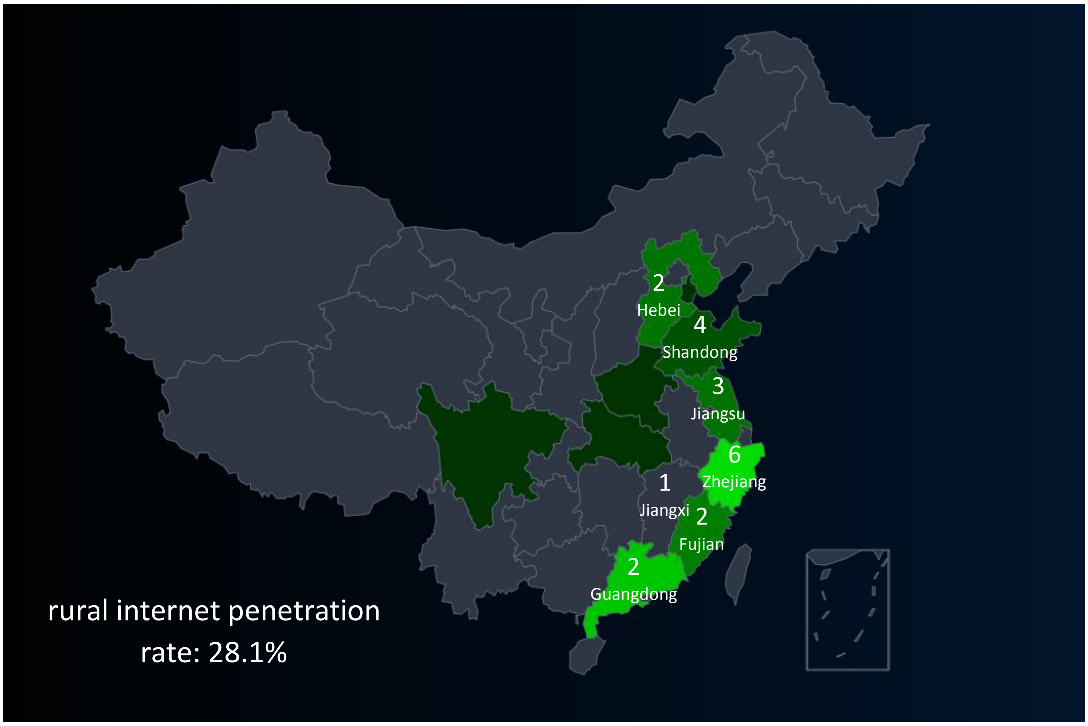

Figure 1.1 Distribution of Taobao villages in China of 2013

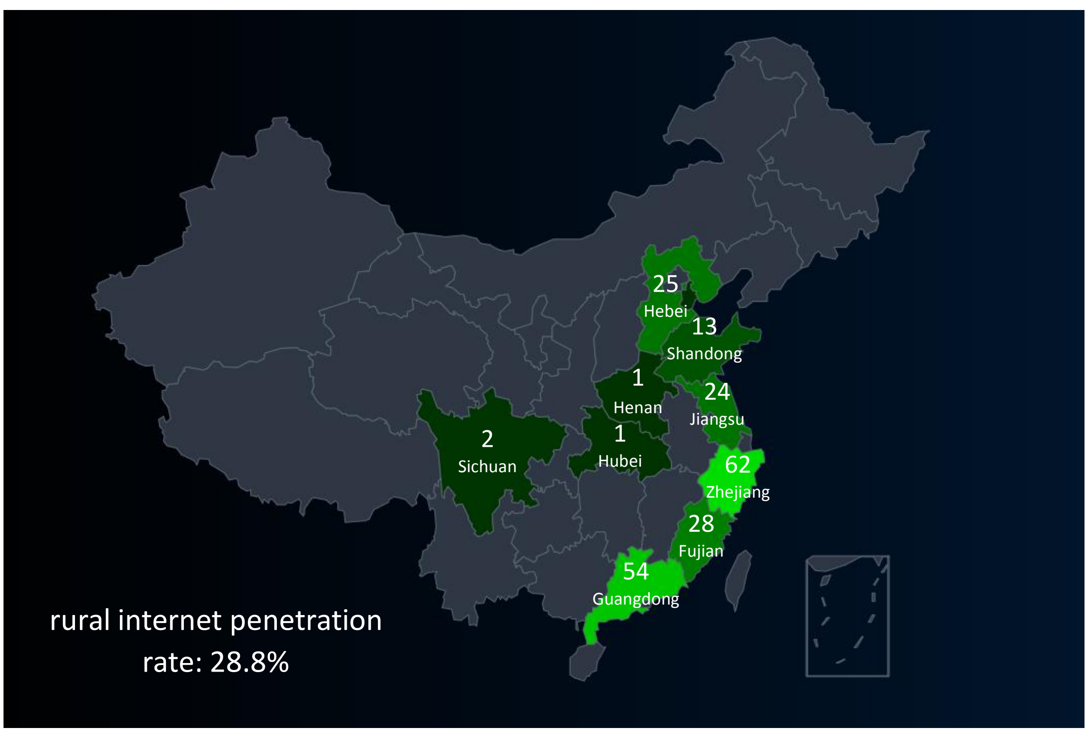

Figure 1.2 Distribution of Taobao villages in China of 2014

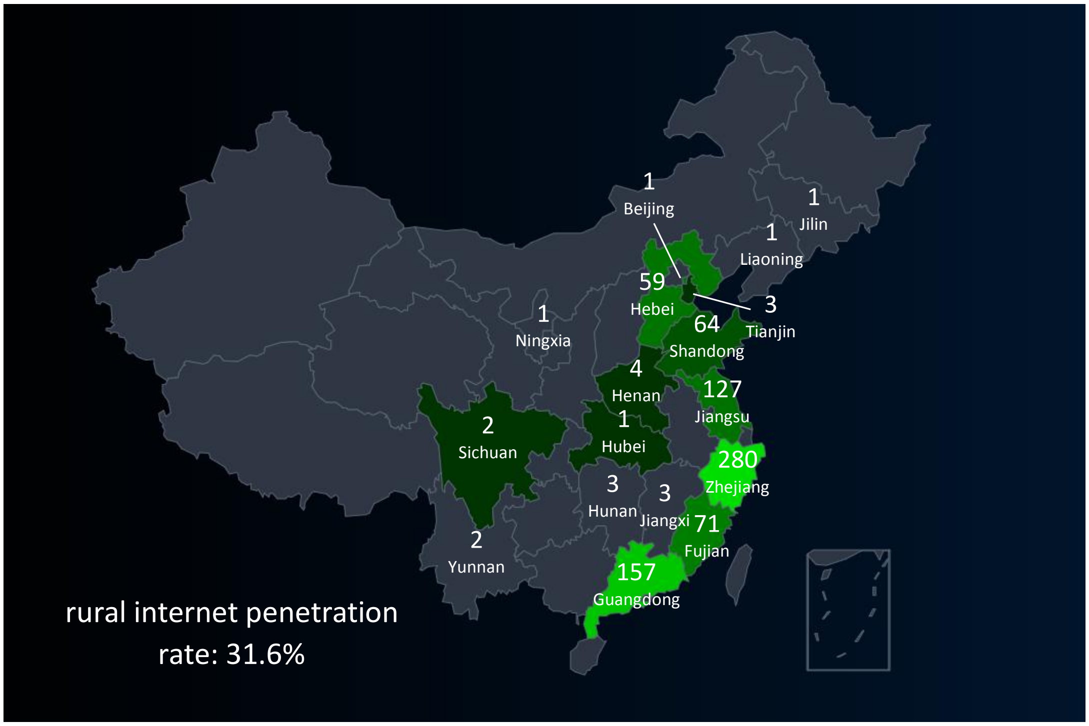

Figure 1.3 Distribution of Taobao villages in China of 2015

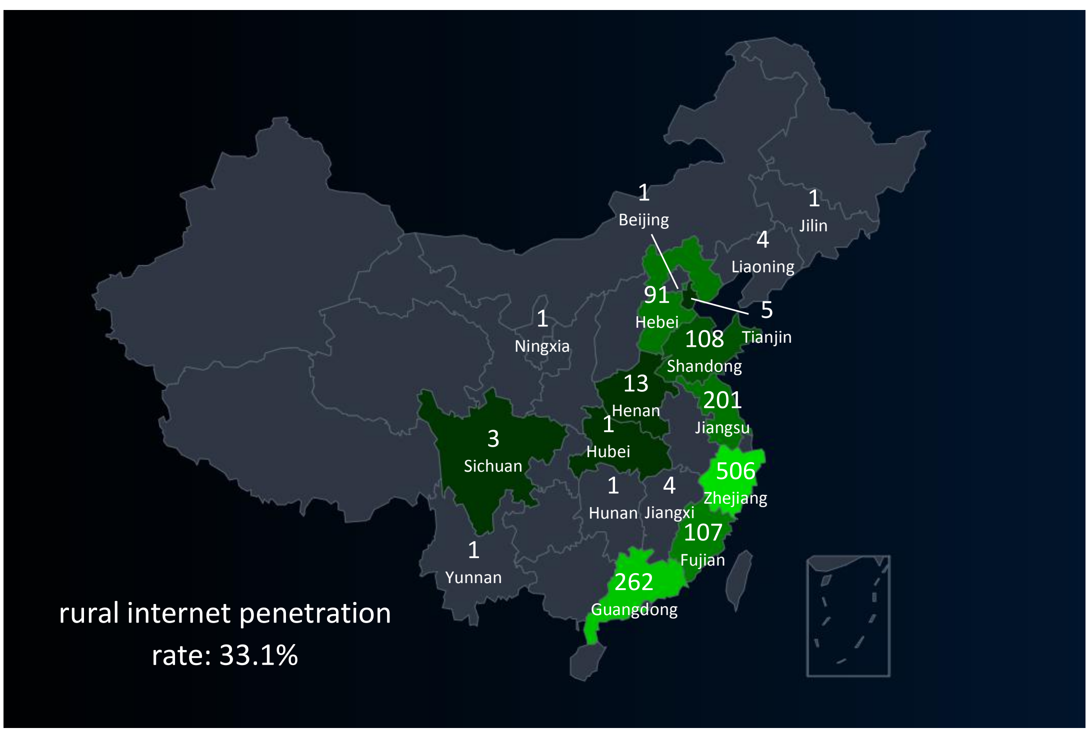

Figure 1.4 Distribution of Taobao villages in China of 2016

Chapter 2 The Effect of Agribusiness Clusters on Farmers' Migration Decisions in China

Introduction

Economic development is often accompanied by the transfer of a large number of workers from the rural-based traditional agricultural sector to the urban-based industrial sector (Demurger and Xu, 2015). This rural-to-urban migration is necessary for urban transformation to occur, which in turn is a key component of the economic growth process (Taylor and Martin, 2001). Migration is also one of the main ways poverty is reduced in developing countries, as remittances from migrant workers help reduce income inequality between rural and urban residents (Todaro, 1989; McKenzie and Rapoport, 2007; Du, Park, and Wang, 2005). The massive internal rural-to-urban migration since the mid-1980s has played a vital role in China's rapid urbanization and industrialization process. The number of migrants increased from about 30 million in the late 1980s to 130 million in 2006, and to over 220 million in 2010.[14] According to the most recent national survey, the total number of rural migrants working in cities was about 282 million at the end of 2016 (National Bureau of Statistics of China, 2017). The survey also indicated that migrants accounted for over 60% of the rural labor force at that time.

However, starting in the 1990s, more and more migrant workers in China decided to return to their home villages. For example, the number of return migrants from Guangdong province (the most popular migration destination) to Hunan province (which sends the largest number of

[14] Here, migrants are defined as people who work in a different locality than that of their registered addresses for more than six months in a year (Fan, 2009; Reuters, 2011).

inter-provincial migrants to Guangdong) increased from 28,000 to 157,000 between the censuses of 2000 and 2010. The rates of inter-provincial return migration from Guangdong to Hunan province increased from 1.6 percent (1995-2000) to 4.0 percent (2005-2010) (Liang, Li and Ma, 2014).[15]

There are several reasons migrant workers chose to stop migrating: the global financial crisis in 2008, China's household registration system, and the social problems surrounding left-behind children. First, the recession that accompanied the 2008 global financial crisis caused massive job losses for migrant workers in urban areas, and large portions of that population decided to return home after becoming unemployed. For example, Chan (2010) reported that in Guangdong province a total of 62,000 factories closed in 2008 due to lack of export demand. It was also reported that roughly 20 million workers lost their jobs in China as a whole during the recession (LaFraniere, 2009). Between December 31, 2012 and February 22, 2013, the large coastal city of Xiamen "lost" 36,200 registered migrant workers, and hourly salaries for cooks and cleaners increased steadily (Shi, 2013). The National Bureau of Statistics reported that the national migrant population decreased by 5.68 million in 2015 which was the first decline in about three decades (Minter, 2016), and the number was further reduced by 1.71 million the following year. The share of migrants as a percentage of the total rural labor force also fell from 65% in 2011, to 60% in 2016 (National Bureau of Statistics of China, 2017).

Second, the internal migration in China is unique compared to those of other countries as well as migration across country borders (Demurger and Xu, 2011). The primary reason for this is the complex household Hukou registration system in China which makes it difficult for migrants to live in cities permanently. Individuals without household registration in cities do not

[15] Note these numbers represent the percentages of migrants who were in Guangdong in 1995 (or 2005) but returned to Hunan by 2000 (or 2010), so the actual return migration rate could be even higher.

have access to many urban social services such as health insurance, schooling for their children, and social security. As a result, migrants have less access to public health facilities, poorer mental health, lower income and other non-income welfare measures (Yang, 2014a; Park and Wang, 2010). On the other hand, in 2003 the government of China established a new cooperative medical insurance which covers rural individuals without urban residency, and allows them to get reimbursed for medical spending in their home villages and cities nearby.

Third, because of the Hukou system, children of migrants tend to not have access to urban schools. These children were often left behind in rural areas, and this separation motivated their migrating parents to return. According to data from the 2010 Census, there were a total of 61 million left-behind children in rural China in 2010, and approximately 43 million of them were younger than 14. The practice of leaving children behind has been found to have negative impacts on children's physical health (de Brauw and Mu, 2011), mental health (He et al., 2012), and academic performance (Amuedo-Dorantes and Pozo, 2006; Kandel and Kao, 2001). Wang and Fan (2006) underscored family demand as an important reason for return migration, especially for female migrants. Among all family demands, taking care of the left-behind children is the most important. Previous studies show that mothers who do not take their children with them return home sooner than those who migrate with their children. A female migrant's timing of return is also strongly linked to the age of their children, as many of them choose to return around the time that their children begin formal schooling (Connelly, Roberts and Zheng, 2012).

Other reasons for return migration that have been identified and studied in the literature include the decreased urban and rural income difference due to the rising per capita income in rural areas (Demurger and Xu, 2015), the rising cost of living in urban areas (Durand, Massey, and Zenteno, 2001; Davin, 2012), and wage and job discrimination migrants face in cities (Davin, 2012). In this paper, we examine whether more agribusiness opportunities in rural areas, specifically whether the home villages of migrants being Agro-Taobao villages, is another reason for migrants to stop migrating. E-commerce has developed rapidly in China in recent years. In terms of total online sales, China surpassed the U.S. in 2013 and became the largest market for e-commerce in the world (Tan, 2015). The rapid growth of e-commerce in China has been largely driven by Taobao.com, which was founded in 2003 by Alibaba.[16] While initially most of the Taobao sellers were urban residents, the internet penetration rate in China's rural areas increased from 3.1% in 2006 to over 31.6% in 2015 (China Internet Network Information Center, 2016). E-commerce has now become one of the most popular businesses in China's rural areas due to rural areas' low labor and capital costs. Given the improved infrastructure in China's rural areas and the large market provided by Taobao, it is not surprising that more and more rural residents started engaging in e-commerce. In November 2013, there were 1.05 million online shops registered from villages and towns, representing an increase of 76.3% over the same number at the end of 2012 (Liu, Li and Liu, 2015).

Very often, many sellers from the same village started selling the same or very similar products. Then quickly, related businesses along the supply chain of the main products the village sells and other support businesses such as web page design and maintenance, shipping, financial services, photography, video and image processing, brand and management consulting and legal services appeared in the same village. Therefore, they are essentially a new generation

[16] Alibaba is the parent company of Taobao.com and many other business lines including business to business (B2B) trading platforms, online retailing, cloud computing services, and third-party payment services. In 2014, Alibaba's IPO (BABA) on NYSE raised $21.8 billion and was the biggest IPO in history at that time. As the predominant firm in China's online market, Taobao.com occupied 81.2% of the Chinese e-commerce market in terms of the gross merchandise volume (or GMV) in 2012 (Tan, 2015). In comparison, the largest e-commerce company, Amazon, only accounted for 27.1% of the U.S. online market in the same year (Yang, 2014b).

of low-cost manufacturing or factor endowment clusters that are common in a wide range of countries and sectors (Nadvi and Schmitz, 1994; Altenburg and Meyer-Stamer, 1999). Similar to the old generation of business clusters in developing countries such as the cotton knitwear industry in Tiruppur, India (Cawthorne, 1995) and the shoe industry in Sinos Valley, Brazil (Schmitz, 1995), the clusters appear and produce their goods in less-developed and low-cost countries or areas and then sell goods to more developed ones. In the case here, clusters appear in the less developed rural areas of China and their markets are mainly the more prosperous urban areas of China. Unlike the old generation of clusters, businesses in these villages rely on the internet and e-commerce platforms to market and sell their products.

Once Taobao businesses reach a certain scale within an administrative village, the village becomes a "Taobao village." More precisely, a "Taobao village" is defined as an administrative village in which the number of active online shops is over 100, or higher than the total number of households divided by 10, and the annual e-commerce sales exceed 10 million RMB (Gao et al., 2014). The number of Taobao villages had increased from 20 in 2013, to 211 in 2014, 780 in 2015, and finally 1311 in 2016 (Chen et al., 2016). Different Taobao villages usually specialize in selling different products. Those which specialize in selling agricultural products are further termed as Agro-Taobao villages. Agro-Taobao villages are essentially factor endowment agribusiness clusters as these villages often appear in regions which are especially conducive to produce certain agricultural products due to land, water, climate and other conditions in the regions. The number of Agro-Taobao villages changed from 3 in 2013, to 7 in 2014, further to 61 in 2015 and 101 by the end of 2016 (Zeng and Guo, 2015; Chen et al., 2016).

Migrating villagers from Agro-Taobao villages have several reasons to stop migrating: the lower cost of starting a new e-commerce agribusiness themselves, the special skills they possess to start such an agribusiness and job opportunities created by other e-commerce agribusinesses. First, the e-commerce agribusiness environment or culture of Agro-Taobao villages significantly lowers the difficulty for migrants to start their e-commerce agribusinesses upon returning. Migrants are able to learn from the experiences of their successful relatives and neighbors, and as a result their own agribusinesses are more likely to be successful. In addition, savings accumulated when they migrated provide them the start-up fund, further increasing the chance to start their own e-commerce agribusinesses.

Second, villagers from Agro-Taobao villages possess the skills needed to start an e-commerce agribusiness in their home villages. As discussed above, Agro-Taobao villages are essentially factor endowment agribusiness clusters for specific agricultural products and thus only a small portion of Taobao villages are Agro. Farmers from Agro-Taobao villages have the skills and knowledge on the best practice of producing the agricultural products in question. Very often, their families have grown certain agricultural goods for many generations and have gained specific knowledge about the production process for the lands they have, which is difficult for outsiders to learn and master.

Third, e-commerce agribusinesses in Agro-Taobao villages create many job opportunities, which allows return migrants to earn an income while taking care of dependents and other family members at the same time. Online shops create jobs directly as they need to employ workers to run their shops, but also indirectly in related upstream (e.g. input production, webpage maintenance) and downstream (e.g. logistics and shipping) industries. By the end of August 2016, more than 840,000 job opportunities had been created in rural areas by 300,000 active online stores from Taobao villages nationwide (Chen et al., 2016). Therefore, migrants who originated from Agro-Taobao villages can now have access to local jobs and are more likely to stay in their home villages.

In this paper, using a new and self-collected survey data, we conduct the first econometric analysis to examine whether a migrant's home village being an Agro-Taobao village is a factor for the migrant workers to stop migrating. We find that originating from an Agro-Taobao village decreases a villager's likelihood of migrating by 18%, after controlling for various factors that are likely to influence the migration decision such as the individual's demographic characteristics, migration experience, educational background, and health statuses of the individual and his/her spouse. Our study contributes to the literature that examines and identifies the factors that influence the migration decisions of farmers, especially those in China (Connelly, Roberts and Zheng, 2012; Demurger and Xu, 2011; Demurger and Xu, 2015; Chunyu, Liang, and Wu, 2013). The remainder of the paper is organized as follows. Section 2 describes our survey, introduces the dataset used in the empirical analysis, discusses the expected effects of the explanatory variables, and provides descriptive statistics. Empirical results are presented and discussed in Section 3, while the final Section concludes.

Survey and Data

The data set used in this study comes from a household survey conducted in December, 2017 in six villages of Shuyang county, which is located in the city of Suqian of Jiangsu province. Jiangsu was chosen because it is one of the most developed provinces in China. It had the second highest GDP and the highest GDP per capita among all provinces except province-level cities

Beijing, Shanghai and Tianjin from 2008 to 2016, on average.[17] E-commerce is also very active in Jiangsu. For example, on the Taobao Online Shopping Festival Day of November 11, 2016, the total gross merchandise value (GMV) of the goods transacted through Taobao.com was 120.7 billion RMB (Alizila Staff, 2016) with Jiangsu ranked as the third highest province. Furthermore, four of the top 10 counties in terms of GMV were located in Jiangsu.

Jiangsu also has many Taobao and Agro-Taobao villages. One of the three earliest Taobao villages, Dongfeng village, is located in Shaji township of Jiangsu. Furthermore, Jiangsu had the third highest numbers of Taobao villages in both 2015 and 2016 with 127 and 201, respectively (Gao et al., 2015; Chen et al., 2016). The provincial government has been very supportive of e-commerce development and has issued policies that allow registered online sellers to have access to discounted business loans (Wang, 2015). So far, Jiangsu is the first and only province in the nation to have this kind of supportive policy.

The city of Suqian was selected because it is a relatively less developed city with low income and living standards and many local residents still migrate to other cities for better employment opportunities. In 2015, Suqian had the highest proportion of workers employed in the primary industry (35.37% compared with the provincial average of 18.4%), and the highest ratio of people working in the rural area (79.99% compared with the provincial average of 54.65%). In addition, statistics from Jiangsu Provincial Bureau of Statistics (2016) show that Suqian and its rural area had the lowest disposable income per capita, and the highest Engel coefficient in Jiangsu.[18] On the other hand, e-comerce has grown very fast in Suqian. Suqian had

[17] "List of Chinese Administrative Divisions by GDP per Capita," available at: https://en.wikipedia.org/w/index.php?title=List_of_Chinese_administrative_divisions_by_GDP_per_capita&oldid=783263022.
[18] In 2015, Suqian's disposable income per capita was 17342 RMB compared with the provincial average of 29539 RMB; Suqian's Engel coefficient was 35.3% compared with the provincial average of 28.9%; the disposable income

four Taobao villages in 2014, 26 in 2015, and 51 in 2016. The rate of increase (550% and 96.2%) was much higher than that of Jiangsu province (429.2% and 58.3%), and that of China as a whole (270% and 68.1%). These statistics show that Suqian is an excellent site to conduct our research.

Within the city of Suqian, we chose six villages. As our goal is to study the effects of Agro-Taobao villages on farmers' migration decisions, we chose three Agro-Taobao villages. They are Shahe village, Weixin village and Xinhuai village of Shuyang county. The main agricultural products these three villages sell are flower seedlings as Shuyang county is called the China's capital of flower seedling. We then selected three other villages, each of which is close to one Agro-Taobao village geographically. They are Tongxi village, Yangping village and Zhouzhuang village of Shuyang county.

We surveyed the residents in these six villages. Our questionnaire asked detailed questions about family and personal characteristics, whether family members between the age of 16 and 60 were currently migrating or had migration experience before, their income and occupations while migrating, the length of migration, and reasons for returning. See the Appendix for an English version of the survey (the original survey is in Chinese). Within a village, we randomly selected households to complete the survey. We excluded households without children at home (we define children as family members who are less than 16 years old) as we are interested in studying whether Agro-Taobao villages helps alleviate left-behind children problem in many parts of rural China. For families without children or migrating with children, migration is less of a concern in this aspect.

---

per capita of rural Suqian was 12772 RMB compared with the provincial average of 16257 RMB; Engel coefficient, which defined as the proportion of income spent on food, in rural Suqian was 35.9% compared with the provincial average of 31.7%.

For each survey, the enumerator read all the questions to the respondent and wrote down the answers. Many of the respondents were seniors in the family as their adult children were away from home at the time of the survey. In total, 207 completed questionnaires were collected, 106 of which were from the three Agro-Taobao villages. Among the households in Agro-Taobao villages, 71 of them had income from an e-commerce agribusiness run through Taobao and related businesses while the rest did not. Given our focus on individual return decisions, we split household level data into individual level data since husband and wife in one household can make different migration decisions. As a result, the 207 questionnaires yielded 404 individual level observations. This became the working sample for our empirical analysis below.

Variables

From the collected survey data, we created the following dependent variable and explanatory variables for our empirical analysis. Our dependent variable *migrant* is a dummy variable that equals 1 if the individual was a migrant worker at the time of the survey and 0 otherwise. Our main explanatory variable of interest, *Taobao_village*, is also a dummy variable that equals 1 if the individual is from an Agro-Taobao village and 0 otherwise. This variable is expected to have a negative effect on the probability of migrating. As discussed above, rural e-commerce agribusinesses help narrow the income gap between urban and rural areas (Zheng, 2007) and create more employment opportunities. As a result, local residents have less incentive to migrate.

We then include a vector of personal characteristic variables that are likely to affect the migration decision. First, the variable *mother* is a dummy variable that equals 1 if the individual is a female (since we focus on households with children, a female is also a mother) and 0 otherwise. In our sample, over 69% of the respondents that never migrated are females. Therefore, the expected effect is negative as well since women are more likely to stay in their homes because of family responsibilities such as giving birth and taking care of other family members.

Second, household health status variables were created as follows: *health_ excellent* is a dummy variable that equals 1 if both the individual and the spouse reported their health status as excellent and 0 otherwise; *health_good* is a dummy variable that equals 1 if one reports the health status as good while the other reports it as excellent and 0 otherwise; *health_fair* is a dummy variable that equals 1 if at least one of the couple reported the health status as fair or bad and 0 otherwise. *Health_fair* is excluded from the regression to avoid multi-collinearity. We used the worse health condition in a couple to represent the health status of the household because 1) an individual's migration decision is often influenced by both the individual's own and his/her spouse's health condition and 2) the health status of the relatively unhealthy one is often the binding constrain the household faces. Migrants with poor health are not able to engage in highly demanding jobs such as construction, housekeeping, or waiting tables at restaurants, and are thus less likely to remain in their migration destinations. On the other hand, rural entrepreneurs, including those engaged in e-commerce agribusinesses in Agro-Taobao villages, usually need help from their spouses with the operation of their businesses, hence poor health could decrease the chance of success for their businesses and hence more incentives to migrate. Also, working in large cities provides households with poor health access to better health-care. Therefore, the expected effect of health on the probability of migrating is ambiguous.

Third, the household's education level variables were created as follows: *educ_college* is a dummy variable that equals 1 if at least one of the couple had completed college or more advanced study and 0 otherwise; *educ_secondary* is a dummy variable that equals 1 if the individual or the spouse had completed high school or middle school but neither had completed college and 0 otherwise; *educ_primary* is a dummy variable that equals 1 if neither the individual and his or her spouse had not finished middle school. Again, to avoid multi-collinearity, the variable *educ_primary* is omitted from the regression. Villagers with more education are more likely to discover the agribusiness opportunities in rural areas and also have the skills to start an e-commerce agribusiness of their own and thus they are less likely to migrate. On the other hand, more education also helps them overcome the discrimination of being migrant workers in their host cities and increases their chances of getting well-paid jobs. This makes them more likely to migrate. As a result, the expected effect of education on migration is also ambiguous.

Fourth, the variable regarding migration experience, *m_length,* is defined as the length of time spent as a migrant worker. Specifically, for migrants who had already returned, the length of migration was calculated as the total cumulated number of years spent in migration, from the year of the individual's first migration to the year he or she returned; for individual who were still migrating at the time of the survey, the length was calculated as the number of years from the individual's first migration to the time of the survey; and it was counted as zero for those individuals who never migrated. With more time spent as a migrant, the migrant usually has higher income, more friends and social relationships, and other resources in their host cities, making them more likely to continue migrating. Hence, the length of migration is expected to have a positive impact on the probability of migrating.

Finally, we created several household characteristic variables that are likely to influence the migration decision. First, the *spouse_migrant* variable, which is a dummy variable that equals to 1 if the individual's spouse was a migrant worker at the time of the survey and 0 otherwise. Individuals with a migrant spouse are more likely to migrate than those without a migrant spouse (Demurger and Xu, 2015). Connelly, Roberts and Zheng (2012) found that for women, having a migrant husband increases the length of the last completed migration episode by eight months. Therefore, the *spouse_ migrant* variable is expected to have a positive effect on the probability of migrating. Second, the variable *grandparents* is a dummy variable that equals 1 if grandparents were available to take care of the children in the home village at the time of the survey and 0 otherwise. Grandparents taking care of children makes migration easier, particularly for those having young children below the age of 12 (Demurger and Xu, 2015). We expect this variable to have a positive impact on the migration probability. Third, several children-related variables were created: the number of pre-school children 5 years old or younger is *kids_5*; the number of primary school age children between 6 and 11, *kids_11*; and the number of secondary school age children between 12 and 15, *kids_15.* Studies in the literature show that having left-behind children is one of the most important determinants of a migrant's decision to extend their length of stay in cities, as migrating parents need to accumulate money for their children's education (Demurger and Xu, 2015). On the other hand, Connelly, Roberts and Zheng (2012) report that females with left-behind children return sooner than the those without, and the difference is larger when younger children (pre-school) are present, since it is more difficult for grandparents to take care of younger children than older ones. Li (2004) finds that school performance of migrant children is adversely affected by their parents' migration, so parents have motivation to stay in their home villages to supervise the studies of their primary and secondary school-age children. Therefore, we expect the effects of all three children variables on the migration decision to be ambitious.

Summary Statistics

Table 2.1 lists the definitions of all variables used in estimation and Table 2.2 reports the summary statistics of these variables for our working sample. As Table 2.2 shows, 52.5 percent of the surveyed households come from an Agro-Taobao village, which is consistent with our survey design and 47 percent of the children's parents in these households were migrating at the time of the survey. The average age of these parents were 32.17 at the time of the survey and slightly less than half of these parents are females as there are some divorced parents in our surveyed households. In addition, on average, these parents had worked as a migrant worker for four years and had about 1.53 children younger than 16 at the time of the survey. About 45% of these parents had a migrating spouse at the time of the survey. Most of them reported excellent or good health while only 2 percent of them reported fair or bad health. In about 75% of the surveyed households, the highest education achieved is high school, while in 12 percent and 14 percent of the households, the highest education achieved are primary schools and college or above, respectively. Finally, in 48.5 percent of the surveyed households, grandparents were available to take care of the children at the time of the survey.

To further compare the differences in demographics between migrants and non-migrants, Table 2.3 reports the summary statistics of the same variables by migration status. Several interesting features of the data are worth mentioning. First, 78% of the non-migrants come from an Agro-Taobao village while only 24% of the migrants come from an Agro-Taobao village and this difference is statistically significant. This is the first piece of evidence that the home village has becoming an Agro-Taobao village is likely to have an impact on farmers' migrations decisions. Second, 59 % of the non-migrants are females while only 37% of the migrants are females. This is consistent with the fact that mothers are more likely to stay in home villages to take care of the children and other family members. Third, for 79% of the migrants, grandparents were available to take care of the children at the time of the survey while this was the case for only 21% of the non-migrants. Again, this is consistent with the observation that farmers are more likely to migrate when there are other care takers of their children. Fourth, as expected, migrants had had more years of migration experience than non-migrants and also more likely to have a migrating spouse at the time of the survey. Finally, the two groups have similar statistics regarding age, health status and education level with the exception that a statistically significantly larger percentage of farmers had completed college education or above among the non-migrants than the migrants at the time of the survey. The latter may be due to the fact that college educated farmers know how to operate a computer and run an e-commerce agribusiness and hence are more likely to stay in their home villages and run the agribusinesses.

For those non-migrants who had stopped migrating at the time of survey, we also asked them about the reason that they stopped migrating. 64.8% of them reported family-related matters as their primary reason for stop migrating, while 16.6% of them stopped migrating because they were not satisfied with the life as a migrant worker. In addition, over 13% of them stopped migrating because they wanted to start a new business, including e-commerce agribusiness in their home villages. The latter can be compared to the result from a survey of return migrants in Sichuan and Anhui province conducted by the Research Center for the Rural Economy of the Ministry of Agriculture in 1999. Results there show that at that time, only 2.6%

of the migrants returned because of the willingness to start a new business in their home villages (Wang and Fan, 2006). The difference in the percentage of return migrants citing starting a new business as their main motivation of stopping migration across the two surveys is another piece of evidence that the home village being an Agro-Taobao villages influences farmers' migration decisions as three villages we surveyed are Agro-Taobao villages.

Estimation and Results

In the previous section, our summary statistics analysis showed some preliminary evidences that the home village being an Agro-Taobao villages affects farmers' migration decisions. In this section, we further examine this by estimating the ceteris-paribus effect of home village being an Agro-Taobao village on the migration probability using regressions. As our dependent variable, the migration decision, is a binary dummy variable, we use both the linear probability and Probit models to estimate the effect.

Both the linear probability model and the Probit model have their own advantages and disadvantages. Compared with the Probit model, the advantages of the linear probability model are as follows. First, it does not require a distributional assumption for the error term hence the model is less likely to suffer from specification errors. Second, the computation of the linear probability model is more straightforward, and the estimated coefficients have the direct interpretation as the marginal effects.

The linear probability model also has several disadvantages compared with the Probit model. First, it imposes the restriction that the marginal effects of explanatory variables do not change for different values of the explanatory variables. This may not be true. Second, at extreme values of the explanatory variables the predicted probability could be less than zero or larger than one, making the interpretation difficult. For these reasons, we estimate both models and compare the results.

The estimation results are reported in Table 2.4. OLS results are collected in the first column while columns 2 and 3 report the Probit estimation results and the associated marginal effects. The results are robust across the two estimation methods, both in terms of the signs of the coefficients estimates and the statistical significance. Therefore, we focus our discussion on the Probit results below. Several results are worth discussing. First, individuals from Agro-Taobao villages are 18% less likely to start or continue migrating than those from non-Taobao villages and this effect is statistically significant at the 1% significance level. On average, engaging in e-commerce raises the rural household's annual income by 20.5 thousand RMB in China (Chen et al., 2016). According to our survey, over 61% of the respondents in Agro-Taobao villages were engaged in Taobao and related businesses at the time of the survey. Thus, it is not surprising that individuals from Agro-Taobao villages are less likely to migrate.

Second, females are 20% less likely to migrate than their male counterparts, while individuals with grandparents taking care of the children are 20% more likely to migrate. These results are consistent with our expectations, as women and the elderly in rural China are usually obligated to take care of other family members especially the children.

Third, individuals with a migrating spouse is 16.8% more likely to migrate than those without a migrating spouse. Moreover, an additional year of migration experience makes the individual 2.87% more likely to migrate. Again, these conform to our prior expectations.

Fourth, there is evidence that individuals with excellent and good health are more likely to migrate than those in fair or poor health. This implies that individuals prefer to stay at home rather than migrate to cities if their health are fair or poor, despite the better health care facilities available in cities.

Finally, the migration decision is found to be influenced by education as well. Compared with families where the highest education completed is only the primary school, those with secondary school and college or higher education are 9.64% and 13.1% less likely to migrate, respectively. These results are also consistent with our expectations. Education plays an important role in an individual's migration decision. As discussed above, education has two opposite effects on the migration decision. On the one hand, individuals with more education have the necessary skills and knowledge about computer and internet to start and operate an e-commerce and therefore are less likely to migrate. On the other hand, individuals with more education are also more likely to find better paying jobs in cities and hence are more likely to migrate. Our results here show the first effect dominates the second one. This could be explained by the recent deteriorating job opportunities for college graduates in China. For example, the unemployment rates of college graduates were more than 10% from 2007 to 2010 and nearly 100,000 new college graduates could not find jobs in 2011 (Mai and He, 2015). Some of these unemployed college graduates probably returned to their home villages and become e-commerce entrepreneurs.

Conclusions

In this paper, we investigate whether home village being an Agro-Taobao village, a cluster of e-commerce agribusinesses, influences farmers' migration decisions in China. Using a self-collected survey data from three Agro-Taobao villages and three nearby non-Taobao villages in Shuyang county of Jiangsu province, we find that originating from an Agro-Taobao village decreases a villager's likelihood of migrating by 18%, after controlling for various factors that are likely to influence the migration decision such as the individual's demographic characteristics, migration experience, educational background, and health statuses of the individual and his/her spouse.

Our results have important policy implications. Rural-to-urban migration has played an important role in the growth and development of China in recent decades. However, migration has also caused numerous social and economic problems such as overcrowding of cities, rising inequality between urban and rural areas, lack of sufficient care for children and seniors, as well as the depletion of brain and labor resources in rural areas. The findings of our study indicate that fostering development of e-commerce activities in general, and Agro-Taobao villages in particular, offers a promising way to convince villagers to stay in their home villages, and is a potential solution to the social and economic problems rural China is currently facing.

REFERENCES

Alizila Staff (2016). "Live Updates: Alibaba's 11.11 Global Shopping Festival," available at: http://www.alizila.com/live-updates-alibabas-11-11-global-shopping-festi val/ [Accessed 13 Jun. 2017].

Altenburg, T. and Meyer-Stamer, J. (1999). "How to Promote Clusters: Policy Experiences from Latin America," *World Development*, *27*, 1693-1713.

Amuedo-Dorantes, C. and Pozo, S. (2006). "International Migration, Remittances and the Education of Children: The Dominican Case," Working Paper, Department of Economics, Western Michigan University.

Cawthorne, P. (1995). "Of Networks and Markets: The Rise and Rise of a South Indian Town, the Example of Tiruppur's Cotton Knitwear Industry," *World Development*, *23*, 43-56.

Chan, K. (2010). "The Global Financial Crisis and Migrant Workers in China: 'There is no Future as a Laborer; Returning to the Village has no Meaning,'" *International Journal of Urban and Regional Research, 34*, 659-677.

Chen, L., Cui, L., Jiang, Z., Li, Z, Luo, Z., Sheng, Z., Wang, J. and Zeng, Y. (2016). "China Taobao Village Research Report of 2016," available at: http://i.aliresearch.com/img/20170414/20170414113512.pdf [Accessed 5 Jun. 2017].

China Internet Network Information Center. (2016). "Research Report on China's Rural Internet Development, " available at: http.//www.199it.com/archives/512565.html [Accessed 6 Jun. 2017]. (in Chinese).

Chunyu, M., Liang, Z. and Wu, Y. (2013). "Interprovincial Return Migration in China: Individual and Contextual Determinants in Sichuan Province in the 1990s," *Environment And Planning A, 45*, 2939 - 2958.

Connelly, R., Roberts, K. and Zheng, Z. (2012). "The Role of Children in the Migration Decisions of Rural Chinese Women," *Journal of Contemporary China, 21,* 93-111.

Davin, D. (2012). "Internal Migration and Inequality in China," available at: http://www.worldpoliticsreview.com/articles/11820/internal-migration-and-inequality-in-china [Accessed 4 Jul. 2017]

de Brauw, A. and Mu, R. (2011). "Migration and the Overweight and Underweight Status of Children in Rural China," *Food Policy, 36*, 88-100.

Demurger, S. and Xu, H. (2011). "Return Migrants: The Rise of New Entrepreneurs in Rural China," *World Development, 39*, 1847-1861.

Demurger, S. and Xu, H. (2015). "Left-Behind Children and Return Migration in China," *Journal of Migration, 4*, 1-21.

Du, Y., Park, A. and Wang, S. (2005). "Migration and rural poverty in China," *Journal of Comparative Economics, 33,* 688-709

Durand, J., Massey D. and Zenteno R. (2001). “Mexican Immigration to the United States: Continuities and Changes,” *Latin American Research Review, 36*, 107-127.

Fan, C. (2009). “Flexible Work, Flexible Household: Labor Migration and Rural Families in China,” *Research in the Sociology of Work, 19*, 377-408.

Gao, H., Chen, L., Sheng, Z., Zhang, R. and Jiang, Z. (2015). “China Taobao Village Research Report of 2015,” available at. http.//www.jianshu.com/p/6a24c126d8f8 [Accessed 5 Jun. 2017]. (in Chinese).

Gao, H., Wang, X., Chen, L., Sheng, Z., Zhang, R. and Jiang, Z. (2014). “China Taobao Village Research Report of 2014,” available at: http.//www.jianshu.com/p/f2c28a4f53a0 [Accessed 5 Jun. 2017]. (in Chinese).

He, B., Fan, J., Li, H., Wang, Y., Williams, J. And Wong, K. (2012). “Depression Risk of ‘Left-Behind Children’ in Rural China,” *Psychiatry Research, 200*, 306-312.

Jiangsu Provincial Bureau of Statistics. (2016). “Jiangsu Statistical Yearbook 2016,” available at: http://www.jssb.gov.cn/2016nj/indexc.htm [Accessed 30 Jun. 2017]. (in Chinese).

Kandel, W. and Kao, G. (2001). “The Impact of Temporary Labor Migration on Mexican Children's Educational Aspirations and Performance,” *International Migration Review, 35*, 1205-1231.

LaFraniere, S. (2009). “China Puts Joblessness for Migrants at 20 Million,” *The New York Times,* February 2, A10.

Li, X. (2004). “The investigation on the rural migrant children,” *Journal of Women Study in China, 10*, 35-37. (in Chinese).

Liang, Z., Li, Z. and Ma, Z. (2014). “Changing Patterns of the Floating Population in China, 2000-2010,” *Population And Development Review, 40*, 695-716.

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 Agricultural Research, 7*, 35-37.

Mai, Y. and He, Y. (2015). “Higher Education Expansion in China and the ‘Ant Tribe’ Problem,” *Higher Education Policy, 28*, 333-352.

McKenzie, D. and Rapoport, H. (2007). “Network Effects and the Dynamics of Migration and Inequality: Theory and Evidence from Mexico,” *Journal of Development Economics, 84*, 1-24.

Minter, A. (2016). “China's Migrants Go Home -- and Stay There,” available at: https://www.bloomberg.com/view/articles/2016-01-25/china-s-migrants-go-home-and-stay-there [accessed June 13, 2017].

Nadvi, K. and Schmitz H. (1994). “Industrial Clusters in Less Developed Countries: Review of Experiences and Research Agenda,” Discussion Paper, No. 339. Institute of Development Studies, University of Sussex, Brighton.

National Bureau of Statistics of China. (2017). “report of migrant workers of 2016” available at: http://www.stats.gov.cn/tjsj/zxfb/201704/t20170428_1489334.html [Accessed 22 Aug. 2017]. (in Chinese).

Park, A. and Wang, D. (2010). “China Economic Journal Migration and urban poverty and inequality in China,” *China Economic Journal, 3*, 49-67.

Reuters. (2011). “Factbox: China’s 2010 Census,” available at: http://www.reuters.com/article/2011/04/28/us-china-census-fact-box-idUSTRE73R0XM20110428 [accessed June 13, 2017].

Schmitz, H. (1995). “Collective Efficiency: Growth Path for Small-Scale Industry,” *Journal of Development Studies*, *31*, 529-566.

Shi, Y. (2013). “Why Chinese Migrant Workers Are Abandoning the Country's Top Cities,” available at: https://www.theatlantic.com/china/archive/2013/04/why-chinese-migrant-workers-are-abandoning-the-countrys-top-cities/274860/ [accessed July 04, 2017].

Tan, J. (2015). “Alibaba: The Giant With 81.5% Market Share In China’s Online Shopping Market,”, available at: http://www.sharesinv.com/articles/2015/11/09/ alibaba-the-giant-with-81-5-market-share-in-chinas-online-shopping-market [accessed 01 September 2016].

Taylor, J. and Martin, P. (2001). “Human Capital: Migration and Rural Population Change,” *Handbook for Agricultural Economics*. Elsevier Science, New York.

Todaro, MP. (1989).” Economic Development in the Third World,” Longman: New York.

Wang, B. (2015). “Taobao Shopkeeper can also get Entrepreneurial Loans in Jiangsu,” available at: http://www.ayrc.com.cn/news/news-show-10105.htm [accessed June 13, 2017]. (in Chinese).

Wang, W. and Fan, C. (2006). “Success or Failure: Selectivity and Reasons for Return Migration in Sichuan and Anhui, China,” *Environment and Planning, 38*, 939-958.

Yang, G. S. (2014b). “Potentials and Perils of E-Business in China,” In I. Lee, ed. *Encyclopedia of E-commerce Development, Implementation, and Management,* Hershey, PA, USA. Business Science Reference pp. 1250-1262.

Yang, X. (2014a). “Rural–Urban Migration and Mental and Sexual Health: a Case Study in Southwestern China,” *Health Psychology and Behavioral Medicine, 2*, 1-15

Zeng, Y. and Guo, H. (2015). “The Formation Mechanism of Agro-Taobao Village: A Multiple-Case Study,” *Agricultural Economy, 4*, 39-48. (in Chinese).

Zheng, Y. (2007). “Principal Component Cluster Analysis of the Development of Rural E-Commerce Regional Infrastructure,”*Forum on Science and Technology in China, 1*, 119-122. (in Chinese)

Table 2.1: Definition of Variables

Variable name	Definition
migrant	Whether individual was a migrant worker at the time of the survey
Taobao_village	Whether home village is an Agro-Taobao village
mother	Whether the individual is a female
grandparents	Whether grandparents were taking care of the children at the time of the survey
kids_5	Number of children 5 years old or younger
kids_11	Number of children between 6 and 11
kids_15	Number of children between 12 and 15
age	Respondent's age
m_length	Years spent as a migrant worker
spouse_migrant	Whether having a migrating spouse
health_excellent	Whether both the respondent's and spouse's health status were excellent at the time of the survey
health_good	Whether one of the couple's health status was excellent and the other's was good at the time of the survey
health_fair	Whether either the respondent's or the spouse's health status was fair or bad at the time of the survey
educ_primary	Whether the respondent or the spouse had completed primary school, but neither had completed secondary school or higher
educ_secondary	Whether the respondent or the spouse had completed secondary school, but neither had completed college or higher
educ_college	Whether the respondent or the spouse had completed college or higher

Table 2.2: Summary Statistics

Variable name	Mean	S.D.	min	max
Taobao_village	0.525	0.500	0	1
migrant	0.470	0.500	0	1
mother	0.488	0.500	0	1
grandparents	0.485	0.500	0	1
kids_5	0.730	0.690	0	3
kids_11	0.582	0.650	0	3
kids_15	0.218	0.448	0	2
age	32.17	6.360	21	56
m_length	4.001	3.720	0	25
spouse_migrant	0.448	0.498	0	1
health_excellent	0.691	0.463	0	1
health_good	0.290	0.454	0	1
health_fair	0.020	0.139	0	1
educ_primary	0.119	0.324	0	1
educ_secondary	0.745	0.426	0	1
educ_college	0.136	0.343	0	1

Notes: N=404.

Table 2.3: Summary Statistics by Migration Status

Variable	Non-migrants	Migrants	T-Stats for group diff.
Taobao_village	0.78	0.24	12.97***
mother	0.59	0.37	4.41***
grandparents	0.21	0.79	-14.05***
kids_5	0.75	0.71	0.68
kids_11	0.50	0.68	-2.86***
kids_15	0.21	0.22	-0.14
age	32.59	31.71	1.40
m_length	2.73	5.43	-7.83***
spouse_migrant	0.23	0.69	-10.60***
health_excellent	0.78	0.60	3.99***
health_good	0.21	0.38	-4.02***
health_fair	0.019	0.021	-0.17
educ_primary	0.070	0.17	-3.25***
educ_secondary	0.76	0.73	0.58
educ_college	0.17	0.095	2.30**
N	**214**	**190**	

Notes: * p<0.10, ** p<0.05, *** p<0.01 for two tailed test.

Table 2.4: Regression Results

Variables	OLS	Probit	Probit marginal effects
	(1)	(2)	(3)
Taobao_village	-0.259***	-0.908***	-0.179***
	(-5.75)	(-4.65)	(-5.05)
Mother	-0.202***	-1.033***	-0.203***
	(-5.42)	(-4.97)	(-5.32)
Grandparents	0.285***	1.023***	0.202***
	(6.14)	(5.03)	(5.63)
Kids_5	-0.000345	0.000555	0.000109
	(-0.01)	(0.00)	(0.00)
Kids_11	-0.0111	-0.0551	-0.0109
	(-0.35)	(-0.35)	(-0.35)
Kids_15	-0.016	-0.0871	-0.0172
	(-0.34)	(-0.35)	(-0.35)
Age	-0.00624	-0.026	-0.00513
	(-1.63)	(-1.36)	(-1.37)
M_length	0.0269***	0.146***	0.0287***
	(5.21)	(4.93)	(5.36)
Spouse_migrant	0.176***	0.852***	0.168***
	(3.79)	(3.68)	(3.81)
Health_excellent	0.257**	0.913*	0.180*
	(1.97)	(1.73)	(1.75)
Health_good	0.329**	1.356**	0.267**
	(2.51)	(2.48)	(2.54)
Educ_secondary	-0.105*	-0.489*	-0.0964*
	(-1.85)	(-1.80)	(-1.82)
Educ_college	-0.134*	-0.667*	-0.131*
	(-1.83)	(-1.87)	(-1.89)
Constant	0.415***	-0.229	
	(2.10)	(-0.25)	

Notes: *N*=404; T-statistics are in parentheses;
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Chapter 3 Estimating the Effects of Taobao Villages on Rural Net Income and Urban-Rural Income Gap

Introduction

Chinese regional policy has undergone several decades of trials, adjustments, and readjustments (Fan, 1997). During the first three decades since the establishment of the People's Republic of China in 1949, "Maoist development strategy" attempted to equalize regional economic development with collectivized agricultural production in the countryside, and concentration on heavy industry in urban areas (Long, Zou and Liu, 2009). Specifically, central government diverted investments from more developed coastal urban areas to poorer interior rural areas, not only to correct the inherent uneven growth, but also to tap into inland raw materials and resources (Yang, 1990), and foreign investment was not allowed anywhere in China. However, these policies were found to lack economic efficiency and failed to accelerate national economic growth, since support of enterprises in the developed coastal region was much more efficient than making new investments in interior areas that lacked infrastructural support.

In the early 1980s, Xiaoping Deng indicated that scarcity of Chinese capital required that it be concentrated in select coastal cities rather than allocated evenly across regions (Li, 1988). Deng's post-Mao policies have reversed the Maoist model of regional development strategy and priorities. The new strategy emphasized regional comparative advantage, and encouraged foreign investment and technological innovation. Specifically, the coastal region was selected by central government to be developed first, with a host of open zones for foreign investment and rapid economic growth (Fan, 1997). Compare with other area, those open zones are better endowed with raw materials, energy, and natural resources. These policy bring China out of economic backwardness (Yang, 1990), and the resulting growth has persisted for the last 35 years, and China’s gross domestic product (GDP) between 1983 and 2013 saw an average annual growth rate of 10.12%, making China's economy the second-largest in the world (Bajpai, 2013).

Although Deng’s policy has contributed to strong overall national economic growth, it has also created substantial inequality both between provinces, due to the enormous growth in eastern provinces, as well as between urban and rural areas (Zhao and Tong, 2000; Sicular et al., 2007). China has gradually become a country with diverse regional economic development levels and incomes. There is a widely accepted view that the urban-rural income gap is the major driver of income inequality in China (Jiang et al., 2011; Sicular et al., 2007, Su et al., 2015). The per capita disposable income in urban areas was, on average, 2.63 times that of rural areas between 1978 and 2010, and has lingered around 3.00 since 2000; the Gini coefficient of China increased from 0.32 in 1978 to 0.55 in 2012 (Xie and Zhou, 2014). Also, in 1978, the coefficient of variation of per capita rural income among 28 provinces in China was 0.33, but increased to 0.49 in 1993 (Meng and Wu, 1998). These numbers are both very high by international standards.

In China, the large rural-urban income gap would lead to a sectoral reallocation of labor from rural to urban areas, which could potentially cause various social and economic problems including: 1) left-behind children and empty nest elderly, as most of internal migrants are between the age of 16 and 50 (Yang and Zhou, 1999); 2) rural human capital drain, as migrants tend to be more talented and productive (Gao, Zheng and Bu, 2014). Because of the rising number of internal migrant workers, income inequality will also endanger social stability[19], and make it more difficult for the left-behind poor in rural areas to escape relative poverty (Su and Heshmati, 2013). Furthermore, internal migrants in urban China tend to be unhappier, more likely to have symptoms of depression, and have a lower quality of life on average than local residents (Qiu et al., 2011; Wang et al., 2010). One part of the explanation is the different distribution of time in daily life: Migrants allocated less time to active leisure, as well as activities outside of home, work, and transit (Hendriks,Ludwigs and Veenhoven, 2016), and because of their stressful lives, migrants have less mental and physical energy to engage in activities that promote happiness. Given these problems caused by income gap, it will be important to examine and identify factors that can contribute to a rise in rural income and a reduction in the urban-rural income gap.

Studies find E-commerce to be an effective method to raise income in many countries. On average, holding other factors constant, engaging in e-commerce raises a household's annual income by 20.5 thousand RMB in China (Chen et al., 2016); In the U.S., e-commerce retail jobs pay, on average, about 30 percent more than traditional retail jobs (Sorkin, 2017); the average income of online business owners was over 59,000 dollars in 2016, while the average wage for department store workers was only slightly above 20,000 dollars (Mu, 2017); Anvaria and Norouzi (2016) also showed that e-commerce in 21 selected European countries[20] played a significant positive role in per capita income and GDP. Therefore, rural e-commerce is expected to be a potential way to raise rural income, and reduce rural-urban income inequality in China.

[19] In 2013, the proportion of crimes committed by migrants accounted for more than 70% of the total number of crimes in China (Chang, 2016).

[20] Austria, Belgium, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Luxembourg, Netherlands, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, and United Kingdom

E-commerce, and rural e-commerce in particular, have recently developed rapidly in China. In 2013, China's total online sales surpassed that of U.S. and became the largest market for e-business in the world (Tan, 2015). Growth thus far has mostly been driven by the largest online trading platform, Taobao.com, founded in 2003 by Alibaba. Running an e-business requires the seller to have internet access, and in 2006, internet penetration rate in China's rural areas was only 3.1%, which meant very few people in these areas had access to the internet. Thus it is not surprising that in the early stages a large percentage of Taobao online sellers lived and operated their businesses in cities. However, because of improving rural infrastructure, internet penetratio rate increased to over 31.6% by the end of 2015 (China Internet Network Information Center, 2016). Given the substantial online market mainly provided by Taobao.com, it is not surprising that more and more rural residents started to engage in e-commerce, and e-business gradually became one of the most popular businesses in rural China, as it usually has low labor and capital requirements. In November 2013, there were 1.05 million online shops registered in villages and towns, representing an increase of 76.3% compared to the number at the end of 2012 (Liu, Li and Liu, 2015). Most of these online sellers were selling on Taobao.com, since Taobao, and its associated website Tmall, have, in term of gross merchandise volume, occupied at least 75% of the online shopping market in China since 2011 (Tan, 2015). Additionally, China's rural villages are often regarded as "acquaintances societies," which have their own widely accepted rules and the information is relatively public (Chen, Zhang and Li, 2014). Once an individual in a village is operating a successful Taobao store, his/her experience will be learned by neighbors and relatives in the same village very quickly. Because of the low entry barrier to starting a Taobao store, many of these neighbors and relatives will soon start Taobao businesses of their own, which will then attract even more people to the business. When Taobao businesses reach a certain scale within a village, the village becomes a rural e-commerce cluster, or Taobao village. Alibaba defines a Taobao village as an administrative village where the number of active online shops is over 100, or accounts for more than 10% of the total number of households, and the annual e-commerce sales exceed 10 million RMB (Gao et al., 2014). The first Taobao village appeared in 2009, and the number increased to 20 in 2013, 211 in 2014, and 780 in 2015 (Chen et al., 2016).

Taobao villages are expected to affect income in rural China in two ways: by helping to increase the income of e-commerce employers through Taobao businesses, and by increasing employees' income through job opportunities that Taobao and related industries create in Taobao villages. First, the e-commerce environment of Taobao villages helps employers start their own e-commerce and related businesses. In Taobao villages, villagers can learn from the successful experiences of their relatives and neighbors, and as a result, their businesses are more likely to be successful. Rural Taobao business helps to raise villagers' income, which would attract internal migrants to return to their home village. Take two famous Taobao villages: Junpu village[21] and Dongfeng village[22] as examples, in both of which Taobao businesses within the village were started by returning migrant workers. On December 2017, we conducted a survey in three Taobao villages and three non-Taobao villages located in Shuyang of Jiangsu province. We found that 81.1% of households that engaged in Taobao and related businesses have at least one

[21] Junpu village is located in Jieyang, Guangdong, with a total area of 0.53 square kilometers, 490 households, and a population of 2,695 in 2014. The Taobao business in Junpu was started by migrant worker Haijin Huang and his friends in July 2012 and became one of the first Taobao villages. By the end of 2014, 348 of the 490 households, and nearly 2000 villagers in Junpu were engaged in e-commerce or related businesses. They opened more than 3,000 Taobao shops which had annual sales of 1.68 billion RMB, and net profit of 15 million RMB.

[22] Dongfeng village is located in Xuzhou, Jiangsu, with a total area of 6 square kilometers, 1180 households, and a population of 4,849 in 2014. It is also one of the first Taobao villages. The first Taobao shop in Dongfeng was opened by migrant worker Han Sun in July 2006. In 2014, over 60% of the households were participating in Taobao. They opened more than 3,000 Taobao shops which had annual sales that over 1.2 billion RMB.

return migrant worker. This number is not surprising since return migrants are more likely to opt for higher ability jobs than non-migrants, and both the savings they accumulated and the frequency of job changes during migration increase the likelihood that return migrants will become self-employed (Demurger and Xu, 2011). Ma (2001) also found that the skills and entrepreneurial abilities acquired by migrants in the urban labor market can greatly facilitate their occupational transition into non-farming jobs after returning home.

Second, job opportunities created by Taobao and related business in Taobao villages can raise employees' income. From December 2007 to May 2017, the e-commerce industry created 397,000 jobs in the United States, while the traditional retail industry lost 76,000 jobs (Sorkin, 2017). Online shops create jobs both directly, as they need to employ workers to run their shops, but also indirectly in related upstream (e.g., input production, webpage maintenance) and downstream (e.g., logistics and shipping) industries. These jobs usually provide salaries much higher than traditional agricultural jobs. By the end of August 2016, 300,000 active online stores from Taobao villages created more than 840,000 direct job opportunities in rural areas by (Chen et al., 2016), E-commerce and related businesses created over 20 million jobs nationwide in China, and the stock keeping unit (SKU), which is a product and service identification code for a store or product, of rural online retail reached 293 million, which accounted for 20.3% of the total online retail in China (Meng, 2017).

By August 2016, there were over 1,300 Taobao villages in China, and that number is expected to continue to rise. Thus, it will be critical to clarify what specific benefits Taobao villages will bring to their residents. We focus on rural net income, the most important benefit for rural residents, as well as the rural-urban income gap which is also one of the most serious issues facing rural families in China. Previous studies show that participating in Taobao or related businesses provides a potential source of alternative income for people in rural areas. Studies of clusters in other industries and countries have also shown that clusters benefit locals by increasing income (Guo, Liang and Luo, 2014; Fleisher et al., 2010; Martin et al., 2015; Rivera, Gligor and Sheffi, 2016). In this paper we focus on the effect of Taobao villages on rural regional economies at county and district level. Specifically, we try to answer the following question: Do rural residents living in or close to the counties with Taobao villages earn more on average, and face a smaller intra-county urban-rural income gap than those living in other counties, after controlling for various other factors which affect rural net income? Existing literature foucses on explaining either determinants of rural-urban income gap (Su and Heshmati, 2013; Sicular et al., 2007; Gao, Zheng and Bu, 2014; Wei and Zhao, 2012; Jiang et al., 2011) or the development of rural e-commerce in China (Guo, Liang and Luo, 2014; Wu and Huang, 2014; Liu and Liu, 2015; Li, 2017). This paper attempts to explain the composition of intra-county income gap by rural e-commerce growth, using China's county and district level panel data from 2012 to 2016. It is also the first study to provide an econometric analysis of the determinants of China's urban-rural income gap at county and district level.

The remainder of the paper is organized as follows. The dataset used in the empirical analysis is laid out in Section two. Section three introduces our empirical strategy, discusses the expected effects of the explanatory variables, and provides descriptive statistics, while Section four presents and discusses the results. The final section concludes this chapter.

Data

We used county and district level panel data of five provinces of China from 2012 to 2016. They are Zhejiang, Jiangsu, Shandong, Fujian, and Hebei which are consistently five of the top six provinces in terms of number of Taobao villages. Guangdong provinces are excluded from our research because their provincial and municipal Bureau of Statistics did not collect rural per capita net income at county and district level. Each year these five provinces contain over 75% of the total Taobao villages in China (Chen et al., 2016). In China, both the disposable income per capita of urban residents and the net income per capita of rural residents reflect their real levels of incomes, respectively (Cai, 2005). Given this, we will take rural net income as one of our dependent variables, and use the two income variables to calculate the urban-rural income gap, which is our second dependent variable. Our major data source, including our dependent variables, is statistical yearbooks published by provincial and municipal governments. The statistical yearbooks are edited and published by local Bureaus of Statistics which report statistics on population, economy, employment, investment in fixed assets, energy, transportation, finance, foreign economic relations, wages and other income, household characteristics, as well as other social economic variables at the province, city, and county level. We chose this this time period because 2012 was the first year Alibaba began to both track the number of Taobao villages, and publish a list of them. The yearbooks of 2016 are the newest available.

Our key variable of interest, the number of Taobao villages in each county, was constructed using the lists of Taobao villages published annually in Taobao Villages Annual Research Reports by Alibaba since 2012 (Gao et al., 2014; Gao et al., 2015; Chen et al., 2016). Furthermore, to examine potential influence of Taobao villages in neighboring counties, for each county in each year, we also counted the number of Taobao villages in all surrounding counties.

The surrounding counties are defined as those that share a common border with the county of interest.

The administrative divisions of China are different than most other countries. Each province in China consists of five to 23 prefecture-level cities, each of which in turn consists of two to 15 districts and counties. There are 200 to 300 administrative villages in a county or district on average. Within a prefecture-level city, districts are usually located in urban areas, while counties are usually located in rural areas. In addition, many districts do not have their own Statistics Bureau, so those districts will not have data at county and district levels. Therefore, we excluded the districts that lack data, and those that cover little no or rural areas since our study focuses on Taobao villages, which by definition, are located in rural areas. Table 1 lists all the districts that were dropped.

Empirical Strategy

Since the data are panel data, pooled OLS, random effect and fixed effect panel data models will be used. The focus will be on interpreting the fixed effect model while we estimated other two models for robustness check.

The Rural Income

The regression equation for rural income is $\ln rinc_{it} = \alpha_i + X_{it}\beta + T_{it}\gamma + u_{it}$. The dependent variable is nature log of per capita rural net income for county i at year t. However, it is not precise enough to estimate using the number on the yearbook if we fail to consider the spatial differences in the cost of living among different counties and districts (Yang and Zhou, 1999).

Therefore, we deflated the nominal rural per capita net income into the real value with the consumer price index (CPI) of each prefecture- level city which is also shown in the yearbooks. Specifically, we set 2013 as the base year, which means the CPI of all counties/districts in our sample equal to 100 in 2013.

Let X_{it} denote a vector of control variables that vary across both counties and years, T_{it} denote the two Taobao village variables, which are the number of Taobao villages in one county and all counties nearby, α_i the unobserved time-invariant individual effect that only varies across counties, and u_{it} denote an idiosyncratic error term.

The Intra-County Rural-Urban Income Gap

Previous studies about rural-urban income gaps in China (Su and Heshmati, 2013; Sicular et al., 2007; Gao, Zheng and Bu, 2014) usually estimated the income gap by the ratio of the per capita disposable income of urban residents to the per capita net income of rural residents. The intra-county rural-urban income gap (also called the observed urban-rural income ratio) for county i at year t gap_{it} can be calculated as:

$$gap_{it}=unic_{it}/rinc_{it}$$

where $uinc_i$ represents per capita disposable income of urban residents, and $rinc_i$ denotes the per capita net income of rural residents in county i. Similarly, in order to calculate income gap precisely, we deflated both nominal rural and urban per capita income using the CPI as the proxy for inflation rate. Next, we take the nature log for both sides of the equation above: $\ln gap_{it}=\ln unic_{it}-\ln rinc_{it}$, and obtain the regression equation for income gap:

$\ln gap_{it} = \alpha_i + X_{it}\beta + T_{it}\gamma + u_{it}$,where explanatory variables on the right-hand side are identical with the regression equation for rural income.

Explanatory Variables

In this section, we discuss the explanatory variables used in the estimation, and briefly describe how each variable is expected to affect the rural net income and the urban-rural income gap. Table 2 shows the definitions of explanatory variables of our empirical models. They were chosen according both to economic theories and previous empirical works. First, it is important to control for the time trend, and some of the most basic differences among counties. As our data come from five provinces over four years, we create four provincial dummy variables to control for factors at the provincial level that determine the rural income and income gap, and define the trend variable of years 2013 to 2016 as 1 to 4, respectively. Also, we include the variable $density_{it}$, which is defined as the population density in the county (in thousands of people per squared kilometers) into the estimation. Since Glaeser, Resseger and Tobio (2008) found a 45 percent correlation between population density and income inequality, measured with the Gini coefficient across all counties in U.S., we expect the effect to be similar for China, and counties with high population density tend to have a large urban-rural income gap. In addition, the time-invariant provincial dummy variables are subsumed in the county-specific effects in fixed effect models, thus those dummies will be excluded from the FE estimation.

Second, our variables of interest, the number of Taobao villages in each county, n_taobao_{it}, and the number of Taobao villages in all surrounding counties, nei_taobao_{it}, could both raise China's rural income significantly because of the self-employment and income-raising opportunities which Taobao villages provide. On the other hand, the number of Taobao villages has very little effect on changing the urban income, thus these two variables are expected to have negative impact on the urban-rural income gap. We further create a new variable, $taobao_{it}$ $=n\_taobao_{it}+nei\_taobao_{it}$, which is defined as the number of Taobao villages in one county and all other counties that share a common boarder. The expected effect of $taobao_{it}$ will be the same as n_taobao_{it} and nei_taobao_{it}.

Third, education level is one of the most important determinants of one's level of income, while human capital is also significant for economic growth (Barro, 1991). In our paper, we use the number of high school students per thousand people in the county, $high_{it}$, as a proxy for the county's average level of education and human capital. It is expected to have positive effects on both urban and rural income, however, the positive effect will be larger on the urban income since high school graduates are more likely to work in urban areas as most of the jobs that require more education are in urban areas. Therefore, in our sample, the higher regional average level of education is expected to widen the income gap.

Fourth, we include several county level economic indicators in the estimation to control for differences in levels of economic development among counties. Lu and Chen (2004) claimed that governmental fiscal expenditure is one of the reasons urban-rural income gap widened in China between 1987 and 2001. Within this time period, the primary goal of the local government is to promote regional GDP, which is predominantly contributed by non-farm industries rather than agriculture. That is why urban cities were guaranteed a priority in the support of local fiscal expenditure, and cities will benefit more from a higher ratio of local fiscal expenditure in regional GDP, which will widen the income gap between urban and rural areas. Therefore, we used the lagged proportion of local fiscal expenditure to GDP, $fiscal_{it-1}$, in our regression as a proxy for government intervention. This variable could also measure fiscal policies and the effectiveness of government (Burnside and Dollar, 2000). The reasons we lag the fiscal expenditure variable are: 1) most of the projects that government funded takes a while to complete, and those projects usually start to affect the per capita income when they are completed. 2) in order to avoid the endogeneity concern caused by reverse causality as rural income and urban-rural income inequality are important considerations while government making the decision of expenditure. The lagged total amount of outstanding loans lent by local banks as a share of GDP, $loan_{it-1}$, is also included as a control. When the local economy experience a recession, both the rural per capita net income and urban per capita disposable income will drop, and the rural-urban income differential will widen since urban income usually drop faster. In order to stimulate the economy and raise income, local government is likely to increase loans and relax the restriction of obtaining loans. The rural income and income inequality will affect the amount of outstanding loans, thus the variable is suffered from the reverse causality and we need to lag it to avoid the endogeneity. This variable could be seen as a proxy for financial sector development (Anwar and Cooray, 2015). Lu and Chen (2004) also found evidence that, between 1987 and 2001, financial development helps widen the income gap. The major reason is China's financial development was focus on cities and large companies especially since 1990s. Due to a relative elevated risk in agriculture, there are usually information asymmetry between rural households and financial institutions. Meanwhile, financial institutions are low in efficiency of providing support for agriculture as well as rural areas. Therefore, they are reluctant to grant a loan for rural households. By contrast, these institution are more willing to provide financing services for large and medium-sized enterprises. Since these enterprises are mostly clustered in urban, financial development in China will broaden the urban-rural income gap. According to previous studies, we expect they will have positive relationships with both dependent variables.

Fifth, Zhou (2009) provides empirical evidence supporting the claim that business and trade openness widens the income gap. In order to reduce the impact of the difference in openness, we also add foreign direct investment (FDI) as a proportion of GDP, $foreign_{it}$, to the empirical estimation[23]. A higher value for a county or district indicates a more open business and trade environment and a more developed regional economy. Existing literature (Nair-Reichert and Weinhold, 2001; Balasubramanyam, Salisu, and Dapsoford, 1999) argues that there is a positive relationship between FDI and economic growth. Given the obvious relationship between economic growth and income growth, the foreign trade openness should have a positive relationship with the rural income. Moreover, very few enterprises with foreign investment are located in rural areas, so the impact of foreign direct investment is expected to be larger on urban income, and the impact on income gap will also be positive.

Sixth, infrastructure, especially that in rural areas, will help narrow the urban-rural income gap given that these projects usually require having to hire many rural workers with good pay (Lu and Chen, 2004). The variables about infrastructure which we introduce to our model are: 1) lagged fixed asset investment to GDP ratio, $asset_{it-1}$, which represents relatively how much the government spends on an economic industrial basis since fixed asset investment refers to investments in infrastructure and real estate. Similar with the fiscal expenditure, the fixed asset investment usually take some time to affect income except the salaries of the workers. The variable is also likely to suffer from the reverse causality problem, since local government and

[23] FDI is originally in dollars in the yearbooks. We converted it to yuan using the annual average exchange rate.

enterprises will make their decision of investment according to the level of development of rural and urban area which can be represented by the per capita income; and 2) per capita mileage of roads, $mile_{it}$, which also represents the development of rural infrastructure since most new roads are built in rural areas. Agrawal, Galasso and Oettla (2016) find that a 10% increase in a region's highway mileage causes a 1.7% increase in regional innovation growth over a five-year period, suggesting that the length of road per capita may be positively correlated with regional growth. These two variables are thus expected to have positive impact on rural income and negative impact on urban-rural income gap.

Finally, the GDP share of primary and tertiary sectors, $prim_{it-1}$ and $tert_{it-1}$, are included as control variables. These two variables represent the importance of agriculture and service industries in the local economy. However, in China, the compositions of urban disposable income and rural net income are quite different: in 2011, agricultural income accounted for 36.1% (2520 out of 6977 yuan) of the rural net income, compared with 0% of urban disposable income. On the other hand, 87% of the decline in China's rural poverty between 1980 and 1995 can be attributed to non-farm activities, and rural poverty and income inequality would also be much higher and deeper without non-farm employment (de Janvry, Sadoulet and Zhu, 2005). Because non-farm employment usually bring villagers higher income compare with traditional agricultural work, rural individuals living in the counties with low rural per capita net income or large urban-rural gap tend to participate more in non-farm activities. The GDP share of primary sector of that county will fall, while the share of other two sectors will raise since there are less villagers engage in farming. Therefore, we lag these variables which are likely to be endogenous in the regression due to reverse causality. We omit the secondary sector from our regression. Existing literature claim that primary and tertiary sector growth has had greater impact on poverty and inequality than secondary sector growth (Ravallion and Datt, 1996); in rural China, growth in the primary sector (primarily agriculture) did more to reduce poverty (raise rural income) and inequality (narrow the urban-rural income gap) than either the secondary or tertiary sectors (Ravallion and Chen, 2007; Chaudhuri and Ravallion, 2006); the impact of growth in the primary sector on poverty reduction was roughly four times higher than the other two sectors (Chaudhuri and Ravallion, 2006). The major reasons agriculture could help to reduce poverty are: 1) the incidence of poverty tends to be higher in agricultural and rural populations; and 2) most of the poor live in rural areas and a large share of them depend on agriculture for survival (Christiaensen and Demery, 2007; Ravallion and Chen, 2007). On the other hand, the non-farm activity becomes increasingly important in rural China, the share of non- farm income was rising from 57.9% in 2007 to 63.9% in 2011, and as menitioned before, off-farm employment is the major contributor of the reduction of rural poverty and urban-rural income inequality. The reason for the growth in off-farm employment in rural China is straightforward: the worker moves when her/his contribution to the household income is higher in off-farm employment than if s/he stays and works on the farm (Su et al., 2016). Therefore, the effects of GDP share of primary sector (agricultural development) on both rural income and urban-rural income gap are ambiguous. Furthermore, the tertiary sector is the major source of job opportunities and income growth in both urban and rural China since the 1980s. It has doubled in size over the three decades since 1983, and accounted for roughly 46% of GDP in 2013 (Bajpai, 2013). However, it is hard to determine and compare the magnitude of income growth by the developing service sector for rural and urban areas. Hence, we expect the effects of tertiary sector's GDP share on rural income and income distribution to be positive and uncertain, respectively.

We are interested in whether lagging some of the variables discussed above by one period solve the problem of endogeneity. One way to do so is to test for whether there is serial correlation in the error term u_{it} above. If there is still serial correlation in u_{it}, then the lagged variables are still correlated with the error term and violates the exogeneity assumption. We apply the method of Wooldridge test for autocorrelation from Drukker (2003) and the null hypothesis that there is no first-order autocorrelation is rejected at 10% level. The problem might be solved by lagging those variables for two or more periods, but this can't be done because we will not have enough observations if lagging the variables further. This is a shortcoming of my study and results below should be interpreted with caution.

Empirical Results

The estimation results are reported in Tables 3 and 4. Our working sample suffers from missing values caused by the unavailability of data (e.g. there are a number of cities in Shandong province that did not report their mileage of roads and urban per capita disposable income at county level; similarly, some counties of Hebei provinces lack data about the total amount of loans given by local banks). In order to solve this issue, we apply the method of listwise deletion which is the default for handling missing data for a lot of software including SAS and STATA, so there are 1117 and 1053 observations left for estimations of rural income and income gap, respectively.

We first discuss the estimation results of real rural income reported on Table 3.3. As the table shows, most results are robust in terms of magnitude and significance to the estimation method used. The coefficients of the four province dummy variables are all negative for both the

OLS and random effect models, which means the province we omitted: Zhejiang has the highest average per capita rural net income among the five provinces. Zhejiang had been the top among all provinces in terms of rural net income for 32 years beginning in 1983, and it is the first province ever to have a rural net income over 20,000 yuan. The underlying reasons are: 1) the off-farm income accounted for 61.95% of the net income, the share of which is about 20% higher than the national average; and 2) the off-farm employment ratio is 85.3% which is also the highest in China. Next, we look at the time trend variable. The coefficient of the fixed effect model can be interpreted as: on average, the deflated per capita rural net income grew by 6.68% each year between 2013 and 2016 in our sample. Given that the annual growth rate of real rural net income in China within the same time period was 5.38%, we can say the villagers' income grew faster in the five province we selected.

That the coefficient of $taobao_{it}$ is positive and significant even at the 1% level, shows that rural Taobao business was successful in poverty reduction. The estimation of the fixed effect model gives us the idea that an increase of one Taobao village in a county, or all counties nearby, will cause a 0.0828% rise in the real rural net income. According to the study of the effect of Taobao villages in the previous chapter, Taobao and related businesses provide opportunity for employment and self- employment, which motivates migrant workers to return home. Furthermore, encouraging migrants to return to their rural home in order to start businesses, including Taobao, is found to be one solution for solving the issue of rural poverty in China (Demurger and Xu, 2011).

Although the reason for adding the variable $density_{it}$ is to control for the natural differences among counties, its positive and significant coefficient is worth discussing. In response to an increase in population density of 1000 per square kilometers, the rural net income will raise by 8.36%. The U.S. Department of Veteran Affairs determines an area is either "Urban," "Rural," or "Highly Rural" according to its population density[24], and a high population density will reflect a relatively low level of rurality. Villagers who live in areas with a high degree of unurbanized space have little access to income-raising opportunities, which could potentially explain the positive correlation between rural income and population density.

The coefficients of $prim_{it-1}$ and $tert_{it-1}$ provide evidence that the lagged GDP share of primary and tertiary sectors will affect rural income. The fixed effects estimation results show that a one unit increase in the percentage of primary share last year will cause a 0.153% decline in rural net income within a county, while the same amount of increase in tertiary in the previous year will result in a 0.111% growth in rural income this year, and both coefficients are significant at 10% level. These results are consistent with our expectations. The negative coefficient of primary share leads us to believe that in the five provinces we selected, the non-farm sector dominated the agricultural sector in terms of the effect on poverty reduction.

Lastly we expected government intervention to have a positive effects on rural income growth. The coefficients of the fixed effect model suggest that a one percent improvement on the proportion of local fiscal expenditure to GDP last year, $fiscal_{it-1}$, will raise the rural net income this year by 0.582%. In addition, the results of this variable are not robust between the random effects models and the fixed effects model. The discrepancy can be attributed to the random effects models assuming the unobserved county level effects are uncorrelated with other explanatory variables, while the fixed effects model does not rely on this assumption.

---

[24] **Urban Area**: Census Bureau-defined urbanized area, which is any block or block group having a population density of at least 1000 people per square mile; **Rural Area**: Any non-urban or non-highly rural area; **Highly Rural Area**: An area having < 7 civilians per square mile.

The estimation results of urban-rural income gap when adjusted for inflation is reported in Table 4. As we can see, the results are quite different from the previous regression. First, the urban-rural income gap of Zhejiang is not significantly different from Jiangsu for both OLS and random effect models, but the negative coefficients of the other three provinces in the OLS model suggest that Zhejiang has a narrower income gap compared with those three provinces. This is not surprising since in 2016 Zhejiang and Jiangsu provinces had the lowest ratios of urban and rural disposable income among all provinces in China at 2.07 and 2.28, respectively.

Second, the estimation of time trend provides evidence that, between 2013 and 2016, the urban-rural income ratio of the counties in our sample decreased by 1.59% each year on average. This result is consistent with the fact that the urban-rural income ratio of China reached 3.33 in 2009, fell each year after that, and finally decreased to 2.72 by the end of 2016.

Third, the coefficient $taobao_{it}$ is negatively significant at a 1% level, showing that rural Taobao could also effectively improve income distribution. Unlike our interpretation in the previous section, this number can be explained as follows: for every extra Taobao village in a county we expect a decrease in the income ratio of the county, and others in the area, by 0.0598%. Our finding is consistent with both our expectation and existing literature which shows that rural e-commerce business helps to narrow the income gap between urban and rural areas (Zheng, 2007).

Fourth, as previoulsy discussed, population density can reflect the level of urbanization. The coefficient of the fixed effect model is significant at a 1% level, which we interpret to mean an increase in population density of 1000 per square kilometer will cause the urban-rural income ratio to fall by 10.8%.

Fifth, areas with more high schools tend to have a larger gap between urban and rural disposable income. In counties within the five selected provinces, each added high school student per thousand civilians will lead to a 0.0988% increase in income ratio. Because of there are only a few high schools are located in rural areas. Besides, Han, Li, and Zhang (2014) claim that improving high-level education is less effective in lowering the national income inequality than promoting basic education exclusively in rural areas.

Finally, it is interesting to find that an increase in fiscal expenditure actually helps equalize the income distribution, since a one unit increase in the lagged proportion of local fiscal expenditure to GDP will result in a 0.207% fall in the urban-rural income ratio. The estimation results conflict with our expectation and the findings of Lu and Chen (2004). However, China's government started to invest more in rural areas in the early 21st century, especially after the global financial crisis of 2008. The different ways governments arrange their expenditures are one explanation for this contradiction. Other studies also found positive relationships between urban-rural income gap and fiscal expenditure in Australia (Gunasinghe, 2016), Pakistan (Bhatti, Batool and Naqvi, 2015), and European countries (Brinca, 2017).

Conclusion

In this paper, we use panel data models to examine the effects of Taobao villages and various economic indexes on rural net income and urban-rural income inequality. We find that the development of Taobao villages is an effective method to raise income for rural residents, and to narrow the income gap between urban and rural China, while noting that fiscal expenditures play a similar role. Also, compared with the agricultural sector, the development of non-farm sectors is a major contributor to rising incomes in rural China. In addition, county's average education level will actually widen the income gap.

Our results have important policy implications. In order to achieve the goal of reducing rural poverty and income inequality, local governments should foster development of rural e-commerce activities, increase fiscal expenditures, and invest in non-farm sectors over agriculture for rural areas.

For future work, an ideal dataset will include the number of Taobao villages and the economic and infrastructural indexes for only rural portions of China. With that data, we could focus on how Taobao villages affect rural areas of the country. On the other hand, similar to what is mentioned in the first chapter, Taobao transaction volume within the county could better represent the level of development of Taobao and related industries. However, we do not currently have access to that data. If the yearbooks begin to collect data about the county's rural economy, and Taobao transaction volume becomes available in the future, a more detailed study can be conducted to examine how Taobao businesses change villagers' lives, and China's economic structure.

REFERENCES

Agrawal, A., Galasso, A. and Oettl, A. (2016). “Roads and Innovation,” *Review of Economics and Statistics, 99,* 417-434

Anvari, R. and Norouzi, D. (2016). “The Impact of E-commerce and R&D on Economic Development in Some Selected Countries ,” *Procedia Social And Behavioral Sciences, 229*, 354-362.

Anwar, S. and Cooray, A. (2015). “Financial Flows and Per Capita Income in Developing Countries,” *International Review of Economics and Finance, 35,* 304-314.

Bajpai, P. (2013). “China's GDP Examined: A Service-Sector Surge,” available at: https://www.investopedia.com/articles/investing/103114/chinas-gdp-examined-servicesector-surge.asp [Accessed 21 Feb. 2018].

Balasubramanyam, V., Salisu, M. and Dapsoford, D. (1999). “Foreign Direct Investment as an Engine of Growth,” *Journal of International Trade and Economic Development, 8*, 27-40.

Barro, R. (1991). “Economic Growth in a Cross Section of Countries,” *The Quarterly Journal of Economics, 106,* 407-443.

Bhatti, A., Batool, Z. and Naqvi, H. (2015). “Fiscal Policy and Its Role in Reducing Income Inequality: A CGE Analysis for Pakistan,” *The Pakistan Development Review, 54*, 843-864.

Brinca, P., Ferreira, M., Franco, F., Holter, H. and Malafry, L. (2017). “Fiscal Consolidation Programs and Income Inequality,” *Working paper*, Universidade Nova de Lisboa, Lisbon.

Burnside, C. and Dollar, D. (2000). “Aid, Policies and Growth,” *American Economic Review, 90,* 847-868.

Cai, F. (2005). “Why did not Labor Migration Bridge Urban-rural Income Gap?”*Theoretical Front, 20*, 18-20. (in Chinese).

Chang, H. (2014). “Report on the crime of migrant workers in China,” available at: http://world.people.com.cn/n/2014/0818/c1002-25488734.html [Accessed 23 Oct. 2017]. (in Chinese).

Chaudhuri, S. and Ravallion, M. (2006). “Partially Awakened Giants: Uneven Growth in China and India,” *Policy Research Working Paper, No. 4069*. World Bank, Washington, DC.

Chen, L., Cui, L., Jiang, Z., Li, Z, Luo, Z., Sheng, Z., Wang, J. and Zeng, Y. (2016). “China Taobao Village Research Report of 2016,” available at: http://i.aliresearch.com/img/20170414/20170414113512.pdf [Accessed 5 Jun. 2017].

Christiaensen, L. and Demery, L. (2007). “Down to Earth: Agriculture and poverty reduction in Africa,” *Directions in Development*. World Bank, Washington D.C..

de Janvry, A., Sadoulet, E. and Zhu, N. (2005). “The role of non-farmincomes in reducing poverty and inequality in China,” *CUDARE Working Paper 1001*. Berkeley: University of California.

Demurger, S. and Xu, H. (2011). “Return Migrants: The Rise of New Entrepreneurs in Rural China,” *World Development, 39*, 1847-1861.

Drukker, D.M. (2003). “Testing for serial correlation in linear panel-data models,” *The Stata Journal, 3*, 168–177.

Fan, C. (1997). “Uneven development and beyond: Regional development theory in post-Mao China,” *International Journal Of Urban And Regional Research, 21*, 620- 639.

Fleisher, B., Hu, D., Mcguire, W. and Zhang, X. (2010). “The Evolution of an Industrial Cluster in China,” *China Economic Review, 21*, 456-469.

Gao, H., Liang, Q. and Luo, G. (2014). “Effects of Clusters on China’s E-Commerce: Evidence from the Junpu Taobao Village,” *International Journal of Business and Management, 9*, 180-186.

Gao, Y., Zheng, J. and Bu, M. (2014). “Rural-urban Income Gap and Agricultural Growth in China ,” *China Agricultural Economic Review, 6*, 92-107.

Glaeser, E., Resseger, M. and Tobio K. (2008). “Urban Inquality,” *NBER Working paper No. 14419* , National Bureau of Economic Research, Cambridge.

Gunasinghe, C. (2016). “Fiscal Policy, Economic Growth and Income Inequality: A Case Study on Australia,” *Working paper*, Griffith University , Nathan.

Guo, G., Liang, Q. and Luo, G. (2014). “Effects of Clusters on China’s E-Commerce: Evidence from the Junpu Taobao Village,” IJBM International Journal of Business and Management, 9, 51-72.

Han, Q., Li, J. and Zhang, J. (2014). “Education and Income Inequality in China,” *Journal of Global Business and Trade, 10*, 13-36

Hendriks, M., Ludwigs, K. and Veenhoven R. (2016). “Why are Locals Happier than Internal Migrants? The Role of Daily Life,” *Social Indicators Research, 125*, 481-508.

Jiang, Y., Shi, X., Zhang, S. and Ji, J. (2011). “The Threshold Effect of High-Level Human Capital Investment on China's Urban-Rural Income Gap,” *China Agricultural Economic Review, 3*, 297-320.

Li, A. (2017). “E-commerce and taobao villages: A promise for China’s rural development?” *China Perspectives, 3*, 57-62.

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 Agricultural Research, 7*, 35-37.

Long, H., Zou, J. and Liu, Y. (2009). “Differentiation of Rural Development Driven by Industrialization and Urbanization in Eastern Coastal China,” *Habitat International, 33*, 454-462.

Lu, M. and Chen, Z. (2004). “Urbanization, Urban-biased Economic Policies and Urban-rural Inequality,” *Economic Research Journal, 6*, 50-58 (in Chinese).

Martin, R., Florida, R., Pogue, M. and Mellander, C. (2015). "Creativity, Clusters and the Competitive Advantage of Cities," *Competitiveness Review, 25*, 482-496.

Meng, J. (2017). "Economic Watch: How has E-commerce Changed Life in Rural China?" available at: http://news.xinhuanet.com/english/2017-07/10/ c_136432492.htm [accessed 18 Jan. 2018].

Meng, X. and Wu, H. (1998). "Household Income Determination and Regional Income Differential in Rural China," *Asian Economic Journal, 12*, 65-88.

Mu, X. (2017). "E-commerce not to Kill Jobs but to Enlarge Inequality: Research," available at: http://news.xinhuanet.com/english/2017-10/07/c_136662506.htm [accessed 18 Jan. 2018].

Nair-Reichert, U. and Weinhold, D. (2001). "Causality Tests for Cross-country Panels: A New Look at FDI and Economic Growth in Developing Countries," *Oxford Bulletin of Economics and Statistics, 63*, 153-171.

Qiu, P., Caine, E., Yang, Y., Chen, Q., Li, J. and Ma, X. (2011). "Depression and Associated Factors in Internal Migrant Workers in China," *Journal Of Affective Disorders, 134*, 198-207.

Ravallion, M. and Chen, S. (2007). "China's (Uneven) Progress against Poverty," *Journal Of Development Economics, 82*, 1-42.

Ravallion, M. and Datt, G. (1996). "How important to India's Poor is the Sectoral Composition of Economic Growth?" *World Bank Economic Review, 10*, 1-25.

Rivera, L., Gligor, D. and Sheffi Y. (2016). "The Benefits of Logistics Clustering," *International Journal of Physical Distribution & Logistics Management, 46*, 242-268.

Sicular, T., Yue, X., Gustafsson, B. and Li, S. (2007). "the Urban-Rural Income Gap and Inequality in China," *Review Of Income And Wealth, 53,* 93-126.

Sorkin, A. (2017). "E-Commerce as a Jobs Engine? One Economist's Unorthodox View," available at: https://www.nytimes.com/2017/07/10/business/dealbook /e-commerce-jobs-retailing.html [accessed 18 Jan. 2018].

Su, B. and Heshmati, A. (2013). "Analysis of the Determinants of Income and Income Gap between Urban and Rural China," *China Economic Policy Review, 2*, 1-29.

Su, C., Liu, T., Chang, H. and Jiang, X. (2015). "Is urbanization narrowing the urban-rural income gap? A cross-regional study of China," *Habitat International, 48*, 79-86.

Su, W., Eriksson, T., Zhang, L. and Bai, Y. (2016). "Off-farm Employment and Time Allocation in On-farm Work in Rural China from Gender Perspective," *China Economic Review, 41*, 34-45.

Tan, J. (2015). "Alibaba: The Giant With 81.5% Market Share In China's Online Shopping Market,", available at: http://www.sharesinv.com/articles/2015/11/09/ alibaba-the-giant-with-81-5-market-share-in-chinas-online-shopping-market [accessed 05 September 2017].

Wang, B., Li, X., Stanton, B. and Fang, X. (2010). "The Influence of Social Stigma and Discriminatory Experience on Psychological Distress and Quality of Life among Rural-to-Urban Migrants in China," *Social Science & Medicine, 71*, 84-92.

Wei, H. and Zhao, C. (2012). "Effects of International Trade on Urban-Rural Income Gap in China," *Finance and Trade Economics, 1*, 78-86.

Wu, X. and Huang, C. (2014). "Research on the Development of Chinese Agriculture-Related E-Commerce," Working Paper, Zhejiang University College of Management, Zhejiang University, Hangzhou. (in Chinese).

Xie, Y. and Zhou, X. (2014). " Income Inequality in Today's China,"*Proceedings Of The National Academy Of Sciences Of The United States Of America, 111*, 6928-6933.

Yang, D. (1990). "Patterns of China's Regional Development Strategy," *The China Quarterly, 122*, 230-257.

Yang, D. and Zhou, H. (1999). "Rural-urban disparity and sectoral labor allocation in China," *Journal of Development Studies, 35*, 105-133.

Zheng, Y. (2007). "Principal Component Cluster Analysis of the Development of Rural E-Commerce Regional Infrastructure,"*Forum on Science and Technology in China, 1*, 119-122. (in Chinese)

Zhou, Y. B. (2009), "Urbanization, Urban-rural Income Gap and Overall Income Inequality in China: an Empirical Test of the Inverse-U Hypothesis," *China Economic Quarterly, 4*, 1239-1256.

Table 3.1 Districts Dropped from the Sample

Zhejiang Province	
Hangzhou City district	Gongshu, Shangcheng, Xiacheng, Jianggan, Xihu, Binjiang
Ningbo City district	Haishu, Jiangdong, Jiangbei, Beilun, Zhenghai
Wenzhou City district	Lucheng, Longwan, Ouhai, Dongtou
Jiaxing City district	Nanhu, Xiuzhou
Huzhou City district	Wuxing, Nanxun
Shaoxing City district	Yuecheng
Jinhua City district	Wucheng, Jindong
Quzhou City district	Kecheng, Qujiang
Zhoushan City district	Dinghai, Putuo
Taizhou City district	Jiaojiang, Huangyan, Luqiao
Lishui City district	Liandu
Jiangsu Province	
Nanjing City district	Xuanwu, Gulou, Qinhuai, Jianye, Yuhuatai, Xixia, Jiangning, Pukou, Liuhe, Lishui, Gaochun
Wuxi City district	Liangxi, Xishan, Huishan, Binhu, Xinwu
Xuzhou City district	Gulou, Yunlong, Jiawang, Quanshan, Tongshan
Changzhou City district	Tianning, Zhonglou, Xinbei, Wujin
Suzhou City district	Gusu, Huqiu, Wuzhong, Xiangcheng, Wujiang
Nantong City district	Chongchuan, Gangzha
Lianyungang City district	Lianyun, Haizhou, Ganyu
Huaian City district	Huaian, Huaiyin, Qingjiangpu, Hongze
Yancheng City district	Tinghu, Yandu, Dafeng
Yangzhou City district	Guangling, Hanjiang, Jiangdu
Zhengjiang City district	Jingkou, Runzhou, Dantu
Taizhou City district	Hailing, Gaogang, Jiangyan
Suqian City district	Sucheng, Suyu
Shandong Province	
Jinan City district	Lixia, Shizhong, Huaiyin, Tianqiao, Licheng, Changqing
Qingdao City district	Shinan, Shibei, Huangdao, Laoshan, Licang, Chengyang
Zibo City district	Zichuan, Zhangdian, Boshan, Linzi, Zhoucun
Zaozhuang City district	Shizhong, Xuecheng, Yicheng, Taierzhuang, Shanting
Dongying City district	Dongying, Hekou
Yantai City district	Zhifu, Fushan, Mouping, Lanshan
Weifang CIty district	Weicheng, Hanting, Fangzi, Kuiwen
Taian City district	Taishan, Daiyue
Rizhao City district	Donggang, Lanshan

Table 3.1 (continued)

Laiwu City district	Laicheng, Gangcheng
Linyi City district	Lanshan, Luozhuang, Hedong
Fujian Province	
Fuzhou City district	Gulou, Taijiang, Cangshan, Jinan, Mawei
Xiameng City district	Siming, Huli, Jimei, Haicang, Tongan, Xiang'an
Putian City district	Chengxiang, Hanjiang, Licheng, Xiuyu
Sanming City district	Meilie, Sanyuan
Quanzhou City district	Licheng, Fengze, Luojiang, Quangang
Zhangzhou City district	Xiangcheng, Longwen
Hebei Province	
Shijiazhuang City district	Changan, Qiaoxi, Xinhua, Yuhua, Gaocheng, Luquan, Luancheng
Chengde City district	Shuangqiao, Shuangluan
Zhangjiakou City district	Qiaodong, Qiaoxi, Xuanhua, Xiahuayuan, Wanquan, Chongli
Qinhuangdao City district	Haigang, Shanhaiguan, Beidaihe
Tangshan City district	Lunan, Lubei, Guye, Kaiping, Fengnan, Fengrun, Caofeidian
Langfang City district	Anci, Guangyang
Baoding City district	Jingxiu, Lianchi
Cangzhou City district	Xinhua, Yunhe
Hengshui City district	Taocheng, Jizhou
Xingtai City district	Qiaodong, Qiaoxi
Handan City district	Hanshan, Congtai, Fuxing, Feixiang

Table 3.2 Summary Statistics and Definition of Variables

Variables	Obs	Mean	Std. Dev.	Min	Max
$Deflated\_uincome_{it}$	1442	26713	7999	10205	57769
$Deflated\_rincome_{it}$	1603	12853	4653	3686	30160
$Deflated\_income\_D_{it}$	1446	13879	4042	2751	28709
N_taobao_{it}	1624	1.003	4.173	0	65
Nei_taobao_{it}	1624	4.722	11.584	0	109
$Taobao_{it}$	1624	5.725	13.706	0	123
$Zhejiang_{i}$	1624	0.163	0.369	0	1
$Jiangsu_{i}$	1624	0.111	0.314	0	1
$Shandong_{i}$	1624	0.236	0.425	0	1
$Fujian_{i}$	1624	0.165	0.371	0	1
$Hebei_{i}$	1624	0.325	0.469	0	1
$Trend_{t}$	1624	2.5	1.118	1	4
$Pop\_density_{it}$	1617	0.569	0.435	0.05	4.3
Pr_per_{it-1}	1620	14.94	9.009	0.42	59.3
Ti_per_{it-1}	1621	36.34	7.881	14.37	65.09
$Pcmileage_{it}$	1410	3.366	2.147	0.25	17.92
$High_{it}$	1615	43.64	9.978	10.11	111.64
$Loan_{it-1}$	1442	71.18	40.03	12.57	288.54
$Asset_{it-1}$	1621	77.76	30.23	6.23	245.6
$Fiscal_{it-1}$	1617	14.11	7.230	3.67	69.41
$Foreign_{it}$	1425	1.337	1.627	0	30.06

$Deflated\_uincome_{it}$: deflated urban per capita disposable income of the county in yuan for county or district i at year t;

$Deflated\_rincome_{it}$: deflated rural per capita net income of the county in yuan for county or district i at year t;

$Deflated\_income\_D_{it}$: the difference between the two variables above in yuan for county or district i at year t;

N_taobao_{it}: number of Taobao villages in county i at year t;

Nei_taobao_{it}: number of Taobao villages in all counties and districts share a common boarder with county i at year t;

$Taobao_{it}$: the total number of Taobao villages in county i and all other counties and districts that share a common boarder at year t;

$Zhejiang_{i}$: whether the county or district i is in Zhejiang province;

Table 3.2 (continued)

$Jiangsu_i$: whether the county or district i is in Jiangsu province;
$Shandong_i$: whether the county or district i is in Shandong province;
$Fujian_i$: whether the county or district i is in Fujian province;
$Hebei_i$: whether the county or district i is in Hebei: province;
$Trend_t$: time trend variable which equals to the year minus 2012;
$Pop\_density_{it}$: number of people in thousands per square kilometer of county or district i at year t;
Pr_per_{it-1}: lagged primary sector share of GDP of county or district i at year t-1;
Ti_per_{it-1}: lagged tertiary sector share of GDP of county or district i at year t-1;
$Pcmileage_{it}$: per capita mileage of roads in kilometers of county or district i at year t;
$High_{it}$: number of high school students per thousand residents of county or district i at year t;
$Loan_{it-1}$: lagged total amount of loans lent by local banks and financial institution to GDP ratio of county or district i at year t-1;
$Asset_{it-1}$: lagged fixed asset investment to GDP ratio of county or district i at year t-1;
$Fiscal_{it-1}$: lagged local fiscal expenditure to GDP ratio of county or district i at year t-1;
$Foreign_{it}$: foreign direct investment (FDI) as a proportion of GDP of county or district i at year t;

Table 3.3 Estimation of Rural Income

Variables	OLS	Random Effect	Fixed Effect
	(1)	(2)	(3)
$Jiangsu_i$	-0.221***	-0.173***	
	(-10.85)	(-4.56)	
$Shandong_i$	-0.408***	-0.338***	
	(-18.23)	(-9.66)	
$Fujian_i$	-0.372***	-0.320***	
	(-17.66)	(-8.96)	
$Hebei_i$	-0.564***	-0.592***	
	(-25.79)	(-17.25)	
$Trend_t$	0.0933***	0.0712***	0.0668***
	(16.94)	(28.48)	(26.87)
$Taobao_{it}$	-0.000353	0.000758***	0.000828***
	(-0.85)	(5.00)	(5.86)
$Density_{it}$	0.111***	0.144***	0.0836**
	(7.82)	(6.54)	(2.54)
Pr_per_{it-1}	-0.0006	-0.00499***	-0.00153*
	(-0.68)	(-4.59)	(-1.71)
Ti_per_{it-1}	0.00291***	-0.000671	0.00111*
	(3.44)	(-0.80)	(1.91)
$High_{it}$	-0.00202***	-0.00000678	-0.000311
	(-3.58)	(-0.02)	(-0.83)
$Asset_{it-1}$	-0.000341	0.000112	0.0000767
	(-1.46)	(0.59)	(0.40)
$Loan_{it-1}$	0.000578***	0.00028	0.000102
	(3.19)	(1.57)	(0.54)
$Fiscal_{it-1}$	-0.0201***	-0.00147	0.00582***
	(-17.61)	(-1.44)	(5.34)
$Foreign_{it}$	0.0142***	0.000282	-0.000536
	(4.44)	(0.15)	(-0.29)
$Pcmileage_{it}$	0.00464	-0.00706	0.00465
	(1.34)	(-1.47)	(0.73)
Constant	9.719***	9.618***	9.228***
	(207.25)	(196.10)	(175.60)

Notes: *N*=1117; T-statistics are in parentheses;
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.4. Estimation of Urban-rural Income Gap

Variables	OLS	Random Effect	Fixed Effect
	(1)	(2)	(3)
$Jiangsu_i$	0.0157	-0.0568*	
	(0.91)	(-1.85)	
$Shandong_i$	0.126***	0.0226	
	(5.81)	(0.72)	
$Fujian_i$	0.111***	0.0356	
	(6.20)	(1.22)	
$Hebei_i$	0.224***	0.161***	
	(12.02)	(5.63)	
$Trend_t$	-0.0308***	-0.0203***	-0.0159***
	(-6.39)	(-8.20)	(-5.94)
$Taobao_{it}$	0.000243	-0.000523***	-0.000598***
	(0.69)	(-3.45)	(-3.98)
$Density_{it}$	-0.0179	0.00712	0.108***
	(-1.48)	(0.37)	(3.11)
Pr_per_{it-1}	-0.00374***	-0.00108	-0.00165
	(-4.80)	(-1.07)	(-1.03)
Ti_per_{it-1}	-0.00149**	0.000871	0.0014
	(-2.05)	(1.10)	(1.42)
$High_{it}$	-0.00100**	0.000186	0.000988**
	(-2.06)	(0.50)	(2.44)
$Asset_{it-1}$	0.0000781	0.000104	0.0000922
	(0.39)	(0.56)	(0.44)
$Loan_{it-1}$	0.000811***	0.0000353	-0.000278
	(5.23)	(0.21)	(-1.39)
$Fiscal_{it-1}$	0.00552***	0.000966	-0.00207*
	(5.64)	(0.99)	(-1.77)
$Foreign_{it}$	0.000617	-0.000178	-0.00122
	(0.23)	(-0.10)	(-0.63)
$Pcmileage_{it}$	0.00961***	0.00874**	-0.00522
	(3.23)	(2.04)	(-0.77)
Constant	0.689***	0.662***	0.701***
	(17.08)	(14.91)	(12.55)

Notes: N=1053; T-statistics are in parentheses;
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Conclusions

This dissertation aims to provide valuable information on the development of Taobao villages, which are China's unique rural e-commerce clusters. We are interested in what is Taobao village, how Taobao villages are formed, what kind of area have the potential to have more Taobao villages, and how Taobao villages affect China's economy. Specifically, we are focusing on: 1) the determinants of the formation of Taobao villages, 2) whether labors migrating from Agro-Taobao village make different decisions on migration compare with those from nearby villages and 3) the effects of Taobao villages on rural net income and urban-rural income inequality.

We begin with using Poisson and Negative binomial count regression models to examine the determinants of the formation of Taobao villages. We find that education plays an important and positive role in the formation of Taobao villages, while credit support for other businesses and existing Taobao villages nearby make the formation of new Taobao villages less likely.

Moreover, this chapter also identify that to help new Taobao villages to form and develop, local governments should focus on promoting high school graduation rates and give less credit support to other businesses. Also, support for areas that are far away from existing Taobao villages are more likely to be successful. Since the subprime mortgage crisis in 2008, China tried to transit from investment oriented growth model to consumption based model. The contribution of consumption to GDP growth was 51.8% in 2012, which surpassed the contribution of investment for the first time since 2000, and the number increase further to 64.5% in 2017. On the other hand, online retail accounts for 18.54% of the total retail sales in China and the number is expected to rise in the future, which means to foster the development of e-commerce and Taobao villages is one of the most important goal for the local governments.

In the second chapter, we investigate whether home village being an Agro-Taobao village, a cluster of e-commerce agribusinesses and the Taobao villages which most of online sellers are selling the agricultural products, influences farmers' migration decisions in China. Using a self-collected survey data from three Agro-Taobao villages and three nearby non-Taobao villages in Shuyang county of Jiangsu province, we find that originating from an Agro-Taobao village decreases a villager's likelihood of migrating by 18%, after controlling for various factors that are likely to influence the migration decision such as the individual's demographic characteristics, migration experience, educational background, and health statuses of the individual and his/her spouse.

Rural-to-urban migration has played an important role in the growth and development of China in recent decades. However, migration has also caused numerous social and economic issues such as overcrowding of cities, rising inequality between urban and rural areas, lack of sufficient care for children and seniors, as well as the depletion of brain and labor resources in rural areas. The findings of our study indicate that the develop of Agro-Taobao villages or e-commerce activities in general, offers a promising way to convince villagers to stay in their home villages, and can potential be a solution to the problems rural China is currently facing.

In this last chapter, we use panel data models to test for the effects of Taobao villages and various economic indexes on rural net income and urban-rural income inequality in five coast provinces. We find that the development of Taobao villages is an effective method to raise income for rural residents, and to narrow the income gap between urban and rural China, while noting that fiscal expenditures play a similar role. Also, compared with the agricultural sector, the development of non-farm sectors is a major contributor to rising incomes in rural China. In addition, county's average education level will actually widen the income gap.

Reducing rural poverty and income inequality is usually the ultimate goal for local governments. Our results show that the rural e-commerce activities did help to rise China’s rural income and narrow the urban-rural income gap in our sample provinces. Meanwhile, increase fiscal expenditures, and invest in non-farm sectors over agriculture for rural areas have similar effects.

Chapter 1 and 3 are both suffered from the data availability, the econometric analyses of this two chapters are at the county level and we are only able to include a limited set of variables. Also, the online transaction volumes is the best index to represent the development of Taobao business and e-commerce, but it is not available and we use the number of Taobao villages in a county instead. If data at the town or village level or data on Taobao transaction volume become available in the future, a more detailed study can be conducted to further examine the determinants of the formation of Taobao villages or e-commerce clusters in general, and to examine how Taobao businesses change villagers' lives, and China's economic structure. .

APPENDICES

Appendix A The Questionnaire for Chapter 2

Date: _______/_______/2017 Enumerator: __________________

Location: ________ City________ County________ Township________ Village

Family information

1、Family size ______ Number of family members in labor force (aged between 16 and 60) ________

For children under 16:

1). Age______ Gender ______ Education:______

A. Kindergarten B. Primary school C. Middle school D. High school E. Do not go to school

Whether live in the villavge ___________ Whether live with parents ____________ Distance between school and home _________ km

2). Age______ Gender ______ Education:______

A. Kindergarten B. Primary school C. Middle school D. High school E. Do not go to school

Whether live in the villavge ___________ Whether live with parents ____________ Distance between school and home _________ km

3). Age______ Gender ______ Education:______

A. Kindergarten B. Primary school C. Middle school D. High school E. Do not go to school

Whether live in the villavge ___________ Whether live with parents ____________ Distance between school and home _________ km

4). Age______ Gender ______ Education:______

A. Kindergarten B. Primary school C. Middle school D. High school E. Do not go to school

Whether live in the villavge ___________ Whether live with parents ____________ Distance between school and home _________ km

2、Your total household income in 2017 is____________Yuan. The major source of income is______________________________, The percentage of Taobao related income is ________%.

3、Your total household spending in 2017 is_________Yuan.

4、The number of family members engaged in Taobao and related businesses_______. They work as **(Multiple Choice)**___________

(1) Taobao business owner (2) Taobao shop staff (3) working in Taobao related businesses (packaging, delivery, website maintenance) (4) Other (please specify)__________________

5、The number of relatives and friends engaged in Taobao and related businesses_______. They work as **(Multiple Choice)**___________

(1) Taobao business owner (2) Taobao shop staff (3) working in Taobao related businesses (packaging, delivery, website maintenance) (4) Other (please specify)__________________

6、distance from home to the administrative village office _______ km

Parents' information

1. Father's information

Age ___________ Personal annual income last year ____________Yuan

Whether he is currently a migrant worker ___________

Education level	Health status
A. No education	A. Excellent
B. Primary school	B. Good
C. Middle school	C.Fair
D. High school or technical secondary school	D. Poor
E. College, technical college or higher	

1) If currently a migrant worker:

Number of years being a migrant worker ________

What kind of jobs he has worked as a migrant worker

____

Whether migrating with kids ________ if yes, how many kids _________

Whether kids are attending school in the current migration destination ________

2) If currently not a migrant worker：

Had he ever been a migrant worker ________ If answer yes, what kind of jobs he had worked before ___________________________ How many years had he migrated _______

When did he return _______________ For what reason ____________________________

The last job worked before return _______________ The annual income from that job __________Yuan

Whether he is currently engaged in e-commerce _____, if yes, he works as **(Multiple Choice)**

(1) Taobao business owner (2) Taobao shop staff (3) working in Taobao related businesses (packaging, delivery, website maintenance) (4) Other (please specify)__________________

The reason for starting Taobao business ________________________________

If he is not engaged in e-commerce, what is his current occupation

____

2. Mother's information

Age ___________ Personal annual income last year ____________Yuan

Whether she is currently a migrant worker ___________

Education level	Health status
A. No education	A. Excellent
B. Primary school	B. Good
C. Middle school	C.Fair
D. High school or technical secondary school	D. Poor
E. College, technical college or higher	

1) If currently a migrant worker：

Number of years being a migrant worker ________

What kind of jobs she has worked as a migrant worker

____

Whether migrating with kids ________ if yes, how many kids _________

Whether kids are attending school in the current migration destination ________

2) If currently not a migrant worker：

Had she ever been a migrant worker ________ If answer yes, what kind of jobs she had worked before ____________________________ How many years had she migrated _______

When did she return _______________ For what reason _____________________________

The last job worked before return _______________ The annual income from that job __________Yuan

Whether she is currently engaged in e-commerce _____, if yes, he works as **(Multiple Choice)**

(1) Taobao business owner (2) Taobao shop staff (3) working in Taobao related businesses (packaging, delivery, website maintenance) (4) Other (please specify)__________________

The reason for starting Taobao business _________________________________

If she is not engaged in e-commerce, what is her current occupation _____________________________

Primary non-parent caregiver information (no need to fill out if the primary caregiver is one of the parents)

The relationship with kids _________ age _________ Whether retired __________

Education level	Health status
A. No education	A. Excellent
B. Primary school	B. Good
C. Middle school	C.Fair
D. High school or technical secondary school	D. Poor
E. College, technical college or higher	

Current annual income (including pension) ____________

The number of years he/she has been taking care of the children ___________