DISCUSSION PAPER SERIES

DP14506

GEOGRAPHIC CLUSTERING AND RESOURCE REALLOCATION ACROSS FIRMS IN CHINESE INDUSTRIES

Di Guo, Kun Jiang, Cheng-Gang Xu and Xiyi Yang

DEVELOPMENT ECONOMICS

INDUSTRIAL ORGANIZATION

ISSN 0265-8003

GEOGRAPHIC CLUSTERING AND RESOURCE REALLOCATION ACROSS FIRMS IN CHINESE INDUSTRIES

Di Guo, Kun Jiang, Cheng-Gang Xu and Xiyi Yang

Discussion Paper DP14506
Published 18 March 2020
Submitted 06 March 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

GEOGRAPHIC CLUSTERING AND RESOURCE REALLOCATION ACROSS FIRMS IN CHINESE INDUSTRIES

Abstract

We examine the effects of China's industrial clustering on resource reallocation efficiency across firms. Based on our county-industry level DBI index panel, we find that industrial clustering significantly increases local industries' productivity by lifting the average firm productivity and reallocating resources from less to more productive firms. Moreover, we find major mechanisms through which resource reallocation is improved within clusters: (i) clusters facilitate higher entry rates and exit rates; and (ii) within clusters' environment the dispersion of individual firm's markup is significantly reduced, indicating intensified local competition within clusters. The identification issues are carefully addressed by instrumental variable (IV) regressions.

JEL Classification: D2, H7, L1, O1, R1, R3

Keywords: Industrial Cluster, Productivity Growth, Resource reallocation, Competition

Di Guo - guodi2007@gmail.com
Brunel University London

Kun Jiang - kun.jiang@nottingham.ac.uk
the University of Nottingham

Cheng-Gang Xu - cgxu@ckgsb.edu.cn
Cheung Kong Graduate School of Business and CEPR

Xiyi Yang - yangxiyi@gmail.com
ShanghaiTech University

Acknowledgements
Chenggang Xu acknowledges the research grants from the CKGSB; and the support and hospitality of the Brevan Howard Finance Centre at the Imperial College and the Institute of Global Affairs at the LSE.

Geographic Clustering and Resource Reallocation Across Firms in Chinese Industries

Di Guo, Brunel University London
Kun Jiang, The University of Nottingham
Chenggang Xu, CKGSB, HKU
Xiyi Yang, ShanghaiTech University

Abstract: We examine the effects of China's industrial clustering on resource reallocation efficiency across firms. Based on our county-industry level DBI index panel, we find that industrial clustering significantly increases local industries' productivity by lifting the average firm productivity and reallocating resources from less to more productive firms. Moreover, we find major mechanisms through which resource reallocation is improved within clusters: (i) clusters facilitate higher entry rates and exit rates; and (ii) within clusters' environment the dispersion of individual firm's markup is significantly reduced, indicating intensified local competition within clusters. The identification issues are carefully addressed by instrumental variable (IV) regressions.

Keywords: Industrial Cluster, Productivity Growth, Resource Reallocation, Competition
JEL Classification: D2, H7, L1, O1, R1, R3

1.Introduction

The fast growth of the Chinese economy under its piculiar institutions since the 1980s has attacted intensive interests from scholars, who attempt to explore the relationship between the growth and institutions (Xu, 2011). The emergence of industrial clusters in numerous small towns in the recent three decades has been recognized as one of the major forces driving China's growth. The organization and coordination of entrepreneurial firms within clusters were identified as an insitututional innovation to overcome institutional impediments (Long and Zhang, 2011; Xu, 2011; Guo et al., 2020). At the same time, ersource misallocation that negatively affects China's productivity growth has been evident in recent studies (Brandt et al., 2012; Brandt et al., 2013; Hsieh and Klenow, 2009). Taking the United States as a benchmark, Hsieh and Klenow (2009) discover that China could have boosted its total factor productivity (TFP) by 2% per year between 1998 and 2005 if capital and labor were reallocated to the extent observed in the United States. Brandt et al. (2013) estimate that, on average, for the period between 1985 and 2007, the distortions in factor allocation reduced non-agricultural TFP in China by at least 20%.

How clustering affects resource allocation and productivity, however, is not studied in the literature. Does clustering help to mitigate resource misallocation and enhance productivity within clusters? If so, what are the mechanisms through which such effects work on-site?

To address these questions, in this paper, based on the density-based index (DBI) approach (a la Guo et al., 2020), which measures industrial clusters in China, we create a county-level industrial clustering dataset. The reason we adopt the DBI approach is because measuring clustering in China is a challenging issue. The problem is caused by the fact that most of the production factors (e.g., labor, capital, and land) are not freely mobile in China. All standard approaches of measuring clusters are based on the assumption of free mobility of factors. By defining clustering and the characteristics of clusters based on the density of firms of each industry within a geographical location, the DBI measurement captures the distinctive features of the industrial clusters created and developed under the institutional restrictions in China. It reduces the noise in the data created by industrial agglomerations of specialized large state-owned enterprises (SOEs). We construct a panel of county-industry level DBI cluster indices that measure the existence and strength of industrial clusters based on firm-level data from the Above-Scale Industrial Firm Panel (ASIFP)[1] between 1998 and 2007.

With the DBI county-industry panel data, we apply Olley and Pakes' (1996) approach to calculate the aggregate TFP for all county-industry pairs. We then decompose the aggregate TFP into firm-level average TFP and reallocation TFP. Our panel and IV estimations show that industrial clustering not only increases the aggregate and average TFP of firms but also improves reallocation efficiency across firms within the clusters. The existence of an industrial cluster in any county-industry is associated with a 2% increase of reallocation TFP per year between 1998 and 2007, indicating that industrial clustering can explain 8.5% of the increase in reallocation TFP in county-industries of China (Table 3)[2].

Moreover, we provide evidence on the mechanisms through which clustering mitigates resource misallocation. We find firm entry and exit are much more active in clusters than

---

[1] ASIFP is composed of virtually all manufacturing firms in China with annual sales of RMB 5 million (US$ 750,000) or more between 1998 and 2007. The database provides detailed financial information and other firm-specific information, including location, industry, age, and ownership structure.

[2] Ended with 2007, findings from our panel data are complement with the observations that China's resource misallocation problem was alleviated before 2008 (Bai, Hsieh and Song, 2016; Song and Xiong, 2018) in the sense that we provide some mechanisms. But our data does not allow us to explore why the situation was worsened after.

outside clusters, and firm markup dispersion is significantly reduced within clusters. These findings indicate local competition is intensified within clusters, which reduces resource misallocation across individual firms.

This study is the first, which links the literature of agglomeration to the literature on resource allocation and productivity. It fills the existing knowledge gaps in several aspects. First, our discoveries contribute to the literature on economic geography and urban economics by estimating the reallocation of resources across heterogeneous firms within a cluster and its impact on productivity. Existing theories of agglomeration economics focus on the "trinity" of labor pooling, lower transportation costs, and information spillover as an explanation that firm-level productivity can be improved by clustering many firms together (Marshall, 1890; Jacobs, 1969; Krugman, 1991; Rotemberg and Saloner, 2000; Combes and Duranton, 2006; Ellison et al., 2010). Empirical examinations generally find positive effects of agglomeration on firm productivity and the growth of the local economy.[3] Yet, except for a few studies which emphasize that the effects of agglomeration would vary depending on geographic distance (Rosenthal and Strange, 2003; Arzaghi and Henderson, 2008) and the maturity of the industries (Henderson et al., 1995; Rosenthal and Strange, 2003), this literature generally assumes that agglomeration effects are homogeneous to the firms within a cluster. Our study, for the first time, examines how clustering improves resource reallocation across firms with heterogeneous productivity levels and consequently affects the aggregate productivity of a

[3] Some estimations find positive effects of collocation of firms from diverse industries within a locality (e.g. Feldman and Audretsch, 1999; Glaeser et al., 1992), favoring Jacobs (1969) urban diversity theory. Some others discover positive effects of regional specialization (e.g. Cingano and Schivardi, 2004; Dekle, 2002; Delgado et al., 2014; Jofre-Monseny, 2005; Rosenthal and Strange, 2003; van Oort and Stam, 2006), favoring the claim of Marshall-Arrow-Romer (MAR) model (Marshall, 1890; Arrow, 1962; Jacobs, 1969). At the same time, some studies find evidence for both the Jacobs and MAR externalities (Henderson et al., 1995; Rosenthal and Strange, 2003).

locality. Our findings that the effects of clustering are attributed to the intensified competition among firms within the clusters also complements the arguments of Porter (1990).

Second, our findings contribute to the literature on economic development and new institutional economics. Studies have documented a large, persistent and ubiquitous degree of productivity dispersion across production units, emphasizing the role of resource reallocation across firms in explaining aggregate productivity growth (Foster et al., 2001; Melitz, 2003; Bartelsman et al., 2009; Collard-Wexler and Loecker, 2014). Specifically, scholars suggest that the low aggregate TFP in developing countries is mainly due to the micro-level resource misallocation (Caselli and Coleman II, 2001; Banerjee and Duflo 2005; Restuccia and Rogerson, 2008; Gancia and Zilibotti 2009; Hsieh and Klenow 2009). In the case of China, the development of clusters is a result of institutional evolution during the post-Mao reforms. By providing the first evidence on the relationship between clustering and resource reallocation at the micro-level, this study sheds new light on the relationship between institutions and resource misallocation in China, the largest developing economy in the world. This study suggests that cluster-based production, at least in the context of China, may alleviate the problem of resource misallocation for firms operating within clusters.

The rest of the paper is organized as follows. In the next section, we discuss the institutional features of industrial clustering in China and review the relevant literature. In section 3, we discuss the data and samples and introduce how we construct the DBI measurements of clusters. Section 4 presents the empirical findings on clustering and aggregate, average as well as resource reallocation productivities and addresses the identification concerns using IV regression. In section 5, we examine the mechanisms through which clustering affects the resource reallocation efficiency with the focus on product market competition. Section 6 concludes this study.

2. Clustering of Chinese industries, firm productivity, and resource reallocation

2.1 Clustering in China under institutional constraints

The clustering of firms within a geographic location is a ubiquitous phenomenon that has been studied by economists and geographers for more than a century (Marshall, 1890; Weber, 1929; Fujita, Krugman and Venables, 1999; Ellison, Glaeser and Kerr, 2010). A critical condition for "clustering" to happen in market economies is factor mobility: labor and capital are mobile, and the land is freely tradable in the market. Under this condition, the market prices of mobile factor inputs will affect firms' co-location decisions that are essential to forming clusters. The development path of clusters in China, however, is significantly different from that in free-market economies. Instead, it has been a consequence of joint efforts of entrepreneurs and local government in overcoming institutional restrictions on factor mobility.

First, peasants, individually or collectively, are not allowed by law to trade "their" land for non-agricultural purposes. According to the Constitution of China, urban land is state-owned, whereas rural land is collectively-owned at the village level. Nationalization is required as the first legal step for trading collectively-owned rural land for non-agricultural use. 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).[4] Since the late 1990s, when political and legal restrictions to private ownership were gradually relaxed, many TVEs have become privatized (Xu, 2011). Many of the clusters nowadays started from privatized TVEs or their spin-offs (Huang et al., 2008). However, there is no change in the property rights of the land. Peasants who intend to use their rural land for industrial purposes, therefore, normally set up industrial firms on the land within their villages or towns. As a result, most rural enterprises are owned and set up by local peasants and cannot be relocated easily.

Second, the *Hukou* system continues to restrict labor mobility, particularly the movement of peasants from rural to urban areas. *Hukou* is a household registration system that officially identifies a person as a resident of a specific area, and the social welfare the person may be entitled. A peasant who seeks to move from a rural to an urban area and take up a non-agricultural job should gain the approval of various bureaucracies. Meanwhile, people who work outside the geographical area of their *Hukou* are unqualified for local social welfares, including housing, health care, education benefits, and pensions (Au and Henderson, 2006). Although the *Hukou* system has been relaxed over time that peasant migrants are allowed to work in cities as de-facto lower-class citizens, moving businesses to urban areas remain extremely difficult for most rural entrepreneurs.

Third, the capital market in China is highly underdeveloped and is particularly biased against lending to private enterprises (Allen et al., 2005). Although the share of the private sector on the total national GDP soared to 50% in 2009, the share of the short-term bank loans issued

---

[4] Land ownership restriction was somewhat relaxed in the recent 15 years, such that non-local entrepreneurs can lease a piece of "collectively owned" land to develop rural industrial firms by recruiting local peasants who collectively "own" the land. Nevertheless, developing real estate for urban residences before nationalization is strictly forbidden by the constitution.

to the private sector was only 4.9% of the national total (Guo et al., 2014).

Under the above-mentioned institutional constraints, the development and features of industrial clusters in China diametrically differ from the concepts of "clustering" or "geographical agglomeration" defined in existing studies. First, industrial clusters in China tend to be defined by administrative boundaries. The firms comprising the clusters are usually owned and set up by local residents who cannot easily move their business elsewhere. Second, under strong financial constraints and other constraints on factor mobility, firms are usually very small in size and highly specialized. Specifically, production processes, which are usually integrated within a single firm in developed countries, are segmented into many small "firms," each of them narrowly specialized in one process. Suppliers, manufacturers, and merchants are coordinated and organized in a dynamic network (Huang et al. 2008; Long and Zhang, 2011).

2.2 Clustering, firm productivity and resource reallocation in China

A central idea of the economics of agglomeration is that firms can enjoy the local scale economies from co-locating with each other and thereby increase productivity. Different explanations are provided for the sources of advantages of agglomeration, mainly 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 the same or similar industries clustered in a region can enjoy the advantages of knowledge spillovers from each other, reduce transportation costs of customer-supplier interactions, and benefit a large common labor pool. On the other hand, Jacobs (1969) highlights the benefits of urban diversity. The theory of Jacobs argues that knowledge exchanges across firms in diverse industries co-located in urban cities create knowledge spillover externalities. The third theory is proposed by Porter (1990), who argues that the advantage of agglomeration comes from the strong competition of firms in a locality, which provides significant incentives for firms to innovate,which in turn accelerates the rate of technical progress and hence of productivity growth. Porter (1990) shares with the MAR model by emphasizing the benefits of regional specialization while he favors Jacobs in highlighting the positive impacts of local competition on knowledge spillover.

Empirically, many studies evident positive effects of agglomeration on local economic growth and firm productivity. Yet, explanations vary regarding which kind of externalities matters. Several studies find positive effects of agglomeration of firms from diverse industries on the growth of employment, wage, and firm productivity (e.g., Feldman and Audretsch, 1999; Glaeser et al., 1992), supporting the argument of Jacobs externalities. Some other studies, however, support the claims of MAR model and evident positive effects of regional specialization (e.g., Cingano and Schivardi, 2004; Dekle, 2002; Delgado et al., 2014; Jofre-Monseny, 2005; Rosenthal and Strange, 2003; van Oort and Stam, 2006). At the same time, also some studies find evidence for both the Jacobs and MAR externalities, depending on the maturity of the industries (Henderson et al., 1995). Yet, using cross-country panel data for 70 countries, Henderson (2003) failed to observe growth-promoting effects from agglomeration in any means. Existing studies have substantially improved our understanding of the economics of agglomeration. However, most of the studies assume (often as an implicit assumption) that agglomeration effects are homogeneous to the firms within a cluster and industry except for a few studies which evident the decay of agglomeration effects over geographic distance (Rosenthal and Strange, 2003; Arzaghi and Henderson, 2008).

In this study, we attempt to examine the interactions between clustering and firm productivity in China from a new angle. That is, we look at how the clustering affects the resource reallocation among firms and thereby improves the aggregate productivity in general. The dispersion in firm productivity within the same industry or the same market is well documented. A growing literature has emphasized the role of resource reallocation across firms in explaining aggregate productivity growth (Foster et al., 2001; Melitz, 2003; Bartelsman et al., 2009). Some suggest that aggregate productivity can rise not only because the firms on average become more productive (usually because of the upgrades in the technology, management or investment in R&D) but also due to shifts in production factors from less to more productive firms (Restuccia and Rogerson, 2008; Collard-Wexler and Loecker, 2014). Institutional changes such as deregulation or trade liberalization that lead to the exit of less productive firms or the expansion of more productive firms can improve aggregate productivity (Pavcnik, 2002; Melitz, 2003; Schmitz, 2005; Bustos, 2011; Bernard et al., 2009). At the same time, financial frictions may substantially reducethe level of TFP, output, and consumption (Buera, Kaboski, and Shin 2011; Buera and Shin 2013; Caselli and Gennaioli 2013; Midrigan and Xu, 2014; Gopinath et al., 2017). On the other hand, in the presence of institutional distortions such as market imperfections, monopoly power, the lack of protection of property rights as well as discretionary provision of production factors, highly productive firm may not have sufficient access to resources, and such restriction to further development of these firms could lower the aggregate TFP of the economy (Banerjee and Duflo, 2005; Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009a; Acemoglu et al., 2018).

In the case of China, as mentioned-above substantial production factors are not allocated through the market. Evidence of negative effects of resource misallocation on aggregate TFP, and the connection of such resource misallocation to institutional problems are discovered in the literature (Brandt et al., 2012; Hsieh and Klenow, 2009; Adamopoulos et al., 2017). For instance, Hsieh and Klenow (2009) find that productive firms are much smaller in China than they would be in an undistorted economy. At the same time, state-owned enterprises (SOEs) are much larger than they should be by the standard of efficiency. Brandt et al. (2012) evident that the differences in productivities between the entering and exiting-firms explain a substantial part of the aggregate TFP growth in China between 1998 and 2005. Additionally, Adamopoulos et al. (2017) evident that eliminating resource misallocation because of the restrictions on land ownership and labor mobility in rural China may have increased agricultural productivity by 1.84 times. Finally, Song et al. (2011) suggest that entrepreneurial firms use more productive technologies to overcome the problems related to the imperfection of the financial market. Such results reflect a significant linkage between institutional distortions and resource misallocation in China.

The development of industrial clusters in China is not only an economic geography phenomenon but also an institutional arrangement. As we have discussed, the organization of production within clusters is designed by entrepreneurs and local governments to overcome the institutional restrictions over the mobility of production factors. Specifically, clustering deepens the division of labor in the production process. It makes it possible for small entrepreneurial firms to enter the market by focusing on a narrowly defined stage of production. These highly specialized entrepreneurial firms closely co-ordinate alongside the value chain within the clusters. With such division of labor, the capital and technical barriers to entry are lowered, resulting in increased competition within the clusters (Xu and Zhang, 2009; Long and Zhang, 2011). We, therefore, expect to observe the positive effects of clustering on aggregate TFP of a locality. More importantly, we suggest that such improved aggregate TFP is accredited to the improved resource reallocation among firms within the industrial clusters in China. In particular, we expect that the intensified competition serves as a major channel through which the clustering improves resource reallocation within the clusters.

3. Data and sample

Our primary dataset is the ASIFP in China from 1998 to 2007. This dataset provides detailed firm-level information including the industry, location, age, size, ownership, and financial information of all SOEs and non-SOEs with annual sales of 5 million RMB or above. Admittedly, this dataset would miss smaller firms, almost entirely non-state firms with annual sales below 5 million RMB. Compared with Economic Census Data, as of 2004, the enterprises covered by ASIFP account for 90% of the total sales of all industries in China[5]. Furthermore, 84% of the firms in ASIFP are officially labeled as small enterprises, defined by no more than 300 employees.

[5] In the first Chinese Economic Census conducted in 2004, the amount of the total sales for all industrial firms was RMB 218 billion, whereas that of the total sales for all ASIFP firms was RMB 196 billion.

A key dependent variable in our study is county-industry level resource reallocation efficiency, which is measured for each industry located in each county. To compute reallocation efficiency, we first calculate the TFP of each firm within each county-industry. We use three TFP measures to ensure the robustness of the results. The first measure TFP_ols_{it} is the ordinary least square (OLS) regression residual from a log-linear transformation of the general Cobb-Douglas production function with year fixed effects and industry fixed effects. The OLS approach considers only tangible inputs, while ignores unobservable shocks, and assumes that all types of inputs are exogenous and hence have no correlation with the error term, that is, the computed TFP itself. To account for these shortcomings, we also calculate firm-level TFP following Olley and Pakes (1996), which is a semi-parametric method to account for both the unobservable production shocks and the non-random sample selection. Specifically, we calculate TFP_op1_{it} with industry fixed effects, and TFP_op2_{it} with year fixed effects and industry fixed effects. The details of the TFP calculation are summarized in Appendix 1.

The resource reallocation efficiency in each county-industry is obtained following a standard decomposition method of Olley and Pakes (1996). Concretely, the county-industry TFP for industry *j* county *k* at time *t*, tfp_{jkt}^{AGG}, is calculated as the sum of each firm *i*'s TFP in the county-industry, tfp_{ijkt}, weighted by the market share of this firm, $share_{ijkt}$. Olley and Pakes (1996) show that aggregate TFP can be decomposed in the following way:

$$tfp_{jkt}^{AGG} = \sum_{i}^{n} tfp_{ijkt} * share_{ijkt} = \overline{tfp_{jkt}} + \sum_{i}^{n} [tfp_{ijkt} - \overline{tfp_{jkt}}] * [share_{ijkt} - \overline{share_{jkt}}]$$

$$= tfp_{jkt}^{AVG} + tfp_{jkt}^{RAL}, \quad (1)$$

where $\overline{tfp_{jkt}}$ and $\overline{share_{jkt}}$ are respectively un-weighted average firm-level TFP and average firm market share in industry *j*, county *k*, and year *t*. The first component, tfp_{jkt}^{AVG}, is the un-weighted average firm-level TFP of the county-industry. The second component, tfp_{jkt}^{RAL}, measures the covariance between firm productivity and market share. Changes in the latter measure represent a reallocation of market share among firms of different productivity levels: a higher level of tfp_{jkt}^{RAL} would represent a higher level of resource reallocation efficiency.

Our key explanatory variable is the existence and strength of the DBI cluster in industry *j,* county *k,* and year *t*. As discussed in Guo et al. (2020), employing standard regional specialization or inter-connectedness measurements to identify industrial clusters in China is not the most suitable method due to institutional constraints on factor mobility and location decisions of firms in China. The onset of the post-Mao reform and at the early stages of the reform, all the firms were owned or controlled by the state or local governments; thus, governments make decisions on firms' locations. The situation was changed gradually, but the legacy is substantial. The concentration of heavy industries in certain regions of China was mostly driven by political concerns. Regions with giant SOEs are likely to be highly specialized when measured by standard clustering measurements such as Herfindahl-Hirschman Index (HHI), Gini coefficient, Krugman Dissimilarity Index (KDI), or location quotient (LQ). Therefore, clusters identified by applying standard indices to China are often located in regions dominated by giant SOEs.[6]

As discussed previously, the development of industrial clusters in China is characterized by the clustering of a large number of small and medium-sized firms within a region, implying

---

[6] When applying standard indices to measure clustering in China, Xinjiang, Shanxi and Qinghai are the provinces with the highest HHI, Gini, KDI, or LQ scores (Guo et al., 2020). These regions have concentration of SOEs, underdevelopment of entrepreneurial firms, and lower development level.

that the density of firms in the industry within a locality is one of the most important features of entrepreneurial clustering in China. Hence, we apply DBI (a la Guo et al., 2020) to measure clusters in China. The DBI counts the number of firms in the same industry within a county. Specifically, we define a county to have an industrial cluster of a particular industry if the county is among the top α percentile of all counties regarding firm density for that industry, and we assign α = 5.[7] We then construct a dummy variable $Cluster_{jkt}$, which equals 1 if firms of industry *j* have formed a cluster in county *k* in year *t*, and 0 otherwise. Moreover, we measure the strength of each cluster based on its relative contribution to the national total industrial output or establishment number, following Guo et al. (2020).[8]

Table 1a reports the summary statistics of DBI clusters. Each year there were about 2,000 industrial clusters in all the counties in China, counted for 5% of the county-industry observations (given α=5). These clusters consisted of more than 30% of the manufacturing firms, contributed to 35-40% of the national industrial output from 1998-2007.

Besides, to define whether a county-industry has formed an industrial cluster or not, we further measure the strength of each cluster based on its relative contribution to the national total industrial output or total establishment number. When measuring cluster strength based on industrial output, we first calculate the contribution of each cluster of industry *j* in county *k* to the national total industrial output of industry *j* at time *t* by $S\_V_{jkt} = \frac{Output_{jkt}}{Output_{jt}}$, which is the percentage. Based on S_V_{jkt} we can distinguish weak clusters versus strong clusters. Specifically, we construct a categorical variable $Strength\_V_{jkt}$. It equals 0 for non-clusters if firms from industry *j* have not formed a cluster in county *k* in time *t*. It equals 1 for clusters

[7] In the remaining part of the paper, we simply call an α-industrial cluster as a cluster. For testing robustness, we also try other α values, such as 3 or 10, and our main results are not affected by the choice of α.
[8] For detailed discussions on the construction of the clustering measurements, please refer Guo et al. (2020).

with below-median S_V_{jkt} compared with other clusters from the same industry *j* at time *t*, and equals 2 for clusters with exact or above median S_V_{jkt}.

Similarly, when measuring cluster strength based on total establishment number, we define $Strength\_E_{jkt}$, which equals 0 if firms from industry *j* have not formed a cluster in county *k* in time *t*. It equals 1 for clusters with below-median S_E_{jkt} compared with other clusters from the same industry *j* at time *t*, where $S\_E_{jkt} = \frac{Establishment_{jkt}}{Establishment_{jt}}$ is the contribution of the cluster in county *k* of industry *j* to the total number of firms of industry *j* at time *t*. $Strength\_E_{jkt}$ equals 2 for clusters with exact or above median S_E_{jkt}. The summary statistics of the constructed cluster variables are reported in Table 1b.

When estimating the effect of clustering on county-industry TFP, we control for a set of industrial firm characteristics including *Average firm age*, *Average firm size*, *Average firm state-ownership*, and *Average firm leverage* of the firms within each county-industry. We also control for the size effect of the local industry using *County-industry employment*, which is the total number of employees for each county-industry. We further include in our regressions *County per capita GDP* and *County total GDP* to control for the effects of regional development level and regional economic size. These data are from the China Socio-Economic Development Statistical Database. Finally, in the panel estimations, we also include year dummies and county×industry dummies to control for time trends and time-invariant heterogeneities across county-industries. Detailed definitions of our variables are summarized in Appendix 2.

Our sample covers firms in more than 2,800 counties and 39 2-digit industries in China from 1998-2007. During our sample period, some counties changed their names or judiciary boundaries. 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). We tract the four-digit industry codes that have become either more disaggregated or more aggregated after 2002 and use the more aggregated codes to group the industries from 1998 to 2007.

Table 2 reports the summary statistics of dependent variables and control variables for clusters and non-clustered county-industries (which we call non-clusters). On average , clusters appear significantly higher in aggregate, average, and reallocation TFP than non-clusters. Furthermore, firms in clusters tend to be younger, larger in size, have less state-ownership, and lower leverage ratio than firms outside of clusters. Finally, counties with clusters are more likely to have higher per capita GDP and total GDP than counties without clusters.

4. Findings on industrial clustering and resource reallocation across firms

In the subsequent sections, we estimate clustering effects on aggregate, average, and cross-firm resource reallocation TFP, after controlling for various industrial and regional characteristics. The identification issue will be addressed by IV regressions. And our IVs for clusters are retail activities, scarcity of arable land, and abundance of mineral products.

4.1 Industrial clustering and reallocation TFP

The following equation (2) is our baseline regression model. We use it to estimate the effect of clustering on local-industry productivities.

$$TFP_{jkt} = \alpha + \beta Cluster_{jkt} + \boldsymbol{\delta Z_{jkt}} + \theta_{jk} + \theta_t + \varepsilon_{jkt}, \quad (2)$$

TFP_{jkt} is measured by tfp_{jkt}^{AGG}, tfp_{jkt}^{AVG}, or tfp_{jkt}^{RAL}, which are decomposed county-industry level TFP elements; $Cluster_{jkt}$ is the DBI cluster dummy variable that equals 1 if firms from industry *j* have formed an industrial cluster in county *k* in year *t*, or 0 otherwise. The decomposed county-industry level TFP elements are calculated with equation (1) based on firm-level TFP, which is measured by *TFP_ols*, *TFP_op1*, and *TFP_op2*, respectively. $\boldsymbol{Z_{jkt}}$ is a set of control variables that include *Average firm age* and *Average firm size*, *Average firm state-ownership*, *Average firm leverage*, *Local-industry employment*, *County per capita GDP*, and *County total GDP*. θ_{jk} and θ_t are county×industry and year dummies, respectively, and we cluster the standard errors at the county-industry level.

The baseline regression results are reported in Table 3. The cluster variable $Cluster_{j,k,t}$ is always significantly associated with higher county-industry aggregate productivity tfp_{jkt}^{AGG} regardless of how firm-level TFP is measured. The increased aggregate productivity in clusters seems to be derived from higher average firm productivity tfp_{jkt}^{AVG}, and more efficient resource reallocation across firms within clusters tfp_{jkt}^{RAL}. In particular, except for the case of *TFP_ols* (TFP is estimated by the OLS method), $Cluster_{j,k,t}$ is significantly associated with higher reallocation productivity tfp_{jkt}^{RAL}. One possible mechanism behind this phenomenon is expansions of higher-productivity firms and exits of lower-productivity firms in clusters, which we will further study in the next section. Model (6) & (9) indicate an about 2% increase of reallocation TFP per year between 1998 and 2007. Given the mean of the reallocation TFP (measured by TFP_op1 or TFP_op2) is approximately 0.235 during this period, industrial clustering alone can explain 8.5% of the increase in reallocation TFP in county-industries of China.

Tables 4a and 4b report the effects of clusters with different strengths on local-industry productivities. As defined in section 3, we construct two categorical variables to differentiate weak and strong clusters based on clusters' output value or establishment number. The first variable, $Strength\_V_{jkt}$, equals 0 for non-clustered county-industries. It equals 1 for clusters with a below-median contribution to national total industrial output compared with other clusters from the same industry and equals 2 for clusters with the median or above-median contribution to national total industrial output.

As shown in Table 4a, for aggregate productivity and average productivity, no matter how TFP is measured, the coefficients of $Strength\_V_{jkt}\_1$ and $Strength\_V_{jkt}\_2$ are always positive and significant. Moreover, the magnitudes of the coefficients for $Strength\_V_{jkt}\_2$ are always larger than those t of $Strength\_V_{jkt}\_1$, and are also larger than those of $Cluster_{jkt}$ (see Table 3). For instance, for aggregate productivity calculated using the OLS method, the coefficient of $Cluster_{jkt}$ (not differentiating strength) is about 0.215. It is larger than the coefficient of weak clusters ($Strength\_V_{jkt}\_1$, β=0.191), but smaller than the coefficient of strong clusters ($Strength\_V_{jkt}\_2$, β=0.269). As for reallocation productivity, no matter how TFP is measured, the coefficients of $Strength\_V_{jkt}\_1$ are always insignificant. On the contrary, the coefficients of strong clusters, $Strength\_V_{jkt}\_2$, are always positive and highly significant. When productivity is measured by TFP_op1 or TFP_op2, the coefficients of $Strength\_V_{jkt}\_2$ are about twice as big as those of $Cluster_{jkt}$ (see Table 3). As shown in Model (6) & (9), the coefficient of $Strength\_V_{jkt}\_2$ is about 0.04. Given that the mean value of reallocation TFP is about 0.235 during our sample period, the presence of strong clusters can explain 17% of the increase in reallocation TFP in the county-industries of China.

Table 4b reports similar results when cluster strength is measured by its contribution to the national total establishment number. The variable of interest, $Strength\_E_{jkt}$, equals 0 for non-clustered county-industries. It equals 1 for weak clusters and equals 2 for strong clusters. Similarly, for aggregate productivity and average productivity, no matter how TFP is measured, the coefficients of $Strength\_E_{jkt}\_1$ and $Strength\_E_{jkt}\_2$ are always positive and significant. The magnitudes of the coefficients of $Strength\_E_{jkt}\_2$ is larger than those of $Cluster_{jkt}$ (see Table 3). The magnitudes of the coefficients of $Strength\_E_{jkt}\_1$ are the smallest. For reallocation productivity, weak clusters do not seem to have a significant effect. Strong clusters, on the other hand, have a positive and significant effect and the coefficients of $Strength\_E_{jkt}\_2$ range from 0.021 to 0.045, depending on how TFP is measured.

Taking together, the results in Table 3 and Table 4a&b imply that industrial clustering, especially clustering with strong presence of output and firm establishment, is associated with higher aggregate productivity, and such increase in productivity not only comes from the lift of average firm productivity but also the improvement of resource reallocation across firms within the local industry. The significant correlations between clustering and improved productivities are very interesting, but that alone is not good enough for inferring causality between clustering and productivity improvement. Alternative explanations for the baseline estimations remain. For example, one could argue that the existence or the strength of the clustering are a result rather than the cause of the productivity improvement, although such possibility is slim in the context of China given the location choices of firms are constrained by institutions generally. Moreover, omitted variables, such as the local entrepreneurship culture or the management or production skills of the local people, may have contributed to the improved productivity, and the rise of industrial clusters simultaneously. To address such concerns, we employ two-stage estimations using two sets of instrument variables (IVs) to identify the effect of clustering on aggregate productivity and resource reallocation productivity, respectively.

Our first IV, *County retail sales ratio,* is the ratio of retail sales of consumer goods to total GDP in each county. The retail sales of consumer goods are defined as the sales value of physical commodities sold by firms to individuals and organizations for consumption, i.e., not for production or business purposes. These data are from the China Socio-Economic Development Statistical Database. We believe local retail activities are positively associated with clustering development for the following reason: industrial clusters are usually national, if not international, production centers of certain commodities. Hence the production capacity of clusters usually far exceeds the local demand, and the majority of the products are sold through retail or wholesale[9] to other places. We, therefore, expect that the higher the level of retail sales a county has, the more likely that the county has formed industrial clusters. However, no direct links shall exist between a county's retail activities and the individual firm productivities or the resource allocation across firms. Therefore, this instrument is likely to satisfy both the relevancy and exogeneity conditions.

Our second IV, *Provincial mineral output ratio*, is the annual mineral output over GDP ratio at the provincial level. This data is obtained from the China Mineral Yearbooks. We believe

[9] Much of the wholesales are sold to individuals, knowns as the "individual enterprises" in China that employ less than 8 people. These wholesales, are therefore, included in the retail activities as well.

this instrumental variable satisfies the conditions of relevancy and exogeneity. On the one hand, we expect mineral resources in each province to be negatively associated with the development of industrial clusters. This anticipation is based on an observation that mine-rich regions are often dominated by large companies due to large fixed investments and high returns to scale in the mining industry. Thus, smaller businesses are often crowded out, and entrepreneurship is often depressed (Chinitz, 1961)[10] as supported by empirical evidence (e.g., Rosenthal and Strange, 2003; Glaeser et al., 2015). On the other hand, mineral resources should be exogenous, particularly to the resource reallocation process between firms, because the mine-richness of a region is geologically determined.

The results of the two-stage estimations for aggregate and reallocation productivity are reported in Table 5. Panel A presents the results for the first-stage regressions. Consistent with our expectation, *County retail sales ratio* is significantly and positively correlated with industrial clustering, and the coefficient of *Provincial mineral output ratio* is significant and negative. The under-identification tests and weak-identification tests reject the hypotheses that the instruments are irrelevant or weak. Furthermore, except for the regression on aggregate productivity using *TFP_op2*, the Sargan tests indicate that we cannot reject the hypothesis that the instruments are jointly exogenous, confirming the exogeneity of the two instruments we used to identify the effects of clustering on local industry's aggregate and reallocation productivity. The second-stage regression results are reported in Panel B. In all

---

[10] 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 companies in regions with 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 low costs, which consequently depresses the local supply development of small entrepreneurial firms.

the regressions, the coefficients of the instrumented cluster dummy are always positive and significant, which suggests that industrial clustering can indeed increase the aggregate productivity and resource reallocation efficiency across firms of the local-industry.

Regarding average firm productivity, as it may be related to innovation and technology, one might challenge the exogeneity of the *Provincial mineral output ratio* as an IV. We, therefore, replace it with *City per capita arable land*, which is the ratio of the arable land over the total population in each city in a given year. . This data is from the China Socio-Economic Development Statistical Database. We expect this instrument to be negatively correlated with clusters, as the scarcity of natural resources in agriculture may drive the labor force from agriculture to industry. Anecdotally, most coastal regions with industrial clusters have dense rural populations. For instance, located in the mountainous south-eastern part of Zhejiang Province, per capita arable land in Wenzhou is about one-third of the national average, only 0.52 mu per person (Zhang and Li, 1990). However, Wenzhou has developed several industrial clusters (Huang et al., 2008) and a commodity cluster (Hessler, 2007). Furthermore, this instrument shall be exogenous to the average level of productivity of the firms within the local industry, given it is a proxy for the abundance of agricultural resources.

The results of the IV regressions for average firm productivity are reported in Table 6. The instrument of *County retail sales ratio* is again positive and significant. The coefficient of the second instrument, *City per capita arable land*, is negative and marginally significant. Both the under-identification and weak-identification tests suggest that the two instruments are relevant. Finally, except for the case where average firm productivity is measured by *TFP_op2* (p=0.0821), the statistics from the Sargan test suggest we cannot reject the hypothesis that the two instruments are jointly exogenous to the regressions' residuals, and hence the instruments satisfy the exclusion criterion. The second-stage results are presented in Panel B. It clearly shows that the positive effect of clustering on average firm productivity is robust regardless of how TFP is measured.

In sum, the results of the IV regressions are consistent with the baseline regressions. Thus the causal relationships between industrial clustering and increased aggregate, average, and reallocation productivity of the local industries are confirmed.

4.2 Additional robustness checks

In this subsection, we conduct some additional robustness checks to rule out alternative explanations for the estimation results. Industrial profile varies across locations that some regions may have a concentration of heavy industries, and some other regions may have a concentration of high-tech industries. The productivities of different industries may vary a lot. As our cluster measurement is the density of firms, one may concern the impacts of regional specialization on the productivity of localities. Regarding this kind of concern, we run a set of regressions, in which we control for the three largest industries in each county. Moreover, we control for regional specialization at the county level defined by standard regional specialization measurement, i.e., location quotient (Glaeser et al., 1992; Porter, 2003). The results in Table 7a and 7b show that with the three largest industries and the location quotient variables controlled, the effects of clustering on the aggregate, average, and resource reallocation TFP stay robust.

Another major concern is the impacts of the megacities on clusters and productivities. It is known that the economic growth in China has been concentrated in the coastal regions, and all Chinese megacities are also located in coastal regions. It might be possible that the clustering effect in improving productivity we discover is driven by the megacities, which are clusters at a much larger scale than those defined in our study. In our baseline and two-stage estimations, we have controlled county fixed effects. However, if the megacity effects overwhelm the county fixed effects, the effects of clustering we have observed from the baseline estimations may have been inflated. To address such concerns, we run two sets of additional regressions. First, we look at the subsample of the counties located outside the megacities (Beijing, Shanghai, Guangzhou, and Shenzhen). Second, we control the megacity effects. As we have controlled the county fixed effects for the estimations, it is unpractical to control the megacity dummy variables. We, therefore, choose to employ the population of the megacities. The estimations, presented in Table 8a and 8b, show that with the effects of megacities controlled, the results we present in the baseline estimations remain as robust.

To summarize, the estimations presented in Tables 3 to 8 confirm that clustering improves aggregate and average productivities of firms; moreover, it reduces resource misallocation across firms within clusters.

5. Mechanism: industrial clustering and local-industry competition

In this section, we explore the mechanisms through which industrial clusters in China alleviate the problem of resource misallocation across firms and hence improve productivity. Previous studies on Chinese clusters suggest that clustering may have reduced entry barriers and, therefore, improved competition within the clusters (Huang et al. 2008; Xu and Zhang, 2009; Long and Zhang, 2011). Specifically, clustering deepens the division of labor in the production process. It makes it possible for small entrepreneurial firms to enter the market by focusing on a narrowly defined stage of production. Yet, previous studies are either based on case studies or cross-sectional data, and competition is not systematically measured. In the following, we systematically evaluate the pro-competitive effects of industrial clustering by investigating its relationship with firm entry and exit patterns, and how it affects firm markup dispersion within the local-industry.

5.1 Industrial clustering and firm entry and exit

To examine the firm entry and exit within clusters, we follow Dunne, Roberts, and Samuelson (1988) to calculate the entry and exit patterns of firms in all county-industries in China. We then compare the statistics between clusters and non-clusters. The entry and exit statistics are defined in the following:

NE_{jkt}= number of firms that enter industry *j* of county *k* between years *t-1* and *t*;

NT_{jkt}= total number of firms in industry *j* of county *k* in year *t*, including firms that enter industry *j* of county *k* between years *t-1* and *t*;

NX_{jkt-1}= number of firms that exit industry *j* of county *k* between years *t-1* and *t*;

QE_{jkt}= total output of firms that enter industry *j* of county *k* between years *t-1* and *t*;

QT_{jkt}= total output of all firms in industry *j* of county *k* in year *t*;

QX_{jkt-1}= total year *t-1* output of firms exiting industry *j* of county *k* between years *t-1* and *t*.

The entry and exit rates of industry *j* in county *k* between year *t-1* and *t* are defined as the following:

$$ER_{jkt} = NE_{jkt}/NT_{jkt-1}$$

$$XR_{jkt-1} = NX_{jkt-1}/NT_{jkt-1}.$$

Moreover, to measure the relative size of the entrants and exiting-firms, we calculate the average size of entering firms relative to incumbents (ERS) and the average size of exiting-firms relative to non-exiting-firms (XRS) as:

$$ERS_{jkt} = \frac{QE_{jkt}/NE_{jkt}}{(QT_{jkt} - QE_{jkt})/(NT_{jkt} - NE_{jkt})}$$

$$XRS_{jkt-1} = \frac{QX_{jkt}/NX_{jkt-1}}{(QT_{jkt-1} - QX_{jkt-1})/(NT_{jkt-1} - NX_{jkt-1})}.$$

Table 9 reports the comparative statistics of firm entry and exit patterns within clusters and non-clusters based on the ASIFP data. The results indicate clearly that firm entry and exit is much more active in clusters than in non-clusters. Moreover, the entry rate in the cluster (ER=0.3118) is more than twice that of non-clusters (ER=0.1429). Furthermore, the exit rate in clusters (XR=0.1366) is significantly higher than that in non-clusters (XR=0.0876). These results suggest a higher competition level within industrial clusters since the firm turnover is significantly higher. Moreover, on average, the larger turnover in clusters seems to be mainly driven by small firms, given that the mean value of ERS and XRS are smaller than one in clusters, and they are much smaller than those in non-clusters.

Table 10 reports the regression results for the effects of clustering on the firm entry and exit of a county. As shown in columns (1) and (2), the existence of clusters is significantly and positively associated with the total number of new entrants and exiting-firms. Meanwhile, columns (3) and (4) show a similar pattern of clustering effects on entry and exit rates of firms. By average, a county with a cluster in a given industry may have 4.2 more new entrants and 1.8 more exiting-firms in the industry comparing to a county without a cluster, after controlling factors such as the average age, size, ownership, and leverage of firms within the same industry in the county as well as the size and development level of the counties. Similarly, by average, a county with a cluster in a given industry have 0.028 (8.97% of the mean) higher entry rate and 0.018 higher (12.6% of the mean) exit rate of firms within the same industry than a county without a cluster in the industry.

Table 11a and 11b show the relationship between the strength of the clusters and the entry and exit of firms. It is clear that the strength of the clusters, no matter measured by the total outputs, or the number firms in the industry, is significantly and positively correlated with the new entrants, exiting-firms, as well as the entry and exit rate of firms. Overall, the results of Table 10 and Table 11 confirm our conjecture that industrial clustering exposes firms to greater competition and therefore facilitates the reshuffling of market shares from less to more productive firms.

Finally, in Tables 10 and 11, the relationship between clusters and the exit of smaller firms is insignificant, although it is positive. However, results shown in those two tables are based on the ASIFP data, which only contains firms with annual sales of 5 million RMB or above. So, the “entry” into the panel data may include cases where an existing firm’s sales grow to exceed 5 million RMB, and the “exit” may include cases where an existing firm’s sales decrease to below 5 million RMB. As a further robustness check, we utilize another firm-level data from the State Administration for Industry and Commerce, which contains information on the establishment date and deregistration date (if applicable) of all the registered firms in China during our sample period from 1998-2007. Using this dataset, we can calculate the total number of newly-established firms and de-registered firms in each county-industry. Meanwhile, the registration database provides information on the registration capital of the firm at the time of incorporation that allows us to estimate the financial situation of the startup firms. However, due to data limitations, we do not have information on the surviving firms during this period. We are unable to calculate firm entry rate, exit rate, or relative size to incumbents. Table 12 reports the OLS results for the effects of clustering on the firm entry and exit, based on the firm (de)registration data. As shown in the table, the existence of clusters in a given industry is significantly and positively associated with the number of new entrants and existing firms in any local-industry between 1998 and 2007. Moreover, the estimations on the registration capital show that the startup capital for firms in a county with a cluster in a given industry is lower than that of firms in a country without a cluster in this industry. Such results further confirm that clustering lowers the entry barriers of firms and thereby intensifies the competition in a locality.

5.2 Industrial clustering and firm markup dispersion within the local-industry

Hsieh and Klenow (2009) and Edmond et al. (2015, 2018) show that markups increase with firm size, which leads to misallocation. In this sub-section, we explore if China's industrial clustering mitigates markups, thus reduces resource misallocation.

The economic discussion of the efficiency costs of markups can be traced back to Lerner (1934), which shows that in a world with markup dispersion, firms with higher markups employ resources at less than optimal levels, while those with lower markups produce more than optimal, resulting in efficiency losses (Opp, Parlour, and Walden, 2014). Some recent papers (Baqaee and Farhi, 2018; Edmond et al., 2018) show that for heterogeneous firms engaging in monopolistic competition[11], in equilibrium, more productive firms will be larger, choose to deal with less elastic demands, and so charge higher markup than less productive firms. Whereas, in a more competitive environment in which resources or market shares are allowed to be reallocated freely from less productive to more productive firms, more productive (and higher markup) firms will produce more, leading to a reduction of their markups. Similarly, lower productivity (and lower markup) firms will produce less, and their markup will increase. Hence, if China's industrial clusters provide a more competitive environment, there should be reduced firm markup dispersion within clusters than in non-clustered local-industries. Furthermore, within clusters, the individual firm markup at higher-quantile should be reduced while that at lower-quantile should be increased[12].

For this purpose, we first look at the relationship between firm size, productivity, and measured markup in our data. Following De Loecker and Warzynski (2012) and Lu and Yu (2015), we use firm sales to measure its size and calculate individual firm markup. Details of the calculation on markup are described in Appendix 3. As shown in Table 13, consistent with Hsieh and Klenow (2009) and Edmond et al. (2015, 2018), the Chinese ASIFP data do feature strong positive correlations ($p<0.001$) among the three variables: firm sales value is positively associated with firm productivity (measured by the three TFP indices), which is in turn positively associated with the markup it charges.

The formal test the effect of industrial clustering on firm markup distribution within county-industries using panel regressions is in the following. We use two measurements of

[11] Edmond et al. (2018) have shown that the same pattern can derive from alternative model of oligopolistic competition among a finite number of heterogeneous firms, as in Atkeson and Burstein (2008).

[12] While more competition reduces firm markup in general, it also reallocates market share to towards more productive firms, and hence the net effect on average markup can be ambiguous (Edmond et al., 2018).

firm markup dispersion to ensure the robustness of our results. The first one is the Theil index ($Theil_{jkt} = \frac{1}{n_{jkt}} \sum_{i=1}^{n_{jkt}} \frac{y_{ijkt}}{\overline{y_{jkt}}} log \frac{y_{ijkt}}{\overline{y_{jkt}}}$), where y_{ijkt} is the markup of firm *i* of industry *j* in county k at year t. $\overline{y_{jkt}}$ is the average firm markup of industry *j* in county *k* at year *t*, and n_{jkt} is the total number of firms of that county-industry in year *t*. The second measure of markup dispersion is the relative mean deviation of each county-industry during our sample period ($RMD_{jkt} = \frac{1}{n_{jkt}} \sum_{i=1}^{n_{jkt}} \left| \frac{y_{ijkt}}{\overline{y_{jkt}}} - 1 \right|$). In addition to investigating the effect of industrial clustering on firm markup dispersion, we also look into markup responses at different quantiles along with the distribution. Specifically, we pin down the firm markup at the 10th, 25th, 50th, 75th, and 90th percentile of each county-industry in every year, and then estimate the effect of clustering on firm markup at different percentiles separately. The estimation model is the following:

$$Y_{jkt} = \alpha + \beta Cluster_{jkt} + \boldsymbol{\delta Z_{jkt}} + \theta_{jk} + \theta_t + \varepsilon_{jkt}, \quad (3)$$

Y_{jkt} is $Theil_{jkt}$, RMD_{jkt}, as well as the firm markup at different percentiles. $Cluster_{jkt}$ is the dummy indicator of clustering in industry *j* of county *k* in year *t*. $\boldsymbol{Z_{jkt}}$, θ_{jk}, and θ_t are the same control variables and fixed effects as defined in section 4.

Table 14 reports the regression results on the effect of clustering on markup dispersion and markup distribution. Columns (1) and (2) demonstrate that firm markup dispersion is significantly lower in clusters than in non-clustered county-industries. Given the mean of $Theil_{jkt}$ and RMD_{jkt} being 0.0065 and 0.047, industrial clustering alone can explain 15.38% and 21.28% of decrease of $Theil_{jkt}$ and RMD_{jkt}, respectively. The rest columns illustrate that the clustering effects on markups of firms vary with sizes. For smaller firms at the lower quintiles (Columns 3 & 4), the clustering effect is significant and positive, implying enlargement of these firms' markups. But for the larger firms at higher quintiles (Columns 6 & 7), the effect is the opposite, i.e., significant and negative, indicating the reduction of their markups. Whereas for the middle-sized firms (Column 5), the clustering effect is insignificant. These findings provide evidence that China's clusters provide a more competitive environment, which reduces the gap of markups between large firms and small firms, and therefore mitigates resource misallocation across firms within clusters.

6. Conclusions

One of the most striking developments in China during the post-Mao reforms is the emergence of numerous specialized industrial clusters in small towns that transformed a large part of the Chinese economy from agriculture to industry. It is an institutional innovation in overcoming institutional impediments, which are overwhelming in China, and it contributed to China's growth substantially (Long and Zhang, 2011; Xu, 2011; Guo et al., 2020). However, associated with all kinds of institutional impediments, resource misallocation is a prevailing problem in the Chinese economy (Hsieh and Klenow, 2009), and it causes severe problems at the national level (Song and Xiong, 2018).

In this paper, we investigate the effect of China's industrial clusters on resource misallocation. Under the institutional constraints, the way the clusters are formed and coordinated distinguishes the Chinese clusters from the common phenomenon of clustering observed in the rest of the world. The development of such clusters reflects the institutional changes during the post-Mao reform. Based on a systematic analysis of county-industry panel data, we find China's clustering has significantly improved the local-industry productivity. And the increased productivity not only comes from higher average firm productivity than that outside of clusters but also comes from more efficient resource reallocation across firms within the clusters than their counterparts outside of clusters.

Besides, we find concrete mechanisms through which China's clusters mitigate resource misallocation problem. There is a higher level of firm turnover in clusters than outside clusters, and the startup capital for firms in a cluster is lower than that outside a cluster. Moreover, markups of firms in clusters are lower than those outside of clusters. Furthermore, markups of firms in clusters have smaller dispersions than that outside of clusters. All of these imply that industrial clustering in China is a more competitive environment than that the rest of the Chinese economy.

References

Acemoglu, Daron, Ufuk Akcigit, Harun Alp, Nicholas Bloom, and William R. Kerr, 2018. Innovation, reallocation and growth. American Economic Review. 2018

Adamopoulos, Tasso, Loren Brandt, Jessica Leight, and Diego Restuccia, 2017. "Misallocation, selection and productivity: A quantitative analysis with panel data from China." No. w23039. National Bureau of Economic Research.

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.

Arzaghi, Mohammad, and Vernon Henderson, 2008. "Networking off Madison Avenue." *Review of Economic Studies* 75(4): 1011-1038.

Atkeson, A. and Burstein, A., 2008. "Pricing-to-market, trade costs, and international relative prices." *American Economic Review*, 98(5): 1998-2031.

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, Chong-En, Chang-Tai Hsieh and Zheng Song, 2016, "The Long Shadow of China's Fiscal Expansion," *Brookings Papers on Economic Activity*, 2016, Fall, 129-165

Banerjee, A.V. and Duflo, E., 2005. "Growth theory through the lens of development economics." *Handbook of Economic Growth*, 1: 473-552.

Baqaee, David Rezza and Emmanuel Farhi, 2018. "Productivity and Misallocation in General Equilibrium." LSE working paper.

Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta. 2009. "Cross-country Differences in Productivity: The Role of Allocation and Selection." *American Economic Review* 103(1):

305-334.

Bernard, A.B., Jensen, J.B., Redding, S.J. and Schott, P.K., 2009. “The margins of US trade.” *American Economic Review*, 99(2): 487-93.

Brandt, Loren, Johannes Van Biesebroeck, Yifan Zhang. 2012. “Creative Accounting or Creative Destruction: Firm Level Productivity Growth in Chinese Manufacturing.” *Journal of Development Economics* 97 (2): 339–351

Brandt, Loren, Trevor Tombe, Xiaodong Zhu. 2013. “Factor Market Distortions Across Time, Space and Sectors in China.” *Review of Economic Dynamics* 16(1): 39–58.

Bustos, P., 2011. “Trade liberalization, exports, and technology upgrading: Evidence on the impact of MERCOSUR on Argentinian firms.” *American economic review*, 101(1): 304-340.

Buera, F. J., Kaboski, J. P., & Shin, Y. (2011). Finance and development: A tale of two sectors. *American economic review*, *101*(5), 1964-2002.

Buera, F. J., & Shin, Y. (2013). Financial frictions and the persistence of history: A quantitative exploration. *Journal of Political Economy*, *121*(2), 221-272.

Caselli, F. and Coleman II, W.J., 2001. “The US structural transformation and regional convergence: A reinterpretation.” *Journal of Political Economy*, 109(3): 584-616.

Caselli, F., & Gennaioli, N. (2013). Dynastic management. *Economic Inquiry*, *51*(1), 971-996.

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

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

Collard-Wexler Allen, and Jan De Loecker, 2014. “Reallocation and Technology: Evidence from the US Steel Industry.” *American Economic Review* 105(1): 131-171.

Combes, Pierre-Philippe, and Giles Duranton, 2006. "Labour Pooling, Labour Poaching, and

Spatial Clustering." *Regional Science and Urban Economics* 36(1): 1-28.

De Loecker, Jan, and Frederic Warzynski, 2012. “Markups and Firm-level Export Status.” *American Economic Review* 102: 2437-2471.

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.

Dunne, Timothy, Mark J. Roberts, and Larry Samuelson, 1988. “Patterns of firm entry and exit in US manufacturing industries.” *The RAND Journal of Economics* 19(4): 495-515.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu, 2015. “Competition, markups, and the gains from international trade.” *American Economic Review* 105(10): 3183-3221.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu, 2018. “How Costly Are Markups?”. National Bureau of Economic Research Working Paper.

Ellison, Glenn, Edward Glaeser, and William Kerr, 2010. “What causes industry agglomeration? Evidence from coagglomeration patterns.” *American Economic Review* 100(3): 1195-1213.

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.

Foster, Lucia, John C. Haltiwanger, and C. J. Krizan, 2001. “Aggregate Productivity Growth. Lessons from Microeconomic Evidence.” In Hulten, Dean, and Harper (eds.), *New Developments in Productivity Analysis*, Chicago: NBER Books.

Fleisher, Belton, Dinghuan Hu, William McGuire, and Xiaobo Zhang, 2010. “The evolution of an industrial cluster in China.” *China Economic Review* 4: 1-14.

Fujita, Masahisa, Paul Krugman, and Anthony Venables, 1999. *The Spatial Economy*.

Cambridge, MA: MIT Press.

Gancia, G. and Zilibotti, F., 2009. “Technological change and the wealth of nations.” *Annu. Rev. Econ.*, 1(1): 93-120.

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

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.

Gopinath, G., Kalemli-Özcan, Ş., Karabarbounis, L., & Villegas-Sanchez, C. (2017). Capital allocation and productivity in South Europe. *The Quarterly Journal of Economics*, *132*(4), 1915-1967.

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.

Guo, Di, Kun Jiang, Chenggang Xu, and Xiyi Yang, 2020. “Industrial Clustering, Growth and Inequality in China.” *Journal of Economic Geography*, 2020.

Henderson, Vernon, 2003. “Marshall’s Scale Economies.” *Journal of Urban Economics* 53(1): 1-28.

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.

Hsieh, Chang-Tai, and Peter L. Klenow, 2009. " Misallocation and Manufacturing TFP in China and India." *The Quarterly Journal of Economics* 124 (4): 1403-1448.

Huang, Zuhui, Xiaobo Zhang, and Yunwei Zhu, 2008. “The Role of Clustering in Rural Industrialization: A Case Study of Wenzhou’s Footwear Industry.” *China Economic Review* 19: 409-420.

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

Jofre-Monseny, J., 2009. “The scope of agglomeration economies: Evidence from Catalonia.” *Papers in Regional Science* 88(3), pp.575-590.

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

Lerner, A. P., 1934. “The Concept of Monopoly and the Measurement of Monopoly Power.” *Review of Economic Studies* 1(3): 157-175.

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

Lu, Yi, and Linhui Yu. 2015. "Trade Liberalization and Markup Dispersion: Evidence from China's WTO Accession." *American Economic Journal: Applied Economics* 7(4): 221-253.

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

Melitz, Marc, 2003. “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity.” *Econometrica* 71(6): 1695-1725.

Midrigan, V., & Xu, D. Y. (2014). Finance and misallocation: Evidence from plant-level data. *American economic review*, *104*(2), 422-58.

Olley, Steven, and Ariel Pakes, 1996. “The Dynamics of Productivity in the Telecommunications Equipment Industry.” *Econometrica* 64(6): 1263-1297.

Opp, Marcus M., Christine A. Parlour, and Johan Walden. 2014. “Markup Cycles, Dynamic Misallocation, and Amplification.” *Journal of Economic Theory* 154: 126-161.

Pavcnik, Nina, 2002. “Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants.”*Review of Economic Studies* 69(1): 245-276.

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.

Restuccia Diego, and Richard Rogerson, 2008. “Policy Distortions and Aggregate Productivity with Heterogeneous Plants.” *Review of Economic Dynamics* 11(4): 707-720.

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.

Rotemberg, Julio, and Garth Saloner, 2000. "Competition and Human Capital Accumulation: A Theory of Interregional Specialization and Trade." *Regional Science and Urban Economics* 30(4): 373-404.

Schmitz, James, 2005. “What Determines Productivity? Lessons from the Dramatic Recovery of the US and Canadian Iron Ore Industries Following Their Early 1980s Crisis.” *Journal of Political Economy* 113(3): 582–625.

Song, Z., Storesletten, K. and Zilibotti, F., 2011. “Growing like China.” *American Economic Review*, 101(1) 196-233.

Van Oort, F.G. and Stam, E., 2006. “Agglomeration economies and entrepreneurship in the ICT industry.” Working Paper.

Song, Zheng and Wei Xiong, 2018, “Risks in China's Financial System,” *Annual Review of Financial Economics*, 2018, vol 10(1), pages 261-286.

Weber, Alfred, 1929. *Theory of the Location of Industries*. Chicago: University of Chicago Press.

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 Wu and Yao (eds.), *Reform and Development in China*, London and New York: Routledge.

Zhang, Renzhou, and Hong Li, 1990. *Wenzhou Model Research*, Beijing: China Social Sciences Publishing House.

Table 1a: Summary statistics of identified industrial clusters and their contribution to the national economy

Year	Numb of county-industries	Numb of industrial clusters	Numb of clustered firms	Numb of non-clustered firms	Ratio of clustered firms	Clusters' contribution to national output	Clusters' contribution to national employment
1998	41899	2024	50,963	98,855	0.3402	0.3838	0.2797
1999	41571	2037	46,960	102,246	0.3147	0.3527	0.2546
2000	40272	1958	46,557	100,655	0.3163	0.3733	0.2821
2001	38712	1896	51,675	103,941	0.3321	0.3633	0.2760
2002	39432	1931	57,447	112,266	0.3385	0.3680	0.2905
2003	40207	1971	65,258	120,621	0.3511	0.3811	0.3124
2004	41996	2118	106,402	160,437	0.3987	0.4035	0.3650
2005	41809	2125	98,907	159,625	0.3826	0.3914	0.3636
2006	43177	2157	111,411	176,058	0.3876	0.3911	0.3748
2007	44175	2213	126,820	193,947	0.3954	0.3802	0.3668

Table 1b: Summary statistics of variables related to cluster existence and strength

Variables	$Cluster_{jkt}$	S_V_{jkt}	$Strength\_V_{jkt}$	S_E_{jkt}	$Strength\_E_{jkt}$
mean	0.0532	0.0010	0.0798	0.0010	0.0798
median	0	0.0002	0	0.0004	0
s.d.	0.2244	0.0032	0.3559	0.0023	0.3559
minimum	0	0	0	0.0000	0
p10	0	0.0000	0	0.0001	0
p25	0	0.0001	0	0.0002	0
p75	0	0.0007	0	0.0009	0
p90	0	0.0023	0	0.0020	0
maximum	1	0.1788	2	0.1660	2
N	371222	371222	371222	371222	371222

Table 2: Summary statistics of dependent and control variables by clusters and non-clusters

	using TFP_ols			**using TFP_op1**			**using TFP_op2**		
Variables	tfp^{AGG}	tfp^{AVG}	tfp^{RAL}	tfp^{AGG}	tfp^{AVG}	tfp^{RAL}	tfp^{AGG}	tfp^{AVG}	tfp^{RAL}
Clusters									
mean	0.7574	0.1850	0.5721	3.1991	2.7252	0.4736	2.8812	2.3524	0.5286
median	0.7628	0.1915	0.5167	3.2807	2.8037	0.4141	2.9284	2.4155	0.4659
s.d.	0.5629	0.5112	0.4354	0.9454	0.8905	0.4188	1.1408	1.0296	0.4643
minimum	-4.0188	-5.5013	-2.4600	-2.1186	-3.6379	-1.3970	-2.5272	-3.6034	-1.9529
p10	0.0642	-0.3976	0.1013	1.9821	1.5797	0.0410	1.3587	0.9559	0.0419
p25	0.3901	-0.1093	0.2820	2.6496	2.2488	0.2038	2.0941	1.6573	0.2283
p75	1.1365	0.4858	0.8076	3.8266	3.3024	0.6811	3.7523	3.1484	0.7663
p90	1.4699	0.7909	1.1146	4.3220	3.7683	0.9853	4.3355	3.6222	1.0985
maximum	2.5061	2.4719	4.4854	5.7981	5.4387	4.1938	5.7362	5.4597	4.6741
N	19736	19736	19736	19736	19736	19736	19736	19736	19736
Non-Clusters									
mean	0.0845	-0.1829	0.2455	2.5760	2.3515	0.2110	2.2111	1.9682	0.2292
median	0.1522	-0.0806	0	2.6815	2.4848	0	2.2623	2.0521	0
s.d.	0.9913	1.0089	0.4567	1.3185	1.3319	0.4481	1.4690	1.4416	0.4831
minimum	-6.6572	-6.6573	-1.7629	-4.2792	-4.2792	-2.6438	-4.2492	-4.2492	-2.6225
p10	-1.0821	-1.3055	0	0.8835	0.6922	-0.0119	0.3429	0.1779	-0.0094
p25	-0.4244	-0.6281	0	1.8321	1.6521	0	1.2790	1.1014	0
p75	0.7082	0.4091	0.3586	3.4538	3.2089	0.2992	3.2497	2.9754	0.3301
p90	1.2319	0.8866	0.8274	4.1440	3.8658	0.7546	4.0617	3.6978	0.8259
maximum	2.5484	2.5484	5.9708	5.8587	5.8587	6.2820	5.8362	5.8362	6.0499
N	351486	351486	351486	351486	351486	351486	351486	351486	351486
Mean Difference	0.6729 ***	0.3679 ***	0.3266 ***	0.6230 ***	0.3738 ***	0.2626 ***	0.6701 ***	0.3842 ***	0.2994 ***

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

Table 2: Summary statistics of dependent and control variables by clusters and non-clusters (Continued)

Variables	Average firm age	Average firm size	Average firm state ownership	Average firm leverage	County-industry employment	County per capita GDP	County total GDP
Clusters							
mean	8.9524	59657.08	0.17163	0.5666	7635.763	223.488	16.6624
median	8.2539	37811	0.0197	0.5756	4404.5	166.37	12.054
s.d.	3.7150	72422.1	0.2843	0.1411	11386.69	187.6909	14.0597
minimum	0	0	0	0.0080	20	1.56	0.0257
p10	5.0270	13652.31	0	0.3847	898	54.27	2.8407
p25	6.4907	22463.85	0	0.4893	1983.5	90.7	5.9531
p75	10.6667	67833.34	0.2134	0.6540	9141	296.07	23.081
p90	13.92	121976.8	0.6592	0.7270	17022	470.94	42.091
maximum	29	1104813	1	1.5061	338800	1024.67	49.25
N	19736	19736	19734	19736	19736	10666	14753
Non-Clusters							
mean	11.1264	48555.34	0.2961	0.6058	848.4793	113.7241	6.3714
median	9.5	18400	0	0.6039	337	76.52	3.9431
s.d.	7.3913	102090	0.4116	0.2629	1384.711	122.8165	7.4402
minimum	0	0	0	0.0080	8	1.56	0.0257
p10	2.5	2848	0	0.2689	50	30.23	0.9421
p25	5	7643.8	0	0.4418	120	46.1343	1.939
p75	16.3333	45034	0.6971	0.7610	948	134.29	7.7481
p90	22	106664	1	0.9176	2279	228.96	14.2161
maximum	29	1286673	1	1.5061	28492	1024.67	49.25
N	351486	351486	349348	351480	351486	226302	276241
Mean Difference	-2.1740 ***	11101.74 ***	-0.1244 ***	-0.0392 ***	6787.284 ***	109.7639 ***	10.2910 ***

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

Table 3: Industrial clustering and county-industry productivities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	using TFP_ols			**using TFP_op1**			**Using TFP_op2**		
	tfp^{AGG}	tfp^{AVG}	tfp^{RAL}	tfp^{AGG}	tfp^{AVG}	tfp^{RAL}	tfp^{AGG}	tfp^{AVG}	tfp^{RAL}
$Cluster_{j,k,t}$	0.215***	0.256***	0.007	0.261***	0.272***	0.019**	0.258***	0.273***	0.016*
	(0.011)	(0.011)	(0.009)	(0.012)	(0.012)	(0.009)	(0.012)	(0.012)	(0.010)
Average firm age	-0.013***	-0.012***	-0.002***	-0.012***	-0.011***	-0.002***	-0.007***	-0.006***	-0.003***
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)
Average firm size	0.372***	0.554***	-0.018***	0.414***	0.537***	-0.020***	0.421***	0.545***	-0.020***
	(0.007)	(0.005)	(0.002)	(0.005)	(0.005)	(0.002)	(0.006)	(0.006)	(0.002)
Average firm state-ownership	-0.106***	-0.115***	0.027***	-0.199***	-0.209***	0.022***	-0.178***	-0.194***	0.028***
	(0.010)	(0.009)	(0.005)	(0.010)	(0.010)	(0.005)	(0.011)	(0.011)	(0.006)
Average firm leverage	-0.154***	-0.201***	0.050***	-0.144***	-0.193***	0.051***	-0.134***	-0.179***	0.049***
	(0.012)	(0.011)	(0.006)	(0.013)	(0.012)	(0.006)	(0.014)	(0.013)	(0.006)
County-industry employment	-0.122***	-0.270***	0.108***	-0.170***	-0.288***	0.093***	-0.157***	-0.282***	0.099***
	(0.004)	(0.004)	(0.002)	(0.004)	(0.004)	(0.002)	(0.005)	(0.005)	(0.002)
County per capita GDP	0.022***	0.009*	0.002	0.035***	0.024***	0.003	0.028***	0.021***	-0.000
	(0.006)	(0.005)	(0.004)	(0.006)	(0.006)	(0.004)	(0.007)	(0.006)	(0.004)
County total GDP	0.204***	0.138***	0.050***	0.202***	0.142***	0.049***	0.188***	0.119***	0.059***
	(0.012)	(0.011)	(0.007)	(0.013)	(0.012)	(0.008)	(0.014)	(0.013)	(0.008)
Constant	-3.212***	-3.878***	-0.453***	-1.422***	-1.728***	-0.389***	-1.356***	-1.647***	-0.412***
	(0.075)	(0.067)	(0.036)	(0.075)	(0.073)	(0.037)	(0.081)	(0.077)	(0.039)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County×Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	233780	233780	233780	233780	233780	233780	233780	233780	233780
adj. R-sq	0.199	0.359	0.037	0.381	0.452	0.027	0.206	0.307	0.028

Note: values in parentheses are robust standard errors; *** = $p < 0.01$; ** = $p < 0.05$; * = $p < 0.1$.

Table 4a: Clustering strength (measured by output) and county-industry productivities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	using TFP_ols			**using TFP_op1**			**Using TFP_op2**		
	tfp^{AGG}	tfp^{AVG}	tfp^{RAL}	tfp^{AGG}	tfp^{AVG}	tfp^{RAL}	tfp^{AGG}	tfp^{AVG}	tfp^{RAL}
$Strength\_V_{jkt}\_1$	0.191***	0.254***	-0.004	0.239***	0.267***	0.009	0.235***	0.266***	0.007
	(0.011)	(0.011)	(0.009)	(0.012)	(0.012)	(0.010)	(0.013)	(0.013)	(0.010)
$Strength\_V_{jkt}\_2$	0.269***	0.261***	0.032**	0.311***	0.284***	0.042***	0.311***	0.291***	0.037***
	(0.015)	(0.016)	(0.013)	(0.017)	(0.018)	(0.013)	(0.018)	(0.018)	(0.014)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County×Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	233780	233780	233780	233780	233780	233780	233780	233780	233780
adj. R-sq	0.199	0.359	0.038	0.381	0.452	0.027	0.206	0.307	0.028

Table 4b: Clustering strength (measured by establishment number) and county-industry productivities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	using TFP_ols			**using TFP_op1**			**Using TFP_op2**		
	tfp^{AGG}	tfp^{AVG}	tfp^{RAL}	tfp^{AGG}	tfp^{AVG}	tfp^{RAL}	tfp^{AGG}	tfp^{AVG}	tfp^{RAL}
$Strength\_E_{jkt}\_1$	0.180***	0.225***	-0.003	0.220***	0.244***	0.002	0.219***	0.246***	-0.000
	(0.011)	(0.011)	(0.009)	(0.012)	(0.012)	(0.009)	(0.013)	(0.012)	(0.010)
$Strength\_E_{jkt}\_2$	0.266***	0.304***	0.021*	0.321***	0.313***	0.045***	0.317***	0.315***	0.040***
	(0.014)	(0.015)	(0.011)	(0.015)	(0.016)	(0.012)	(0.016)	(0.017)	(0.013)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County×Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	233780	233780	233780	233780	233780	233780	233780	233780	233780
adj. R-sq	0.199	0.359	0.038	0.382	0.452	0.027	0.206	0.307	0.028

Note: For convenience, we do not present all the control variables in the table. *** = p < 0.01; ** = p < 0.05; * = p < 0.1.

Table 5: IV regressions for the effects of industrial clustering on county-industry aggregate and reallocation productivities

Panel A: First-Stage	(1) $Cluster_{j,k,t}$		(2) $Cluster_{j,k,t}$		(3) $Cluster_{j,k,t}$	
County retail sales ratio	0.0404***		0.0404***		0.0404***	
	(0.0074)		(0.0074)		(0.0074)	
Provincial mineral output ratio	-0.0002*		-0.0002*		-0.0002*	
	(0.0001)		(0.0001)		(0.0001)	
Panel B: Second-Stage	**using TFP_ols**		**using TFP_op1**		**using TFP_op2**	
	tfp^{AGG}	tfp^{RAL}	tfp^{AGG}	tfp^{RAL}	tfp^{AGG}	tfp^{RAL}
$Cluster_{j,k,t}$	4.934***	1.261**	4.234***	1.035**	4.183***	1.025*
	(1.154)	(0.505)	(1.075)	(0.500)	(1.129)	(0.533)
Average firm age	-0.008***	-0.001	-0.007***	-0.001	-0.002	-0.002**
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Average firm size	0.481***	-0.004	0.501***	-0.011	0.507***	-0.010
	(0.018)	(0.007)	(0.016)	(0.007)	(0.017)	(0.008)
Average firm state-ownership	-0.068***	0.028***	-0.156***	0.024***	-0.150***	0.030***
	(0.015)	(0.007)	(0.014)	(0.007)	(0.015)	(0.007)
Average firm leverage	-0.083***	0.043***	-0.064***	0.042***	-0.073***	0.041***
	(0.018)	(0.008)	(0.018)	(0.008)	(0.019)	(0.008)
County-industry employment	-0.286***	0.075***	-0.289***	0.069***	-0.274***	0.077***
	(0.036)	(0.016)	(0.033)	(0.015)	(0.035)	(0.017)
County per capita GDP	-0.010	-0.007	0.008	-0.004	0.002	-0.008*
	(0.010)	(0.005)	(0.010)	(0.005)	(0.010)	(0.005)
County total GDP	0.113***	0.025**	0.114***	0.027**	0.097***	0.038***
	(0.026)	(0.013)	(0.025)	(0.013)	(0.026)	(0.013)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
County×Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	171879	171879	171879	171879	171879	171879
Second-stage F-test p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Underidentification test p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Weak identification F-statistics	31.515	31.515	31.515	31.515	31.515	31.515
Hansen J statistics	0.3135	0.2300	0.3024	0.2416	0.0699	0.2173

Note: values in parentheses are robust standard errors; *** = p < 0.01; ** = p < 0.05; * = p < 0.1.

Table 6: IV regressions for the effect of industrial clustering on county-industry average TFP

Panel A: First-Stage	(1) $Cluster_{j,k,t}$	(2) $Cluster_{j,k,t}$	(3) $Cluster_{j,k,t}$
County retail sales ratio	0.0411***	0.0411***	0.0411***
	(0.0072)	(0.0072)	(0.0072)
City per capita arable land	-0.0049☆	-0.0049☆	-0.0049☆
	(0.0034)	(0.0034)	(0.0034)
Panel B: Second-Stage	**using TFP_ols** tfp^{AVG}	**using TFP_op1** tfp^{AVG}	**using TFP_op2** tfp^{AVG}
$Cluster_{j,k,t}$	3.812***	3.345***	3.453***
	(0.875)	(0.872)	(0.925)
Average firm age	-0.007***	-0.006***	0.000
	(0.001)	(0.001)	(0.001)
Average firm size	0.606***	0.578***	0.587***
	(0.015)	(0.015)	(0.016)
Average firm state-ownership	-0.087***	-0.195***	-0.177***
	(0.014)	(0.015)	(0.016)
Average firm leverage	-0.155***	-0.158***	-0.150***
	(0.018)	(0.019)	(0.020)
County-industry employment	-0.397***	-0.401***	-0.399***
	(0.033)	(0.033)	(0.035)
County per capita GDP	-0.005	0.012	0.011
	(0.008)	(0.008)	(0.009)
County total GDP	0.071***	0.088***	0.062***
	(0.023)	(0.022)	(0.024)
Year fixed effects	Yes	Yes	Yes
County×Industry fixed effects	Yes	Yes	Yes
N	176163	176163	176163
Second-stage F-test p-value	0.0000	0.0000	0.0000
Underidentification test p-value	0.0000	0.0000	0.0000
Weak identification F-statistics	31.376	31.376	31.376
Hansen J statistics	0.4757	0.1280	0.0821

Note: values in parentheses are robust standard errors; *** = $p < 0.01$; ** = $p < 0.05$; * = $p < 0.1$; ☆ = $p<0.15$

Table 7a: Robustness checks control for the three largest industries in the county

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	using TFP_ols			**using TFP_op1**			**Using TFP_op2**		
	tfp^{AGG}	tfp^{AVG}	tfp^{RAL}	tfp^{AGG}	tfp^{AVG}	tfp^{RAL}	tfp^{AGG}	tfp^{AVG}	tfp^{RAL}
$Cluster_{j,k,t}$	0.213***	0.254***	0.006	0.258***	0.269***	0.019**	0.256***	0.271***	0.016*
	(0.011)	(0.011)	(0.009)	(0.012)	(0.012)	(0.009)	(0.012)	(0.012)	(0.010)
Average firm age	-0.013***	-0.012***	-0.002***	-0.012***	-0.011***	-0.002***	-0.007***	-0.006***	-0.002***
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)
Average firm size	0.372***	0.554***	-0.018***	0.414***	0.536***	-0.020***	0.421***	0.545***	-0.020***
	(0.007)	(0.005)	(0.002)	(0.005)	(0.005)	(0.002)	(0.006)	(0.006)	(0.002)
Average firm state-ownership	-0.107***	-0.115***	0.026***	-0.198***	-0.208***	0.021***	-0.178***	-0.193***	0.027***
	(0.010)	(0.009)	(0.005)	(0.010)	(0.010)	(0.005)	(0.011)	(0.011)	(0.006)
Average firm leverage	-0.151***	-0.198***	0.050***	-0.141***	-0.190***	0.051***	-0.130***	-0.176***	0.049***
	(0.012)	(0.011)	(0.006)	(0.013)	(0.012)	(0.006)	(0.014)	(0.013)	(0.006)
County-industry employment	-0.123***	-0.271***	0.108***	-0.170***	-0.288***	0.092***	-0.157***	-0.283***	0.099***
	(0.004)	(0.004)	(0.002)	(0.004)	(0.004)	(0.002)	(0.005)	(0.005)	(0.002)
County per capita GDP	0.022***	0.009*	0.002	0.034***	0.023***	0.003	0.027***	0.020***	-0.000
	(0.006)	(0.005)	(0.004)	(0.006)	(0.006)	(0.004)	(0.007)	(0.006)	(0.004)
County total GDP	0.200***	0.137***	0.047***	0.201***	0.143***	0.047***	0.187***	0.120***	0.057***
	(0.012)	(0.011)	(0.007)	(0.013)	(0.012)	(0.007)	(0.014)	(0.013)	(0.008)
Constant	-3.471***	-4.095***	-0.547***	-1.635***	-1.889***	-0.473***	-1.529***	-1.770***	-0.495***
	(0.136)	(0.126)	(0.050)	(0.131)	(0.123)	(0.052)	(0.134)	(0.126)	(0.054)
Dummies for the three largest industries in the county	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County×Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	233780	233780	233780	233780	233780	233780	233780	233780	233780
adj. R-sq	0.200	0.360	0.038	0.382	0.452	0.027	0.207	0.308	0.029

Note: values in parentheses are robust standard errors; *** = p < 0.01; ** = p < 0.05; * = p < 0.1; ☆ = p<0.15

Table 7b: Robustness checks control for the location quotient of the county

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	using TFP_ols			**using TFP_op1**			**Using TFP_op2**		
	tfp^{AGG}	tfp^{AVG}	tfp^{RAL}	tfp^{AGG}	tfp^{AVG}	tfp^{RAL}	tfp^{AGG}	tfp^{AVG}	tfp^{RAL}
$Cluster_{j,k,t}$	0.215***	0.257***	0.007	0.262***	0.273***	0.019**	0.259***	0.274***	0.016*
	(0.011)	(0.011)	(0.009)	(0.012)	(0.012)	(0.009)	(0.012)	(0.012)	(0.010)
Location quotient	-0.002***	-0.002***	-0.000	-0.003***	-0.003***	-0.000	-0.003***	-0.002***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Average firm age	-0.013***	-0.012***	-0.002***	-0.012***	-0.011***	-0.002***	-0.007***	-0.006***	-0.003***
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)
Average firm size	0.372***	0.554***	-0.018***	0.414***	0.536***	-0.020***	0.421***	0.545***	-0.020***
	(0.007)	(0.005)	(0.002)	(0.005)	(0.005)	(0.002)	(0.006)	(0.006)	(0.002)
Average firm state-ownership	-0.106***	-0.115***	0.027***	-0.199***	-0.209***	0.022***	-0.178***	-0.194***	0.028***
	(0.010)	(0.009)	(0.005)	(0.010)	(0.010)	(0.005)	(0.011)	(0.011)	(0.006)
Average firm leverage	-0.154***	-0.201***	0.050***	-0.144***	-0.193***	0.051***	-0.134***	-0.179***	0.049***
	(0.012)	(0.011)	(0.006)	(0.013)	(0.012)	(0.006)	(0.014)	(0.013)	(0.006)
County-industry employment	-0.122***	-0.271***	0.108***	-0.170***	-0.288***	0.093***	-0.157***	-0.283***	0.099***
	(0.004)	(0.004)	(0.002)	(0.004)	(0.004)	(0.002)	(0.005)	(0.005)	(0.002)
County per capita GDP	0.022***	0.009*	0.002	0.035***	0.025***	0.003	0.029***	0.022***	-0.000
	(0.006)	(0.005)	(0.004)	(0.006)	(0.006)	(0.004)	(0.007)	(0.006)	(0.004)
County total GDP	0.204***	0.138***	0.050***	0.202***	0.142***	0.049***	0.188***	0.119***	0.059***
	(0.012)	(0.011)	(0.007)	(0.013)	(0.012)	(0.008)	(0.014)	(0.013)	(0.008)
Constant	-3.191***	-3.859***	-0.452***	-1.394***	-1.703***	-0.387***	-1.331***	-1.624***	-0.411***
	(0.075)	(0.067)	(0.036)	(0.076)	(0.073)	(0.037)	(0.081)	(0.077)	(0.039)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County×Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	233780	233780	233780	233780	233780	233780	233780	233780	233780
adj. R-sq	0.199	0.359	0.037	0.382	0.452	0.027	0.206	0.307	0.028

Note: values in parentheses are robust standard errors; *** = $p < 0.01$; ** = $p < 0.05$; * = $p < 0.1$; ☆ = $p<0.15$

Table 8a: Robustness checks using the subsample of counties outside mega cities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	using TFP_ols			**using TFP_op1**			**Using TFP_op2**		
	tfp^{AGG}	tfp^{AVG}	tfp^{RAL}	tfp^{AGG}	tfp^{AVG}	tfp^{RAL}	tfp^{AGG}	tfp^{AVG}	tfp^{RAL}
$Cluster_{j,k,t}$	0.216***	0.255***	0.011	0.261***	0.270***	0.023**	0.260***	0.272***	0.019**
	(0.011)	(0.011)	(0.009)	(0.012)	(0.012)	(0.009)	(0.013)	(0.013)	(0.010)
Average firm age	-0.013***	-0.012***	-0.002***	-0.012***	-0.011***	-0.002***	-0.007***	-0.006***	-0.003***
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)
Average firm size	0.372***	0.555***	-0.018***	0.413***	0.537***	-0.020***	0.421***	0.545***	-0.020***
	(0.007)	(0.005)	(0.002)	(0.005)	(0.005)	(0.002)	(0.006)	(0.006)	(0.002)
Average firm state-ownership	-0.106***	-0.115***	0.026***	-0.199***	-0.209***	0.021***	-0.177***	-0.193***	0.028***
	(0.010)	(0.009)	(0.005)	(0.010)	(0.010)	(0.005)	(0.011)	(0.011)	(0.006)
Average firm leverage	-0.154***	-0.201***	0.050***	-0.144***	-0.193***	0.051***	-0.134***	-0.179***	0.049***
	(0.012)	(0.011)	(0.006)	(0.013)	(0.012)	(0.006)	(0.014)	(0.013)	(0.006)
County-industry employment	-0.122***	-0.271***	0.108***	-0.170***	-0.288***	0.093***	-0.157***	-0.283***	0.100***
	(0.004)	(0.004)	(0.002)	(0.004)	(0.004)	(0.002)	(0.005)	(0.005)	(0.002)
County per capita GDP	0.022***	0.009*	0.002	0.035***	0.025***	0.003	0.029***	0.022***	-0.001
	(0.006)	(0.005)	(0.004)	(0.006)	(0.006)	(0.004)	(0.007)	(0.006)	(0.004)
County total GDP	0.202***	0.136***	0.051***	0.199***	0.139***	0.050***	0.186***	0.116***	0.060***
	(0.012)	(0.011)	(0.007)	(0.013)	(0.012)	(0.008)	(0.014)	(0.013)	(0.008)
Constant	-3.203***	-3.872***	-0.452***	-1.413***	-1.722***	-0.387***	-1.344***	-1.639***	-0.410***
	(0.075)	(0.067)	(0.036)	(0.076)	(0.073)	(0.037)	(0.081)	(0.077)	(0.039)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County×Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	232608	232608	232608	232608	232608	232608	232608	232608	232608
adj. R-sq	0.198	0.359	0.038	0.382	0.452	0.027	0.206	0.307	0.028

Note: values in parentheses are robust standard errors; *** = p < 0.01; ** = p < 0.05; * = p < 0.1; ☆ = p<0.15

Table 8b: Robustness checks control for the population of mega cities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	using TFP_ols			**using TFP_op1**			**Using TFP_op2**		
	tfp^{AGG}	tfp^{AVG}	tfp^{RAL}	tfp^{AGG}	tfp^{AVG}	tfp^{RAL}	tfp^{AGG}	tfp^{AVG}	tfp^{RAL}
$Cluster_{j,k,t}$	0.214***	0.255***	0.007	0.259***	0.271***	0.019**	0.257***	0.272***	0.016*
	(0.011)	(0.011)	(0.009)	(0.012)	(0.012)	(0.009)	(0.012)	(0.012)	(0.010)
Mega city population	-1.660***	-1.386***	0.221	-1.964***	-1.713***	0.056	-1.599***	-1.606***	0.320
	(0.350)	(0.324)	(0.238)	(0.364)	(0.315)	(0.227)	(0.481)	(0.395)	(0.293)
Average firm age	-0.013***	-0.012***	-0.002***	-0.012***	-0.011***	-0.002***	-0.007***	-0.006***	-0.003***
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)
Average firm size	0.372***	0.554***	-0.018***	0.414***	0.537***	-0.020***	0.421***	0.545***	-0.020***
	(0.007)	(0.005)	(0.002)	(0.005)	(0.005)	(0.002)	(0.006)	(0.006)	(0.002)
Average firm state-ownership	-0.106***	-0.115***	0.027***	-0.199***	-0.209***	0.022***	-0.178***	-0.193***	0.028***
	(0.010)	(0.009)	(0.005)	(0.010)	(0.010)	(0.005)	(0.011)	(0.011)	(0.006)
Average firm leverage	-0.154***	-0.201***	0.050***	-0.143***	-0.193***	0.051***	-0.133***	-0.179***	0.049***
	(0.012)	(0.011)	(0.006)	(0.013)	(0.012)	(0.006)	(0.014)	(0.013)	(0.006)
County-industry employment	-0.121***	-0.270***	0.108***	-0.169***	-0.288***	0.093***	-0.156***	-0.282***	0.099***
	(0.004)	(0.004)	(0.002)	(0.004)	(0.004)	(0.002)	(0.005)	(0.005)	(0.002)
County per capita GDP	0.023***	0.009*	0.002	0.036***	0.025***	0.003	0.029***	0.022***	-0.000
	(0.006)	(0.005)	(0.004)	(0.006)	(0.006)	(0.004)	(0.007)	(0.006)	(0.004)
County total GDP	0.202***	0.136***	0.051***	0.200***	0.140***	0.049***	0.186***	0.117***	0.059***
	(0.012)	(0.011)	(0.007)	(0.013)	(0.012)	(0.008)	(0.014)	(0.013)	(0.008)
Constant	-3.155***	-3.830***	-0.461***	-1.355***	-1.670***	-0.391***	-1.301***	-1.592***	-0.423***
	(0.076)	(0.068)	(0.037)	(0.077)	(0.074)	(0.038)	(0.083)	(0.078)	(0.040)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County×Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	233780	233780	233780	233780	233780	233780	233780	233780	233780
adj. R-sq	0.199	0.359	0.037	0.382	0.452	0.027	0.206	0.307	0.028

Note: values in parentheses are robust standard errors; *** = p < 0.01; ** = p < 0.05; * = p < 0.1; ☆ = p<0.15

Table 9. Firm entry and exit patterns within clusters and non-clusters using ASIFP data

Variables	NE	NX	ER	XR	ERS	XRS
Clusters						
mean	11.2218	4.2489	0.3118	0.1366	0.7715	0.7341
median	6	2	0.1739	0.0938	0.4221	0.3908
s.d.	20.8483	8.4441	0.8297	0.3717	1.5660	1.4255
minimum	0	0	0	0	0.0070	0
p10	0	0	0	0	0.1214	0.0828
p25	2	0	0.0625	0	0.2330	0.1862
p75	13	5	0.3478	0.1842	0.7542	0.7432
p90	26	11	0.6364	0.3	1.4105	1.4099
maximum	637	394	34	29	22.9403	18.5946
N	19736	19736	13502	13502	15423	9868
Non-clusters (non-clustered county-industry)						
Mean	1.0701	0.4761	0.1429	0.0876	1.3836	1.1672
median	0	0	0	0	0.4541	0.3749
sd	1.8953	1.2675	0.3667	0.1922	3.2218	2.6415
minimum	0	0	0	0	0.0070	0
p10	0	0	0	0	0.0705	0.0344
p25	0	0	0	0	0.1831	0.1242
p75	1	1	0.1111	0.0455	1.0794	0.9753
p90	3	1	0.5	0.3333	2.7801	2.4450
maximum	41	58	20	9	22.9403	18.5946
N	351486	351486	242680	242680	100069	60109
Mean Difference	10.1517***	3.7728***	0.1689***	0.0491***	-0.6121***	-0.4331***

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

Table 10: Industrial clustering and firm entry and exit patterns

	(1) NE_{jkt}	(2) NX_{jkt}	(3) ER_{jkt}	(4) XR_{jkt}	(5) ERS_{jkt}	(6) XRS_{jkt}
$Cluster_{j,k,t}$	4.211***	1.758***	0.028*	0.018***	0.070	-0.075
	(0.151)	(0.090)	(0.015)	(0.006)	(0.049)	(0.057)
Average firm age	-0.075***	0.000	-0.013***	0.000***	0.042***	-0.003
	(0.002)	(0.001)	(0.000)	(0.000)	(0.007)	(0.007)
Average firm size	-0.481***	-0.249***	-0.054***	-0.027***	0.431***	0.428***
	(0.015)	(0.010)	(0.002)	(0.001)	(0.043)	(0.048)
Average firm state-ownership	-0.048**	-0.045***	-0.022***	-0.017***	0.238**	0.363***
	(0.022)	(0.017)	(0.005)	(0.003)	(0.112)	(0.118)
Average firm leverage	-0.206***	-0.021	-0.012**	0.015***	0.562***	-0.066
	(0.024)	(0.019)	(0.006)	(0.003)	(0.140)	(0.162)
County-industry employment	0.919***	0.482***	0.113***	0.044***	-0.256***	-0.102***
	(0.024)	(0.015)	(0.002)	(0.001)	(0.034)	(0.038)
County per capita GDP	0.118***	0.059***	0.011**	-0.004**	-0.041	-0.022
	(0.023)	(0.016)	(0.005)	(0.002)	(0.036)	(0.085)
County total GDP	0.375***	0.477***	-0.003	0.022***	-0.192**	0.073
	(0.042)	(0.028)	(0.007)	(0.004)	(0.077)	(0.103)
Constant	0.854***	-2.573***	0.075	-0.027	-1.345**	-2.336***
	(0.252)	(0.184)	(0.051)	(0.022)	(0.615)	(0.779)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
County×Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	233780	185465	164369	154563	70061	38324
adj. R-sq	0.093	0.058	0.055	0.049	0.021	0.014

Note: *** = $p < 0.01$; ** = $p < 0.05$; * = $p < 0.1$; ✧=$p<0.15$.

Table 11a: Clustering strength (measured by output) and firm entry and exit patterns

	(1) NE_{jkt}	(2) NX_{jkt}	(3) ER_{jkt}	(4) XR_{jkt}	(5) ERS_{jkt}	(6) XRS_{jkt}
$Strength\_V_{jkt}\_1$	3.635***	1.479***	0.035*	0.010	0.067	-0.069
	(0.184)	(0.102)	(0.019)	(0.010)	(0.050)	(0.059)
$Strength\_V_{jkt}\_2$	5.547***	2.411***	0.012	0.036	0.079	-0.091
	(0.320)	(0.170)	(0.036)	(0.022)	(0.064)	(0.075)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
County×Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	233780	185465	164369	154563	70061	38324
adj. R-sq	0.095	0.059	0.055	0.049	0.021	0.014

Table 11b: Clustering strength (measured by establishment number) and firm entry and exit patterns

	(1) NE_{jkt}	(2) NX_{jkt}	(3) ER_{jkt}	(4) XR_{jkt}	(5) ERS_{jkt}	(6) XRS_{jkt}
$Strength\_E_{jkt}\_1$	3.198***	1.143***	0.048**	0.012*	0.091**	-0.042
	(0.177)	(0.105)	(0.019)	(0.006)	(0.045)	(0.053)
$Strength\_E_{jkt}\_2$	5.727***	2.704***	-0.003	0.027***	0.021	-0.155*
	(0.270)	(0.160)	(0.015)	(0.008)	(0.074)	(0.086)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
County×Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	233780	185465	164369	154563	70061	38324
adj. R-sq	0.096	0.062	0.055	0.049	0.021	0.014

Note: For convenience, we do not present all the control variables in the table. *** = p < 0.01; ** = p < 0.05; * = p < 0.1.

Table 12: Industrial clustering and firm entry and exit patterns using firm (de)registration data

	(1) NE_{jkt}	(2) NX_{jkt}	(3) $Avg\ Regis\ K_{jkt}$
$Cluster_{j,k,t}$	0.024✧	0.050**	-0.053✧
	(0.015)	(0.023)	(0.035)
Average firm age	-0.004***	-0.003***	-0.005***
	(0.001)	(0.001)	(0.002)
Average firm size	-0.003	-0.014***	-0.006
	(0.003)	(0.005)	(0.009)
Average firm state-ownership	0.028***	0.005	0.055*
	(0.009)	(0.015)	(0.031)
Average firm leverage	-0.008	0.004	-0.180***
	(0.011)	(0.018)	(0.036)
County-industry employment	0.025***	0.005	-0.009
	(0.003)	(0.006)	(0.009)
County per capita GDP	-0.001	0.015	0.042**
	(0.006)	(0.009)	(0.018)
County total GDP	-0.018	0.060***	-0.039
	(0.012)	(0.019)	(0.036)
Constant	1.272***	1.195***	4.346***
	(0.069)	(0.113)	(0.198)
Year fixed effects	Yes	Yes	Yes
County×Industry fixed effects	Yes	Yes	Yes
N	119943	75640	119943
adj. R-sq	0.043	0.006	0.018

Note: *** = p < 0.01; ** = p < 0.05; * = p < 0.1, ✧=p<0.15.

Table 13: Statistical correlation between firm size, productivity and markup

	firm sales	firm TFP_ols	firm TFP_op1	firm TFP_op2	firm markup
firm sales	1				
firm TFP_ols	0.2512***	1			
firm TFP_op1	0.1850***	0.8320***	1		
firm TFP_op2	0.1687***	0.7377***	0.7970***	1	
firm markup	0.0211***	0.3408***	0.2388***	0.1823***	1

Note: *** = p < 0.001.

Table 14. Industrial clustering and firm markup distribution within county-industries

	(1) Theil	(2) RMD	(3) p10	(4) p25	(5) p50	(6) p75	(7) p90
$Cluster_{j,k,t}$	-0.001*	-0.010***	0.020***	0.021***	-0.004	-0.042***	-0.045***
	(0.000)	(0.001)	(0.003)	(0.003)	(0.004)	(0.006)	(0.010)
Average firm age	-0.000***	-0.001***	0.001**	-0.000	-0.001***	-0.003***	-0.005***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Average firm size	-0.001***	-0.007***	0.020***	0.015***	0.005***	-0.011***	-0.029***
	(0.000)	(0.000)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Average firm state-ownership	0.000**	0.001	0.001	0.002	-0.000	-0.003	-0.000
	(0.000)	(0.001)	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)
Average firm leverage	0.000	-0.000	-0.051***	-0.052***	-0.051***	-0.050***	-0.050***
	(0.000)	(0.001)	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)
County-industry employment	0.003***	0.020***	-0.051***	-0.039***	-0.013***	0.023***	0.059***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
County per capita GDP	-0.000**	-0.001	0.005***	0.006***	0.006***	0.005	0.000
	(0.000)	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)
County total GDP	0.002***	0.007***	0.011**	0.016***	0.025***	0.040***	0.073***
	(0.000)	(0.001)	(0.004)	(0.004)	(0.005)	(0.007)	(0.008)
Constant	-0.002	-0.024***	1.253***	1.228***	1.219***	1.222***	1.178***
	(0.001)	(0.006)	(0.025)	(0.025)	(0.027)	(0.037)	(0.046)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County×Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	233780	233780	177514	177514	177514	177514	177514
adj. R-sq	0.019	0.054	0.021	0.018	0.013	0.016	0.030

Note: values in parentheses are robust standard errors; *** = p < 0.01; ** = p < 0.05; * = p < 0.1