

## Grain processing clusters and income inequality: evidence from rural China

### RESEARCH ARTICLE

Yuwei Zhang<sup>a</sup>, Jianqing Ruan<sup>b</sup> and Yinze Zhang<sup>a</sup>

<sup>a</sup>PhD candidate, <sup>b</sup>Professor, China Academy for Rural Development, School of Public Affairs,  
Zhejiang University, Hangzhou 310058, PR China

---

### Abstract

In developing economies, rural income inequality remains a persistent and pressing challenge, this study identifies grain processing clusters as a pivotal factor for mitigating disparities. Based on the National Fixed Point Survey (NFP) from 2004 to 2017, we constructed an unbalanced panel dataset covering 2300 counties. Using the National Economic Industry Classification Code, we identified information on rough and deep grain processing enterprises in the China Academy for Rural Development-Qiyang China Agri-research Database (CCAD), thereby generating a unique dataset for grain processing clusters. Employing a two-way fixed effects model, we empirically analyze how grain processing clusters affect rural income inequality in China and explore the underlying mechanisms. The baseline regression results indicate heterogeneity in the impact of grain processing clusters on income inequality. Deep processing clusters significantly reduce both farm and non-farm income inequality, whereas rough processing clusters have a significant effect only on non-farm income inequality. Mechanism analysis reveals that grain processing clusters primarily enhance the income of low-income households through three channels: providing inclusive employment opportunities, accelerating the deepening of capital, and increasing grain revenues, thereby reducing the extent of inequality. Heterogeneity analysis further demonstrates that this effect is more pronounced in China's major grain-producing regions and central areas. The research results of this paper comprehensively analyze the effective path to reduce income inequality in rural China while ensuring food security, providing valuable insights for policymakers.

**Keywords:** capital deepening, grain processing clusters, grain revenue, inclusive employment opportunities, rural income inequality

**JEL codes:** D31; D63; L22; Q12

---

<sup>①</sup>Corresponding authors: [zhangyuweiwy@126.com](mailto:zhangyuweiwy@126.com); [ruanjq@126.com](mailto:ruanjq@126.com); [yinzezhang@zju.edu.cn](mailto:yinzezhang@zju.edu.cn)

## 1. Introduction

Since the 1980s the level of income inequality within rural China has continued to increase. The deterioration in the distribution of households' income is detrimental to the stable development of the economy and society (Guo *et al.*, 2022; Luo *et al.*, 2020). Scholars generally believe that the transfer of labor can narrow the income gap (Wang and Yu Benjamin, 2019), but in many underdeveloped areas, the transfer of rural labor is not sufficient. From 2016 to 2020, nearly 70% of China's sown area was dedicated to food crops, yet the average annual net profit of major food crops was only 32.35 yuan per mu.<sup>1</sup> The low comparative earnings of grain not only undermines the enthusiasm for planting but also increases the difficulty for households to increase their income (Howell, 2017). Clusters are considered an effective way to improve the economic benefits of agriculture (Galvez-Nogales, 2010). In many developing countries, enterprises in clusters enjoy many competitive advantages in market access, labor concentration, and easy learning, which can provide employment opportunities for the abundant labor force (Marshall, 1920; Ruan and Zhang, 2009). This point is vital importance for developing countries (Minten *et al.*, 2020; Wardhana *et al.*, 2020; Zhang, 2023). Therefore, our study, which focuses on grain processing clusters, offers a novel perspective for reducing inequality in rural areas of China.

Grain clusters are an extension and application of industrial cluster theory in the field of grain. The extant studies have confirmed that industrial clusters can increase income and reduce income inequality among rural households simultaneously (Ali and Peerlings, 2012; Guo *et al.*, 2022). In the agricultural sector, agro-clusters are also considered an effective way to reduce risk, create employment opportunities, and increase the income of small farmers (Galvez-Nogales and Webber, 2017; Jr Tabe Ojong and Dureti, 2023). However, few studies have discussed how grain clusters affect income inequality among farmers. Based on operational activities, the grain industry chain can be divided into segments such as production, processing, retail, and consumption, with the processing industry being a key component (Hadachek *et al.*, 2024). Grain processing is typically categorized into rough processing and deep processing based on the degree of processing and the capacity for value addition. Rough processing refers to basic physical treatment of raw grains that primarily modifies external characteristics without substantially changing molecular structures or functional attributes, thus offering limited value addition. Deep processing, conversely, involves comprehensive biochemical or physical modification of raw materials, fundamentally altering their compositional or functional properties to achieve significant value enhancement. The enhancement of product added value can directly influence the income distribution mechanisms for households. Based on this, we propose the research question of this paper: Can grain processing clusters reduce income inequality in rural China? What is the difference between rough processing and deep processing? What types of income gaps and types of households are specifically affected? What is its internal mechanism?

To address these issues, we primarily utilize the National Fixed Point Survey (NFP) data from 2004 to 2017 to obtain household income data. This dataset provides detailed information on non-farm employment, agricultural operations, and grain production, which adequately meets the research requirements. Data on grain processing clusters are mainly derived from the China Academy for Rural Development-Qiyang China Agri-research Database (CCAD).<sup>2</sup> We identify rough processing and deep processing enterprises in the business registration system based on keywords and construct cluster indicators based on the number of enterprises per county unit area. After data cleaning and matching, we applied a fixed-effects model to test our hypothesis, supplemented by two-stage estimation, quantile regression, and other robustness checks.

This study contributes to the general literature in the following areas. First, this study bridges grain industry with geographic agglomeration theory, advancing research on agro-based clusters. Galvez-Nogales (2010) argues that agro-based clusters refer to an interconnected value network comprising producers, agro-industries, traders, and other entities involved in related industries. Within the cluster, each sector can effectively promote

<sup>1</sup> The data can be accessed at <http://www.moa.gov.cn/>. These data do not include agricultural subsidies.

<sup>2</sup> The data can be accessed at <https://www.qiyandata.com/home>.

the value-added of agricultural products, the extension of the agricultural industry chain, and the improvement of agricultural economic returns through horizontal, vertical, or supportive interactions (Jr Tabe Ojong and Dureti, 2023; Poulton *et al.*, 2010). Within the realm of grain cultivation, clusters also play a significant role. However, in comparison, there is a relative dearth of research focusing on grain clusters. Some studies have indicated that rice milling clusters can promote the adoption of improved milling technologies, thereby enhancing the competitiveness of domestically produced rice (Mano *et al.*, 2022). Additionally, research has shown that the development of the food manufacturing industry can alleviate local poverty and increase household income (Cazzuffi *et al.*, 2017; Takahashi and Barrett, 2014). Although the academic community has conducted preliminary explorations into the role of grain processing clusters, processing can be further differentiated into rough processing and deep processing. The impact of different types of grain processing clusters on farmers' behavioral decisions may vary. We systematically analyze grain processing clusters and their heterogeneous effects.

Second, our study contributes to the literature on rural inequality. There are many factors that affect rural income inequality, which can be summarized as physical capital, financial capital, human capital, political and social capital, and institutions (Bou Dib *et al.*, 2018; Batuo *et al.*, 2022; Glomm and Ravikumar, 1992). We focus on the new perspective provided by grain processing clusters for studying rural income inequality. The study most closely related to ours is Guo *et al.* (2022), which examines the relationship between industrial clusters and rural inequality, focusing primarily on how industrial clusters influence income inequality through non-farm employment channels, while overlooking their impacts on farm income and production. Parallel research by Berkes and Gaetani (2023) demonstrates that knowledge-intensive clusters may exacerbate inequality through three mechanisms: the geographic clustering of knowledge-intensive jobs, the sorting of residents by income, occupation, and education, and the feedback loop between rents and amenities. However, grain processing clusters, as a typical form of labor-intensive industrial agglomeration, exhibit distinct mechanisms and effects compared to the industrial or knowledge-intensive clusters emphasized in the existing literature. This study provides a comprehensive analysis of how grain processing clusters affect income inequality through both farm and non-farm income channels, while also identifying the underlying mechanisms, thereby addressing a critical gap in the existing literature.

Third, our study advances the understanding of poverty reduction by employing a novel “county-household” dual-level analytical framework. Using macro-level data to study inequality offers the advantages of broad research coverage and ease of cross-regional comparison. For instance, Cerina and Mureddu (2014) developed a two-region dynamic model to analyze the impact of agglomeration on efficiency, equity, and growth pathways. However, relying solely on macro-level data may obscure micro-level variations and fail to capture underlying mechanisms. Qiu *et al.* (2025), utilizing the China Household Finance Survey (CHFS) database, found that economic agglomeration benefits low-income rural populations by creating inclusive non-farm employment opportunities, thereby mitigating income inequality. Building on existing research, this paper first examines the impact of grain processing clusters on income inequality at the county level and then uses household data to explore micro-level mechanisms. This approach provides new empirical evidence for understanding the intrinsic mechanisms of poverty reduction and income growth among rural residents.

The rest of this paper is organized as follows: Section 2 introduces the background and theoretical framework of grain processing clusters. In Section 3, we describe the data sources, model settings, and variable descriptions. Sections 4 and 5 are the empirical results sections. Section 6 is the summary and recommendations.

## 2. Background and theoretical analysis

### 2.1 Development of agro-based clusters in China

Industrial clusters, an ancient and ubiquitous phenomenon, have always attracted the attention of numerous scholars. The most commonly used definition for industrial clusters refers to the geographical concentration of

interrelated firms and institutions (Porter, 1990). In the context of the agricultural sector, the term “agro-based clusters” is the preferred terminology, defined as the concentration of agricultural activities within a specific area that generates income and employment opportunities (Galvez-Nogales and Webber, 2017).

Given the presence of transaction costs, heavy dependence on inherent endowments, and intensive inputs of land and labor resources in agricultural production processes, agricultural activities tend to concentrate increasingly in advantaged regions. Within these regions, agricultural producers typically specialize in the production of a specific agricultural commodity, leading to rising levels of specialization, expanding scales of agricultural production and markets, and thereby attracting agricultural production-related sectors to agglomerate in these advantaged areas. Processing enterprises constitute a crucial component of this agglomeration mechanism. The geographical agglomeration of processing enterprises can facilitate the integration of smallholder and large-scale economies by reducing transaction costs, building regional and corporate brands, sharing public resources, and promoting technological innovation.

Due to restrictions on the free flow of land and labor factors, rural industrialization in China began with the development of township enterprises, which are defined as collectively owned (Weitzman and Xu, 1994), allowing rural residents to engage in industrial activities using collectively owned land and surplus labor. Since the late 1990s, with the gradual weakening of resistance to private ownership, many rural enterprises have been privatized, and the growth of the private sector has been associated with trends towards specialization and clustering of small enterprises.

With the concentration of numerous small specialized enterprises, many townships have become national or international centers for specific products, forming industrial clusters typically composed of privatized township enterprises or their derivative companies. In rural industries, the agricultural processing industry is the largest in scale, with the highest degree of industrial linkage and the broadest benefits for farmers; it is a key component of rural industries. By the end of 2023, more than 90 000 large-scale agricultural processing enterprises generated 14.35 trillion yuan in total revenue.<sup>3</sup> In 2023, China established 40 new advantaged industrial clusters and 200 strong agricultural towns. Driven by industry and employment, the per capita disposable income of rural residents reached 21,691 yuan, reflecting a real increase of 7.6% over the previous year.<sup>4</sup> These trends confirm that the agricultural processing industry has the greatest development potential in rural industries, plays a crucial role in extending the industrial chain, and significantly impacts agricultural production and farmer welfare. Therefore, studying the impact of processing clusters on inequality is of significant practical importance.

## *2.2 Potential channels for grain processing clusters to narrow income inequality*

Different income sources contribute differently to household income inequality. Household income comprises non-farm income, farm income, property income, and transfer income, with non-farm and farm incomes being the largest components (Cheng *et al.*, 2016; Wan and Zhou, 2005). Therefore, we primarily categorize farmer incomes into two types: non-farm and farm income. This allows us to examine how grain processing clusters affect income inequality in rural areas.

First, grain processing clusters can reduce the non-farm income gap among farmers by creating local employment opportunities. The division of labor within the cluster lowers entry barriers for entrepreneurs, thus generating more jobs (Ruan *et al.*, 2011). Rural low-income groups are often disadvantaged populations who “cannot move out” of their localities. When non-farm employment opportunities increase within the county, these groups are spared the migration costs of working elsewhere, and local employment is

<sup>3</sup> The data can be accessed at [https://www.gov.cn/lianbo/bumen/202312/content\\_6922349.htm](https://www.gov.cn/lianbo/bumen/202312/content_6922349.htm). In China, “large-scale agricultural processing enterprises” generally refer to agricultural processing enterprises with an annual main business revenue of 20 million yuan or more.

<sup>4</sup> The data can be accessed at [https://www.gov.cn/zhengce/202401/content\\_6927914.htm](https://www.gov.cn/zhengce/202401/content_6927914.htm).

more aligned with their employment needs (Liu *et al.*, 2014). Moreover, the grain processing industry is labor-intensive and low-tech (World Bank, 2008), generating non-farm jobs characterized by ease of learning and operation. This lowers the threshold for non-farm employment and is compatible with the human capital endowment of rural low-income groups (Duranton and Puga, 2024). Under such circumstances, rural low-income groups gain equitable access to non-farm employment opportunities. In contrast, high-income rural households typically hold stable jobs in urban areas or high-value-added sectors, where migrant wages often surpass local employment wages (Benjamin *et al.*, 2005). When the expected income from local non-farm employment is lower than their current income, groups with high human capital tend to migrate for work (Xing, 2014). Therefore, compared to high-income rural groups, grain processing clusters make a greater marginal contribution to the non-farm income of low-income groups. Based on this, we propose the following hypothesis:

**Hypothesis 1:** Grain processing clusters provide inclusive employment opportunities, narrowing the non-farm income gap among households.

Second, grain processing clusters can reduce the farm income gap among households by generating positive agglomeration externalities. In the agri-food value chain, processors play a crucial role (Reardon *et al.*, 2021). The backward linkages between grain processing and production create a multiplier effect on local farmers' income (Maertens *et al.*, 2012). Grain processing clusters influence farmers' income from grain cultivation through two primary channels. First, they foster a competitive market environment, preventing farmer exploitation by intermediaries, facilitating efficient product circulation, and boosting farmers' income (Barrett, 1997; Dorward *et al.*, 2004). Second, these clusters connect farmers with input suppliers, provide technical support, and offer premiums for high-quality products, thereby helping farmers benefit from trade (Bold *et al.*, 2022). As the opportunity cost of agricultural labor rises, it is rational for farmers to replace labor with capital. This shift accelerates agricultural capital deepening, improving production efficiency and product quality, which enables farmers to command higher market prices and increase non-farm income. Low-income farmers often lack access to market information, sales channels, and economies of scale, rendering them more susceptible to price undercutting by intermediaries. Additionally, constrained by credit availability, they face difficulties in investing in labor-saving mechanization. Grain processing clusters, therefore, can benefit low-income farmers by reducing intermediate links and lowering the threshold for technology adoption (Liu *et al.*, 2019). In contrast, high-income farmers typically have established stable sales channels and possess the capacity to independently invest in mechanization and advanced agricultural technologies, leading to relatively limited marginal gains from such clusters. Thus, increased farm income for low-income farmers narrows the income gap with other households. Based on this, we propose the following hypothesis:

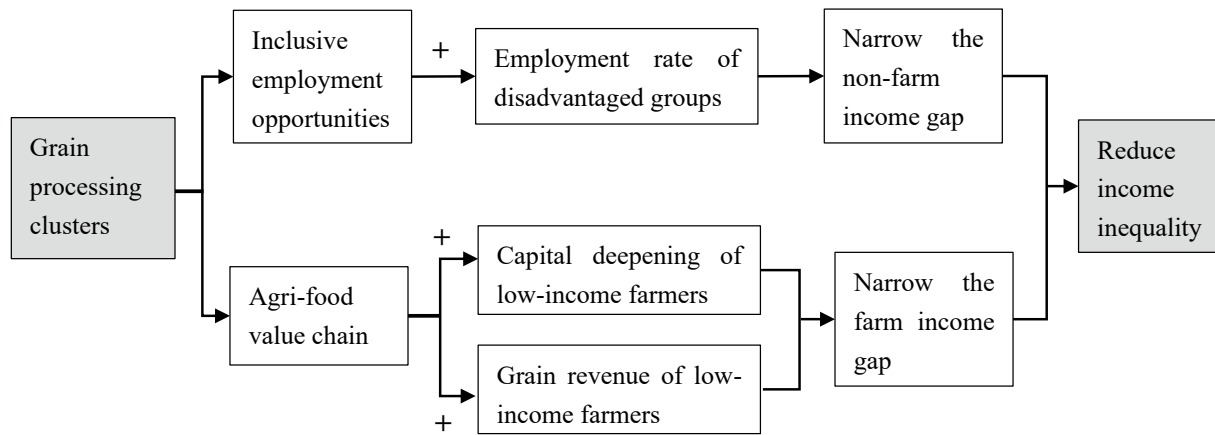
**Hypothesis 2:** Grain processing clusters have accelerated the deepening process of agricultural capital for low-income farmers, increased their grain returns, and thereby narrowed the farm income gap among households.

In summary, grain processing clusters can influence both local non-farm and farm sectors by providing local employment opportunities, accelerating the process of agricultural capital deepening, and increasing grain returns, facilitating low-income groups to access income growth pathways, and thereby reducing income inequality in rural areas. We summarize the mechanism framework in Figure 1.

### 3. Research design

#### 3.1 Data sources

Our primary data sources are the National Fixed Point Survey (NFP) and the China Academy for Rural Development-Qiyan China Agri-research Database (CCAD). We also use county-level control variables from the China Statistical Yearbook and County Statistical Yearbook. NFP is a nationwide survey database led by



**Figure 1.** Analytical framework.

the Ministry of Agriculture and Rural Affairs (MARA), encompassing longitudinal household-level survey data since 1986. It comprehensively documents various aspects of households, including family demographic structure, labor employment, income and expenditure, land contracting and transfer, agricultural production, social security, and village-level basic information, providing a holistic record of farmers' agricultural production, livelihoods, and employment status. Given that around 2003, China implemented a series of significant rural policy reforms that profoundly impacted grain production incentives and employment decisions, and considering that the calculation methods for several core variables in the NFP were adjusted and updated after 2003, we retain survey samples from 2004 to 2017 as our research dataset to ensure data continuity and research validity. CCAD contains micro-data of agricultural enterprises. Grain processing enterprises are categorized based on the "Classification of Agriculture and Related Industries",<sup>5</sup> and we used Python to clean and classify the data. Given the study's focus on household income, we excluded NFP samples with missing codes, unmatched records, or incomplete income variables. We then matched NFP farm household data with CCAD and other county-level data using county codes. The final sample includes 2300 county-years and 64 619 household-years.

### 3.2 Empirical strategy

We first investigate whether grain processing clusters affect income inequality in rural areas. The estimated equation is as follows:

$$\text{Inequality}_{j,t} = \alpha_1 + \beta_1 \text{Clusters}_{j,t-1} + \mathbf{X}_{j,t-1}\beta_2 + \text{Region}_j + \text{Year}_t + e_{j,t} \quad (1)$$

In the equation (1),  $j$  denotes the county and  $t$  denotes the year. The dependent variables are total income inequality, non-farm income inequality, and farm income inequality at the county level. We use the Gini coefficient to measure income inequality, ranging from 0 to 1, with higher values indicating greater inequality.<sup>6</sup>  $\text{Clusters}_{j,t-1}$  represents the density measure of grain processing clusters in county  $j$ , lagged by one period.<sup>7</sup> As a representative measure of industrial cluster level, density reflects the concentration of economic

<sup>5</sup> The "Classification of Agriculture and Related Industries" is a document issued by the National Bureau of Statistics of China, which is used to scientifically define the scope of statistics for agriculture and related industries, and to fully and accurately reflect the value of the entire industrial chain, including production, processing, manufacturing, distribution, and services in the forestry, animal husbandry, and fishery sectors. Accordingly, we identified grain rough processing and deep processing enterprises based on this document. For example, enterprises with the industry classification code "1311" (rice processing) under GB/T 4754 were categorized as rough processing enterprises; those with the code "1411" (bread manufacturing) under the same standard were identified as deep processing enterprises.

The document can be accessed at [https://www.stats.gov.cn/sj/tjbz/gjtjbz/202302/t20230213\\_1902783.html](https://www.stats.gov.cn/sj/tjbz/gjtjbz/202302/t20230213_1902783.html).

<sup>6</sup> The formula for calculating the Gini coefficient is detailed in formula (A1) in the Appendix.

<sup>7</sup> To mitigate potential reverse causality issues, we apply a lag treatment to the core explanatory variables and control variables in the equation.

activities in a region. Specifically, we measure grain processing clusters by the number of enterprises per unit area.  $\mathbf{X}_{j,t-1}$  is a vector of county characteristics influencing income inequality besides clusters.  $\text{Region}_j$  are region dummies for counties, and  $\text{Year}_t$  are year dummies.  $\alpha_1$  is the intercept term.  $\beta_1$  and  $\beta_2$  are estimated coefficients, and  $e_{j,t}$  is the random disturbance term.

To investigate which income groups are specifically affected by grain processing clusters, thus leading to income inequality, our estimated equation is as follows:

$$\text{Income}_{i,j,t} = \alpha_2 + \gamma_1 \text{Clusters}_{j,t-1} + \mathbf{Z}_{i,t}\gamma_2 + \mathbf{X}_{j,t-1}\gamma_3 + \text{Region}_j + \text{Year}_t + \varepsilon_{i,j,t} \quad (2)$$

In the equation (2),  $i$  represents the  $i$ -th rural household in county  $j$ . The dependent variables are per capita non-farm income and per capita farm income (logarithm) of rural households. Non-farm income refers to household wage income, while farm income is income from agricultural operations.<sup>8</sup> We ranked rural households in each county by income, dividing them into: the richest 20% by non-farm income (Nonfarm R20), the poorest 20% by non-farm income (Nonfarm P20), the richest 20% by farm income (Farm R20), and the poorest 20% by farm income (Farm P20). We used  $\text{Clusters}_{j,t-1}$  to regress the corresponding income of these four groups.  $\mathbf{Z}_{i,t}$  is a vector of household and village characteristics that might influence income.  $\alpha_2$  is the intercept term.  $\gamma_1$ ,  $\gamma_2$  and  $\gamma_3$  are the estimated coefficients.  $\varepsilon_{i,j,t}$  is the random disturbance term. The definitions of the core independent variables,  $\mathbf{X}_{j,t-1}$  and  $\text{Year}_t$  are the same as in Equation (1).  $\text{Region}_j$  are region dummies for provinces.

### 3.3 Sample characteristics and descriptive statistics

Table 1 provides the definitions and descriptions of the selected variables. The variables are grouped into two categories: county-level characteristics and household-level characteristics of rural households and their villages. At the county level, our model controls for GDP, city scale, financial development, and the share of secondary and tertiary industries. At the household level, we control for the household head's age, gender, and education, as well as household size, family health status, and geographic factors such as distance to the nearest highway, terrain, and village location. In terms of personal characteristics, approximately 95% of the household heads in the survey sample were male, the average age of the household heads was approximately 54 years old, they had received an average of 7 years of education, and each sample household had an average of approximately four family members, half of whom were in good health.

## 4. Empirical findings

### 4.1 Main results

Table 2 presents the estimated effects of grain processing clusters on income inequality. Columns (1) and (2), respectively, present the regression results of rough processing and deep processing on total income inequality, both indicating significantly negative correlations. This suggests that both cluster types can effectively reduce total income inequality. Specifically, a one-standard-deviation increase (0.02 units) in rough processing clusters is associated with an approximately 0.012-unit reduction in total income inequality. The magnitude of this equalizing effect is comparable for deep processing clusters, where a similar one-standard-deviation expansion (0.02 units) yields an equivalent 0.012-unit decrease in inequality.

Next, Columns (3) and (4) of Table 2 present the effects of the two types of clusters on non-farm income inequality. A one-standard-deviation increase in rough processing clusters is associated with a 0.008-unit reduction in non-farm income inequality (significant at the 10% level), whereas the same increase in

<sup>8</sup> This paper focuses on income directly related to rural industrialization, hence when defining non-farm income, income from interests, pensions, government transfers, and subsidies is not considered.

**Table 1.** Summary statistics

| VarName                                       | Definition                                                                                                        | Mean  | SD    | Min   | Max   |
|-----------------------------------------------|-------------------------------------------------------------------------------------------------------------------|-------|-------|-------|-------|
| County level variables ( <i>N</i> =2300)      |                                                                                                                   |       |       |       |       |
| Total income inequality                       | County-level total income inequality measured by the Gini coefficient (0–1)                                       | 0.33  | 0.10  | 0.08  | 0.81  |
| Nonfarm income inequality                     | County-level nonfarm income inequality measured by the Gini coefficient (0–1)                                     | 0.33  | 0.09  | 0.13  | 0.58  |
| Farm income inequality                        | County-level farm income inequality measured by the Gini coefficient (0–1)                                        | 0.46  | 0.16  | 0.01  | 0.99  |
| Rough processing clusters                     | The number of rough processing enterprises per unit area of the county (households/square kilometer)              | 0.01  | 0.02  | 0.00  | 0.25  |
| Deep processing clusters                      | The number of deep processing enterprises per unit area of the county (households/square kilometer)               | 0.01  | 0.02  | 0.00  | 0.18  |
| GDP                                           | Regional GDP (logarithm)                                                                                          | 13.97 | 0.97  | 11.00 | 18.20 |
| City scale                                    | The total population at the end of the year in the region (logarithm)                                             | 4.00  | 0.60  | 1.61  | 5.63  |
| Finance                                       | The ratio of the total amount of loans from financial institutions at the end of the year to the regional GDP (%) | 0.55  | 0.33  | 0.08  | 3.71  |
| Secondary industry                            | The proportion of the added value of the secondary in the regional GDP (%)                                        | 0.44  | 0.14  | 0.06  | 0.90  |
| Tertiary industry                             | The proportion of the added value of the tertiary in the regional GDP (%)                                         | 0.35  | 0.09  | 0.08  | 0.67  |
| Household level variables ( <i>N</i> =64 541) |                                                                                                                   |       |       |       |       |
| Nonfarm income                                | Per capita non-farm income of rural households (logarithm)                                                        | 8.30  | 1.02  | 2.30  | 12.94 |
| Farm income                                   | Per capita farm income of rural households (logarithm)                                                            | 7.95  | 1.21  | 0.34  | 21.63 |
| Age                                           | Age of the household head                                                                                         | 53.75 | 10.38 | 18.00 | 96.00 |
| Gender                                        | Gender of the household head (male=1)                                                                             | 0.95  | 0.21  | 0.00  | 1.00  |
| Education                                     | Years of education received by the household head (years)                                                         | 6.82  | 2.53  | 0.00  | 20.00 |
| Household size                                | The number of household members                                                                                   | 4.14  | 1.53  | 1.00  | 8.00  |
| Health status                                 | The proportion of people in good health in the household (%)                                                      | 0.56  | 0.38  | 0.00  | 1.00  |
| Distance to highway                           | The distance from the household's village to the highway (km)                                                     | 2.52  | 3.80  | 0.00  | 26.00 |
| Terrain                                       | Whether the village terrain is hilly (yes=1)                                                                      | 0.62  | 0.49  | 0.00  | 1.00  |
| Village location                              | Whether the village is located in a suburban area (yes=1)                                                         | 0.14  | 0.35  | 0.00  | 1.00  |

**Table 2.** Grain processing clusters and income inequality

|                           | Total income inequality |                     | Nonfarm income inequality |                      | Farm income inequality |                     |
|---------------------------|-------------------------|---------------------|---------------------------|----------------------|------------------------|---------------------|
|                           | (1)                     | (2)                 | (3)                       | (4)                  | (5)                    | (6)                 |
| Rough processing clusters | −0.597***<br>(0.197)    |                     | −0.403*<br>(0.232)        |                      | −0.524<br>(0.552)      |                     |
| Deep processing clusters  |                         | −0.591**<br>(0.284) |                           | −0.408***<br>(0.121) |                        | −1.948**<br>(0.840) |
| Controls                  | Yes                     | Yes                 | Yes                       | Yes                  | Yes                    | Yes                 |
| Year                      | Yes                     | Yes                 | Yes                       | Yes                  | Yes                    | Yes                 |
| Region                    | Yes                     | Yes                 | Yes                       | Yes                  | Yes                    | Yes                 |
| Constant                  | 0.337<br>(0.200)        | 0.332<br>(0.207)    | 0.550**<br>(0.203)        | 0.263***<br>(0.087)  | 0.079<br>(0.364)       | 0.079<br>(0.371)    |
| <i>N</i>                  | 2143                    | 2143                | 2112                      | 2112                 | 2133                   | 2133                |
| <i>R</i> <sup>2</sup>     | 0.035                   | 0.033               | 0.483                     | 0.178                | 0.109                  | 0.118               |

Robust standard errors clustered at the provincial level are reported in parentheses. \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

deep processing clusters corresponds to a more precisely estimated 0.008-unit decrease (significant at the 1% level). This result indicates that the effect of deep processing clusters is more statistically significant.

We further examined the impact of clusters on farm income inequality. Columns (5) and (6) shows that the rough processing cluster has no significant effect, while the deep processing cluster significantly reduces farm income inequality at the 5% level. A one-standard-deviation increase in deep processing clusters decreases farm income inequality by 0.039 units. This suggests that deep processing clusters significantly affect farm income distribution within counties.

In sum, the results from Table 2 show that both rough and deep processing clusters have a negative impact on income inequality, but the types of effects differ. Rough processing clusters mainly reduce total income inequality by lowering non-farm income inequality. In contrast, deep processing clusters exert a comprehensive effect on both non-farm and farm income inequality, ultimately leading to a reduction in total income inequality.

Rough processing clusters and deep processing clusters exert heterogeneous impacts on reducing farm income inequality, which constitutes an intriguing finding. On the one hand, rough processing enterprises, primarily engaged in basic milling and threshing, face low technological barriers and severe product homogenization. This results in relatively weak market bargaining power and limited capacity to boost farmers' income. In contrast, deep processing clusters can enhance the added value of grain crops through premiums from differentiated products and enable participating farmers to share in the value increments of the industrial chain via contractual models, thereby effectively increasing smallholders' grain cultivation earnings. On the other hand, rough processing clusters primarily demand standardized raw grains, allowing farmers to engage in production without additional investment. Conversely, deep processing clusters require specific crop varieties and quality control, which incentivizes farmers to adopt improved seeds, precision fertilization, and other technologies. Through knowledge spillover effects, these practices enhance farmers' production efficiency (Duranton and Puga, 2004), ultimately yielding positive impacts on farmers' income.

In order to further analyze which type of households are affected by the grain processing clusters and thus lead to income inequality, we used household-level data to regress clusters on non-farm income and farm income, respectively. This analysis is divided into two parts in Table 3. The results for non-farm income are shown first. For the Nonfarm R20 group, neither type of cluster significantly affects per capita non-farm income. However, for the Nonfarm P20 group, both clusters are significantly correlated with per capita non-farm income growth.

**Table 3.** Grain processing clusters and income of different groups

|                           | Nonfarm R20         |                     | Nonfarm P20         |                     |
|---------------------------|---------------------|---------------------|---------------------|---------------------|
|                           | (1)                 | (2)                 | (3)                 | (4)                 |
| Non-farm income           |                     |                     |                     |                     |
| Rough processing clusters | 0.338<br>(1.594)    |                     | 4.597*<br>(2.483)   |                     |
| Deep processing clusters  |                     | −1.511<br>(1.452)   |                     | 7.216**<br>(3.278)  |
| Controls                  | Yes                 | Yes                 | Yes                 | Yes                 |
| Year dummy                | Yes                 | Yes                 | Yes                 | Yes                 |
| Region dummy              | Yes                 | Yes                 | Yes                 | Yes                 |
| Constant                  | 6.982***<br>(0.576) | 6.884***<br>(0.592) | 4.432***<br>(1.069) | 4.464***<br>(1.073) |
| <i>N</i>                  | 12562               | 12562               | 10317               | 10317               |
| <i>R</i> <sup>2</sup>     | 0.639               | 0.639               | 0.532               | 0.532               |
| Farm income               |                     |                     |                     |                     |
| Rough processing clusters | −3.404<br>(2.350)   |                     | 1.916<br>(2.888)    |                     |
| Deep processing clusters  |                     | −1.991<br>(1.550)   |                     | 7.110***<br>(1.975) |
| Controls                  | Yes                 | Yes                 | Yes                 | Yes                 |
| Year dummy                | Yes                 | Yes                 | Yes                 | Yes                 |
| Region dummy              | Yes                 | Yes                 | Yes                 | Yes                 |
| Constant                  | 3.330***<br>(0.790) | 3.437***<br>(0.788) | 5.481***<br>(1.132) | 5.658***<br>(1.086) |
| <i>N</i>                  | 15249               | 15249               | 13290               | 13290               |
| <i>R</i> <sup>2</sup>     | 0.425               | 0.425               | 0.398               | 0.405               |

Robust standard errors clustered at the county level are reported in parentheses. \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

A one-standard-deviation increase in rough processing clusters leads to a 9.194% rise in per capita non-farm income (significant at the 10% level), equivalent to an average increase of approximately 569.608 yuan. Meanwhile, a one-standard-deviation expansion in deep processing clusters results in a more substantial 14.432% growth (significant at the 5% level), corresponding to an average gain of around 894.124 yuan.<sup>9</sup> These results suggest that both cluster types primarily reduce inequality by boosting non-farm income in the *Nonfarm P20* group, with deep processing clusters having a stronger effect. This difference may stem from the more complex, multi-step nature of deep processing, which generates more job opportunities and increases wage or business income for farmers.

The analysis now considers the effects on farm income in Table 3. The results show that rough processing clusters have no significant effect on per capita farm income, while deep processing clusters positively affect the per capita farm income of the Farm P20 group. A one-standard-deviation increase in deep processing clusters leads to a 14.220% increase in per capita farm income (significant at the 1% level), translating to an average increase of approximately 769.320 yuan. These findings suggest that deep processing clusters primarily reduce inequality by boosting farm income in the Farm P20 group, while rough processing clusters have no significant impact. This difference highlights that deep processing enterprises add more value to

<sup>9</sup> In the sample, the average per capita non-farm income is 6195.427 yuan, while the average per capita farm income stands at 5410.126 yuan. By multiplying the growth rate with the mean income values, we can derive the monetary equivalent of income growth at the average level. Similar calculations will be performed hereafter.

products and elevate farmers' status in the agri-food value chain, increasing their profit-sharing opportunities through localized production and sales.

#### 4.2 Endogeneity discussion

To ensure the core estimated coefficients in the baseline regression model are consistent estimators, two conditions must be met: first, the equation must control for confounding factors that simultaneously affect both clusters and income inequality; second, reverse causality must be ruled out. Although we have addressed these concerns in our baseline model specification by lagging the independent variables by one period and incorporating potential confounders as control variables to the greatest extent possible, real-world conditions make it challenging to fully satisfy the exogeneity requirements. Therefore, this study employs the Two-Stage Least Squares (2SLS) method to address potential endogeneity issues. Specifically, we used agro-based clusters at the city level as instrumental variables for county-level grain processing clusters (IV).<sup>10</sup> In economic research, aggregation data at higher levels are often used as instruments for more localized explanatory variables (Card and Krueger, 1996). Municipal-level agro-based clusters, as an instrumental variable, satisfy three conditions: First, in terms of relevance, within the administrative jurisdiction of a prefecture-level city, municipal-level industrial policy support, infrastructure investment, and factor market integration directly radiate and drive the development of industrial clusters in subordinate counties. Thus, municipal-level agro-based clusters are closely associated with the development of county-level grain processing clusters. Second, regarding exogeneity, planning decisions for municipal-level agro-based clusters are typically based on regional development strategies rather than the income distribution status of specific counties, making them uncorrelated with the error term. Third, in terms of exclusion restriction, municipal-level industrial policies must generate economic effects through county-level administrative implementation and responses from market entities, and thus have no direct impact on county-level income distribution.

Tables 4 and 5 present the 2SLS regression results for deep processing clusters on three types of income inequality. Table 4 shows that the IV coefficients are significantly positive at the 1% level, confirming the relevance of the instruments. Table 5 presents the second-stage 2SLS estimates. The findings for deep processing clusters, shown in columns (1)–(3), indicate a significant negative effect on inequality. In contrast, the results for rough processing clusters in columns (4)–(6) show that they significantly reduce total and non-farm income inequality, but have an insignificant effect on farm income inequality. These results align with those in Table 2. Additionally, the Cragg-Donald Wald F statistics exceed the critical value of 16.38 for the Stock-Yogo weak instrument test, indicating no issues with weak instruments. Therefore, after addressing potential endogeneity with instrumental variables, the regression results remain robust.

#### 4.3 Robustness checks

We also use additional robustness checks to validate the findings. First, we re-estimated the equation using the Theil index to measure income inequality.<sup>11</sup> As shown in Table 6, after redefining the dependent variable, clusters continue to reduce income inequality in rural areas. Deep processing clusters decrease total income inequality and significantly reduce both non-farm and farm income inequality. Rough processing clusters primarily affect total income inequality by reducing non-farm income inequality. These findings are consistent with the results presented in Table 2.

Second, following Guo (2022), we use the DBI method to measure grain processing clusters. This index is based on the relative density of firms in the same industry within a geographic area, capturing the clustering characteristic. Unlike the density method, which treats clusters as a continuous variable, the DBI index is a

<sup>10</sup> The agro-based clusters are measured by the number of agriculture and agriculture-related industry enterprises within the administrative area of city level, with data sourced from CCAD.

<sup>11</sup> The formula for calculating the Theil index is detailed in formula (A2) in the Appendix.

**Table 4.** Results of the first-stage regression of 2SLS

|              | Deep processing clusters |                     |                     | Rough processing clusters |                    |                     |
|--------------|--------------------------|---------------------|---------------------|---------------------------|--------------------|---------------------|
|              | (1)                      | (2)                 | (3)                 | (4)                       | (5)                | (6)                 |
| IV           | 0.010***<br>(0.003)      | 0.010***<br>(0.003) | 0.010***<br>(0.003) | 0.009**<br>(0.003)        | 0.009**<br>(0.003) | 0.011***<br>(0.004) |
| Controls     | Yes                      | Yes                 | Yes                 | Yes                       | Yes                | Yes                 |
| Year dummy   | Yes                      | Yes                 | Yes                 | Yes                       | Yes                | Yes                 |
| Region dummy | Yes                      | Yes                 | Yes                 | Yes                       | Yes                | Yes                 |

Robust standard errors clustered at the municipal level are reported in parentheses. \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

**Table 5.** Results of the second-stage regression of 2SLS

|                           | Total income inequality<br>(1) | Nonfarm income inequality<br>(2) | Farm income inequality<br>(3) | Total income inequality<br>(4) | Nonfarm income inequality<br>(5) | Farm income inequality<br>(6) |
|---------------------------|--------------------------------|----------------------------------|-------------------------------|--------------------------------|----------------------------------|-------------------------------|
| Deep processing clusters  | -2.353*<br>(1.300)             | -2.762**<br>(1.338)              | -4.847***<br>(1.708)          |                                |                                  |                               |
| Rough processing clusters |                                |                                  |                               | -2.715<br>(1.691)              | -3.099*<br>(1.782)               | 1.674<br>(1.684)              |
| Controls                  | Yes                            | Yes                              | Yes                           | Yes                            | Yes                              | Yes                           |
| Year dummy                | Yes                            | Yes                              | Yes                           | Yes                            | Yes                              | Yes                           |
| Region dummy              | Yes                            | Yes                              | Yes                           | Yes                            | Yes                              | Yes                           |
| <i>N</i>                  | 1941                           | 1918                             | 1932                          | 1941                           | 1918                             | 1938                          |
| Cragg–Donald–Wald         | 174.552                        | 167.180                          | 173.392                       | 73.030                         | 74.531                           | 105.519                       |
| <i>F</i> statistic        |                                |                                  |                               |                                |                                  |                               |
| <i>R</i> <sup>2</sup>     | 0.006                          | -0.002                           | 0.073                         | -0.033                         | -0.083                           | 0.238                         |

Robust standard errors clustered at the municipal level are reported in parentheses. \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

**Table 6.** Robustness tests for inequality measured by the Theil inde

|                           | (1)<br>Total income inequality | (2)<br>Nonfarm income inequality | (3)<br>Farm income inequality | (4)<br>Total income inequality | (5)<br>Nonfarm income inequality |
|---------------------------|--------------------------------|----------------------------------|-------------------------------|--------------------------------|----------------------------------|
| Deep processing clusters  | -0.820**<br>(0.314)            | -0.697*<br>(0.381)               | -1.781**<br>(0.845)           |                                |                                  |
| Rough processing clusters |                                |                                  |                               | -0.707**<br>(0.268)            | -0.351*<br>(0.203)               |
| Controls                  | Yes                            | Yes                              | Yes                           | Yes                            | Yes                              |
| Year dummy                | Yes                            | Yes                              | Yes                           | Yes                            | Yes                              |
| Region dummy              | Yes                            | Yes                              | Yes                           | Yes                            | Yes                              |
| Constant                  | 0.244<br>(0.261)               | 0.689**<br>(0.278)               | -0.534<br>(0.509)             | 0.252<br>(0.256)               | 0.368***<br>(0.129)              |
| <i>N</i>                  | 2047                           | 2047                             | 2047                          | 2047                           | 2047                             |
| <i>R</i> <sup>2</sup>     | 0.037                          | 0.020                            | 0.043                         | 0.038                          | 0.415                            |

Robust standard errors clustered at the provincial level are reported in parentheses. \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

dummy variable. If a county ranks in the top  $\alpha$  percentile of firm density, with  $\alpha$  set to 3 or 10, it indicates the presence of at least one cluster, and the variable takes a value of 1; otherwise, it is 0. Table 7 presents regression results using the DBI index to assess the impact of grain processing clusters on inequality. Columns (1)–(3) show that deep processing clusters are significantly negatively correlated with total, non-farm, and farm income inequality. For example, counties with deep processing clusters have an average inequality level 0.034 lower than those without. Columns (4)–(5) show that rough processing clusters are significantly negatively correlated with total and non-farm income inequality. These results confirm the robustness of the baseline regression findings.

Third, we re-examined the impact of grain processing clusters on households' income using the quantile regression method.<sup>12</sup> The theoretical basis of the quantile regression model was proposed by Koenker and Bassett (1978), and can be used to explain the impact of explanatory variables on the explained variables at different quantiles. As can be seen in Figure 2, as the quantile increases, the quantile regression coefficients for clusters exhibit a declining trend, indicating that the impact of clusters on the lower tail of the income distribution is less than their impact on the upper segments. In other words, the increase in cluster density has a greater effect on low-income individuals compared to high-income individuals. This finding supports the research conclusions presented in Table 3. The above tests confirm that grain processing clusters can indeed reduce income inequality by increasing the income of low-income households.

#### 4.4 Mechanisms of grain processing clusters on household incomes

The previous text has explored the correlation between grain processing clusters and non-farm income of farmers. Building on this, we investigate the underlying mechanisms. According to theoretical analysis, grain processing clusters primarily affect non-farm income by providing local employment opportunities. The empirical results are shown in Table 8. Specifically, we calculated the proportion of local employment in the household (ratio of local\_emp), the proportion of local employment among vulnerable groups in the household (ratio of vulnerable\_emp).<sup>13</sup> Columns (1) and (3) in Table 8 display the impact of two types of processing

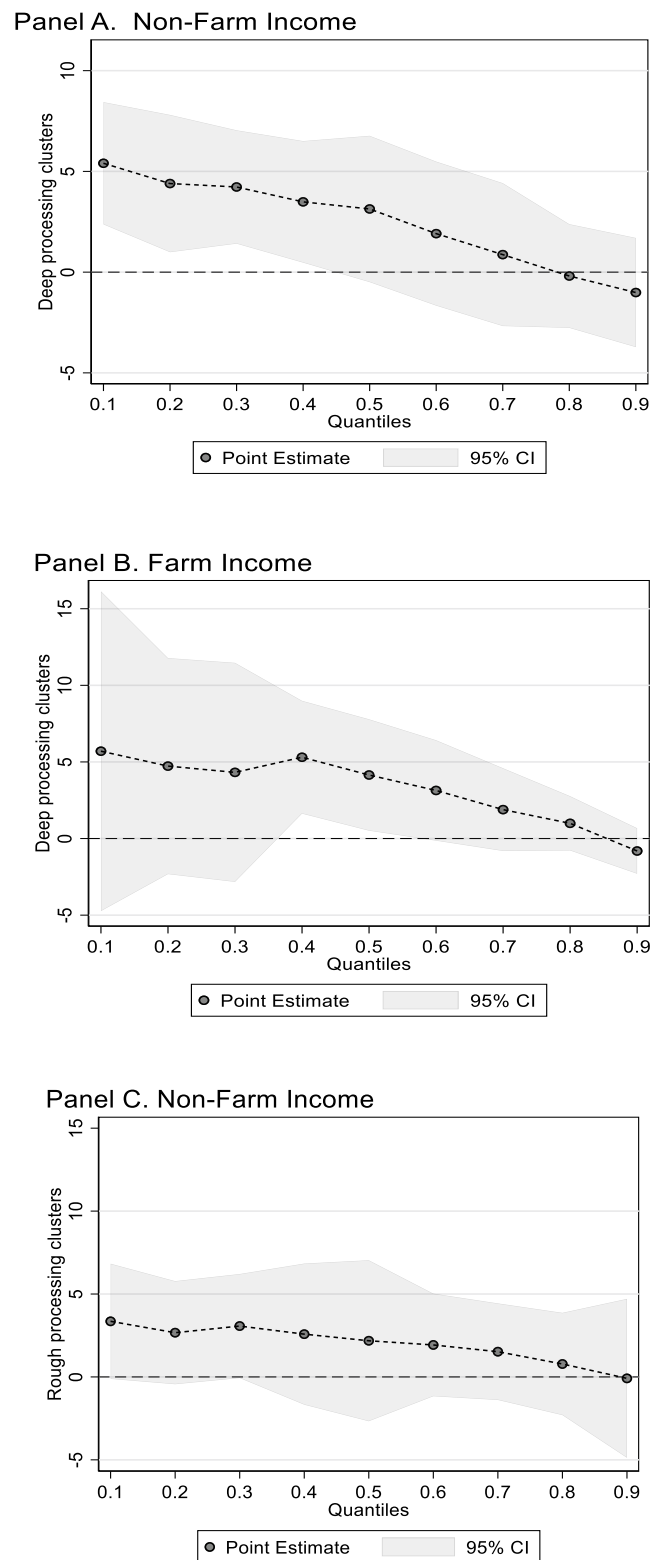
**Table 7.** Robustness tests for cluster measurement using the DBI index

|                           | (1)<br>Total income<br>inequality | (2)<br>Nonfarm income<br>inequality | (3)<br>Farm income<br>inequality | (4)<br>Total income<br>inequality | (5)<br>Nonfarm income<br>inequality |
|---------------------------|-----------------------------------|-------------------------------------|----------------------------------|-----------------------------------|-------------------------------------|
| Deep processing clusters  | −0.034**<br>(0.012)               | −0.028**<br>(0.014)                 | −0.064**<br>(0.026)              |                                   |                                     |
| Rough processing clusters |                                   |                                     |                                  | −0.016**<br>(0.008)               | −0.018**<br>(0.008)                 |
| Controls                  | Yes                               | Yes                                 | Yes                              | Yes                               | Yes                                 |
| Year dummy                | Yes                               | Yes                                 | Yes                              | Yes                               | Yes                                 |
| Region dummy              | Yes                               | Yes                                 | Yes                              | Yes                               | Yes                                 |
| Constant                  | 0.321<br>(0.206)                  | 0.533***<br>(0.161)                 | 0.056<br>(0.361)                 | 0.322<br>(0.203)                  | 0.883***<br>(0.297)                 |
| <i>N</i>                  | 2143                              | 2112                                | 2133                             | 2143                              | 2112                                |
| <i>R</i> <sup>2</sup>     | 0.034                             | 0.483                               | 0.113                            | 0.033                             | 0.052                               |

Robust standard errors clustered at the provincial level are reported in parentheses. \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

<sup>12</sup> The formula of quantile regression is detailed in formula (A3) in the Appendix.

<sup>13</sup> The ratio of local\_emp is calculated as the number of non-farm employed workers in the local area divided by the total household population. The ratio of low-skill\_emp is calculated as the number of vulnerable groups in non-farm local employment divided by the total household population. Vulnerable groups primarily include laborers with education levels below junior high school, individuals aged 55 and above, and laborers with poor health conditions.



**Figure 2.** Quantile regression of grain processing clusters on households income. (A) Results of a quantile regression of deep processing clusters on household nonfarm income. (B) Results of a quantile regression of deep processing clusters on household farm income. (C) Results of a quantile regression of rough processing clusters on household nonfarm income. Shown are the regression coefficients and the associated 95% confidence intervals.

**Table 8.** Impact of grain processing clusters on the employment of households

|                           | (1)<br>Nonfarm R20 ratio<br>of local_emp | (2)<br>Nonfarm P20 ratio<br>of vulnerable_emp | (3)<br>Nonfarm R20 ratio<br>of local_emp | (4)<br>Nonfarm P20 ratio<br>of vulnerable_emp |
|---------------------------|------------------------------------------|-----------------------------------------------|------------------------------------------|-----------------------------------------------|
| Deep processing clusters  | 0.043<br>(0.669)                         | 0.956***<br>(0.281)                           |                                          |                                               |
| Rough processing clusters |                                          |                                               | 0.429<br>(0.509)                         | 0.302<br>(0.464)                              |
| Controls                  | Yes                                      | Yes                                           | Yes                                      | Yes                                           |
| Year dummy                | Yes                                      | Yes                                           | Yes                                      | Yes                                           |
| Region dummy              | Yes                                      | Yes                                           | Yes                                      | Yes                                           |
| Constant                  | 0.383<br>(0.242)                         | 0.560***<br>(0.166)                           | 0.410*<br>(0.239)                        | 0.520***<br>(0.169)                           |
| <i>N</i>                  | 4118                                     | 3196                                          | 4118                                     | 3196                                          |
| <i>R</i> <sup>2</sup>     | 0.180                                    | 0.375                                         | 0.180                                    | 0.373                                         |

Robust standard errors clustered at the municipal level are reported in parentheses. \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

clusters on the Ratio of local\_emp for Nonfarm R20 group households. The results indicate that neither type of cluster significantly promotes non-farm local employment among high-income groups, suggesting that employment opportunities in the grain processing industry do not have strong appeal for high-income individuals. Columns (2) and (4) in Table 8 show the impact of the two types of processing clusters on the Ratio of vulnerable\_emp for Nonfarm P20 group households. It is found that deep processing clusters can significantly promote local employment among vulnerable groups in low-income households, while the effect of rough processing clusters is not significant. Grain deep processing involves more operational stages, which enables it to provide more inclusive employment opportunities for local farmers. Consequently, deep processing clusters can exert a significant impact on facilitating non-farm employment among disadvantaged groups. Synthesizing the conclusions from Table 8, we find that deep processing clusters, by providing inclusive employment opportunities, promote local employment for vulnerable groups, thereby increasing non-farm income for low-income households and reducing non-farm income inequality. Thus, Hypothesis 1 is confirmed.

Second, previous research found a correlation between deep processing clusters and farm income, hence in this section we explore the potential mechanisms through which deep processing clusters affect farm income. Theoretical analysis suggests that deep processing clusters primarily affect farm income by promoting the deepening of agricultural capital and increasing grain returns. The empirical results are shown in Table 9. Specifically, we use the growth rate of the ratio of capital input to labor input in grain crops as a proxy indicator for the deepening of agricultural capital (Capital deepening),<sup>14</sup> and the logarithm of grain income per unit area as a measure of grain returns (Grain revenue), and conducted regressions on the Farm R20 and Farm P20 samples, respectively. It can be seen that in the Farm R20 group, deep processing clusters have no significant impact on capital deepening and grain returns. However, in the Farm P20 group, the coefficients of deep processing clusters are significantly positive, indicating that the externalities of deep processing clusters have accelerated the capital deepening of low-income farmers and increased their grain returns, which helps to narrow the farm income gap among households. Thus, Hypothesis 2 is confirmed.

<sup>14</sup> The capital investment in grain crops includes expenses for seeds and seedlings, fertilizers, agricultural films, pesticides, water and electricity, as well as irrigation costs, and machinery operation fees. The labor input refers to the number of days of direct labor engaged in the production and operation of grain. Since the NFP data do not have ready-made data for grain, the values for rice, wheat, and corn, the three main grain crops, are calculated separately first, and then a weighted calculation is performed. The weights are based on the proportion of the areas of the three main grain crops to the total area.

**Table 9.** Impact of deep processing clusters on the operations of households

|                          | Capital deepening  |                     | Grain revenue       |                     |
|--------------------------|--------------------|---------------------|---------------------|---------------------|
|                          | (1)<br>Farm R20    | (2)<br>Farm P20     | (3)<br>Farm R20     | (4)<br>Farm P20     |
| Deep processing clusters | 0.232<br>(0.494)   | 0.794*<br>(0.465)   | 2.515<br>(1.601)    | 3.183***<br>(1.115) |
| Controls                 | Yes                | Yes                 | Yes                 | Yes                 |
| Year dummy               | Yes                | Yes                 | Yes                 | Yes                 |
| Region dummy             | Yes                | Yes                 | Yes                 | Yes                 |
| Constant                 | 0.459**<br>(0.173) | 0.623***<br>(0.205) | 5.364***<br>(0.826) | 5.418***<br>(0.815) |
| <i>N</i>                 | 9902               | 7957                | 11024               | 9776                |
| <i>R</i> <sup>2</sup>    | 0.043              | 0.042               | 0.245               | 0.282               |

Robust standard errors clustered at the provincial level are reported in parentheses. \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

Deep processing clusters systematically narrow income disparities among farmers through two mutually reinforcing pathways: employment creation and agricultural value addition. Specifically, for rural surplus labor with limited employment advantages, deep processing clusters offer inclusive employment opportunities, enabling low-income groups to secure stable wage income (Pathway 1). Growth in non-farm income prompts farmers to reinvest in agriculture by optimizing factor input structures, such as increasing short-term investments or purchasing mechanized services to replace labor, thereby indirectly strengthening the agricultural capital deepening effect of Pathway 2. Meanwhile, through contractual models like contract farming, deep processing clusters allow farmers who have upgraded their technologies to obtain price premiums (Pathway 2). The subsequent rise in agricultural income reduces low-income groups' dependence on non-farm employment, enabling them to focus on leveraging their comparative advantages and thus optimizing the labor supply structure under Pathway 1.

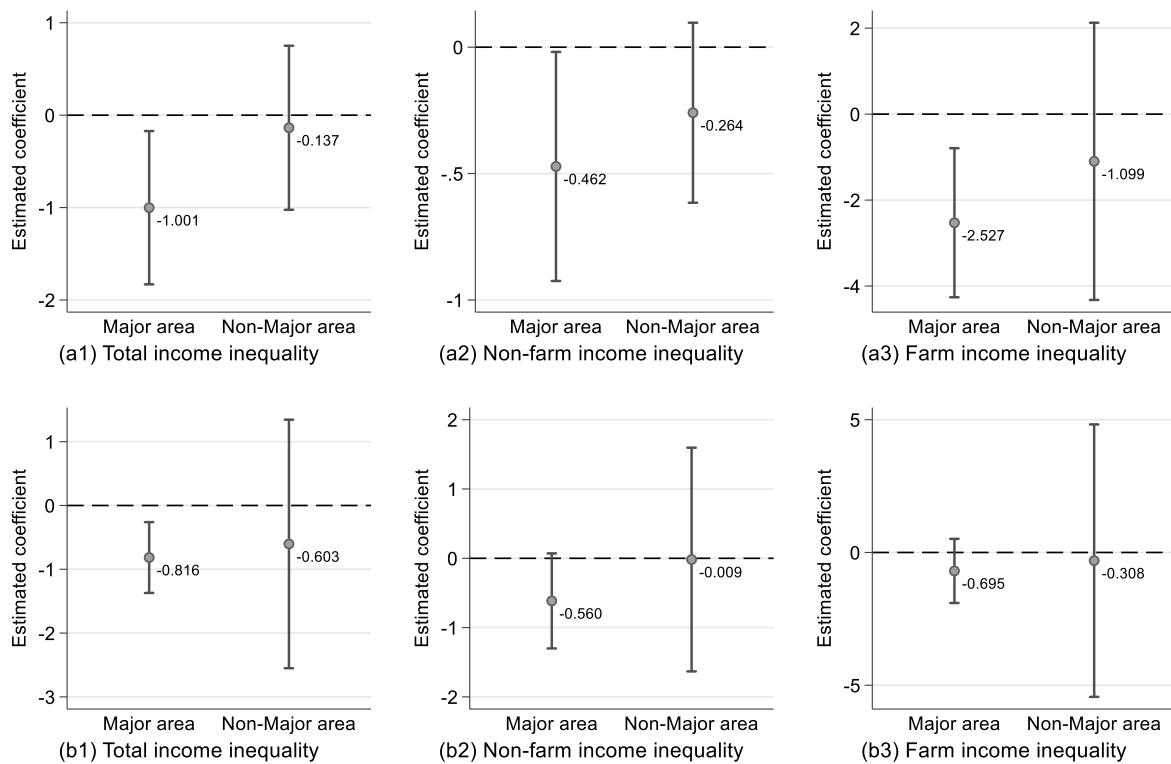
## 5. Heterogeneity test

### 5.1 Heterogeneity analysis by grain functional areas

Previous studies have confirmed that grain processing clusters have a negative impact on income inequality, but whether this effect holds across different regions still warrants further exploration. To adapt to the new pattern of grain production changes, China officially designated 13 provinces as major grain-producing areas by the end of 2003, with the aim of ensuring food security and promoting the growth of grain output. Data shows that the contribution of the thirteen major grain-producing areas to domestic grain supply has continued to strengthen, with grain output accounting for a consistently high proportion of the national total, increasing from 76.86% in 2012 to 78.25% in 2022.<sup>15</sup> However, the pursuit of grain output has also led to the neglect of the operational efficiency of grain-producing areas, causing misallocation of agricultural production resources (Clapp, 2017) and resulting in significant efficiency losses (Magnan *et al.*, 2011). Policies for major grain-producing areas, including national food security industry belt construction policies and grain crop cultivation and seed subsidy policies, help promote the concentration of various production factors in advantageous regions, enhancing the concentration and competitiveness of grain enterprises (Gao and Wei, 2021). Therefore, we divide the county samples into major and non-major grain-producing areas for subsample estimation, and Figure 3 presents the grouped regression results.

<sup>15</sup> Data source: China Statistical Yearbook.

Figure 3a1–a3 shows that for both total income inequality and disaggregated income inequality, the coefficients for deep processing clusters in major grain-producing areas are negative and significant, while those in non-major areas are not significant. Figure 3b1–b3 shows that, except for farm income inequality, the estimated coefficients for rough processing clusters in major grain-producing areas are negative and significant, with coefficients in non-major areas also being insignificant. A plausible explanation is that major grain-producing areas typically have grain cultivation as their dominant industry. However, given the relatively low comparative returns of grain crops, the income growth of smallholders is constrained, resulting in more prominent income inequality in these regions. Nevertheless, when grain processing clusters emerge in such regions, they can boost smallholders' income by raising grain procurement prices and creating local employment opportunities, thereby alleviating rural internal income disparities. In contrast, non-major grain-producing areas are characterized by low grain yields, small and fragmented cultivation scales, and their grain processing clusters primarily depend on external raw material procurement. Consequently, farmers in non-major grain-producing areas have limited access to the value-added benefits from the processing stage. Furthermore, due to fluctuations in raw material supply, grain processing clusters in non-major grain-producing areas are generally smaller in scale and less stable. Unlike their counterparts in major grain-producing areas, they cannot cover rural low-income groups through large-scale employment, thus exerting a limited effect on reducing inequality in non-major grain-producing areas.

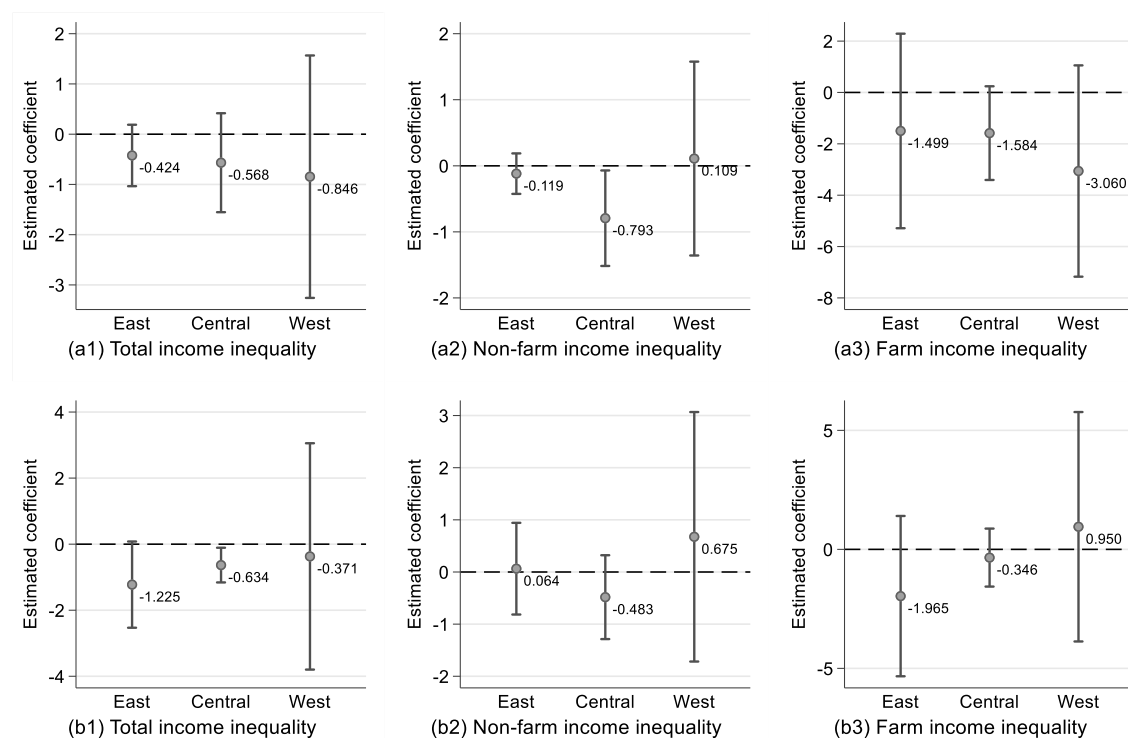


**Figure 3.** Heterogeneity analysis by grain functional areas. (a1–a3) Heterogeneous analysis results of deep processing clusters on income inequality; (b1–b3) heterogeneous analysis results of rough processing clusters on income inequality. From left to right in the panels, the dependent variables are total income inequality, non-farm income inequality, and farm income inequality, respectively. Shown are the regression coefficients and the associated 95% confidence intervals.

## 5.2 Heterogeneity analysis by geographical location

Based on factors such as geographical location, level of economic development, and policy orientation, China's coastal provinces are categorized as the eastern region, inland provinces as the central region, and western provinces as the western region. The eastern region is the most economically developed, with a greater number of industrial bases and economic centers. The central region, as an important base for agriculture and resources, has seen a gradual acceleration of industrialization in recent years. The western region's economic development lags, and its infrastructure is relatively weak. Accordingly, we divide the county samples into eastern, central, and western regions for subsample estimation, and Figure 4 presents the grouped regression results.

The results presented in Figure 4 reveal that deep processing clusters significantly mitigated inequality in both non-farm and farm income only in the central region. By contrast, rough processing clusters significantly reduced total income inequality in both the eastern and central regions, with no significant impact observed in the western region. This empirical evidence indicates that the impact of grain processing clusters on income inequality exhibits distinct regional heterogeneity, with its underlying mechanisms potentially explained as follows. First, both rough and deep processing clusters exerted significant effects in the central region, a major grain-producing area in China. The high level of specialization in grain production ensures a stable raw material supply, endowing the central region with advantages in developing grain processing clusters. These clusters can improve income distribution by absorbing rural surplus labor and raising grain procurement prices. Second, the impacts of deep and rough processing clusters differ in the eastern region. In the eastern



**Figure 4.** Heterogeneity analysis by geographical location. (a1–a3) Heterogeneous analysis results of deep processing clusters on income inequality; (b1–b3) heterogeneous analysis results of rough processing clusters on income inequality. From left to right in the panels, the dependent variables are total income inequality, non-farm income inequality, and farm income inequality, respectively. Shown are the regression coefficients and the associated 95% confidence intervals.

region, where factor prices are higher, deep processing clusters typically require substantial fixed-asset investments. Consequently, the relatively high opportunity costs have constrained the development of deep grain processing clusters to some extent. In contrast, rough processing clusters are characterized by lower technological thresholds and smaller capital demands. Despite the generally higher economic level in eastern rural areas, a large number of low-skilled individuals remain unable to access technical positions. Rough processing clusters, however, can directly absorb this group into employment, thereby increasing the income of the lowest rural strata and significantly reducing income inequality in eastern rural areas. Finally, neither type of processing cluster showed significant effects in the western region. This may be attributed to the region's relatively underdeveloped infrastructure, fragmented agricultural production scales, and inadequate supporting facilities. These factors limit the scale effects and radiation capacity of grain processing clusters, preventing them from significantly alleviating income inequality in western rural areas. The findings of this study suggest that the role of grain processing clusters in improving income distribution is not universally applicable. Their effectiveness depends on regional agricultural resource endowments, industrial supporting capabilities, and economic development levels.

## 6. Conclusions and discussion

### 6.1 Conclusions

Reducing income disparities in rural areas of developing countries contributes to food security and promotes sustainable economic and social development. Based on this, this study provides a new perspective on reducing rural inequality in China from the perspective of grain processing clusters. The study's findings indicate that, firstly, the grain processing industry can be divided into rough processing and deep processing industries. At the county level, both rough and deep processing clusters can reduce total income inequality, but there are differences in their impact on different types of income inequality. Specifically, the deep processing cluster has a greater suppressive effect on non-farm income inequality than the rough processing cluster; the deep processing cluster significantly reduces farm income inequality, while the rough processing cluster has no significant impact. Secondly, based on the analysis of household-level data, we find that grain processing clusters mainly reduce income disparities by increasing the income levels of low-income households. Thirdly, mechanism analysis shows that deep processing clusters provide inclusive employment opportunities for vulnerable groups, enabling low-income households to participate more equally in local non-farm employment; deep processing clusters can also accelerate the deepening of agricultural capital for low-income households and increase their grain cultivation returns. The backward linkages between grain processing and grain production enhance the status of low-income households in the value chain, ultimately narrowing the income gap among farmers. Lastly, regional heterogeneity analysis shows that the negative impact of grain processing clusters on income inequality is more significant in China's major grain-producing areas and the central region of China.

This study constructs an analytical framework for grain processing clusters and inequality, contributing to the literature on development economics and poverty. Our research indicates that industrial development does not have to come at the cost of increasing regional inequality. On the contrary, grain processing clusters, especially deep processing clusters, can effectively reduce income inequality in rural areas by providing more inclusive employment opportunities and improving the comparative benefits of grain cultivation. This finding offers a new interpretation of the complex relationship between industrialization and social equality. Furthermore, in underdeveloped areas where labor migration is limited, grain processing clusters can also provide income-increasing pathways for local farmers, thus promoting income distribution fairness without relying on labor migration.

Due to the focus of the study and data availability, future research can be further expanded in the following areas. Firstly, this paper currently focuses only on the impact of the grain processing segment on farmers, but the grain industry chain can be divided into multiple segments, and future research can further explore

the linkages with other segments of value chain. Secondly, this study mainly examines the impact of grain processing clusters on local farmers' income, but clusters may have spatial spillover effects, and their driving effect on employment and economic development in surrounding areas remains to be explored. Lastly, this paper's analysis focuses on how grain processing clusters reduce income inequality within rural areas, but grain processing clusters will inevitably affect regional economic development through influencing labor agglomeration and industrial value-added. Therefore, future research can explore the impact of grain processing clusters on regional development inequality and provide more comprehensive recommendations for agricultural industry development by synthetically evaluating the role of grain processing clusters.

## *6.2 Policy implications*

Reducing inequality is an important issue related to social stability and sustainable economic development, and grain production is an important strategic industry for the country, so our research results are of great significance for policy making.

First, strengthen policy support systems by expanding rough grain processing, upgrading deep grain processing, and enhancing the value-added potential of grain product conversion and processing. Provide grain processing enterprises with subsidies for equipment procurement, technological R&D, and new product promotion to reduce their initial investment costs. Implement tax relief policies, granting eligible grain processing enterprises reductions or exemptions from corporate income tax, value-added tax, and other relevant taxes. Given that deep processing enterprises significantly drive smallholders' income growth through high-value-added products, greater emphasis should be placed on developing such enterprises. Organize deep processing enterprises to participate in domestic and international agricultural product exhibitions and brand promotion events to boost brand visibility and market influence. For eligible deep-processed grain products, assist enterprises in applying for geographical indication certification; establish a quality traceability system for geographical indication products to ensure their quality and distinctiveness, thereby enhancing their market competitiveness and added value. Second, establish an inclusive development support framework. Remove labor market barriers by offering skill training services to migrant workers engaged in non-farm employment and legislating to guarantee equal pay for equal work. Promote the deep integration of smallholders into the grain industry chain by supporting processing enterprises in providing contracted farmers with shared access to smart agricultural machinery and variety improvement services, ensuring stable returns for small-scale households. Finally, optimize regional industrial layout by focusing on developing grain processing clusters in China's major grain-producing areas and the central region. Support the establishment of grain processing industry associations in major grain-producing counties to prevent vicious competition, provide decision-making references for enterprises, and secure more resources and support for their development. Simultaneously, encourage grain processing enterprises to actively utilize e-commerce platforms for sales. Governments should build public service platforms for agricultural product e-commerce, offering supporting services such as logistics, quality inspection, and financial services to expand product distribution channels. Additionally, for underdeveloped regions, prioritize comprehensive investments in agriculture, including strengthening agricultural technological innovation and its promotion, upgrading transportation networks, and improving irrigation systems to enhance water resource utilization. These integrated measures will effectively improve the essential conditions required for the development of grain processing clusters.

## **Acknowledgements**

The authors acknowledge the support of the National Social Science Fund of China (Grant No. 22&ZD080) and the Ministry of Education's Major Project "Theoretical Innovation in Chinese-style Rural Modernization and Rural Development" (2024JZDZ061). 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.

## References

- Ali, M. and J. Peerlings. 2012. Farm households and nonfarm activities in Ethiopia: does clustering influence entry and exit? *Agricultural Economics* 43(3): 253–266. <https://doi.org/10.1111/j.1574-0862.2012.00580.x>
- Barrett, C. 1997. Food marketing liberalization and trader entry: evidence from Madagascar. *World Development* 25(5): 763–777. [https://doi.org/10.1016/S0305-750X\(96\)00132-5](https://doi.org/10.1016/S0305-750X(96)00132-5)
- Batuo, M.E., G. Kararach and I. Malki. 2022. The Dynamics of Income Inequality in Africa: An Empirical Investigation on the Role of Macroeconomic and Institutional Forces. *World Development* 157: 105939. <https://doi.org/10.1016/j.worlddev.2022.105939>
- Benjamin, D., L. Brandt and J. Giles. 2005. The evolution of income inequality in rural China. *Economic Development and Cultural Change* 53(4): 769–824. <https://doi.org/10.1086/428713>
- Berkes, E., and R. Gaetani. 2023. Income segregation and the rise of the knowledge. *American Economic Journal: Applied Economics* 15(2): 69–102. <https://doi.org/10.1257/app.20210074>
- Bold, T., S. Ghisolfi, F. Nsonzi and J. Svensson. 2022. Market access and quality upgrading: evidence from four field experiments. *American Economic Review* 112(8): 2518–2552. <https://doi.org/10.1257/aer.20210122>
- Bou Dib, J., Z. Alamsyah and M. Qaim. 2018. Land-use change and income inequality in rural Indonesia. *Forest Policy and Economics* 94: 55–66. <https://doi.org/10.1016/j.forpol.2018.06.010>
- Card, D. and A. Krueger. 1996. School resources and student outcomes: an overview of the literature and new evidence from North and South Carolina. *Journal of Economic Perspectives* 10(4): 31–50. <https://doi.org/10.1177/0002716298559001004>
- Cazzuffi, C., M. Pereira-López and I. Soloaga. 2017. Local poverty reduction in Chile and Mexico: the role of food manufacturing growth. *Food Policy* 68: 160–185. <https://doi.org/10.1016/j.foodpol.2017.02.003>
- Cerina, F. and F. Mureddu. 2014. Is agglomeration really good for growth? Global efficiency, interregional equity and uneven growth. *Journal of Urban Economics* 84: 9–22. <https://doi.org/10.1016/j.jue.2014.08.006>
- Cheng, M., Y. Jin, Q. Gai and Q. Shi. 2016. Accounting for income inequality between households in rural China: a regression based approach. *Economics Quarterly* 15(3): 1253–1274. (in Chinese)
- Clapp, J. 2017. Food Self-sufficiency: Making Sense of it, and When it Makes Sense. *Food Policy* 66: 88–96. <https://doi.org/10.1016/j.foodpol.2016.12.001>
- Dorward, A., J. Kydd, J. Morrison and I. Urey. 2004. A policy agenda for pro poor agricultural growth. *World Development* 32(1): 73–89. <https://doi.org/10.1016/j.worlddev.2003.06.012>
- Duranton, G. and D. Puga. 2004. Micro-foundations of urban agglomeration economies. *Handbook of Regional and Urban Economics* 4: 2063–2117. [https://doi.org/10.1016/S1574-0080\(04\)80005-1](https://doi.org/10.1016/S1574-0080(04)80005-1)
- Duranton, G. and D. Puga. 2024. Urban growth and its aggregate implications. *Econometrica* 91(6): 2219–2259. <https://doi.org/10.3982/ECTA17936>
- Galvez-Nogales, E. 2010. *Agro-based clusters in developing countries: staying competitive in a globalized economy*. FAO Agricultural Management, Marketing and Finance Occasional Paper 25. FAO, Rome.
- Galvez-Nogales, E. and M. Webber. 2017. *Territorial tools for agro-industry development: a sourcebook. Agrobased clusters*. FAO, Rome.
- Gao, M. and J. Wei. 2021. Accelerating the construction of national grain security industrial belt: development orientation and strategic conception. *China's Rural Economy* 11: 16–34. (in Chinese)
- Glomm, G. and B. Ravikumar. 1992. Public versus private investment in human capital: endogenous growth and income inequality. *Journal of Political Economy* 100(4): 818–834. <https://doi.org/10.1086/261841>
- Guo, D., K. Jiang, C. Xu and X. Yang. 2022. Industrial Clustering, Income and Inequality in Rural China. *World Development* 154: 105878. <https://doi.org/10.1016/j.worlddev.2022.105878>
- Hadachek, J., M. Ma and R. J. Sexton. 2024. Market structure and resilience of food supply chains under extreme events. *American Journal of Agricultural Economics* 106(1): 21–44. <https://doi.org/10.1111/ajae.12393>

- Howell, A. 2017. Impacts of migration and remittances on ethnic income inequality in rural China. *World Development* 94: 200–211. <https://doi.org/10.1016/j.worlddev.2017.01.005>
- Jr Tabe Ojong, M.P. and G.G. Dureti. 2023. Are agro-clusters pro-poor? Evidence from Ethiopia. *Journal of Agricultural Economics* 74(1): 100–115. <https://doi.org/10.1111/1477-9552.12497>
- Koenker, R. and G. Bassett. 1978. Regression quantiles. *Econometrica* 46(1): 33–50. <https://doi.org/10.2307/1913643>
- Liu, D., X. Yu, Y. Wang, H. Zhang and G. Ren. 2014. The Impact of perception of discrimination and sense of belonging on the loneliness of the children of Chinese migrant workers: a structural equation modeling analysis. *International Journal of Mental Health Systems* 8(1): 1–6. <https://doi.org/10.1186/1752-4458-8-52>
- Liu, J., Z. Xu, Q. Zheng and L. Hua. 2019. Is the Feminization of labor harmful to agricultural production? The decision-making and production control perspective. *Journal of Integrative Agriculture* 18(6): 220–229. [https://doi.org/10.1016/S2095-3119\(19\)62649-3](https://doi.org/10.1016/S2095-3119(19)62649-3)
- Luo, C., S. Li and T. Sicular. 2020. The long-term evolution of national income inequality and rural poverty in China. *China Economic Review* 62: 101465. <https://doi.org/10.1016/j.chieco.2020.101465>
- Maertens, M., B. Minten and J. Swinnen. 2012. Modern food supply chains and development: evidence from horticulture export sectors in sub-Saharan Africa. *Development Policy Review* 30(4): 473–497. <https://doi.org/10.1111/j.1467-7679.2012.00585.x>
- Magnan, N., T.J. Lybbert, A.F. McCalla and J. A. Lampietti. 2011. Modeling the limitations and implicit costs of cereal self-sufficiency: the case of Morocco. *Food Security* 3(1): 49–60. <https://doi.org/10.1007/s12571-010-0103-2>
- Mano, Y., T. Njagi and K. Otsuka. 2022. An inquiry into the process of upgrading rice milling services: the case of the Mwea irrigation scheme in Kenya. *Food Policy* 106: 102195. <https://doi.org/10.1016/j.foodpol.2021.102195>
- Marshall, A. 1920. *Principles of economics*. Macmillan, London.
- Minten, B., B. Mohammed and S. Tamru. 2020. Emerging medium scale tenant farming, gig economies, and the COVID-19 disruption: the case of commercial vegetable clusters in Ethiopia. *The European Journal of Development Research* 32(5): 1402–1429. <https://doi.org/10.1057/s41287-020-00315-7>
- Porter, M. 1990. *The competitive advantage of nations*. Macmillan, London.
- Poulton, C., A. Dorward and J. Kydd. 2010. The Future of small farms: new directions for services, institutions, and intermediation. *World Development* 38(10): 1413–1428. <https://doi.org/10.1016/j.worlddev.2009.06.009>
- Qiu, F., H. Yang and T. Zhuang. 2025. Mitigation or Intensification: The impact of modern agricultural industrial park on income inequality of rural low-income groups. *Chinese Rural Economy* 3: 101–121. (in Chinese)
- Reardon, T., L.S.O. Liverpool-Tasie and B. Minten. 2021. Quiet revolution by SMEs in the midstream of value chains in developing regions: wholesale markets, wholesalers, logistics, and processing. *Food Security* 13: 1577–1594. <https://doi.org/10.1007/s12571-021-01224-1>
- Ruan, J. and X. Zhang. 2009. Finance and cluster-based industrial development in China. *Economic Development and Culture Change* 58(1): 143–164. <https://doi.org/10.1086/605208>
- Ruan, J., X. Zhang and L. Wei. 2011. Imperfect capital market and choice of production organization form—evidence from rural industrial clusters in China. *Management World* 8: 79–91. (in Chinese)
- Takahashi, K. and C.B. Barrett. 2014. The System of rice intensification and its impacts on household income and child schooling: evidence from rural Indonesia. *American Journal of Agricultural Economics* 96(1): 269–289. <https://doi.org/10.1093/ajae/aat086>
- Wan, G.H. and Z.Y. Zhou. 2005. Income inequality in rural China: regression-based decomposition using household data. *Review of Development Economics* 9(1): 107–120. <https://doi.org/10.1111/j.1467-9361.2005.00266.x>
- Wang, S.X. and F.U. Yu Benjamin. 2019. Labor mobility barriers and rural-urban migration in transitional China. *China Economic Review* 53: 211–224. <https://doi.org/10.1016/j.chieco.2018.09.006>

- Wardhana, D., R. Ihle and W. Heijman. 2020. Farmer cooperation in agro-clusters. evidence from Indonesia. *Agribusiness* 36(4): 725–750. <https://doi.org/10.1002/agr.21637>
- Weitzman, M. L. and C. Xu. 1994. Chinese Township-Village enterprises as vaguely defined cooperatives. *Journal of Comparative Economics* 18(2): 121–145.
- World Bank. 2008. *World development report 2008: agriculture for development*. World Bank, Washington, DC.
- Xing, C. (2014). Migration, self-selection and income distributions: evidence from rural and urban China. *Economics of Transition* 3: 539–576. <https://doi.org/10.1111/ecot.12041>
- Yi, Q., M. Chen, Y. Sheng and J. Huang. 2019. Mechanization services, farm productivity and institutional innovation in China. *China Agricultural Economic Review* 11(3): 536–554. <https://doi.org/10.1108/CAER-12-2018-0244>
- Zhang, X. 2023. Cluster-based agricultural development: a comparison between China and Africa. *International Food and Agribusiness Management Review* 26(1): 1–9. <https://doi.org/10.22434/IFAMR2022.0041>

## Appendix

### Additional calculation formulas

The Gini coefficient is defined as follows:

$$\text{Gini} = \frac{1}{n} \left( n + 1 - 2 \left( \frac{\sum_{i=1}^n (n+1-i)y_i}{\sum_{i=1}^n y_i} \right) \right) \quad (\text{A1})$$

where  $n$  is the total number of rural households in the county, and  $y_i$  is the total income, non-farm income or farm income of the rural household.

The Theil index is defined as follows:

$$\text{Theil} = \sum_{i=1}^n \frac{y_i}{Y} \log \left( \frac{\frac{y_i}{Y}}{\frac{1}{n}} \right) \quad (\text{A2})$$

where  $n$  is the total number of households in the county.  $y_i$  is the total income, non-farm income or farm income of the household.  $Y$  presents the total income, non-farm income or farm income of the population, with  $Y = \sum_{i=1}^n y_i$ .

Quantile regression is defined as follows:

$$y_i = \beta_0(\theta) + \beta_1(\theta)x_i + \varepsilon_i \quad (\text{A3})$$

Where  $\theta \in (0,1)$ ,  $y_i$  is the explained variable,  $x_i$  is the explanatory variable, and  $\beta_1(\theta)$  is the regression coefficient of the explained variable at the  $\theta$  quantile. Further define  $y_i = \beta_0(\theta) + \beta_1(\theta)x_i + \varepsilon_i$  as the sample conditional quantile function at the  $\theta$  quantile.