Industrial Clustering, Income, and Inequality in Rural China

Abstract:

This study provides evidence that links industrial clusters and rural income in China. Based on a pooled cross-sectional dataset composed of rural households from 109, 121 and 313 counties in 1995, 2002 and 2007, respectively, as well as a unique density-based index measuring the existence of industrial clusters calculated from firm-level data, we identify the mechanisms by which industrial clusters simultaneously increase rural income and reduce income inequality among rural households in China. Furthermore, we provide systematic evidence that specialization, urbanization and industrialization, measured in standard ways, do not have such effects on rural household income or inequality. Our evidence suggests that in China's context, industrial clusters developed under joint efforts of entrepreneurs and local governments have lessened institutional constraints and provided relatively equal opportunities for rural residents to participate in nonfarm activities. As a result, in those areas, rural household income is increased, and income inequality is reduced. The findings of this study have important policy implications for reducing poverty and inequality, and, smoothing income gaps between socioeconomic groups in economic transitions.

Keywords: industrial clustering, income and inequality, transition economies

1. Introduction

China in Mao's era had the largest population in absolute poverty in the world. In the post-Mao reforms, China became the world champion in poverty reduction with rapid industrialization. However, the progress is far from even, and poverty is still prevalent.[i] The vast majority of poverty is concentrated in rural areas, where income inequality has worsened over time and is much higher than that in cities (Benjamin, Brandt, & Giles, 2005; Ravallion and Chen, 2007; Li et al., 2015; Wu, Zheng, & Wei, 2017; Gao et al., 2019). Moreover, studies find that rural poverty and income inequality have stable Spatio-temporal patterns, i.e., rural poverty rate and income inequality have been higher in less developed regions than those in developed eastern coastal regions (Ravallion and Chen, 2007; Li et al., 2015; Han et al., 2021). Most studies highlight the decisive impact of nonfarm income, particularly from migrant workers and urbanization, on rural income and inequality (Yao, 1999; Benjamin, Brandt, & Giles, 2005; Wan & Zhou, 2005; Howell, 2017). Our study, however, provides evidence for mechanisms through which China has reduced rural poverty and inequality through a particular type of industrialization, namely, entrepreneurial industrial clusters.

Different industrialization processes have different implications for the rural population (e.g., Kuznets, 1955, 1963; Lee & Sissons, 2016). This study provides evidence that links industrial clusters, one of the most important features of industrialization in China, and rural income. China's economic reform began in the countryside with the development of township and village enterprises (TVEs) in full swing, with hundreds of millions of farmers no longer engaging in agricultural production. These TVEs, while no longer a major sector of the Chinese economy, have served as a steppingstone for private sector development and the basis for rural industrial clusters (Xu, 2011). With hundreds or even thousands of small firms clustering in rural towns, some of which have been transformed into national or international production centers for specific products.[ii] However, the high degree of income inequality and poverty in rural areas stands in stark contrast with the rapid industrialization of rural China. How does industrial clustering impact rural income, poverty and inequality? Which part of the income is most affected by industrial clustering? Which types of households are affected more? What are the mechanisms of the clustering effects?

To address these questions empirically, we use a pooled cross-sectional dataset consisting of approximately 7,998, 9,200, and 13,000 households in 109, 121, and 313 counties covered by the China Household Income Project (CHIP) survey in 1995, 2002, and 2007, respectively. To measure industrial clusters, we deploy a density-based index (DBI) (a la Guo et al., 2020) to calculate the presence of clusters based on firm-level data from the Above-Scale Industrial Firm Panel (ASIFP). The DBI is constructed based on the relative density of firms in the same industry in a county. We use such an index to capture the distinctive features of industrialization in China during the economic reform, i.e., the agglomeration of a large number of small and specialized entrepreneurial firms within a geographical territory.

We find that industrial clusters contribute to rural income growth, poverty reduction and lower income inequality. First, rural households' income, especially nonfarm income, is higher in counties with industrial clusters than in counties without industrial clusters. Similarly, the poverty rates in counties with industrial clusters are also significantly lower than those in counties without industry clusters. Second, counties with industrial clusters have significantly lower rural household income inequality than counties without clusters. Interestingly, a similar result holds for nonfarm income inequality, whereas it does not hold for farm income. Third, households with more members of disadvantaged groups, including the elderly, under-educated and/or those with health issues, benefit from clustering more than others. Finally, we provide systematic evidence that specialization, urbanization and industrialization, measured in standard ways, do not have such effects on rural household income or inequality. Our evidence suggests that in China's context, industrial clusters developed at the early stages of economic reforms under joint efforts of entrepreneurs and local governments have lessened institutional constraints and provided relatively equal opportunities for rural residents to participate in nonfarm activities. As a result, in those areas, rural household income is increased, and income inequality is reduced.

To address identification concerns, we employ a series of estimations, including two-stage estimations with instrumental variables (IVs) and additional examinations in which we replace the DBI with traditional measurements for specialization, urbanization and industrialization to verify the effects of industrial clusters. For the two-stage estimation, the first IV is the development of TVEs in the early years of the economic reform, measured by the total employment of TVEs in the region. Regions with a strong TVE sector in the 1980s are more likely to have industrial clusters (Mukherjee & Zhang, 2007). Therefore, the size of the TVEs in the early years of economic reforms should be related to the development of clusters. At the same time, it should not be correlated to the within-county income and inequality after more than ten years except through its effect on industrial development. The second IV is the historical importance of Chinese lineage value, measured by the total number of genealogies that appeared in a county before 1949. It was argued that social trust and close coordination embedded in the lineage system could be important elements in fostering the TVEs (Weitzman & Xu, 1994) and later development of clusters. The lineage system was an important institution in rural Chinese societies, but it had been eliminated by the Chinese Communist Party since the early 1950s. Thus, the historical importance of this institution should not affect rural income in contemporary China directly unless through economic activities based on trust and coordination. Statistical analyses confirm the validity of the IVs. We find significantly positive relationships between the employment of TVEs and clustering and the number of genealogies in history and clustering. Meanwhile, Sargan tests suggest the joint exogeneity of the two IVs from the error terms of the regressions. More importantly, the effects of clustering on household income and income inequality stay as robust after the clustering variable is instrumented. Finally, by replacing the DBI with specialization, urbanization and industrialization measured by traditional ways in the estimates, we confirm that the effects of industrial clusters on rural income and inequality we have evident are not equivalent to those of specialization or general nonfarm shocks of urbanization or manufacturing booms.

We believe this is the first time such systematic evidence has been provided in the development economics literature. The closest work to the present study is Guo et al. (2020), which examines how industrial clusters affect Chinese regional disparities and urban-rural income inequality. However, the present study differs from Guo et al. (2020) because it examines rural inequality from household data and identifies the underlying mechanisms by which industrial clusters affect rural income and income inequality among rural households, using rigorous research approaches.

In addition to improving our understanding of poverty and inequality in China, this study also contributes to the general literature in the following areas. First, it complements research on rural poverty and inequality in developing countries. Rural poverty and inequality have been major concerns of policymakers and scholars because such issues will ultimately affect the sustainability of global development. Various factors, including public policy, local leadership, human capital, access to markets, credit and social infrastructure, technology transfer, and migration, are considered in explaining rural poverty and inequality, most of which suggest that nonfarm activities are the ultimate solution for rural poverty, especially in developing counties (Barham & Boucher, 1998; Khan, 2001; Barret, Lee, & McPeak, 2005; De Janvry, Sadoulet & Zhu, 2005; Shen, Docquier, & Rapoport, 2010; Zhu & Luo, 2010; Naschold, 2012; Emran & Hou, 2013; Howell, 2017; Adamopoulos et al., 2017; Gibson et al., 2017; Giles & Mu, 2018). However, most empirical studies find that driven by nonfarm income, poverty reduction and income growth in rural areas often go hand in hand with widening income inequality among rural residents (Rozelle, 1994; Benjamin, Brandt, & Giles, 2005; Bramall, 2008; Haggblade, Hazell, & Reardon, 2010; Iqbal et al., 2018). Focusing on a particular type of institution, we discover that nonfarm income generated by rural industrial clusters does not necessarily increase inequality while bringing growth. The key is whether the institution provides relatively equal opportunities for local rural residents to be engaged in nonfarm activities.

Additionally, this study extends the existing literature on economic geography by linking agglomeration with rural income distribution. The literature on agglomeration mainly focuses on economic efficiencies (Marshall, 1890; Jacobs, 1969; Glaeser et al., 1992; Rotemberg & Saloner, 2000; Cingano & Schivardi, 2004; Combes & Duranton, 2006; Ellison, Glaeser, & Kerr, 2010). However, there is growing recognition regarding the tradeoffs of agglomeration, focusing on the spatial aspects of poverty and inequality (Krugman, 1991; Waltz, 1996; Fujita, Krugman, & Venables, 1999; Martin & Ottaviano, 1999, 2001; Kanbur & Venables, 2005; Castells-Quintana & Royuela, 2014; Combes, Duranton, & Gobillon, 2019). These existing studies mainly focus on urban issues or migration because urbanization and clustering are inextricably linked in a market economy with free movement of production factors. Our study departs from this literature by focusing on the effects of industrial clusters in China, which are featured with restricted factor mobility, on the income distribution within rural areas.

Finally, this study also contributes to the long-standing debate on the relationship between industrialization, growth, and income distribution. The seminal work of Kuznets (1955, 1963) shows an inverse U-shaped relationship between industrialization and income inequality, arguing that a positive relationship between growth and income inequality at the early stages of industrialization is inevitable. The fact that income inequality is higher in rural China than in urban areas is a significant challenge for the Kuznets hypothesis, which suggests that rural income inequality should be lower than that in urban areas before a turning point comes when the overall inequality lowers. Therefore, this study complements the existing discussions that suggest a different trajectory of income inequality in the early stage of industrialization in developing countries (Ravallion & Chen, 1999; Deininger & Squire, 1998; Lee & Sissons, 2016). Our empirical investigation suggests that different institutions may shape the path of industrialization, affecting income growth and inequality, adding evidence for the endogenous growth theory (Acemoglu, Johnson, & Robinson, 2002, 2005; Easterly, 2007).

The rest of the paper is organized as follows. Section 2 introduces the industrial clusters and income inequality in rural China. In Section 3, we discuss the data sources, variables and sample. Section 4 presents the empirical findings for the effects of industrial clusters on rural income and inequality and the mechanisms for such effects. Section 5 reports some additional robustness checks identifying the effects of industrial clusters. Section 6 concludes this study.

2. Industrial clusters and rural household income in China

2.1. Industrial clusters as the engine of industrialization in China

In a market economy, the essential conditions for "clusters" to occur are the predominance of private ownership of assets and the mobility of input factors: labor and capital are mobile, while land can be traded freely in the marketplace. Under these conditions, the market price of input factors will influence the co-location decisions of firms and labor, forming clusters of firms and urban cities. Therefore, industrial clusters are inextricably linked to industrialization and urbanization (Marshall, 1890; Weber, 1929; Fujita, Krugman, & Venables, 1999; Ellison, Glaeser, & Kerr, 2010).

However, all of these primary conditions are not satisfied in China, and as a result, industrial clusters in China have developed under different institutions with different mechanisms. Despite the rapid growth of the de facto private sector since 1997, when state-owned enterprises were in deep trouble, it was not until 2004 that private enterprises were constitutionally recognized (Xu, 2011). Moreover, at the onset of reform at the end of the 1970s, the government completely controlled all input factors, restricting entrepreneurial firms' freedom in choosing their locations. Of all the factors, the prohibition of free trade of rural land is the most significant constraint. Farmers, individually or collectively, are not allowed by law to trade "their" land for non-agricultural purposes till now.[iii] Associated with the government control of land is the *Hukou* system (residence registration system), which restricts the labor mobility of peasants, particularly their movement from rural to urban areas (Meng, 2012). Although the *Hukou* system has been relaxed over time that peasant migrants are allowed to work in cities as de facto lower-class citizens, people who work outside the geographical area of their *Hukou* are not eligible for or discriminated against in accessing local social welfares, including housing, health care, education, childcare benefits, and pensions (Smart & Smart, 2001; Au & Henderson, 2006). Moreover, peasants and small-medium enterprises in rural areas have limited access to formal financial resources (Liu & Yu, 2008).

With the institutional constraints described above, the industrialization in rural China began with the development of TVEs, which are vaguely defined as collectively owned as all township or village residents "set up" the TVE and own the firm collectively (Weitzman & Xu, 1994). The township or village community government "represents" the communal collective owners and is the de facto executive owner of the TVE (Byrd & Lin, 1990). In this way, rural residents could use collectively owned land and surplus agricultural labor to engage in industrial activities without violating the public ownership system. The performance of TVEs was spectacular. Between 1981 and 1990, the total industrial output of TVEs grew at an average annual rate of 28.1%, twice the national average and more than three times that of the state-owned sector. The share of TVEs in the gross domestic product (GDP) increased from 14.3% in 1980 to 37.5% in 1995 (Xu & Zhang, 2009). Despite differences in details, TVEs share the following key characteristics: all were led by rural entrepreneurs; all had vague definitions of ownership at the incipient stage, reflecting certain institutional constraints (Weitzman & Xu

1994; Li, 1996); and all had close ties with local governments (Qian & Xu, 1993; Chang & Wang, 1994; Che & Qian, 1998).

Since the late 1990s, many rural enterprises have been privatized as the resistance against private ownership has waned. Subsequently, the development of private firms has been even faster. Associated with the growth of the private sector is a trend toward increased specialization and clustering of small firms. These specialized small firms are linked together through subcontracting networks where a collection of many specialized firms produces a final product. With the concentration of a vast number of small and specialized firms, many townships have become national or international centers of specific products, forming industrial clusters. These clusters often consist of many privatized TVEs or their derivative companies.

In the past decade, with the rapid development of information technology, the widespread use of E-commerce has provided further entrepreneurial opportunities, and such E-commerce activities have shown a robust spatial aggregation pattern (Zhu et al., 2016; Lin, 2019; Qi, Zheng, & Guo, 2019; Liu et al., 2020; Couture et al., 2021). It has been reported that rural residents benefit from E-commerce more than urban residents because it helps rural residents break the geographical boundaries of traditional markets to buy and sell goods and services, thus relaxing institutional constraints on factor mobility (Liu, Li, & Liu, 2015; Aker et al., 2016; Deichmann, Goyal, & Mishra, 2016).

To summarize, industrial clusters that originate from the concentration of small private firms are a consequence of China's unique institutions, especially in rural areas. The firms in the clusters are usually owned and set up by local residents who have no right to sell their land and cannot easily relocate their business elsewhere. That is why many of them are located in officially defined rural areas, although many could be considered urban areas by general economic geography criteria. Moreover, in association with the *Hukou* system, many employees of the firms within clusters are officially defined as peasants, although they are manufacturing or service workers. These constraints imply that administrative boundaries are critically crucial for China's industrial clusters.

2.2. Industrial clusters and rural income growth and distribution

Since the post-Mao reform, rural China has experienced significant development that absolute poverty has fallen substantially. Rural residents' per capita disposable income increased from RMB 134 in 1978 to RMB 16,021 in 2019. However, many rural residents are lagged behind, and poverty is still a significant issue in many Chinese rural areas, despite the government's efforts. Income inequality among rural residents has also widened, with the Gini coefficient for income distribution increasing from 0.24 in 1979 to 0.39 in 2011 (Li & Sicular, 2014).

Many studies have tried to identify factors that affect poverty and income inequality in rural China. For example, Jalan and Ravallion (2002) suggest that publicly provided goods and services and private investment may have contributed to the geographical poverty trap in rural China. Consistently, Emran and Hou (2013) discover that the lack of access to domestic and international markets plays an essential role in explaining the rural poverty in China. Some other studies have highlighted the *Hukou* system as a significant factor influencing rural incomes (Smart & Smart, 2001; Whalley & Zhang, 2004; Au & Henderson, 2006). From a political economy perspective, He, Guo, and Wang (2017) find that the human capital of local leadership has a significant impact on the income of rural residents, while Giles and Mu (2018) find that village election and expected land tenure affect the migration decisions of rural residents and thus their income. Similarly, Adamopoulos et al. (2017) discover that resource misallocation obstructs the productivity of the rural residents and thereby contributes to their low income. Finally, some studies suggest that the collapse of farm incomes causes income inequality in rural China due to falling agricultural prices and rising nonfarm activities such as migration (Yao, 1999; Benjamin, Brandt, & Giles, 2005; Wan & Zhou, 2005; Howell, 2017).

This paper complements the previous research by revealing concrete mechanisms that drive different income growth patterns and distributions in rural China. We argue that the emergence and development of industrial clusters in China, as a response to institutional constraints, has distinctive impacts on rural income distribution. With a high degree of specialization, the entry barrier to entrepreneurship is lowered within clusters, and the close collaboration of all relevant stakeholders within clusters ensures that coordination costs are reduced without the need for integration. Such inclusive clusters provide rural residents with increased and relatively equal access to entrepreneurial and employment opportunities in their hometowns. As a result, not only local elites but also the vulnerable groups of rural residents can benefit from local entrepreneurial and employment opportunities. The disadvantaged groups, such as the elderly, under-educated, and those with health problems, are often not able to leave their hometowns to earn nonfarm income in faraway big cities and fall into absolute poverty. Therefore, we expect to observe that in rural areas where industrial clusters exist, rural households enjoy higher income through increased nonfarm income and that income inequality among rural households decreases due to more equal opportunities to participate in nonfarm activities.

3. Data and sample

This study investigates the effects of industrial clustering on rural household income and inequality. To measure income and inequality, we construct a pooled cross-sectional dataset composed of approximately 7,998, 9,200, and 13,000 rural households from 109, 121, and 313 counties covered by the China Household Income Project (CHIP)[iv] in 1995, 2002 and 2007, respectively. To measure industrial clustering across Chinese regions, we follow the DBI method in Guo et al. (2020) and utilize data from ASIFP, a census-type firm survey data that provide detailed firm-level information including the industry, location, age, size, ownership, and financial information of all SOEs and non-SOEs with annual sales of 5 million RMB or above.

It is worth noting that both datasets have some limitations for our research purpose. We, therefore, use different ways, including data cleaning and robustness checks to avoid potential biases caused by such limitations. First, although CHIP, the most comprehensive income survey across China over the most extended period, covers many regions, in some regions, the number of households included in the surveys is reasonably small, which may hinder the accuracy in the calculation of income inequality. We therefore drop all the counties with fewer than 20 households covered by CHIP, leaving a total of 413 counties for estimations. Such a data cleaning method leads to the result that, on average, there are 77 households in a county in 1995, 70 in 2002, and 67 in 2007. Second, ASIFP does not cover non-SOEs with less than 5 million RMB annual sales. Since firms in industrial clusters are mainly non-SOEs and small, the DBI calculated based on ASIFP may lead to biases. We use various ways to check the robustness of our estimations, mainly based on the comparisons between the results drawn from the DBI calculated from ASIFP and those calculated based on the Economic Census of 1995 (details are presented in Section 3.1 and Section 4).

Data for control variables such as per capita GDP and total GDP at the county level and other county-specific variables such as expenditure in education and agriculture come from the China Socio-Economic Development Statistical Database and National Prefecture and County Public Finance Statistical Yearbooks.

3.1 Measuring industrial clusters

The key explanatory variable in this study is the DBI measure of cluster existence. Existing clustering indices focus on either regional specialization or inter-connectedness of local industries (Porter, 1990; Krugman, 1991; Glaeser et al., 1992). Most studies on regional specialization have applied the Herfindahl–Hirschman Index (HHI), Gini coefficient (Gini), location quotient (LQ), or Krugman index to measure clustering.[v] However, these measures hardly capture the characteristics of China's industrial clusters because the industrial clusters in China are by themselves products of the significant legacy of the centrally planned system rather than developed under a free market economy.

This paper applies the DBI measure of Chinese clusters developed in Guo et al. (2020), which is constructed based on the relative density of firms in the same industry within a geographical territory. We use such an index to capture the distinctive features of industrialization in China during the economic reform, i.e., the agglomeration of a large number of small and specialized entrepreneurial firms within a geographical territory. Firm density is particularly important to understanding the Chinese economy due to the institutional constraints and entry barriers faced by entrepreneurs, as we have discussed earlier. As a result of these constraints, many firms are specialized and limited in size and mobility. They achieve competitiveness by clustering together, coordinating closely and sharing resources.

Figure 1 presents how the DBI reflects the unique features of industrial clusters in China and differs from the agglomeration measured in traditional ways in the existing literature. If we apply the existing agglomeration measures to the Chinese data, as shown in Figure 1 panel (a) to (d), inland regions, particularly those with a concentration of SOEs, are identified with a higher level of industrial clustering. However, these regions have long been recognized for lacking entrepreneurial firms and clusters. To construct the DBI measure, we define a county to have an industrial cluster of a particular industry if the county is among the top α percentile of all counties regarding firm density for that industry, and we assign $\alpha = 5$.[vi] We then construct a dummy variable $Cluster_{i,t}$, which equals 1 if there is at least one industrial cluster in county i in year t, and equals 0 otherwise.

[Figure 1 here]

Ideally, census data should better capture industrial clusters in China, given that it contains all firms, including small family workshops. However, the census data is only available for the years 1995 and 2004 during the examination period of this study. So, we use the ASIFP data (available from 1995 to 2007) in our empirical analysis. Note that as of 2004, the enterprises covered by ASIFP can account for 90% of the total sales of all the firms covered by the Economic Census, suggesting a slim chance of data biases. However, to check whether using the ASIFP leads to serious biases for our estimations, we compare the spatial patterns of the cluster existence generated from ASIFP with those generated from the Economic Census data. We also use the DBI calculated from the Economic Census of 1995 to conduct regression estimations on rural income and inequality as a robustness check (details are discussed in Section 4.1 and 4.2).

Figure 1 panel (e) and (f) plot the geographic distribution of DBI clusters calculated from both the 2007 ASIFP data and the 2004 Economic Census data across Chinese regions (the figures of the comparison for 1995 are available by request). It shows that the spatial patterns of clusters are highly identical for DBIs calculated from ASIFP and the Economic Census data. For instance, the identified clusters are concentrated in the coastal areas, particularly in Zhejiang, Jiangsu, and Guangdong Province. Inland regions have much fewer clusters, and there are virtually no clusters operating in provinces like Tibet and Xingjiang. Note that this pattern is consistent with the general perceptions of spatial distributions of entrepreneurial clusters in China. For instance, existing studies have documented the existence of clusters in Wenzhou City (Huang, Zhang, & Zhu, 2008), Tongxiang County (Ruan & Zhang, 2009), and Wuxing County (Sonobe, Hu, & Otsuka, 2002), and all these clusters have been identified using the DBI method. Given that the DBI better captures clustering in China than standard regional specialization measure or inter-connectedness measure, in this study, we apply the DBI to measure industrial clusters in Chinese counties.

3.2 Measuring rural household income and within-county income inequality

Our dependent variables include the measurements of rural household income and income inequality at the county level. Above all, total household income is measured by the net disposable household per capita income, which equals the net disposable household total income divided by the total number of family members (denoted as *Total income* and in logarithm). Such measurement automatically corrects for the household size. However, it does not consider that the needs of a household with each additional member may vary due to the economies of scale in consumption. We, therefore, construct an alternative measure employing the square root scale (denoted as *Total income_sqrt*) that divides household income by the square root of household size to calculate per capita income. This measure implies that a household of four persons has consumption needs twice as large as one composed of a single person. We use it as an alternative measure of household per capita income.

Besides the per capita total income, we are also interested in whether clustering affects different components of income, particularly that from industrial activities and agricultural activities, respectively. *Nonfarm income* is defined as (the logarithm of) the average household per capita income from wage or/and business activities within a county.[vii] *Farm income* is (the logarithm of) the average household per capita income from agricultural activities in a county. We also examine how the clustering affects the share of the two components to the total income, which can tell to what degree clustering facilitates the rural transformation from an agricultural to an industrial economy.

The third category of our dependent variables is within-county household income inequality. We mainly focus on intra-county household income inequality, which contributes to more than two-thirds of the disparity in rural areas (Benjamin, Brandt, & Giles, 2005). We measure inequality from several different dimensions. First, we calculate inequality not only for total income but also for the two major income components. Then we can test how clustering affects total income inequality, nonfarm income inequality, and farm income inequality simultaneously. Second, we measure inequality by different indexes, which are Gini coefficient, Theil index, mean log deviation index (MLD), the share of income going to the 10% richest and 10% poorest households, and the extreme poverty rate, to capture different aspects of the income distribution. Gini coefficient is defined as follows:

$$Gini = \frac{1}{n}\left[n + 1 - 2\left(\frac{\sum_{i=1}^{n}(n+1-i)y_i}{\sum_{i=1}^{n} y_i}\right)\right] \quad (1)$$

where *n* is the total number of households in the county, and y_i is the household income or income components indexed by *i* in non-decreasing order ($y_i \leq y_{i+1}$).

Two counties may have the same Gini coefficient but different inequalities because the two Lorenz curves can have the same area yet different shapes. Therefore, we use the Theil index to measure inequality as well. Theil index is calculated as:

$$Theil = \sum_{i=1}^{n} \frac{y_i}{Y} * log\left[\left(\frac{y_i}{Y}\right)/\left(\frac{1}{n}\right)\right] \quad (2)$$

where, similarly, *n* is the total number of households, and y_i is the household income or income components indexed by *i*, and *Y* represents the total income or income components of the population, with $Y = \sum_{i=1}^{n} y_i$.

Finally, MLD is defined as:

$$MLD = \frac{1}{n}\sum_{i=1}^{n} ln\frac{\bar{y}}{y_i} \quad (3)$$

where, *n* is the total number of households, and y_i is the household income or income components, and $\bar{y}$ is the mean of y_i. MLD equals 0 if every household has the same income and takes larger values as the income becomes more unequal.

Although it is widely accepted that *Gini*, *Theil*, and *MLD* are qualified inequality measures, they give little information on the extreme values of the income distribution. To investigate whether the effect of clustering is mainly significant for the top or bottom decile respectively, we generate another two variables, *R10* and *P10*, defined as the share of income going to the top 10% wealthiest households and the share of income going to the top 10% poorest households within a county, respectively. The first three inequality measures (i.e., *Gini*, *Theil*, and *MLD*) can tell how cluster affects the distributions of total income and different income components in general, while the last two measures (i.e., *R10* and *P10*) can determine whether clusters affect the more disadvantaged group more or less than, the more affluent group. To avoid potential measurement errors, we also construct variables that capture the share of income going to the top 15% and 20% of most affluent and poorest households to conduct the robustness checks (details are discussed in Section 4.2).

Finally, we also look at the extreme poverty rate to capture income inequality. Precisely, we measure extreme poverty using the ratio of the population living under the international poverty line (IPL) over the total population in each sampled county in a given year. The IPL is a monetary threshold under which an individual is considered to be living in extreme poverty. The IPL was set at USD 1 before 2001 and USD 1.25 between 2002 and 2008 in purchasing power parity (PPP) terms. In our study, given that we have three cross-sectional data across 1995 and 2007, we use both thresholds for calculating the extreme poverty rate in each county each year. *PovertyR_1* is the percentage of the population living under USD 1 per day (PPP), while *PovertyR_1.25* is the percentage of the population living under USD 1.25 per day (PPP) in each sampled county in a given year.

We also include several control variables in our estimations. First, we control the per capita GDP of a given county in the previous year to capture the economic development level of a given region because it may directly affect the household income level. Second, we include the expenditure in education over the total fiscal expenditure of a given county in the previous year as a control variable because education variation may significantly contribute to regional inequality (Wan & Zhou, 2005). Third, we control the ratio of expenditure in agriculture (including forestry and fisheries) over the total fiscal expenditure of a given county in the previous year as this may be directly related to rural development. Fourth, we control the number of special economic zones (SEZs) in a given county in the previous year to disentangle the effect of clusters from that of SEZs. Fifth, we control the regional dummy variables that define whether the county is located in western, central, or eastern China to capture the cross-regional variances in institutional environments, industrialization, and economic development. Finally, our study combines three national surveys; we therefore control the year effect in all our empirical models.

Figure 2 shows the geographic distribution of counties covered by the CHIP survey in 1995, 2002, and 2007. The majority of the surveyed counties are concentrated in the eastern and central regions, though the 2007 data covers more inland counties than other years. About one-third of the CHIP surveyed counties have at least one cluster operating there, and this ratio is relatively stable across time. Table 1 provides summary statistics for the variables of our interests, such as the household income, income components including nonfarm income, farm income, and their shares, and household income inequality indexes. It shows that, on average, the total household income is much higher in counties with clusters than that in counties without clusters. In particular, in counties with clusters, the nonfarm income is much higher than that in counties without clusters. Meanwhile, the income inequality also seems to be higher in counties with clusters, though the poverty rate is much lower than in counties without clusters.

[Figure 2 and Table 1 here]

4. Empirical findings

4.1 Clustering and rural household income

Our first research question is whether the development of industrial clusters affects household income in rural China, particularly the composition of nonfarm and farm income. We estimate the following equation to answer this question:

$Income_{i,t} = \alpha + \beta_1 Cluster_{i,t} + \beta_2 X_{i,t-1} + Region_i + Year_t + e_{i,t}$ (4).

The dependent variable is the county average of household per capita income (*Total income*), household per capita nonfarm income (*Nonfarm income*), household per capita farm income (*Farm income*), or their respective shares in the total income (*Share nonfarm* or *Share farm*). $Cluster_{i,t}$ is the DBI measure of cluster existence in county *i* in year *t*. $X_{i,t-1}$ is a vector of control variables including county per capita GDP, fiscal expenditure ratio in education, fiscal expenditure ratio in agriculture, and the number of SEZs in the county in the previous year. $Region_i$ are the region dummies for eastern, middle and western regions, and $Year_t$ are the year dummies.

Table 2 presents the OLS regression results. Model (1) to (3) present the estimations on how industrial clustering affects total income, nonfarm income, and farm income. These three variables are transformed into a natural logarithmic format because the data are skewed.[viii] In this case, the estimated coefficients for these three variables indicate the growth rate of the income. Moreover, information on per capital GDP, fiscal expenditure on education and agriculture for some counties is missing for specific years. Eventually, when we pool all the years together, the total number of observations for the final estimations is dropped to 284.

[Table 2 here]

As shown in Table 2, industrial clustering is significantly associated with higher growth of the average household per capita income and per capita nonfarm income. Households in counties with clusters have, on average, 8.5% higher growth in per capita total income and 18.7% higher growth in per capita nonfarm income. There is, however, no significant effect of the cluster on the growth of per capita farm income. On the other hand, models (4) and (5) show that clusters have a significantly positive impact on the share of household income from nonfarm activities, which can be either wage income from working for someone else or/and business income from establishing small businesses. Households in counties with clusters have, on average, 3.5% more income from nonfarm activities compared with those in counties without clusters. However, the effect of clusters on the share of household farm income is statistically insignificant. The insignificant impact of cluster development on rural households' farm income may seem intriguing. On the one hand, if the development of clusters shifts too much labor force from their original agricultural work, it shall negatively impact the farm income. On the other hand, clustering may impact farm income positively because the household can use the income from nonfarm jobs supplied by industrial clusters to increase their capital input in agriculture activities to improve productivity. Hence, how clustering affects household farm income can be ambiguous.

Furthermore, consistent with the existing literature, we find that county per capita GDP in the previous year is significantly and positively correlated with total household income and nonfarm income. It is also significantly associated with higher nonfarm and lower farm income shares. Meanwhile, we observe a significantly positive relationship between the ratio of fiscal expenditure on agriculture activities (including agriculture, forestry, and fisheries) and farm income share, implying that government fiscal expenditure that supports agriculture indeed raises the importance of agricultural income. However, the ratio of fiscal spending on education in the previous year is significantly and negatively correlated to the growth of nonfarm income and its share in total income. Such results are consistent with existing studies that find no consensus on the relationship between formal education and entrepreneurship (Simoes, Crespo, & Moreira, 2016). In the case of China, many studies have found that years of education have significant negative impacts on entrepreneurial choices (Lu & Tao, 2010), in particular, the self-employed type of entrepreneurial choice (Chu & Wen, 2019), which is the focus of our study. Finally, the coefficients of middle and western region dummies are significantly associated with lower total income and nonfarm income.

We use the square root scale to adjust the household income as a further robustness check. The results are presented in Table A-1. Our main conclusions that clustering is associated with higher household total income and nonfarm income remain robust. Furthermore, as we have discussed earlier, to check whether using the ASIFP leads to serious biases for our estimations, we use the DBI calculated from the Economic Census of 1995 to conduct regression estimations on rural income and inequality as robustness checks. The reason to use DBI calculated based on the Economic Census of 1995 is that the spatial pattern of the industrial clusters does not change a lot over time, and 1995 is the starting period for our estimations. The results are presented in Table A-2. It shows that all the results stay robust regarding the clustering effects on rural and nonfarm income.

4.2 Clustering and rural household income inequality

The previous analysis reveals a positive relationship between clustering and rural household income. In the following, we examine how industrial clustering affects the household income distribution within a county. The model we estimated is the following:

$$Inequality_{i,t} = \alpha + \beta_1 Cluster_{i,t} + \beta_2 X_{i,t-1} + Region_i + Year_t + e_{i,t} \quad (5).$$

The dependent variable is the within-county household per capita income inequality measure, including *Gini*, *Theil*, *MLD*, *R10*, *P10*, *PovertyR_1*, and *PovertyR_1.25*. The definitions of $Cluster_{i,t}$, $X_{i,t-1}$, $Region_i$, and $Year_t$ are the same as they are in equation (4). In addition to measuring the inequality of total household income, we also investigate how clustering affects inequality of different income components, i.e., nonfarm income and farm income.

Table 3 reports the estimations on the effects of clustering on household per capita total income inequality. Models (2) to (3) show that the existence of cluster is significantly and negatively correlated to the Theil index and MLD of household per capita income within a county, implying that a county with an industrial cluster may significantly reduce the household total income inequality. For instance, on average, the Theil index is significantly lower in counties with clusters than those without clusters by 0.023, which is about 16.08% of the average Theil index for all the counties in our sample during the examination period. Similarly, the MLD is significantly lower in counties with clusters than those without clusters by 0.018, which corresponds to 12.95% of the average MLD for all the counties during the examination period.

[Table 3 here]

Moreover, in Model (4), we observe that the existence of clusters is significantly and negatively correlated to *R10*, the ratio of the income earned by the wealthiest 10% households over the total income of all the households in the county. The point estimate on $Cluster_{i,t}$ is -0.017 (and statistically significant at the p<0.05 level), corresponding to 7.17% of the average value of *R10*. Such results suggest that the existence of industrial clusters may reduce income inequality by soothing the income gap between affluent households and others. However, we do not observe a statistically significant relationship between the existence of clusters and *P10* in Model (5), the ratio of income earned by the poorest 10% of households in the county.

Finally, the existence of clusters is significantly and negatively associated with the extreme poverty rate when we use USD 1.25 as the threshold. When we use USD 1 as the threshold, the effect of the cluster is significant at the margin (p = 0.116). On average, keeping other things being equal, the estimated extreme poverty rate measured with an IPL of UDS 1.25 is significantly lower in counties with cluster than in those without cluster by 0.032, which is about 22.70% of the average poverty rate for all the counties of our sample during the examination period.

In sum, Table 3 shows a significantly negative association between local industrial clustering and within-county household total income inequality in rural China. The existence of clusters is associated with the reduction of most inequality measures, namely, *Theil*, *MLD*, *R10*, and *PovertyR_1.25*. Furthermore, clusters mainly affect the income inequality of rural households within a county by smoothing the income gap between the wealthiest rural household and others.

We then estimate the relationships between clustering and inequality of different income components. Table 4 reports how clustering affects household per capita nonfarm income inequality. It shows that the existence of clusters is negatively correlated with the inequality measured by *Gini*, *Theil*, and *MLD*, at the $p<0.01$ statistical level. Specifically, Model (1) shows that the Gini coefficient of nonfarm income is lower by 0.044 in counties with clusters than that in those without a cluster, which corresponds to 8.40% of the average Gini coefficient (or 28.39% of one standard deviation of the Gini coefficient) of all the counties in our sample. Model (2) and (3) show that the point estimate on $Cluster_{i,t}$ is -0.131 or -0.609 when inequality is measured by Theil index or MLD, respectively. Such results suggest that counties with industrial clusters, on average, have a 22.51% and 25.51% lower nonfarm income inequality measured by Theil index or MLD. Note that the magnitude of the effect is more significant for nonfarm inequality than for total income inequality. As shown in Model (4) and (5), similar to what we find for total income inequality, the development of clusters does not seem to have a statistically significant effect on the nonfarm income of the top 10% poorest households. In comparison, it has a significantly negative impact on the nonfarm income of the top 10% of the wealthiest households.

[Table 4 here]

We further look at the relationship between clustering and household farm income inequality. As shown in Table 5, the relationship between industrial clustering and farm income inequality is not statistically significant in most estimations except for Model (1), where the dependent variable is the Gini coefficient of farm income. Such results suggest that industrial clusters generally have no statistically significant impact on the farming income distribution of rural residents within a county, implying that the reduction of total income inequality in counties with clusters is mainly driven by the reduction of nonfarm income inequality.

[Table 5 here]

Once again, we use the DBI calculated based on the Economic Census of 1995 to conduct the estimations for income inequality to check the robustness of our estimations. The results are presented in Tables A-3 and A-4. It shows that the patterns in terms of the clustering effects on income and inequality stay robust. Furthermore, as we have discussed earlier, in order to check the robustness of our study, we also construct variables measuring the ratio of the income earned by the wealthiest 15% and 20% as well as the poorest 15% and 20% and examine how industrial clusters affect these inequality variables. Again, the estimates show that the results stay robust (results are available by request) when we use different cut-offs to define rich or poor.

4.3 Two-stage estimations with IVs

In the previous analyses, we find a statistically significant association between industrial clustering, the growth of household income, and the reduction of household income inequality within a region. However, although reverse causality is not a significant concern in our context, given that clusters are hardly formed because of a county's income level or income distribution, the above-mentioned significant relationships may be inflated by omitted variables coexisting with clustering. To address the identification issue, we employ two-stage estimations with two IVs to identify the effect of clustering. Since we have two IVs, our regressions are overestimated that we can test the relevance and the exogeneity of the IVs at the same time.

Our first IV is the early development of TVEs measured by the total employment of TVEs in 1987 in logarithm in a given province where a county was located (denoted as *TVE_emp*). As discussed, many industrial clusters were rooted in TVEs, which emerged as a form of adaptation to the weak legal protection of property rights to capture market opportunities when private property rights were illegal (Weitzman & Xu, 1994). With vaguely defined property ownership, TVEs functioned as another local institution in nurturing entrepreneurship, spreading managerial skills, and accumulating physical and human capital in China's early years of economic reform. Many private firms were spun off directly from TVEs before forming industrial clusters (Xu & Zhang, 2009). Therefore, we suggest that this variable may predict whether a county within the province has developed clusters while given the data is at province level and is for the period of late 1980s, it should not be related to the error terms of the income level and income inequality of a given county during our sample period.

The second IV is the historical importance of Chinese lineage value measured by the total number of genealogies (denoted as *Genealogies*) that appeared in a county before 1949. Information on genealogy is obtained from http://ouroots.nlc.cn/, an online database covering 30,581 genealogies nationwide constructed by the National Library of China. It is well known that lineage groups play an important role in the social life of Chinese society, especially in Southern China's countryside. As a special kind of community organization, lineage groups coordinate and regulate the behavior of their members and provide protection for them in traditional China (Fei, 1953; Freedman, 1966). Such lineage groups with intensive interpersonal relationships and close ties have a high degree of trust and better internal coordination (Coleman, 1990; Putnam, 2000). Meanwhile, studies have found that in regions with strong impacts of lineage value, local officials, including the cadres in communist China, work closely with the lineage groups because they rely on these community organizations to manage some public matters (Fei, 1953; Tsai, 2002). As we have discussed, social trust and coordination and close cooperation between entrepreneurs and local governments are fundamentally important factors in the development of TVEs and, later, industrial clusters in China (Weitzman & Xu, 1994; Xu & Zhang, 2009). Meanwhile, the lineage system was an important institution in rural Chinese societies but was eliminated by the Chinese Communist

Party in the early 1950s. Only those regions with deep roots of the lineage culture would have survived radical political changes (Peng, 2010). Thus, the historical importance of this institution should not affect rural income in contemporary China directly unless through economic activities based on trust and coordination.

The results of the two-stage estimations are presented in Table 6. Panel A reports the first-stage estimations where we regress the cluster existence, $Cluster_{i,t}$, on the two IVs and other control variables. The estimation shows that both the size of TVE employment and the number of genealogies are statistically and positively correlated to the DBI measure of clustering. It is suggested that clusters are more likely formed in a county located in a province with a stronger TVE endowment in 1987. Furthermore, a county with stronger impacts of the lineage value in history is also more likely to have formed industrial clusters. Such results confirm the relevance of the IVs. Additionally, the Sargan test shows that the null hypothesis, which states that the two IVs are uncorrelated to the residuals of the estimation, cannot be rejected. Thus, the results statistically prove that both the IVs satisfy the conditions of relevancy and joint exogeneity.

[Table 6 here]

Panel B of Table 6 reports the second-stage estimation results on how clustering affects household income. Model (1) shows that the instrumented cluster existence is significantly associated with higher household per capita income growth. Model (2) shows that the instrumented cluster existence is significantly correlated with higher growth of household per capita nonfarm income, while Model (3) shows that its impact on household per capita farm income growth remains insignificant. Moreover, Model (4) shows that instrumented cluster variable has a significantly positive effect on the share of household income from nonfarm activities. Such results are consistent with the OLS estimation results shown in Table 2. These results confirm our expectation that industrial clusters may bring more employment opportunities to the local rural residents, which leads to higher nonfarm income and total income to local households.

The results of the second-stage estimations on clustering and household total income inequality are reported in Table 6 Panel C. Model (2) and (3) show that the instrumented $Cluster_{i,t}$ remains significantly associated with reduced *Theil* and *MLD* of household per capita total income within a county. These results confirm that the development of industrial clusters effectively reduces local household income inequality. Furthermore, the instrumented cluster existence is significantly and negatively associated with *R10*, the share of income of the wealthiest 10% of households in a county. Finally, as shown in Models (6) and (7), we find the negative effects of clustering on the extreme poverty rate remain robust after the existence of industrial clusters is instrumented.

In sum, the results shown in Table 6 confirm the causal relationships between industrial clustering and household income growth. Furthermore, the income distribution of rural households in counties with industrial clusters is significantly more equal than that of rural households in counties without industrial clusters.

4.4 The mechanisms of the clustering effects

In the previous subsections, we have provided solid evidence for the effects of the industrial clusters on income growth and reduced inequality in rural China. In the next, we explore the potential mechanisms for such effects. As we have discussed, the major difference between the industrial clusters in China and the usual manufacturing booms or industrialization lies in the fact that this type of industrialization is concentrated in rural areas and with a large number of small, specialized firms coordinating closely and sharing resources to achieve the economic efficiency. In such a way, the entry barriers are reduced and firm competitiveness is achieved without integration. However, the usual manufacturing booms usually are concentrated in cities with integrated firms. Indeed, based on the 2004 and 2008 economic census data, Guo et al. (2020) find that the number of new entrants to non-agricultural businesses in regions with clusters is significantly higher than that in other regions. For instance, the number of new businesses established between 2004 and 2008 is 821.55 in counties with clusters, while it is 186.74 in counties without clusters.

Without local industrial work opportunities, rural people often have to migrate in order to seek nonfarm employment and higher income, usually with enormous social costs such as being discriminated against in welfare entitlements and income, as well as the psychological issues caused by being separate from family members and working in cities as de factor lower-class citizens (Liu et al., 2014; Wang, Guo, & Cheng, 2015). The development of industrial clusters in rural areas, on the other hand, provides a large number of local entrepreneurial or employment opportunities that substitute the need for migration. Indeed, we find supporting evidence for the substitution between working in clusters and migration from the CHIP data we used in the paper. Table 7 compares the share of households with remittance income from migrant family members and the share of the remittance over the total income between counties with and without clusters. It shows that counties with clusters have significantly fewer households with remittance income from migrant family members than counties without clusters. Furthermore, the household share of income from migrant remittance is significantly lower in counties with clusters.

[Table 7 here]

More importantly, while both working in a cluster and migration can increase income, they affect inequality differently. Many studies have found that migration typically widens income inequality for rural households because migrants earn substantially more than those who stay at home. Thus, the income gap between households with and without migrants (and their remittance income) may widen in areas with high outflows of migrant workers (Barham & Boucher, 1998; Yao, 1999; Benjamin, Brandt, & Giles, 2005; Wan & Zhou, 2005; Howell, 2017; Young, 2013). However, our estimates suggest that rural households enjoy higher income growth and lower income inequality in regions with industrial clusters than those in other areas. We suggest that such increased income and reduced income inequality in counties with clusters are driven by the local entrepreneurial and employment opportunities. A simple comparison of the income gaps for households with and without migrant remittance in different regions supports our conjecture. As shown in Table 8, the household per capita income in counties with clusters is RMB 4,798, while counties without clusters are RMB 3,019. More importantly, in counties with clusters, the per capital income gap between the rural households with and without remittance income is RMB 180 (4% of the mean), while it is RMB 635 (21% of the mean) in counties without clusters. Thus, one way in which clusters reduce rural inequality is through substituting for migration.

[Table 8 here]

Finally, suppose the relatively equal entrepreneurial and employment opportunities provided by industrial clusters are the primary channels through which industry clusters help increase rural income and reduce income inequality. In that case, we should observe that where industry clusters exist, disadvantaged groups will have more opportunities to participate in the nonfarm economy than their counterparts elsewhere. To check our hypothesis, we further look at the household-level data and examine which types of households benefit more from the industrial clusters. Based on the CHIP survey in 2002 and 2007, which provide detailed information of household members, we construct a series of variables to measure household characteristics related to disadvantaged groups. Specifically, we calculate the ratio of household members above the age of 55 (*Ratio elderly*), and the ratio of members with primary school education or below (*Ratio less-educated*) and construct a dummy variable indicating whether the household has any member identified as "very unhealthy" in the survey (*Dummy unhealthy*). We then conduct estimates for the household-level nonfarm income by adding these household variables and the interaction terms of these variables with the variable indicating the existence of clusters. We also include all the control variables as in the baseline county-level analysis and control for the ratio of male household members and the total number of household members in the regression. The results are presented in Table 9.

[Table 9 here]

As shown in the table, the existence of industrial clusters is significantly and positively correlated with the nonfarm income of a household, consistent with the results shown in Table 2. Moreover, all the indicators concerning disadvantaged groups are significantly and negatively correlated with the nonfarm income of a household, suggesting that households with more members of disadvantaged groups are less likely to be engaged in nonfarm activities. However, the interaction terms of clustering and the household characteristics are significantly and positively associated with the nonfarm income of a household. Such results suggest that households with more members of disadvantaged groups can engage in more nonfarm activities and earn more nonfarm income in counties with clusters than their counterparts in counties without clusters. These findings support our hypothesis that providing relatively equal nonfarm opportunities is the primary channel for industrial clusters to help increase household income and reduce income inequality in rural China simultaneously.

5. Additional robustness checks

In the previous section, we have employed various approaches to examine the effects of industrial clusters on rural income and inequality and identify the mechanisms through which such effects work on site. In this section, we will conduct additional robustness checks to confirm our findings.

5.1 The effects of industrial cluster or the effects of other nonfarm shocks?

Although we have discussed in detail how industry clusters, as defined by the DBI, have different impacts on rural income and inequality from other types of nonfarm shocks, we are aware that the DBI is not exclusively about rural industrial firms. Therefore, it might be challenged that the effects of industrial clusters defined by the DBI we have observed are similar to some other nonfarm shocks defined by standard ways. To further verify whether the increase in nonfarm income and the reduction in rural income inequality that we have found are the results of what we call the effects of unique industrial clusters measured by the DBI or general nonfarm shocks or other types of agglomeration, we have conducted a series of additional empirical tests.

First, we replicate all the estimates by excluding city districts from our sample. It is noted that our samples are rural households surveyed by CHIP, and the majority of the samples are from regions concentrated with rural residents. The surveys are distributed in 543 counties, among which only 13% are city districts (*qu*). Tables A-5 and A-6 report the results. In order to save space, we present only some of the main estimates for the growth and inequality of total and nonfarm incomes in Table A-5. It shows that when we exclude clusters in the city districts, industrial clusters defined by the DBI are still significantly associated with higher nonfarm income and lower income inequality. Moreover, as shown in Table A-6, the positive moderating effects of industrial clusters on the nonfarm income of disadvantaged rural residents remain robust.

Second, we replace the DBI with traditional indicators of industrialization and urbanization used in the existing literature to see whether the impacts of industrial clusters we have discovered on rural income are the effects of a general nonfarm shock. "Urbanpop_ratio" is the proportion of a county's urban population to its total population in the year prior to the CHIP survey year, indicating the level of urbanization in a county. "Industrialization" is the proportion of manufacturing output to the total output of a county in the year prior to the CHIP survey year, measuring a county's level of industrialization. Table A-7 presents the results of the estimates on rural income and distribution, where the DBI index is replaced by measures of industrialization and urbanization in Panels A and B, respectively. As the table shows, we do not observe significantly positive impacts of urbanization and industrialization on total or nonfarm incomes of rural residents within a county, which is consistent with the findings of Fan, Li & Zhang (2012). More importantly, the measure of industrialization is significantly associated with higher levels of rural income inequality as measured by the Gini coefficient and the MLD index, indicating that an increase of rural income inequality is associated with a general manufacturing boom, consistent with the prediction of "Kuznets Curve" (Kuznets, 1963). Furthermore, we do not observe that more urbanized counties are associated with a reduction in rural income inequality measured by any means. The estimates of rural income inequality and poverty measured by other means are similar to the results shown in this table (results are provided by request). Finally, based on household-level data, we replace the DBI with urbanization and industrialization indicators for estimations on the nonfarm income of the disadvantaged group of rural residents. The results are reported in Table A-8. It shows that industrialization is significantly and positively correlated with the nonfarm income of rural residents in general. However, the interaction terms of industrialization and the indicators defining disadvantaged groups of rural residents, i.e., less educated groups and/or older people, are significantly and negatively correlated with the nonfarm income, indicating that rural residents of disadvantaged groups are benefited less than others in more industrialized regions, contrasting to the results we have found with the moderating effects of industrial clusters defined by the DBI. Panel B shows that urbanization is significantly and negatively correlated with the nonfarm income of rural residents in general, differing from what we have found with the industrial clusters defined by the DBI. Meanwhile, we do not observe statistically significant moderating effects of urbanization regarding the nonfarm income for disadvantaged groups in rural China, except for the less educated group.

Third, besides the estimates mentioned above, we further verify whether the effects of the industrial clusters defined by the DBI differ from those of clusters measured by traditional agglomeration indicators used in the existing literature. We first construct a series of specialization indicators used in the existing literature (e.g., Porter, 1990; Krugman, 1991; Glaeser et al., 1992). We then replace the DBI with these agglomeration indicators and investigate whether the impacts of these indicators on rural income are the same as those of the clusters defined by the DBI. Tables A-9 and A-10 present the results. It shows that most standard specialization indicators, including the Herfindahl–Hirschman Index (HHI) and the location quotient (LQ), are significantly and negatively correlated with rural residents' total income and nonfarm income at the county level. Meanwhile, we do not observe a statistically significant relationship between these specialization indicators and income inequality of rural residents' total income or nonfarm income within a county. At last, the coefficients of these specialization indicators and the interaction terms of these indicators and measures of disadvantaged groups are either significantly negative or statistically insignificant for the estimations of nonfarm income of rural residents based on household-level data. Such results suggest that rural people in general (and disadvantaged groups in particular) in more specialized regions defined by the traditional specialization indicators do not do any better in nonfarm income than their counterparts in less specialized regions.

To summarize, institutional restrictions make China's industrial clusters differ from the concepts of "urbanization", "manufacturing boom", or "geographical agglomeration" studied in the existing literature. These industrial clusters simultaneously increase rural income and reduce income inequality among rural households by providing relatively equal work opportunities for rural residents. Empirical estimations presented in Tables A-5 to A-10 provide systematic evidence that specialization, urbanization and industrialization, measured in standard ways, do not have such effects on rural household income or inequality.

5.2 The rural migration effects

Many existing studies have found the significant impacts of migration on rural income and inequality. Therefore, it is natural to ask whether we have considered the remittance from migrant workers and how the effects of industrial clusters on rural income may be influenced by migration. Although we have discussed the difference between rural migration and working in local industrial clusters in terms of the impacts on rural residents in the previous section, we conduct a series of additional estimates to empirically estimate whether our original findings may remain when considering the rural migration.

First, we replicate the estimations of Tables 2, 3, and 5 by including an additional control variable, the out-migration of residents from the county. "Out-migration" is the ratio of the number of people with Hukou (residence registration) in a county to the number of people who have been living there for more than six months. We obtain such information from China's 2000 population census. A higher ratio of this variable indicates that a county has more out-migrants employed outside the county. We focus on total income, nonfarm income, and major income inequality variables. The results are presented in Tables A-11 and A-12. As shown in the tables, our main results regarding the effects of industrial clusters defined by the DBI on rural income, income inequality and the nonfarm income of disadvantaged groups remain when we control for the outflow of migrant workers.

Second, we replicate the estimations of Tables 2, 3 and 5 by including an additional control variable, the county-level ratio of households that received remittance income from family members that work outside the county in the year of the CHIP survey (denoted by Remittance_ratio). The results are reported in Tables A-13 and A-14. It shows that our main results stay after controlling the ratio of households receiving remittance from migrant workers.

We believe that one of the main contributions of this study lies precisely in our finding that the unique type of industrial clusters in China provides an alternative mechanism for a large number of rural residents to engage in nonfarm activities without having to migrate far, thereby reducing income inequality among rural households while stimulating income growth at the same time. Moreover, the estimations in which we replace the DBI with indicators of urbanization and industrialization and the estimations in which we control for the out-migration or the ratio of households receiving remittance have supported our arguments.

6. Conclusion

In this study, we link three significant phenomena in the industrialization process in China, i.e., industrial clustering, poverty reduction, and income inequality. First of all, we find that clustering increases the total income of rural households, mainly through the increase of their nonfarm income from either wage or business income. More importantly, the intra-county household income inequality is significantly reduced in counties with industrial clusters. Industrial clusters also help reduce the extreme poverty rate and facilitate smoothing income gaps between the wealthiest households and others. Finally, our empirical analysis suggests that clustering creates more equal opportunities for rural households, especially those with disadvantaged groups, to participate in non-agricultural production that reduces their inequality. We further provide systematic evidence that specialization, urbanization and industrialization, measured in standard ways, do not have such effects on rural household income or inequality.

This study contributes to the literature on development economics, growth, inequality, and economic geography. The discoveries suggest that growth and worsened income inequality do not necessarily coexist in the early stages of industrialization. Instead, institutions that shape industrialization and the relationship between the players in the process simultaneously determine growth and income distribution. By nature, such discussions eventually address an important while under-investigated question, i.e., when is growth pro the poor? Such findings also enrich the understanding of the tradeoffs of agglomeration by suggesting that the effects of clustering on growth and inequality are conditional on institutions.

Several challenging questions arising from our discoveries require further research. Our finding on the simultaneous existence of the strong growth-enhancing effect and inequality-reducing effect of industrial clusters indicates a possibility that the Schumpeterian growth mechanism (e.g., Aghion, Howitt, & Violante, 2002) is at work. However, what will happen when the economy further develops and the country steps into a relatively industrialized society? What will happen when the institutional constraints are relaxed and production factors are entirely mobile? Will such development in economy and institutions change the structure and the way of coordination of the industrial clusters, and whether such changes lead to a more equal or more unequal society? These future research directions promise further contribution to the literature on economic development and institutions and knowledge on growth and inequality, economic geography, urban economics, and China's economic development.

Finally, the findings of this study have significant implications for policymaking. Reduction of poverty in rural China and smoothing income gaps between socioeconomic groups have been one of the central goals of the leadership in recent years. Various policies such as reducing agricultural tax, fiscal transfer, and living standard guarantee programs have been implemented to achieve such goals, while to what extent the government policies help relax the poverty and income inequality are not clear (Li & Sicular, 2014). As the first study which links industrial clusters and rural income in China with systematic analysis on the mechanisms, our findings shed some light on the discussion regarding the factors underlying income distributions in rural China. This study implies that institutions that create equal opportunities for peasants to participate in nonfarm activities are crucial to increasing rural income and reducing inequality. Of course, many other dimensions of development and income inequality, such as the policies mentioned above, social capital, education, and economic endowment, are essential while beyond this study's scope. As a starting point, this study leaves more questions to be answered and calls for further examination on industrial clusters and links the phenomenon to more general macroeconomic realities of China's growth and income inequality.

References

Acemoglu, D., Johnson, S., & Robinson, J. A. (2002). Reversal of Fortune: Geography and institutions in the making of the modem world income distribution. *Quarterly Journal of Economics 118*, 1231-1294.

Acemoglu, D., Johnson, S., & Robinson, J. A. (2005). Institutions as a fundamental cause of long-run growth. *Handbook of Economic Growth 1*, 385-472.

Adamopoulos, T., Brandt, L., Leight, J., & Restuccia, D. (2017). Misallocation, selection and productivity: A quantitative analysis with panel data from China. National Bureau of Economic Research Working Paper No. w23039.

Aghion, P., Howitt, P., & Violante, G. L. (2002). General purpose technology and wage inequality. *Journal of Economic Growth*, *7*(4), 315-345.

Ahluwalia, Montek S, and Hollis Chenery. 1974. The economic framework. *Redistribution with Growth*, 38-51.

Aker, J. C., Ghosh, I., & Burrell, J. (2016). The promise (and pitfalls) of ICT for agriculture initiatives. *Agricultural Economics*, *47*(S1), 35-48.

Au, C. C., & Henderson, J. V. (2006). How migration restrictions limit agglomeration and productivity in China. *Journal of Development Economics, 80*(2), 350-388.

Barham, B., & Boucher, S. (1998). Migration, remittances, and inequality: estimating the net effects of migration on income distribution. *Journal of Development Economics*, *55*(2), 307-331.

Barrett, C., Lee, D., & McPeak, J. (2005). Institutional arrangements for rural poverty reduction and resource conservation. *World Development, 33*(2), 193-197.

Benjamin, D., Brandt, L., & Giles, J. (2005). The evolution of income inequality in rural China. *Economic Development and Cultural Change*, *53*(4), 769-824.

Bramall, C. (2008). Rural industrialisation and spatial inequality in China, 1978-2006. *Economic and Political Weekly, 43*(52), 43-50.

Byrd, W. & Lin, Q. (Eds.). (1990). China's rural industry: Structure, development, and reform. Oxford University Press.

Castells-Quintana, D., & Royuela, V. (2014). Agglomeration, inequality and economic growth. *The Annals of Regional Science, 52*(2), 343-366.

Chang, C., & Wang, Y. (1994). The nature of the township-village enterprise. *Journal of Comparative Economics, 19*(3), 434-452.

Che, J., & Qian, Y. (1998). Institutional environment, community government, and corporate governance: Understanding China's township-village enterprises. *Journal of Law, Economics, & Organization, 14*(1), 1-23.

Chu, T., & Wen, Q. (2019). Does College Education Promote Entrepreneurship in China?. *Journal of Labor Research, 40*(4), 463-486.

Cingano, F., & Schivardi, F. (2004). Identifying the sources of local productivity growth. *Journal of the European Economic Association, 2*(4), 720-742.

Coleman, J. (1990). Commentary: Social institutions and social theory. *American Sociological Review, 55*(3), 333-339.

Combes, P. P., Duranton, G. & Gobillon, L. (2019). The costs of agglomeration: House and land prices in French cities. *The Review of Economic Studies, 86*(4), 1556-1589.

Couture, V., Benjamin, F., Gu, Y., & Liu, L. (2021). Connecting the countryside via e-commerce: evidence from China. *American Economic Review: Insights, 3*(1), 35-50.

De Janvry, A., Sadoulet, E. & Zhu, N. (2005). The role of non-farm incomes in reducing rural poverty and inequality in China. CUDARE Working Paper No. 1001.

Deichmann, U., Goyal, A., & Mishra, D. (2016). Will digital technologies transform agriculture in developing countries?. *Agricultural Economics, 47*(S1): 21-33.

Deininger, K., & Squire, L. (1998). New ways of looking at old issues: inequality and growth. *Journal of Development Economics, 57*(2), 259-287.

Easterly, W. (2007). Inequality does cause underdevelopment: Insights from a new instrument. *Journal of Development Economics, 84*(2), 755-776.

Ellison, G., Glaeser, E. L., & Kerr, W. R. (2010). What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review, 100*(3), 1195-1213.

Emran, M. S., & Hou, Z. (2013). Access to markets and rural poverty: evidence from household consumption in China. *Review of Economics and Statistics, 95*(2), 682-697.

Fei, H. (1953). *China's Gentry: Essays on Rural-Urban Relations*. University of Chicago Press.

Freedman, M. (1966). *Chinese Lineage and Society: Fujian and Guandong*. London: Athlone.

Fujita, M., Krugman, P. R., & Venables, A. (1999). *The spatial economy: Cities, regions, and international trade.* MIT press.

Gibson, J., Datt, G., Murgai, R., & Ravallion, M. (2017). For India's rural poor, growing towns matter more than growing cities. *World Development, 98*, 413-429.

Giles, J., & Mu, R. (2018). Village political economy, land tenure insecurity, and the rural to urban migration decision: evidence from China. *American Journal of Agricultural Economics, 100*(2), 521-544.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A. & Shleifer, A. (1992). Growth in cities. *Journal of Political Economy, 100*(6), 1126-1152.

Guo, D, Jiang, K., Xu, C. & Yang, X. (2020). Clustering, growth and inequality in China. *Journal of Economic Geography, 20*(5), 1207-1239.

Haggblade, S., Hazell, P., & Reardon, T. (2010). The rural non-farm economy: Prospects for growth and poverty reduction. *World Development, 38*(10), 1429-1441.

Hausmann, R., & Klinger, B. (2007). The structure of the product space and the evolution of comparative advantage. CID Working Paper Series.

He, Guo., & Wang, S. (2017). Do college graduates serving as village officials help rural China? *American Economic Journal: Applied Economics, 9*(4), 186-215.

Howell, A. (2017). Impacts of migration and remittances on ethnic income inequality in rural China. *World Development, 94,* 200-211.

Huang, Z., Zhang, X. & Zhu, Y. (2008). The role of clustering in rural industrialization: A case study of the footwear industry in Wenzhou. *China Economic Review, 19*(3), 409-420.

Iqbal, M. A., Abbas, A., Ullah, R., Ahmed, U. I., Sher, A., & Akhtar, S. (2018). Effect of non-farm income on poverty and income inequality: farm households evidence from Punjab province Pakistan. *Sarhad Journal of Agriculture, 34*(2), 233-239.

Jacobs, J. (1969). *The economy of cities.* Vintage.

Jalan, J., & Ravallion, M. (2002). Geographic poverty traps? A micro model of consumption growth in rural China. *Journal of Applied Econometrics, 17*(4), 329-346.

Kanbur, R., & Venables, A. J. (Eds.). (2005). *Spatial inequality and development.* OUP Oxford.

Khan, M. M. H. (2001). *Rural poverty in developing countries: Implications for public policy.* International Monetary Fund.

Krugman, P. R. (1991). *Geography and trade.* MIT press.

Kuznets, S. (1955). Economic growth and income inequality. *The American Economic Review, 45*(1), 1-28.

Kuznets, S. (1963). Quantitative aspects of the economic growth of nations: VIII. Distribution of income by size. *Economic Development and Cultural Change, 11*(2), 1-92.

Lee, N., & Sissons, P. (2016). Inclusive growth? The relationship between economic growth and poverty in British cities. *Environment and Planning A: Economy and Space, 48*(11), 2317-2339.

Li, D. D. (1996). A theory of ambiguous property rights in transition economies: The case of the Chinese non-state sector. *Journal of Comparative Economics, 23*(1), 1-19.

Li, S., & Sicular, T. (2014). The distribution of household income in China: Inequality, poverty and policies. *The China Quarterly, 217,* 1-41.

Lin, Y. (2019). E-urbanism: E-commerce, migration, and the transformation of Taobao villages in urban China. *Cities, 91*, 202-212.

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*(11), 35-37, 46.

Liu, R., Wong, T. C., & Liu, S. (2012). Peasants' counterplots against the state monopoly of the rural urbanization process: urban villages and 'small property housing' in Beijing, China. *Environment and Planning A, 44*(5), 1219-1240.

Liu, M., & Yu, J. (2008). Financial structure, development of small and medium enterprises, and income distribution in the People's Republic of China. *Asian Development Review, 25*(1-2), 137-155.

Liu, D., Yu, X., Wang, Y., Zhang, H., & Ren, G. (2014). The impact of perception of discrimination and sense of belonging on the loneliness of the children of Chinese migrant workers: A structural equation modeling analysis. *International Journal of Mental Health Systems, 8*(1), 1-6.

Liu, M., Zhang, Q., Cao, S., & Huang, J. (2020). The spatial aggregation of rural e-commerce in China: An empirical investigation into Taobao Villages. *Journal of Rural Studies, 80,* 403-417.

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

Long, C., & Zhang, X. (2012). Patterns of China's industrialization: Concentration, specialization, and clustering. *China Economic Review, 23*(3), 593-612.

Lu, J., & Tao, Z. (2010). Determinants of entrepreneurial activities in China. *Journal of Business Venturing, 25*(3), 261-273.

Marshall, A. (1890). *Principles of economics: An introductory volume.* London: Macmillan.

Martin, P., & Ottaviano, G. I. (1999). Growing locations: Industry location in a model of endogenous growth. *European Economic Review, 43*(2), 281-302.

Martin, P., & Ottaviano, G. I. (2001). Growth and agglomeration. *International Economic Review, 42*(4), 947-968.

Meng, X. (2012). Labor market outcomes and reforms in China. *Journal of Economic Perspectives, 26*(4), 75-102.

Midelfart-Knarvik, K. H., Overman, H. G., Redding, S. J., & Venables. A. J. (2000). The location of European industry. Working paper.

Mukherjee, A., & Zhang, X. (2007). Rural industrialization in China and India: role of policies and institutions. *World Development, 35*(10), 1621-1634.

Naschold, F. (2012). The poor stay poor: Household asset poverty traps in rural semi-arid India. *World Development, 40*(10), 2033-2043.

OECD. (2017). OECD Economic Surveys: China 2017. Retrieved from http://www.oecd.org/economy/surveys/china-2017-OECD-economic-survey-overview.pdf. Accessed March 16, 2020.

Paukert, F. (1973). Income distribution at different levels of development: A survey of evidence. *Int'l Lab. Rev, 108*, 97.

Peng, Y. (2010). When formal laws and informal norms collide: Lineage networks versus birth control policy in China. *American Journal of Sociology, 116*(3), 770-805.

Porter, M. (1990). *The competitive advantage of nations.* London: Macmillan.

Porter, M. (2003). The economic performance of regions. *Regional Studies, 37*(6-7), 549-578.

Putnam, R. (2000). Bowling alone: America's declining social capital. In *Culture and politics* (pp. 223-234). New York: Palgrave Macmillan.

Qi, J., Zheng, X., & Guo, H. (2019). The formation of Taobao villages in China. *China Economic Review, 53*, 106-127.

Qian, Y., & Xu, C. (1993). Why China's economic reforms differ: The M-form hierarchy and entry/expansion of the non-state sector. *The Economics of Transition, 1*(2), 135-170.

Ravallion, M., & Chen, S. (1999). When economic reform is faster than statistical reform: Measuring and explaining income inequality in rural China. *Oxford Bulletin of Economics and Statistics, 61*(1), 33-56.

Rotemberg, J. J., & Saloner, G. (2000). Visionaries, managers, and strategic direction. *RAND Journal of Economics*, 693-716.

Rozelle, S. (1994). Rural industrialization and increasing inequality: Emerging patterns in China′ s reforming economy. *Journal of Comparative Economics, 19*(3), 362-391.

Ruan, J., & Zhang, X. (2009). Finance and cluster-based industrial development in China. *Economic Development and Cultural Change, 58*(1), 143-164.

Shen, I., Docquier, F., & Rapoport, H. (2010). Remittances and inequality: a dynamic migration model. *The Journal of Economic Inequality, 8*(2), 197-220.

Simoes, N., Crespo, N., & Moreira, S. (2016). Individual determinants of self-employment entry: What do we really know?. *Journal of Economic Surveys, 30*(4), 783-806.

Smart, A., & Smart, J. (2001). Local citizenship: welfare reform urban/rural status, and exclusion in China. *Environment and Planning A, 33*(10), 1853-1869.

Sonobe, T., Hu, D., & Otsuka, K. (2002). Process of cluster formation in China: A case study of a garment town. *Journal of Development Studies, 39*(1), 118-139.

Sulemana, I., Nketiah-Amponsah, E., Codjoe, E. A., & Andoh, J. A. N. (2019). Urbanization and income inequality in Sub-Saharan Africa. *Sustainable Cities and Society, 48*, 101544.

Tsai, L. (2002). Cadres, temple and lineage institutions, and governance in rural China. *The China Journal,* (48), 1-27.

Walz, U. (1996). Transport costs, intermediate goods, and localized growth. *Regional Science and Urban Economics, 26*(6), 671-695.

Wan, G., & Zhou, Z. (2005). Income inequality in rural China: Regression-based decomposition using household data. *Review of Development Economics, 9*(1), 107-120.

Wang, H., Guo, F., & Cheng, Z. (2015). Discrimination in migrant workers' welfare entitlements and benefits in urban labour market: Findings from a four-city study in China. *Population, Space and Place, 21*(2), 124-139.

Weber, A. (1929). *Theory of the Location of Industries.* University of Chicago Press.

Weitzman, M. L., & Xu, C. (1994). Chinese township-village enterprises as vaguely defined cooperatives. *Journal of Comparative Economics, 18*(2), 121-145.

Whalley, J., & Zhang, S. (2004). Inequality change in China and (Hukou) labour mobility restrictions. National Bureau of Economic Research Working Paper No. 10683.

World Bank. (2017). China: systematic country diagnostic. Retrieved from http://documents.worldbank.org/curated/en/147231519162198351/pdf/China-SCD-publishing-version-final-for-submission-02142018.pdf. Accessed March 16, 2020.

Xu, C. (2011). The fundamental institutions of China's reforms and development. *Journal of Economic Literature, 49*(4), 1076-1151.

Xu, C., & Zhang, X. (2009). The evolution of Chinese entrepreneurial firms: Township-village enterprises revisited. Vol. 854. Intl Food Policy Res Inst Working Paper No.00854.

Yao, S. (1999). Economic growth, income inequality and poverty in China under economic reforms. *The Journal of Development Studies, 35*(6), 104-130.

Young, A. (2013). Inequality, the urban-rural gap, and migration. *The Quarterly Journal of Economics, 128*(4), 1727-1785.

Zhu, B., Song, Y., Li, G., & Yu, T. (2016). Spatial aggregation pattern and influencing factors of Taobao village in China under the C2C e-commerce mode. *Economic Geography*, *36*(4), 92-98.

Zhu, N., & Luo, X. (2010). The impact of migration on rural poverty and inequality: a case study in China. *Agricultural Economics*, *41*(2), 191-204.

Table 1. Summary Statistics

Panel A: Income Statistics in Counties with and without Clusters

	Mean	Median	Min	Max	SD	Obs
Counties with Clusters						
Total income	4959.73	4724.23	937.99	20409.19	2887.22	174
Nonfarm income	2651.42	2047.56	86.95	12032.51	2227.34	174
Farm income	1528.17	1415.10	0.00	3976.12	796.55	174
Share nonfarm	0.47	0.47	0.09	0.83	0.18	174
Share farm	0.38	0.39	0.00	0.83	0.19	174
Counties without Clusters						
Total income	3087.10	2779.67	704.03	10807.75	1520.96	239
Nonfarm income	1253.83	979.35	42.55	8342.63	1060.49	239
Farm income	1480.15	1273.41	307.74	4532.85	784.59	239
Share nonfarm	0.37	0.38	0.04	0.87	0.17	239
Share farm	0.50	0.51	0.08	0.89	0.16	239
Full sample						
Total income	3876.05	3366.37	704.03	20409.19	2386.15	413
Nonfarm income	1842.64	1309.22	42.55	12032.51	1791.70	413
Farm income	1500.38	1346.25	0.00	4532.85	789.04	413
Share nonfarm	0.41	0.41	0.04	0.87	0.18	413
Share farm	0.45	0.47	0.00	0.89	0.18	413

Panel B: Total Income Inequality Statistics in Counties with and without Clusters

	Mean	Median	Min	Max	SD	Obs
Counties with Clusters						
Gini	0.28	0.27	0.16	0.46	0.06	174
Theil	0.14	0.12	0.05	0.53	0.08	174
MLD	0.14	0.13	0.05	0.37	0.07	174
R10	0.24	0.22	0.15	0.41	0.05	174
P10	0.04	0.04	0.01	0.07	0.01	174
PovertyR_1	0.05	0.01	0	0.76	0.11	174
PovertyR_1.25	0.09	0.03	0	0.90	0.16	174
Counties without Clusters						
Gini	0.27	0.26	0.11	0.55	0.07	239
Theil	0.14	0.12	0.02	0.64	0.09	239
MLD	0.14	0.12	0.02	0.51	0.08	239
R10	0.24	0.23	0.15	0.49	0.06	239
P10	0.04	0.04	0.01	0.09	0.01	239
PovertyR_1	0.11	0.04	0	0.82	0.16	239
PovertyR_1.25	0.18	0.10	0	0.92	0.21	239
Full Sample						
Gini	0.28	0.27	0.11	0.55	0.07	413
Theil	0.14	0.12	0.02	0.64	0.09	413
MLD	0.14	0.12	0.02	0.51	0.07	413
R10	0.24	0.23	0.15	0.49	0.05	413
P10	0.04	0.04	0.01	0.09	0.01	413
PovertyR_1	0.08	0.03	0	0.82	0.14	413
PovertyR_1.25	0.14	0.05	0	0.92	0.19	413

Table 2. Industrial clustering and household per capita income

	(1)	(2)	(3)	(4)	(5)
	Total income	Nonfarm income	Farm income	Share nonfarm	Share farm
Cluster	0.085**	0.187**	0.249	0.035*	-0.022
	(2.030)	(2.441)	(1.486)	(1.755)	(-1.122)
Per capita GDP	0.016***	0.017***	-0.060	0.002**	-0.004***
	(6.932)	(4.835)	(-1.622)	(2.409)	(-4.825)
Education expenditure	-0.919**	-2.181***	-0.846	-0.350*	0.400**
	(-2.569)	(-3.305)	(-0.836)	(-1.949)	(2.166)
Agricultural expenditure	-0.691	-3.685***	3.553	-0.867**	0.994***
	(-1.035)	(-2.974)	(1.419)	(-2.544)	(2.730)
Number of SEZs	0.079**	0.118**	0.300	0.011	-0.022
	(2.414)	(2.297)	(1.205)	(0.724)	(-1.395)
Middle-region dummy	-0.289***	-0.474***	-0.226	-0.068***	0.094***
	(-5.229)	(-4.934)	(-1.351)	(-2.777)	(3.631)
Western-region dummy	-0.447***	-0.698***	-0.309*	-0.084***	0.121***
	(-6.820)	(-6.095)	(-1.871)	(-3.366)	(4.476)
Year-2002 dummy	0.209***	0.382***	0.318***	0.074**	0.023
	(3.083)	(3.033)	(2.677)	(2.525)	(0.772)
Year-2007 dummy	0.476***	1.007***	0.644***	0.177***	-0.024
	(5.140)	(6.139)	(3.561)	(4.245)	(-0.570)
Constant	8.139***	7.568***	7.188***	0.504***	0.236***
	(42.956)	(22.235)	(16.306)	(6.087)	(2.739)
N	284	284	284	284	284
R square	0.743	0.679	0.288	0.389	0.367

Note: Values in parentheses are t- statistics; * = $p < 0.1$; ** = $p < 0.05$; *** = $p < 0.01$.

Table 3. Industrial clustering and household per capita total income inequality

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Gini	Theil	MLD	R10	P10	PovertyR_1	PovertyR_1.25
Cluster	-0.011	-0.023*	-0.018*	-0.017**	-0.001	-0.022+	-0.032*
	(-1.136)	(-1.879)	(-1.757)	(-2.329)	(-0.489)	(-1.576)	(-1.872)
Per capita GDP	0.001**	0.001**	0.001**	0.001***	-0.000	0.000	-0.000
	(2.066)	(2.259)	(2.103)	(2.669)	(-1.636)	(0.682)	(-0.054)
Education expenditure	-0.015	-0.022	-0.041	0.007	0.010	0.359**	0.408**
	(-0.213)	(-0.248)	(-0.528)	(0.124)	(0.828)	(2.417)	(2.146)
Agricultural expenditure	-0.047	-0.045	-0.038	0.026	0.023	0.359*	0.498**
	(-0.274)	(-0.213)	(-0.198)	(0.199)	(0.756)	(1.912)	(2.000)
Number of SEZs	0.011*	0.011	0.014**	0.004	-0.001	-0.005	-0.009
	(1.699)	(1.402)	(1.980)	(0.875)	(-1.022)	(-0.715)	(-0.852)
Middle-region dummy	-0.019*	-0.020	-0.025**	-0.007	0.005***	0.037**	0.059**
	(-1.731)	(-1.423)	(-1.998)	(-0.748)	(2.681)	(2.179)	(2.508)
Western-region dummy	-0.010	-0.012	-0.014	-0.007	0.002	0.120***	0.163***
	(-0.821)	(-0.792)	(-1.066)	(-0.702)	(1.052)	(4.363)	(4.632)
Year-2002 dummy	-0.015	-0.043**	-0.030*	-0.018	0.002	-0.116***	-0.174***
	(-1.023)	(-2.008)	(-1.794)	(-1.397)	(1.170)	(-4.320)	(-5.197)
Year-2007 dummy	-0.020	-0.052*	-0.028	-0.027	0.001	-0.141***	-0.225***
	(-0.960)	(-1.834)	(-1.219)	(-1.557)	(0.299)	(-4.533)	(-5.579)
Constant	0.298***	0.192***	0.180***	0.249***	0.036***	0.015	0.093
	(8.356)	(4.204)	(4.538)	(8.689)	(6.138)	(0.228)	(1.052)
N	284	284	284	284	284	284	284
R square	0.059	0.072	0.085	0.052	0.085	0.390	0.457

Note: Values in parentheses are t- statistics; * = $p < 0.1$; ** = $p < 0.05$; *** = $p < 0.01$; + = $p<0.15$.

Table 4. Industrial clustering and household per capita nonfarm income inequality

	(1)	(2)	(3)	(4)	(5)
	Gini	Theil	MLD	R10	P10
Cluster	-0.044***	-0.131***	-0.609***	-0.029**	0.002
	(-2.823)	(-3.467)	(-2.876)	(-2.141)	(0.641)
Per capita GDP	-0.000	0.002	0.002	0.000	-0.000
	(-0.603)	(1.119)	(0.235)	(0.770)	(-0.229)
Education expenditure	0.237	0.693**	4.931**	-0.029	-0.000
	(1.635)	(2.008)	(2.515)	(-0.253)	(-0.015)
Agricultural expenditure	0.599**	1.344**	6.487**	0.202	-0.015
	(2.402)	(2.405)	(2.050)	(0.948)	(-0.380)
Number of SEZs	0.000	-0.013	-0.053	-0.000	0.001
	(0.041)	(-0.598)	(-0.421)	(-0.019)	(0.515)
Middle-region dummy	0.009	0.044	0.107	0.015	0.005*
	(0.476)	(0.959)	(0.402)	(0.906)	(1.918)
Western-region dummy	0.017	0.046	-0.112	-0.002	0.008**
	(0.822)	(0.923)	(-0.418)	(-0.137)	(2.599)
Year-2002 dummy	-0.131***	-0.383***	-1.251***	-0.146***	0.012***
	(-4.935)	(-5.152)	(-3.435)	(-5.056)	(2.913)
Year-2007 dummy	-0.214***	-0.558***	-1.871***	-0.167***	0.011**
	(-6.358)	(-6.141)	(-4.028)	(-4.925)	(2.299)
Constant	0.542***	0.610***	1.577*	0.391***	0.017*
	(7.868)	(3.779)	(1.671)	(6.838)	(1.715)
N	284	284	284	284	284
R square	0.367	0.405	0.275	0.235	0.073

Note: Values in parentheses are t- statistics; * = p < 0.1; ** = p < 0.05; *** = p < 0.01.

Table 5. Industrial clustering and household per capita farm income inequality

	(1)	(2)	(3)	(4)	(5)
	Gini	Theil	MLD	R10	P10
Cluster	0.028*	0.039	0.114	-0.018	0.002
	(1.724)	(1.110)	(0.731)	(-1.406)	(0.576)
Per capita GDP	0.003	0.012**	0.066***	0.000	0.000
	(1.379)	(2.408)	(2.765)	(0.269)	(1.243)
Education expenditure	-0.159	-0.219	-1.270	0.073	-0.022
	(-1.309)	(-0.821)	(-0.909)	(0.689)	(-0.644)
Agricultural expenditure	-0.129	0.157	4.802	-0.100	0.064
	(-0.476)	(0.264)	(1.381)	(-0.333)	(0.601)
Number of SEZs	0.044**	0.103**	0.455**	0.001	-0.004
	(2.307)	(2.344)	(2.120)	(0.043)	(-1.007)
Middle-region dummy	-0.100***	-0.174***	-0.733***	-0.011	-0.002
	(-5.457)	(-4.249)	(-3.708)	(-0.690)	(-0.436)
Western-region dummy	-0.091***	-0.153***	-0.609***	-0.005	-0.009*
	(-5.085)	(-3.933)	(-3.331)	(-0.327)	(-1.962)
Year-2002 dummy	0.035**	0.016	-0.028	0.007	0.004
	(2.141)	(0.450)	(-0.182)	(0.462)	(1.153)
Year-2007 dummy	0.056**	-0.013	-0.591	0.051	-0.009
	(1.992)	(-0.206)	(-1.506)	(1.648)	(-0.895)
Constant	0.420***	0.318***	0.603	0.165***	0.061***
	(8.036)	(2.717)	(1.032)	(3.173)	(4.300)
N	284	284	284	284	284
R square	0.499	0.504	0.479	0.041	0.057

Note: Values in parentheses are t- statistics; * = p < 0.1; ** = p < 0.05; *** = p < 0.01.

Table 6. Addressing endogeneity using IV estimation

Panel A: First-Stage Regression

	(1)	(2	(3)	(4)	(5)
	Cluster	Cluster	Cluster	Cluster	Cluster
TVE_emp	0.640***	0.593***	0.533***	0.434***	0.472***
	(3.751)	(3.492)	(2.693)	(2.858)	(2.730)
Genealogies	0.003***	0.003***	0.002**	0.003***	0.003***
	(4.427)	(4.575)	(2.012)	(4.151)	(3.507)
Controls	Yes	Yes	Yes	Yes	Yes
N	224	224	224	224	224
Sargan test p-value	0.725	0.594	0.559	0.307	0.251

Panel B: Second-Stage Regressions on Clustering and Household Per Capita Income

	(1)	(2)	(3)	(4)	(5)
	Total Income	Nonfarm income	Farm income	Share nonfarm	Share farm
Cluster	0.475***	0.950***	0.314	0.300***	-0.245***
	(6.201)	(6.758)	(1.082)	(6.418)	(-4.028)
Per capita GDP	0.017***	0.017***	-0.007*	0.002*	-0.003**
	(6.449)	(3.537)	(-1.669)	(1.851)	(-2.283)
Education expenditure	-0.790**	-2.445***	0.437	-0.491***	0.610***
	(-2.178)	(-3.567)	(0.700)	(-2.728)	(3.116)
Agricultural expenditure	-0.049	-3.328**	2.764**	-0.957***	1.192***
	(-0.071)	(-2.544)	(2.294)	(-2.826)	(3.187)
Number of SEZs	0.039	0.059	-0.044	-0.001	-0.019
	(1.255)	(1.000)	(-0.825)	(-0.053)	(-1.106)
Middle-region dummy	-0.264***	-0.463***	-0.041	-0.069***	0.098***
	(-5.819)	(-5.443)	(-0.520)	(-3.133)	(4.071)
Western-region dummy	-0.385***	-0.622***	-0.155*	-0.070**	0.104***
	(-7.175)	(-6.101)	(-1.680)	(-2.539)	(3.531)
Year-2002 dummy	0.183***	0.223*	0.313***	0.030	0.060*
	(2.753)	(1.776)	(2.721)	(0.912)	(1.660)
Year-2007 dummy	0.384***	0.831***	0.457***	0.143***	-0.011
	(4.325)	(4.969)	(2.990)	(3.259)	(-0.242)
Constant	7.950***	7.459***	6.556***	0.475***	0.227**
	(47.106)	(23.013)	(21.053)	(5.525)	(2.419)
N	224	224	224	224	224
Wald Chi-sqr	493.45	413.05	50.04	185.34	111.70
Prob > F	0.0000	0.0000	0.0000	0.0000	0.0000

Panel C: Second-Stage Regressions on Clustering and Household Per Capita Total Income Inequality

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Gini	Theil	MLD	R10	P10	PovertyR_1	PovertyR_1.25
Cluster	-0.068*	-0.076**	-0.067**	-0.046*	0.018***	-0.117***	-0.185***
	(0.035)	(0.032)	(0.029)	(0.025)	(0.004)	(0.033)	(0.040)
Per capita GDP	0.001	0.001*	0.001	0.001**	-0.000	0.00003	-0.001
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
Education expenditure	0.026	0.022	-0.008	0.033	0.011	0.284**	0.305*
	(0.091)	(0.115)	(0.099)	(0.073)	(0.015)	(0.140)	(0.176)
Agricultural expenditure	-0.136	-0.110	-0.152	-0.038	0.043	0.166	0.185
	(0.176)	(0.220)	(0.191)	(0.141)	(0.029)	(0.268)	(0.335)
Number of SEZs	0.009	0.007	0.012	0.001	-0.002	0.002	0.003
	(0.008)	(0.010)	(0.009)	(0.006)	(0.001)	(0.012)	(0.015)
Middle-region dummy	-0.015	-0.016	-0.022*	-0.005	0.004**	0.027	0.044**
	(0.011)	(0.014)	(0.012)	(0.009)	(0.002)	(0.017)	(0.022)
Western-region dummy	-0.003	-0.002	-0.006	-0.000	0.002	0.098***	0.131***
	(0.013)	(0.017)	(0.015)	(0.011)	(0.002)	(0.021)	(0.026)
Year-2002 dummy	-0.015	-0.045**	-0.031*	-0.019	0.003	-0.093***	-0.143***
	(0.017)	(0.021)	(0.018)	(0.013)	(0.003)	(0.026)	(0.032)
Year-2007 dummy	-0.006	-0.042	-0.015	-0.019	-0.001	-0.097***	-0.161***
	(0.022)	(0.028)	(0.024)	(0.018)	(0.004)	(0.034)	(0.043)
Constant	0.311***	0.201***	0.195***	0.256***	0.027***	0.062	0.171**
	(0.044)	(0.055)	(0.048)	(0.035)	(0.007)	(0.066)	(0.082)
N	224	224	224	224	224	224	224
Wald Chi-sqr	12.79	19.31	19.51	12.31	36.38	99.59	141.95
Prob > F	0.1725	0.0227	0.0212	0.1965	0.0000	0.0000	0.0000

Note: Values in parentheses are t- statistics; * = p < 0.1; ** = p < 0.05; *** = p < 0.01.

Table 7. Remittance income in counties with and without clusters

	Share of households with remittance income	Obs	Share of remittance income to total household income	Obs
Counties with cluster	0.289	195	0.101	10,986
Counties without cluster	0.334	322	0.127	13,655
Difference	-0.045**		-0.026***	

Note: * = p < 0.1; ** = p < 0.05; *** = p < 0.01.

Table 8. Household per capita income in counties with and without clusters: Comparison between households with and without migrant remittance

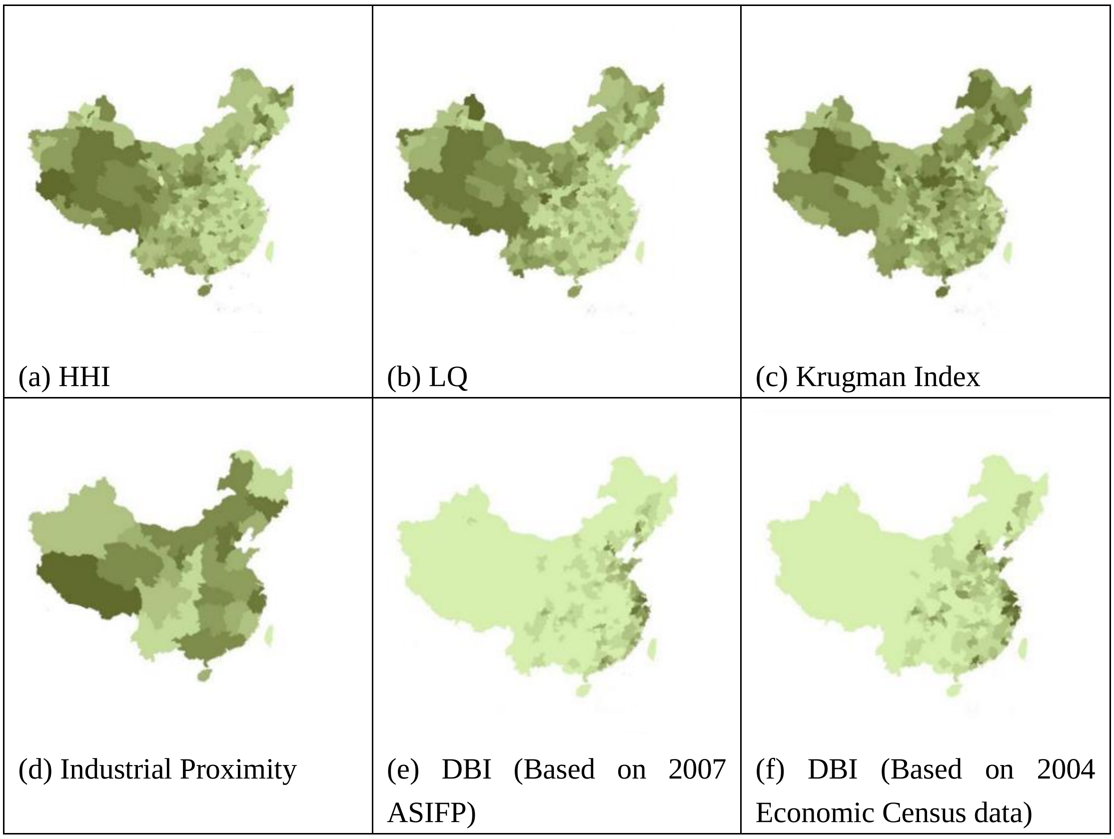
Average per capita income of households	Counties with cluster	Obs	Counties without cluster	Obs
All households	4,798	10,986	3,019	13,655
Households with migrant remittance	4,927	3,138	3,430	4,815
Households without migrant remittance	4747	7,848	2,795	8,840
Difference between households with and without remittance	180*		635***	

Note: * = p < 0.1; ** = p < 0.05; *** = p < 0.01.

Table 9. Industrial clustering, disadvantaged groups, and household nonfarm Income

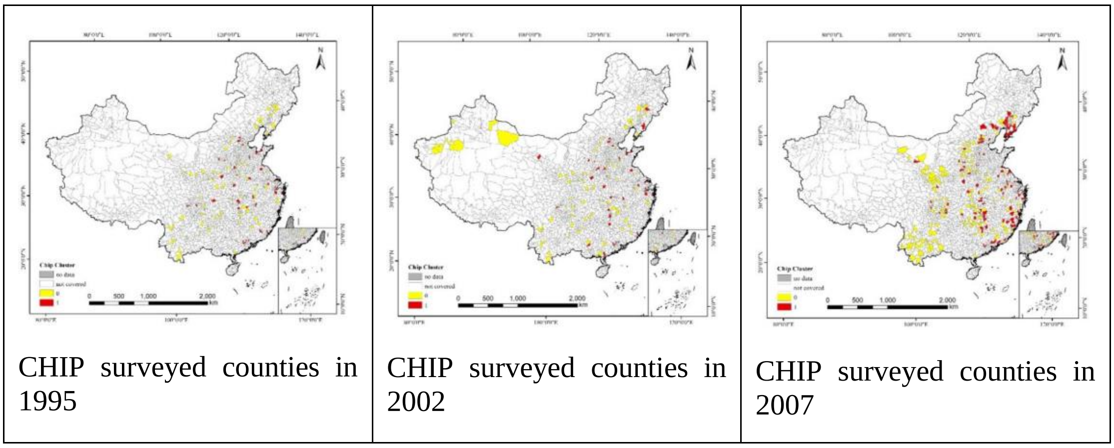
	(1)	(2)	(3)
Cluster	0.451***	0.361***	0.483***
	(0.087)	(0.103)	(0.079)
Cluster*Ratio elderly	0.377*		
	(0.220)		
Cluster*Ratio less-educated		0.514**	
		(0.223)	
Cluster*Dummy unhealthy			0.751**
			(0.357)
Ratio elderly	-1.279***	-1.133***	-1.117***
	(0.145)	(0.120)	(0.121)
Ratio less-educated	-0.316***	-0.501***	-0.305**
	(0.119)	(0.138)	(0.119)
Dummy unhealthy	-0.078	-0.069	-0.347*
	(0.168)	(0.168)	(0.199)
Ratio male	0.374**	0.385**	0.371**
	(0.163)	(0.163)	(0.163)
Total household members	0.113***	0.114***	0.114***
	(0.025)	(0.025)	(0.025)
Per capita GDP	0.022***	0.023***	0.022***
	(0.004)	(0.004)	(0.004)
Education expenditure	0.097	0.131	0.130
	(0.608)	(0.607)	(0.607)
Agricultural expenditure	-12.109***	-12.071***	-12.088***
	(1.138)	(1.138)	(1.138)
Number of SEZs	-0.079	-0.081	-0.080
	(0.055)	(0.055)	(0.055)
Middle-region dummy	-0.353***	-0.346***	-0.350***
	(0.084)	(0.084)	(0.084)
Western-region dummy	0.136	0.155*	0.140
	(0.092)	(0.092)	(0.092)
Year-2007 dummy	-0.302***	-0.315***	-0.304***
	(0.115)	(0.115)	(0.115)
Constant	3.317***	3.327***	3.285***
	(0.269)	(0.269)	(0.268)
N	13925	13925	13925
R square	0.043	0.044	0.044

Note: Values in parentheses are t- statistics; * = $p < 0.1$; ** = $p < 0.05$; *** = $p < 0.01$.



Note: (i) the HHI, LQ, and Krugman Index are calculated from the ASIFP 2007 data, and the Industrial Proximity is from Long and Zhang (2011) and is based on the 2004 Economic Census data; (ii) The DBI cluster shown in the maps are at city-level, which is the summation of county-level clusters within the same city; (iii) darker area represents a higher level of measured industrial clustering.

Figure 1: Industrial clustering measured by existing indexes and DBI



Note: (i) counties marked in either red or yellow are surveyed by CHIP; (ii) red indicates counties with clusters, and yellow indicates counties without clusters.

Figure 2: Geographic distribution of CHIP surveyed counties in the three waves

Endnote

[i] In May of 2020, Premier Li Keqiang stated that there are over 600 million people whose monthly income is barely RMB 1,000 (USD 140), meaning approximately 40 percentage of the total population live under USD 5 per day, and, the majority of the poor are rural residents. As of 2015, approximately 10 million (0.7% of the total population) and 100 million people were still living below the international poverty line of $1.90 and $3.20 (in 2011 purchasing power parity), respectively. In May of 2020, Premier Li Keqiang stated that there are over 600 million people whose monthly income is barely RMB 1,000 (USD 140), meaning approximately 40 percentage of the total population live under USD 5 per day, and, the majority of the poor are rural residents.

[ii] For instance, one-third of the world's socks, 40% of the world's neckties, and 60% of China's cashmere sweaters were produced in the towns of Datang, Shengzhou, and Puyuan, respectively (Xu & Zhang, 2009).

[iii] According to the constitution of China, rural land is collectively owned at the village level. Nationalization is required as the first legal step for trading collectively owned rural land for non-agricultural use (Liu, Wong & Liu, 2012).

[iv] The purpose of the Chinese Household Income Project was to measure and estimate the distribution of personal income in both rural and urban areas of China. Definition of income is based on cash payments and on a broad range of additional components such as agricultural output produced for self-consumption valued at market prices, the value of ration coupons and other direct subsidies, and the imputed value of housing. Data were collected through a series of questionnaire-based interviews conducted in rural and urban areas in 1988, 1995, 2002, 2007 and 2013. The reason we focus on the year of 1995, 2002, and 2007 is that ASIFP consistently covers these years while the data collection criteria becomes different for ASIFP after 2007.

[v] For instance, the Gini measures how far away a country or region is from an equal distribution in which each industry produces the same share of output or value added.

Midelfart–Knarvik et al. (2000) use Gini to explore industrial location changes in terms of spatial concentration in Europe. LQ is an analytical statistic that measures a region's industrial specialization relative to a larger geographic unit (usually the nation). Glaeser et al. (1992) apply LQ as a specialization measure of an industry in a city and test its effect on city-industry employment growth. Porter (2003) utilizes LQ as an important criterion in defining traded industries that form clusters. Krugman (1991) constructs a dissimilarity index focusing on the deviation of a region's industry structure from the average industry structure of a regional reference group to reveal a region's comparative advantage.

[vi] For testing robustness, we also try other α values, such as 3 or 10, and our main results are not affected by the choice of α.

[vii] In this study, when we define the nonfarm income, we focus on the wages and income from business activities while we do not count subsidies, transfers, pensions, interest, etc., because we are interest in the direct income related to the cluster-related activities.

[viii] Taking net disposable household per capita income as an example, the skewness of this variable is as high as 1.73 indicating that the distribution is highly skewed, while the skewness of its logarithmic transformation is only 0.1 showing that the distribution is approximately symmetric after the logarithmic transformation.