



Geographic clusters, regional productivity and resource reallocation across firms: Evidence from China

Di Guo^{a,*}, Kun Jiang^b, Chenggang Xu^c, Xiyi Yang^d

^a Brunel University London, Kingston Lane, Uxbridge UB8 3PH, United Kingdom

^b University of Nottingham, Jubilee Campus, Nottingham NG8 1BB, United Kingdom

^c Stanford University, Encina Hall, 616 Jane Stanford Way, Stanford, CA 94305, United States of America

^d ShanghaiTech University, 393 Middle Huaxia Road, Pudong, Shanghai 201210, China

ARTICLE INFO

JEL classification:

D2
L1
O1

Keywords:

Industrial cluster
Productivity growth
Resource reallocation
Competition

ABSTRACT

We link industrial clusters, regional productivity and resource reallocation efficiency with geographical and sectoral disaggregated data. Based on a county-industry level panel from 1998 to 2007 in China, we find that industrial clusters significantly increase 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 are associated with a higher firm turnover with increased entry and exit rates simultaneously; and (ii) within clusters' environment, the dispersion of individual firm's markup is significantly reduced, indicating intensified local competition within clusters. Such results suggest that industrial clusters in China help improve regional productivity and resource allocation efficiency with intensified competition and accelerated firm dynamics. The identification issues are carefully addressed by two-stage estimations with instrumental variables and other robustness checks.

1. Introduction

Economic activity in many countries around the world tends to be spatially concentrated. Dating back to Marshall (1890), theorists have highlighted the benefits of local economies of scale and inter- or intra-industry externalities arising from firms being co-located together, leading to improved regional growth (Arrow, 1962; Jacobs, 1969; Krugman, 1991; Porter, 1990). While empirical estimates for the effects of spatial agglomeration on economic growth have been flourishing over the last few decades, the findings are far from conclusive (e.g., Cingano and Schivardi, 2004; Dekle, 2002; Feldman and Audretsch, 1999; Glaeser et al., 1992; Henderson, 2003; Rosenthal and Strange, 2003). The conflicting results suggest that the net effect of geographical agglomeration remains ambiguous and may depend on different factors of which we still have little understanding (Grashof, 2020; McCann and Folta, 2008).

Productivity growth driven by technological progress and other productivity-enhancing factors beyond traditional input factors, which is usually measured by total factor productivity (TFP),¹ has been

recognized as a major contributor to economic growth, and it explains a large proportion of the variation in economic development between regions (Caselli, 2005; Galor and Tsiddon, 1997; Hall and Jones, 1999; Klenow and Rodriguez-Clare, 1997). At the same time, it is evident that resource misallocation problems can significantly reduce aggregate TFP (Bartelsman et al., 2013; Collard-Wexler and De Loecker, 2015; Restuccia and Rogerson, 2008). If agglomeration positively affects economic growth, what is the relationship between agglomeration, regional TFP and resource misallocation? Does agglomeration affect regional TFP through resource reallocation? The answers are unknown.

This study aims to fill the existing knowledge gap by linking industrial clusters and region-level TFP in China, focusing on resource reallocation among firms in a locality. Specifically, we ask whether industrial clusters improve resource allocation and thereby enhance the TFP of a local industry and, if so, what the mechanisms are through which such effects work on-site.

To address the questions mentioned above, we employ a large firm-level panel dataset in China's manufacturing sector to create a county-industry level industrial cluster panel from 1998 to 2007, using a

* Corresponding author.

E-mail addresses: di.guo@brunel.ac.uk (D. Guo), Kun.Jiang@Nottingham.ac.uk (K. Jiang), cgxu@stanford.edu (C. Xu), yangxy@shanghaitech.edu.cn (X. Yang).

¹ TFP is the leading measure for the impact of technological progress and other productivity-enhancing factors on output without increase of traditional input factors (Olley and Pakes, 1996; Solow, 1957).

density-based Index (DBI, a la Guo et al., 2020).² Based on both county- and firm-level data, applying the methodology of Olley and Pakes (1996), we find that industrial clusters not only increase the aggregate and average TFP but also improve the resource reallocation TFP³ of a local industry. The existence of an industrial cluster in any county-industry is associated with a 2 % increase in reallocation TFP per year between 1998 and 2007, indicating that the existence of industrial clusters can explain 8.5 % of the increase in reallocation TFP in county-industries of China.⁴ Moreover, we provide evidence on how clusters mitigate resource misallocation. We find that firm entry and exit are much more active in clusters than outside clusters, and firm markup dispersion is significantly reduced within clusters. These findings indicate that local competition is intensified within clusters, which reduces resource misallocation across individual firms, supporting the arguments of Porter (2003).

Our discoveries contribute to the research on economic geography in several ways, taking one step forward to explore the net effect of agglomeration and the potential mechanisms. Above all, this study is among the few studies that connect agglomeration and regional TFP using micro-level data, complementing Cingano and Schivardi (2004) and Greenstone et al. (2010). Because of data constraints, empirical studies on the regional effects of agglomeration have typically focused on regional employment, wage growth, or innovation (e.g., Delgado et al., 2014; Feldman and Audretsch, 1999; Glaeser et al., 1992; Henderson, 2003; Henderson et al., 1995; Porter, 2003; Rosenthal and Strange, 2003), assuming the proportional employment gains or innovation improvement in a region are the results of overall productivity increase. Such an assumption is problematic because increases in input in one or two dimensions may be accompanied by decreases in others, and the net productivity growth may be balanced or even reversed due to the elasticity of substitution of input factors (Hicks, 1932; Hicks and Allen, 1934). Therefore, we propose that TFP, a comprehensive measurement of productivity improvement focusing on how input factors are utilized, is appropriate to capture the net effects of agglomeration. It is important to address this issue because it is a key identification problem that brings empirical research closer to the much-discussed theoretical predictions on cluster spillovers.

In addition, this work complements existing research by linking clusters and regional TFP growth with a focus on agglomeration-driven resource reallocation. Resource sharing, knowledge transfer and competition have been identified as major benefits of agglomeration (Arrow, 1962; Jacobs, 1969; Krugman, 1991; Porter, 1990). However, firms are heterogeneous in their capability to take advantage of resource sharing or knowledge transfer (Grillitsch and Nilsson, 2017, 2019; Hervás-Oliver et al., 2018; Knoben et al., 2016). Moreover, intensified competition in clusters may cause damage to some firms due to the increased congestion costs and lowered incentives for innovation under certain conditions. Therefore, the net effects of agglomeration should be related to how resources are shifted among firms under the competition in clusters. Unfortunately, most empirical studies on the regional effects of agglomeration assume that firms within a cluster benefit equally from agglomeration. Such a knowledge gap may bring about conflicting

findings in existing studies. This study demonstrates that certain types of clusters may help improve regional growth by enhancing the shift of resources from less productive to more productive firms, filling the existing knowledge gap and opening a new dimension for agglomeration study. Furthermore, our discoveries showing that intensified competition serves as an important channel for achieving agglomeration externalities in China, where institutional constraints are significantly strong, call for further insightful examinations of Porter's (1990) externalities at the regional level under different conditions. At the same time, it makes theoretical contributions to the debates on the impacts of competition on growth.

Finally, our findings contribute to the literature on economic development and China studies. Scholars suggest that the low aggregate TFP in developing countries is mainly due to micro-level resource misallocation (Banerjee and Duflo, 2005; Caselli and Coleman, 2001; Gancia and Zilibotti, 2009; Restuccia and Rogerson, 2008). Studies have suggested that weak institutions and policies, including labor market regulations, licensing and size restrictions, and the dominance of state ownership, may potentially cause such misallocation (Hsieh and Klenow, 2009; Kochhar et al., 2006; Lewis, 2005). The organization and coordination of entrepreneurial firms within clusters in China have been identified as an institutional innovation to overcome institutional impediments (Guo et al., 2020; Long and Zhang, 2011; Xu, 2011). However, we have little understanding of the mechanisms under which clusters work in China. By providing the first evidence on the relationship between clusters, resource reallocation and regional productivity growth based on micro-level data, 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 through intensified competition. It enriches the evidence to explore the mechanisms that mitigate the resource misallocation problem.

The rest of the paper is organized as follows. The next section discusses the institutional features of industrial clusters in China and reviews the relevant literature. Section 3 discusses the data and samples and introduces how we construct the variables of regional TFP growth. Section 4 presents the empirical findings on clusters and aggregate, average and resource reallocation productivity and addresses the identification concerns using IV regression. Section 5 examines the mechanisms through which clusters affect resource reallocation efficiency, focusing on product market competition. Finally, Section 6 concludes this study.

2. Industrial clusters in China and related literature

2.1. Clusters in China under institutional constraints

The critical conditions for “clustering” in market economies are the protection of property rights and factor mobility. Under these conditions, the market prices of mobile factor inputs will affect firms' collocation decisions that are essential to forming clusters (Ellison et al., 2010; Fujita et al., 1999; Marshall, 1890). Unfortunately, these primary conditions are not met in China due to institutional constraints. Above all, private property rights were not constitutionally recognized until 2004. Moreover, law enforcement for private property rights remains weak (Guo et al., 2014). Meanwhile, almost all input factors are not mobile or tradable. The number one issue is land ownership. According to the constitution, urban land is state-owned, whereas rural land is collectively owned by villages and is not tradable for non-agricultural usage. Associated with the government control of land is the *Hukou*

² The details of why the DBI index captures the distinctive features of the industrial clusters created and developed under the institutional restrictions in China are discussed in Guo et al. (2020).

³ Here, average TFP is the average TFP of all the firms in a county-industry and reallocation TFP measures TFP derived from reallocating resources among firms of different productivity levels within the same county-industry. The sum of the average and reallocation TFP is aggregate TFP, which is the total TFP of all the firms in that county-industry.

⁴ Ended in 2007, findings from our panel data complement the observations that China's resource misallocation problem was alleviated before 2008 (Bai et al., 2016; Song and Xiong, 2018) in the sense that we provide some mechanisms.

system (residence registration system), which restricts labor mobility (Au and Henderson, 2006; Meng, 2000).⁵ Furthermore, the financial market in China is highly underdeveloped and is particularly biased against lending to private enterprises (Allen et al., 2005; Guo et al., 2014). Under the above institutional constraints, firms' co-location decisions are not made solely based on market prices. Instead, the development path and organization of industrial clusters in China differ significantly from those in market economies.

Above all, the emergence and development of industrial clusters are the consequence of the joint efforts of local entrepreneurs and governments to overcome institutional constraints. Research has shown that most of China's industrial clusters have origins in township-village enterprises (TVEs). TVEs are essentially the creatures of local governments and entrepreneurs in the face of the suppression of the private economy (Xu and Zhang, 2009). At the beginning of the economic reform, when the private sector was not recognized and restrictions on rural land use were strict, the only way for rural entrepreneurs to engage in business activities was to use the 'red hat' (*Hongmaozi*) strategy, i.e., to register a business as a collectively owned TVE to seek legal protection and mitigate the discriminations against the private sector (Xu, 2011). Although the TVEs differ in detail across regions, they share the following key characteristics: all were led by rural entrepreneurs, all had vague definitions of ownership at the incipient stage, reflecting certain institutional constraints, and all had close ties with local governments (Chang and Wang, 1994; Che and Qian, 1998; Qian and Xu, 1993). Since the late 1990s, when political and legal resistance to private ownership was gradually relaxed, many TVEs have become privatized. In regions where local governments provided continuous support to those subsequently privatized firms, local residents actively engaged in entrepreneurial activities and eventually formed industrial clusters observed today (Xu and Zhang, 2009). The key role local governments have played was to help relax some of the institutional restrictions (e.g., providing the 'red hat' for private enterprises, relaxing restrictions on private lending and land usage, etc.), provide public goods and coordinate local entrepreneurs (Long and Zhang, 2011). In the past two decades, industrial clusters with a concentration of private entrepreneurial firms coordinated by local governments have emerged rapidly in vast rural areas in coastal provinces.

In addition, firms within clusters are often small in size and highly specialized. In particular, with limited access to formal financial and land resources, private entrepreneurial firms generally start small and use their profits for reinvestment (Ruan and Zhang, 2009). Production processes, which are usually integrated within a single firm in developed countries, are segmented into many small "firms," with each narrowly specialized in one production step. These specialized small firms are linked together through subcontracting networks where a collection of many specialized firms produces a final product. With repeated close interactions within a cluster, the members build trust that forms a basis for coordination and mutual support in many aspects of the business. In such a way, the requirements for input factors such as financial, technological and skilled-labor resources are lowered for entrepreneurial firms in clusters without sacrificing productivity improvement (Ruan and Zhang, 2009). With the concentration of a vast number of small and specialized firms, many cluster-centered townships have become national or international centers of specific products. These clusters often consist of a large number of privatized TVEs or their derivative companies.

⁵ The *Hukou* system has been relaxed gradually that migrant workers are allowed to work outside the geographical area of their *Hukou*. However, migrant workers, especially those from rural areas, are not eligible for or discriminated against in accessing local social welfares, including housing, health care, education childcare benefits, and pensions (Au and Henderson, 2006; Smart and Smart, 2001).

2.2. Clusters, regional productivity and resource reallocation

A central idea of agglomeration economics is that firms can enjoy local economies of scale from co-locating with each other, thereby increasing productivity. The well-known Marshall-Arrow-Romer (MAR) model emphasizes the advantages of regional specialization, while Jacobs (1969) highlights the benefits of knowledge spillovers across industries caused by urban diversity. A more recent theory proposed by Porter (1990) argues that the advantage of agglomeration comes from the intense competition of firms in a locality. This competition provides significant incentives for firms to innovate, accelerating the rate of technological progress and hence 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.

However, empirically, the effects of agglomeration are found to be inconclusive. Many studies document the positive effects of agglomeration on local economic growth and firm productivity though the explanations vary regarding which kind of externalities matters. For example, 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 the MAR model and document the positive effects of regional specialization (e.g., Cingano and Schivardi, 2004; Dekle, 2002; Delgado et al., 2014; Howell, 2017; Jofre-Monseny, 2009; Rosenthal and Strange, 2003; Van Oort and Stam, 2006). At the same time, some studies find evidence for both the Jacobs and MAR externalities, depending on the maturity of the industries (Henderson et al., 1995). However, using cross-country panel data from 70 countries, Henderson (2003) fails to observe the growth-promoting effects of agglomeration by any means. Moreover, some studies find no evidence of the positive effects of agglomeration on firm growth (Globerman et al., 2005; Van Geenhuizen and Reyes-Gonzalez, 2007), firm market value (Zaheer and George, 2004) or employment rate of firms in related industries (Beaudry and Swann, 2009).

The conflicting results from the empirical assessments suggest that the net effect of geographical agglomeration remains ambiguous and may depend on different factors of which we still have little understanding (Grashof, 2020; McCann and Folta, 2008). One of the major knowledge gaps is that the agglomeration literature is disconnected from recent studies on productivity growth and related mechanisms. Empirical studies on the regional effects of agglomeration have typically focused on regional employment, wage growth, or innovation outputs (e.g., Delgado et al., 2014; Feldman and Audretsch, 1999; Glaeser et al., 1992; Henderson, 2003; Henderson et al., 1995; Porter, 2003; Rosenthal and Strange, 2003), assuming the proportional employment gains or innovation improvement in a region are the result of overall productivity increase. However, the elasticity of substitution of input factors is well evident (Hicks, 1932; Hicks and Allen, 1934). An increase in one or two input factors may be accompanied by a decrease in the others. For instance, labor and capital demands may be reduced when technology progresses. Therefore, the net productivity may be balanced or even reversed due to the elasticity of input factor substitutions (Cingano and Schivardi, 2004). Therefore, TFP, a comprehensive measurement of productivity improvement focusing on how input factors are utilized (through technological progress or other productivity-enhancing strategies), is more appropriate to capture the net effects of agglomeration. However, due to data constraints, studies linking agglomeration and regional TFP with geographical and sectoral disaggregated data are limited, with a few exceptions focusing on agglomerations in the US (Greenstone et al., 2010; Henderson, 2003) or Italy (Cingano and Schivardi, 2004). As a result, how agglomeration affects regional TFP in developing counties is unknown.

Another knowledge gap left by the existing literature is that most studies on the effects of agglomeration, whether on firm or regional

economic performance, assume (often as an implicit assumption) that agglomeration effects are homogeneous to firms within a cluster. Such an assumption is problematic. It is well documented that firms differ in resources and capability, which may determine that they generate different outputs even with the same inputs. Indeed, some recent studies have provided evidence that the effects of clusters on firm performance depend on the region or firm-level factors (Arzaghi and Henderson, 2008; Greenstone et al., 2010; Hervas-Oliver et al., 2018; Knoblen et al., 2016; Lee, 2018; Rosenthal and Strange, 2003; Speldekamp et al., 2020).⁶ Identifying which types of firms benefit or suffer from clusters based on firm-level data is essential for understanding the insights of agglomeration. However, we have little knowledge about how resources are shifted among firms within clusters and how the heterogeneous effects of agglomeration on different types of firms relate to its net effect on regional productivity growth.

Finally, Porter's (1990) competition arguments are under-investigated in the agglomeration literature. According to Porter's theory, the most important agglomeration economies are dynamic efficiencies. The intense competition of firms in a locality provides significant incentives for firms to learn and innovate, thereby accelerating technological progress and productivity growth. However, theoretically, the relationship between agglomeration, competition and regional growth is debated. First, the intensified competition within a cluster may reduce firms' incentives to innovate. Following Schumpeter (1950), Aghion and Howitt (1992) argue that when competition is intensified, the laggard's reward for catching up with the technological leader may fall. Therefore, innovation incentives may be reduced by competition. Second, the increasing density of firms in a region may lead to congestion costs and thereby cause the diseconomies of agglomeration (Prevezer, 1997). Such congestion costs may be reflected by the increased transportation, labor, capital, and land costs when firms compete for such input factors (Audretsch and Feldman, 1996; Folta et al., 2006; Stuart and Sorenson, 2003). Third, firms within clusters may be trapped in the regional competition and lose the capability to explore growth opportunities outside the region (Asheim and Isaksen, 2002; Boschma, 2005).

Recognizing the above knowledge gaps, we seek to examine the effects of industrial clusters on regional growth in China from a new perspective to take one more step toward exploring the net effects of agglomeration and related mechanisms. Specifically, based on a large-scale firm-level panel, we examine the effects of industrial clusters on regional TFP, focusing on resource reallocation among firms, taking firm heterogeneity into consideration. Moreover, we explore the mechanisms through which the agglomeration effects are achieved, focusing on the dynamic efficiencies of agglomeration.

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 (Bartelsman et al., 2013; Foster et al., 2001; Melitz, 2003). 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, improvement in management or investment in R&D) but also due to shifts in production factors from less to more productive firms (Collard-Wexler and De Loecker, 2015; Restuccia and Rogerson, 2008). Institutional changes such as deregulation or trade liberalization that lead to the exit of less productive firms

⁶ For instance, Arzaghi and Henderson (2008) and Rosenthal and Strange (2003) document the decay of agglomeration effects over geographic distance. Lee (2018) finds that domestic technological leaders, which have significant technological distance to the global pioneers, benefit from clustering more. Speldekamp et al. (2020) discover that weak firms may take the advantages of strong network and urbanization to compensate the limited internal resources they have. On the contrary, firms with strong internal resources may benefit from urbanized clusters when they have many local partners.

or the expansion of more productive firms can improve aggregate productivity (Bernard et al., 2009; Bustos, 2011; Melitz, 2003; Pavcnik, 2002; Schmitz, 2005). On the contrary, in the presence of institutional distortions such as market imperfections, monopoly power, the lack of protection of property rights as well as the discretionary provision of production factors, highly productive firms may not have sufficient access to resources, and such restrictions to the further development of these firms could lower the aggregate TFP of the economy (Acemoglu et al., 2018; Banerjee and Duflo, 2005; Hsieh and Klenow, 2009; Restuccia and Rogerson, 2008).

In the case of China, it has been known that a substantial amount of production factors is not allocated through the market. Evidence of the negative effects of resource misallocation on aggregate TFP and the connection of such resource misallocation to institutional problems are discovered in the literature (Adamopoulos et al., 2017; Brandt et al., 2012; Hsieh and Klenow, 2009). 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 the optimal level. Brandt et al. (2012) evident that the differences in productivity 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) evidence that eliminating resource misallocation caused by restrictions on land ownership and labor mobility in rural China could increase agricultural productivity by 1.84 times. These 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 evolved with the efforts of entrepreneurs and local governments to overcome the institutional restrictions on the mobility of production factors. Specifically, clustering deepens the division of labor in the production process. As a result, it makes it possible for small entrepreneurial firms to enter the market by focusing on a narrowly defined production stage. These highly specialized entrepreneurial firms closely coordinate alongside the value chain within the clusters. With such a division of labor, the capital and technical barriers to entry are lowered, resulting in increased competition within the clusters (Long and Zhang, 2011; Xu and Zhang, 2009). Studies have systematically shown that these industrial clusters have significantly positive effects on local economic growth (Guo et al., 2020). Therefore, we expect to observe the positive effects of clustering on the aggregate TFP of a locality. Moreover, with lowered entry barriers and a more market-oriented ecosystem, we expect to observe more efficient resource reallocation among firms within the industrial clusters in China. In particular, we expect intensified competition to serve as a major channel through which the clustering improves resource reallocation among individual firms.

3. Data and sample

Our primary dataset is the Above-scale Industrial Firm Panel (ASIFP) 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 micro firms, i.e., non-state firms with annual sales below 5 million RMB. Aware of possible data bias, we use the 2004 Economic Census Data (ECD),⁷ which covers all business entities in China in that year, to cross-check the accuracy of cluster identification. Results show that identified clusters using the DBI measurement are consistent across different datasets, including both the ASIFP and the ECD (see Appendix A).

⁷ 2004 ECD is the only available census data for the examination period.

Compared with the census data, as of 2004, the enterprises covered by ASIFP account for 90 % of the total sales of all manufacturing industries in China.⁸ Furthermore, 84 % of the firms in ASIFP are officially labelled as small enterprises, defined by with no >300 employees. We therefore suggest that the estimations are reliable and not biased by the data used.

A key dependent variable in our study is county-industry level resource reallocation efficiency, measured for each industry 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 and industry-fixed effects. The OLS approach considers only tangible inputs while ignoring unobservable shocks and assuming that all types of inputs are exogenous and hence have no correlation with the error term, 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 both year-fixed and industry-fixed effects. The details of the TFP calculation are summarized in Appendix B.

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 in 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 the aggregate TFP can be decomposed in the following way:

$$\begin{aligned}
 tfp_{jkt}^{AGG} &= \sum_i^n tfp_{ijkt} * share_{ijkt} = \overline{tfp_{jkt}} + \sum_i^n [tfp_{ijkt} - \overline{tfp_{jkt}}] * [share_{ijkt} - \overline{share_{jkt}}] \\
 &= tfp_{jkt}^{AVG} + tfp_{jkt}^{RAL}
 \end{aligned}
 \tag{1}$$

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 by [Guo et al. \(2020\)](#), employing standard regional specialization or interconnectedness 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. At the onset of the post-Mao reform, all the firms were owned or controlled by the state or local governments; thus, governments made decisions on firms' locations. The situation changed gradually, but the legacy is substantial. The concentration of heavy industries in certain regions of China was primarily driven by political concerns. Regions with giant SOEs are likely to be highly specialized when measured by standard cluster 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.⁹

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 that the density of firms in the industry within a locality is one of the essential features of entrepreneurial clusters in China. Hence, we apply the 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 (the construction of the DBI is discussed in detail in [Appendix A](#)). Explicitly, we define a county to have an industrial cluster of a particular industry if the county is among the top α percentile of all counties regarding firm density for that industry, and we assign $\alpha = 5$.¹⁰ We then construct a dummy variable $Cluster_{jkt}$, which equals one if firms of industry j have formed a cluster in county k in year t , and 0 otherwise. We further measure the strength of each cluster based on its relative contribution to the national total industrial output or 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 $S_{-}V_{jkt} = \frac{Output_{jkt}}{Output_j}$, which is a 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 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 the 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 $S_{-}E_{jkt} = \frac{Establishment_{jkt}}{Establishment_j}$ 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 $S_{-}E_{jkt}$.

[Table 1a](#) reports the summary statistics of clusters measured by the DBI. Each year, there are about 1500–2000 industrial clusters in all counties in China, accounting for 5 % of the observed number of county industries (given $\alpha = 5$). These clusters comprised >30 % of the manufacturing firms, contributing to around 30–40 % of the national industrial output and employment from 1998 to 2007. The summary statistics of variables related to cluster existence and strength are reported in [Table 1b](#).

When estimating the effect of clusters 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 County per capita GDP and County total GDP in our regressions to control the effects of regional development level and regional economic size. These data are from the China Socio-Economic Development Statistical Database.

Besides the above-mentioned firm and county-level factors, we also control the industry diversity level of a county because the industry structure of a location may have impacts on regional firm TFP and resource reallocation. The industry diversity level is measured by the

⁸ In the first Chinese Economic Census conducted in 2004, the amount of the total sales for all industrial firms was RMB218 billion, whereas that of the total sales for all ASIFP firms was RMB196 billion.

⁹ 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.

¹⁰ 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 .

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 other firms	Share of clustered firms	Clusters' contribution to national output	Clusters' contribution to national employment
1998	30,178	1509	37,751	67,677	0.3581	0.4057	0.3084
1999	37,218	1861	43,423	93,800	0.3164	0.3595	0.2757
2000	34,356	1718	43,426	88,404	0.3294	0.3663	0.2906
2001	34,783	1739	49,486	96,615	0.3387	0.3641	0.2989
2002	36,251	1813	54,210	104,372	0.3418	0.3655	0.3062
2003	37,238	1862	62,948	113,476	0.3568	0.3751	0.3251
2004	38,634	1932	100,245	142,720	0.4126	0.4080	0.3703
2005	39,852	1993	96,199	152,577	0.3867	0.3862	0.3631
2006	41,118	2056	108,762	168,685	0.3920	0.3886	0.3753
2007	41,594	2080	123,778	187,137	0.3981	0.3852	0.3831

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	371,222	371,222	371,222	371,222	371,222

HHI of industries' output in a given county each year.

In addition, local governments play an essential role in shaping a region's institutions, which may affect the TFP and resource reallocation in a locality. We therefore control the market-based institutions in a region. Precisely, we control the overall marketization index score (*Marketization index1*) obtained from China Marketization Index (CMI, 1998–2007). The CMI is a composite index based on an annual assessment of 25 factors across five aspects of marketization in each province, including the relationship between market and government, the development of the non-public economy, the development of the factor market, the development of factor market, and the development of legal and market services (Fan et al., 2011).

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 C.

Our sample covers firms in >2800 counties and 39 two-digit industries in China from 1998 to 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. In addition, China also modified its industry coding system in 2002 (from GB/T 4754-1994 to GB/T 4754-2002). Therefore, 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 and control variables for clusters and non-clustered county-industries (non-clusters). On average, clusters have significantly higher aggregate, average, and reallocation TFP than non-clusters. Furthermore, firms in clusters tend to be younger, larger in size, have less state-ownership, and have a lower leverage ratio than firms outside clusters. Finally, counties with clusters are more likely to have higher per capita and total GDP than counties without clusters and a lower level of industrial concentration measured by HHI.

4. Findings on industrial clusters and resource reallocation across firms

In the subsequent section, we estimate cluster effects on aggregate, average, and cross-firm resource reallocation TFP. The identification issue will be addressed by two-stage regressions.

4.1. Industrial clusters and reallocation TFP

The following eq. (2) is our baseline regression model. We use it to estimate the effect of clusters on local-industry productivity.

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

TFP_{jkt} is measured by tfp_{jkt}^{AGG} , tfp_{jkt}^{AVG} , or tfp_{jkt}^{RAL} , which are decomposed county-industry level TFP elements. These TFP elements are calculated with eq. (1) based on firm-level TFP, measured by TFP_{ols} , TFP_{op1} , and TFP_{op2} , respectively. $Cluster_{jkt}$ is the DBI cluster dummy variable that equals one if firms from industry j have formed an industrial cluster in county k in year t , or 0 otherwise.

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} (where 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 the expansion of higher-productivity firms and the exits of lower-productivity firms in clusters, which we will further study in the next section. Models (6) and (9) indicate a 2 % increase in reallocation TFP per year between 1998 and 2007. Given that the mean of the reallocation TFP (measured by TFP_{op1} or TFP_{op2}) is approximately 0.235 during this period, industrial clusters 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

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

Variables	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}
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	19,736	19,736	19,736	19,736	19,736	19,736	19,736	19,736	19,736
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	351,486	351,486	351,486	351,486	351,486	351,486	351,486	351,486	351,486
Mean Diff	0.6729	0.3679	0.3266	0.6230	0.3738	0.2626	0.6701	0.3842	0.2994
	***	***	***	***	***	***	***	***	***

Variables	Average firm age	Average firm size	Average firm state ownership	Average firm leverage	County-industry employment	County per capita GDP (in 100 RMB)	County total GDP (in billion RMB)	Marketization index1	County industry HHI
Clusters									
mean	8.95	59,657.08	0.17	0.57	7635.76	223.49	16.66	7.41	0.14
median	8.25	37,811	0.02	0.58	4404.5	166.37	12.05	7.4	0.11
s.d.	3.72	72,422.1	0.28	0.14	11,386.69	187.69	14.06	2.16	0.11
minimum	0	0	0	0.01	20	1.56	0.03	1.49	0.05
p10	5.03	13,652.31	0	0.38	898	54.27	2.84	4.64	0.07
p25	6.49	22,463.85	0	0.49	1983.5	90.7	5.95	5.73	0.08
p75	10.67	67,833.34	0.21	0.65	9141	296.07	23.08	9.1	0.16
p90	13.92	121,976.8	0.66	0.73	17,022	470.94	42.09	10.25	0.27
maximum	29	1,104,813	1	1.51	338,800	1024.67	49.25	11.71	0.99
N	19,736	19,736	19,734	19,736	19,736	10,666	14,753	19,736	19,736
Non-clusters									
mean	11.13	48,555.34	0.30	0.61	848.48	113.72	6.37	6.08	0.22
median	9.5	18,400	0	0.60	337	76.52	3.94	5.76	0.17
s.d.	7.39	102,090	0.41	0.26	1384.71	122.82	7.44	1.98	0.14
minimum	0	0	0	0.01	8	1.56	0.03	0	0
p10	2.5	2848	0	0.27	50	30.23	0.94	3.9	0.09
p25	5	7643.8	0	0.44	120	46.13	1.94	4.63	0.12
p75	16.33	45,034	0.70	0.76	948	134.29	7.75	7.29	0.26
p90	22	106,664	1	0.92	2279	228.96	14.22	8.94	0.39
maximum	29	1,286,673	1	1.51	28,492	1024.67	49.25	11.71	1
N	351,486	351,486	349,348	351,480	351,486	226,302	276,241	351,288	351,486
Mean Diff	-2.18***	11,101.74***	-0.13***	-0.04***	6787.28***	109.77***	10.29***	1.33***	-0.08***

*** $p < 0.01$.

on local-industry productivity. 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 2 for clusters with the median or above-median contribution to national total industrial production.

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 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 0.222. It is larger than the coefficient of weak clusters ($Strength_{V_{jkt}1}$, $\beta=0.196$), but smaller than the coefficient of strong clusters ($Strength_{V_{jkt}2}$, $\beta=0.283$). 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 Models (6) and (9), the coefficient of

Table 3
Industrial clusters and county-industry productivity.

	(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}
$Cluster_{j, k, t}$	0.222*** (0.011)	0.260*** (0.011)	0.009 (0.009)	0.270*** (0.012)	0.278*** (0.012)	0.022** (0.009)	0.266*** (0.012)	0.277*** (0.012)	0.019** (0.010)
Average firm age	-0.040*** (0.005)	-0.059*** (0.005)	0.007*** (0.002)	-0.026*** (0.006)	-0.039*** (0.005)	0.006** (0.002)	0.012* (0.006)	0.001 (0.006)	0.004 (0.002)
Average firm size	0.374*** (0.007)	0.556*** (0.005)	-0.017*** (0.002)	0.415*** (0.005)	0.538*** (0.006)	-0.020*** (0.002)	0.422*** (0.006)	0.546*** (0.006)	-0.020*** (0.002)
Average firm state-ownership	-0.161*** (0.010)	-0.154*** (0.009)	0.011** (0.005)	-0.255*** (0.010)	-0.251*** (0.010)	0.007 (0.005)	-0.228*** (0.011)	-0.227*** (0.011)	0.011** (0.006)
Average firm leverage	-0.181*** (0.012)	-0.216*** (0.011)	0.040*** (0.006)	-0.174*** (0.013)	-0.213*** (0.012)	0.042*** (0.006)	-0.159*** (0.014)	-0.194*** (0.013)	0.039*** (0.006)
County-industry employment	-0.125*** (0.004)	-0.273*** (0.004)	0.107*** (0.002)	-0.173*** (0.004)	-0.290*** (0.004)	0.092*** (0.002)	-0.162*** (0.005)	-0.287*** (0.005)	0.099*** (0.002)
County per capita GDP	0.021*** (0.006)	0.007 (0.005)	0.002 (0.004)	0.035*** (0.006)	0.024*** (0.006)	0.003 (0.004)	0.027*** (0.007)	0.020*** (0.006)	-0.001 (0.004)
County total GDP	0.200*** (0.012)	0.132*** (0.011)	0.050*** (0.007)	0.199*** (0.013)	0.139*** (0.012)	0.049*** (0.008)	0.183*** (0.014)	0.113*** (0.013)	0.059*** (0.008)
Marketization index1	0.049*** (0.005)	0.058*** (0.004)	0.004 (0.003)	0.026*** (0.005)	0.031*** (0.005)	0.003 (0.003)	0.039*** (0.005)	0.042*** (0.005)	0.006** (0.003)
County industry HHI	-0.535*** (0.029)	-0.467*** (0.028)	-0.046*** (0.015)	-0.570*** (0.031)	-0.520*** (0.030)	-0.035** (0.015)	-0.542*** (0.032)	-0.494*** (0.031)	-0.033** (0.016)
Constant	-3.304*** (0.078)	-3.983*** (0.071)	-0.485*** (0.038)	-1.436*** (0.079)	-1.742*** (0.077)	-0.417*** (0.039)	-1.444*** (0.085)	-1.724*** (0.081)	-0.453*** (0.041)
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	233,780	233,780	233,780	233,780	233,780	233,780	233,780	233,780	233,780
adj. R-sq	0.197	0.359	0.037	0.381	0.452	0.027	0.207	0.309	0.028

Note: Standard errors are clustered at the county-industry level.

- *** $p < 0.01$.
- ** $p < 0.05$.
- * $p < 0.1$.

Table 4a
Cluster strength (measured by output) and county-industry productivity.

	(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}
$Strength_{V_{jkl}1}$	0.196*** (0.011)	0.256*** (0.011)	-0.002 (0.009)	0.245*** (0.012)	0.270*** (0.012)	0.012 (0.009)	0.240*** (0.013)	0.267*** (0.013)	0.010 (0.010)
$Strength_{V_{jkl}2}$	0.283*** (0.015)	0.271*** (0.016)	0.035*** (0.013)	0.326*** (0.017)	0.296*** (0.018)	0.046*** (0.013)	0.325*** (0.018)	0.301*** (0.018)	0.040*** (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	233,780	233,780	233,780	233,780	233,780	233,780	233,780	233,780	233,780
adj. R-sq	0.198	0.359	0.037	0.381	0.452	0.027	0.207	0.309	0.028

Note: To save space, we do not present all the control variables in the table; standard errors are clustered at the county-industry level.

- *** $p < 0.01$.

$Strength_{V_{jkl}2}$ 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, $Strength_{E_{jkl}}$, 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 $Strength_{E_{jkl}1}$ and $Strength_{E_{jkl}2}$ are always positive and significant. The magnitudes of the coefficients of $Strength_{E_{jkl}2}$ are larger than those of $Cluster_{jkl}$ (see Table 3). The magnitudes of the coefficients of $Strength_{E_{jkl}1}$ are the smallest. Weak clusters do not seem to have a significant effect on reallocation productivity. Strong clusters, on the other hand, have a positive and

significant impact on the reallocation TFP and the coefficients of $Strength_{E_{jkl}2}$ range from 0.024 to 0.048, depending on how TFP is measured.

The results in Table 3 and Table 4a and 4b imply that industrial clusters, especially clusters with a strong presence of output and firm establishment, are associated with higher aggregate productivity, and such an increase in productivity not only comes from the lift of average firm productivity but also the improvement of resource reallocation efficiency across firms within the local industry. The significant correlation between clusters and improved productivity is interesting, but that alone is insufficient for inferring causality between clusters and productivity improvement. Alternative explanations for the baseline estimations remain. For example, one could argue that the existence or strength of the cluster is a result rather than the cause of the productivity improvement, although such a possibility is slim in China, given that the

Table 4b
Cluster strength (measured by establishment number) and county-industry productivity.

	(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}
<i>Strength_{E_{jk_t}}</i>	0.187*** (0.011)	0.228*** (0.011)	-0.000 (0.009)	0.228*** (0.012)	0.249*** (0.012)	0.005 (0.009)	0.225*** (0.013)	0.249*** (0.012)	0.003 (0.010)
<i>Strength_{E_{jk_t}}</i>	0.275*** (0.014)	0.309*** (0.015)	0.024** (0.011)	0.332*** (0.015)	0.321*** (0.016)	0.048*** (0.012)	0.326*** (0.016)	0.320*** (0.017)	0.043*** (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	233,780	233,780	233,780	233,780	233,780	233,780	233,780	233,780	233,780
adj. R-sq	0.198	0.359	0.037	0.381	0.452	0.027	0.207	0.309	0.028

Note: To save space, we do not present all the control variables in the table; standard errors are clustered at the county-industry level.

*** $p < 0.01$.

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 IVs to identify the effect of clusters.

Specifically, the first IV we use is the per capita number of non-government organizations (NGOs) in each city in a given year. The number is the per capita sum of the stock number for three types of NGOs, including foundations, private non-enterprise entities, and social groups. As we have discussed, a key feature of Chinese industrial clusters is a large number of small specialized firms linked through subcontracting networks to produce the final product. With repeated close interactions within the cluster, trust is built between members, forming the basis for coordination and mutual support among firms in many ways. Social trust and social coordination are therefore essential elements in the development of Chinese industrial clusters (Weitzman and Xu, 1994; Xu and Zhang, 2009). Many empirical studies have shown substantial evidence of the positive impact of NGOs on fostering social trust and community coordination (Coleman, 1990; Putnam, 1993; Fukuyama, 1995; Knack and Keefer, 1997). In the case of China, NGOs were abandoned entirely after 1949 and before the economic reforms. However, they recovered gradually from the mid-1980s and grew dramatically from 1998 onwards when the central government relaxed the laws regulating the non-government sector (Teets, 2009). These NGOs are involved in various activities such as education, poverty alleviation, environment, health, and community development and provide services to vulnerable groups such as the rural poor, migrants and women. Given the critical role social trust and coordination of society plays in fostering the development of industrial clusters, we expect that the presence of industrial clusters is associated with the development of NGOs in a region. On the other hand, the development of NGOs in a region should not directly influence the average firm-level TFP or resource relocation TFP in a specific industry of a county unless the impact is achieved through the presence of industrial clusters in such an industry. Above all, it is well documented that the objectives of NGOs are pretty different from those of for-profit firms. More importantly, in this study, it is noted that the major dependent variables are the average firm TFP and reallocation TFP at the county-industry level. NGOs may be more active in areas where TFP is higher. However, this in no way suggests that NGO development in a city may be directly related to the aggregate or average TFP of firms in a particular industry in a county unless this association is achieved through an increase in the county's economic performance caused by the development of industrial cluster in that specific industry. Moreover, it is even harder to relate the development of NGOs in a city with the resource reallocation of firms within a specific industry in a county, which is the key focus of this study. We therefore suggest that the density of NGOs in a city is a qualified IV.

The second IV we use is the per capita inter-city passenger traffic, measured by the ratio of the total number of inter-city passengers by all transportation means (including embarking and disembarking) over the total population of each city in each year. Inter-city passenger traffic should be closely related to the presence of industrial clusters. First, the development of industrial clusters is often associated with a large wave of population migration, increasing a region's per capita passenger traffic (Fu and Gabriel, 2012; Kerr et al., 2017). Second, more business-related passenger flows will naturally occur with increased trade activity led by industrial clusters. At the same time, we suggest that this IV satisfies the exclusion condition that the per capita passenger traffic should be exogenous from the error terms of the estimations for the average firm TFP or reallocation TFP. While the relationship between productivity and intercity passenger traffic in a region is evident in both theoretical and empirical studies, all available research suggests that this relationship is achieved through agglomeration, regardless of whether the relationship is positive or negative or the agglomeration is defined by urbanization or specialization (Beeson, 1990; Ciccone and Hall, 1996; Eberts and McMillen, 1999; Fujita et al., 1999; Combes and Overman, 2004; Graham, 2007; Ke, 2010). Therefore, we suggest that this IV satisfies both the relevance and exclusion conditions.

The two-stage estimations for aggregate, average and reallocation productivity are reported in Table 5. With two IVs being introduced into the two-stage estimates, we provide statistical evidence for both the relevance and exogeneity of the IVs. Panel A presents the results for the first-stage regressions. Panel A presents the results for the first-stage regressions. Consistent with our expectation, both Per Capita Passenger Traffic and Per capita NGO are significantly and positively correlated with the existence of industrial clusters, confirming the relevance of the IVs. Meanwhile, the Sargan-Hansen tests statistically confirm that these two IVs satisfy the exclusion condition. The second-stage regression results are reported in Panel B. In all the regressions, the coefficients of the instrumented cluster dummy are always positive and significant, except for the average TFP calculated following Olley and Pakes (1996) with an industry-fixed effect. Such results confirm that industrial clusters can increase overall productivity, average productivity, and resource reallocation efficiency across firms of the local industry. In sum, the results of the two-stage estimates are consistent with those of the baseline regressions. Thus, the causal relationships between industrial clusters and increased aggregate, average, and reallocation productivity are confirmed.

4.2. Additional robustness checks

Using the two-stage estimates, we have provided statistical evidence for the causal relationship between industrial clusters, productivity and resource reallocation. In this subsection, we conduct additional robustness checks to rule out alternative explanations for the estimation results to establish the causality further.

Table 5
IV regressions for the effects of industrial clusters on county-industry productivity.

Panel A: First-Stage	(1)	(2)	(3)
	$Cluster_{j, k, t}$	$Cluster_{j, k, t}$	$Cluster_{j, k, t}$
Per capita NGO	0.723*** (0.199)	0.723*** (0.199)	0.723*** (0.199)
Per capita passenger traffic	0.005*** (0.001)	0.005*** (0.001)	0.005*** (0.001)
Control variables	Yes	Yes	Yes

Panel B: Second-Stage	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}
$Cluster_{j, k, t}$	1.551*** (0.551)	0.884* (0.509)	1.690*** (0.638)	1.205** (0.476)	0.455 (0.465)	1.171** (0.552)	1.143*** (0.366)	0.938*** (0.333)	1.032*** (0.367)
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	206,919	206,919	206,919	206,919	206,919	206,919	206,919	206,919	206,919
Hansen J statistic	0.5610	0.6330	0.5762	0.1918	0.1390	0.1293	0.2996	0.1191	0.1465
R-sq	0.1404	0.3767	0.1497	0.3364	0.4550	0.2880	-0.1278	-0.0789	-0.0882
Prob > F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: To save space, we do not present all the control variables in the table; standard errors are clustered at the county-industry level.

- *** $p < 0.01$.
- ** $p < 0.05$.
- * $p < 0.1$.

First, we check whether our results may be rejected if we include some alternative variables measuring market-based institutions in a region. Specifically, we focus on two sub-indices of the CMI to replace the original overall marketization index, i.e., the development of factor market (*Marketization index2*) and the protection of property rights (*Marketization index3*) in a province, which are particularly relevant to a locality's productivity and resource allocation. [Tables A-1a and A-1b](#) present the estimates in which the two sub-indices are controlled, respectively. As shown in the tables, the significantly positive relationships between the existence of industrial clusters and average firm TFP and the resource reallocation TFP stay robust no matter which specific marketization index we include in the estimations.

Second, to further confirm the causal relationship between the clusters and the productivity and resource allocation efficiency of a local industry, we add the interaction terms of the marketization indices and the existence of industrial clusters in the estimation. The rationale is that we emphasize that industrial clusters help firms overcome various institutional constraints in China; if this is indeed the case, we should observe that the impact of industrial clusters diminishes in regions with better institutions because those firms, whether in clusters or non-clusters, face fewer institutional constraints than those in regions with weaker institutions. So if the coefficients of the interaction terms of marketization indices and the existence of industrial clusters are negative in the estimates of average and reallocation TFP, it should have confirmed implications for the causality proposed. We present the estimates in [Tables A-2a, A-2b and A-2c](#), in which we add the interaction terms of industrial clusters and three marketization indices (i.e., the overall marketization index and the indices measuring the development of the local factor market and the protection of property rights in a province in a given year), respectively. As shown in the tables, both the existence of the industrial clusters and marketization indices stay significantly and positively related to the average firm TFP and reallocation TFP within a local industry. However, the coefficients of the interaction terms are constantly and significantly negative, supporting our propositions. Such results suggest that in regions where the institutions are better, the gaps in average firm TFP and reallocation TFP between clusters and non-clusters are narrower, confirming that by overcoming institutional constraints, industrial clusters help improve local firm TFP and the resource reallocation among firms in a locality.

Third, the industrial profile varies across locations so that some

regions may have a concentration of heavy industries, and others may have a concentration of high-tech industries. As a result, the productivity of different industries may vary a lot. As our cluster measurement is derived from the density of firms, one may be concerned with the impacts of industry distribution on the productivity of localities. Although we have controlled for county industry HHI in our analysis, we further run a set of regressions in which we control the three largest industries in each county. Specifically, for each county in a given year, we first identify the three largest industries in terms of output. We then include a set of dummy variables in the regressions indicating whether the industries concerned belong to these largest ones. The results in [Table A-3](#) show that with the three largest industries controlled, the effects of clusters on the aggregate, average, and resource reallocation TFP stay robust.

Fourth, another concern is the impacts of megacities on clusters and productivity. It is known that all Chinese megacities are located in coastal areas. Therefore, it might be possible that the cluster 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×industry fixed effects. However, if the megacity effects overwhelm the county-industry effects, the effects of clusters 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 directly, measured by the population of the megacities. The estimates presented in [Tables A-4a and A-4b](#) show that with the effects of megacities controlled, the results we present in the baseline estimations remain robust.

To summarize, the estimations presented in [Tables A-1a to A-4b](#) confirm that clusters improve the aggregate and average productivity of firms; moreover, they reduce resource misallocation across firms by improving resource reallocation efficiency.

5. Mechanism: industrial clusters and local-industry competition

In this section, we explore how industrial clusters in China alleviate the problem of resource misallocation across firms and improve

productivity. Previous studies on Chinese clusters suggest that clusters may have reduced entry barriers and improved competition within the local industries (Huang et al., 2008; Long and Zhang, 2011; Xu and Zhang, 2009). Specifically, clusters deepen the division of labor in the production process. As a result, it makes it possible for small entrepreneurial firms to enter the market by focusing on a narrowly defined production stage. However, 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 clusters by investigating their relationship with firm entry and exit patterns and how they affect firm markup dispersion within the local industries.

5.1. Industrial clusters and firm entry and exit

To examine the firm entry and exit within clusters, we follow Dunne et al. (1988) to calculate firms' entry and exit patterns 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 output of firms exiting industry j of county k between years $t-1$ and t .

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

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

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

$$Turnover_{jkt} = (NE_{jkt} + NX_{jkt}) / NT_{jkt}$$

Moreover, to measure the relative size of the new entries 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:

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

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

Table 6 reports the comparative statistics of firm entry and exit patterns within clusters and non-clusters based on the ASIFP data. The results indicate that firm entry and exit are more active in clusters than those in non-clusters. Moreover, the entry rate in the cluster ($ER = 0.3118$) is more than twice as high as that in non-clusters ($ER = 0.1429$). Furthermore, the exit rate in clusters ($XR = 0.1646$) is significantly higher than that in non-clusters ($XR = 0.0879$). These results suggest a higher competition level within industrial clusters since the firm turnover is significantly higher. Moreover, on average, the more significant turnover in clusters seems to be mainly driven by small firms, given that the mean values of ERS and XRS are smaller than 1 in clusters, and they are much lower than those in non-clusters.

Table 7 reports the regression results for the effects of clusters 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 entries and exiting firms. Meanwhile, columns (3) and (4) show a similar pattern of cluster effects on firms' entry and exit rates. On average, a county with a cluster in a given industry may have 4.2 more new entries and 1.8 more exiting firms in the industry

compared to a county without a cluster in this industry after controlling for factors such as the average age, size, ownership, and leverage of firms within the county-industry as well as the size, development level, marketization level and industry diversity of the counties. Similarly, on average, a county with a cluster in a given industry has a 0.029 (19.1 % of the mean) higher entry rate and 0.019 higher (20.6 % of the mean) exit rate of firms than a county without a cluster in the industry. Finally, Column (5) shows that the turnover rate of firms is significantly higher in counties with a cluster in a given industry than others without a cluster in the industry. On average, a county with a cluster in a given industry has a 0.027 (5.7 % of the mean) higher firm turnover rate than a county without a cluster.

Tables 8a and 8b show the relationships between the strength of the clusters and firm dynamics. Again, it is clear that the strength of the clusters, no matter measured by the total outputs or the number of firms in the industry, is significantly and positively correlated with the number of new entries and exiting firms, and the turnover of firms. Overall, the results of Tables 7, 8a and 8b confirm our conjecture that industrial clusters expose firms to greater competition and therefore facilitate the reshuffling of market shares from less to more productive firms.

Results shown in Tables 7 and 8 are based on the ASIFP data, which only contains SOEs or non-state firms with annual sales of 5 million RMB or above. So, the "entry" into the panel data may include cases where an existing non-state firm's sales grow to exceed 5 million RMB, and the "exit" may consist of cases where an existing non-state 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 to 2007. Using this dataset, we can calculate the number of newly established 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, allowing 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. Therefore, we cannot calculate the firm entry rate, exit rate, or relative size to incumbents. Instead, we calculate the growth of firm entry or exit from the current to the following year. Table 9 reports the OLS results for the effects of clusters on these measures of firm entry and exit patterns based on the firm (de) registration data. As shown in the table, the existence of clusters is significantly and positively associated with the number of new entries and existing firms in a county-industry between 1998 and 2007 and the growth rate of exiting firms. 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 for firms in a county without a cluster in this industry. Such results further confirm that clusters lower the entry barriers of firms and thereby intensify the competition in a locality.

Although we use registration data to test further the robustness of estimates for the impact of industrial clusters on resource reallocation, we note that there may be self-selection, i.e., some firms choose to deregister themselves rather than due to elimination by the market competition. In order to address such concerns, we conduct a series of robustness checks as follows.

Above all, a central question underlying our resource reallocation argument is whether less productive firms are more likely to lose market share or even drop out in clusters than in non-clusters. To directly address the concern of which firms lose market shares or are deregistered, we compare which type of active firms lose market shares in and outside clusters using our ASIFP data. We suggest that such an empirical strategy provides further support for our argument. At the same time, it reduces the concern of self-selected deregistration (i.e., some entrepreneurs may choose to close their businesses for other reasons rather than that firms are less productive). We will discuss these further estimates

Table 6
Firm entry and exit patterns within clusters and non-clusters using the ASIFP data.

Variables	NE	NX	ER	XR	Turnover	ERS	XRS
Clusters							
mean	11.2218	5.3495	0.3118	0.1646	0.5193	0.7715	0.7122
median	6	3	0.1739	0.1111	0.44	0.4221	0.3796
s.d.	20.8483	9.0588	0.8297	0.3823	0.3464	1.5660	1.3581
minimum	0	0	0	0	0	0.0070	0
p10	0	0	0	0	0.1364	0.1214	0.0785
p25	2	1	0.0625	0.0313	0.2532	0.2330	0.1795
p75	13	7	0.3478	0.2143	0.7391	0.7542	0.7327
p90	26	13	0.6364	0.3571	1	1.4105	1.3795
maximum	637	394	34	29	2	22.9403	17.9958
N	19,736	17,377	13,502	14,047	17,377	15,423	10,896
Non-clusters (non-clustered county-industry)							
Mean	1.0701	0.5584	0.1429	0.0879	0.4741	1.3836	1.1718
median	0	0	0	0	0.4	0.4541	0.3713
sd	1.8953	1.1826	0.3667	0.1917	0.4713	3.2218	2.6178
minimum	0	0	0	0	0	0.0070	0
p10	0	0	0	0	0	0.0705	0.0345
p25	0	0	0	0	0	0.1831	0.1243
p75	1	1	0.1111	0.0555	1	1.0794	0.9840
p90	3	3	0.5	0.3333	1	2.7801	2.4793
maximum	41	38	20	9	2	22.9403	17.9958
N	351,486	283,333	242,680	236,476	283,333	100,069	59,081
Mean difference	10.152***	4.791***	0.169***	0.077***	0.045***	-0.612***	-0.460***

Note: Standard errors are clustered at the county-industry level.
*** $p < 0.01$.

Table 7
Industrial clusters and firm entry and exit patterns.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	NE_{jkt}	NX_{jkt}	ER_{jkt}	XR_{jkt}	$Turnover_{jkt}$	ERS_{jkt}	XRS_{jkt}
$Cluster_{j, k, t}$	4.206*** (0.150)	1.761*** (0.090)	0.029* (0.015)	0.019*** (0.006)	0.027*** (0.007)	0.067 (0.049)	-0.077 (0.056)
Average firm age	-0.712*** (0.017)	0.049*** (0.012)	-0.130*** (0.003)	0.017*** (0.001)	-0.163*** (0.003)	0.322*** (0.059)	-0.054 (0.071)
Average firm size	-0.471*** (0.015)	-0.248*** (0.010)	-0.055*** (0.002)	-0.028*** (0.001)	-0.082*** (0.002)	0.430*** (0.043)	0.436*** (0.049)
Average firm state-ownership	-0.095*** (0.022)	-0.073*** (0.016)	-0.029*** (0.005)	-0.021*** (0.003)	-0.042*** (0.006)	0.269** (0.112)	0.361*** (0.118)
Average firm leverage	-0.192*** (0.025)	-0.035* (0.019)	-0.017*** (0.006)	0.010*** (0.003)	-0.020*** (0.007)	0.579*** (0.140)	-0.044 (0.162)
County-industry employment	0.940*** (0.024)	0.475*** (0.014)	0.118*** (0.002)	0.044*** (0.001)	0.055*** (0.002)	-0.257*** (0.035)	-0.112*** (0.039)
County per capita GDP	0.106*** (0.023)	0.057*** (0.016)	0.012** (0.005)	-0.003 (0.002)	-0.002 (0.004)	-0.043 (0.036)	-0.028 (0.085)
County total GDP	0.365*** (0.041)	0.468*** (0.028)	-0.000 (0.007)	0.025*** (0.004)	0.042*** (0.007)	-0.195** (0.077)	0.063 (0.103)
Marketization index1	0.241*** (0.029)	0.049*** (0.016)	-0.007** (0.003)	-0.020*** (0.001)	-0.033*** (0.003)	0.010 (0.041)	0.064 (0.043)
County HHI	-0.072 (0.082)	-0.192*** (0.063)	-0.094*** (0.015)	-0.051*** (0.008)	0.027 (0.017)	0.685** (0.297)	-0.298 (0.346)
Constant	0.486 (0.302)	-2.774*** (0.213)	0.271*** (0.055)	0.032 (0.024)	1.948*** (0.042)	-1.846*** (0.689)	-2.428*** (0.798)
Year-fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County×Industry-fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	233,780	185,465	164,369	154,563	185,465	70,061	38,324
adj. R-sq	0.094	0.058	0.054	0.052	0.291	0.020	0.015

Note: Standard errors are clustered at the county-industry level.
*** $p < 0.01$.
** $p < 0.05$.
* $p < 0.1$.

below.

Firstly, as shown in Table A-5, a significant and positive correlation exists between firms' TFP and their market shares in clusters ($* = p < 0.01$), indicating that more productive firms gain more market share in clusters on average. However, in non-clusters, this relationship is not clear. In particular, when TFP is measured by the OP method, there

seems to be a negative association between firm TFP and market share in non-clusters. Secondly, we focus on exiting firms' TFP in clusters and non-clusters. As we use the ASIFP data, exiting composes both "real exiting" and a decrease in sales to below 500 million RMB for non-state firms. The results are presented in Table A-6. It shows that no matter how TFP is measured, exiting firms in clusters always have higher TFP

Table 8a
Cluster strength (measured by output) and firm entry and exit patterns.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	NE_{jkt}	NX_{jkt}	ER_{jkt}	XR_{jkt}	$Turnover_{jkt}$	ERS_{jkt}	XRS_{jkt}
<i>Strength_{V_{jkt}}</i> 1	3.626*** (0.183)	1.476*** (0.102)	0.035* (0.019)	0.010 (0.010)	0.027*** (0.007)	0.070 (0.050)	-0.069 (0.059)
<i>Strength_{V_{jkt}}</i> 2	5.547*** (0.320)	2.413*** (0.170)	0.013 (0.036)	0.036 (0.022)	0.034*** (0.009)	0.072 (0.064)	-0.087 (0.076)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County×Industry-fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	233,780	185,465	164,369	154,563	185,465	70,061	38,324
adj. R-sq	0.095	0.059	0.055	0.051	0.282	0.021	0.015

Note: To save space, we do not present all the control variables in the table; standard errors are clustered at the county-industry level.

*** $p < 0.01$.

* $p < 0.1$.

Table 8b
Cluster strength (measured by establishment number) and firm entry and exit patterns.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	NE_{jkt}	NX_{jkt}	ER_{jkt}	XR_{jkt}	$Turnover_{jkt}$	ERS_{jkt}	XRS_{jkt}
<i>Strength_{E_{jkt}}</i> 1	3.192*** (0.176)	1.142*** (0.105)	0.048** (0.019)	0.012* (0.006)	0.022*** (0.007)	0.092** (0.045)	-0.042 (0.053)
<i>Strength_{E_{jkt}}</i> 2	5.719*** (0.270)	2.702*** (0.160)	-0.002 (0.015)	0.027*** (0.008)	0.041*** (0.009)	0.020 (0.073)	-0.151* (0.086)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County×Industry-fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	233,780	185,465	164,369	154,563	185,465	70,061	38,324
adj. R-sq	0.097	0.062	0.056	0.051	0.282	0.021	0.015

Note: To save space, we do not present all the control variables in the table; standard errors are clustered at the county-industry level.

*** $p < 0.01$.

** $p < 0.05$.

* $p < 0.1$.

Table 9
Industrial clusters and firm entry and exit patterns using firm (de)registration data.

	(1)	(2)	(3)	(4)	(5)
	NE_{jkt}	NX_{jkt}	$Avg\ Regis\ K_{jkt}$	NE_growth_{jkt}	NX_growth_{jkt}
<i>Cluster_{j, k, t}</i>	2.155*** (0.589)	1.452** (0.668)	-0.060* (0.035)	-0.022 (0.029)	0.144* (0.086)
Average firm age	-0.374*** (0.115)	0.021 (0.144)	-0.074*** (0.014)	0.043*** (0.014)	0.009 (0.039)
Average firm size	-0.144** (0.058)	-0.065 (0.092)	-0.002 (0.009)	0.025*** (0.009)	-0.013 (0.029)
Average firm state-ownership	1.072*** (0.190)	-0.165 (0.233)	0.059* (0.031)	0.034 (0.029)	-0.007 (0.083)
Average firm leverage	-0.146 (0.207)	-0.494 (0.314)	-0.164*** (0.036)	0.026 (0.036)	-0.156 (0.102)
County-industry employment	0.771*** (0.108)	-0.112 (0.139)	-0.013 (0.010)	-0.040*** (0.009)	0.025 (0.027)
County per capita GDP	0.152 (0.105)	0.148 (0.180)	0.036** (0.018)	0.001 (0.022)	0.048 (0.041)
County total GDP	1.066*** (0.267)	0.754 (0.463)	-0.054 (0.036)	-0.112*** (0.037)	0.170* (0.099)
Marketization index1	1.874*** (0.217)	0.302 (0.233)	0.094*** (0.015)	0.028** (0.013)	-0.165*** (0.044)
County HHI	1.336** (0.579)	-1.072 (0.917)	-0.130 (0.109)	-0.282*** (0.104)	-0.085 (0.330)
Constant	-13.080*** (2.580)	3.528 (3.672)	4.116*** (0.207)	0.758*** (0.214)	0.788 (0.606)
Year-fixed effects	Yes	Yes	Yes	Yes	Yes
County×Industry-fixed effects	Yes	Yes	Yes	Yes	Yes
N	119,943	75,640	119,943	77,057	42,763
adj. R-sq	0.016	0.001	0.019	0.006	0.004

Note: Standard errors are clustered at the county-industry level.

*** $p < 0.01$.

** $p < 0.05$.

* $p < 0.1$.

than their counterparts in non-clusters. That is, firms with the same TFP are more likely to lose market share if they are in clusters. Thirdly, we compare the TFP difference between survivors and exiting firms in clusters and non-clusters, presented in Table A-7. It shows that within clusters, the gap in TFP between those surviving and those exiting the ASIFP is substantially smaller than the gap in non-clusters, which to a certain extent, indicates that the competition among firms with different TFP levels is fiercer in clusters than in non-clusters.

The summary statistics discussed above provide further evidence that in counties with clusters, less productive firms lose more market shares than those in regions without clusters though they may not necessarily be deregistered. Such results support the regression analysis of the relationship between industrial clusters and resource reallocation among the firms of specific industries within a county.

5.2. Industrial clusters and firm markup dispersion within the local industry

Early inquiries into the relationship between resource misallocation and aggregate TFP tend to focus on the misallocation of production factors and generally assume that all firms have the same markup within industries (e.g., Hsieh and Klenow, 2009; Restuccia and Rogerson, 2008). In contrast, recent studies in international trade and industrial organizations have developed models with endogenous markup (De Loecker and Warzynski, 2012; Edmond et al., 2015, 2018). In their models, markup increases with firm size, generating the same misallocation captured by Restuccia and Rogerson (2008) and Hsieh and Klenow (2009). In this sub-section, we explore if China's industrial clusters mitigate markup dispersion, thus reducing resource misallocation.

The discussion of the efficiency costs of markups can be traced back to the study of 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 et al., 2014). Furthermore, some recent papers (Baqae and Farhi, 2020; 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. In a more competitive environment where resources or market shares can be reallocated freely from less productive to more productive firms, more productive (and higher markup) firms will produce more, reducing their markups. Similarly, lower productivity (and lower markup) firms will produce less, increasing their markups. Hence, if China's industrial clusters provide a more competitive environment, there should be reduced firm markup dispersion within clusters compared with non-clustered local industries. Furthermore, within clusters, the individual firm markup at the higher quantile should be decreased while that at the 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. Then, 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 of markup are described in Appendix D. As shown in Table 10, 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

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 toward more productive firms, and hence the net effect on average markup can be ambiguous.

charges.

The following is the formal test on the effect of industrial clusters on firm markup distribution within county-industries using panel regressions. We use two measurements of firm markup dispersion to ensure the robustness of our results. The first one is the Theil index ($Theil_{jkt} = \frac{1}{n_{jkt}} \sum_{i=1}^{n_{jkt}} \frac{y_{ijkt}}{\bar{y}_{jkt}} \log \frac{y_{ijkt}}{\bar{y}_{jkt}}$), where y_{ijkt} is the markup of firm i of industry j in county k at year t . \bar{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 in 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 ($RMD_{jkt} = \frac{1}{n_{jkt}} \sum_{i=1}^{n_{jkt}} \left| \frac{y_{ijkt}}{\bar{y}_{jkt}} - 1 \right|$). In addition to investigating the effect of industrial clusters 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 every year and then estimate the effect of clusters on firm markup at different percentiles separately. The estimation model is the following:

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

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

Table 11 reports the regression results on the effect of clusters 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 clusters alone can explain 15.38% and 19.15% of the decrease of $Theil_{jkt}$ and RMD_{jkt} , respectively. The rest columns illustrate that the cluster effects on firms' markups vary with size. For smaller firms at the lower quintiles (Columns 3 & 4), the cluster effect is significant and positive, implying the enlargement of these firms' markups. However, 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 middle-sized firms (Column 5), the cluster effect is insignificant. These findings prove that China's clusters offer a more competitive environment, which reduces the gap in markups between large and small firms and mitigates resource misallocation across firms within clusters.

6. Discussions and conclusion

Based on a systematic analysis of county-industry panel data, we find that China's industrial clusters have significantly improved the productivity of local industries. The increased productivity not only comes from higher average firm productivity but also comes from more efficient resource reallocation across firms within the clusters. Additionally, we find concrete mechanisms through which China's clusters mitigate the resource misallocation problem. For example, there is a higher level of firm turnover within clusters than outside clusters, and the startup capital for firms within a cluster is lower than that outside a cluster. Moreover, firms' markups within clusters have smaller dispersion than those outside clusters. These findings imply that industrial clusters in China provide a more competitive environment that contributes to more efficient resource reallocation within clusters.

This study fills three important knowledge gaps in agglomeration research by linking regional TFP and resource reallocation among firms within clusters. The three gaps refer to the disconnection between the agglomeration literature and the recent regional growth literature, the lack of understanding of the heterogeneity of agglomeration effects, and the under-investigation of Porter's competition arguments in regional growth settings. It sheds light on some challenging questions in agglomeration and development studies. Precisely, by estimating regional aggregate TFP and TFP changes driven by resource

Table 10
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

*** $p < 0.001$.

Table 11
Industrial clusters and firm markup distribution within county-industries.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Theil	RMD	p10	p25	p50	p75	p90
$Cluster_{j, k, t}$	-0.001 [◇] (0.000)	-0.009*** (0.001)	0.019*** (0.003)	0.019*** (0.003)	-0.004 (0.004)	-0.042*** (0.006)	-0.045*** (0.010)
Average firm age	-0.001*** (0.000)	-0.003*** (0.000)	-0.003 (0.002)	-0.006*** (0.002)	-0.008*** (0.002)	-0.012*** (0.003)	-0.024*** (0.003)
Average firm size	-0.001*** (0.000)	-0.006*** (0.000)	0.021*** (0.002)	0.016*** (0.002)	0.006*** (0.002)	-0.010*** (0.002)	-0.027*** (0.002)
Average firm state-ownership	0.000 (0.000)	-0.002** (0.001)	0.003 (0.004)	0.003 (0.004)	-0.004 (0.004)	-0.015*** (0.006)	-0.015** (0.006)
Average firm leverage	0.000 (0.000)	-0.002*** (0.001)	-0.047*** (0.005)	-0.049*** (0.005)	-0.051*** (0.005)	-0.054*** (0.007)	-0.055*** (0.007)
County-industry employment	0.003*** (0.000)	0.020*** (0.000)	-0.052*** (0.001)	-0.040*** (0.001)	-0.015*** (0.002)	0.022*** (0.002)	0.057*** (0.002)
County per capita GDP	-0.000** (0.000)	-0.001 (0.001)	0.004** (0.002)	0.005*** (0.002)	0.005*** (0.002)	0.004 (0.003)	-0.001 (0.004)
County total GDP	0.002*** (0.000)	0.007*** (0.001)	0.009** (0.004)	0.014*** (0.004)	0.023*** (0.005)	0.038*** (0.007)	0.071*** (0.008)
Marketization index1	0.000*** (0.000)	0.001*** (0.000)	0.020*** (0.002)	0.018*** (0.002)	0.019*** (0.002)	0.022*** (0.003)	0.024*** (0.003)
County industry HHI	-0.002*** (0.000)	-0.016*** (0.002)	0.025* (0.015)	0.009 (0.015)	-0.016 (0.016)	-0.058*** (0.019)	-0.106*** (0.021)
Constant	-0.003** (0.001)	-0.026*** (0.006)	1.182*** (0.026)	1.168*** (0.026)	1.157*** (0.028)	1.150*** (0.039)	1.118*** (0.047)
Year-fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County×Industry-fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	233,780	233,780	177,514	177,514	177,514	177,514	177,514
adj. R-sq	0.019	0.052	0.022	0.019	0.014	0.016	0.030

Note: Standard errors are clustered at the county-industry level and reported in parentheses.

*** $p < 0.01$.

** $p < 0.05$.

* $p < 0.1$.

◇ $p < 0.15$.

reallocation, we capture the spillover effects generated by agglomeration in the absence of structural changes in input factors, thus bringing the empirical estimates closer to the theoretical predictions of the agglomeration literature. Moreover, such results may shed some light on explaining the heterogeneous effects of agglomeration (e.g., [Arzaghi and Henderson, 2008](#); [Lee, 2018](#); [Rosenthal and Strange, 2003](#); [Speldekamp et al., 2020](#)). In particular, it helps explain a challenging paradox found in the existing studies, namely, the seemingly contradicting observations that higher performance and lowered survival rates exist simultaneously in clusters (e.g., [Folta et al., 2006](#); [Myles Shaver and Flyer, 2000](#); [Staber, 2001](#)). Moreover, the findings on the lowered markup dispersions and the accelerated firm turnover within clusters suggest that both the entry and exit barriers are lowered within clusters in China. It therefore enriches evidence for [Porter's \(1990\)](#) competition arguments in two ways. First, it is among the few empirical studies examining agglomeration effects on regional growth by applying Porter's competition framework. Second, it provides a rigorous analysis of firm dynamics and markup dispersions among firms within a region simultaneously for the first time in the agglomeration literature. By adding this evidence, this study contributes to the existing theoretical debates on whether the benefits of competition may offset the costs of competition within clusters in certain contexts.

Finally, it complements existing discussions on economic growth by identifying agglomeration, a particular type of production organization, as a contributor to resource reallocation for the first time. It therefore adds some thoughts on why firms choose to co-locate in some cities with extremely high production costs and why the variations in economic development levels persist across regions (e.g., [Banerjee and Duflo, 2005](#); [Gancia and Zilibotti, 2009](#); [Hsieh and Klenow, 2009](#); [Restuccia and Rogerson, 2008](#)).

The findings from this study have significant policy implications. Many governments worldwide try to promote local economic growth by providing incentives for firms to co-locate, believing that firms may benefit from the spillovers from the agglomeration. However, our discovery suggests that firms are not equally benefited from the agglomeration. Therefore, policymakers should consider the industry structure, general environments, and organizational structure of firms in a region when designing urban or agglomeration policies.

Some limitations of this study are worthwhile to mention. Above all, the estimations are up to 2007 because of the constraints of data sources. Both the institutional environments within China and the global trade have experienced significant changes since then, so more updated data may help explore more insights regarding the institutional impacts on the effect of agglomeration on regional growth in the country. However,

it is important to note that some theoretical nature of the research questions focused on in this study and the corresponding empirical findings (i.e., the lowered entry and exit barrier and intensified competition within clusters may help improve resource reallocation and thereby improve regional productivity) are time- and context-independent. Therefore, we suggest that such limitations should not significantly impact the contributions of this study. In addition, the primary sample of this study does not cover micro non-state-owned firms with sales below RMB5 million, which may have caused some potential biases. However, the robustness checks based on the 2004 ECD, which covers all business entities in China, provide evidence that the potential issue of the data does not change the findings of our estimations qualitatively. Finally, we focus only on manufacturing firms in this study because of the data constraints. If the effects of agglomeration on the service sector can be studied in future research, our understanding will be further improved.

Several challenging questions arising from our discoveries require further research. One of our major findings is that industrial clusters enhance regional productivity through improved resource reallocation in China. Our findings also suggest that the accelerated firm turnover and intensified competition within clusters serve as a potential channel for achieving such effects. Further exploration of alternative mechanisms and why particular mechanisms are more prevalent than others in certain regions requires much more research. At the same time, examinations on the reallocation of which specific resources are more likely to be affected by agglomeration deserve more scrutiny. Further studies on the characteristics of new entries to and exiting firms from clusters under different conditions may enrich our understanding of why some clusters last long while others decline with age. Furthermore, a systematic analysis of which aspects of technological progress or organizational learning are more likely to be affected by agglomeration and

how such effects are related to the net effects of agglomeration deserves more investigation. Finally, our findings indicate that the collaboration between entrepreneurs and local governments to overcome institutional constraints is important for the development of clusters in China. Therefore, it is worth investigating whether it is also an important element in a market economy based on comparative analysis to further contribute to the literature on economic development and institutions.

CRediT authorship contribution statement

Di Guo: Conceptualization, Resources, Investigation, Writing – original draft, Writing – review & editing, Project administration. **Kun Jiang:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Visualization, Writing – original draft, Writing – review & editing. **Chenggang Xu:** Conceptualization, Resources, Methodology, Writing – review & editing, Supervision, Funding acquisition. **Xiyi Yang:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Construction of the DBI

As discussed by Guo et al. (2020), in addition to regional specialization, the density of firms in an industry within a locality is one of the most important features of entrepreneurial clusters in China. Hence, the authors propose a density-based index (DBI) to measure entrepreneurial clusters in China. This study uses the same DBI measurement to identify country-level clusters. Formally, we define:

$$Cluster_{j,k,t} = \begin{cases} 1, & \text{if } fn_{j,k,t} \geq (100 - \alpha) \text{ percentile of } \{fn_{j,1,t}, fn_{j,2,t}, \dots, fn_{j,K,t}\}; \\ 0, & \text{otherwise} \end{cases}$$

where we denote the number of firms in any 2-digit¹³ industry $j \in \{1, 2, \dots, J\}$ in county $k \in \{1, 2, \dots, I\}$ at time t as $fn_{j,k,t}$. We define county k as “a county with an α cluster of industry j ” if the number of firms in this county is among the top α percentile of all counties for any given industry at time t . In this paper, we focus on the top five percentile county-level clusters ($\alpha = 5$).¹⁴ We use this dummy variable to capture the existence of the DBI cluster in any county-industry in year t .

Fig. A-1 shows the distribution of the DBI clusters in Chinese counties calculated using both the Above-Scale Industrial Firm Panel (ASIFP) and the 2004 Economic Census Data (ECD). As Guo et al. (2020) have discussed, measured by the DBI cluster count, the top five provinces were Zhejiang, Jiangsu, Guangdong, Shandong, and Shanghai in 2007. Such results are consistent with the general perceptions of spatial distributions of entrepreneurial clusters in China. Furthermore, the identified clusters using the DBI measurement are generally consistent across the two datasets.

Appendix B. Estimation of firm TFP

TFP is measured using two methods to ensure that the conclusions of this study are not driven by a specific TFP measure.

The first measure is a straightforward OLS residual from a log-linear transformation of a general Cobb–Douglas production function. However, the OLS production function estimates may be biased when unobservable shocks correlate with input levels. Meanwhile, the OLS method lacks dynamic consideration. In the second OP method, we follow Olley and Pakes (1996), who use investment as a proxy for unobservable production shocks. This semi-parametric method is applied to control for the simultaneity caused by unobserved productivity and the non-random sample selection induced by

¹³ We choose a two-digit SIC to construct the DBI for two main reasons. First, industrial clusters are defined at the county-industry level. As we have discussed, industrial clusters in China often include networks of many specialized firms. If we use a narrower definition of industry (e.g., a three-digit or four-digit SIC), it may be too hard to capture the characteristics of Chinese industrial clusters or the resource reallocation of industries within a county. Second, two-digit industry classifications are commonly used in the literature to define industrial clusters or agglomerations (e.g., Glaeser et al., 1992; Baptista and Swann 1998). Using the same measures for industry classification allows us to compare our findings with existing literature.

¹⁴ Our results stay robust when we assign other values for α , such as 3 and 8.

the different probabilities of exits for small and large low-productivity firms.

B.1. OLS method explanation

The OLS method is straightforward. In the OLS regression, the TFP, which denotes the effects on the total output that are not caused by the tangible inputs in the production and represents the technological dynamism, is estimated as the error term.

The equation below demonstrates the estimation of TFP through the OLS method.

$$Y = A \times K^\alpha \times L^\beta$$

$$\rightarrow \ln Y = \ln A + \alpha \ln K + \beta \ln L$$

As the residual of the OLS regression, $\ln A$ is the TFP we intend to measure. The firm-level TFP estimation considers the year and two-digit SIC code fixed effects. The robustness of the estimation results is verified by relaxing the year effects.

The disadvantage of the OLS method is that it only considers tangible inputs, such as labor and capital, and not unobservable shocks. This aspect results in a static model, in which all types of inputs are exogenous and have no correlation with the error term (i.e., TFP). The limitation of the OLS method is obvious, and the associated coefficients are biased.

B.2. OP method explanation

In the presence of selection bias and simultaneity, the OP estimation allows for the endogeneity of some input factors and unobserved productivity differences among firms. Moreover, such an estimation also considers the exit of firms from the market. Hence, the OP estimation has several advantages over the simple OLS method.

The Olley and Pakes' (1996) approach is characterized by the Bellman equation and assumes that the firm constantly maximizes the expected discounted value of future profits. Thus, stay-or-quit and investment decisions in each period are formulated.

For estimation purposes, this study uses the Cobb–Douglas production function. In particular, gross output and value-added production functions are adopted. The equation below denotes the production function in the OP method.

$$\text{Total Output}_{it} = \beta_0 + \beta_1 L_{it} + \beta_2 K_{it} + \beta_3 I_{it} + w_{it} + \varepsilon_{it}$$

where Total Output_{it} is deflated by the producer price index for manufactured products, L_{it} is the labor input by firm i at time t (either the number of employees or the total payment of employees of a firm can be a proxy for this variable), K_{it} is the capital input by firm i at time t and is deflated by the price index of investment in fixed assets, I_{it} denotes the intermediate inputs by firm i at time t and is deflated by the producer price index for purchasing products, w_{it} is the productivity shock known by a firm when it makes its liquidation decision and investment decision, and ε_{it} is the true error term.

In this study, all variables in the equations are in their logarithm form, and the time trend and two-digit industry heteroskedasticity are controlled.

B.3. Data source

The firm-level data of this study, together with associated financial information, are derived from the Above-scale Industrial Firm Panel (ASIFP). ASIFP comprises virtually all the manufacturing firms in China, including all the state-owned enterprises and non-state firms with annual sales of at least 5 million RMB between 1996 and 2007. This database covers input information, such as labor, fixed asset, and intermediate inputs, as well as other firm-specific characteristics such as location, industry, and age. The dataset is an unbalanced panel data with gaps.

As a prerequisite to TFP calculation, real capital stock stimulates discussion and dispute. The lack of firm-level capital stock data causes difficulty in constructing a series of real capital stocks, which are comparable across time and firms. Following Brandt et al. (2012), the perpetual inventory method (PIM) is applied. Through the PIM method, the effective capital stock in production is measured as a weighted sum of previous fixed asset investments in constant price terms.

$$RCS_t = \sum_{\tau=0}^{\infty} d_\tau I_{t-\tau}$$

where RCS_t is real capital stock in year t , d_τ is the efficiency of fixed assets in the τ th year, and $I_{t-\tau}$ is the fixed asset investment flow τ years ago.

With the additional assumption of d_τ declining in a geometric pattern, we write the PIM equation as follows:

$$d_\tau = (1 - \delta)^\tau$$

$$RCS_t = RCS_{t-1} + I_t - \delta RCS_{t-1}.$$

This study formulates fixed asset growth at the two-digit SIC code level as a recursive step back to when a firm was established. Applying the preceding PIM method, together with the series of investment deflators from China Urban Life and Price Yearbook (2009), this study constructs the series of real capital stocks. 1978 is set as the starting point of the initial capital stock for the series calculation, and 9 % is applied as the fixed depreciation rate, to be specific. Finally, all the nominal values are deflated by price indices with the benchmark 100 set in 1996.

In the OP model, a firm's decision-making process, whether or not a firm opts to remain in the market, must be clarified. However, this information is not contained in the dataset used by this study. Accordingly, the panel data themselves are used to verify this exit variable. Using the unbalanced panel data with gaps ranging from 1996 to 2007, we define that a firm exits from the market when the observation record is not continuous. The dummy variable *exit* is equal to 1 if the firm exits from the market in the current period or 0 if otherwise.

Appendix C. Definitions of variables

Variables	Definitions
Dependent variables	
tfp_{jkt}^{AGG}	The aggregate productivity of industry j in county k at year t
tfp_{jkt}^{AVG}	The average productivity of firms of industry j in county k at year t
tfp_{jkt}^{REAL}	Reallocation productivity of industry j in county k at year t
NE_{jkt}	Number of 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
ER_{jkt}	The entry rates between year $t-1$ and t of industry j in county k
XR_{jkt}	The exit rates between year $t-1$ and t of industry j in county k
$Turnover_{jkt}$	Sum of entries and exiting firms divided by the total number of firms of industry j in county k in year t
ERS_{jkt}	The average size of entering-firms relative to incumbents of industry j in county k in year t
XRS_{jkt}	The average size of exiting-firms relative to non-exiting-firms of industry j in county k in year t
$Avg\ Regis\ K_{jkt}$	The average registration capital of newly established firms of industry j in county k in year t
$Theil_{jkt}$	The Theil index of firm markup of industry j in county k at year t
RMD_{jkt}	The relative mean deviation of firm markup of industry j in county k at year t
$p\#_{jkt}$	Firm markup at the # percentile of industry j in county k at year t
Independent variables	
$Cluster_{jkt}$	A dummy variable that equals 1 if firms of industry j form a cluster in county k in year t , and it equals 0 if otherwise
$Strength\ V_{jkt}$	A categorical variable, where 0 indicates industry j in county k in year t does not form a cluster, 1 indicates a weak cluster, and 2 indicates a strong cluster. Weak or strong is measured by the cluster's relative contribution to the total national output
$Strength\ E_{jkt}$	A categorical variable, where 0 indicates industry j in county k in year t does not form a cluster, 1 indicates a weak cluster, and 2 indicates a strong cluster. Weak or strong is measured by the cluster's relative contribution to the national total establishment number
Control variables	
Average firm age	The average age of firms within the same county-industry
Average firm size	Average sales (in logarithm) of firms within the same county-industry
Average firm state-ownership	The average ratio of state ownership to the total paid-in capital of the firms within the same county-industry
Average firm leverage	The average ratio of total liability to the total asset of the firms within the same county-industry
County-industry employment	Total number of employment (in logarithm) of each county-industry
County per capita GDP	County per capita GDP (in logarithm)
County total GDP	County total GDP (in logarithm)
Marketization index1	The overall marketization score at the provincial level
Marketization index2	The sub-index measuring the development of the factor market at the provincial level
Marketization index3	The sub-index measuring private property right protection at the provincial level
County industry HHI	County Herfindahl-Hirschman Index (HHI) of industries' output
Instrumental variables	
Per capita NGO	The per capita stock number for three types of NGOs, including foundations, private non-enterprise entities, and social groups in each city of each year
Per capita passenger traffic	The ratio of the total number of inter-city passengers by all transportation means (including embarking and disembarking) over the total population of each city in each year

Appendix D. Estimations of markup

Generally speaking, demand-based methods and production-based methods are two standard ways to estimate firm-level markup. Demand-based methods estimate the residual demand curve, while production-based methods estimate the production function. Following [De Loecker and Warzynski \(2012\)](#) and [Lu and Yu \(2015\)](#), we use production-based methods to calculate markup. They have proved that markup is the difference between an input's output elasticity and the input's revenue share.

Suppose firm i at time t has the following production function that is assumed to be continuous and twice differentiable w.r.t. all of its variables,

$$Q_{it} = F_{it}(L_{it}, K_{it}, M_{it}, \omega_{it})$$

where L_{it} , K_{it} , and M_{it} are labor, capital, and intermediate goods inputs, and ω_{it} is total factor productivity.

The firm i has to minimize its cost at time t

$$\min_{\{L_{it}, K_{it}, M_{it}\}} p_{it}^l L_{it} + p_{it}^k K_{it} + p_{it}^m M_{it}$$

$$\text{s.t. } F_{it}(L_{it}, K_{it}, M_{it}, \omega_{it}) \geq \bar{Q}_{it}$$

where p_{it}^l , p_{it}^k , and p_{it}^m are the prices for labor, capital, and intermediate goods, respectively, and \bar{Q}_{it} is the reservation value for the firm.

Following [Lu and Yu \(2015\)](#), we focus on the optimal choice of intermediate goods to estimate markup. We suggest using the Lagrangian method to solve the cost minimization problem; we define the Lagrangian as,

$$\mathcal{L}(L_{it}, K_{it}, M_{it}, \lambda_{it}) = p_{it}^l L_{it} + p_{it}^k K_{it} + p_{it}^m M_{it} + \lambda_{it} (\bar{Q}_{it} - F_{it}(L_{it}, K_{it}, M_{it}, \omega_{it}))$$

The FOC for intermediate goods is

$$\frac{\partial \mathcal{L}}{\partial M_{it}} = p_{it}^m - \lambda_{it} \frac{\partial F_{it}}{\partial M_{it}} = 0$$

Based on the above equation, we first transpose, then multiply both sides by $\frac{M_{it}}{Q_{it}}$, and then multiply RHS by $\frac{P_{it}}{P_{it}}$. We get the following equation, and P_{it} is the price of the final good.

$$\frac{\partial F_{it}}{\partial M_{it}} \frac{M_{it}}{Q_{it}} = \frac{1}{\lambda_{it}} \frac{p_{it}^m M_{it}}{Q_{it}} = \frac{P_{it}}{\lambda_{it}} \frac{p_{it}^m M_{it}}{P_{it} Q_{it}}$$

According to De Loecker and Warzynski (2012), $\frac{P_{it}}{\lambda_{it}}$ is the markup defined as the ratio of price over marginal cost. And then we can get,

$$\mu_{it} = \theta_{it}^m (\alpha_{it}^m)^{-1}$$

where μ_{it} is the markup, $\theta_{it}^m \equiv \frac{\partial F_{it}}{\partial M_{it}} \frac{M_{it}}{Q_{it}}$ is the output elasticity of intermediate goods, and $\alpha_{it}^m \equiv \frac{p_{it}^m M_{it}}{P_{it} Q_{it}}$ is the intermediate goods' revenue share. α_{it}^m can be calculated conveniently because the information on the expenditure of intermediate goods and total sales are available in the ASIFP dataset. However, the output elasticity of intermediate goods must be obtained from the estimated production function. We follow Olley and Pakes (1996) to estimate the production function, and how our production function is estimated can be found in Appendix B for the estimation of TFP. Specifically, the output elasticity of intermediate goods is the first-order derivative of our production function w.r.t intermediate goods.

Appendix E. Tables for robustness checks

Table A-1a

Control for marketization index related to the development of factor market.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	using TFP_ols			using TFP_op1			Using TFP_op2		
	<i>tffp</i> ^{AGG}	<i>tffp</i> ^{AVG}	<i>tffp</i> ^{RAL}	<i>tffp</i> ^{AGG}	<i>tffp</i> ^{AVG}	<i>tffp</i> ^{RAL}	<i>tffp</i> ^{AGG}	<i>tffp</i> ^{AVG}	<i>tffp</i> ^{RAL}
<i>Cluster_{j, k, t}</i>	0.211*** (0.011)	0.254*** (0.012)	0.007 (0.009)	0.255*** (0.012)	0.265*** (0.012)	0.022** (0.009)	0.251*** (0.013)	0.266*** (0.013)	0.018* (0.010)
Average firm age	-0.040*** (0.005)	-0.061*** (0.005)	0.009*** (0.002)	-0.023*** (0.006)	-0.039*** (0.006)	0.007*** (0.007)	0.014** (0.007)	0.001 (0.006)	0.005* (0.003)
Average firm size	0.371*** (0.007)	0.557*** (0.006)	-0.020*** (0.002)	0.413*** (0.006)	0.540*** (0.006)	-0.022*** (0.002)	0.419*** (0.006)	0.546*** (0.006)	-0.022*** (0.002)
Average firm state-ownership	-0.148*** (0.010)	-0.153*** (0.010)	0.013** (0.006)	-0.246*** (0.011)	-0.251*** (0.011)	0.010* (0.006)	-0.224*** (0.012)	-0.232*** (0.011)	0.014** (0.006)
Average firm leverage	-0.173*** (0.013)	-0.220*** (0.012)	0.045*** (0.006)	-0.162*** (0.013)	-0.209*** (0.013)	0.046*** (0.006)	-0.154*** (0.014)	-0.197*** (0.014)	0.042*** (0.007)
County-industry employment	-0.123*** (0.005)	-0.276*** (0.004)	0.112*** (0.002)	-0.163*** (0.005)	-0.286*** (0.005)	0.097*** (0.002)	-0.151*** (0.005)	-0.282*** (0.005)	0.104*** (0.002)
County per capita GDP	0.019*** (0.006)	0.007 (0.006)	0.001 (0.004)	0.036*** (0.006)	0.026*** (0.006)	0.003 (0.004)	0.027*** (0.007)	0.022*** (0.007)	-0.002 (0.004)
County total GDP	0.197*** (0.012)	0.128*** (0.011)	0.046*** (0.008)	0.194*** (0.013)	0.134*** (0.013)	0.045*** (0.008)	0.176*** (0.015)	0.106*** (0.014)	0.055*** (0.008)
Marketization index2	0.048*** (0.004)	0.036*** (0.003)	0.014*** (0.002)	0.040*** (0.004)	0.029*** (0.004)	0.012*** (0.002)	0.042*** (0.004)	0.031*** (0.004)	0.012*** (0.002)
County industry HHI	-0.533*** (0.031)	-0.459*** (0.029)	-0.047*** (0.016)	-0.555*** (0.032)	-0.502*** (0.031)	-0.035** (0.016)	-0.527*** (0.034)	-0.476*** (0.032)	-0.034** (0.017)
Constant	-3.947*** (0.095)	-4.686*** (0.086)	-0.541*** (0.047)	-1.211*** (0.095)	-1.547*** (0.093)	-0.458*** (0.048)	-2.187*** (0.102)	-2.517*** (0.098)	-0.474*** (0.050)
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	217,914	217,914	217,914	217,914	217,914	217,914	217,914	217,914	217,914
adj. R-sq	0.194	0.356	0.038	0.369	0.445	0.027	0.205	0.306	0.028

Note: standard errors are clustered at the county-industry level.

- *** $p < 0.01$.
- ** $p < 0.05$.
- * $p < 0.1$.

Table A-1b

Control for marketization index related to the protection of property rights.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	using TFP_ols			using TFP_op1			Using TFP_op2		
	<i>tffp</i> ^{AGG}	<i>tffp</i> ^{AVG}	<i>tffp</i> ^{RAL}	<i>tffp</i> ^{AGG}	<i>tffp</i> ^{AVG}	<i>tffp</i> ^{RAL}	<i>tffp</i> ^{AGG}	<i>tffp</i> ^{AVG}	<i>tffp</i> ^{RAL}
<i>Cluster_{j, k, t}</i>	0.214*** (0.011)	0.256*** (0.012)	0.008 (0.009)	0.260*** (0.012)	0.269*** (0.012)	0.023** (0.009)	0.255*** (0.013)	0.269*** (0.013)	0.018* (0.010)
Average firm age	-0.041*** (0.005)	-0.062*** (0.005)	0.008*** (0.002)	-0.025*** (0.006)	-0.040*** (0.006)	0.007*** (0.007)	0.013* (0.007)	0.000 (0.006)	0.005* (0.003)
Average firm size	0.370*** (0.007)	0.557*** (0.006)	-0.020*** (0.002)	0.411*** (0.006)	0.538*** (0.006)	-0.022*** (0.002)	0.418*** (0.006)	0.546*** (0.006)	-0.022*** (0.002)

(continued on next page)

Table A-1b (continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	using TFP_ols			using TFP_op1			Using TFP_op2		
	<i>tfp</i> ^{AGG}	<i>tfp</i> ^{AVG}	<i>tfp</i> ^{RAL}	<i>tfp</i> ^{AGG}	<i>tfp</i> ^{AVG}	<i>tfp</i> ^{RAL}	<i>tfp</i> ^{AGG}	<i>tfp</i> ^{AVG}	<i>tfp</i> ^{RAL}
Average firm state-ownership	-0.148*** (0.010)	-0.154*** (0.010)	0.013** (0.006)	-0.242*** (0.011)	-0.247*** (0.011)	0.010* (0.006)	-0.222*** (0.012)	-0.230*** (0.011)	0.013** (0.006)
Average firm leverage	-0.174*** (0.013)	-0.220*** (0.012)	0.044*** (0.006)	-0.163*** (0.013)	-0.210*** (0.013)	0.045*** (0.006)	-0.156*** (0.014)	-0.197*** (0.014)	0.042*** (0.007)
County-industry employment	-0.122*** (0.005)	-0.276*** (0.004)	0.112*** (0.002)	-0.160*** (0.005)	-0.284*** (0.005)	0.097*** (0.002)	-0.149*** (0.005)	-0.281*** (0.005)	0.104*** (0.002)
County per capita GDP	0.021*** (0.006)	0.008 (0.006)	0.002 (0.004)	0.038*** (0.006)	0.028*** (0.006)	0.003 (0.004)	0.029*** (0.007)	0.024*** (0.007)	-0.001 (0.004)
County total GDP	0.205*** (0.012)	0.135*** (0.011)	0.049*** (0.008)	0.198*** (0.013)	0.137*** (0.013)	0.047*** (0.008)	0.181*** (0.015)	0.110*** (0.014)	0.057*** (0.008)
Marketization index3	0.010*** (0.002)	0.012*** (0.002)	0.004*** (0.001)	-0.005** (0.002)	-0.005** (0.002)	0.004*** (0.001)	0.003 (0.003)	0.003 (0.002)	0.005*** (0.001)
County industry HHI	-0.539*** (0.031)	-0.464*** (0.029)	-0.049*** (0.016)	-0.558*** (0.032)	-0.505*** (0.031)	-0.037** (0.016)	-0.532*** (0.034)	-0.480*** (0.032)	-0.036** (0.017)
Constant	-3.805*** (0.095)	-4.612*** (0.086)	-0.511*** (0.046)	-1.000*** (0.095)	-1.386*** (0.093)	-0.432*** (0.048)	-2.028*** (0.102)	-2.401*** (0.098)	-0.454*** (0.050)
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	217,914	217,914	217,914	217,914	217,914	217,914	217,914	217,914	217,914
adj. R-sq	0.193	0.356	0.038	0.369	0.445	0.027	0.204	0.305	0.028

Note: standard errors are clustered at the county-industry level.

*** $p < 0.01$.

** $p < 0.05$.

* $p < 0.1$.

Table A-2a

Interaction effects of marketization index and industrial clusters.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	using TFP_ols			using TFP_op1			Using TFP_op2		
	<i>tfp</i> ^{AGG}	<i>tfp</i> ^{AVG}	<i>tfp</i> ^{RAL}	<i>tfp</i> ^{AGG}	<i>tfp</i> ^{AVG}	<i>tfp</i> ^{RAL}	<i>tfp</i> ^{AGG}	<i>tfp</i> ^{AVG}	<i>tfp</i> ^{RAL}
<i>Cluster_{j, k, t}</i>	0.386*** (0.027)	0.302*** (0.028)	0.085*** (0.022)	0.367*** (0.030)	0.304*** (0.030)	0.064*** (0.023)	0.404*** (0.032)	0.343*** (0.032)	0.061** (0.024)
Cluster * Marketization index1	-0.023*** (0.003)	-0.006* (0.004)	-0.011*** (0.003)	-0.014*** (0.004)	-0.004 (0.004)	-0.006** (0.003)	-0.019*** (0.004)	-0.009** (0.004)	-0.006** (0.003)
Marketization index1	0.053*** (0.005)	0.059*** (0.004)	0.006** (0.003)	0.028*** (0.005)	0.032*** (0.005)	0.004 (0.003)	0.042*** (0.006)	0.043*** (0.005)	0.007** (0.003)
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	233,780	233,780	233,780	233,780	233,780	233,780	233,780	233,780	233,780
adj. R-sq	0.198	0.359	0.037	0.381	0.452	0.027	0.207	0.309	0.028

Note: To save space, we do not present all the control variables in the table; Standard errors are clustered at the county-industry level.

*** $p < 0.01$.

** $p < 0.05$.

* $p < 0.1$.

Table A-2b

Interaction effects of marketization index (development of factor market) and industrial clusters.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	using TFP_ols			using TFP_op1			Using TFP_op2		
	<i>tfp</i> ^{AGG}	<i>tfp</i> ^{AVG}	<i>tfp</i> ^{RAL}	<i>tfp</i> ^{AGG}	<i>tfp</i> ^{AVG}	<i>tfp</i> ^{RAL}	<i>tfp</i> ^{AGG}	<i>tfp</i> ^{AVG}	<i>tfp</i> ^{RAL}
<i>Cluster_{j, k, t}</i>	0.352*** (0.024)	0.299*** (0.026)	0.070*** (0.021)	0.379*** (0.026)	0.326*** (0.028)	0.064*** (0.021)	0.395*** (0.028)	0.344*** (0.029)	0.062*** (0.022)
Cluster* Marketization index2	-0.026*** (0.004)	-0.008** (0.004)	-0.011*** (0.003)	-0.023*** (0.004)	-0.011*** (0.004)	-0.008** (0.003)	-0.026*** (0.004)	-0.014*** (0.004)	-0.008** (0.003)
Marketization index2	0.051*** (0.004)	0.037*** (0.003)	0.015*** (0.002)	0.043*** (0.004)	0.030*** (0.004)	0.013*** (0.002)	0.045*** (0.004)	0.032*** (0.004)	0.013*** (0.002)
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	217,914	217,914	217,914	217,914	217,914	217,914	217,914	217,914	217,914
adj. R-sq	0.194	0.356	0.038	0.370	0.445	0.027	0.205	0.306	0.028

Note: To save space, we do not present all the control variables in the table; Standard errors are clustered at the county-industry level.

*** $p < 0.01$.

** $p < 0.05$.

Table A-2c

Interaction effects of marketization index (protection of property rights) and industrial clusters.

	(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}
$Cluster_{j, k, t}$	0.250*** (0.019)	0.257*** (0.020)	0.038** (0.015)	0.296*** (0.021)	0.289*** (0.022)	0.035** (0.016)	0.292*** (0.022)	0.298*** (0.022)	0.023 (0.017)
Cluster* Marketization index3	-0.007** (0.003)	0.000 (0.003)	-0.006*** (0.002)	-0.008** (0.003)	-0.005 (0.003)	-0.003 (0.002)	-0.008** (0.004)	-0.006* (0.003)	-0.001 (0.003)
Marketization index3	0.010*** (0.002)	0.010*** (0.002)	0.000 (0.001)	0.007*** (0.002)	0.007*** (0.002)	0.000 (0.001)	0.008*** (0.002)	0.008*** (0.002)	0.001 (0.001)
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	217,914	217,914	217,914	217,914	217,914	217,914	217,914	217,914	217,914
adj. R-sq	0.193	0.356	0.038	0.369	0.445	0.027	0.204	0.305	0.028

Note: To save space, we do not present all the control variables in the table; standard errors are clustered at the county-industry level.

*** $p < 0.01$.

** $p < 0.05$.

* $p < 0.1$.

Table A-3

Robustness check that controls 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}
$Cluster_{j, k, t}$	0.220*** (0.011)	0.259*** (0.011)	0.009 (0.009)	0.268*** (0.012)	0.276*** (0.012)	0.022** (0.009)	0.264*** (0.012)	0.275*** (0.012)	0.019** (0.010)
Dummies for the three largest industries in the county	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
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	233,780	233,780	233,780	233,780	233,780	233,780	233,780	233,780	233,780
adj. R-sq	0.199	0.360	0.038	0.381	0.452	0.027	0.208	0.310	0.028

Note: To save space, we do not present all the control variables in the table; Standard errors are clustered at the county-industry level.

*** $p < 0.01$.

** $p < 0.05$.

Table A-4a

Robustness check using the subsample of counties outside megacities.

	(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}
$Cluster_{j, k, t}$	0.223*** (0.011)	0.259*** (0.011)	0.013 (0.009)	0.270*** (0.012)	0.275*** (0.012)	0.025*** (0.009)	0.267*** (0.013)	0.276*** (0.013)	0.022** (0.010)
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	232,608	232,608	232,608	232,608	232,608	232,608	232,608	232,608	232,608
adj. R-sq	0.197	0.359	0.037	0.381	0.452	0.027	0.207	0.309	0.028

Note: To save space, we do not present all the control variables in the table; Standard errors are clustered at the county-industry level.

*** $p < 0.01$.

** $p < 0.05$.

Table A-4b

Robustness check that controls for the population of megacities.

	(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}
$Cluster_{j, k, t}$	0.221*** (0.011)	0.259*** (0.011)	0.010 (0.009)	0.268*** (0.012)	0.277*** (0.012)	0.022** (0.009)	0.264*** (0.012)	0.276*** (0.012)	0.019** (0.010)

(continued on next page)

Table A-4b (continued)

	(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}
Megacity population	-2.020*** (0.357)	-1.805*** (0.327)	0.179 (0.238)	-2.153*** (0.369)	-1.929*** (0.319)	0.020 (0.227)	-1.895*** (0.488)	-1.910*** (0.401)	0.264 (0.294)
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	233,780	233,780	233,780	233,780	233,780	233,780	233,780	233,780	233,780
adj. R-sq	0.198	0.359	0.037	0.381	0.452	0.027	0.207	0.309	0.028

Note: To save space, we do not present all the control variables in the table; Standard errors are clustered at the county-industry level.

*** $p < 0.01$.

** $p < 0.05$.

Table A-5

The correlation between the TFP and market share of firms inside and outside clusters.

Market share	TFP	TFP_OLS	TFP_OP1	TFP_OP2
The full sample		0.0560*	-0.0436*	-0.0495*
In clusters		0.2545*	0.0993*	0.0814*
in non-clusters		0.0954*	-0.0031*	-0.0085*

* $p < 0.1$.

Table A-6

The comparison of the TFP of firms exiting the ASIFP in clusters and non-clusters.

TFP measured by different means		Mean	p50	sd	min	max	N
TFP_OLS	In clusters	-0.02	0.05	1.29	-11.80	5.76	95,912
	In non-clusters	-0.47	-0.21	1.73	-13.65	8.04	190,764
TFP_OP1	In clusters	2.65	2.79	1.47	-11.05	9.19	95,912
	In non-clusters	2.06	2.35	1.94	-11.97	10.04	190,764
TFP_OP2	In clusters	2.42	2.58	1.63	-11.22	8.69	95,912
	In non-clusters	1.82	2.05	2.07	-12.84	9.73	190,764

Table A-7

The average gap between the TFP of surviving and exiting firms in clusters and non-clusters.

The average gap between the TFP measured by different means		Mean	p50	sd	min	max	N
TFP_OLS	In clusters	0.28	0.20	1.02	-9.27	11.89	13,800
	In non-clusters	0.42	0.25	1.67	-11.33	12.81	76,466
TFP_OP1	In clusters	0.36	0.25	1.07	-9.64	12.17	13,800
	In non-clusters	0.49	0.33	1.72	-11.24	12.62	76,466
TFP_OP2	In clusters	0.26	0.18	1.09	-10.10	11.89	13,800
	In non-clusters	0.39	0.23	1.77	-11.51	11.80	76,466

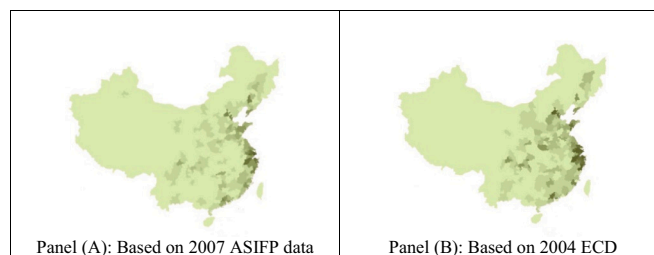


Fig. A-1. Distribution of the DBI clusters across Chinese counties.

References

Acemoglu, D., Akcigit, U., Alp, H., Bloom, N., Kerr, W., 2018. Innovation, reallocation, and growth. *Am. Econ. Rev.* 108 (11), 3450–3491.

Adamopoulos, T., Brandt, L., Leight, J., Restuccia, D., 2017. Misallocation, Selection and Productivity: A Quantitative Analysis With Panel Data From China (No. w23039).

Aghion, P., Howitt, P., 1992. A model of growth through creative destruction. *Econometrica* 60 (2), 323–351.

Allen, F., Qian, J., Qian, M., 2005. Law, finance, and economic growth in China. *J. Financ. Econ.* 77 (1), 57–116.

Arrow, K.J., 1962. The economic implications of learning by doing. *Rev. Econ. Stud.* 29 (3), 155–173.

- Arzaghi, M., Henderson, J.V., 2008. Networking off Madison avenue. *Rev. Econ. Stud.* 75 (4), 1011–1038.
- Asheim, B.T., Isaksen, A., 2002. Regional innovation systems: the integration of local 'sticky' and global 'ubiquitous' knowledge. *J. Technol. Transf.* 27 (1), 77–86.
- Atkeson, A., Burstein, A., 2008. Pricing-to-market, trade costs, and international relative prices. *Am. Econ. Rev.* 98 (5), 1998–2031.
- Au, C.C., Henderson, J.V., 2006. How migration restrictions limit agglomeration and productivity in China. *J. Dev. Econ.* 80 (2), 350–388.
- Audretsch, D.B., Feldman, M.P., 1996. Innovative clusters and the industry life cycle. *Rev. Ind. Organ.* 11 (2), 253–273.
- Bai, C.E., Hsieh, C.T., Song, Z.M., 2016. The Long Shadow of a Fiscal Expansion (No. w22801). National Bureau of Economic Research.
- Banerjee, A.V., Duflo, E., 2005. Growth theory through the lens of development economics. In: *Handbook of Economic Growth*, 1, pp. 473–552.
- Baqae, D.R., Farhi, E., 2020. Productivity and misallocation in general equilibrium. *Q. J. Econ.* 135 (1), 105–163.
- Baptista, R., Swann, P., 1998. Do firms in clusters innovate more? *Res. Policy* 27 (5), 525–540.
- Bartelsman, E., Haltiwanger, J., Scarpetta, S., 2013. Cross-country differences in productivity: the role of allocation and selection. *Am. Econ. Rev.* 103 (1), 305–334.
- Beaudry, C., Swann, G.P., 2009. Firm growth in industrial clusters of the United Kingdom. *Small Bus. Econ.* 32 (4), 409–424.
- Beeson, P.E., 1990. Sources of the decline of manufacturing in large metropolitan areas. *J. Urban Econ.* 28 (1), 71–86.
- Bernard, A.B., Jensen, J.B., Redding, S.J., Schott, P.K., 2009. The margins of US trade. *Am. Econ. Rev.* 99 (2), 487–493.
- Boschma, R., 2005. Proximity and innovation: a critical assessment. *Reg. Stud.* 39 (1), 61–74.
- Brandt, L., Van Biesebroeck, J., Zhang, Y., 2012. Creative accounting or creative destruction? Firm-level productivity growth in Chinese manufacturing. *J. Dev. Econ.* 97 (2), 339–351.
- Bustos, P., 2011. Trade liberalization, exports, and technology upgrading: evidence on the impact of MERCOSUR on argentinian firms. *Am. Econ. Rev.* 101 (1), 304–340.
- Caselli, F., 2005. Accounting for cross-country income differences. *Handbook of Economic Growth* 1, 679–741.
- Caselli, F., Coleman II, W.J., 2001. The US structural transformation and regional convergence: a reinterpretation. *J. Polit. Econ.* 109 (3), 584–616.
- Chang, C., Wang, Y., 1994. The nature of the township-village enterprise. *J. Comp. Econ.* 19 (3), 434–452.
- Che, J., Qian, Y., 1998. Insecure property rights and government ownership of firms. *Q. J. Econ.* 113, 467–496.
- Ciccone, A., Hall, R.E., 1996. Productivity and the density of economic activity. *Am. Econ. Rev.* 86, 54–70.
- Cingano, F., Schivardi, F., 2004. Identifying the sources of local productivity growth. *J. Eur. Econ. Assoc.* 2 (4), 720–742.
- Coleman, S., 1990. *Foundations of social theory*. Harvard University Press, Cambridge, MA.
- Collard-Wexler, A., De Loecker, J., 2015. Reallocation and technology: evidence from the US steel industry. *Am. Econ. Rev.* 105 (1), 131–171.
- Combes, P.P., Overman, H.G., 2004. The spatial distribution of economic activities in the European Union. In: *Handbook of Regional and Urban Economics*, Vol. 4. Elsevier.
- De Loecker, J., Warzynski, F., 2012. Markups and firm-level export status. *Am. Econ. Rev.* 102 (6), 2437–2471.
- Dekle, R., 2002. Industrial concentration and regional growth: evidence from the prefectures. *Rev. Econ. Stat.* 84 (2), 310–315.
- Delgado, M., Porter, M.E., Stern, S., 2014. Clusters, convergence, and economic performance. *Res. Policy* 43 (10), 1785–1799.
- Dunne, T., Roberts, M.J., Samuelson, L., 1988. Patterns of firm entry and exit in US manufacturing industries. *RAND J. Econ.* 19 (4), 495–515.
- Eberts, R.W., McMillen, D.P., 1999. Agglomeration economies and urban public infrastructure. In: *Handbook of Regional and Urban Economics*, Vol. 3. Elsevier.
- Edmond, C., Midrigan, V., Xu, D.Y., 2015. Competition, markups, and the gains from international trade. *Am. Econ. Rev.* 105 (10), 3183–3221.
- Edmond, C., Midrigan, V., Xu, D.Y., 2018. How Costly Are Markups? (No. w24800). National Bureau of Economic Research.
- Ellison, G., Glaeser, E.L., Kerr, W.R., 2010. What causes industry agglomeration? Evidence from coagglomeration patterns. *Am. Econ. Rev.* 100 (3), 1195–1213.
- Fan, G., Wang, X., Zhu, H., 2011. NERI Index of Marketization of China's Provinces (Zhongguo Shichang Jingji Zhishu). Economic Science Press, Beijing.
- Feldman, M.P., Audretsch, D.B., 1999. Innovation in cities: science-based diversity, specialization and localized competition. *Eur. Econ. Rev.* 43 (2), 409–429.
- Folta, T.B., Cooper, A.C., Baik, Y.S., 2006. Geographic cluster size and firm performance. *J. Bus. Ventur.* 21 (2), 217–242.
- Foster, L., Haltiwanger, J.C., Krizan, C.J., 2001. Aggregate productivity growth: lessons from microeconomic evidence. In: *New Developments in Productivity Analysis*. University of Chicago Press, pp. 303–372.
- Fu, Y., Gabriel, S.A., 2012. Labor migration, human capital agglomeration and regional development in China. *Reg. Sci. Urban Econ.* 4 (3), 473–484.
- Fujita, M., Krugman, P.R., Venables, A., 1999. *The Spatial Economy: Cities, Regions, and International Trade*. MIT Press.
- Fukuyama, F., 1995. Social capital and the global economy. *Foreign Aff.* 74, 89.
- Galor, O., Tsiddon, D., 1997. Technological progress, mobility, and economic growth. *Am. Econ. Rev.* 87 (3), 363–382.
- Gancia, G., Zilibotti, F., 2009. Technological change and the wealth of nations. *Ann. Rev. Econ.* 1 (1), 93–120.
- Glaeser, E.L., Kallal, H.D., Scheinkman, J.A., Shleifer, A., 1992. Growth in cities. *J. Polit. Econ.* 100 (6), 1126–1152.
- Globerman, S., Shapiro, D., Vining, A., 2005. Clusters and intercluster spillovers: their influence on the growth and survival of Canadian information technology firms. *Ind. Corp. Chang.* 14 (1), 27–60.
- Graham, D.J., 2007. Variable returns to agglomeration and the effect of road traffic congestion. *J. Urban Econ.* 62 (1), 103–120.
- Grashof, N., 2020. Firm-specific cluster effects: a meta-analysis. *Pap. Region. Sci.* 99 (5), 1237–1260.
- Greenstone, M., Hornbeck, R., Moretti, E., 2010. Identifying agglomeration spillovers: evidence from winners and losers of large plant openings. *J. Polit. Econ.* 118 (3), 536–598.
- Grillitsch, M., Nilsson, M., 2017. Firm performance in the periphery: on the relation between firm-internal knowledge and local knowledge spillovers. *Reg. Stud.* 51 (8), 1219–1231.
- Grillitsch, M., Nilsson, M., 2019. Knowledge externalities and firm heterogeneity: effects on high and low growth firms. *Pap. Reg. Sci.* 98 (1), 93–114.
- Guo, D., Jiang, K., Kim, B.Y., Xu, C., 2014. Political economy of private firms in China. *J. Comp. Econ.* 42 (2), 286–303.
- Guo, D., Jiang, K., Xu, C., Yang, X., 2020. Clustering, growth and inequality in China. *J. Econ. Geogr.* 20 (5), 1207–1239.
- Hall, R.E., Jones, C.I., 1999. Why do some countries produce so much more output per worker than others? *Q. J. Econ.* 114 (1), 83–116.
- Henderson, J.V., 2003. Marshall's scale economies. *J. Urban Econ.* 53 (1), 1–28.
- Henderson, V., Kuncoro, A., Turner, M., 1995. Industrial development in cities. *J. Polit. Econ.* 103 (5), 1067–1090.
- Hervas-Oliver, J.L., Sempere-Ripoll, F., Rojas Alvarado, R., Estelles-Miguel, S., 2018. Agglomerations and firm performance: who benefits and how much? *Reg. Stud.* 52 (3), 338–349.
- Hicks, J.R., 1932. Marginal productivity and the principle of variation. *Economica* 35, 79–88.
- Hicks, J.R., Allen, R.G., 1934. A reconsideration of the theory of value. Part I. *Economica* 1 (1), 52–76.
- Howell, A., 2017. Marshallian sources of relatedness and their effects on firm survival and subsequent success in China. *Econ. Geogr.* 93 (4), 346–366.
- Hsieh, C.T., Klenow, P.J., 2009. Misallocation and manufacturing TFP in China and India. *Q. J. Econ.* 124 (4), 1403–1448.
- Huang, Z., Zhang, X., Yunwei, Z.H.U., 2008. The role of clustering in rural industrialization: a case study of the footwear industry in Wenzhou. *China Econ. Rev.* 19 (3), 409–420.
- Jacobs, J., 1969. *The Economy of Cities*. Random House, New York.
- Jofre-Monseny, J., 2009. The scope of agglomeration economies: evidence from Catalonia. *Pap. Reg. Sci.* 88 (3), 575–590.
- Ke, S., 2010. Agglomeration, productivity, and spatial spillovers across Chinese cities. *Ann. Reg. Sci.* 45 (1), 157–179.
- Kerr, S.P., Kerr, W., Özden, Ç., Parsons, C., 2017. High-skilled migration and agglomeration. *Annu. Rev. Econ.* 9, 201–234.
- Klenow, P.J., Rodriguez-Clare, A., 1997. Economic growth: a review essay. *J. Monet. Econ.* 40 (3), 597–617.
- Knack, S., Keefer, P., 1997. Does social capital have an economic payoff? *Q. J. Econ.* 112, 1251–1288.
- Knoben, J., Arikian, A., van Oort, F., Raspe, O., 2016. Agglomeration and firm performance: one firm's medicine is another firm's poison. *Environ. Plan. A* 48 (1), 132–153.
- Kochhar, K., Kumar, U., Rajan, R., Subramanian, A., Tokatlidis, I., 2006. India's pattern of development: what happened, what follows? *J. Monet. Econ.* 53 (5), 981–1019.
- Krugman, P.R., 1991. *Geography and Trade*. MIT Press.
- Lee, C.Y., 2018. Geographical clustering and firm growth: differential growth performance among clustered firms. *Res. Policy* 47 (6), 1173–1184.
- Lerner, A., 1934. The concept of monopoly and the measurement of monopoly power. *Rev. Econ. Stud.* 1 (3), 157–175.
- Lewis, W., 2005. *The Power of Productivity*. University of Chicago Press.
- Long, C., Zhang, X., 2011. Cluster-based industrialization in China: financing and performance. *J. Int. Econ.* 84 (1), 112–123.
- Lu, Y., Yu, L., 2015. Trade liberalization and markup dispersion: evidence from China's WTO accession. *Am. Econ. J. Appl. Econ.* 7 (4), 221–253.
- Marshall, A., 1890. *Principles of Economics*. Macmillan, London.
- McCann, B.T., Folta, T.B., 2008. Location matters: where we have been and where we might go in agglomeration research. *J. Manag.* 34 (3), 532–565.
- Melitz, M.J., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica* 71 (6), 1695–1725.
- Meng, X., 2000. *Labour market reform in China*. Cambridge University Press.
- Myles Shaver, J., Flyer, F., 2000. Agglomeration economies, firm heterogeneity, and foreign direct investment in the United States. *Strateg. Manag. J.* 21 (12), 1175–1193.
- Olley, S., Pakes, A., 1996. The dynamics of productivity in the telecommunications equipment industry. *Econometrica* 64, 1263–1298.
- Opp, M.M., Parlour, C.A., Walden, J., 2014. Markup cycles, dynamic misallocation, and amplification. *J. Econ. Theory* 154, 126–161.
- Pavcnik, N., 2002. Trade liberalization, exit, and productivity improvements: evidence from Chilean plants. *Rev. Econ. Stud.* 69 (1), 245–276.
- Porter, M., 1990. *The Competitive Advantage of Nations*. Macmillan, London.
- Porter, M., 2003. The economic performance of regions. *Reg. Stud.* 37 (6–7), 549–578.
- Prevezer, M., 1997. The dynamics of industrial clustering in biotechnology. *Small Bus. Econ.* 9 (3), 255–271.

- Putnam, P., 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton University Press, Princeton, NJ.
- Qian, Y., Xu, C., 1993. Why China's economic reforms differ: the M-form hierarchy and entry/expansion of the non-state sector. *Econ. Transit.* 1 (2), 135–170.
- Restuccia, D., Rogerson, R., 2008. Policy distortions and aggregate productivity with heterogeneous establishments. *Rev. Econ. Dyn.* 11 (4), 707–720.
- Rosenthal, S.S., Strange, W.C., 2003. Geography, industrial organization, and agglomeration. *Rev. Econ. Stat.* 85 (2), 377–393.
- Ruan, J., Zhang, X., 2009. Finance and cluster-based industrial development in China. *Econ. Dev. Cult. Chang.* 58 (1), 143–164.
- Schmitz Jr., J.A., 2005. What determines productivity? Lessons from the dramatic recovery of the US and Canadian iron ore industries following their early 1980s crisis. *J. Polit. Econ.* 113 (3), 582–625.
- Schumpeter, J.A., 1950. The march into socialism. *Am. Econ. Rev.* 40 (2), 446–456.
- Smart, A., Smart, J., 2001. Local citizenship: welfare reform urban/rural status, and exclusion in China. *Environ Plan A* 33 (10), 1853–1869.
- Solow, R.M., 1957. Technical change and the aggregate production function. *Rev. Econ. Stat.* 39 (3), 312–320.
- Song, Z., Xiong, W., 2018. Risks in China's financial system. *Ann. Rev. Financ. Econ.* 10, 261–286.
- Speldekamp, D., Knoblen, J., Saka-Helmhout, A., 2020. Clusters and firm-level innovation: a configurational analysis of agglomeration, network and institutional advantages in European aerospace. *Res. Policy* 49 (3), 103921.
- Staber, U., 2001. Spatial proximity and firm survival in a declining industrial district: the case of knitwear firms in Baden-württemberg. *Reg. Stud.* 35 (4), 329–341.
- Stuart, T., Sorenson, O., 2003. The geography of opportunity: spatial heterogeneity in founding rates and the performance of biotechnology firms. *Res. Policy* 32 (2), 229–253.
- Teets, J.C., 2009. Post-earthquake relief and reconstruction efforts: the emergence of civil society in China? *China Q.* 198, 330–347.
- Van Geenhuizen, M., Reyes-Gonzalez, L., 2007. Does a clustered location matter for high-technology companies' performance? The case of biotechnology in the Netherlands. *Technol. Forecast. Soc. Chang.* 74 (9), 1681–1696.
- Van Oort, F.G., Stam, E., 2006. Agglomeration economies and entrepreneurship in the ICT industry. In: ERIM Report Series Reference No. ERS-2006-016-ORG.
- Weitzman, M., Xu, C., 1994. Chinese township village enterprises as vaguely defined cooperatives. *J. Comp. Econ.* 18 (2), 121–145.
- Xu, C., 2011. The fundamental institutions of China's reforms and development. *J. Econ. Lit.* 49 (4), 1076–1151.
- Xu, C., Zhang, X., 2009. The evolution of Chinese entrepreneurial firms: Township-village enterprises revisited. In: Wu and Yao (Ed.), *Reform and Development in China*. Routledge, London and New York.
- Zaheer, A., George, V.P., 2004. Reach out or reach within? Performance implications of alliances and location in biotechnology. *Manag. Decis. Econ.* 25 (6–7), 437–452.