Clustering, Growth, and Inequality in China*

Di Guo, Brunel University London
Kun Jiang, The University of Nottingham
Chenggang Xu, Cheung Kong Graduate School of Business
Xiyi Yang, ShanghaiTech University

Abstract

This study examines the effects of China's industrial clusters on regional economic growth and urban-rural income inequality within a region. A density-based index (DBI) is developed to capture the unique features of cluster development in China. From a county-level DBI panel data constructed based on firm-level and county-level datasets, we find that clusters enhance local economic growth substantially. Moreover, the existence of entrepreneurial clusters (clusters mainly consist of non-state-owned firms) helps to reduce local urban-rural income inequality by increasing the income of local rural residents. We also find that the clustering effects on growth and reduction of inequality are less significant in more urbanized regions or megacities. Identification issues are carefully addressed by deploying two-stage estimations with instrumental variables and Granger test.

Keywords: China, clustering, geography, growth and inequality, institutions
JEL Classification: D2, H7, O1, R1, R3

---

* Corresponding author: Chenggang Xu, Cheung Kong Graduate School of Business, Oriental Plaza 2/F, Tower E2, 1 East Changan Ave., Beijing, CN 100738 (e-mail: cgxu@ckgsb.edu.cn). Di Guo, Kun Jiang, and Chenggang Xu acknowledge the financial support from the RGC Theme-based Research Scheme (TRS) (Project #: T31-717 *112*-R). Chenggang Xu acknowledges the financial support from the RGC General Research Fund (Project #: 756711) and CKGSB research support. Kun Jiang acknowledges the financial support from Nottingham University Business School 4*Publication Boost Small Grant. Xiyi Yang acknowledges the financial support from the ShanghaiTech Faculty Startup Funding.

1. Introduction

This study investigates the co-existence of industrial clustering, economic growth, and income inequality and their interactions in China. The post-Mao economic reforms have transformed the world's largest developing country from one of the poorest nations into a major power. The emergence of industrial clusters in numerous towns, mostly along China's coastal areas, is among the most striking developments throughout the said reforms. Considering the rise of industrial clusters as one of the primary engines of China's growth is a warranted assertion (Sonobe and Otsuka, 2006; Long and Zhang, 2011).[1] However, along with China's record-breaking growth was the rapid increase in inequality. China now has become one of the least equal economies in the world (Sicular, 2013), a status that may threaten the country's social stability and economic sustainability. The urban-rural income gap has been a dominant component of the overall inequality (Li et al., 2013). Meanwhile, regional disparity, particularly inland–coastal disparity, remains an important dimension of the increased inequality during the reform era (Kanbur and Zhang, 1999).

This study links the phenomena mentioned above by examining how industrial clusters affect economic growth and inequality in China simultaneously. First, we determine whether clustering associated with different strengths and ownership structures affects local economic growth and its implications for regional disparity. Second, we investigate whether such clustering influences urban-rural income inequality within a region, and if so, through which channel. Third, we examine the heterogeneous effects of industrial clusters focusing on the urbanization level of the regions. In our paper, we are particularly interested in the impacts of clustered entrepreneurial firms, which we call entrepreneurial clusters. Based on available data, the best statistical proxy for entrepreneurial firms in China are non-state-owned firms (non-state firms in short) as discussed in the literature (Xu, 2011; Long and Zhang, 2011).[2]

To address the above research questions, we first define the measurements of clustering in the context of China. In a market economy, the industrial agglomeration is an outcome of the co-location decisions of firms in the spatial equilibrium model (Ellison and Glaeser, 1997; Glaeser and Gottlieb, 2009) or the new economic geography model (Krugman, 1991b). However, production factors such as land, labor, and capital are not freely mobile or accessible in China, because of the state control over ownership of land and major financial institutions, as well as labor mobility through the residence registration (*Hukou*) system. To capture the distinctive features of the industrial clusters created and developed under the institutional restrictions in China, we create a county-level density-based index (DBI) of clusters based on the density of firms of each industry within a county. The indices are constructed based on the firm-level panel dataset from the Above-Scale Industrial Firm Panel (ASIFP) between 1998 and 2007. Applying the DBI measurement, we further construct a panel of county-level cluster indices that measure the existence, strength, and ownership structure of industrial clusters.

Based on our county-level DBI cluster indices, which is a panel of 2,815 Chinese counties from 1998 to 2007, we find that counties with clusters, particularly strong

[1] According to Long and Zhang (2011), 62% of the growth of the number of firms in China from 1995 to 2004 was caused by the rise of these clusters; 14% of China's total industrial GDP growth during the same period was attributed to the firms within clusters.

[2] In the rest of the paper, our discussions designate clusters composed mainly of non-state firms as entrepreneurial clusters. However, when merely presenting statistics, we simply use the term non-state firms or non-state clusters when referring to the data.

clusters (measured by clusters' outputs or establishment) or with entrepreneurial clusters (clusters composed of mainly non-state firms), grow significantly faster than other counties. On average a 1% increase in the clusters' contribution to national industrial output within a country will result in a 1% increase of per capita GDP growth in that county; and a 1% increase in the non-state firms' contribution to the clusters' outputs or establishment number will result in approximately 1.6% increase in the per capita GDP growth. More interestingly, we discover that entrepreneurial clusters not only promote economic growth but also substantially reduce local urban-rural income inequality, and this outcome is driven by the increased income of rural residents in the counties where these clusters are located. An increase of 1% in the outputs or establishment number of non-state firms within the clusters is associated with a 3% reduction in the urban-rural inequality in the county. Moreover, the clustering effects are less significant in mega-city regions (e.g., Beijing, Shanghai, Shenzhen) or more urbanized regions according to *Hukou* registration.

Two-stage estimations and Granger tests are deployed to address identification concerns. Regarding clustering effects on local economic growth, we use per capita mining outputs in a region as an instrumental variable (IV). Mine-rich regions are typically dominated by large companies and entrepreneurship is often weakened (Chinitz, 1961; Rosenthal and Strange, 2003; and Glaeser et al., 2010, 2015). Therefore, provinces with higher per capita mining outputs have weaker industrial clusters and fewer private firms in the clusters. The mine richness of a region is determined geologically so that it is exogenous to the county-level growth. Two-stages estimations using this IV confirm the causal effect of clusters on regional economic growth.

For identifying clustering effects on within-county urban-rural income inequality, we use two IVs: the per capita length of classified highways in a city and the number of Christian churches in a county in a given year. The access to transportation network is expected to be related to the development of clusters because transportation infrastructure helps to reduce the trade and shipping costs, increase the market size, and facilitate knowledge diffusion within a region, which are ultimately important for clustering of firms. Additionally, Christianity culture plays an important role in fostering entrepreneurship and trade and facilitating coordination of local society that are major roots for the development of clusters. These two IVs are relevant to the clustering while are exogenous to the error terms of the estimations for local urban-rural income inequality. The two-stage estimations confirm the validity of the IVs and the causal relationships between clustering, urban-rural inequality and per capita household income of rural residents.

Finally, Granger causality tests further confirm our findings on the clustering effects on regional growth and local urban-rural income inequality. Additional robustness checks further rule out alternative explanations.

Our discoveries contribute to the literature on economic geography and urban economics by providing new evidence for the clustering effects on economic growth. There is no empirical consensus in prior studies on the effects of clusters on growth (Glaeser et al., 1992; Cingano and Schivardi, 2004; Dekle, 2002). Our study complements the existing debates by suggesting that the effects of clustering on growth are conditional on institutions. In our context, it depends on the ownership type of firms within the clusters and the nature of urbanization of the regions. Moreover, using county-level data, this study complements the new micro-geography literature, which suggests that it is important to 'zoom in' to a smaller scale of the geographical territories to gain the insights of the local advantages (e.g., Feldman, 2014; Catalini, 2018; Mudambi et al., 2018).

More importantly, perhaps, this study is the first one which examines the effects of clustering on income inequality, shedding some light on discussions of the tradeoffs of agglomeration (Fujita and Thisse, 2013; Combes et al., 2018). Additionally, the present study enriches the literature on economic development and inequality by suggesting an alternative trend between growth and inequality under different institutions. The relationship between economic growth and inequality raises challenging questions since the beginning of industrialization (Kuznets, 1963; Persson and Tabellini, 1994; Barro, 2000; Knowles, 2001; Banerjee and Duflo, 2003). The findings of this study indicate that the underlying institutions deeply influence the relationship between development and inequality, consisting with the arguments that good institution simultaneously promotes development and reduces inequality (Acemoglu et al., 2002, 2005; Easterly, 2007).

Furthermore, this study contributes to the literature on income inequality in China. From 1984 to 2005, both inter-regional income gap and urban-rural income disparity increase significantly (Fujita and Hu, 2001; Kanbur and Zhang, 1999, 2005). The findings in this paper complement previous studies by providing mechanisms of different types of inequalities. On the one hand, we find clustering increases regional disparity by widening the gap of the growth rates between counties with and without clusters. On the other hand, we find there is a reduction in urban-rural income inequality within a county if there are entrepreneurial clusters in the county.

Finally, our study provides a methodological contribution to the research on economic geography and urban economics. Given the strong institutional constraints to factor mobility in China, directly applying the agglomeration indices used in the existing literature is not the most suitable approach because entrepreneurial industrial clusters would be mixed with or even overwhelmed by the agglomerations dominated by state-owned enterprises (SOEs). The DBI indices of clustering we constructed, however, can capture the institutionally constrained entrepreneurial clusters in China.

The rest of the paper is organized as follows. Section 2 introduces the institutional background of industrial clusters in China. Section 3 discusses the related literature. Section 4 constructs the density-based indices and presents the data. Section 5 reports our empirical findings on clustering, economic growth, and urban-rural inequality with the identification issues addressed. Section 6 concludes this study.

2. Industrial Clusters in China

The market economy in China has been anything but well-functioned that almost all major production factors were not sufficiently mobile, particularly at the beginning of the economic reform when entrepreneurial industrial clusters first emerged. Such institutions determine that the clusters in China vary from those in the market economies significantly.

Above all, entrepreneurs in China face serious institutional constraints that prevent them from choosing locations freely for setting up their businesses. The number one issue is land ownership. According to the constitution, urban land is state-owned, whereas rural land is collectively owned by villages and is not tradable for non-agricultural usage. Peasants, individually or collectively, remain prohibited from trading "their" land for non-agricultural purposes. Before the mid-1990s, the only way for peasants to use their collectively-owned land beyond agriculture activities was to establish industrial firms within their villages or towns, i.e., township–village enterprises (TVEs, see Weitzman and Xu, 1994). Since the late 1990s, when political and legal resistance to private ownership was gradually relaxed, many TVEs have become privatized (Xu, 2011). Such firms have become increasingly specialized and clustered together. With the concentration of a vast number of small and specialized firms, many townships have become national or international production centers of specific products. For instance, in Zhejiang Province, the Songxia Township produces 350 million umbrellas annually, the Qiaotou Township supplies 70% of the buttons for clothing made in China (Hessler, 2007), and the Puyuan Township produces over 500 million cashmere sweaters per year (Ruan and Zhang, 2009). Many of these clusters consist of privatized TVEs or their spin-offs.

The second issue is the restriction of labor mobility. China has been implemented a *Hukou* system which is a household registration system that officially identifies a person as a resident of a specific region, as well as the duties and the social welfare that the person may be obligated and entitled to. Under the *Hukou* system, individuals are classified as "rural" or "urban" residents and "local" or "non-local" residents. Whyte (2010) characterizes the *Hukou* system as "socialist serfdom". A peasant who seeks to move from a rural to an urban area and takes up a non-agricultural job used to require an approval. Meanwhile, "non-local" citizens are rendered unqualified for local social welfare, including housing, health care, and education benefits, etc. (Au and Henderson, 2006). Moving businesses to urban areas remains extremely difficult because of discriminatory policies imposed on rural residents, although the *Hukou* system has been relaxed over time such that peasant migrants are allowed to work in cities as lower-level residents.

The third restriction is the underdeveloped capital market in China (Allen et al., 2005). China's banking system is particularly biased against lending to private enterprises. Although the share of the private sector in the national GDP soared to 50% in 2009, the short-term bank loans issued to the private sector was only 4.9% of the national total (Guo et al., 2014). Consequently, the size and scope of entrepreneurial firms in China are, in general, constrained by difficulties in external financing.

To summarize, the aforementioned institutional restrictions make China's industrial clusters differ from the concept of "clustering" or "geographical agglomeration" studied in the existing literature. First, industrial clusters in China tend to be defined by administrative boundaries. As discussed, clustered TVEs before the mid-1990s and the subsequently privatized firms were the steppingstones for various entrepreneurial clusters today. They are concentrated within the administrative boundaries of certain local governments, which facilitate and protect the interests of private firms. Second, in association with the *Hukou* system, "rural" and "urban" are official terms standing for official recognition and describing social status rather than economic reality. Therefore, a large percentage of employees of the firms within clusters are officially defined as peasants, although they are manufacturing or service workers (in officially defined urban areas they are called peasant-workers). Third, firms in clusters are usually small and highly specialized given that the cluster-based production effectively decomposes the production process of a product into many small steps, which lowers both technical and capital barriers to entry (Ruan and Zhang, 2009; Xu and Zhang, 2009; Long and Zhang, 2011).

3. Related Literature: Industrial Clusters, Growth, and Inequality

The central idea in the economics of agglomeration is that firms can benefit from co-locating with each other. Both anecdotal and systematic evidence shows that in general clustering areas are more productive than other areas (Ciccone and Hall, 1996). There are different explanations for the sources of advantages of industrial clusters, focusing on regional specialization and urbanization. The well-known Marshall-Arrow-Romer (MAR) model which is formalized by Glaeser et al. (1992) based on the studies of Marshall (1890), Arrow (1962), and Romer (1986), emphasizes on regional specialization. This model claims that firms of similar industries clustered in a region can enjoy the advantages of knowledge spillovers from each other, lowered transportation costs of customer-supplier interactions and a sizeable common labor pool. On the other hand, Jacobs (1969) highlights the benefits of urban diversity. The theory of Jacobs externalities claims that the primary source of knowledge spillovers comes from exchanges and competition among firms in diverse industries co-located in metropolitan cities. Supporting regional specialization, Porter (1990), however, emphasizes the benefits from intensified competitions of firms specialized geographically, sharing with the spirit of Jacobs externalities.

While individual firms benefit from being clustered, the manner in which clustering or geographical agglomeration affects regional economic growth remains a contested issue. A perfectly competitive market for production resources results in diminishing returns to scale, and there will be economic convergence (Solow, 1956; Baumol, 1986; Barro and Sala-i-Martin, 1992). When firms cluster together, the diminishing returns of clustering is expected given the intense competition and increasing prices of production inputs. By contrast, endogenous growth theory explains growth through the Schumpeterian processes of creative destruction with a focus on entrepreneurship and innovation (Aghion and Howitt, 1992). Thus, the convergence or divergence across regions depends on how various regions adapt to new technologies or innovation. Aghion et al. (1999) emphasize that the role of organizational change in the production process (which specifies the way in which workers or organizations interact and learn from each other) may be crucial in determining productivity, and thus economic growth. A significant advantage of clustering is the strong cross-firm spillovers generated by sharing knowledge, innovation, and entrepreneurial culture (Kerr, 2010; Glaeser et. al., 2015). When both the convergence effects and the endogenous growth effects are present, the net effect of agglomeration on regional economic growth depends on the tradeoffs of these different forces.

Empirical studies on clustering and regional growth are mostly focused on identifying whether regional specialization or urbanization stimulates growth. So far, there is a little consensus reached. Several studies find that the growth of employment, wage and entrepreneurship is positively correlated to clusters composed of firms from diverse industries (e.g., Glaeser et al., 1992; Feldman and Audretsch, 1999), supporting the argument of Jacobs externalities. Others find that the clustering of firms in the same or related industries stimulates the growth of employment and wage, favoring the claims of MAR model (e.g., Porter, 2003; Delgado et al., 2014). At the same time, some studies find evidence for both the Jacobs and MAR externalities (Henderson et al., 1995; Rosenthal and Strange, 2003). Conversely, Cingano and Schivardi (2004), Dekle (2002) and Henderson (2003) fail to observe growth-promoting effects from agglomeration in any means.

Growth and inequality are among the most important social welfare issues concerned by economists. The relationship between growth and inequality has been a debated subject since Kuznets (1955). Kuznets finds that the relationship between the two in the US is an inverted U-shape between 1770 and 1970 (Kuznets, 1963). The interpretation is that in the transition from a rural to an industrial economy, income inequality should increase during the early stages of development (because of urbanization and industrialization) and decrease later (because industries would have already attracted a significant fraction of the rural labor force). However, recent studies find high income-inequality is associated with the deceleration of economic growth in developed countries since the 1970s (Persson and Tabellini, 1994; Perotti, 1996). At the same time, additional factors (such as human capital, social capital, capital mobility and institutions) may affect either growth (e.g., Barro and Sala-i-Martin, 1992; McCleary and Barro, 2006) or inequality (e.g., Heckman and Hotz, 1986; Krugman and Venables, 1995). Finally, a good (bad) institution may simultaneously protect (violate) private property rights, promote (prevent) development, and reduce (widen) inequality. Factors which may simultaneously affect growth and inequality are still not sufficiently investigated except some recent studies on institutions (e.g., Acemoglu et al., 2002, 2005; Easterly, 2007). Whereas there are extensive studies on the effect of clustering on economic growth, the general unanswered question is how specialization or urbanization affects inequality.

As discussed before, a large proportion of entrepreneurial industrial clusters in today's China can be traced to locations with a concentration of TVEs. Such regions are normally those with the most active entrepreneurial activities and most developed private sector. We, therefore, expect to observe industrial clusters may drive local economic growth. Moreover, the rapid development of entrepreneurial clusters in China has brought business opportunities and created jobs for those officially defined as rural residents. Therefore, we expect that entrepreneurial clusters in China may help to reduce officially defined urban-rural inequality within the region. Meanwhile, China is vast in size and regions vary significantly in institutions, economic structure, and the development level, among other aspects. As suggested by the new micro-geography literature, the potential functions of clusters may differ depending on the ecosystems of a region (Feldman, 2014; Catalini, 2018; Mudambi et al., 2018). We therefore expect to observe heterogeneity in the clustering effects.

4. Data and the Construction of DBI

4.1 Data Sources

First, our key explanatory variables, including the existence and features of clusters, are constructed based on data in the ASIFP, which covers all state- and non-state-owned industrial firms with annual sales of 5 million RMB or above, including firm-level data on industry, location, age, ownership, and financial information. The enterprises included in this database account for 90% of the total sales of all industrial firms in China.[3]

Second, data on county-level per capita GDP, per capita household income, and other general county-level economic and demographic variables (e.g., total GDP, rural and urban populations, and investment in fixed assets), are all from the China Socio-Economic Development Statistical Database.[4] According to the description of the database, the income data of rural and urban households are collected based on the surveys of randomly selected local residents, who have been residing in a place for more than 6 months, regardless of being migrant workers or permanent residents.

[3] ASIFP excludes non-state-owned firms with annual sales under 5 million RMB. Thus, the potential bias of our findings should be underestimating the impacts of entrepreneurial clusters. As a robustness check, we apply the same DBI methodology to identify clusters using the Chinese Economic Census data in 2004, which includes industrial firms of all sizes. The identified clustering patterns with the census data are qualitatively consistent with those identified based on the ASIFP.

[4] Per capita household income statistics include rural household per capita net income and urban household per capita disposable income in the said database. Rural household net income is defined as the total family income excluding the family business expenses, depreciation of productive fixed assets, taxes, and land contract fees. Urban household disposable income is the total family income minus personal income tax and expenditures on social security.

Third, we control a series of county-specific variables. Specifically, we include *fraction of industrial outputs* to total GDP for capturing the local economic structure; *fraction of non-state owned firms* and *fraction of micro firms*[5] for controlling the effect of privatization and small businesses; *fraction of education expenditure* to GDP for capturing the human capital development; *fraction of investment in fixed assets* to GDP (including investment in infrastructure, renovation, and real estate among others) for controlling the local investment in physical capital; and *fraction of government expenditure* to GDP to control for government administrative expenditure. All the above ratios are in log form. Data on local fiscal expenditures come from the National Prefecture and County Public Finance Statistical Yearbooks for the same period. Furthermore, we construct a panel of officially designated "National Poor Counties".[6] Each year, officially designated poor counties received sizeable amounts of fiscal transfers from the central government. This subsidy may affect the local income and the urban-rural inequality that we aim to investigate. The list of the poor counties is obtained from the official website of the State Council. Finally, to differentiate the effect of clusters from that of Special Economic Zones (SEZs), we construct a panel that indicates the existence and number of provincial-level SEZs in each county from 1998 to 2007. The list of SEZs is obtained from the website of the Ministry of Commerce of the People's Republic of China.

Forth, we control the inflow of the migrant workers, measured by the percentage of inflow migrant rural labor over the total employment in each province between 1998 and 2007. The data for this variable is extracted from Fan et al. (2011). It is known that industrial clusters employ large numbers of migrants from other regions. Such inflow of migrant workers may have affected the economic growth, urban-rural income inequality and the development of industrial clusters simultaneously. There is no county-level panel data for migration available. We thus use the provincial level panel data for the inflow of migrant workers as a proxy to control such effects. We fully understand that such measurement of migration may not be the best for our estimations and better measures of migration should be incorporated in future research.

All the above data are deflated to 1998 price level when applicable. Some counties changed their names or judiciary boundaries during our examination period. We identify the changes and convert the corresponding county codes into a benchmark system. China also modified its industry coding system in 2002 (from GB/T 4754-1994 to GB/T 4754-2002). The four-digit industry codes that have become either more disaggregated or more aggregated after 2002 were tracked and the aggregated codes are used to group industries from 1998 to 2007.

All variables used in this study are defined in Table A-1of online appendix.

4.2 Measuring Industrial Clusters in China

Constructing clustering indices to capture entrepreneurial clusters in China is a major challenge. Clustering indices constructed in existing studies focus either on regional specialization or inter-connectedness of local industries (Porter, 1990; Krugman, 1991a; Glaeser et al., 1992). Most studies on regional specialization apply

[5] The fractions of non-state-owned and micro firms are derived from firm-level data from the ASIFP. For instance, for any county during the sample period, we calculate the total number of non-state-owned firms or micro firms and divide it by the total number of firms in the county in that year to obtain the fraction.

[6] Two rounds of the poverty reduction program (known as the 8-7 Plan) were conducted in China during 1986-1993 and 1994-2000, aiming to promote local economic development through targeted public investments with fiscal transfer. In 1986 and 1994, the Poverty Reduction and Development Team supervised by the State Council (*Guo-wu-yuan Fu-pin Kai-fa Ling-dao Xiao-zu*) published two National Poor County lists. The 1994 list was modified further in 2006 and 2012. As of 2012, there were 592 national level poor counties in the list.

the Herfindahl–Hirschman index (HHI), Gini coefficient (Gini), location quotient (LQ), or Krugman index to measure clustering.[7] Based on the revealed comparative advantage in product export, Hausmann and Klinger (2007) construct a proximity measure for all four-digit Standard International Trade Classification products. Long and Zhang (2011, 2012) employ this proximity index to measure clusters in China.

Employing regional specialization or inter-connectedness measurements directly to study entrepreneurial clustering in China may not be the most suitable method for our purpose. This is because restrictions on factor mobility and firm location decisions may create biases and potential measurement errors. At the onset of the economic reform, all firms in China were owned or controlled by national or local governments and their locations were chosen as administrative decisions. As such, the concentration of heavy industries in certain areas of China was driven mostly by political concerns, including those related to national security.

Today, in commanding heights sectors such as energy, mining, railway, airlines, and communication, state ownership still dominates. These firms usually are large so that regions with such giant SOEs will be recognized as specialized, with high specialization scores measured by HHI, Gini, or LQ. For instance, oil refining and processing SOEs contributed 24.5% to the local industrial output in Daqing City in 2007. Changchun City is highly specialized in manufacturing transportation equipment, which contributed to 68.26% of the industrial output there in 2007, with 79.62% of the outputs coming from 13 gigantic SOEs. However, the co-location decisions of these firms have little to do with markets, and this kind of regional specialization is not of interest for this study.

As discussed in Section 2, the entrepreneurial clusters in China are characterized by the emergence of numerous "specialty towns," each consists of a large number of small firms and family workshops producing a particular type of products. Outputs of these clusters comprise significant shares of national or global markets. These observations suggest that in addition to specialization, the density of firms in an industry within a locality is one of the most important features of entrepreneurial clustering in China. Hence, we propose a density-based index (DBI) to measure entrepreneurial clusters in China. For the DBI at the county level, we denote the number of firms in any 2-digit industry $j \in \{1, 2, \ldots J\}$ in county $i \in \{1,2, \ldots I\}$ at time t as $fn_{j,i,t}$. We define county i as "a county with an α cluster of industry j" if the number of firms in this county is among the top α percentile of all counties in this industry at time t. Formally, we define

$$c_{j,i,t} = \begin{cases} 1, \ if\ fn_{j,i,t} \geq (100-\alpha)\ percentile\ of\ \{fn_{j,1,t}, fn_{j,2,t}, \ldots, fn_{j,I,t}\} \\ 0, \qquad otherwise \end{cases};$$

And

$$C_{i,t}=\sum_{j \in C_{it}} c_{j,i,t}.$$

In this paper, we focus on the top five percentile county-level clusters. Thus, for

[7] For instance, Glaeser et al. (1992) focus on the contribution of a region's top five largest industries to the local economy to reveal the extent to which a given region is specialized or diversified, regardless of how the economic structure of the country as a whole evolves. 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 (1991a) 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.

the remainder of this paper, the term cluster means α = 5, and we will omit to mention this unless a definition is specified.[8] We use the following dummy variable to capture the existence of the DBI cluster in any county *i* of any industry in year *t*.

$$Cluster_{it} = \begin{cases} 1, & if\ C_{i,t} \neq \emptyset \\ 0, & otherwise \end{cases}$$

Identifying the strength of clusters is essential because counting the number of firms treats all the firms as the same, ignoring their differences. We define the output[9] strength of any cluster of industry *j* located in county *i* at time *t* by the ratio of its output contribution to the total industrial output over the national county average level. Specifically, the output strength of each cluster *ji* at time *t* is defined as $OutputStrength_{jit} = \frac{OutputShare_{jit}}{\frac{1}{I}\sum_{i=1}^{I} OutputShare_{jit}}$, where $OutputShare_{jit} = \frac{Output_{jit}}{Output_{jt}}$ is the output share of the cluster *ji* in the national total output of industry *j* in time *t*, and $\frac{1}{I}\sum_{i=1}^{I} Output\_Share_{jit}$ is the average number of all counties for industry *j* in time *t*. $OutputStrength_{jit} > 1$, if output strength of cluster *ji* at time *t* is larger than the national average; otherwise, $OutputStrength_{jit} \leq 1$. Similarly, we define the establishment strength of each cluster ji at time *t* as $EstablishmentStrength_{jit} = \frac{EstablishmentShare_{jit}}{\frac{1}{I}\sum_{i=1}^{I} EstablishmentShare_{jit}}$, where $EstablishmentShare_{jit} = \frac{Establishment_{jit}}{Establishment_{jt}}$ is cluster ji's share in the national total establishments of industry j, and $\frac{1}{I}\sum_{i=1}^{I} EstablishmentShare_{jit}$ is the national average establishment share of industry j in each county.

To measure the cross-industry aggregate strength of clustering in each county *i* based on the strength ratios defined above, we construct the overall strengths of clusters in the following. The overall output strength of the clusters in county *i* is defined as the weighted average of the strength of each cluster (if any):

$$OutputStrength_{it} = \begin{cases} \frac{\sum_{j \in c_{jit}} Output_{jit} OutputStrength_{jit}}{\sum_{j \in c_{jit}} Output_{jit}}, & if\ C_{it} \neq \varnothing \\ 0, & if\ C_{it} = \varnothing \end{cases}.$$

Similarly, the overall establishment strength of the clusters in county *i* is defined as

$$EstablishmentStrength_{it} = \begin{cases} \frac{\sum_{j \in c_{jit}} Establishment_{jit} EstablishmentStrength_{jit}}{\sum_{j \in c_{jit}} Establishment_{jit}}, & if\ C_{it} \neq \varnothing \\ 0, & if\ C_{it} = \varnothing \end{cases}.$$

To capture the development of entrepreneurial clusters, we calculated two indices based on the ownership structure of clusters in county *i* at time *t*. First, we measure the share of non-state firms'[10] outputs in the total outputs of clusters in county *i* at time *t* as the following,

[8] Our results stay robust when we assign other values for α, such as 3 and 8.
[9] The output and establishment data used in calculating cluster strength indices are based on the firm-level data from the ASIFP. For our county level cluster measurement, the output of a given county in a given year is the aggregated output of all firms located in the county in that year.
[10] Non-state-owned firms refer to firms where state capital constitutes less than 50% of the total paid-in capital.

$$OutputNonstate_{it} = \begin{cases} \frac{\sum_{x \in X_{it}} \sum_{j \in c_{jit}} Output_{xjit}}{\sum_{j \in c_{jit}} Output_{jit}}, & if\ C_{it} \neq \emptyset \\ 0, & if\ C_{it} = \emptyset \end{cases},$$

where $\boldsymbol{X_{it}}$ is the set of non-state firms in county *i* at time *t*. This index captures the development of non-state firms within the clusters. Similarly, we calculate $EstablishmentNonstate_{it}$ as the share of non-state firms in the total establishments of the clusters in county *i* at time *t*:

$$EstablishmentNonstate_{it} = \begin{cases} \frac{\sum_{x \in X_{it}} \sum_{j \in c_{jit}} Establishment_{xjit}}{\sum_{j \in c_{jit}} Establishment_{jit}}, & if\ C_{it} \neq \emptyset \\ 0, & if\ C_{it} = \emptyset \end{cases}.$$

The summary statistics of DBI clusters is shown in Table A-2 in online appendix. Taking the year 2007 as an example, among 2,734 counties, 739 (about 27%) have clusters. These clusters contribute to 38% of the total national outputs and 37% of the total national employment. On average, the output strength of a cluster is about 6.5 times that of an average local industry within a county. Also, the number of establishments in a cluster is about 5.9 times that of the national average number. On average, about 80% of the clusters' outputs are from non-state firms, and more than 83% of firms in the clusters are non-state owned. This table also shows the dynamics of cluster development in Chinese counties. In our sample period (1998–2007), 294 counties (about 10% of all the counties) always have some clusters in at least one industry. Conversely, 1,576 counties (about 56% of all the counties) have never developed any industrial cluster. A total of 317 counties initially did not have clusters in 1998, but they developed industrial clusters by 2007. By contrast, 292 counties had clusters operating in 1998, but clusters in these counties disappeared by 2007. Most of these backward developments occurred in inland areas.

The comparisons between provincial level DBI cluster indices and some standard cluster indices used in the existing literature are presented in Table A-3 of online appendix. All measurement results are based on the 2007 ASIFP data. Measured by DBI cluster count, the top five provinces are Zhejiang, Jiangsu, Guangdong, Shandong, and Shanghai. Measured by DBI output strength, the top five provinces (in descending order) are Shanghai, Tianjin, Zhejiang, Shandong, and Jiangsu. Finally, when measured by the DBI establishment strength, the top five provinces are Zhejiang, Shanghai, Jiangsu, Shandong, and Tianjin. The strongest five provincial regions in DBI non-state cluster indices are Zhejiang, Shanghai, Jiangsu, Shandong, and Tianjin. This is consistent with the general perceptions of spatial distributions of entrepreneurial clusters in China.

In contrast, applying standard measurements to the Chinese data tends to capture agglomerations of highly specialized large SOEs, many of which are located in interior regions. For example, measured by HHI, the top 5 provincial regions will be Xinjiang, Shanxi, Hainan, Jilin, Gansu; or Xinjiang, Qinghai, Gansu, Yunnan, Hainan if measured by Gini; or Tibet, Ningxia, Xinjiang, Qinghai, Yunnan if measured by LQ; and Shanxi, Tibet, Xinjiang, Qinghai, Yunnan if measured by Krugman Index. Except for Tibet (which is one of the most under-developed regions in China), all these provinces are known for a concentration of SOEs and their weakness regarding entrepreneurial activities. Some of the provincial level cluster measurements shown in Long and Zhang (2011), which apply the Industrial Proximity measurement (Hausmann and Klinger, 2007), are close to ours; but some other results reflect noises. The top five provinces by their measurement are Tibet, Beijing, Jilin, Zhejiang, and Ningxia. We further compare the average weights of SOEs in the regions defined by different clustering measurements as shown in Table A-4 (online appendix). The comparison confirms that standard clustering measurements tend to capture the specialization or concentration of SOEs under the centrally planned economy. Figure A-1.a and Figure A-1.b (online appendix) illustrate how standard approaches and DBI capture China's clusters geographically by showing maps of regional clustering levels. The map shows that the DBI clusters are concentrated heavily along coastal line regions, which is highly consistent with a satellite night vision of China (Figure A-2 of online appendix). By contrast, coastal line regions are not captured adequately by standard clustering measurements.

Table 1(a) presents the summary statistics of the dependent variables for counties with and without DBI clusters. Compared with counties without clusters, on average, counties with clusters grow faster and have lower urban-rural income inequality. Between 1998 and 2007, the average growth rate of the counties with clusters is 1.3% higher than other counties, whereas the urban-rural per capita income ratio in counties with clusters is lower by about 20% than other counties. Table 1(b) provides the summary statistics of other characteristics of counties with and without DBI clusters. On the one hand, counties with clusters have higher per capita and total GDP on average, are more industrialized and have more private firms than other counties. On the other hand, fractions of government and education expenditures in counties with clusters are about half of those in other counties.

5. Clustering, Regional Economic Growth, and Urban-Rural Inequality

In this section, we present the estimations for the effects of clustering on regional growth and urban-rural inequality within a region, with identification issues addressed.

5.1 Clustering and Regional Growth

Our hypothesis is that clusters should be positively and significantly associated with economic growth. Moreover, regions should grow faster if their clusters are stronger, or if their clusters are entrepreneurial ones. We test this hypothesis by estimating a type of Barro growth model (Barro, 2000) as follows:

$$\ln\left(\frac{GDPpercapita_{it+1}}{GDPpercapita_{it}}\right) = \alpha + \beta Clustering_{it} + \gamma \ln(GDPpercapita_{it}) + \mu[\ln(GDPpercapita_{it})]^2 + \tau \ln(CPI_{pt}) + \boldsymbol{\delta Z_{it}} + \varepsilon_{it} \quad (1),$$

where $GDPpercapita_{it}$ is the per capita GDP of county i at year t, and the dependent variable represents a county level annual growth rate of per capita GDP. Our major explanatory variables are $Cluster_{it}$, which is a dummy variable that equals to one if at least one cluster operates in county i in year t and zero otherwise; the strength of clusters ($OutputStrength_{it}$, and $EstablishmentStrength_{it}$); and the ownership structure of clusters ($OutputNonstate_{it}$ and $EstablishmentNonstate_{it}$). The initial level of per capita GDP controls for any convergence or divergence effect, and the square term of per capita GDP controls the speed of convergence or divergence. Provincial level consumer price index (CPI) is included to control for the inflation effect. $\boldsymbol{Z_{it}}$ is a vector of other control variables as outlined in Section 4.1. County and year fixed effects are controlled in the panel analysis to address county- and time-specific effects. Meanwhile, as each county's observations in the panel data are auto-correlated, following Peterson (2009), we group the standard errors within counties.

The major estimation results are presented in Table 2. Panel A reports the baseline estimations. Column (1) of Panel A indicates that the coefficient of *Cluster* is positive and significant, indicating that the existence of industrial clusters has a positive impact on local economic growth. Moreover, Columns (2) and (3) indicate that the strengths of clusters (*OutputStrength* and *EstablishmentStrength*) are also positively and significantly associated with growth. A 1% increase in the clusters' output strength will result in about a 1% increase of per capita GDP growth. Finally, Columns (4) and (5) demonstrate that entrepreneurial clusters, i.e., clusters dominated by non-state firms (measured by *OutputNonstate* or *EstablishmentNonstate*) are positively and significantly associated with economic growth. A 1% increase in the contribution of the non-state sector (either in output or establishment) will result in approximately 1.6% increase in the per capita GDP growth.

For all regressions, both initial levels of per capita GDP and the squared term of per capita GDP are significantly and negatively associated with growth, suggesting a mean convergence in the economic growth among Chinese regions, and the convergence accelerates over time. Regarding the structure of the local economies, estimated coefficients of *fraction of industrial output* and *fraction of non-state firms* are all significant and positive, thereby implying that counties with higher levels of industrialization and more non-state firms experience higher economic growth rates. Furthermore, expenditures for education and fixed investments are significantly and positively correlated to local economic growth, indicating that regions that invest more in human and physical capitals experience higher economic growth. Local government expenditure is positively correlated to growth, which denotes a positive correlation between the size of the government and growth rate. Lastly, the presence of SEZ (special economic zone) within a county is positively correlated with growth.

Our baseline regressions demonstrate strong statistical associations between industrial clusters, particularly strong and entrepreneurial clusters, and economic growth. However, the causalities for the relationships are yet to be established because the clustering might be a result rather than a cause of economic growth. Moreover, the existence of clusters or the features of clustering might coincide with other unobservable variables that might influence local economic growth. To address identification concerns, we employ two empirical strategies: the two-stage estimations and Granger test of causalities.

First, we employ the two-stage least squares estimation procedure with an IV to identify the clustering effects on growth. The IV, per capita mining output in each province, is obtained from the China Mining Yearbook (2001–2007). We believe per capita mining output is relevant to the existence of clusters and the strength and the ownership structure of clusters, because we expect provinces with higher per capita mining outputs to have weaker industrial clusters and a smaller share of the private sector in the clusters. This anticipation is based on an observation that mine-rich regions are often dominated by large companies due to large fixed investments and scale economies in the mining industry. Thus, smaller businesses are often crowded out, and entrepreneurship is often depressed as argued by Chinitz (1961)[11] and as supported by empirical evidence (e.g., Rosenthal and Strange, 2003; and Glaeser et al., 2010, 2015).

[11] Chinitz (1961) also argues that when a region is dominated by large mining companies, the culture of entrepreneurship is weak because the executives of large mining companies are less likely to transfer entrepreneurial knowledge to the next generations. Moreover, in such regions, the financial and labor constraints for entrepreneurial firms may be severe, because both financial institutions and labor may easily access large firms with low levels of risks and uncertainty. Furthermore, large companies are more likely to internalize supplies or source them outside the region to enjoy lower costs, which consequently depresses the local supply development of small entrepreneurial firms.

Moreover, given the state-ownership of mining rights in China, mine-rich regions should have higher shares of the state sector, which in general would affect all industries in those regions. We also believe that the per capita mining output of a province is exogenous because the mine-richness of a region is geologically determined, and thus by itself is exogenous to regional economic growth. Given our panel data is at the county level, whereas this particular IV is at the provincial level, it should not be directly correlated to county-level economic growth.[12]

Panel B of Table 2 presents the regression results of the two-stage estimations for economic growth when clustering variables are instrumented. Columns (1) to (5) report the first-stage estimations, showing that the per capita mining output is, as expected, significantly and negatively associated with the existence, strength, and non-state ownership of clusters within a county. Columns (6) to (10) show the second-stage estimation results. Consistent with our baseline OLS regression results, the instrumented clustering variables remain significantly and positively associated with economic growth. These outcomes confirm that the presence of industrial clusters, particularly strong clusters and entrepreneurial clusters drive local economic growth.

Besides the two-stage estimations, we further conduct a Granger causality test (Granger, 1969) to identify the dynamic causal relationships between industrial clustering and economic growth. Granger causality test initially focuses on examining the predicting power of one time series for another through T-tests and F-tests on lagged values of relevant variables. As we are dealing with panel data, we follow the Dumitrescu and Hurlin (2012) procedure for panel data analysis. Ideally, we should find out the lag value to make the average value of the Akaike, the Bayesian, or the Hannan-Quinn information criteria minimized. However, the largest number of lags is constrained by the length of a panel (T) where: T>5+3×lag. Given that we have a ten-year panel, the only lag we are able to choose is one year. The results of the Granger causality tests are presented in Panel A of Table A-5 of online appendix. As shown in the table, the null hypothesis that industrial clustering (including all variables that measure the existence, strength and non-state ownership of clustering) does not lead to economic growth is rejected. In other words, industrial clustering does have predictive power for the economic growth of counties between 1998 and 2007.

5.2 Clustering and Urban-Rural Income Inequality within a County

In this subsection, we examine the relationship between clustering and urban-rural income inequality and explore how clustering affects household income of urban and rural residents within a county. We use urban-rural household per capita income ratio as a proxy for urban-rural inequality. In our sample, the county-level urban-rural income ratio increased from 2.08 to 2.69 (on average) from 1998 to 2007. Our hypothesis is that industrial clustering, measured by different DBIs, should be negatively and significantly associated with urban-rural income ratio. Our baseline regression model of urban-rural inequality vs. clustering is the following equation:

$$\ln\left(Ratio_{it}\right) = \alpha + \beta Clustering_{it} + \boldsymbol{\delta W_{it}} + \varepsilon_{it} \quad (2),$$

where $Ratio_{it}$ refers to the urban-rural household income inequality measured by the ratio of urban over the rural household per capita income in county i in year t. The clustering variables and major control variables $\boldsymbol{W_{it}}$ are the same as those in Equation

[12] Indeed, even in the case of dealing with the relationship between provincial level mining endowment and provincial economic growth, evidence provided in the literature (e.g. Glaeser et al., 2015) still supports our conjecture.

(1) except for the inclusion of the total GDP of the county and dropping the square term of the per capita GDP and CPI.

The estimation results are reported in Table 3. Panel A presents the baseline estimations. Columns (1), (2), and (3) of the table indicate that by pooling all clusters together without differentiating the ownership structure of clusters, the cluster is negatively but insignificantly associated with local urban-rural inequality. The correlation becomes statistically significant only when we focus on entrepreneurial clusters which are dominated by non-state firms, as shown in Columns (4) and (5). These findings indicate that an increase of 1% in the number and outputs of non-state firms within clusters is associated with a 3% reduction in the urban-rural inequality in that county. Furthermore, in all five columns, investments in both fixed assets and education and the presence of SEZs are positively and significantly correlated to urban-rural inequality, while the inflow of migrants within the province is negatively and significantly associated with urban-rural inequality.

In order to establish the causalities between entrepreneurial clustering and urban-rural income inequality, we conduct two-stage estimations and the Granger causality tests. We use two IVs for the two-stage estimations. The first IV we use is the per capita number of Christian churches of a county in a given year. Christianity culture plays an important role in fostering trade and entrepreneurship, which is essential for cluster formation in modern China. Christian missionaries brought the ideas of modern commerce and trade to China a century ago. Meanwhile, religious activities organized by churches serve as a mechanism for the local people to communicate and coordinate with each other. Although religious activities were banned during 1950 and 1980, they were revived since the post-Mao reform. Arguably, the formation of industrial clusters is rooted in the entrepreneurship culture and coordination capacity of the local people. We, therefore, expect that the local density of churches is positively related to our clustering measurements. On the other hand, religions in China do not play any role in wealth distribution as the scale of religious activities is tightly controlled by the government. We hence do not expect Christian activities to be directly related to the urban-rural income inequality in a county, unless through the related economic activities.

The second IV is the per capita length of classified highways[13] in a city. The so-called city here is an administrative level within the government hierarchy. Typically a city government controls a dozen counties, but itself is under the control of a provincial government. Access to the transportation network is expected to be related to the development of clusters because of several reasons. First, transportation infrastructure reduces the trade and shipping costs for the firms within the region that is ultimately important for clustering of firms. Second, with reduced transportation costs, the scale and scope of the market may be increased that more firms within the region may benefit. Third, with better transportation infrastructure, technology and knowledge may be transferred more easily that helps firms' specialization and development within the region. We, therefore, expect that the density of classified highways to be correlated to the probability of a county to have industrial clusters. However, the construction of classified highways in China is a decision made at the national and provincial governments, and intra-county inequality issue cannot be their major concern. County governments do not have a voice in such decision. Hence, the per capita length of classified highways at city level should not have a direct relationship with the urban-

[13] Classified highways here refer to national and provincial level highways. These highways are planned, financed and constructed by national and provincial level governments.

rural income inequality at the county level, unless through the economic activities within the county.

The two-stage estimations for the clustering effects on urban-rural income inequality with the two IVs are presented in Panel B of Table 3. Our focus is the effects of entrepreneurial clusters while ignoring the effects of other measurements of clusters, such as *OutputStrength* and *EstablishmentStrength,* as they are statistically insignificant in baseline OLS regressions (Table 3 Panel A). Columns (1) to (2) report the first-stage estimation results while Columns (3) to (4) report the second-stage estimation results, respectively. As shown in the table, both per capita number of Christian churches and per capita length of classified highways in a city are significantly and positively associated with entrepreneurial cluster measurements. Thus, these two IVs are relevant. Moreover, the Sargan tests indicate that the IVs are jointly exogenous. Finally, the second-stage estimations show a negative and significant relationship between the development of entrepreneurial clusters and urban-rural income inequality, confirming the causality between the two.

Besides, we conduct Granger tests to further identify the clustering effects on urban-rural income inequality. The results are reported in Panel B of Table A-5 of online appendix. As shown in the table, the null hypothesis that there is no causal relationship between cluster measurements and urban-rural income inequality within the county is rejected, confirming the causal relationship between the entrepreneurial clustering and reduced urban-rural inequality within the county.

A positive association between growth-enhancing efforts and the widening of regional inequality may be unsurprising. However, the negative association between entrepreneurial clusters and urban-rural income inequality might be counter-intuitive, as it is a popular view that a market economy dominated by privately owned firms tends to worsen inequality and one of the benefits of maintaining state-owned firms is to contain inequality. To understand the mechanism of how the development of non-state firms in clusters might reduce inequality, we subsequently study impacts of clusters on rural and urban income separately.

Noticing that today's entrepreneurial clusters are mostly located in areas with active TVEs in the 1990s, it is likely that entrepreneurial clusters create more business opportunities for rural residents (Long and Zhang, 2012) and employ more rural labors that contributes to the rural residents' income and reduces urban-rural inequality. The 2007 Chinese Household Income Project (CHIP) database, which covers 92 rural counties in China, reveals that in counties with clusters, individuals' average non-agricultural income is about 8% (1,398RMB vs. 1,293RMB in Table A-6.a of online appendix) higher than that in counties without clusters, and the ratio of individuals engaged in non-agricultural activities in counties with clusters is about 20% (42.76% vs. 35.58%) higher than that in counties without clusters. Additionally, using 2004 and 2008 economic census data which include enterprises of all sizes to calculate the entry of new businesses in counties with and without clusters, we find that clusters create significantly more new business opportunities for local people. As shown in Table A-6.b (online appendix), on average, in counties with clusters, the number of new businesses established between 2004 and 2008 is 821.55 comparing to 186.74 in counties without clusters. Moreover, the survival rate of businesses in counties with clusters is much higher than that in counties without clusters in the same period. The results presented in Table A-6.a and Table A-6.b to some extent suggest that clustering lowers the entry barriers for starting up new businesses for rural residents and creates more non-farming jobs for the local people. Therefore, we hypothesize that clustering lifts rural residents' income, which reduces the urban-rural income inequality within a region.

Our baseline regression models for testing this hypothesis are as follows:

$$\ln\left(Rural\ household\ income_{it}\right) = \alpha + \beta Entrepreneurial\ Cluster_{it} + \boldsymbol{\delta W_{it}} + \varepsilon_{it} \quad (3)$$

$$\ln\left(Urban\ household\ income_{it}\right) = \alpha + \beta Entrepreneurial\ Cluster_{it} + \boldsymbol{\delta W_{it}} + \varepsilon_{it} \quad (4),$$

where $Rural\ or\ Urban\ household\ income_{it}$ are the rural or urban household per capita income in each county during our sample period. $OutputNonstate_{it}$ and $EstablishmentNonstate_{it}$ are the entrepreneurial cluster measurements and $\boldsymbol{W_{it}}$ are the same control variables as those in Equation (2).

The regression results are summarized in Panel A of Table 4. Columns (1) and (2) show that, in the OLS model, the development of non-state firms in clusters measured by both output and establishment is significantly and positively associated with the rural household per capita income, suggesting that local rural residents' income is positively associated with entrepreneurial clusters. Everything else being equal, a 1% increase in the non-state firms' output or establishment number within clusters is associated with 3% increase of the rural household per capita income. By contrast, Columns (3) and (4) indicate that the development of non-state sectors in the clusters is not statistically related to urban household income.

Table 4 also shows that the share of the private sector, the total GDP, the investments in fixed assets and education, and provincial inflow of migrants are all significantly and positively correlated to rural household income. Interestingly, government policies on special economic zones are negatively correlated with rural household income and poverty eradication does not demonstrate any statistical significance on rural household income. Furthermore, the expenditures on administration seem to have a negative effect on rural household income but positively affect its urban counterpart.

Finally, we conduct two-stage estimations for the clustering effects on per capita income of rural households. The IVs and the specifications are the same as those in the two-stage estimations for the clustering effects on inequality. The estimation results are reported in Panel B of Table 4. Columns (1) and (2) present the first-stage estimation results, confirming the relevance of the two IVs in general. Second-stage estimations from Columns (3) and (4) confirm that the development of non-state firms in the clusters (measured by *OutputNonstate* and *EstablishmentNonstate*) is significantly and positively correlated with rural household per capita income after these independent variables are instrumented. The estimations verify that clusters with more non-state firms lead to a higher income of rural residents, which confirms that entrepreneurial clusters reduce urban-rural income inequality (Table 3) by increasing the income of rural residents.

5.3 Additional Robustness Checks

Besides the identification estimations, we conduct additional robustness checks to test the reliability of the clustering effects. Above all, the economic activities in China have been concentrated in the coastal regions, whereas there are fast-growing megacities such as Beijing, Shanghai, Guangzhou, and Shenzhen. In our baseline and two-stage estimations, we have controlled county fixed effects. However, if megacities' impacts on economic activities, such as creating a huge inflow of FDI or other public policies, are so large that they overwhelm the county fixed effects, the estimation results we obtained from the baseline estimations might be biased. In addressing such concerns, we control the megacity effects (measured by the megacity population) for the estimations. As shown in Table A-7 (online appendix), the effects of clustering variables stay robust after controlling for megacity population.

Additionally, industrial profile varies considerably across locations and might affect local growth and urban-rural inequality as well as the development of clusters substantially. Such concerns are relevant in our case as we did not consider regional specialization in certain industries. To address such concerns, we control the local industry profile by identifying the three largest industries in each county in a given year. As shown in Tables A-8a and A-8b of the online appendix, after controlling these largest industries in the county, the clustering effects on growth and income inequality stay robust.

To summarize, by employing different identification strategies and exercising robustness checks, our estimations confirm the positive effects of industrial clustering on regional economic growth as well as the effect of entrepreneurial clustering on the reduction of intra-region urban-rural income inequality in China.

5.4 Clustering Effects under Different Urbanization Levels

As all Chinese large cities are national or regional political centers, the driving forces of these cities' expansions are more of political powers, including SOEs, than markets (Bai and Jia, 2018). City boundaries (sizes) and entitlement of city residents are assigned and defined administratively rather than by business activities. Strong distortions and restrictions (e.g., *Hukou* in these cities) associated with political powers might overwhelm the benefits of agglomeration. Furthermore, in rural areas where agricultural activities were the major household activities, clusters may evolve as the most prominent industrial force of a county that have significant impacts on the society from various aspects including the economic growth and urban-rural income inequality. On the contrary, in highly urbanized regions and especially megacities, manufacturing activities are much more diverse and the service sector has increasing impacts on the local economy and society. The Jacobs-type effects of urbanization are likely to overwhelm the clustering effects under complicated landscapes. Combining both the factors discussed above, we expect that clusters in highly urbanized areas or megacities have less significant effects than those in less urbanized areas or non-megacities.

In order to examine clustering effects in regions with different urbanization levels, we first compare regions at the highest urbanization level, i.e. megacities, with the rest. We divide our county samples into two groups: those located in megacities, and those not located in megacities. Our estimations of the baseline regression model for the two subsamples (Table 5) show that in non-megacity counties, the impacts of clusters on growth and urban-rural inequality (Table 5a) are qualitatively the same as our full sample estimations (Tables 2 Panel A and Table 3 Panel A): the existence of clusters, entrepreneurial clusters and strong clusters are significantly and positively associated with the county's growth; and the existence of entrepreneurial clusters is significantly and negatively associated with the urban-rural income inequality within the county. However, for megacity counties, clusters have no significant relationship with growth or with urban-rural income inequality (Table 5b).

Then we compare relatively more urbanized regions with less urbanized regions. We divide all the counties in our sample into two groups based on the ratio of urban population to the total population, using the median value as the cutoff threshold. We then conduct our baseline regressions on economic growth and urban-rural income inequality for the two subsamples separately. Our regression results (Table 6) show that in the sub-sample of less urbanized counties, stronger clusters and entrepreneurial clusters have significant and positive effects on growth; and entrepreneurial clusters are correlated with significantly lower urban-rural inequality (Table 6a). However, clusters located in more urbanized regions are insignificantly related to growth; while entrepreneurial clustering in these regions is still significantly associated with a reduction of urban-rural income inequality (though the effect is only significant for *OutputNonstate*) (Table 6b).

The results presented in Tables 5 and 6 confirm our conjecture that the clustering effects on growth and intra county urban-rural inequality vary depending on the urbanization level of a region. In counties located in more urbanized regions, the clustering effects are weaker.

6. Conclusion

In this study, we develop an industrial cluster measurement, density-based index (DBI), which captures institutionally constrained industrial clusters, particularly entrepreneurial clusters in China. Combining both firm- and county-level data, we create a county-level DBI cluster panel, and find that industrial clustering enhances regional economic growth. Moreover, entrepreneurial clusters reduce urban-rural intra-region income inequality by increasing rural residents' income, which is qualitatively different from the impacts of clusters of SOEs on income inequality. We also find that the clustering effects on growth and reduction of intra-region inequality are insignificant in highly urbanized regions, and particular in megacities. To our knowledge, this is the first study of this kind in the literature.

We carefully address identification concerns through the two-stage least squares approach and Granger causality test. For investigating the effects of clustering on economic growth, we use local per capita mining outputs as an IV. Concerning urban-rural inequality and the income of rural residents, we use the per capita length of classified highways in a city and the number of Christian churches in a county as IVs.

This study contributes to the literature on development economics, growth, inequality, and economic geography. Several challenging questions arising from our discoveries require further research. One of our major findings is that entrepreneurial industrial clusters enhance economic growth and reduce intra-region urban-rural inequality effectively. However, such clusters are only concentrated in some regions and certain industries, and entrepreneurial clusters are non-existent in many provinces. Why is this so? What barriers prevent the creation of entrepreneurial clusters? Evidence on the simultaneous existence of the strong growth-enhancing effect and inequality-reducing effect of entrepreneurial industrial clusters indicates a possibility that Schumpeterian growth mechanism (e.g., Aghion, 2002) is at work. Determining the mechanisms and addressing why particular mechanisms are more prevalent than others in certain regions require much more research. These future research directions promise further contribution to the literature on economic development and institutions (e.g., Acemoglu et al., 2002, 2005; Easterly, 2007), and knowledge on growth and inequality, economic geography, urban economics, and the economic development of China.

Table 1(a): Summary Statistics of County Growth, Income, and Inequality

Variables	GDP per capita growth	Urban household per capita income	Rural household per capita income	Urban–rural income ratio
Counties with Clusters				
Mean	0.135	9855.471	3999.749	2.244
Median	0.124	8803.	3672.771	2.127
Standard deviation	0.150	4098.133	1874.159	0.583
Min	−0.273	1585.087	643.471	1.097
Max	0.9137	31597	14845	6.910
N	4810	1353	4841	1274
Counties without Clusters				
Mean	0.122	6697.413	2628.195	2.741
Median	0.109	6297.533	2357.346	2.569
Standard deviation	0.143	2503.023	1647.037	0.933
Min	−0.273	388	435.097	1.097
Max	0.9137	22384.460	14845.000	6.910
N	13695	2625	13643	2446
Mean Difference	0.013***	3158.057***	1370.554***	-0.497***

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

Table 1(b): Summary Statistics of Other County Characteristics

Variables	GDPpercapita (thousand yuan)	GDP (billion yuan)	Fraction of industrial output	Fraction of non-state firms	Fraction of micro firms
Counties with clusters					
Mean	14.270	8.863	1.097	0.833	0.056
Median	10.084	6.112	0.845	0.904	0.033
Standard deviation	13.205	8.876	0.930	0.184	0.075
Min	0.133	0.026	0.016	0	0
Max	85.599	48.526	5.000	1	0.909
N	5544	5626	5626	6914	6914
Counties without clusters					
Mean	6.923	2.788	0.523	0.640	0.079
Median	4.898	1.765	0.3350	0.708	0.029
Standard deviation	7.461	3.607	0.6577	0.287	0.142
Min	0.133	0.026	0.009	0	0
Max	85.599	48.526	5.000	1	1
N	16194	16367	16367	20572	20572
Mean Difference	0.735***	60.743***	0.574***	0.194***	-0.022***

Variables	Fraction of education expenditure	Fraction of fixed asset investment	Fraction of government expenditure	Provincial inflow of migrant	Number of SEZ	Poor
Counties with clusters						
Mean	0.019	0.363	0.011	5.128	0.465	0.076
Median	0.015	0.312	0.008	3.470	0	0
Standard deviation	0.014	0.246	0.012	3.854	0.620	0.267
Min	0.004	0.007	0.003	0	0	0
Max	0.136	1.756	0.154	17.03	4	1
N	5192	4528	5194	6229	6914	6914
Counties without clusters						
Mean	0.034	0.376	0.025	3.062	0.172	0.254
Median	0.026	0.311	0.016	2.630	0	0
Standard deviation	0.026	0.275	0.026	2.437	0.407	0.436
Min	0.004	0.007	0.003	0	0	0
Max	0.136	1.756	0.154	17.03	3	1
N	14920	13184	14918	18498	20572	20572
Mean Difference	-0.016***	-0.013***	-0.014***	2.066***	0.293***	-0.178***

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

Table 2: Clustering and Economic Growth
Panel A: Baseline OLS Estimations

	Economic growth				
	(1)	(2)	(3)	(4)	(5)
Cluster	0.008*				
	(1.903)				
OutputStrength		0.010***			
		(3.170)			
EstablishmentStrength			0.006**		
			(2.299)		
OutputNonstate				0.017**	
				(2.421)	
EstablishmentNonstate					0.016**
					(2.331)
GDPpercapita	-0.158***	-0.160***	-0.159***	-0.158***	-0.158***
	(-14.549)	(-14.532)	(-14.543)	(-14.562)	(-14.561)
GDPpercapita (square)	-0.011***	-0.012***	-0.011***	-0.011***	-0.011***
	(-5.238)	(-5.367)	(-5.273)	(-5.255)	(-5.254)
CPI	-0.297**	-0.291**	-0.293**	-0.298**	-0.297**
	(-2.446)	(-2.392)	(-2.409)	(-2.454)	(-2.446)
Fraction of industrial output	0.031***	0.030***	0.031***	0.031***	0.031***
	(9.113)	(8.815)	(9.023)	(9.081)	(9.083)
Fraction of non-state firms	0.036***	0.037***	0.036***	0.035***	0.035***
	(2.782)	(2.843)	(2.786)	(2.730)	(2.704)
Fraction of micro firms	0.007	0.008	0.006	0.007	0.007
	(0.448)	(0.482)	(0.389)	(0.451)	(0.434)
Fraction of edu expenditure	0.014*	0.014*	0.014*	0.014**	0.014**
	(1.939)	(1.941)	(1.951)	(1.972)	(1.963)
Fraction of f. a. investment	0.021***	0.022***	0.021***	0.021***	0.021***
	(8.890)	(8.963)	(8.898)	(8.897)	(8.909)
Fraction of gov expenditure	0.014***	0.014***	0.014***	0.014**	0.014**
	(2.592)	(2.614)	(2.586)	(2.566)	(2.573)
Provincial inflow of migrant	-0.002	-0.002	-0.002	-0.003	-0.003
	(-0.211)	(-0.225)	(-0.267)	(-0.279)	(-0.289)
Number of SEZ	0.010***	0.009***	0.010***	0.010***	0.010***
	(2.726)	(2.606)	(2.657)	(2.676)	(2.682)
Poor	-0.013	-0.013	-0.013	-0.013	-0.013
	(-1.057)	(-1.036)	(-1.059)	(-1.045)	(-1.041)
Constant	0.427***	0.425***	0.427***	0.428***	0.428***
	(13.954)	(13.861)	(13.948)	(13.976)	(13.980)
Year fixed effects	Yes	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes	Yes
Clustered standard error at county-level	Yes	Yes	Yes	Yes	Yes
N	14729	14729	14729	14729	14729
R-square	0.382	0.383	0.382	0.382	0.382

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

Panel B: Two-stage Estimations on Clustering and Economic Growth

	1st stage estimations					2nd stage estimations				
	(1) Cluster	(2) Output Strength	(3) Establishment Strength	(4) Output Nonstate	(5) Establishment Nonstate	(6) Economic growth	(7) Economic growth	(8) Economic growth	(9) Economic growth	(10) Economic growth
Per capita mining output	-0.003**	-0.004***	-0.006***	-0.003***	-0.003***					
	(-2.291)	(-2.691)	(-3.091)	(-3.726)	(-4.157)					
Cluster						1.180**				
						(2.138)				
OutputStrength							0.709**			
							(2.479)			
EstablishmentStrength								0.536***		
								(2.742)		
OutputNonstate									1.150***	
									(3.168)	
EstablishmentNonstate										1.088***
										(3.415)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered standard error at county level	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	12089	12089	12089	12089	12089	12089	12089	12089	12089	12089
P-Values of F-tests	0.0221	0.0072	0.0020	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Weak identification Cragg-Donald Wald F statistic	5.561	7.769	9.948	15.222	17.125					
Underidentification test p-value	0.0225	0.0073	0.0021	0.0002	0.0000					

Note: For convenience, we do not present all the control variables in the table. The control variables for both 1st and 2nd stage estimations are the same as those shown in Table 2 Panel A. The values in parentheses are t- statistics; * = $p < 0.1$; ** = $p < 0.05$; *** = $p < 0.01$.

Table 3: Clustering and Urban-Rural Income Inequality
Panel A: Baseline OLS Estimations

	Urban-Rural Inequality				
	(1)	(2)	(3)	(4)	(5)
Cluster	-0.010				
	(-1.229)				
OutputStrength		0.001			
		(0.147)			
EstablishmentStrength			-0.005		
			(-0.977)		
OutputNonstate				-0.030**	
				(-1.988)	
EstablishmentNonstate					-0.028*
					(-1.905)
GDPpercapita	0.053	0.054	0.054	0.054	0.054
	(1.432)	(1.476)	(1.454)	(1.453)	(1.452)
GDP	0.011	0.009	0.010	0.011	0.011
	(0.318)	(0.260)	(0.304)	(0.317)	(0.324)
Fraction of industrial output	0.007	0.006	0.007	0.008	0.008
	(0.755)	(0.616)	(0.751)	(0.850)	(0.845)
Fraction of non-state firms	0.054	0.053	0.054	0.056	0.056
	(1.074)	(1.067)	(1.069)	(1.110)	(1.111)
Fraction of micro firms	-0.010	-0.014	-0.010	-0.009	-0.008
	(-0.211)	(-0.287)	(-0.200)	(-0.179)	(-0.157)
Fraction of edu expenditure	0.044**	0.044**	0.044**	0.044**	0.044**
	(2.420)	(2.419)	(2.413)	(2.398)	(2.409)
Fraction of f. a. investment	0.019**	0.019**	0.019**	0.019**	0.019**
	(2.226)	(2.221)	(2.222)	(2.224)	(2.215)
Fraction of gov expenditure	-0.019	-0.019	-0.019	-0.018	-0.018
	(-1.272)	(-1.291)	(-1.270)	(-1.237)	(-1.242)
Provincial inflow of migrant	-0.131***	-0.131***	-0.131***	-0.131***	-0.131***
	(-3.327)	(-3.324)	(-3.311)	(-3.318)	(-3.315)
Number of SEZ	0.041***	0.041***	0.041***	0.042***	0.041***
	(2.593)	(2.584)	(2.595)	(2.621)	(2.609)
Poor	-0.003	-0.004	-0.003	-0.003	-0.003
	(-0.158)	(-0.183)	(-0.155)	(-0.140)	(-0.150)
Constant	1.086**	1.109**	1.090**	1.087**	1.084**
	(2.338)	(2.450)	(2.351)	(2.328)	(2.328)
Year fixed effects	Yes	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes	Yes
Clustered standard error at county level	Yes	Yes	Yes	Yes	Yes
N	3187	3187	3187	3187	3187
R-square	0.214	0.214	0.214	0.215	0.215

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

Panel B: Two-stage Estimations on Entrepreneurial Clusters and Urban-Rural Income Inequality

	1st stage estimations		2nd stage estimations	
	(1) OutputNonstate	(2) EstablishmentNonstate	(3) Urban-rural inequality	(4) Urban-rural inequality
Per capita church	0.004***	0.004***		
	(3.022)	(3.317)		
Per capita classified highway	0.036	0.088**		
	(1.219)	(2.485)		
OutputNonstate			-0.782*	
			(-1.746)	
EstablishmentNonstate				-0.687**
				(-2.208)
Year fixed effects	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes
Clustered standard error at county level	Yes	Yes	Yes	Yes
N	2536	2536	2536	2536
P-Values of F-tests	0.0064	0.0002	0.0000	0.0000
Weak identification Cragg-Donald Wald F statistic	3.440	6.661		
Underidentification test p-value	0.1560	0.0315		
Sargan test p-value	0.1170	0.4841		

Note: For convenience, we do not present all the control variables in the table. The control variables for both 1st and 2nd stage estimations are the same as those shown in Table 3 Panel A. The values in parentheses are t- statistics; * = $p < 0.1$; ** = $p < 0.05$; *** = $p < 0.01$.

Table 4: Entrepreneurial Clustering and Household Per Capita Income
Panel A: Baseline OLS Estimations

	Rural household income		Urban household income	
	(1)	(2)	(3)	(4)
OutputNonstate	0.031***		-0.000	
	(3.117)		(-0.026)	
EstablishmentNonstate		0.027***		-0.006
		(2.707)		(-0.415)
GDPpercapita	0.004	0.004	0.148***	0.148***
	(0.504)	(0.507)	(3.014)	(3.000)
GDP	0.138***	0.138***	-0.060	-0.059
	(4.885)	(4.885)	(-1.481)	(-1.465)
Fraction of industrial output	0.007	0.007	0.022**	0.023***
	(1.061)	(1.082)	(2.545)	(2.588)
Fraction of non-state firms	0.117***	0.116***	0.035	0.036
	(3.350)	(3.336)	(0.927)	(0.938)
Fraction of micro firms	-0.053	-0.054	-0.042	-0.041
	(-1.451)	(-1.453)	(-0.864)	(-0.836)
Fraction of edu expenditure	0.039***	0.038***	-0.023	-0.023
	(3.024)	(3.010)	(-1.021)	(-1.024)
Fraction of f. a. investment	0.017***	0.017***	0.039***	0.039***
	(4.509)	(4.508)	(6.752)	(6.757)
Fraction of gov expenditure	-0.051***	-0.050***	0.024*	0.025*
	(-3.617)	(-3.611)	(1.882)	(1.896)
Provincial inflow of migrant	0.145***	0.145***	-0.113***	-0.113***
	(6.693)	(6.688)	(-3.893)	(-3.896)
Number of SEZ	-0.019**	-0.019**	0.046***	0.046***
	(-2.572)	(-2.558)	(2.849)	(2.856)
Poor	0.012	0.012	-0.005	-0.005
	(0.226)	(0.229)	(-0.362)	(-0.353)
Constant	6.092***	6.090***	10.147***	10.142***
	(17.180)	(17.148)	(18.841)	(18.785)
Year fixed effects	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes
Clustered standard error at county-level	Yes	Yes	Yes	Yes
N	12788	12788	3326	3326
R-square	0.470	0.470	0.787	0.787

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

Panel B: Two-stage Estimations on Entrepreneurial Clustering and Rural Household Per Capita Income

	1st stage estimations		2nd stage estimations	
	(1) OutputNonstate	(2) EstablishmentNonstate	(3) Rural household income	(4) Rural household income
Per capita church	0.001	0.001		
	(1.234)	(0.753)		
Per capita classified highway	0.079***	0.101***		
	(3.052)	(3.612)		
OutputNonstate			2.742**	
			(2.509)	
EstablishmentNonstate				2.668***
				(2.876)
Year fixed effects	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes
Clustered standard error at county level	Yes	Yes	Yes	Yes
N	10069	10069	10069	10069
P-Values of F-tests	0.0050	0.0012	0.0000	0.0000
Weak identification Cragg-Donald Wald F statistic	5.157	6.805		
Underidentification test p-value	0.0097	0.0047		
Sargan test p-value	0.1789	0.4801		

Note: For convenience, we do not present all the control variables in the table. The control variables for both 1st and 2nd stage estimations are the same as those shown in Table 4. The values in parentheses are t- statistics; * = p < 0.1; ** = p < 0.05; *** = p < 0.01.

Table 5a: Clustering, Economic Growth and Urban-rural Inequality (Sub-sample of Counties Outside Megacities)

	Panel 1: Economic Growth					Panel 2: Urban-rural Income Inequality				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Cluster	0.008*					-0.011				
	(1.837)					(-1.321)				
OutputStrength		0.011***					0.000			
		(3.364)					(0.013)			
EstablishmentStrength			0.006**					-0.005		
			(2.218)					(-1.100)		
OutputNonstate				0.017**					-0.032**	
				(2.367)					(-2.113)	
EstablishmentNonstate					0.016**					-0.030**
					(2.255)					(-2.047)
Constant	0.440***	0.437***	0.440***	0.441***	0.441***	1.265**	1.290***	1.269**	1.268**	1.262**
	(14.465)	(14.358)	(14.458)	(14.487)	(14.494)	(2.530)	(2.651)	(2.539)	(2.528)	(2.519)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered standard error at county-level	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	14584	14584	14584	14584	14584	3106	3106	3106	3106	3106
R-square	0.387	0.388	0.388	0.388	0.388	0.215	0.215	0.215	0.216	0.216

Note: To save space, we do not present all the control variables in the table. The control variables are the same as those shown in baseline estimations. Values in parentheses are t- statistics; * = p < 0.1; ** = p < 0.05; *** = p < 0.01.

Table 5b: Clustering, Economic Growth and Urban-rural Inequality (Sub-sample of Counties in Megacities)

	Panel 1: Economic Growth					Panel: Urban-rural Income Inequality				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Cluster	0.012					0.012				
	(0.213)					(0.386)				
OutputStrength		0.002					-0.003			
		(0.089)					(-0.174)			
EstablishmentStrength			0.014					0.025		
			(0.448)					(0.951)		
OutputNonstate				-0.007					0.024	
				(-0.080)					(0.510)	
EstablishmentNonstate					0.005					0.075
					(0.049)					(1.013)
Constant	1.501	1.509	1.461	1.517	1.512	1.206	1.155	1.249	1.229	1.195
	(1.291)	(1.263)	(1.252)	(1.296)	(1.294)	(1.042)	(1.019)	(1.065)	(1.053)	(1.039)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered standard error at county-level	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	145	145	145	145	145	81	81	81	81	81
R-square	0.560	0.560	0.561	0.560	0.560	0.446	0.446	0.454	0.447	0.457

Note: To save space, we do not present all the control variables in the table. The control variables are the same as those shown in baseline estimations except that the coefficients of “Poor” are omitted because there are no Poor counties in these megacities, the coefficients are absorbed by the county fixed effects. Values in parentheses are t- statistics; * = p < 0.1; ** = p < 0.05; *** = p < 0.01.

Table 6a: Clustering, Economic Growth and Urban-rural Inequality in Less Urbanized Counties (Measured by Ratio of Urban Population)

	Panel 1: Economic growth					Panel 2: Urban-rural Income Inequality				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Cluster	0.009					-0.023**				
	(1.122)					(-2.096)				
OutputStrength		0.021***					-0.003			
		(3.017)					(-0.406)			
EstablishmentStrength			0.009*					-0.007		
			(1.686)					(-1.041)		
OutputNonstate				0.034***					-0.036**	
				(2.699)					(-2.033)	
EstablishmentNonstate					0.029**					-0.041**
					(2.429)					(-2.363)
Constant	0.505***	0.496***	0.503***	0.505***	0.506***	-4.324	-4.170	-4.265	-4.323	-4.382
	(8.656)	(8.534)	(8.650)	(8.632)	(8.645)	(-1.032)	(-0.975)	(-1.008)	(-1.036)	(-1.052)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered standard error at county-level	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	4889	4889	4889	4889	4889	1044	1044	1044	1044	1044
R-square	0.441	0.443	0.442	0.443	0.443	0.334	0.332	0.333	0.334	0.335

Note: To save space, we do not present all the control variables in the table. The control variables are the same as those shown in baseline estimations. Values in parentheses are t- statistics; * = p < 0.1; ** = p < 0.05; *** = p < 0.01.

Table 6b: Clustering, Economic Growth and Urban-rural Inequality in More Urbanized Counties (Measured by Ratio of Urban Population)

	Panel 1: Economic growth					Panel 2: Urban-rural Income Inequality				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Cluster	0.004					-0.023				
	(0.623)					(-0.894)				
OutputStrength		0.004					0.005			
		(0.842)					(0.395)			
EstablishmentStrength			0.001					-0.017		
			(0.268)					(-1.051)		
OutputNonstate				-0.013					-0.106**	
				(-1.183)					(-1.977)	
EstablishmentNonstate					-0.002					-0.077
					(-0.195)					(-1.513)
Constant	0.394***	0.394***	0.394***	0.393***	0.394***	-0.117	-0.032	-0.130	-0.200	-0.144
	(6.305)	(6.302)	(6.306)	(6.293)	(6.303)	(-0.104)	(-0.029)	(-0.114)	(-0.176)	(-0.126)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustered standard error at county-level	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	4455	4455	4455	4455	4455	1103	1103	1103	1103	1103
R-square	0.372	0.372	0.372	0.372	0.372	0.170	0.170	0.171	0.176	0.173

Note: To save space, we do not present all the control variables in the table. The control variables are the same as those shown in baseline estimations. Values in parentheses are t- statistics; * = p < 0.1; ** = p < 0.05; *** = p < 0.01

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson, 2002. "Reversal of Fortune: Geography and institutions in the making of the modem world income distribution." *Quarterly Journal of Economics* 118: 1231-1294.

Acemoglu, Daron, Simon Johnson, and James A. Robinson, 2005. "Institutions as a fundamental cause of long-run growth." *Handbook of Economic Growth* 1: 385-472.

Aghion, Philippe, Eve Caroli, and Cecilia Garcia-Penalosa, 1999. "Inequality and Economic Growth: The Perspective of the New Growth Theories." *Journal of Economic Literature* 37: 1615-1660

Aghion, Philippe, and Peter Howitt, 1992. "A Model of Growth through Creative Destruction." *Econometrica* 60(2): 323-351.

Aghion, Philippe, Peter Howitt, and Giovanni L. Violante, 2002. "General purpose technology and wage inequality." *Journal of Economic Growth 7*(4): 315-345.

Allen, Franklin, Jun Qian, and Meijun Qian. 2005. "Law, Finance, and Economic Growth in China." *Journal of Financial Economics* 77: 57-116.

Arrow, Kenneth J., 1962. "The Economic Implications of Learning by Doing." *The Review of Economic Studies* 29(3): 155-173.

Au, Chun-Chung, and Vernon Henderson, 2006. "How migration restrictions limit agglomeration and productivity in China." *Journal of Development Economics* 80(2): 350-388.

Bai, Ying and Ruixue Jia, 2018, "The Oriental City: Political Hierarchy and Regional Development in China, AD1000-2000," working paper, Chinese University of Hong Kong and UC-San Diego.

Banerjee, Abhijit, and Esther Duflo, 2003. "Inequality and Growth: What Can the Data Say?" *Journal of Economic Growth* 8: 267-299.

Barro, Robert J, 2000. "Inequality and Growth in a Panel of Countries." *Journal of Economic Growth* 5(1): 5-32.

Barro, Robert, and Xavier Sala-i-Martin, 1992. "Convergence." *Journal of Political Economy* 100(2): 223-251.

Baumol, William, 1986. "Productivity Growth, Convergence, and Welfare: What the Long-run Data Show." *American Economic Review* 76(5): 1072-1085.

Catalini, C. 2018. "Microgeography and the direction of inventive activity." *Management Science* 64(9): 4348-4364

Chinitz, Benjamin. 1961. "Contrasts in Agglomeration: New York and Pittsburgh." *The American Economic Review* 51(2): 279-289.

Ciccone, Antonio, and Robert Hall, 1996. "Productivity and the Density of Economic Activity." *American Economic Review* 86(1): 54-70.

Cingano, Federico, and Fabiano Schivardi, 2004. "Identifying the sources of local productivity growth." *Journal of the European Economic Association 2*(4): 720-744.

Combes, P.P., Duranton, G. and Gobillon, L., 2018. "The costs of agglomeration: House and land prices in French cities." *CEPR Discussion Paper No. DP9240 (updated)*.

Dekle, R. (2002). Industrial concentration and regional growth: evidence from the prefectures. *Review of Economics and Statistics 84*(2), 310-315.

Delgado, Mercedes; Michael Porter, and Scott Stern, 2014. “Cluster, convergence, and economic performance.” *Research Policy* 43: 1785-1799.

Dumitrescu, E. I., & Hurlin, C, 2012. “Testing for Granger non-causality in heterogeneous panels.” *Economic Modelling 29*(4), 1450-1460.

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

Ellison, Glenn, and Edward Glaeser, 1997. “Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach.” *Journal of Political Economy* 105(5): 889-892.

Fan, Gang, Xiaolu Wang, and Hengpeng Zhu, 2011. *NERI Index of Marketization of China's Provinces (Zhongguo shichang jingji zhishu)*. Beijing: Economic Science Press.

Feldman, Maryann P., and David B. Audretsch, 1999. “Innovation in cities: Science-based diversity, specialization, and localized competition.” *European economic review* 43(2): 409-429.

Feldman, M. 2014. “The character of innovative places: entrepreneurial strategy, economic development and prosperity.” *Small Business Economics* 3(1): 9-20.

Fujita, Masahisa, and Dapeng Hu, 2001. “Regional disparity in China 1985–1994: the effects of globalization and economic liberalization.” *The Annals of Regional Science* 35(1): 3-37.

Fujita, M. and Thisse, J.F., 2013. *Economics of agglomeration: cities, industrial location, and globalization*. Cambridge university press.

Glaeser, Edward, Hedi Kallal, Jose Scheinkman and Andrei Shleifer, 1992. “Growth in cities.” *Journal of Political Economy* 100(6): 1126-1152.

Glaeser, Edward, and Joshua Gottlieb, 2009. “The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States.” *Journal of Economic Literature* 47(4): 983-1028.

Glaeser, Edward L., Stuart S. Rosenthal, and William C. Strange. 2010 “Urban economics and entrepreneurship.” *Journal of Urban Economics* 67(1): 1-14.

Glaeser, Edward., Sari Pekkala Kerr, and William Kerr. 2015. “Entrepreneurship and urban growth: An empirical assessment with historical mines.” *Review of Economics and Statistics* 97(2): 498-520.

Guo, Di, Kun Jiang, Byung-Yeon Kim, and Chenggang Xu, 2014. “Political economy of private firms in China.” *Journal of Comparative Economics* 42: 286-303.

Granger, C. W., 1969. “Investigating causal relations by econometric models and cross-spectral methods.” *Econometrica*: 424-438.

Hausmann, Ricardo, and Bailey Klinger, 2007. “The structure of the product space and the evolution of comparative advantage.” Working Paper, Harvard University.

Heckman, J. J., & Hotz, V. J. (1986). An investigation of the labor market earnings of Panamanian males evaluating the sources of inequality. *Journal of Human Resources*, 507-542.

Henderson, Vernon, 2003. “Urbanization and Economic Development.” *Annals of*

Economics and Finance 4(2): 275-341.

Henderson, Vernon, Ari Kuncoro, and Matt Turner, 1995. "Industrial development in cities." *Journal of political economy* 103(5): 1067-1090.

Hessler, Peter, 2007. "Boomtowns." *National Geographic*, June: 88-117.

Jacobs, Jane, 1969. *The Economy of Cities*. Random House: New York

Kanbur, Ravi, and Xiaobo Zhang, 1999. "Which Regional Inequality? The Evolution of Urban-rural and Inland-Coastal Inequality in China from 1983 to 1995." *Journal of Comparative Economics* 27(4): 686-701.

Kanbur, Ravi, and Xiaobo Zhang, 2005. "Fifty Years of Regional Inequality in China: A Journey through Central Planning, Reform, and Openness." *Review of Development Economics* 9(1): 87-106.

Kerr, William, 2010. "Breakthrough inventions and migrating clusters of innovation." *Journal of Urban Economics* 67(1): 46-60.

Knowles, Stephen, 2001. "Inequality and Economic Growth: The empirical Relationship Reconsidered in the Light of Comparable Data." *Journal of Development Studies* 41(1): 135-159.

Krugman, Paul, 1991a. *Geography and Trade*. Cambridge MA: MIT Press.

Krugman, Paul, 1991b. "Increasing returns and economic geography." *Journal of Political Economy* 99(3): 484-499.

Krugman, Paul, and Anthony Venables, 1995. "Globalization and the Inequality of Nations." *The Quarterly Journal of Economics* 110(4): 857-880.

Kuznets, Simon, 1955. "Economic Growth and Income Inequality." *American Economic Review* 45(1): 1–28.

Kuznets, Simon, 1963. "Quantitative aspects of the economic growth of nations, VIII: The distribution of income by size." *Economic Development and Cultural Change* 11(2): 1–92.

Li, Shi, Hiroshi Sato, and Terry Sicular (ed.), 2013. *Rising Inequality in China*. Cambridge: Cambridge University Press.

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

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

Marshall, Alfred, 1890. *Principles of Economics: An Introductory Volume.* London: Macmillan.

McCleary, R.M. and R.J. Barro, 2006. "Religion and economy." *Journal of Economic Perspectives* 20(2): 49-72.

Midelfart-Knarvik, Karen Helene, Henry G. Overman, Stephen J. Redding, and Anthony J. Venables, 2000. "The location of European industry." Working Paper.

Mudambi, R., Li, L., Ma, X., Makino, S., Qian, G., & Boschma, R, 2018. "Zoom in, zoom out: Geographic scale and multinational activity." *Journal of International Business Studies*, *49*(8), 929-941.

Perotti, Roberto, 1996. “Growth, Income Distribution, and Democracy: What the Data Say.” *Journal of Economic Growth* 1(2): 149-187.

Persson, Torsten, and Guido Tabellini, 1994. “Is Inequality Harmful for Growth?” *American Economic Review* 84(3): 600-621.

Peterson, Mitchell A, 2009. “Estimating standard error in finance panel data sets: Compering approaches.” *Review of Financial Studies* 22: 435-480.

Porter, Michael, 1990. *The Competitive Advantage of Nations*. London: Macmillan.

Porter, Michael, 2003 “The Economic Performance of Regions.” *Regional Studies* 37(6-7): 549-578.

Romer, Paul M, 1986. “Increasing Returns and Long-run Growth.” *Journal of Political Economy* 94(5): 1002-1037.

Rosenthal, Stuart S., and William C. Strange. 2003. “Geography, industrial organization, and agglomeration.” *Review of Economics and Statistics* 85(2): 377-393.

Ruan, Jianqing, and Xiaobo Zhang, 2009. “Finance and Cluster-Based Industrial Development in China.” *Economic Development and Cultural Change* 58(1): 143-164.

Sicular, Terry, 2013, “The Challenge of High Inequality in China.” Poverty Reduction and Equity Department, The World Bank.

Solow, Robert M., 1956. “A contribution to the theory of economic growth.” *The Quarterly Journal of Economics* 70(1): 65-94.

Sonobe, Tetsushi, Dinghuan Hu, and Keijiro Otsuka, 2002. “Process of cluster formation in China: A case study of a garment town.” *Journal of Development Studies* 39(1): 118-139.

Sonobe, Tetsushi, and Keijiro Otsuka, 2006. *Cluster-based industrial development: An East Asia model*. New York: Palgrave MacMillan.

Weitzman, Martin L., and Chenggang Xu, 1994. “Chinese Township-Village Enterprises as Vaguely Defined Cooperatives,” *Journal of Comparative Economics*, 18(2): 121-145, 1994.

Whyte, Martin King (ed.), 2010, *One Country, Two Societies: Rural-Urban Inequality in Contemporary China*, Harvard University Press. 2010.

Xu, Chenggang, 2011. “The Fundamental Institutions of China’s Reform and Development.” *Journal of Economic Literatures* 49(4): 1076-1151.

Xu, Chenggang, and Xiaobo Zhang, 2009. “The Evolution of Chinese Entrepreneurial Firms: Township-Village Enterprises Revisited.” In *China’s Economic Transformation* edited by Ronald Coase, forthcoming.

Online Appendix

Table A-1 Variable Definitions

Variables	Definitions
Economic growth	Growth rate of per capita GDP in a given county in a given year in logarithm form: $\ln\left(\frac{GDPpercapita_{it+1}}{GDPpercapita_{it}}\right)$
Urban–rural inequality	Ratio of urban household per capita income over rural household per capita income of a given county in a given year: $ratio_{it} = urban\ household\ per\ capita\ income_{it} / rural\ household\ per\ capita\ income_{it}$
Urban household income	Urban household per capita disposable income in a given county in a given year
Rural household income	Rural household per capita net income in a given county in a given year
$Cluster_{it}$	A dummy variable that equals one if county *i* has any industrial cluster in year *t* and equals zero if otherwise.
$OutputStrength_{it}$	Weighted average of output strength of the cluster(s) in county *i* in year *t*
$EstablishmentStrengt$	Weighted average of establishment strength of the cluster(s) in county *i* in year *t*
$OutputNonstate_{it}$	The share of non-state firms in the total output of the clusters in county *i* at time *t*
$EstablishmentNonsta$	The share of non-state firms in the total establishments of the clusters in county *i* at time *t*
GDPpercapita	Per capita GDP of a given county in a given year
GDP	Total GDP of a given county in a given year
CPI	Provincial-level Consumer Price Index in a given year
Fraction of industrial output	Ratio of total industrial output over total GDP of a given county in a given year
Fraction of non-state firms	Ratio of the number of non-state-owned firms over the total number of firms in a given county in a given year
Fraction of micro firms	Ratio of the number of micro firms over the total number of firms in a given county in a given year
Fraction of edu expenditure	Ratio of fiscal expenditure in education over total fiscal expenditure of a given county in a given year
Fraction of f. a. investment	Ratio of total investment in fixed assets over total GDP of a given county in a given year
Fraction of gov expenditure	Ratio of administration expenditure over total fiscal expenditure in a given county in a given year
Provincial inflow of migrant	Ratio of non-local rural labor force over total number of employment in a given province in a given year, from Fan el al. (2011)
Number of SEZ	Number of provincial-level special economic zones that located in a given county in a given year
Poor	A dummy variable that equals one if a county is identified as a national designated poor county in a given year and equals zero if otherwise.
Instrumental Variables	
Per capita mining output	The per capita mining outputs of a province in a given year
Per capita church	The per capita number of Christian churches in a given county in a given year, extracted from the Spatial Explorer on Religion dataset
Per capita classified highway	The per capita number of classified highways in a city in a given year

Table A-2: Summary Statistics of DBI Clusters

Identified Clusters Over Time

Year	Number of counties	County-industry observations	Counties with clusters	Number of industrial clusters	Clusters' share in total output	Clusters' share in total employment
1998	2700	41899	685	2024	0.3838	0.2797
1999	2747	41571	654	2037	0.3527	0.2546
2000	2746	40272	657	1958	0.3733	0.2821
2001	2673	38712	632	1896	0.3633	0.2760
2002	2726	39432	668	1931	0.3680	0.2905
2003	2787	40207	709	1971	0.3811	0.3124
2004	2790	41996	696	2118	0.4035	0.3650
2005	2790	41809	750	2125	0.3914	0.3636
2006	2793	43177	724	2157	0.3911	0.3748
2007	2734	44175	739	2213	0.3802	0.3668

Features of Clusters

	OutputStrength	Establishment Strength	OutputNonstate	Establishment Nonstate
Mean	6.5391	5.9087	0.8016	0.8331
Median	3.9976	4.7278	0.9523	0.9444
Standard deviation	8.1344	3.6840	0.2922	0.2511
Min	0.1041	2.3384	0	0
Max	50.9959	23.4122	1	1
N	6914	6914	6914	6914

Dynamics of Counties and Clusters

Counties always with clusters, 98-07	Counties always without clusters, 98-07	Counties without clusters in 1998, but with clusters in 2007	Counties with clusters in 1998, but without clusters in 2007
294	1,576	317	292

Table A-3: Clustering Measurements at Provincial Level: DBI vs. Standard Approaches

Province	HHI	Gini	LQ	Krugman Index	Industrial Proximity	DBI Cluster Number	DBI OutputStrength	DBI EstablishmentStrength	DBI OutputNonstate	DBI EstablishmentNonstate
Beijing	0.0913	0.6668	2.5211	0.6478	0.22	64	3.8909	3.3761	0.5189	0.5199
Tianjin	0.0808	0.6372	1.7035	0.5681	0.208	57	6.4096	3.9931	0.5569	0.5491
Hebei	0.1205	0.6325	1.9691	0.5956	0.219	44	1.4906	1.3629	0.2108	0.2140
Shanxi	0.1605	0.8014	4.8040	1.0693	0.208	25	0.6417	1.1312	0.1842	0.1859
Inr Mongolia	0.0802	0.6689	2.6402	0.8504	0.214	11	0.6925	0.4150	0.0663	0.0807
Liaoning	0.0656	0.5983	1.3517	0.4519	0.205	145	2.4913	2.8489	0.4274	0.4351
Jilin	0.1527	0.7231	2.6820	0.8912	0.22	24	1.7743	1.5403	0.2677	0.3055
Heilongjiang	0.1217	0.7201	4.7328	0.9622	0.197	8	0.3400	0.2626	0.0473	0.0585
Shanghai	0.0871	0.5938	1.4276	0.4596	0.219	165	9.2155	5.3317	0.5761	0.6106
Jiangsu	0.0684	0.6208	1.2523	0.4317	0.21	284	5.5121	4.8527	0.6344	0.6375
Zhejiang	0.0530	0.5138	1.4781	0.4851	0.22	414	6.2603	7.5372	0.7646	0.7770
Anhui	0.0584	0.5520	1.2671	0.3973	0.211	18	0.5599	0.6983	0.1411	0.1394
Fujian	0.0503	0.4798	1.5538	0.4789	0.202	112	2.9643	3.8209	0.5252	0.5500
Jiangxi	0.0779	0.5705	1.8355	0.4831	0.206	19	0.8680	0.7392	0.1305	0.1483
Shandong	0.0506	0.5183	1.2188	0.3872	0.205	233	5.9145	4.6156	0.6733	0.6896
Henan	0.0560	0.5513	1.5212	0.5771	0.209	51	1.9954	1.3594	0.2078	0.2194
Hubei	0.0706	0.6004	1.3529	0.4179	0.216	27	1.0604	0.9416	0.1654	0.1818
Hunan	0.0556	0.5426	1.4514	0.4850	0.21	47	0.7942	1.5509	0.2114	0.2203
Guangdong	0.0839	0.5813	1.5243	0.5668	0.215	260	4.0450	3.3413	0.4271	0.4533
Guangxi	0.0791	0.6458	1.6292	0.6692	0.214	11	0.3128	0.4710	0.0824	0.0942
Hainan	0.1555	0.7286	3.6836	0.9652	0.207	2	0.2016	0.4371	0.0870	0.0870
Chongqing	0.1581	0.6954	2.7831	0.7327	0.197	17	2.2100	2.6696	0.3624	0.3684
Sichuan	0.0517	0.5339	1.3152	0.3757	0.202	52	0.5788	1.0106	0.1912	0.1978

Table A-3: Clustering Measurements at Provincial Level: DBI vs. Standard Approaches (continued)

Province	HHI	Gini	LQ	Krugman Index	Industrial Proximity	DBI Cluster Number	DBI OutputStrength	DBI EstablishmentStrength	DBI OutputNonstate	DBI EstablishmentNonstate
Guizhou	0.1079	0.7017	2.3672	0.8837	0.196	6	0.4276	0.2974	0.0575	0.0583
Yunnan	0.1209	0.7382	5.0078	0.9727	0.197	12	0.3499	0.3980	0.0831	0.0897
Tibet	0.1330	0.6293	11.3831	1.0326	0.238	0	0.0000	0.0000	0.0000	0.0000
Shaanxi	0.0762	0.6401	2.5095	0.7317	0.192	18	0.4552	0.6820	0.0913	0.1064
Gansu	0.1394	0.7531	3.0550	0.9771	0.205	10	0.4165	0.3509	0.0671	0.0556
Qinghai	0.1296	0.7588	5.1840	1.0117	0.217	0	0.0000	0.0000	0.0000	0.0000
Ningxia	0.1173	0.7146	7.3161	0.8497	0.22	3	0.1385	0.4480	0.0769	0.0884
Xinjiang	0.1799	0.7727	7.1079	1.0229	0.199	5	0.1098	0.1739	0.0300	0.0350

Note: The HHI, Gini, LQ, Krugman Index, and DBI measures are based on the 2007 ASIFP data. The Industrial Proximity is from Long and Zhang (2012). DBI Cluster number is the summation of county-level DBI clusters for each province; DBI OutputStrength, DBI EstablishmentStrength, DBI OutputNonstate, and DBI EstablishmentNonstate are the county averages of the corresponding indices that we use in the regressions of the paper for each province.

Table A-4 Weights of SOEs in Regions Defined by Different Clustering Measurements

Clustering measurements	Identified counties	SOE number /total firm number	SOE output / total output	SOE employment/ total employment
DBI	Clustered	1.63%	12.49%	9.12%
	Non-clustered	5.38%	32.4%	29.69%
CR5	Clustered	5.62%	44.04%	39.58%
	Non-clustered	2.21%	11.83%	10.8%
Gini	Clustered	3.93%	30.88%	27.67%
	Non-clustered	2.08%	9.16%	9.1%
LQ	Clustered	4.25%	39.12%	32.9%
	Non-clustered	2.26%	13.06%	11.37%
Krugman index	Clustered	3.22%	29.44%	24.36%
	Non-clustered	2.24%	9.2%	9.62%

Note: The above statistics are calculated based on 2004 Chinese Economic Census data.

Table A-5: Granger Causality Test

Panel A: Clustering and Economic Growth

Sample: 1998 – 2007
Lags: 1

Null Hypothesis:	Obs	F-Statistic	P-value
Cluster does not Granger Cause Growth	18250	8.32536	0.0039
OutputStrength does not Granger Cause Growth	18250	19.3516	0.0000
EstablishmentStrength does not Granger Cause Growth	18250	17.6825	0.0000
OutputNonstate does not Granger Cause Growth	18250	24.8276	0.0000
EstablishmentNonstate does not Granger Cause Growth	18250	25.8022	0.0000

Panel B: Clustering and Local Urban-Rural Inequality

Sample: 1998 – 2007
Lags: 1

Null Hypothesis:	Obs	F-Statistic	P-value
Cluster does not Granger Cause Inequality	2698	19.4226	0.0000
OutputStrength does not Granger Cause Inequality	2698	25.8558	0.0000
EstablishmentStrength does not Granger Cause Inequality	2698	21.1368	0.0000
OutputNonstate does not Granger Cause Inequality	2698	18.6427	0.0000
EstablishmentNonstate does not Granger Cause Inequality	2698	15.4247	0.0001

Table A-6.a Clustering and Non-agricultural Activity Engagement by Rural Residents

	Counties with clusters		Counties without clusters		Difference between the two groups	
	Obs.	Mean	Obs.	Mean	Difference	p-value
Income from Non-agricultural activities (yuan)	43	1,398	38	1,293	105	0.0764
Ratio of residents (above 16) engaged in non-agricultural jobs in general	43	42.76%	38	35.58%	7.18%	0.0140
Ratio of private business owners in people with non-agricultural jobs	43	4.39%	38	2.94%	1.45%	0.0469
Ratio of residents engaged in non-agricultural jobs in home county	43	50.74%	38	27.39%	23.35%	0.0000

Note: the information shown in this table is based on income and personal characteristics information from Chinese Household Income Project (CHIP) 2007 data.

Table A-6.b: Entry and Exit of Firms from 2004-2008 in Counties With and Without Clusters

Variables	Entrants number	Survivor number	Exiting number	Entry rate	Survival rate	Exit rate
Counties with Clusters						
Mean	821.55	607.25	471.38	0.8734	0.4928	0.5072
Median	537	283	324	0.8071	0.5076	0.4924
S.D.	1047.67	969.46	573.86	0.4028	0.1534	0.1534
Min	37	10	20	0.1697	0.0652	0.1210
Max	14458	11192	10150	4.2070	0.8790	0.9348
N	1017	1017	1017	1017	1017	1017
Counties without Clusters						
Mean	186.74	92.36	100.96	1.0844	0.4585	0.5415
Median	127	57	76	0.9301	0.4641	0.5359
S.D.	202.55	105.20	92.39	0.7726	0.1561	0.1561
Min	0	0	0	0	0	0
Max	1928	801	647	12	1	1
N	1830	1830	1830	1830	1830	1830
Mean Difference	634.81***	514.89***	370.42***	-0.2110***	0.0343***	-0.0343***

Data Source: 2004 and 2008 economic census data

Table A-7 Clustering and Economic Growth (Controlling for the Population of Megacities)

	(1)	(2)	(3)	(4)	(5)	(6)
Megacity populations	-0.251	-0.240	-0.222	-0.233	-0.240	-0.245
	(-1.043)	(-0.994)	(-0.918)	(-0.963)	(-0.993)	(-1.021)
Cluster		0.008*				
		(1.860)				
OutputStrength			0.010***			
			(3.126)			
EstablishmentStrength				0.006**		
				(2.245)		
OutputNonstate					0.017**	
					(2.386)	
EstablishmentNonstate						0.016**
						(2.313)
GDPpercapita	-0.157***	-0.240	-0.222	-0.233	-0.240	-0.245
	(-14.525)	(-0.994)	(-0.918)	(-0.963)	(-0.993)	(-1.021)
GDPpercapita (square)	-0.011***	-0.158***	-0.160***	-0.158***	-0.158***	-0.158***
	(-5.149)	(-14.522)	(-14.503)	(-14.515)	(-14.536)	(-14.534)
CPI	-0.318***	-0.011***	-0.011***	-0.011***	-0.011***	-0.011***
	(-2.638)	(-5.172)	(-5.303)	(-5.207)	(-5.189)	(-5.188)
Fraction of industrial output	0.031***	-0.312***	-0.305**	-0.308**	-0.313***	-0.313***
	(9.272)	(-2.596)	(-2.533)	(-2.555)	(-2.602)	(-2.598)
Fraction of non-state firms	0.036***	0.031***	0.030***	0.030***	0.031***	0.031***
	(2.825)	(9.078)	(8.785)	(8.991)	(9.045)	(9.045)
Fraction of micro firms	0.009	0.036***	0.037***	0.036***	0.035***	0.035***
	(0.586)	(2.777)	(2.837)	(2.781)	(2.724)	(2.698)
Fraction of edu expenditure	0.014*	0.007	0.008	0.006	0.007	0.007
	(1.926)	(0.446)	(0.480)	(0.389)	(0.449)	(0.430)
Fraction of f. a. investment	0.021***	0.014*	0.014*	0.014*	0.014*	0.014*
	(8.817)	(1.916)	(1.920)	(1.928)	(1.949)	(1.939)
Fraction of gov expenditure	0.015***	0.021***	0.021***	0.021***	0.021***	0.021***
	(2.634)	(8.831)	(8.908)	(8.840)	(8.837)	(8.848)
Provincial inflow of migrant	-0.003	0.015***	0.015***	0.015***	0.014***	0.015***
	(-0.277)	(2.656)	(2.674)	(2.649)	(2.630)	(2.638)
Number of SEZ	0.010***	-0.003	-0.003	-0.003	-0.004	-0.004
	(2.898)	(-0.320)	(-0.326)	(-0.371)	(-0.387)	(-0.400)
Poor	-0.013	0.010***	0.010***	0.010***	0.010***	0.010***
	(-1.030)	(2.814)	(2.691)	(2.744)	(2.763)	(2.770)
Constant	0.449***	0.447***	0.443***	0.446***	0.448***	0.449***
	(12.577)	(12.525)	(12.428)	(12.503)	(12.554)	(12.585)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Clustered standard error at county-level	Yes	Yes	Yes	Yes	Yes	Yes
N	14729	14729	14729	14729	14729	14729
R-square	.3818963	.3821492	.3827025	.3822924	.3823103	.3822859

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

Table A-8a Clustering and Economic Growth (Controlling for the Three Largest Industries in Each County)

	Economic growth				
	(1)	(2)	(3)	(4)	(5)
Cluster	0.008**				
	(1.971)				
OutputStrength		0.010***			
		(3.184)			
EstablishmentStrength			0.006**		
			(2.415)		
OutputNonstate				0.017**	
				(2.509)	
EstablishmentNonstate					0.017**
					(2.389)
GDPpercapita	-0.159***	-0.161***	-0.160***	-0.160***	-0.160***
	(-14.581)	(-14.581)	(-14.585)	(-14.598)	(-14.596)
GDPpercapita (square)	-0.011***	-0.012***	-0.011***	-0.011***	-0.011***
	(-5.237)	(-5.380)	(-5.278)	(-5.256)	(-5.254)
CPI	-0.283**	-0.277**	-0.278**	-0.284**	-0.283**
	(-2.338)	(-2.287)	(-2.298)	(-2.347)	(-2.339)
Fraction of industrial output	0.031***	0.030***	0.031***	0.031***	0.031***
	(8.898)	(8.586)	(8.798)	(8.871)	(8.871)
Fraction of non-state firms	0.036***	0.037***	0.036***	0.036***	0.035***
	(2.780)	(2.819)	(2.780)	(2.724)	(2.698)
Fraction of micro firms	0.004	0.005	0.003	0.004	0.004
	(0.278)	(0.305)	(0.209)	(0.279)	(0.263)
Fraction of edu expenditure	0.014*	0.014*	0.014*	0.014*	0.014*
	(1.925)	(1.935)	(1.938)	(1.958)	(1.949)
Fraction of f. a. investment	0.021***	0.021***	0.021***	0.021***	0.021***
	(8.903)	(8.985)	(8.914)	(8.911)	(8.923)
Fraction of gov expenditure	0.014**	0.014***	0.014**	0.014**	0.014**
	(2.576)	(2.597)	(2.573)	(2.552)	(2.558)
Provincial inflow of migrant	-0.002	-0.002	-0.003	-0.003	-0.003
	(-0.252)	(-0.266)	(-0.313)	(-0.321)	(-0.328)
Number of SEZ	0.010***	0.009***	0.010***	0.010***	0.010***
	(2.742)	(2.626)	(2.673)	(2.694)	(2.701)
Poor	-0.013	-0.012	-0.013	-0.013	-0.012
	(-1.005)	(-0.987)	(-1.009)	(-0.994)	(-0.990)
Constant	0.434***	0.429***	0.433***	0.434***	0.435***
	(11.789)	(11.659)	(11.764)	(11.815)	(11.817)
Dummies for the 3 largest industries in each county	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes	Yes
Clustered standard error at county-level	Yes	Yes	Yes	Yes	Yes
N	14729	14729	14729	14729	14729
R-square	0.3845	0.3850	0.3847	0.3847	0.3846

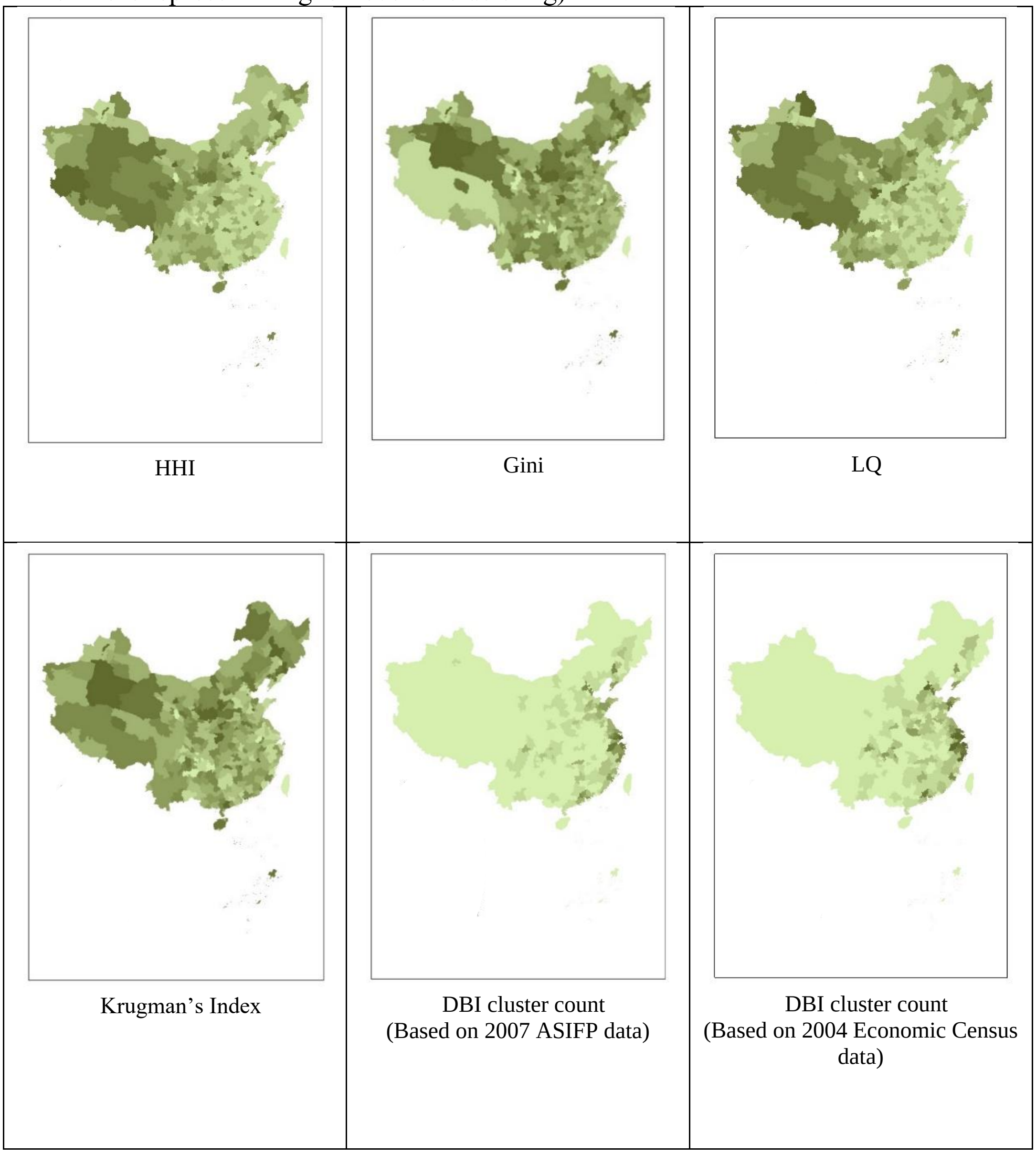
Note: Values in parentheses are t- statistics; * = p < 0.1; ** = p < 0.05; *** = p < 0.01; ☆ = p < 0.15.

Table A-8b Clustering and Urban-Rural Household Income Inequality (Controlling for the Three Largest Industries in Each County)

	Urban-Rural Inequality				
	(1)	(2)	(3)	(4)	(5)
Cluster	-0.012				
	(-1.487)				
OutputStrength		-0.001			
		(-0.115)			
EstablishmentStrength			-0.006		
			(-1.312)		
OutputNonstate				-0.033**	
				(-2.191)	
EstablishmentNonstate					-0.031**
					(-2.122)
GDPpercapita	0.044	0.045	0.045	0.045	0.045
	(1.214)	(1.262)	(1.238)	(1.240)	(1.241)
GDP	0.015	0.013	0.015	0.015	0.015
	(0.448)	(0.397)	(0.437)	(0.445)	(0.451)
Fraction of industrial output	0.011	0.009	0.011	0.012	0.012
	(1.125)	(0.985)	(1.134)	(1.230)	(1.224)
Fraction of non-state firms	0.042	0.042	0.042	0.044	0.045
	(0.836)	(0.832)	(0.833)	(0.876)	(0.881)
Fraction of micro firms	-0.023	-0.026	-0.021	-0.021	-0.020
	(-0.476)	(-0.558)	(-0.450)	(-0.443)	(-0.422)
Fraction of edu expenditure	0.041**	0.041**	0.041**	0.041**	0.041**
	(2.285)	(2.285)	(2.277)	(2.262)	(2.271)
Fraction of f. a. investment	0.017*	0.017*	0.017*	0.017*	0.017*
	(1.911)	(1.903)	(1.905)	(1.907)	(1.897)
Fraction of gov expenditure	-0.018	-0.018	-0.018	-0.018	-0.018
	(-1.249)	(-1.269)	(-1.246)	(-1.215)	(-1.216)
Provincial inflow of migrant	-0.130***	-0.130***	-0.130***	-0.130***	-0.130***
	(-3.348)	(-3.340)	(-3.325)	(-3.337)	(-3.333)
Number of SEZ	0.040**	0.040**	0.040**	0.040**	0.040**
	(2.524)	(2.512)	(2.531)	(2.555)	(2.543)
Poor	-0.005	-0.005	-0.005	-0.004	-0.004
	(-0.199)	(-0.226)	(-0.195)	(-0.180)	(-0.191)
Constant	1.064**	1.083**	1.067**	1.068**	1.065**
	(2.329)	(2.427)	(2.335)	(2.331)	(2.329)
Dummies for the 3 largest industries in each county	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes	Yes
Clustered standard error at county level	Yes	Yes	Yes	Yes	Yes
N	3187	3187	3187	3187	3187
R-square	0.2288	0.2284	0.2288	0.2297	0.2295

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

Figure A-1.a: Prefecture Level Clustering Measured by Standard Approaches and DBI
(Darker areas represent a higher level of clustering)



HHI

Gini

LQ

Krugman's Index

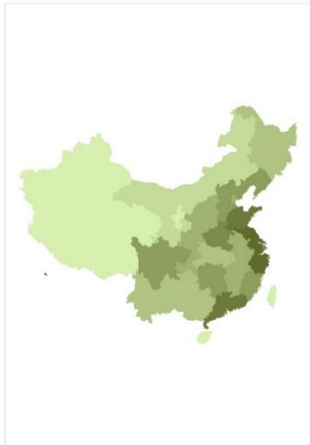
DBI cluster count
(Based on 2007 ASIFP data)

DBI cluster count
(Based on 2004 Economic Census data)

Note: The HHI, Gini, LQ, Krugman Index are calculated from the ASIFP 2007 data.

Figure A-1.b: Provincial Level Clustering Measured by Industrial Proximity and DBI
(Darker areas represent a higher level of clustering)

Industrial Proximity
(Based on 2004 Economic Census data)



DBI cluster count
(Based on 2004 Economic Census data)

Note: The industrial proximity is from Long and Zhang (2011).

Figure A-2: Satellite Night Vision of China vs. Prefecture-Level DBI Cluster Counts

Satellite Photo of China at Night*

DBI cluster count
(Based on 2007 ASIFP data)

* The Satellite Night Vision is taken from the NASA website, and the brightness represents light at night (http://earthobservatory.nasa.gov/NaturalHazards/view.php?id=79790)