China Economic Review 93 (2025) 102505

Contents lists available at ScienceDirect

China Economic Review

journal homepage: www.elsevier.com/locate/chieco

Industrial policies and development of grain clusters: Evidence from China's policy in major grain-producing areas[☆]

Yinze Zhang, Jianqing Ruan [*], Yuwei Zhang

China Academy for Rural Development, Zhejiang University, Yuhangtang Road No.866, 310058 Hangzhou, Zhejiang, China

ARTICLE INFO

JEL classification:
R32
O25
Q13

Keywords:
Grain clusters
Industrial clusters
Industrial policies
Major grain-producing areas

ABSTRACT

Grain shortage remains a prominent concern in many developing countries. While clusters are recognized as effective forms of industrial organization, there is limited empirical studies regarding on which industrial policies effectively promote the development of grain clusters. This study addresses this gap by using a comprehensive dataset CCAD, which covers nearly all agro-related enterprises registered in China, through which we identified the enterprises in each link of the grain industry chain and provided a detailed account of the evolution of grain clusters. Employing a difference-in-differences strategy, we empirically evaluated the impact of China's policy in major grain-producing areas (MGPA) on grain clusters. Our findings indicate a significant contribution of the policy in MGPA to the development of grain clusters, and its impact varies across different links in the grain industry chain. Subsequent analyses reveal that the impact is driven by several mechanisms, including government fiscal support, grain subsidies, profit expectation and the the number and scale of entered grain enterprises. This study also finds that in areas where cooperatives and family farms are fully developed, policy effects are more pronounced. These findings highlight the crucial impact of industrial policies on shaping grain cluster development.

1. Introduction

Grain shortage remains a prominent issue in most of the less developed countries and their governments are looking for effective ways to improve the efficiency of grain production (Fan, 2021). Grain clusters are one possible channel. Porter (1990) defines industrial clusters as geographic concentrations of interconnected companies and institutions in a particular field that enhance the competitive advantage of regions. In the manufacturing industry, clusters have been proved to be an important form of industrial organization for improving production efficiency due to their scale, competition, division of labor and network effects (e.g., Dai et al., 2021; Long & Zhang, 2011; Marshall, 1920; Ruan & Zhang, 2009). These studies on the advantages of manufacturing industrial clusters have extended to agriculture and found that agricultural clusters have the same benefits. For example, they can improve production efficiency, promote innovation among farmers, narrow the rural income gap, and have a certain positive externality on nearby enterprises (e.g., Diez-Vial, 2011; Hailu & James Deaton, 2016; Hu et al., 2019; Jr Tabe-Ojong & Dureti, 2023). Similar to most agricultural products, grain production in the vast majority of countries also has a characteristic of geographic concentration and

[☆] This article is part of a Special issue entitled: 'Agriculture cluster, trade and supply chain resilience' published in China Economic Review.

[*] Corresponding author.
E-mail addresses: yinzezhang@zju.edu.cn (Y. Zhang), ruanjq@zju.edu.cn (J. Ruan).

https://doi.org/10.1016/j.chieco.2025.102505
Received 12 November 2024; Received in revised form 3 June 2025; Accepted 11 July 2025
Available online 11 July 2025

formed de facto grain clusters.[1,2] Due to the high transportation cost of grain and its low price, the clusters bring competitiveness to grain products. Therefore, the question of what measures should be taken to promote clusters has also received attention from economists.

The reasons for the formation and development of industrial clusters have been extensively explored in the academic community. However, such research has not yet explored the grain clusters, nor has it examined the effects of industrial policies from the perspective of the industry chain. Initially, Ellison and Glaeser (1997) considered industrial clusters to be the result of natural advantage and spillovers. Since then there have been many explorations in the literature along the two lines. For natural advantage, Kim (1999) found that factor endowment explains the significant geographic distribution of manufacturing activities, and the empirical evidence provided by Ellison and Glaeser (1999) indicates that about 20 % of observed geographic concentration can be explained by a small set of advantages. However, natural advantage does not always play a role. For example, the impact of land availability (as a natural advantage) on industrial clusters varies across firms with different levels of land dependency (Zheng, 2023). The aforementioned literature on the formation of industrial clusters through natural advantage is essentially an extension of location theory (Hoover, 1936; Weber, 1909). This theory emphasizes how geographic factors, such as resource availability and transportation costs, influence the spatial concentration of economic activities.

For spillovers, both physical spillovers and knowledge-skill spillovers from geographic concentration explain the development of industrial clusters, such as reduction in transportation costs (Krugman, 1991a) and exchange of technical information (Guiso & Schivardi, 2007; Moretti, 2021; Saxenian, 1994). These spillovers promote coagglomeration among firms (Ellison et al., 2010) and serve as the basis for clusters. Guiso and Schivardi (2011) found that external effects similar to spillovers also lead to the concentration of economic activities within the region. One exception is Bischi et al. (2003), which suggests that the attraction basins of coexisting long run equilibria do not depend on the size of the spillovers. In addition to this, crisis-driven quality upgrading, globalization[3] and clan culture have also been cited as important reasons for the development of industrial clusters (Fan, Li, et al., 2023; Fan, Mishra, et al., 2023; Ge, 2009; Hu et al., 2021). However, the relevant literature is not systematic. Overall, spillover effects arise from the geographic proximity of firms, leading to reduced costs and enhanced innovation through shared knowledge. These effects are crucial for the development of industrial clusters, as they foster collaboration and competition, driving regional economic growth.

A third widely studied pathway for clusters formation and development is industrial policy. A large body of literature suggests that appropriate policies are necessary for the development of industrial clusters, such as land incentive policies and place-based policies (e.g., China's special economic zones) will promote the emergence of the agglomeration economy (Lu et al., 2019; Zheng, 2023), and inappropriate industrial policies (e.g., migration restrictions) will hinder the development of industrial clusters (Au & Henderson, 2006). There is also a large amount of literature that explores which industrial policies are beneficial for clusters (e.g., Forslid & Midelfart, 2005; Nathan & Overman, 2013; Rodríguez-Clare, 2007), providing us with important references. In less developed countries, industrial policies are also crucial for agricultural clusters (Gálvez-Nogales, 2010). Taking China as an example, actions taken by the local government to overcome production bottlenecks and establish industrial organizations have promoted the development of tea clusters (Zhao et al., 2021). Policies such as organizing demonstration training courses and negotiations to reduce communication costs by township governments have stimulated the formation of agricultural machinery service clusters (Zhang et al., 2017). In Ethiopia, the government has also actively implemented various strategies to establish Farmer Production Clusters (Jr Tabe-Ojong & Dureti, 2023).

Today, for most developing countries, grain is the most important part of agricultural production. Considering the many advantages of the industrial clusters development pathway, ensuring food security through grain clusters has become urgent. In fact, governments of developing countries like China have taken measures to encourage the establishment of characteristic grain clusters.[4] However, the reasons for the formation of grain clusters are still poorly understood to scholars. An exception is Zhang and Hu (2014), which scrutinizes the potato cluster in Anding County, China, and examines the role of local industrial policies in clusters development, but empirical evidence is still lacking. Although natural advantage explains most of the causes of grain clusters, since grain industry is highly dependent on natural resources such as temperature, precipitation and land, natural advantage is often exogenous to the socio-economic and difficult to shift, so the role of industrial policies cannot be ignored. Additionally, grain is the most fundamental material for human life, being irreplaceable and has a long growth cycle, which implies the impact of industrial policies on other agricultural product clusters cannot be simply transferred to the grain industry. Based on the above facts, how do current industrial policies in less developed countries affect the development of grain clusters? What are the impacts on different production links in the grain cluster? What are the pathways through which it exerts influence? Is the effect heterogeneous across different contexts? These problems urgently need to be addressed.

Therefore, the objective of this study is to systematically analyze the impact of industrial policies on grain clusters. This article selects the policy in major grain-producing areas (hereafter MGPA), which was implemented by the Chinese government in 2004 with the aim of ensuring grain production, as the research object. We chose this industrial policy for two reasons: firstly, since it directly affects China's grain industry and has a close relationship with grain clusters, it can be considered an important measure to stimulate

[1] We adopted a broad definition of grain such as cereals, legume and tubers for food consumption.

[2] For example, wheat in the United States is mainly concentrated in the three major wheat belts. China's grain crops are mainly produced by nine major commodity grain bases.

[3] Hsu et al. (2023) presented opposing evidence, arguing that FDI has intensified industrial competition pressure and is not conducive to the formation of industrial clusters.

[4] Source: Nation Food and Strategic Reserves Administration (NFSRA), http://www.lswz.gov.cn/html/zcfb/2019-08/30/content_246391.shtml.

the formation and development of grain clusters. In addition, one of the important directions of China's agricultural policies is to support MGPA to develop grain clusters based on county areas.[5] Secondly, the policy in MGPA is a special regional agricultural development strategy with the characteristic of adapting to local conditions (Hua et al., 2022). Therefore, this industrial policy can be regarded as a quasi-natural experiment to achieve specialized grain production in certain provinces, which facilitates the selection of an identification strategy in this study.

This paper contributes to two strands of the literature. First, it provides evidence on the key role of industrial policies in shaping and developing grain clusters. The purpose of industrial policies is to promote the development and competitiveness of specific industries, but their effectiveness has been challenged. Some studies argue that industrial policies are effective, especially in developing countries, where short-term industrial policies have contributed to their rise (Goldberg et al., 2024; Lin, 2011; Sylla, 2024). However, other studies oppose industrial policies due to challenges such as difficulties in government information collection and resource misallocation (Bartelme et al., 2019; Barwick et al., 2024; Juhász & Lane, 2024). This study supports the viewpoint of Lin (2011), using specific the policy in MGPA in the agricultural sector to analyze their impact on agricultural clusters, and argues that industrial policies aligned with comparative advantages can promote industrial development.

Second, this study contributes to the literature on measuring industrial clusters. This study is the first to identify enterprises at different links of the grain industry chain based on their operation scope and use this to construct an grain clusters indicator. Although most existing studies construct industrial cluster indicators based on standard industry classifications (SIC) (e.g., Ellison & Glaeser, 1997), SIC often fails to capture the connections between enterprises and industrial chain links (Dai et al., 2021; Porter, 1998), and therefore cannot reflect the important feature of "enterprise linkages" in industrial clusters. As a result, this study innovatively uses text analysis techniques to capture the specific operation scope of enterprises, determine whether they belong to the grain industry chain, and identify which link they belong to. Furthermore, compared to commonly used indicators such as location quotients (LQ), Density-based index (DBI), and spatial GINI coefficients (Glaeser et al., 1992; Guo et al. (2020); Krugman, 1991b), this paper uses the enterprise density indicator, which better captures the "spatial concentration" and "large scale" characteristic of industrial clusters and allows calculation at the county $\times$ link level in each year.

The remaining parts of this study are organized as follows. Section 2 provides an overview of the institutional background and basic facts of China's policy in MGPA, and proposes a theoretical analysis of the impact of this policy on grain clusters. Section 3 introduces the comprehensive dataset CCAD used in this study, and explains the variables and identification strategies. Section 4 presents empirical results and robustness tests, while section 5 provides the results of mechanisms analysis. Finally, the paper concludes with a comment on the entire article.

2. Policy background and theoretical analysis

2.1. China's policy in major grain-producing areas

The major grain-producing areas (MGPA) bear the important task of maintaining China's food security. Before the implementation of the policy in MGPA, grain production had been declining year by year from 1999 to 2003, with inadequate grain yield for needs. In 2003, the sown area of grain had decreased to the lowest level since the founding of China, almost shaking the foundation of China's food security. In this situation, China urgently adopted industrial policies to stabilize grain production.

On December 3rd, 2003, the Ministry of Finance of China issued the *Opinions on Reforming and Improving Several Policy Measures for Agricultural Comprehensive Development*, which explicitly designated 13 provinces as the country's major grain-producing areas.[6] It is worth mentioning that although the document was issued at the end of 2003, nearly all of the supporting policies began to take effect in 2004, and the agricultural economists generally considers 2004 as the starting year of the policy (Luo et al., 2020; Wei et al., 2023). Most of the MGPA provinces are located in plain or shallow hilly areas (the geographic distribution of MGPA provinces in China is shown in Fig. 1). Due to the humid or semi humid climate, abundant rainfall, good conditions for light, heat, and water resources, high soil organic matter content, easy cultivation and soil and water conservation, suitable for crop growth, these provinces possess strong integrated production capacity and great production potential. In addition, China also has 7 major grain-selling areas and 11 producing-selling-balance areas, and these areas have poor agricultural production conditions. According to Lin (2011), the key to the success of industrial policies lies in utilizing regional comparative advantages, and MGPA have unique advantages in developing the grain industry, therefore it can be considered as a positive industrial policy.

The policy in major grain-producing areas encompasses a range of policy tools and is, in fact, a package of policies targeted at these 13 provinces. Based on the food and agriculture policy classification of FAO, Hua et al. (2022) briefly summarized the policies in the MGPA into three types: production support, market management and natural resources management. Du and Mao (2017) summarized relevant policies aimed only at the MGPA as high-quality grain industry policies, major grain-producing counties reward policies, large-scale commodity grain bases policies, and core grain production areas policies. Previous studies have found that the policy in MGPA has promoted the improvement of local agricultural TFP and have had positive spillover effects on surrounding provinces (Ye et al., 2023). The policy in MGPA has a long-run effect of reducing the amount of fertilizers and pesticides (Ji et al., 2023; Luo et al.,

[5] Source: *Economic Daily*, http://paper.ce.cn/jjrb/html/2019-06/22/content_394113.htm.

[6] The 13 MGPA provinces are Heilongjiang, Jilin, Liaoning, Inner Mongolia, Hebei, Henan, Shandong, Jiangsu, Anhui, Jiangxi, Hubei, Hunan, and Sichuan. Source: Ministry of Finance of the People's Republic of China, https://www.mof.gov.cn/gp/xxgkml/gjnyzhkfbgs/200806/t20080625_2502826.htm.

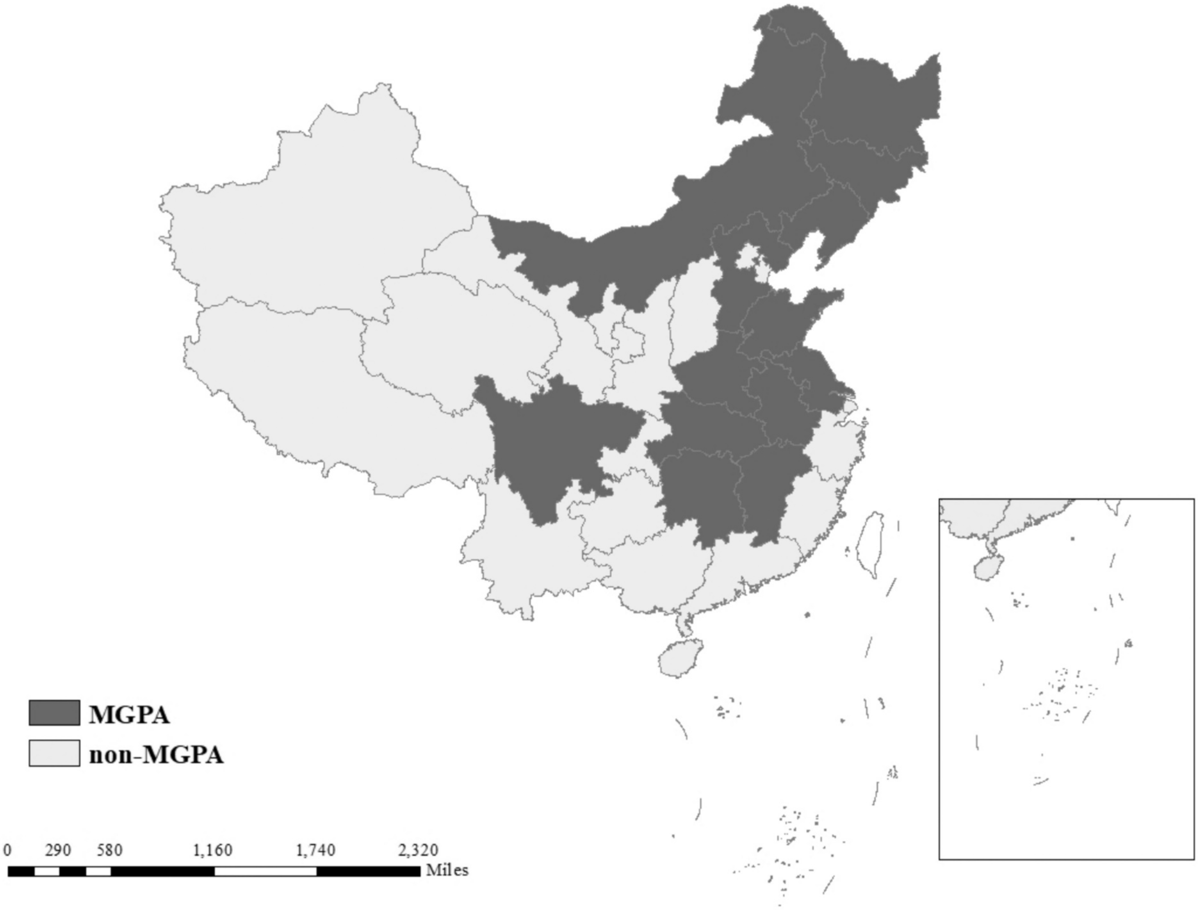

Fig. 1. China's major grain-producing areas. Note: The 13 MGPA provinces are Heilongjiang, Jilin, Liaoning, Inner Mongolia, Hebei, Henan, Shandong, Jiangsu, Anhui, Jiangxi, Hubei, Hunan, and Sichuan.

2020). Therefore, this policy has positive implications for both economic and environmental performance.

The major grain-producing areas have made significant contributions to increasing China's grain production, and they are important foundations for ensuring food security. The special political positioning and unique resource advantages make the MGPA highly concentrated in grain production activities. In the long run, stabilizing the MGPA's production capacity greatly strengthens China's overall food security, which is the core significance of establishing the MGPA. Therefore, under the pressure of food security, the main goal of agricultural development in the MGPA is to increase production. Since 2004, China's grain production has achieved a gratifying "twenty consecutive years of growth", increasing from 431 million tons in 2003[7] to 695 million tons in 2023,[8] with a growth rate of 61.3 %. The MGPA have made significant contributions. However, the MGPA often have underdeveloped economies, and the problem of economic development gap between them and the major grain-selling areas is very prominent.

To sum up, the policy in MGPA is an industrial policy for regions with good grain production conditions and suitable for the development of the grain industry, and it is characterized by the promotion of the agglomeration of grain production activities. Therefore, the policy in MGPA may have the effect of stimulating the development of grain clusters, which is the focus of this study.

2.2. Theoretical analysis: China's policy in MGPA's impact on grain clusters

Geographic concentration and enterprises association are key characteristics of industrial clusters. The geographic concentration will generate scale effect, while the enterprises association will generate division of labor effect (Krugman, 1991b; Marshall, 1920; Ruan & Zhang, 2009). Therefore, the size of regional grain production scale and whether there is collaborative cooperation among related enterprises in the grain industry chain are important indicators of the development of grain clusters. This section attempts to explain through what pathways the policy in MGPA has stimulated grain economies of scale and facilitated the agglomeration of

[7] Source: National Bureau of Statistics (NBS), https://www.stats.gov.cn/xxgk/sjfb/tjgb2020/201310/t20131031_1768610.html.
[8] Source: National Bureau of Statistics (NBS), https://www.stats.gov.cn/sj/zxfb/202402/t20240228_1947915.html.

different links of enterprises in grain industry within the region.

The development of grain industry clusters is reflected in two aspects: the increase in grain sown area and the growth in density of grain enterprises. Therefore, there are two respective mechanisms. First, a series of policy measures focusing on agricultural fiscal support have expanded the grain sown area. In terms of production support policies, the reward for Dafengshou County was added to the package of the policy in MGPA in 2005.[9] In addition, when a county's yield ranks among the top 100 in all MGPA counties, the county will receive additional incentive subsidies from the central government (Hua et al., 2022). As for agricultural risk management policies, a subsidy policy for agricultural disaster prevention and reduction was implemented in 2012. The central government has implemented for the first time a subsidy policy for agricultural disaster prevention and reduction, providing subsidies for six key technologies in the production of winter wheat, northern corn, and southern early rice. Abdur and Wang (2014) found that these policies have played an important role in helping farmers restore production and livelihoods. The above analysis indicates that fiscal support or subsidies are the main form of a package of policy in MGPA. In China, small-scale farming is the main form of grain production, and maximizing profits is the best reference framework for analyzing the behavior of farmers (Moro & Sckokai, 2013). We summarize the expansion of the scale of grain sown by farmers into two ways: The first is to adjust the planting structure to allocate more land for grain cultivation; the second is to reclaim land by converting undeveloped marginal land into arable land for grain production. When the government subsidizes grain cultivation, the marginal cost of grain production will decrease, and farmers will choose to invest more production factors (such as land) into grain production to achieve maximum profits, thereby expanding the area of grain sown. Luo and Zhang (2024) have shown that fiscal incentives are important factors in increasing the enthusiasm of farmers for grain cultivation, especially for developing countries that agriculture is dominated by small-scale farming. In addition, existing empirical studies have also shown that grain subsidies, and an important form of agricultural fiscal support, have greatly expanded the grain sown area. For example, Yi et al. (2015) found that the grain subsidy program improved the grain sown area of households with liquidity constraints. Fan, Li, et al., 2023, Fan, Mishra, et al. (2023) revealed that the new grain subsidy policy encourages farmers to increase the area of grain crops by leasing more land and increasing the share of grain crop cultivation. Guo et al. (2021) indicate raising subsidies for soybean producers will encourage farmers to allocate more land for soybean cultivation, which can be seen as a policy tool to promote soybean production in China. These indicates that subsidies are an effective way to promote large-scale grain production.

The second mechanism is that the policy increases the number and scale of local grain enterprises, thereby enhancing the density of grain enterprises. Specifically, the number of grain enterprises reflects the entry and exit decisions of firms, representing an increase at the extensive margin; while the scale of grain enterprises reflects their production and operational decisions, representing an increase at the intensive margin. Both types of marginal increases can be attributed to the direct or indirect effects of the policy. In terms of direct effects, the implementation of the policy in MGPA will bring higher profit expectations to entrepreneurs, promoting them to locate in the MGPA and leading to an increase in the density of enterprises in each link of the grain industry chain and the formation of grain clusters in the sense of the extensive margin. At the same time, profit expectations encourage entrepreneurs to invest more capital and other factors of production, which increases the scale of new entrants and promotes the development of grain clusters. As regards indirect effects, the increase in enterprise density also benefits from the expansion of grain sown area, i.e. the policy in MGPA promotes cluster development through expansion of grain sown area. Specifically, for prenatal link enterprises such as seed, fertilizer&pesticide, and agricultural machinery enterprises, the expansion of grain sown scale has created demand and increased enterprise revenues for them. For grain planting enterprises and agricultural machinery service enterprises, expanding the sown area provides them with input factors and service targets respectively. However, agricultural machinery service enterprises have the characteristic of cross regional services, and the impact of sown area on the density of such enterprises may be limited. For grain processing enterprises and grain sales enterprises, large-scale grain production provides them with adequate supply. Given that the expansion of sown area brings many benefits to grain enterprises, enterprises in the grain industry chain will prioritize entering MGPA to carry out business activities. In addition, the expansion of grain sown area and the expansion of grain enterprise scale are almost simultaneous, that is, the policy has had both extensive and intensive marginal effects by expanding the grain area sown and promoting the development of grain clusters. In general, when enterprises from different links in the grain industry chain agglomerated in a region, it will generate economies of scale and division of labor, promoting the development of industrial clusters.

3. Data, variables and identification strategy

3.1. Grain clusters

With reference to Porter (1990), we define a grain cluster as geographic concentration of interconnected grain enterprises and other institutions within a specific region. This study investigates the effects of industrial policies on grain clusters, therefore we have sorted out the development of China's grain industry chain through a unique dataset and obtained data on China's grain clusters. This dataset is derived from CCAD,[10] which almost cover all agro-related enterprise registered with China's State Administration for Market

[9] Specifically, when the average annual grain output of a county in the past five years exceeded 2 million tons and the commodity grain output exceeded 5 thousand tons, the county was considered as a Dafengshou county.

[10] China Academy for Rural Development-Qiyan China Agri-research Database (CCAD) is an agricultural research database that serves agriculture and related academic research, including 9 sub databases such as "New Agricultural Operating Entitles", "Agricultural and Related Industry Entities", and "Agricultural Products and Production Factors".

Regulation, including information from over 6.5 million agro-related enterprises. The database clearly displays variables such as the name, region, longitude and latitude, operate scope, registration time, and death time of each enterprise. After obtaining this data, our next step is to identify grain enterprises from agro-related enterprises.

CCAD provides for the basic information of agro-related enterprises, which lists the operational scope of enterprises, which includes the categories of goods, varieties, and service items that are permitted for production and operation in China. From this variable, we can determine whether the enterprise belongs to the grain industry chain and which link in the grain industry chain it belongs to. Specifically, We applied text analysis techniques to identify enterprises in eight key links of the grain industry chain within the CCAD database. These sectors include:

- **Seed**: covering research and development (R&D), manufacturing, and sales.
- **Fertilizer&Pesticide**: covering R&D, manufacturing, and sales.
- **Machine**: focusing on agricultural machinery manufacturing and sales.
- **Planting**: covering grain planting activities.
- **Service**: including agricultural machinery services.
- **Rough processing**: referring to grain rough processing.
- **Deep processing**: referring to grain deep processing.
- **Sale**: including grain purchasing, storage, and sales.

The details of the text analysis method can be found in the appendix A. Fig. 2 shows the grain industry chain and some specific types of grain enterprises in our analysis.

After all grain enterprises were labeled, we calculated the stock of number of grain enterprises for each year form 1992 to 2022. Due to the regional concept of industrial clusters, for counties with the same area, the more enterprises there are, the closer the relative distance between them, and the higher the degree of clustering. Therefore, we use the county-link level grain enterprises density as a proxy variable for the grain clusters,[11] and the formula for the enterprise density of grain industry link l in county c of province p in period t is as follow:

$$Density_{pclt} = \frac{Stock_{pclt}}{Area_{pc}} \tag{1}$$

Where $Area_{pc}$ is the area of county c in province p. Thus, we obtain balanced panel data on the enterprise density of 8 types of links of grain industry chain in 2840 counties from 1992 to 2022. The geographical distributions of the grain industry density in counties are shown in Fig. 3. In the time dimension, the densities of grain enterprises are increasing in almost all regions of China, suggesting that grain production in China is being industrialized; in the regional dimension, the densities of enterprises in MGPA are higher than the densities of enterprises in non-MGPA in any given year; examined in combination with the time and spatial dimensions, the gap in the density of grain enterprises between MGPA and non-MGPA is widening over time, and this gap is particularly pronounced in 2012 and 2022, which may suggest the role of the policy in MGPA.

It should be noted that most links in the aforementioned grain industry chain are primarily driven by enterprises; however, in the planting link, there are other entities involved in addition to enterprises, which means the density of enterprises alone cannot fully reflect the situation of the grain cluster. Therefore, we supplemented it with the variable of grain sown area ($Grain_{pct}$), which comes from China County-Level Economic Research Database (CNT). After doing so, other entities (such as small households, family farms, and agricultural cooperatives) have been included as well. Due to missing observations in the database, we only obtained unbalanced panel data of 1372 counties from 2001 to 2021.

3.2. Control variables and mechanisms

Given that the policy in MGPA are formulated at the province level, this study will also select a series of province level characteristics for control. The reason is that the confounding factors that cause selection bias in estimation results need to simultaneously affect both the outcome variable and the treatment variable, which means that the control variables should also be at least province level. Based on this, we use variables from the China Regional Economy Research Database to address potential omitted variable bias.

The level of regional economic development is an important factor affecting the location selection of enterprises. Existing literature has shown that economic conditions can effectively predict enterprise growth and cluster formation (Bu & Liao, 2022; Guo et al., 2023). Therefore, we have included the per capita GDP of provinces in the specification. In addition, the industrial foundation is also valued by enterprises. For grain enterprises, agricultural products are their production objects or raw materials, so we use the proportion of the primary industry as a proxy for the industrial foundation. Transportation, as a natural advantage, can reduce production costs and promote enterprise agglomeration (Ellison & Glaeser, 1999). For example, bulk commodities such as fertilizers and agricultural machinery need to be transported into the MGPA, and cross regional agricultural machinery services also attach importance to good road conditions. In addition, grain is perishable, and fast transportation conditions are a prerequisite for ensuring smooth processing and sales of grain, so that it is necessary to control the mileage of railways and highways in the model.

[11] This measurement is also used by Moretti (2021).

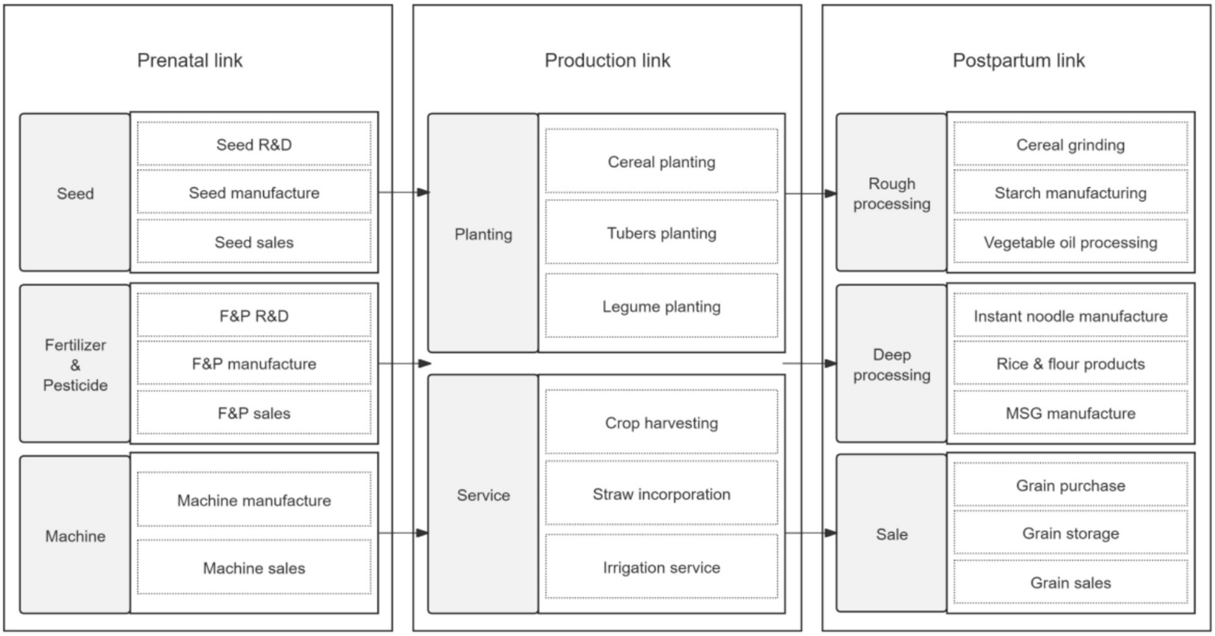

Fig. 2. Grain industry chain and types of grain enterprises.

The mechanism through which the policy in MGPA expands grain sown area is the increase in agricultural fiscal expenditures and the rise in subsidies for farmers' grain planting. First, we manually collected data on agricultural fiscal expenditures in province-level from 2003 to 2022. Second, there is a limited amount of micro-level data on grain planting subsidies for farmers, and the vast majority of data does not cover periods prior to 2004. The only dataset that meets the requirements for our DID design is the Fixed Observation Points System (FOPS) from the Research Center for Rural Economy (RCRE), Ministry of Agriculture of China. Therefore, we use the 2002–2009 farmer sample for analysis, as grain subsidy data for these years is relatively complete. In detail, we aggregated the Direct subsidy, Comprehensive input subsidy, and High-quality seed subsidy received by farmers into the total grain subsidy and used it as the dependent variable in our model.

The mechanism through which the policy in MGPA increases the density of grain enterprises is the increase in profit expectations of enterprises, as well as the rise in the number and scale of new enterprises entering the market. The variables required for the above analysis can all be found in the CCAD. The number and scale of entered enterprises are variables at the county-link level, which were previously calculated together with the stock of enterprises, and measuring scale using average registered capital of entered enterprises in a county. In the model specification, the registered capital is taken as the natural logarithm. It is worth mentioning that in the 2023 version of CCAD, the financial reports of enterprises include net income data. However, this data covers only the period from 2013 to 2022, which does not extend back to the years prior to 2004. As a result, the standard DID method cannot be applied. To address this issue, we adopted a unique specification.

Table 1 presents summary statistics of the main variables used in this study. Due to the use of datasets at different levels in the analysis, we differentiated the datasets and conducted descriptive statistics on the variables. For example, the analysis using enterprise density as the outcome variable uses county-link level data, the analysis using grain sown area as the outcome variable uses county level data, and an analysis using agricultural fiscal expenditure as a mechanism variable uses province level data.

3.3. Difference-in-differences strategy

When we use enterprise density as a proxy for grain clusters, a county-link pair is used as the unit in the analysis, as it is the most disaggregated unit in the data, and smaller than a MGPA province. When we use grain sown area as a proxy for the clusters, the analysis unit becoming county (rather than county-link). Difference-in-differences (DID) estimation compares one county's grain cluster before and after the establishing of a MGPA (the first difference) with the changes among non-MGPA counties during the same period (the second difference). Due to the regional nature of the policy in MGPA, the conditions for using this method are met.[12] Specifically, we use the following specification:

$$Density_{pclt} = \alpha + \beta(MGPA_p \times Post_t) + \mathbf{X}'\varphi + \eta_{pcl} + \theta_{r(p)t} + \varepsilon_{pclt} \tag{2.1}$$

[12] In addition, as other agricultural policies aimed at the whole country are inclusive, it can be seen that there is no significant intergroup difference in their impact on the MGPA and non-MGPA, so there is no interference in the estimation of treatment effects.

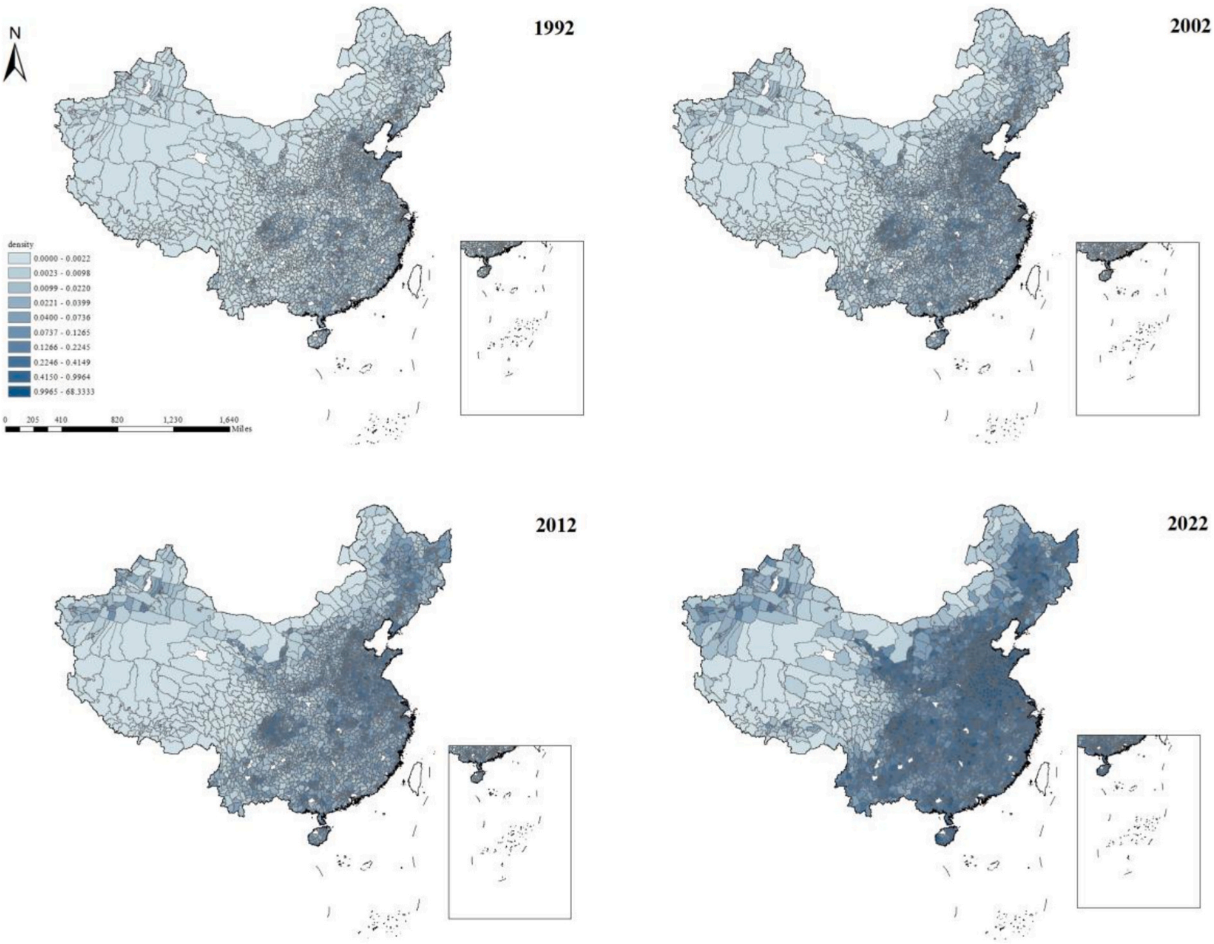

Fig. 3. Geographical distribution of grain enterprise density in China (1992, 2002, 2012, and 2022). Notes: The depth of the colour band indicates the density of grain enterprise and the unit of enterprise density is enterprises/km^2. The gradient of colour bands is divided by the quantiles of density. The white areas indicate missing data. *Source: CCAD (2023).*

$$Grain_{pct} = \alpha + \beta\left(MGPA_p \times Post_t\right) + \mathbf{X}^{'}\varphi + \eta_{pc} + \theta_{r(p)t} + \varepsilon_{pct} \tag{2.2}$$

where $Density_{pclt}$ and $Grain_{pct}$ are the outcomes. $Density_{pclt}$ is the enterprise density of industry chain link l in county c of province p in period t, and $Grain_{pct}$ is the grain sown area in county c of province p in period t; $MGPA_p$ is a dummy variable for the treated group, which denotes the status with 1 if province p is a MGPA province and 0 otherwise; $Post_t$ is the indicator for post-MGPA periods (2004 and beyond); **X** is a vector of province level time-variant characteristics, including the proportion of the primary industry, per capita GDP, railway mileage, and kilometer mileage mentioned in section 3.2; η_{pcl} (η_{pc}) is a county×link (county) fixed effects term capturing time-invariant individual characteristics such as geographic location (part of the natural advantages); in order to further control the confounding factors of regions change over time, we have added a interactive fixed effects between region[13] and year to the model, represented by $\theta_{r(p)t}$; and ε is the error term. α is the intercept and β captures the impact of the policy in MGPA on the outcome variables. To control for potential heteroscedasticity and serial correlation, the standard errors are clustered at the province level.

It is worth noting that, in section 2.1, we clarify that the policy in MGPA is a package of policies. From the perspective of the current policy concept and design for the MGPA, the ultimate goal of this package of policies is to develop grain industry and ensure food security. Therefore, the establishment of the MGPA in 2004 actually included a package of agricultural policies aimed at increasing grain yields. Due to the consistent policy direction, it avoided the difficulty of identifying and separating the impact of each policy. Furthermore, although simultaneity may violate the orthogonality between the policy in MGPA and error term, the identification assumption of DID design does not require policies to be randomly assigned, as long as the results of MGPA counties and non-MGPA

[13] According to the socio-economic development status of different regions in China, the government divides the economic regions into four major regions: eastern, central, western, and northeastern regions.

Table 1
Variable descriptions and summary statistics.

Variable	Level	Obs	Mean	SD
Panel A. County-link level balanced panel data				
Grain enterprise density (enterprises/km2)	County×Link×Year	704,320	0.049	0.262
MGPA (1 = Y, 0 = N)	Province	704,320	0.546	0.498
Post (1 = Y, 0 = N)	Year	704,320	0.613	0.487
Proportion of primary industry	Province×Year	704,320	15.317	7.576
Per capita GDP (thousand yuan)	Province×Year	704,320	27.787	26.844
Railway mileage (thousand km)	Province×Year	704,320	3.402	2.025
Highway mileage (thousand km)	Province×Year	704,320	113.704	72.485
Location quotient	County×Link×Year	696,488	1.230	4.124
Density-based index	County×Link×Year	704,320	0.050	0.218
Number of enterprises entered	County×Link×Year	704,320	5.075	23.282
log(Average scale of enterprises entered)	County×Link×Year	343,400	−8.517	16.811
Number of cooperatives (thousand)	County×Year	494,560	0.334	0.573
Number of family farms (thousand)	County×Year	401,456	0.036	0.182
Panel B. County level unbalanced panel data				
Grain sown area (hectare)	County×Year	21,553	42.398	41.976
MGPA (1 = Y, 0 = N)	Province	21,553	0.546	0.498
Post (1 = Y, 0 = N)	Year	21,553	0.840	0.366
Proportion of primary industry	Province×Year	21,553	12.307	5.124
Per capita GDP (thousand yuan)	Province×Year	21,553	33.210	24.227
Railway mileage (thousand km)	Province×Year	21,553	3.785	1.912
Highway mileage (thousand km)	Province×Year	21,553	129.234	64.602
Panel C. Province level balanced panel data				
Agricultural fiscal expenditure (billion yuan)	Province×Year	620	37.907	32.404
MGPA (1 = Y, 0 = N)	Province	620	0.419	0.494
Post (1 = Y, 0 = N)	Year	620	0.950	0.218
Proportion of primary industry	Province×Year	620	11.026	5.960
Per capita GDP (thousand yuan)	Province×Year	620	44.582	31.661
Railway mileage (thousand km)	Province×Year	620	0.003	0.002
Highway mileage (thousand km)	Province×Year	620	0.122	0.075
Panel D. Household level and Enterprise level unbalanced panel data				
Grain subsidy (thousand yuan)	Household×Year	94,164	0.372	0.482
Net income (10 thousand)	Enterprise×Year	2,765,243	14.987	0.500

Note: Mu is the unit of land used in China. 1 mu = 1/15 ha. Yuan is the currency used in China. 1 yuan = 1/7.18 US dollar in the year 2024. *Source: CCAD (2023), CNT_CropSownArea, China Regional Economy Research Database, Fixed Observation Points System and author manually collected.*

counties are parallel trends after conditioning on control variables.

3.4. Event study

As we mentioned earlier, the consistency of the DID estimator strictly depends on the parallel trend assumption. We plot trends in grain enterprise density and grain sown area for the treatment and control groups. Fig. 4 indicates that until 2004, there was little difference between the average enterprise density in the MGPA counties and in the non-MGPA counties, and that the changes in the average grain sown area were parallel, which provides preliminary results of parallel trends. However, we still need more accurate methods to statistically test parallel trends.

Therefore, in order to test the parallel trend assumption and to better understand the adjustment paths of grain clusters after the policy in MGPA, we employed the event study approach to estimate effects at different years. We used the following specification:

$$Density_{pclt} = \gamma + \sum_{t=1992, t\neq 2003}^{2022} \delta_t \left(MGPA_p \times d_t \right) + \mathbf{X}^{'} \phi + \nu_{pcl} + \xi_{r(p)t} + \mu_{pclt} \tag{3.1}$$

$$Grain_{pct} = \gamma + \sum_{t=2001, t\neq 2003}^{2021} \delta_t \left(MGPA_p \times d_t \right) + \mathbf{X}^{'} \phi + \nu_{pc} + \xi_{r(p)t} + \mu_{pct} \tag{3.2}$$

the definitions of all variables in eqs. (3.1) and (3.2) are consistent with those in eqs. (2.1) and (2.2), and simply replacing the $Post_t$ with a dummy variable d_t, which represents each t period. For example, when a county-link-year pair takes the year as 2021, d_{2021} takes 1 and 0 otherwise. The basic logic of the event study is: if the treatment group and the control group are not significantly different

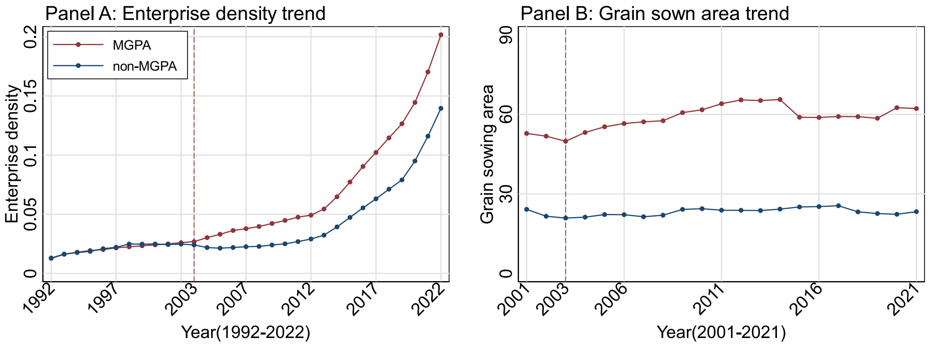

Fig. 4. The time trend of the grain enterprise density and grain sown area. Notes: The markers in Panel A represent the average density of grain enterprise in a certain year, while in Panel B represent the average grain sown area in a certain year. The average density of enterprise is calculated at the county-link level, while the average sown area is calculated at the county level. *Source: CCAD (2023) and CNT_CropSownArea.*

in the trend before the policy intervention, the parallel trend assumption is satisfied. The estimated coefficients on the interaction terms capture the dynamics of treatment effects across periods after the policy intervention.[14] The dummy for d_{2003} is omitted in eq. (4) to represent the reference year. The standard errors are clustered at the province level.

4. Empirical findings

4.1. The impact of the policy in MGPA on grain clusters

This study assesses the impact of the policy in MGPA on grain clusters by applying a DID strategy, and the estimation results are shown in Table 2. Columns (1)–(3) empirically analyze the impact of the policy in MGPA on enterprise density using the stepwise regression. Column (1) controls only for county×link and year fixed effects, and the coefficient of interaction term is significantly positive at the 1 % level; the results in columns (2) and (3) show that the coefficient are still significantly positive after sequentially adding control variables, fixed effects of region×year, respectively, which suggests that the policy in MGPA has a significant positive impact on the enterprise density. Specifically, the policy has resulted in an average increase of 0.025 grain enterprises per km^2 in MGPA counties (approximately 51.07 % of the mean). Columns (4)–(6) use the same method to explore the effect of the policy in MGPA on grain sown area, gradually adding control variables, fixed effects of region and time interaction to the equation. In the case of column (6), for example, the implementation of the policy average increases the grain sown area by 4.35 ha (about 10.26 %), and this effect is significant at the 5 % level. In summary, the policy in MGPA helps to promote the development of grain clusters, which is in line with our expectation that industrial policies that align with comparative advantages can promote the development of clusters.

The previous content explored the impact of the policy in MGPA on the density of all links, but the types of enterprises in each link of the grain industry chain are different, and the development of grain clusters is affected differently by factors such as topography, resource endowment, labor, and capital, so the effect of the policy implementation on the enterprise density of different links may differ. For example, Forslid and Midelfart (2005) points out that optimal policy towards upstream and downstream industries may typically differ radically in an open economy. Thus we classify grain enterprises into eight types according to the grain industry link, and evaluate of policy effects on enterprises in different links. The results are shown in Table 3. The implementation of the policy in MGPA has significantly increased the density of all types of grain enterprises in the grain industry chain, except for agricultural machinery service enterprises. Among them, the implementation of policies has the most significant impact on enterprises engaged in grain planting and sales (74.57 % and 73.98 %, respectively). In addition, the increase in the density of grain rough processing and precision processing enterprises means that policy promoted the development of the secondary industry in MGPA counties. The effect on the density of agricultural machinery service enterprises is negative. Although the negative effect is not significant, we provide a possible explanation: agricultural machinery services have the characteristic of cross regional operation, which means that counties in the MGPA do not need to develop agricultural machinery service enterprises and can also enjoy services from other regions of agricultural machinery, which leads to the fact that agricultural machinery service enterprises are less sensitive to the policy in MGPA than other types of enterprises.

4.2. Parallel trend test and dynamic treatment effects

In order to validate parallel trend before the policy in MGPA counties and non-MGPA counties and to examine the dynamic treatment effects of policy implementation, we adopt an event study approach and the results are shown in Fig. 5. Panel A and B show

[14] Plotting trends in the outcome variables for the treatment and control groups can provide preliminary results for parallel trends, but it is more reasonable and accurate to use event study to test for parallel trend and examine the dynamics of treatment effects. This method has been widely recognized as a standard practice in the DID literature (Moser & Voena, 2012).

Table 2
The impact of the policy in MGPA on grain clusters.

	Outcome: Grain enterprise density			Outcome: Grain sown area		
	(1)	(2)	(3)	(4)	(5)	(6)
MGPA×Post	0.030***	0.027***	0.025**	8.455**	9.580	4.350**
	(0.010)	(0.009)	(0.009)	(3.387)	(5.737)	(1.896)
Percent Change	60.09 %	55.88 %	51.07 %	19.94 %	22.60 %	10.26 %
County×Link FEs	Yes	Yes	Yes			
County FEs				Yes	Yes	Yes
Year FEs	Yes	Yes		Yes	Yes	
Control Variables	No	Yes	Yes	No	Yes	Yes
Region×Year FEs	No	No	Yes	No	No	Yes
Constant	0.039***	0.021	0.064**	38.499***	48.825***	27.343***
	(0.003)	(0.022)	(0.025)	(1.562)	(16.825)	(5.960)
Num of Obs	704,320	704,320	704,320	21,553	21,553	21,553
Num of Counties	2840	2840	2840	1372	1372	1372
F-statistic	8.626	4.908	1.639	6.233	4.146	14.777
Adj-R^2	0.249	0.249	0.252	0.952	0.953	0.961

Notes: Province level clustered standard errors are presented in parentheses. The asterisks indicate statistical significance level: ***$p < 0.01$, **$p < 0.05$. A two-way fixed effect estimation strategy is employed to analyze the effects of the policy in MGPA on grain clusters indicators (enterprise density and grain sown area). Columns (1), (2), and (3) use county-link level panel data, and column (4), (5), and (6) use county level panel data. The control variables include the proportion of primary industry, per capita GDP, railway and highway mileage.

Table 3
Links heterogeneity analysis for the impact of the policy in MGPA on grain clusters.

	Outcome: Grain enterprise density							
	Seed	Fertilizer&Pesticide	Machine	Planting	Service	Rough processing	Deep processing	Sale
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
MGPA×Post	0.020**	0.031**	0.018**	0.015***	−0.007	0.014***	0.016**	0.092*
	(0.009)	(0.014)	(0.007)	(0.004)	(0.006)	(0.005)	(0.007)	(0.049)
Percent Change	48.21 %	43.59 %	42.23 %	74.57 %	−46.81 %	41.09 %	38.38 %	73.98 %
County×Link FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region×Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	0.042**	0.068*	0.055**	0.044**	0.026	0.023**	0.042**	0.191*
	(0.020)	(0.038)	(0.026)	(0.019)	(0.019)	(0.009)	(0.015)	(0.097)
Num of Obs	88,040	88,040	88,040	88,040	88,040	88,040	88,040	88,040
Num of Counties	2840	2840	2840	2840	2840	2840	2840	2840
F-statistic	1.985	2.141	4.691	6.770	1.622	3.054	1.397	0.893
Adj-R^2	0.488	0.416	0.406	0.345	0.262	0.774	0.804	0.562

Notes: Province level clustered standard errors are presented in parentheses. The asterisks indicate statistical significance level: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. A two-way fixed effect estimation strategy, along with county level panel data is employed to analyze the effects of the policy in MGPA on grain clusters indicators (enterprise density). The control variables include the proportion of primary industry, per capita GDP, railway and highway mileage.

the effects of the policy in MGPA on enterprise density and grain sown area, respectively. Regardless of whether the outcome variable is enterprise density or grain sown area, the coefficients are not significant in 2003 and before (95 % confidence intervals contain a value of 0), which means the sample satisfy the parallel trend assumption. After 2003, dynamic effects suggest a sustained and strong role for the policy in MGPA, and the positive and significant results are consistent with the baseline results in Table 3. The above results imply that our data satisfy the prerequisites of the DID strategy for determining causality and that the policy continues to exert a cluster effect.

Once again, we distinguished between industries for analysis. Fig. 6 reveals the use of event study method to analyze the heterogeneous effects of the policy in MGPA on enterprise density in different links. Before the implementation of the policy, all types of grain enterprise densities in the treatment and control groups satisfy the parallel trend assumption, except for agricultural machinery service enterprises. And after 2003, all types of enterprise densities in the treatment group, except for agricultural machinery service enterprises, increased significantly. Analyzing from the perspective of dynamic treatment effects, enterprises at the front end of the industry chain, such as seeds and fertilizers and pesticides, experienced a jump in density immediately after the implementation of the policy. In contrast, agricultural machinery enterprises and planting enterprises experienced a relatively smooth growth process. Grain processing enterprises, both roughing and finishing, experienced a trend similar to linear growth after the implementation of the policy. For grain sales enterprises, despite the very large value of the average treatment effect, the dynamic effect does not seem to support that the effect has a long-run impact. As for agricultural machinery service enterprises, we do not find a significant treatment effect in any period. The result implies the fact that, when considered from the perspective of the industry chain, the effects of

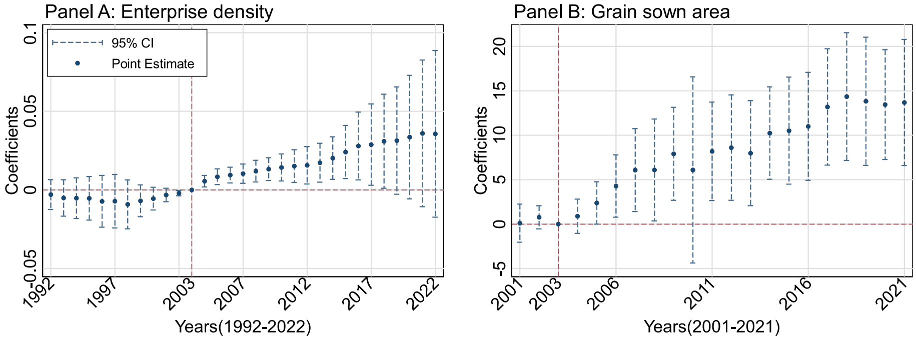

Fig. 5. Event study of the impact of the policy in MGPA on grain clusters. Notes: We have chosen 2003 as the baseline year for the event study approach. Markers represent point estimates, and the dashed lines are 95 % confidence intervals. The regression coefficients before 2003 provided results for parallel trend tests, while the regression coefficients after 2003 provided dynamic treatment effects of the policy in MGPA. The regression controls for individual fixed effects, year fixed effects, region×year fixed effects and all control variables in Table 2, and the standard error are clustered at the provincial level.

industrial policy may be different for each link, and flexible strategies need to be developed to achieve coordinated development.

4.3. Robustness checks

In this section, we perform robustness checks on baseline results. Firstly, we conducted placebo tests, including time and space placebo tests. In this section, we perform robustness checks on baseline results. Firstly, we conducted placebo tests, including time and space placebo tests. For more technical details on the DID placebo test, please refer to Chen et al. (2025). Specifically, the time placebo test takes a certain period before treatment as the "fake treatment time" and only uses the pre-treatment sample for DID estimation. Due to the policy not being implemented before the treatment, if a significant placebo effect is found in a certain period before the treatment, it can be suspected that the treatment effect may be driven by accidental factors or confounding events. Panel A in Fig. 7 shows that regardless of the outcome variable, the placebo effects at the pseudo treatment times we set are not significant, preliminarily verifying that the treatment effects are not accidental. For the spatial placebo test, we randomly selected simulated "treatment group" and "control group" from the entire sample. The sample size of the simulated treatment group and control group was the same as that of the actual treatment group and control group. This sampling was repeated 1000 times, and eq. (2.1) was used to estimate the simulated data sample. The principle of this method is that if the treatment effect is 0, then compared to the distribution of this placebo effect, the estimated value of the treatment effects should not be an extreme value. Panel B shows that the estimated results of the baseline regression are both extreme values and located at the upper tail of the simulated distribution, which means rejecting the assumption of a treatment effect equals to 0 (p-value = 0.000).

Secondly, to examine the sensitivity of using county-link as the unit of analysis, we conducted DID estimates using city-link and province-link panel data. In China, cities comprise many counties, and provinces are made up of cities. This analysis is feasible based on the fact that the policy in MGPA is implemented at the province level. Without spillovers between counties, the province/city level DID estimator is simply the county level version weighted by the outcome shares of the MGPA counties in the province/city (Lu et al., 2019). Fig. 8 shows the results estimated using the province and city level samples. It is clear that they are broadly consistent with the county level sample analyses, whether using the full sample estimates or the sub-link estimates, which suggests that the results of our analyze are not sensitive to the unit.

Finally, we replaced the grain clusters measurements. We replaced enterprises density with location quotient (LQ) and density-based index (DBI), respectively. LQ is a commonly used proxy variable for industrial clusters (Delgado et al., 2010; Porter, 2003), which can reflect the degree of specialization of a certain region in a certain industry. The calculation method is as follow:

$$LQ_{pclt} = \frac{Stock_{pclt} / AgroStock_{pct}}{Stock_{lt} / AgroStock_{t}} \tag{4}$$

Where LQ_{pclt} is the location quotient of link l in county c of province p in period t; $AgroStock_{pct}$ represents the number of agro-related enterprises in county c of province p in period t; $Stock_{lt}$ and $AgroStock_{t}$ represent the number of grain enterprises of link l in period t and the number of agro-related enterprises in period t, respectively. Not all samples entered the specification when analysed using the LQ, as some counties had zero agro-related enterprises in some years, so we could not calculate the LQ (as a consequence sample size drops from 704,320 to 696,488). DBI was proposed by Guo et al. (2020), which is highly consistent with the distribution of industrial clusters in reality and is superior to HHI, Krugman index, etc. in predicting industrial clusters (Guo et al., 2022). The calculation method of DBI is: if the number of enterprises in a certain industry in a certain region can rank in the top 5 % of the country, then the region has a

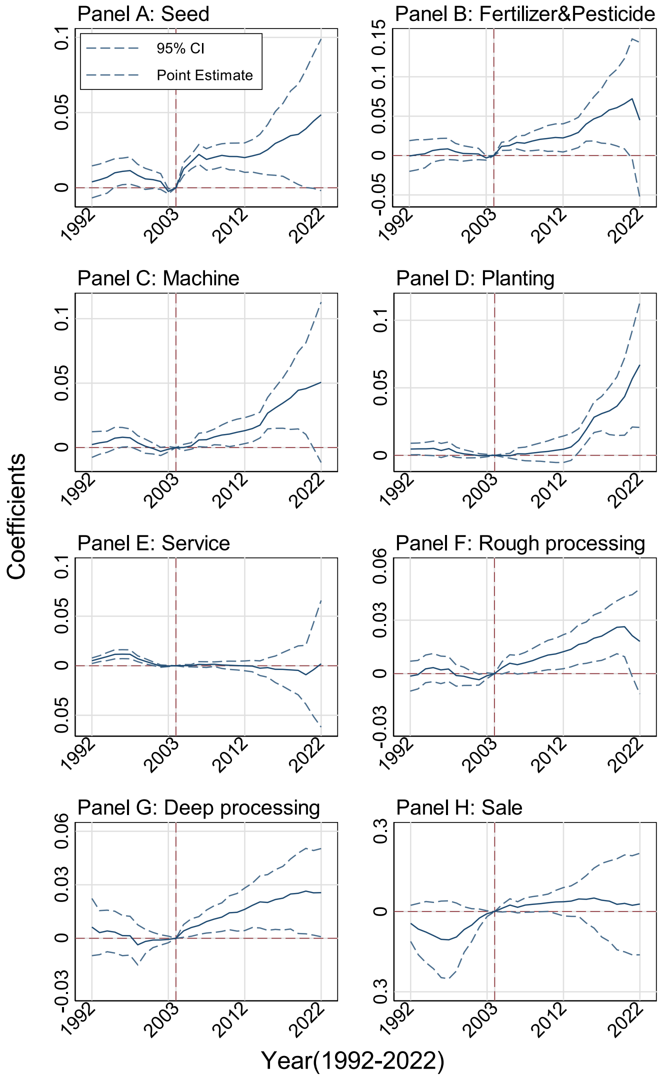

Fig. 6. Event study of the links heterogeneity analysis. Notes: Panels divides the sample by the types of links. We have chosen 2003 as the baseline year for the event study approach. Solid lines represent point estimates, and the dashed lines are 95 % confidence intervals. The regression coefficients before 2003 provided results for parallel trend tests, while the regression coefficients after 2003 provided dynamic treatment effects of the policy in MGPA. The regression controls for county fixed effects, year fixed effects, region×year fixed effects and all control variables in Table 3, and the standard error are clustered at the provincial level.

cluster of that industry ($DBI_{pclt} = 1$). The results of replacing the outcome variable are shown in Fig. 9. When DBI is used as the outcome variable, it is highly consistent with the baseline results, except for the grain processing industry. This may be related to DBI not considering the area of the counties. Data shows that grain processing enterprises are often distributed in smaller counties, so an increase in the number of enterprises does not necessarily mean an increase in density.[15] In other words, the area is assigned different weights to samples. The results of LQ are far from those of enterprise density. This study suggests that LQ completely ignores the important characteristic of scale effect in industrial clusters, and specialization is vastly different from the connotation of clusters.[16] Therefore, LQ may be not a good proxy.

[15] According to DBI, the average area of counties with grain rough processing clusters or deep processing clusters is 1630 km^2, while the average area of counties without such clusters is 3471 km^2. This indicates that the quantity effect of the policy is likely to be offset by the area.

[16] For example, there are very few grain enterprises in Xizang, China, which is difficult to form clusters, but its LQ is very large, because the number of agro-related enterprises in the province is small, too.

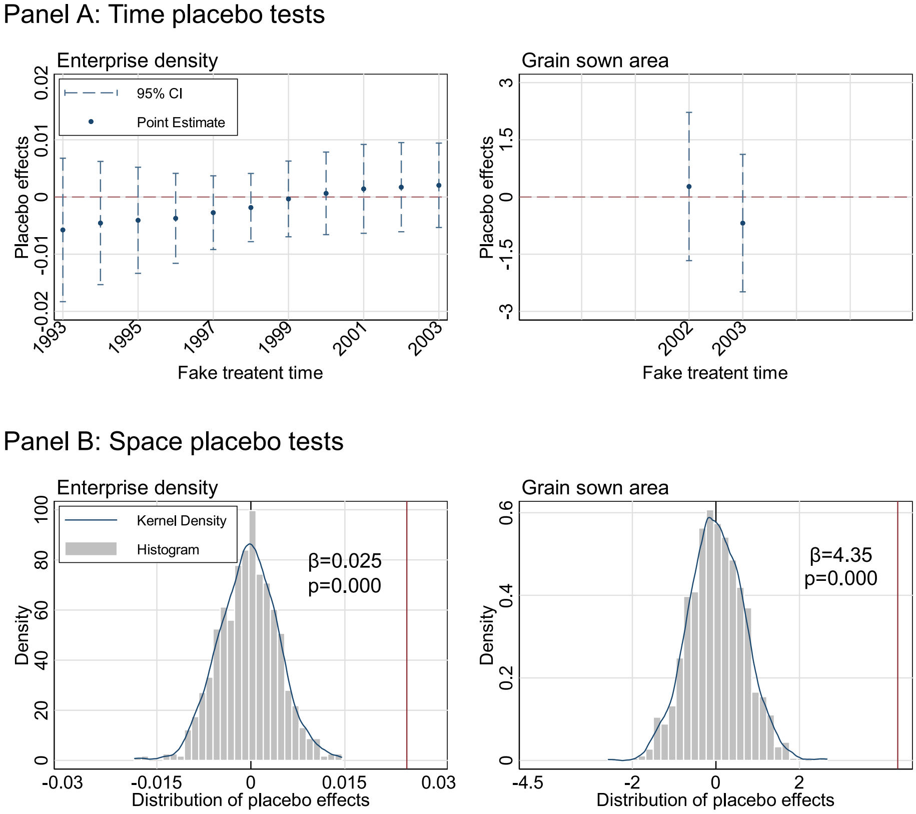

Fig. 7. Placebo tests. Notes: Panel A provided time placebo tests. The two sub-figures used 1993–2003 and 2002–2003 as the fake treatment time, respectively and obtained regression results. Markers represent point estimates, and the dashed lines are 95 % confidence intervals. Panel B provided spatial placebo tests. We randomly selected 1551 and 782 counties as fake treatment counties (consistent with the number of MGPA counties in the original datasets) in both datasets respectively. The sampling was repeated 1000 times, and the histograms and kernel density (epanechnikov kernel function and bandwidths equal to 0.0011 and 0.1471 respectively) of estimates were obtained. The solid red line in the figure represents the benchmark estimation coefficients in columns (3) and (6) of Table 2. 1992 and 2001 are not used as fake treatment years in left panel and right panel of panel A, respectively, because there is no pre-treatment periods before 1992 and 2001. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5. Further analysis

5.1. The mechanisms of the industrial policy effects

This subsection tests the mechanism through which the grain sown area expands. Based on the previous analysis, the increase in agricultural fiscal expenditures and the rise in grain subsidies for farmers could be potential mechanisms. Therefore, we use provincial-level agricultural fiscal expenditure data and household-level grain subsidy data for analysis. Specifically, we replace the previous dependent variables with the mechanism variables to examine the impact of the policy in MGPA on agricultural fiscal expenditures and grain subsidies. Column (1) of Table 4 shows that the policy has brought enormous growth to agricultural fiscal expenditure in MGPA, with a growth rate of up to 69.67 %, and this effect is significant at the 1 % level. The analysis of grain planting subsidies is presented in column (2) of Table 4. We find that after the implementation of the policy in MGPA, the average annual grain subsidy received by farmers increased by 14.7 %, which may have incentivized farmers to expand their grain sown area. These evidences indicate that the Chinese government, based on the unique resource advantages of MGPA, has given it a special political positioning and increased grain subsidies, making the MGPA highly agglomerated in grain production activities, and promoted the development of industrial clusters.

The increase in the grain enterprise density depends on the effect of the policy on the expected profit and the number and scale of entered enterprises, which requires us to consider both the extensive margin and intensive margin. However, the profit expectations of firms are unobservable, we use "the net income of non-newly-entered firms in the current year" as a proxy for the profit expectations of new grain enterprises. This is because a firm's entry and exit decisions, as well as its scale decisions, are typically influenced by the performance of firms in the same industry. Therefore, the variable we selected can be considered as a representation of the firm's profit expectations. In addition, in CCAD, the financial reports of enterprises include net income data. However, this data covers only the period from 2013 to 2022, which does not extend back to the years prior to 2004. As a result, the standard DID method cannot be applied. We took an alternative approach and decided not to focus solely on grain enterprises but to analyze the full sample of

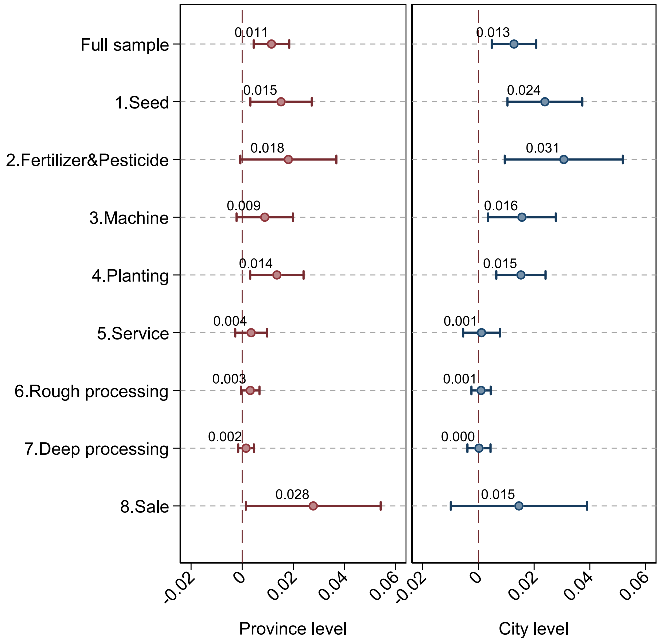

Fig. 8. Unit sensitivity analysis. Notes: The outcome variable is grain enterprise density. The red and blue lines/markers represent the DID estimates of the policy in MGPA on grin clusters using province level and city level data (province-link level and city-link level data), respectively. The model controls for individual fixed effects, year fixed effects, region×year fixed effects and all control variables. The standard error are clustered at the provincial level. The error bars represent the 95 % confidence intervals. The observations are 7688 and 81,344, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

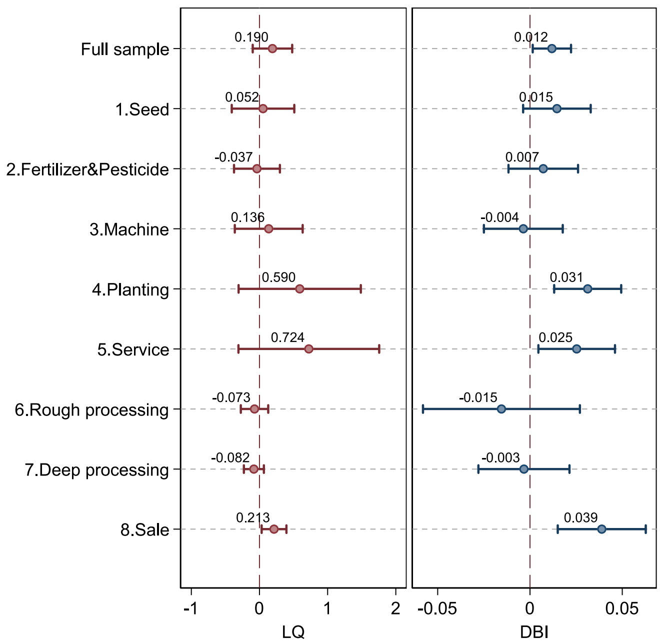

Fig. 9. Replace cluster measurement. Notes: The red and blue lines/markers represent the DID estimates of the policy in MGPA on grin clusters using LQ and DBI as outcome variable, respectively. The model controls for individual fixed effects, year fixed effects, region-year fixed effects and all control variables. The standard error are clustered at the provincial level. The error bars represent the 95 % confidence intervals. The observations are 696,488 and 704,320, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Mechanism: fiscal support, subsidies and profit of grain enterprises.

	Outcome: Agricultural fiscal expenditure	Outcome: ln(Total grain subsidies of household)	Outcome: Net income of grain enterprise
	(1)	(2)	(3)
MGPA×Post	26.509***	0.147**	
	(5.053)	(0.077)	
MGPA×GrainEnterprise			2.360***
			(0.459)
GrainEnterprise			4.375***
			(0.387)
Percent Change	69.67 %		
Province FEs	Yes		Yes
Year FEs		Yes	Yes
Control Variables	Yes	Yes	Yes
Region×Year FEs	Yes		
Household FEs		Yes	
Industry FEs			Yes
Constant	27.346***	0.754**	14.165***
	(2.013)	(0.312)	(0.110)
Num of Obs	620	94,164	2,765,243
F-statistic	27.524	32.489	81.710
Adj-R^2	0.865	0.085	0.017

Notes: Province level clustered standard errors are presented in parentheses. The asterisks indicate statistical significance level: ***$p < 0.01$. The control variables include the proportion of primary industry, per capita GDP, railway and highway mileage.

agricultural enterprises, which includes both grain and non-grain enterprises. This allowed us to introduce an additional dimension of variation, namely "whether the enterprise is a grain enterprise." By incorporating the variations in "whether the enterprise is located in a major grain-producing area" and "whether it is a grain enterprise," we were able to approximate the net effect of the policy in MGPA. The model setup is as follows:

$$Netincome_{ipt} = \alpha + \beta(MGPA_p \times GrainEnterprise_i) + GrainEnterprise_i + \mathbf{X}'\phi + \lambda_{j(i)} + \lambda_p + \lambda_t + \psi_{ipt} \tag{5}$$

In which $Netincome_{ipt}$ is the net income (profit) of enterprise i of province p in year t, and $GrainEnterprise_i$ is a dummy variable indicates that i is a grain enterprise. Through this setup, we address the issue of not having a "pre- and post-event design." The regression results are provided in column (3) of Table 4. The result indicates that the policy in MGPA increased the profits of existing grain enterprises by approximately 23,600 yuan, which is about 15.7 % of the mean. This validates our mechanism regarding profit expectations.

Higher profit expectations will lead to an increase in both the number and scale of new grain enterprises entering the market. Table 5 demonstrates the empirical results, it should be noted that the observations are less when the outcome variable is average scale of enterprises entered. This is because the number of entered enterprises in a county in a given year may be zero, so the average size of

Table 5
Mechanism: number and scale of entered grain enterprises.

	Outcome: Number/average scale of entered enterprises								
	Full sample	Different links							
		Seed	Fertilizer&Pesticide	Machine	Planting	Service	Rough processing	Deep processing	Sale
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A. Number of entered enterprises									
MGPA×Post	2.385***	3.144**	2.938	2.250	4.550***	0.837	0.224	0.143	4.998***
	(0.782)	(1.287)	(1.893)	(1.326)	(1.049)	(0.896)	(0.302)	(0.242)	(1.011)
Num of Obs	704,320	88,040	88,040	88,040	88,040	88,040	88,040	88,040	88,040
Panel B. Average scale of entered enterprises									
MGPA×Post	0.210*	0.243	0.150	0.173	0.252	0.048	0.277*	0.277*	0.061
	(0.117)	(0.169)	(0.168)	(0.115)	(0.149)	(0.193)	(0.150)	(0.150)	(0.099)
Num of Obs	343,400	44,532	52,976	44,509	33,520	27,879	45,572	39,972	54,227
County×Link FEs	Yes								
County FEs		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region×Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Province level clustered standard errors are presented in parentheses. The asterisks indicate statistical significance level: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. A two-way fixed effect estimation strategy, along with county-link level panel data is employed to analyze the effects of the policy in MGPA on the number of enterprises entered or exited. The control variables include the proportion of primary industry, per capita GDP, railway and highway mileage.

enterprises cannot be calculated. The results show that the policy simultaneously increases the number and scale enterprise entered in the MGPA. Specifically, as shown in column (1), the policy resulted in an average of 2.39 new grain enterprises in a given link in grain industry chain in a MGPA county and an average increase of 23.37 % (exp(0.21)-1) in registered capital per new enterprise. This implies, firstly, that the pull of the policy in MGPA on enterprises is significant, as evidenced by the fact that the density of grain enterprises in a county continues to increase despite the accelerated exit of enterprises, the policy stimulate the development of grain clusters in a extensive margin sense and secondly, that policy in MGPA does not lead to grain clusters in a way that simply stacks up enterprises by focusing on quantity without quality. Industrial policies can attract high-quality enterprises to locate and can also lead to an increase in the scale of enterprises. The effect has stimulated industrial upgrading in grain clusters while facilitating the formation of grain clusters. Fig. 10 illustrates the results of the event study, satisfying the parallel trend assumption. Columns (2)–(9) of Table 5 provide results on links heterogeneity. We found that the policy in MGPA has led to an increase in the number and scale of entered enterprises in all links of grain industry chain, albeit some of them insignificant.

5.2. Heterogeneity effects caused by other entitles in grain clusters

We already know that the policy in MGPA has greatly facilitated the development of grain clusters in an average sense, and a natural subsequent thought is whether this effect is conditional? Or would this effect be greater or lesser in certain specific contexts? In fact, agricultural enterprises, farmer professional cooperatives, and family farms are collectively referred to as the "new agricultural operating entity" in China (Zheng, 2024). Coupling effects may also occur between different types of entitles. For example, agricultural enterprises often face the problem of land transfer. Under the organization of farmer professional cooperatives, the land of numerous farmers can be transferred to agricultural enterprises in a simpler way, which significantly reduces transaction costs for enterprises. Therefore, the effect of the policy in MGPA is likely to be more pronounced in areas where the development of new agricultural operating entity is favourable.

To empirically test the heterogeneity of this effect, we first calculated the number of cooperatives and family farms as proxies for their development on a year-by-year, county-by-county basis using the Chinese Farmer Professional Cooperatives and Family Farms segments of the CCAD database. However, because of the late development of cooperatives and family farms in China, there were only a small number of them in the vast majority of counties in 2001 and before. Therefore, we limit the time window of our sample of cooperatives and farms to 2001–2022, with a total of 494,560 and 401,456 observations, respectively. Second, we use an interaction term design in the specification to capture that heterogeneity. Table 6 demonstrates the results of the estimation with the number of cooperatives as an interaction variable. The coefficients of interaction terms show that the policy effects are more pronounced where cooperatives are better developed on average. However, this effect seems to be confined to the front and middle ends of the industry chain, and the impact on the post-production links such as grain processing, acquisition, storage and sales is not significant. This may indicate that at some points in the post-production chain, the cooperatives are in a competitive relationship with the grain enterprises so that the positive coupling effect is cancelled out. Table 7 shows the results of the estimation with the number of family farms as an interaction variable. Except for grain storage and sales enterprises, the results are basically consistent with the heterogeneity effect of cooperatives. The above results indicate that the new agricultural operating entitles can effectively enhance the effect of the policy in MGPA, significantly clustering the pre-production and mid-production links of the grain industry. However, these effects have not yet been reflected in the post-production stage, especially in the grain processing link.

6. Conclusions

Industrial clusters are considered an effective form of industrial organization, also in the grain industry, but there is a lack of empirical evidence on the impact of industrial policy on grain clusters. This study used the CCAD database to identify grain enterprises

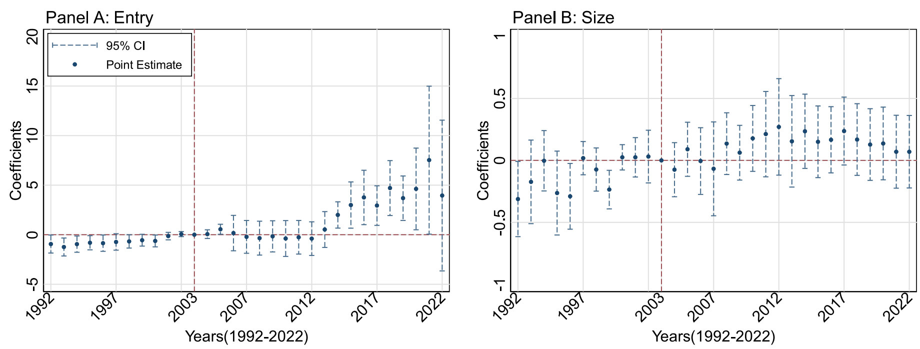

Fig. 10. Event study: the number and scale of entered enterprises. Notes: The outcome variables are the number and scale of entered enterprises, respectively, both use full sample. We have chosen 2003 as the baseline year for the event study approach. Markers represent point estimates, and the dashed lines are 95 % confidence intervals. The regression coefficients before 2003 provided results for parallel trend tests, while the regression coefficients after 2003 provided dynamic treatment effects of the policy in MGPA. The regression controls for county fixed effects, year fixed effects, region-year fixed effects and all control variables, and the standard error are clustered at the provincial level.

Table 6
Heterogeneity: number of cooperatives.

	Outcome: Grain enterprise density								
	Full sample	Different links							
		Seed	Fertilizer&Pesticide	Machine	Planting	Service	Rough processing	Deep processing	Sale
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
MGPA×*Post*×Cooperatives	0.033**	0.037**	0.063**	0.045**	0.036***	0.028**	0.002	0.001	0.054
	(0.016)	(0.014)	(0.025)	(0.017)	(0.012)	(0.013)	(0.007)	(0.008)	(0.040)
MGPA×*Post*	0.009	0.010	0.009	0.006	0.003	−0.011	0.014***	0.015***	0.027
	(0.008)	(0.009)	(0.014)	(0.009)	(0.006)	(0.007)	(0.004)	(0.005)	(0.025)
Cooperatives	−0.065***	−0.053***	−0.117***	−0.090***	−0.028**	−0.044***	−0.023***	−0.027***	−0.139***
	(0.018)	(0.016)	(0.026)	(0.018)	(0.012)	(0.015)	(0.007)	(0.007)	(0.051)
County×Industry FEs	Yes								
County		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region×Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	0.081***	0.070**	0.126**	0.104**	0.085***	0.043	0.024**	0.048***	0.152**
	(0.026)	(0.026)	(0.053)	(0.038)	(0.027)	(0.029)	(0.012)	(0.015)	(0.056)
Num of Obs	494,560	61,820	61,820	61,820	61,820	61,820	61,820	61,820	61,820
F-statistic	7.441	4.771	6.101	7.928	4.734	1.576	6.440	6.769	9.023
Adj-R^2	0.356	0.633	0.557	0.543	0.427	0.363	0.755	0.848	0.629

Notes: Province level clustered standard errors are presented in parentheses. The asterisks indicate statistical significance level: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. A two-way fixed effect estimation strategy, along with county-link level panel data is employed to analyze the heterogeneity of the effects on the number of cooperatives. The control variables include the proportion of primary industry, per capita GDP, railway and highway mileage.

Table 7
Heterogeneity: number of family farms.

	Outcome: Grain enterprise density								
	Full sample	Different links							
		Seed	Fertilizer&Pesticide	Machine	Planting	Service	Rough processing	Deep processing	Sale
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
MGPA×*Post*×Farms	0.084***	0.076***	0.124***	0.085***	0.167***	0.085***	−0.000	0.001	0.137**
	(0.016)	(0.019)	(0.030)	(0.026)	(0.049)	(0.026)	(0.010)	(0.008)	(0.052)
MGPA×*Post*	0.020***	0.024***	0.030***	0.018***	0.008*	0.001	0.013***	0.013***	0.057***
	(0.004)	(0.006)	(0.010)	(0.006)	(0.005)	(0.004)	(0.003)	(0.003)	(0.010)
Farms	−0.018	−0.038***	−0.102***	−0.086***	0.123***	−0.010	0.000	−0.011	−0.022
	(0.013)	(0.012)	(0.026)	(0.025)	(0.035)	(0.015)	(0.009)	(0.007)	(0.032)
County×Industry FEs	Yes								
County		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region×Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	0.075***	0.052***	0.110***	0.106***	0.067***	0.045**	0.026**	0.040***	0.152***
	(0.017)	(0.016)	(0.037)	(0.024)	(0.022)	(0.022)	(0.010)	(0.011)	(0.049)
Num of Obs	401,456	50,182	50,182	50,182	50,182	50,182	50,182	50,182	50,182
F-statistic	20.336	8.934	6.369	6.832	17.439	3.763	5.339	5.216	12.748
Adj-R^2	0.377	0.612	0.558	0.530	0.566	0.329	0.794	0.826	0.681

Notes: Province level clustered standard errors are presented in parentheses. The asterisks indicate statistical significance level: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. A two-way fixed effect estimation strategy, along with county-link level panel data is employed to analyze the heterogeneity of the effects on the number of family farms. The control variables include the proportion of primary industry, per capita GDP, railway and highway mileage.

from agro-related enterprises, and for the first time, reviewed the development of China's grain industry chain and grain clusters. Based on this, an empirical study was conducted on the causal effect of the policy in MGPA, an industrial policy formulated by China for grain production, on grain clusters. The results showed that the policy in MGPA increased the average enterprise density of a certain link of grain industry chain in the county by 51.07 %, and increased the average grain sown area in the county by 10.26 %. This indicates that the policy has produced a scale effect on grain production, which is an important feature of the industrial clusters. In addition, this policy has varying degrees of impact on enterprises in different links of the grain industry chain: seeds, fertilizers and pesticides, agricultural machinery, planting, processing, and sales, which increases the density of enterprises. This indicates that the policy in MGPA has made the grain industry chain more perfect, strengthened enterprise division of labor, which is another characteristic of the industrial cluster. The placebo test, unit sensitivity test, and replace cluster measurement method all indicate that our estimations are robust. Mechanism analysis shows that agricultural fiscal support and the number and scale of entered enterprises are important pathways for the policy in MGPA to affect industrial clusters. Specifically, the policy has increased fiscal support for MGPA by approximately 69.67 %, greatly stimulating the enthusiasm of farmers for grain cultivation and resulting in a highly concentrated nature of grain production activities. This policy promotes both the number and scale of new enterprises, which implies that the policy in MGPA does not lead to grain clusters in a way that simply stacks up enterprises by focusing on quantity without quality. Instead, they attract enterprise site selection while expand their operating scale, guiding capacity upgrading.

Our findings also draw some policy implications. Firstly, industrial policies that leverage regional comparative advantages can promote the development of industrial clusters. The main grain-producing areas usually have abundant land resources, water resources, and climate conditions. The soil in these areas is fertile and suitable for the cultivation of grain crops. In addition, farmers in MGPA usually have rich planting experience and a high degree of agricultural specialization. This helps to improve the efficiency of grain production. The government will increase fiscal support for MGPA, encourage farmers to invest in modern agriculture, and improve production efficiency. This helps to further leverage the advantages of resource endowment and scale. In addition, due to the coupling effect between policy objects and their surrounding environment or entities, such environments or entities may also become a part of comparative advantage. Agricultural industry policies should fully consider these factors to improve the accuracy and validity of policies. Secondly, developing countries generally face the issue of low industrialization in the grain sector. A significant aspect of this problem lies in the lack of a well-established industry chain. Therefore, an essential objective of industrial policies should be to expand the industry chain. For instance, in regions with abundant land resources but relatively small overall grain processing industries and low on-site processing conversion rates, the government should encourage the establishment of processing industrial parks in these regions. This support would extend the grain industry chain, foster grain and feed industry clusters, and promote local value addition. Additionally, providing fiscal support and tax incentives can encourage enterprises to upgrade technology, update equipment, and innovate products. These measures contribute to enhancing the competitiveness and efficiency of the industry chain. Furthermore, promoting collaboration between production and sales areas is crucial. In summary, governments should comprehensively consider all aspects of the industry chain and formulate targeted policies to facilitate the improvement of the grain industry chain and achieve high-quality development.

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.

Acknowledgements

This work was supported by the National Social Science Fund of China [grant number 22&ZD080]; The authors would like to thank the participants of the 2024 China Economist Society (CES) China Annual Conference for their valuable feedback and insightful comments on this paper.

Appendix A. Details of text analysis

The sub database Agricultural and Related Industry Entities (hereafter called AIRE) divides all agro-related enterprises into three categories: agricultural enterprises, agriculture-supporting enterprises, and agro-processing enterprises. The first two categories cover all seed, fertilizer and pesticide, agricultural machinery, grain plant, machinery service and grain sales enterprises, and cover part of grain processing enterprises. The third category covers the vast majority of grain-processing enterprises, including grain rough processing and grain deep processing. In order to more easily identify grain enterprises from the total number of agro-related enterprises, this study applies different treatments to the first two categories and the third category of enterprises, respectively. Specifically, for the first two types of enterprises, we perform text analysis based on their operate scopes variable (OPSCOPE), and when a keyword (or short sentence) is involved, it is defined as a enterprise in one of the links in the grain industry chain. For the third category of enterprises, we judge whether they are grain processing enterprises and whether they are roughing or finishing based on the classification carried out by *Industrial Classification for National Economic Activities.*[17]

[17] Source: National Bureau of Statistics (NBS), https://www.stats.gov.cn/sj/tjbz/gmjjhyfl/202302/P020230213400314380798.pdf.

We begin by describing the text analysis methods performed on the first two types of enterprises. The specific identification method includes the following 3 steps: The first step is to manually mark a large number of keywords (short sentences) according to our subjective judgement. We manually label the keywords (including short sentences) related to the industry nature of grain enterprises from a large number of agro-related enterprises operate scopes. For example, the keywords (short sentences) of fertilizer & pesticide enterprises include "nitrogen fertilizer sales", "organic fertilizer processing and sales", etc. It should be noted that we cannot directly label terms with industry characteristics, such as "fertilizer" and "agricultural inputs", because expressions containing these keywords in the scopes of business may have negative connotations, such as "agricultural inputs (excluding fertilizer)". Therefore, when retrieving the operate scopes from the database, we also need to consider the context of the keywords. The manually marked enterprises sample of is selected based on random sampling.

The second step is to write regular expressions (RegEx) for matching: according to the keywords and short sentences labeled in the first step for different links, and combined with the precautions proposed in the first step, we rewrite the labeled keywords and short sentences to be able to match key descriptions that do not have negative connotations of words from the operate scopes. For example, in the case of grain sales enterprises, the phrases "sales of shaped packaged food, grain and oil" and "grain milling, processing and sales" are labeled with "specific verbs + link keywords" or "link keywords + specific verbs", such as "sale of grain and oil", although there may be other characters between the verb and the noun, which, according to the observation, are mainly other irrelevant words, dunces, bracket characters and Arabic numerals. Therefore, the main work of this step is to write the above text rules into RegEx, the RegEx needs to be written not only to collect and match a variety of key nouns, but also to extend the specific verb. For example, according to the verb "sale" can be extended to "purchase and sale", "retail", "acquisition", "reserve", "storage" and other verbs with similar meanings, while allowing the order of these verbs and key nouns to be switched, as well as allowing zero to multiple characters to appear between them (only Chinese characters, dun, brackets, numbers, etc. are not allowed). In addition, if some modifiers with negative meanings appear before or after a key noun, the keyword will not be matched, for example, "does not include grain and oil", "excluding grain and oil", "except grain and oil", "other than grain and oil", etc. The last thing to note is that RegEx need to be written in such a way that they do not match non-keywords. According to the above ideas, for each type of enterprise to write a RegEx and use Python to the AIRE database in the original data of the enterprise unit to match the conditions of the short sentences or keywords (a part of the keywords can be directly matched without taking into account the verb, the negative word, etc.) and the relevant enterprises to play the tag, and at the same time will be matched to the short sentences is extracted for the subsequent checking and optimization.

Table A1
Grain processing enterprises and industry category.

2-digit code	Middle category	3-digit code	Small category
Panel A. Grain rough processing			
21	Grain and oil processing and soy product manufacturing	211	Grain grinding
		212	Starch and starch product manufacturing
		213	Bean product manufacturing
		214	Edible plant oil processing
		215	Feed processing
Panel B. Grain deep processing			
25	Baking food manufacturing	251	Pastry and bread manufacturing
		252	Biscuits and other baked goods
26	Convenient food manufacturing	261	Rice and flour products manufacturing
		262	Frozen food manufacturing
		263	Instant noodle manufacturing
		264	Canned rice and noodles and other convenient foods manufacturing
27	Food additives and seasonings manufacturing	271	MSG manufacturing
		272	Soy sauce, vinegar manufacturing
		273	Food and feed additives manufacturing
		274	Other seasonings and fermented products manufacturing

Source: Industrial Classification for National Economic Activities.

The third step is to optimize the keywords and matching methods: observe the preliminary matching results, according to the extracted keywords, and combined with the content of the operates scopes to optimize the keywords and short sentences in the first step and add them to the keywords in the second step, in addition to adding some keywords, but also some of the keywords match the implementation of a more stringent restrictions. For example, in the matching of the keyword "seeds", because the original intention is to match grain seeds, so the first few characters of the keyword "seed" can not be "non-conventional crops" and similar expressions and other keywords such as "tree seed" and other keywords. For a more accurate match, before optimizing the RegEx and matching again, write a RegEx that removes all irrelevant words (e.g., "seeds of medicinal herbs") and negative expressions (e.g., "except seeds") from the operate scopes and generate a copy of the operate scopes (a new column), and then use the optimized RegEx to match the keywords and short sentences from the cleaned copy of the operate scopes and tag the enterprises.

For the third category of enterprises, due to *Industrial Classification for National Economic Activities*'s detailed classification of agricultural product processing enterprises, we can simply determine whether they belong to grain processing enterprises based on the variable of enterprise category. The grain processing link and its corresponding industry classification are shown in Table A1. Industries with two-digit code 21 are in the grain rough processing link, and those with 25, 26, and 27 are in the grain deep processing link. Based on this, we have completed the definition of whether a enterprise belongs to the grain industry chain and which link of the chain it belongs to, which formed the basis for delineating the grain industry chain and the evolution of grain clusters.

Data availability

Data will be made available on request.

References

Abdur, R., & Wang, J. (2014). Agricultural natural disasters mitigation in Hebei Province China. *Advanced Materials Research, 962-965*, 1979–1988.
Au, C., & Henderson, J. V. (2006). How migration restrictions limit agglomeration and productivity in China. *Journal of Development Economics, 80*(2), 350–388.
Bartelme, D. G., Costinot, A., Donaldson, D., & Rodríguez-Clare, A. (2019). *The textbook case for industrial policy: Theory meets data*. National Bureau of Economic Research Working Paper Series, No. 26193.
Barwick, P. J., Kalouptsidi, M., & Bin Zahur, N. (2024). Industrial policy: Lessons from shipbuilding. *Journal of Economic Perspectives, 38*(4), 55–80.
Bischi, G. I., Dawid, H., & Kopel, M. (2003). Spillover effects and the evolution of firm clusters. *Journal of Economic Behavior & Organization, 50*(1), 47–75.
Bu, D., & Liao, Y. (2022). Land property rights and rural Enterprise growth: Evidence from land titling reform in China. *Journal of Development Economics, 157*, Article 102853.
Chen, Q., Qi, J., & Yan, G. (2025). Placebo tests for difference-in-differences: A pritical cuide. *Journal of Management World, 41*(02), 181–220 (in Chinese).
Dai, R., Mookherjee, D., Quan, Y., & Zhang, X. (2021). Industrial clusters, networks and resilience to the Covid-19 shock in China. *Journal of Economic Behavior & Organization, 183*, 433–455.
Delgado, M., Porter, M. E., & Stern, S. (2010). Clusters and entrepreneurship. *Journal of Economic Geography, 10*(4), 495–518.
Diez-Vial, I. (2011). Geographical cluster and performance: The case of Iberian ham. *Food Policy, 36*(4), 517–525.
Du, R., & Mao, X. (2017). Estimating the effects of the policies applied in major grain producing areas based on synthetic control method. *China Soft Science, 6*, 31–38 (in Chinese).
Ellison, G., & Glaeser, E. L. (1997). Geographic concentration in U.S. manufacturing industries: A dartboard approach. *Journal of Political Economy, 105*(5), 889–927.
Ellison, G., & Glaeser, E. L. (1999). The geographic concentration of industry: Does natural advantage explain agglomeration? *American Economic Review, 89*(2), 311–316.
Ellison, G., Glaeser, E. L., & Kerr, W. R. (2010). What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review, 100*(3), 1195–1213.
Fan, H., Li, C., Xue, C., & Yu, M. (2023). Clan culture and patterns of industrial specialization in China. *Journal of Economic Behavior & Organization, 207*, 457–478.
Fan, P., Mishra, A. K., Feng, S., Su, M., & Hirsch, S. (2023). The impact of China's new agricultural subsidy policy on grain crop acreage. *Food Policy, 118*, Article 102472.
Fan, S. (2021). Economics in food systems transformation. *Nature Food, 2*(4), 218–219.
Forslid, R., & Midelfart, K. H. (2005). Internationalisation, industrial policy and clusters. *Journal of International Economics, 66*(1), 197–213.
Gálvez-Nogales, E. (2010). *Agro-based clusters in developing countries: Staying competitive in a globalized economy*. Agricultural Management, Marketing and Finance Occasional Paper. Food and Agriculture Organization of the United Nations.
Ge, Y. (2009). Globalization and industry agglomeration in China. *World Development, 37*(3), 550–559.
Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., & Shleifer, A. (1992). Growth in cities. *Journal of Political Economy, 100*(6), 1126–1152.
Goldberg, P. K., Juhász, R., Lane, N. J., Lo Forte, G., & Thurk, J. (2024). *Industrial policy in the global semiconductor sector*. National Bureau of Economic Research Working Paper Series, No. 32651.
Guiso, L., & Schivardi, F. (2007). Spillovers in industrial districts. *The Economic Journal, 117*(516), 68–93.
Guiso, L., & Schivardi, F. (2011). What determines entrepreneurial clusters? *Journal of the European Economic Association, 9*(1), 61–86.
Guo, D., Jiang, K., Xu, C., & Yang, X. (2020). Clustering, growth and inequality in China. *Journal of Economic Geography, 20*(5), 1207–1239.
Guo, D., Jiang, K., Xu, C., & Yang, X. (2022). Industrial clustering, income and inequality in rural China. *World Development, 154*, Article 105878.
Guo, D., Jiang, K., Xu, C., & Yang, X. (2023). Geographic clusters, regional productivity and resource reallocation across firms: Evidence from China. *Research Policy, 52*(2), Article 104691.
Guo, S., Lv, X., & Hu, X. (2021). Farmers' land allocation responses to the soybean rejuvenation plan: Evidence from 'typical farm' in Jilin, China. *China Agricultural Economic Review, 13*(3), 705–719.
Hailu, G., & James Deaton, B. (2016). Agglomeration effects in Ontario's dairy farming. *American Journal of Agricultural Economics, 98*(4), 1055–1073.
Hoover, E. M. (1936). The measurement of industrial localization. *The Review of Economics and Statistics, 18*(4), 162–171.
Hsu, W., Lu, Y., Luo, X., & Zhu, L. (2023). Foreign direct investment and industrial agglomeration: Evidence from China. *Journal of Comparative Economics, 51*(2), 610–639.
Hu, C., Zhang, X., Reardon, T., & Hernandez, R. (2019). Value-chain clusters and aquaculture innovation in Bangladesh. *Food Policy, 83*, 310–326.
Hu, X., Ruan, J., & Zhang, X. (2021). Crisis-induced innovation: Quality upgrading in Chinese industrial clusters. *The Journal of Law, Economics, and Organization, 37*(3), 571–606.
Hua, W., Chen, Z., & Luo, L. (2022). The effect of the major-grain-producing-areas oriented policy on crop production: Evidence from China. *Land, 11*(9), 1375.
Ji, X., Yin, R., & Zhang, H. (2023). Food security and overuse of agrochemicals: Evidence from China's major grain-producing areas policy. *Environmental Science and Pollution Research, 30*(23), 64443–64459.
Jr Tabe-Ojong, M. P., & Dureti, G. G. (2023). Are agro-clusters pro-poor? Evidence from Ethiopia. *Journal of Agricultural Economics, 74*(1), 100–115.
Juhász, R., & Lane, N. (2024). The political economy of industrial policy. *Journal of Economic Perspectives, 38*(4), 27–54.
Kim, S. (1999). Regions, resources, and economic geography: Sources of U.S. regional comparative advantage, 1880-1987. *Regional Science and Urban Economics, 29*(1), 1–32.
Krugman, P. R. (1991a). Increasing returns and economic geography. *Journal of Political Economy, 99*(3), 483–499.
Krugman, P. R. (1991b). *Geography and trade*. Cambridge, Ma: MIT press.
Lin, J. Y. (2011). New structural economics: A framework for rethinking development. *The World Bank Research Observer, 26*(2), 193–221.
Long, C., & Zhang, X. (2011). Cluster-based industrialization in China: Financing and performance. *Journal of International Economics, 84*(1), 112–123.
Lu, Y., Wang, J., & Zhu, L. (2019). Place-based policies, creation, and agglomeration economies: Evidence from China's economic zone program. *American Economic Journal: Economic Policy, 11*(3), 325–360.
Luo, S., He, K., & Zhang, J. (2020). The more grain production, the more fertilizers pollution? Empirical evidence from major grain-producing areas in China. *Chinese Rural Economy, 1*, 108–131 (in Chinese).
Luo, S., & Zhang, J. (2024). The reward of bumper harvests: Fiscal incentives and grain production growth. *Chinese Rural Economy, 8*, 27–46 (in Chinese).
Marshall, A. (1920). *Principles of economics*. London: MacMillan.
Moretti, E. (2021). The effect of high-tech clusters on the productivity of top inventors. *American Economic Review, 111*(10), 3328–3375.
Moro, D., & Sckokai, P. (2013). The impact of decoupled payments on farm choices: Conceptual and methodological challenges. *Food Policy, 41*, 28–38.

Moser, P., & Voena, A. (2012). Compulsory licensing: Evidence from the trading with the enemy act. *American Economic Review, 102*(1), 396–427.
Nathan, M., & Overman, H. (2013). Agglomeration, clusters, and industrial policy. *Oxford Review of Economic Policy, 29*(2), 383–404.
Porter, M. E. (1990). *The competitive advantage of nations*. New York: Free Press.
Porter, M. E. (1998). Clusters and the new economics of competition. *Harvard Business Review, 76*(6), 77–90.
Porter, M. E. (2003). The economic performance of regions. *Regional Studies, 37*(6–7), 549–578.
Rodríguez-Clare, A. (2007). Clusters and comparative advantage: Implications for industrial policy. *Journal of Development Economics, 82*(1), 43–57.
Ruan, J., & Zhang, X. (2009). Finance and cluster-based industrial development in China. *Economic Development and Cultural Change, 58*(1), 143–164.
Saxenian, A. (1994). *Regional advantage: Culture and competition in Silicon Valley and route 128*. Cambridge, Ma: Harvard University Press.
Sylla, R. (2024). Alexander Hamilton's report on manufactures and industrial policy. *Journal of Economic Perspectives, 38*(4), 111–130.
Weber, A. (1909). *Theory of the location of industries*. University of Chicago.
Wei, M., Yan, T., & Luo, S. (2023). The impacts of scale management and technological progress on green and low carbon development of agriculture: A quasi-natural experiment based on the establishment of major grain-producing areas. *Chinese Rural Economy, 02*, 41–65 (in Chinese).
Ye, D., Zhen, S., Wang, W., & Liu, Y. (2023). Spatial double dividend from China's main grain-producing areas policy: Total factor productivity and the net carbon effect. *Humanities and Social Sciences Communications, 10*(1), 459.
Yi, F., Sun, D., & Zhou, Y. (2015). Grain subsidy, liquidity constraints and food security—Impact of the grain subsidy program on the grain-sown areas in China. *Food Policy, 50*, 114–124.
Zhang, X., & Hu, D. (2014). Overcoming successive bottlenecks: The evolution of a potato cluster in China. *World Development, 63*, 102–112.
Zhang, X., Yang, J., & Thomas, R. (2017). Mechanization outsourcing clusters and division of labor in Chinese agriculture. *China Economic Review, 43*, 184–195.
Zhao, L., Ruan, J., & Shi, X. (2021). Local industrial policies and development of agricultural clusters: A case study based on a tea cluster in China. *International Food and Agribusiness Management Review, 24*(2), 267–288.
Zheng, L. (2024). Big hands holding small hands: The role of new agricultural operating entities in farmland abandonment. *Food Policy, 123*, Article 102605.
Zheng, W. (2023). Land supply and industrial agglomeration: Extension of the new economic geography model with land elements incorporated. *Land Use Policy, 132*, Article 106757.