Digital Economic Development and Revenue Concentration in Regional Industries

Shujuan Wu, Ruoying Huang*
School of Economics and Management, Wuyi University, Jiangmen, Guangdong, China
**Corresponding Author*

Abstract: Using a province–industry–year panel covering 28 provincial-level regions and 34 industries in China from 2015 to 2022, this study examines the relationship between digital economic development and regional industry revenue concentration. The Herfindahl–Hirschman Index (HHI) is calculated from listed firms' operating revenue shares within each cell. We estimate models with multiple sets of fixed effects and conduct inference using standard errors clustered at the provincial level and the wild cluster bootstrap. The results show a statistically significant negative association between the digital economy index and industry revenue concentration. A 0.01-point increase in the index is associated with an average reduction of approximately 26.51 HHI points. The coefficient remains negative under alternative variable measures and sample specifications. An algebraic decomposition of the HHI shows that the revenue-share dispersion component accounts for most of the baseline coefficient, while the firm-count component also receives statistical support. Lagged specifications, a lead test, and random permutation tests provide complementary evidence. The findings contribute to understanding how digital economic development relates to the distribution of firm revenues and changes in industry competitive structure.

Keywords: Digital Economy; Regional Industry Revenue Concentration; Herfindahl–Hirschman Index; Listed Firms; Panel Data

1. Introduction

At present, the integration of digital technology and real economy is accelerating, and digital economy has become an important force to change the mode of production, market boundary and competitive structure of enterprises. Big data, cloud computing, artificial intelligence and digital platforms can reduce information search and transaction costs, improve resource allocation efficiency, and change the way companies obtain customers, financing and technical services. In the 2026 government work report, it is proposed to implement the inclusive " cloud-based digital intelligence " service, continue to increase support for the digital intelligence transformation of small and medium-sized enterprises, and cultivate new models of intelligent native new formats, indicating that the digital transformation has risen from the efficiency tool at the enterprise level to an important issue of industrial upgrading and market governance.

The digital economy may have an opposite effect on industry concentration. Digital infrastructure, online commerce, and digital finance can also help smaller firms reach customers, secure financing, and enter new markets at lower cost, allowing them to expand their operations. The direction of change in concentration depends on the relative strength of these two forces. Assessing the outcomes of digital economic development therefore requires examining both efficiency gains and changes in the distribution of revenue shares across firms.

Prior research has primarily examined industries at the national level, individual firms, or buyer–supplier relationships. National measures of industry concentration aggregate across regions with different competitive conditions, while firm-level findings offer limited insight into how revenue shares are distributed among firms within a regional industry. We examine province–industry cells to assess how regional differences in digital economic development relate to the distribution of revenue among listed firms in the same industry across provinces. We also compare estimates across samples defined by different minimum firm counts. We decompose the overall Herfindahl–Hirschman

Copyright @ STEMM Institute Press

Index (HHI) algebraically into a firm-count component and a revenue-share dispersion component. We then use the log number of firms and the firm-count-normalized HHI as complementary outcomes to assess the statistical evidence for each component.

We match listed firms' financial data with a provincial digital economy index to construct a province–industry–year panel covering 2015–2022. The baseline model includes province–industry and industry–year fixed effects, with standard errors clustered at the provincial level. Given the limited number of clusters (28), we use the wild cluster bootstrap to improve finite-sample inference. We also use alternative measures, examine lagged associations, and conduct tests using leads of the digital economy index and province-level random permutations.

The economic interpretation of concentration depends on how industry and geographic boundaries are defined. We examine how revenue is distributed among listed firms within each province and industry. Firms are assigned to their province of registration, although their sales may span multiple regions. We therefore use the dispersion of revenue shares to characterize competitive structure within the sample. Assessing market power would require additional evidence on product market boundaries, prices, and costs.

Observations across industries within a province capture variation in industry revenue distributions, whereas the digital economy index varies at the province–year level. Our inference accounts for potential shocks shared by industries within a province. The permutation procedure also preserves the assignment of a common index value to all industry observations within each province–year. Therefore, the sensitivity of the benchmark results to sample coverage, variable caliber and statistical inference methods can be investigated simultaneously.

2. Literature Review on the Digital Economy and Industry Concentration

2.1 Determinants and Economic Implications of Industry Concentration

Industry concentration is a measure of market structure. Curry and George argue that concentration is jointly determined by demand conditions, costs, and firm behavior [1]. Sutton distinguishes between exogenous and endogenous sunk costs, showing how investments in advertising, research and development (R&D), and product quality can sustain high concentration even as markets expand [2]. When explaining the change of concentration, it is necessary to combine entry conditions and enterprise share distribution to distinguish different sources such as weakening competition and expansion of high-efficiency enterprises.

The study of market power and concentration reveals the economic meaning of income concentration to a few enterprises from different aspects. De Loecker et al. document a long-run rise in firm markups in the United States [3]. Autor et al. use the superstar firm mechanism to explain how rising market shares among highly productive firms can contribute to a decline in the labor share of income [4]. Kwon et al. link long-run changes in concentration to investment in R&D and information technology [5]. Productivity gains can shift revenue shares across firms, while barriers to entry can help dominant firms retain large shares. Distinguishing these explanations requires additional evidence on prices, costs, and entry conditions.

The interpretation of concentration also depends on the geographic scope of the market. In their study of technology diffusion in services, Hsieh and Rossi-Hansberg show how highly productive firms can use technology to expand across regions [6]. The same firm's expansion can raise national industry concentration while reducing local concentration by giving customers more suppliers to choose from. The geographic reach of digital channels can likewise change which markets firms are able to serve. We examine the distribution of revenue among listed firms within province–industry cells, using consistent geographic and industry classifications.

2.2 Cost Reductions and Broader Market Participation

Goldfarb and Tucker characterize the main economic effects of digital technologies through reductions in search, replication, transportation, tracking, and verification costs [7]. Easier access to product information, customer demand signals, and information about trading partners' reputations allows firms to reduce their investment in offline channels and search for

 Copyright @ STEMM Institute Press

customers and suppliers over a wider area. Consumers can also compare prices and quality at lower cost. For small and medium-sized enterprises with limited distribution networks and brand recognition, these changes may weaken the advantages incumbents derive from geographic segmentation and information asymmetries, creating opportunities to expand operations and improve efficiency.

Whether lower costs produce a more dispersed distribution of revenue shares depends on how the gains are distributed across firms. If digital tools improve the operations of firms previously constrained by limited information, distribution channels, or technological capabilities, efficiency gains may increase their sales and narrow the gap between their revenue shares and those of leading firms. Hsieh and Rossi-Hansberg's analysis of technological change and firms' geographic expansion demonstrates how technology can reshape the conditions for participation in local markets [6]. Combining this perspective with the cost reductions associated with digital technologies suggests a potential channel through which digitalization can affect revenue shares by improving market access.

Access to financing can also be examined through changes in information search and verification costs. Extending Goldfarb and Tucker's framework to financial services suggests that digital payments, online transactions, and records of contract performance can provide additional information for credit assessment [7]. When this information reduces screening and verification costs, it may ease financing constraints for smaller firms and potential entrants, giving them greater scope to sustain operations, invest in innovation, and increase sales. This reasoning offers a theoretical basis for linking digital finance to market participation, while its implications for revenue concentration remain an empirical question.

Changes in information costs can also affect firm boundaries and trading networks. Lower search and verification costs make it easier for firms to identify potential partners, compare transaction terms, and establish trust, expanding their choice of customers and suppliers. Digital technologies may consequently reduce reliance on a small number of trading partners and established distribution channels, affecting firms' revenue shares through the reallocation of trading opportunities. We view these changes as potential sources of broader market participation and examine whether they are reflected in the revenue structure of regional industries.

2.3 Economies of Scale and Concentration in Platform Markets

Cross-side network effects can strengthen the scale advantages of platforms. Rochet and Tirole and Armstrong explain that the size of the user base on one side of a platform affects the benefits of participation on the other [8,9]. Platform pricing and users' decisions to participate on multiple platforms, known as multihoming, jointly shape competitive outcomes. Hagiu and Wright link platform competition to how platforms organize interactions among participants [10]. When network effects are strong and switching costs are high, a larger user base can make participation more valuable, attracting further users and concentrating activity on a small number of platforms.

Investment in information technology can also widen cost advantages across firms. Bessen finds that the use of proprietary information technology is associated with rising industry concentration in the United States [11]. Lashkari et al. show that the returns to information technology tend to favor larger firms [12]. In industries where proprietary systems entail substantial fixed costs, leading firms can spread these costs over a larger volume of sales and further increase their market shares.

The economics of data adds another dimension to digital competition. Jones and Tonetti emphasize that data are nonrival and that broad data sharing can generate social benefits [13]. Firms may nonetheless withhold data to reduce the risk of being displaced. Bergemann and Bonatti show that platforms can use data to improve matching and can also exploit informational advantages to strengthen their bargaining position with merchants [14]. When data are difficult to transfer, interfaces are incompatible, or platforms favor their own products and services, control over large datasets, advanced algorithms, and user traffic can create barriers to entry. Open access to public data, data portability, and fair access to platforms can help small and medium-sized enterprises use these resources to expand their market opportunities.

Using evidence on online network centrality, Emery shows that lower search costs and

Copyright @ STEMM Institute Press

network effects can jointly increase the market shares of firms that occupy central positions in online networks [15]. This finding differs from evidence in the supply chain literature that digital technologies expand firms' choice of trading partners. As consumers search more efficiently, demand may shift toward suppliers that were previously difficult to find or toward firms that already hold advantageous positions in online markets. Whether digitalization produces a more even distribution of market shares depends on how broadly the reductions in transaction costs are shared, the strength of platform network effects, and how the study defines the market and measures concentration.

2.4 Measuring the Digital Economy and Examining Its Relationship with Competition

Existing measures of digitalization differ in their level of analysis and focus. Bessen examines proprietary information technology systems [11], Lashkari et al. study firms' information technology inputs and returns to scale [12], and Emery uses online network centrality to capture firms' positions in the digital trading environment [15]. These measures reflect different aspects of technological investment, production conditions, and network advantages. They consequently capture distinct dimensions of how digitalization relates to competitive structure, making the interpretation of empirical estimates dependent on what each measure represents.

We use a provincial composite index to characterize regional conditions for digital economic development, incorporating digital infrastructure, information services, communications applications, and digital finance into a common regional measure. The index primarily captures the local environment in which firms undertake digital activities. Because indicator selection and weighting can affect regional rankings and regression estimates, we use an entropy-weighted index as the baseline measure and a principal component index to assess sensitivity to the weighting method.

Digitization and competitive structure may also influence each other. Technological progress will change the cost and market access conditions of enterprises; the existing scale, market position and resources may affect the ability of enterprises to adopt digital technology and benefit from it. Regional industrial base, human capital accumulation and governance environment may also affect the development of digital economy and the distribution of enterprise income. Therefore, empirical analysis needs to consider these common determinants and assess the scope of application of its conclusions based on the limitations of variable measurement methods, sample coverage, and statistical inference.

2.5 Synthesis and Research Focus

The literature points out that the development of digital economy can affect concentration through two forces. Lower search, transaction and verification costs can broaden the opportunities for companies to participate in the market, while proprietary technology, network effects, and control over data can strengthen the scale advantage of leading companies. These two forces can play a role in the same market at the same time, and their relative strength depends on the characteristics of enterprises, technical conditions and geographical analysis. Therefore, the relationship between digital economy development and concentration needs to be examined within a clearly defined sample and measurement framework. The existing research explains the change of competition structure through the enterprise 's technology investment, productivity, business scope, platform interaction and online network location.

We further extend this analysis by linking provincial digital economic conditions to the distribution of listed companies ' revenues within a province-industry unit. Customer or supplier concentration measures the dependence of a single enterprise on a specific trading partner, and our HHI captures the distribution of the income share of listed companies in the region. The HHI is decomposed into the enterprise quantity component and the income share dispersion component, so that we can further identify where the observed concentration changes are reflected, thus providing additional evidence for the relationship between the development of digital economy and the income structure of regional industries.

3. Theoretical Analysis and Research Hypotheses

3.1 The Digital Economy and Industry Concentration

Digital economy affects the income structure of regional industries through two kinds of forces.

 Copyright @ STEMM Institute Press

The inclusive competition effect enables more enterprises to obtain customers, capital and technical services by reducing the cost of market participation. The scale polarization effect is due to the higher fixed cost of technology, network externalities and data accumulation, which makes it easier for leading enterprises to expand their share. The two types of forces work together on the number of enterprises and the share of income. Based on this, this paper deduces the net change direction of concentration.

Digital infrastructure and online transactions have expanded the reachable market for SMEs. Online channels, digital logistics and online evaluation reduce information search, supply and demand matching and cross-regional operating costs, and also make it easier for consumers to compare products and find alternative supply. As potential participants have access to customers who are originally geographically or channel-constrained, the share of revenue that existing enterprises rely on channel advantages may be diverted.

Digital finance and modular digital tools can ease the financing and technological constraints of SMEs. Transaction and performance data improve credit identification, and cloud services and general software reduce the fixed cost of adopting digital technologies for enterprises. When financing and digital services are more available, non-head enterprises are better positioned to maintain operations, improve efficiency and expand sales, and the income share gap between enterprises may be narrowed.

More adequate information flow will also change the competitive conditions of stock enterprises. Enterprises can identify demand and adjust production in time, and it is easier for consumers to compare quality and price. Therefore, the improvement of market accessibility may not only expand the scope of participation, but also disperse the income share by changing the relative competitiveness of stock enterprises. If the cost of using digital tools decreases more to benefit enterprises with limited original channels, the improvement of their revenue share is expected to narrow the gap between enterprises in regional industries.

The scale polarization effect provides the opposite theoretical expectation. Head enterprises can share the fixed costs of proprietary systems, and use network effects and data accumulation to improve matching and consolidate user relationships. When the switching cost is high, these advantages may further concentrate the share to the leading enterprises. Considering the popularity of China 's digital infrastructure and the expansion of inclusive finance during the sample period, this paper expects that the decentralization effect brought by the decrease of participation cost and the improvement of operating conditions of SMEs is generally stronger.

H1. Digital economic development significantly reduces industry concentration and promotes market competition.

3.2 Firm Counts and the Distribution of Revenue Shares

HHI is also affected by the difference in the number and share of enterprises. When the share of enterprises is equal, HHI is equal to 10000 divided by the number of enterprises. The part where the actual HHI exceeds this minimum value reflects the degree to which the share deviates from the equal distribution.

When digital infrastructure and digital finance lower the threshold of entry and operation, the number of enterprises that can be covered may increase, and the number of enterprises will decrease accordingly. When digital tools improve the operating efficiency and market accessibility of non-head enterprises, the income share between stock enterprises may be more balanced, and the market share dispersion will decrease accordingly. Two changes can occur at the same time, because the change of the scope of enterprise participation will also change the income share of existing enterprises.

If digital economic development operates primarily by lowering operating and financing barriers, we expect the firm-count component to decline and the log number of firms to rise. If digital tools primarily improve the relative efficiency of firms outside the leading group, we expect both the revenue-share dispersion component and the firm-count-normalized HHI to decline. The observed number of listed firms also depends on sample coverage and whether firms report positive revenue, so these results should be interpreted jointly.

H2a. Digital economic development reduces industry concentration by lowering the firm-count component of the HHI.

H2b. Digital economic development reduces industry concentration by making revenue shares more evenly distributed across firms.

Copyright @ STEMM Institute Press

4. Research Design

4.1 Model Specification

The unit of observation is the province–industry–year cell. We estimate the following fixed-effects model.

$$HHI_{pit}=\beta_1 Digital_{pt}+\gamma' Controls_{pit}+\mu_{pi}+\lambda_{it}+\varepsilon_{pit} \quad (1)$$

The subscripts p, i and t denote province, industry, and year, respectively. HHI_{pit} measures revenue concentration within a province–industry–year cell, $Digital_{pt}$ is the provincial digital economy index, and $Controls_{pit}$ is the vector of control variables listed in Table 1. Province–industry fixed effects account for time-invariant differences in industrial bases, locational conditions, and historical market structure. Industry–year fixed effects absorb annual demand, policy, and technology shocks common to firms in the same industry. The coefficient of interest, is estimated from differences in changes over time across provinces within the same industry.

4.2 Sample Selection and Data Sources

We obtain listed firms' financial data from the China Stock Market & Accounting Research (CSMAR) Financial Statements Database. The source data include financial statements from firms listed on the Shanghai, Shenzhen, and Beijing stock exchanges. Comparable data on the digital economy indicators are available through 2022, and our sample covers 2015–2022. We restrict the financial sample to firms whose securities codes identify them as Shanghai or Shenzhen A-share listings, exclude financial firms and observations with missing or nonpositive operating revenue, and aggregate the remaining records by province of registration, industry, and year.

Following Zhao et al. [16], we construct a provincial digital economy index using entropy weighting and an alternative index using principal component analysis (PCA). Both indices draw on five standardized indicators: the number of broadband internet subscribers, the share of employment in the computer and software industries, telecommunications business volume per capita, the number of mobile phone subscribers, and the Digital Financial Inclusion Index. The Digital Financial Inclusion Index comes from the Institute of Digital Finance at Peking University, and the remaining indicators come from the China City Statistical Yearbook. Because composite indices can be sensitive to indicator selection and weighting, we substitute the PCA-based index for the entropy-weighted index in the robustness analysis. Missing annual observations are left unfilled.

To calculate the HHI and the four-firm concentration ratio (CR4) from a sufficient number of firms, we require each province–industry–year cell in the main sample to contain at least five firms with valid operating revenue. The raw financial data contain 41,082 firm-year observations. After applying the year, securities code, industry, and revenue filters, we retain 29,879 firm-year observations, which form 7,420 province–industry–year cells. Of these, 1,610 cells meet the five-firm requirement. After iteratively dropping singleton observations from the fixed-effects model, the baseline regression sample comprises 1,542 observations, 28 provincial clusters, 34 industries, and 237 distinct province–industry pairs.

The main sample retains otherwise eligible firms with special treatment (ST) or delisting flags so that revenue shares are calculated using all valid revenue observations within each cell. In a robustness check, we exclude these firms and recalculate concentration. We also vary the minimum number of firms per cell to three and ten to assess whether the results are sensitive to firm coverage.

With very few firms in a cell, CR4 can mechanically reach or approach 100%, and the HHI can be sensitive to revenue fluctuations at a single firm. A higher threshold, however, excludes provinces and industries with limited coverage of listed firms, changing the composition of the sample. The main analysis therefore uses an unbalanced panel with at least five firms per cell and allows the number of firms to vary over time.

4.3 Variable Definitions

4.3.1 Dependent variable

We measure industry concentration using the Herfindahl–Hirschman Index (HHI). For each listed firm, we calculate its revenue share by dividing its operating revenue by the total operating revenue of all listed firms in the same province–industry–year cell. The HHI is calculated as follows.

$$HHI_{pit}=10000\sum_{j=1}^{N_{pit}}\left(\frac{Revenue_{jpit}}{\sum_{k=1}^{N_{pit}} Revenue_{kpit}}\right)^2 \quad (2)$$

 Copyright @ STEMM Institute Press

where N_{pit} is the number of firms in the cell and $Revenue_{jpit}$ is firm j's operating revenue.

Higher HHI values indicate greater operating revenue concentration. The measure covers the operating revenue of listed firms in the sample, grouped by province of registration. In the robustness analysis, we use CR4 and the firm-count-normalized HHI as alternative measures. CR4 captures the combined revenue share of the four largest firms by operating revenue. The firm-count-normalized HHI allows us to examine whether digital economic development remains associated with the distribution of revenue shares after accounting for the effect of firm counts on the minimum attainable HHI.

4.3.2 Main explanatory variable

The baseline regression uses the entropy-weighted provincial digital economy index. The robustness analysis uses a PCA-based index constructed from the same underlying indicators.

4.3.3 Control variables

We construct the control variables by taking equally weighted averages of firm-level financial measures within each province–industry–year cell. Firm size is the mean of the natural logarithm of total assets. Leverage is the mean ratio of total liabilities to total assets. Return on assets is the mean ratio of net income to end-of-period total assets. The log capital–labor ratio is the mean of the natural logarithm of net fixed assets per employee.

These variables capture firm size, capital structure, profitability, and capital intensity within each cell. Equally weighted averages reduce the extent to which the controls vary mechanically with the number of firms or total operating revenue in the cell. Firm counts, innovation expenditure, and access to financing may themselves respond to digital economic development and help explain the patterns observed in the data. We therefore limit the baseline specification to the four controls defined above.

Table 1. Variable Definitions

variable	symbol	definition	Unit or scope
Industry concentration	HHI	Sum of the squared shares of operating revenue of enterprises within the regional industry, multiplied by 10,000	0—10000
Concentration of the top four enterprises	CR4	Share of operating revenue of the four companies with the highest revenue	0—100
digital economy	Digital	Provincial Composite Index for the Digital Economy Based on the Entropy Method	index
enterprise scale	Size	Unit mean of the natural logarithm of total corporate assets	logarithm
asset-liability ratio	Lev	Mean unit value of the ratio of total corporate liabilities to total assets	ratio
return on assets	ROA	Mean unit value of the ratio of corporate net profit to total assets at the end of the period	ratio
Capital-labor ratio	LnK/L	The unit mean of the logarithm of the ratio of net fixed assets to the number of employees	logarithm

4.4 Identification Limits and Statistical Inference

All industries within a given province and year share the same digital economy index value, and regression errors may be correlated within provinces. Following Cameron and Miller, we cluster standard errors at the provincial level [17]. With only 28 provincial clusters in the baseline regression, conventional cluster-robust inference may be subject to finite-sample distortions. Cameron et al. and MacKinnon and Webb discuss the use of the wild cluster bootstrap when the number of clusters is small [18,19], and Roodman et al. provide an implementation [20]. We use 9,999 bootstrap replications with Rademacher weights and report both conventional province-clustered and wild cluster bootstrap p-values for the coefficient of interest.

The fixed effects account for time-invariant characteristics of province–industry pairs and shocks common to an industry in a given year. However, time-varying factors such as local governance, human capital accumulation, and industrial policy may affect both digital economic development and revenue concentration. Competitive conditions may also shape the pace of regional digitalization. We therefore interpret the coefficient of interest as an association conditional on the controls and fixed effects. Tests using leads of the digital

Copyright @ STEMM Institute Press

economy index provide supplementary evidence on temporal ordering, while province-level permutation tests assess whether an association of the observed magnitude could arise under random assignment of index values across provinces.

5. Empirical Results and Analysis

5.1 Descriptive Statistics and Correlations

Table 2. Descriptive Statistics

variable	mean value	standard deviation	minimum	maximum	Observation count
HHI	3059.441	1691.011	296.660	9330.232	1542
CR4	82.250	16.215	23.037	99.990	1542
Digital	0.390	0.231	0.032	0.961	1542
Size	22.223	0.703	20.570	25.683	1542
Lev	0.428	0.121	0.192	0.951	1542
ROA	0.021	0.026	-0.488	0.162	1542
LnK/L	12.517	0.744	9.718	16.205	1542
number of the enterprise	12.674	14.014	5	195	1542

As shown in Table 2, the HHI has a mean of 3,059.441 and a standard deviation of 1,691.011, indicating substantial variation in revenue concentration across regional industries. The digital economy index has a mean of 0.390 and a standard deviation of 0.231. In the main sample, the number of listed firms per province–industry–year cell averages 12.67 and ranges from 5 to 195, indicating that firm coverage varies widely across cells.

Table 3. Correlation Matrix

variable	(1)	(2)	(3)	(4)	(5)	(6)
(1) HHI	1.000					
(2) Digital	-0.084***	1.000				
(3) Size	0.201***	0.160***	1.000			
(4) Lev	0.247***	0.015	0.628***	1.000		
(5) ROA	-0.093***	-0.030	-0.170***	-0.504***	1.000	
(6) LnK/L	0.075***	-0.193***	0.297***	0.044*	0.092***	1.000

Note. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. The sample contains 1,542 observations.

Table 4. Variance Inflation Factors

variable	VIF	tolerance
Digital	1.013	0.987
Size	1.136	0.880
Lev	1.387	0.721
ROA	1.333	0.750
LnK/L	1.070	0.934
mean	1.188	—

As shown in Table 3, the bivariate correlation between the digital economy index and the HHI is −0.084 and is statistically significant at the 1% level, providing preliminary evidence of a negative association. The largest absolute pairwise correlation among the explanatory variables is 0.628, below levels that typically warrant concern. As,reported in Table 4, all variance inflation factors (VIFs) are below 1.4, suggesting that severe multicollinearity is unlikely.

5.2 Baseline Regression Results

Column (1) of Table 5 includes only province–industry and year fixed effects. The estimated coefficient on the digital economy index is −2,002.817. Adding controls for firm characteristics in column (2) yields an estimate of −1,901.623. Both estimates are statistically significant at the 10% level based on conventional province-clustered tests. Column (3) replaces year fixed effects with industry–year fixed effects to absorb shocks common to each industry in a given year. The coefficient on the digital economy index is −2,651.273, with a conventional province-clustered p-value of 0.025.

Table 5. Baseline Regressions of Industry Concentration on the Digital Economy Index

variable	(1)	(2)	(3)
Digital	-2002.817* (1075.324)	-1901.623* (1047.602)	-2651.273** (1119.624)
Size		-179.170 (194.788)	-355.594* (197.064)

 Copyright @ STEMM Institute Press

Lev		880.452 (755.484)	385.085 (611.948)
ROA		1475.318 (1085.683)	1169.590 (1188.540)
LnK/L		117.834 (181.971)	4.115 (223.294)
controlled variable	NO	YES	YES
Province–Industry Fixed Effects	YES	YES	YES
Year fixed effects	YES	YES	NO
Industry–Year Fixed Effects	NO	NO	YES
Observation count	1542	1542	1542
Provincial cluster count	28	28	28
R^2	0.872	0.873	0.905
Group R^2	0.007	0.013	0.025
Digital's Wild p-value	—	—	0.033

Note. Robust standard errors clustered at the provincial level are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively, based on conventional cluster-robust tests. The wild cluster bootstrap for column (3) uses 9,999 replications.

Based on column (3), Wild Cluster Bootstrap p value is 0.033. According to the estimated coefficient, for every 0.01 increase in the digital economy index, the average HHI of regional industry income is about 26.51 points lower, supporting the negative expectation of H1. This magnitude corresponds to the intra-sample average correlation under a given control variable and fixed effect. The intra-group R2 of column (3) was 0.025, and the model explained about 2.5 % of the remaining changes in the group. The higher overall R2 also reflected the absorption of fixed effects on sample differences.

5.3 Robustness Checks

We assess the robustness of the baseline results by using alternative measures of the dependent variable and the digital economy, adjusting for extreme values, changing which firms are included, and raising the minimum number of firms per cell. Table 6 reports both conventional province-clustered and wild cluster bootstrap p-values so that the results can be assessed under both inference methods.

Table 6. Robustness Checks Using Alternative Measures and Sample Specifications

test	Core coefficient	Cluster Standard Error	Conventional p-value	Wildpprice	N
standardization HHI	-0.244	0.127	0.065	0.073	1542
Replace with CR4	-16.487	9.288	0.087	0.120	1542
Principal Component Digital Index	-4370.488	1859.276	0.026	0.064	1542
Control variables exhibit year-on-year tail contraction.	-2517.927	1119.881	0.033	0.040	1542
Recalculation after excluding ST companies	-2505.782	1307.044	0.067	0.090	1491
At least 10 enterprises	-4489.782	1537.457	0.015	0.090	523

Note. All models include four control variables, province–industry fixed effects, and industry–year fixed effects. Standard errors are clustered at the provincial level. Because the dependent variables are measured on different scales, coefficient magnitudes are not directly comparable across outcomes.

The estimated coefficients on the digital economy index are negative across all specifications in Table 6, although their statistical significance varies. Winsorizing the control variables separately by year yields a wild cluster bootstrap p-value of 0.040, which is significant at the 5% level. The estimates using the firm-count-normalized HHI or the PCA-based digital economy index, as well as those excluding ST firms or requiring at least ten firms per cell, are significant only at the 10% level. The wild cluster bootstrap p-value for CR4 is 0.120, above conventional significance thresholds. The strength of the statistical evidence therefore varies with the concentration measure and sample composition.

CR4 measures the combined revenue share of the four largest firms by revenue, while the HHI

incorporates the revenue shares of all firms in the sample and gives greater weight to larger shares. The HHI thus captures changes in revenue shares outside the four largest firms, so the two measures describe different aspects of market structure. Our main conclusions concern revenue concentration as measured by the HHI.

Table 7. Sensitivity to the Minimum Firm Count per Cell

Sample threshold	Core coefficient	Cluster Standard Error	Conventional p-value	Wildpprice	N
At least 3 enterprises	-261.792	629.043	0.680	0.627	2703
At least 5 enterprises (primary sample	-2651.273	1119.624	0.025	0.034	1542
At least 10 enterprises	-4489.782	1537.457	0.015	0.085	523

As shown in Table 7, the sample requiring at least three firms per cell contains 2,703 observations. The estimated coefficient is −261.792, with a wild cluster bootstrap p-value of 0.627. The main sample requires at least five firms per cell and contains 1,542 observations. The coefficient is statistically significant at the 5% level in this sample. Raising the threshold to ten firms reduces the sample to 523 observations and increases the absolute magnitude of the coefficient, although it is significant only at the 10% level. Changing the threshold affects both the coverage of firms underlying the concentration measure and the composition of the province–industry–year cells in the sample. Changes in the estimates should therefore be interpreted in light of these differences in coverage.

The baseline findings apply primarily to regional industry cells with sufficient coverage of listed firms. Where coverage is limited, revenue changes at an individual firm can have a greater influence on measured concentration. Assessing whether the findings extend to these cells requires data covering a broader set of firms.

5.4 HHI Decomposition and Potential Channels

To show how the two components contribute to the baseline coefficient, we use the HHI identity to decompose the index into a firm-count component and a revenue-share dispersion component.

$$HHI_{pit}=\frac{10000}{N_{pit}}+\left(HHI_{pit}-\frac{10000}{N_{pit}}\right) \quad (3)$$

The first term, $\frac{10000}{N_{pit}}$, is the minimum HHI attainable when the N_{pit} firms have equal revenue shares. The second term is the amount by which the observed HHI exceeds this minimum and captures the extent to which revenue shares depart from an equal distribution. The regressions for both components use the same control variables and fixed effects as the baseline model.

Table 8 shows that the coefficients of the number of enterprises and the dispersion of market share are − 597.072 and − 2054.200, respectively, and the difference between the sum of the two and the benchmark coefficient comes from the decimal rounding. The market share dispersion term contributes more to the benchmark coefficient, and the Wild p value is 0.072. The Wild p value of the number of enterprises is 0.028, and the statistical support is relatively stronger.

The estimated coefficient of the logarithm of the number of enterprises is positive, and the Wild p value is 0.363, which does not show a stable linear correlation of the number of enterprises. The quantity component is combined with the result of the logarithm of the number of enterprises to provide partial support for H2 a. The coefficient of the number of enterprises standardized HHI is still negative, which is significant at the 10 % level. This result is consistent with the estimation of the market share dispersion term, and the two results provide support for the expectation of H2b 's share equalization at the 10 % level.

Table 8. HHI Decomposition and Evidence on Potential Channels

explained variable, variable being explained	Core coefficient	Cluster Standard Error	Conventional p-value	Wildpprice	Group R^2
Number of enterprises component: 10000/N	-597.072	172.566	0.002	0.028	0.052
Market share discrete variable	-2054.200	1063.045	0.064	0.072	0.035
Logarithm of the number of enterprises	0.399	0.396	0.323	0.363	0.073
Enterprise Number Standardization HHI	-0.244	0.127	0.065	0.071	0.031

Note. All models include four control variables, province–industry fixed effects, and industry–year

 Copyright @ STEMM Institute Press

fixed effects. The sample contains 1,542 observations and 28 provincial clusters.

HHI identity decomposition is used to illustrate the statistical sources of changes in income concentration. The number of enterprises may be affected by the coverage of listed companies, whether the income of enterprises is positive and the non-linear change of the number of enterprises. To identify whether the digital economy affects HHI through specific channels such as market entry and relative efficiency of enterprises, it also needs corresponding behavioral data and identification design that can distinguish the order of action.

5.5 Dynamic Analysis

Firms may take time to adopt digital tools and adjust their operations, so changes in revenue shares may emerge in subsequent years. We estimate separate models using the contemporaneous digital economy index, its one-year lag, and its two-year lag. We then include all three terms in a single model and test whether the sum of their coefficients differs from zero to assess whether the negative association persists over time.

Table 9. Dynamic Associations Between Digital Economic Development and Revenue Concentration

effect	estimated value	Cluster Standard Error	Conventional p-value	Wildpprice	N
Current effect	-2651.273	1119.624	0.025	0.033	1542
One-period lag effect	-1992.266	720.922	0.010	0.008	1542
lagged second-phase effect	-1512.446	721.588	0.046	0.023	1542
Current-period and two-period lagged cumulative effects	-2597.807	958.519	0.012	0.008	1542

In Table 9, the coefficients from the separate regressions using one- and two-year lags are −1,992.266 and −1,512.446, with wild cluster bootstrap p-values of 0.008 and 0.023, respectively. In the joint model that includes the contemporaneous index and both lags, the three coefficients sum to −2,597.807, with a wild cluster bootstrap p-value of 0.008. These results provide statistical evidence that the negative association persists in the lagged specifications.

The digital economy index is highly persistent over time. In the joint model, correlation among its contemporaneous and lagged values can reduce the precision of individual coefficient estimates. We therefore use the sum of the coefficients from the joint model as our primary measure of the cumulative association. Identifying the timing of responses to specific policy shocks would require a separate analysis. The separate lag regressions and the sum from the joint model capture different aspects of the relationship. Together, they show how the findings depend on the timing of the explanatory variable.

5.6 Placebo Tests

5.6.1 Lead test

We include the one-year lead of the digital economy index alongside its contemporaneous value. A statistically significant association between the future index and the current HHI could indicate a pre-existing trend or suggest that changes in concentration precede changes in the digital economy index. The estimated coefficient on the one-year lead is −218.575, with a conventional province-clustered p-value of 0.885 and a wild cluster bootstrap p-value of 0.948. The contemporaneous coefficient remains negative. The lead test thus provides no evidence of an anticipatory association. Because the 2023 index is unavailable, the one-year lead cannot be constructed for 2022, leaving 1,294 observations in this specification.

5.6.2 Random permutation test

In order to test whether the results may be caused by the random matching of the provincial digital economy index and the regional industry sample, this paper randomly replaces the provincial digital economy index in each year, and copies the replacement value of the same province-year to each industry in the province, and repeats the estimation 1000 times. The design retains the annual distribution and provincial level structure of digital economic indicators.

Figure 1 presents a complementary within-year provincial permutation test, showing the observed coefficient relative to the distribution of coefficients obtained through random permutations.

As reported in Table 10, the observed coefficient lies below the lower bound of the central 95% of

Copyright @ STEMM Institute Press

the permutation distribution, with a two-sided randomization p-value of 0.001. Under this permutation design, randomly matching provincial index values to regional industry observations rarely produces a coefficient as extreme as the baseline estimate. The lead and permutation tests provide complementary evidence on temporal ordering and random matching, respectively. Time-varying omitted factors and the possibility of influence in both directions remain unresolved.

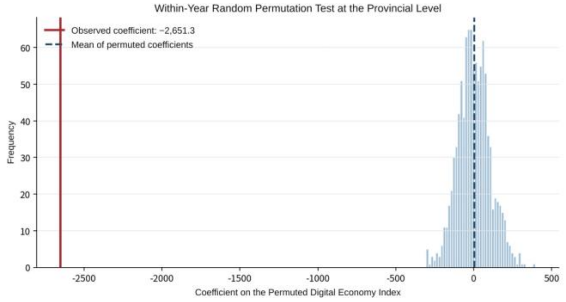


Figure 1. Within-Year Random Permutation Test at the Provincial Level

Table 10. Random Permutation Test Results

index	figure
Estimated coefficient	-2651.273
Mean of the 1,000 replacement coefficients	1.480
Standard deviation of the replacement coefficient	104.419
Replacement distribution: 95% interval	[-196.430, 217.328]
Bilateral randomization p-value	0.001

6. Conclusions and Policy Implications

6.1 Main Findings

This paper constructs a regional industry HHI based on the income share of listed companies in 28 provincial-level regions and 34 industries from 2015 to 2022. After controlling the characteristics of enterprises and absorbing the fixed effects of province-industry and industry-year, the development of digital economy is significantly related to the decline of income concentration. For every 0.01 increase in the digital economy index, the average HHI is about 26.51 points lower, and the Wild Cluster Bootstrap p value is 0.033.

Under the substitution index and sample processing, the direction of the core coefficient is generally consistent, and the statistical support varies by measure and coverage threshold. The sample results of CR4 and at least three companies are not significant. The main evidence focuses on HHI indicators and regional industry units with sufficient coverage of listed companies.

HHI decomposition shows that the discrete term of market share contributes more to the benchmark coefficient, and the statistical support of the number of enterprises is stronger. The logarithmic regression of the number of enterprises is not significant, and the evidence of quantity change is limited. The lag setting and cumulative coefficient test support the time continuity of negative correlation. No significant early correlation was found in the leading item test, and the actual estimated coefficient was at the tail of the random distribution in the provincial replacement within the year.

The sample results are consistent with the theoretical expectation that the inclusive competition effect is dominant. The conclusion is applicable to the income distribution of listed companies according to the place of registration during the study period. The specific digital behavior and its channels of action still need to be directly tested. The evidence mainly shows that when the digital economy develops rapidly, the income distribution of sample enterprises tends to be more dispersed. The judgment of other market levels or industry development stages still needs corresponding data.

6.2 Policy Implications

First, improve inclusive digital infrastructure and digital services for SMEs. The affordability of broadband, cloud services and public data services should be improved, and the focus should be on reducing the fixed and learning costs of small and medium-sized enterprises using digital tools. On the premise of ensuring trade secrets and personal information, credit evaluation can be improved by combining real transaction and performance information to support enterprises to obtain financing that matches the digital investment cycle. The policy effect can be combined with the actual use of the enterprise, the scope of customers and the evaluation of operating efficiency.

Second, reduce barriers to entry in the data and platform ecosystem. Promote data portability, platform interface compatibility and fair access, and focus on exclusive transactions, unreasonable traffic distribution, and excessive data migration costs. Competitive governance can judge the impact of relevant arrangements on market participation opportunities by

 Copyright @ STEMM Institute Press

combining platform network effects, business conversion costs and user choice space. For the limited links of cross-platform operation, the interface opening and merchant complaint mechanism can be improved. When evaluating the platform rules, the matching efficiency and the actual cost of the cross-platform operation of the trading party can be examined at the same time.

Third, continuously monitor the regional industry competition structure. The negative correlation between digital economy and concentration still exists in the lag setting, and policy evaluation can track regional and industrial changes across the year. In addition to HHI and CR4, the market competition state can be judged by combining enterprise entry and exit, price quality and innovation investment, especially focusing on the impact of listed company coverage change on income concentration index. The economic meaning of market share redistribution also needs to be judged in combination with changes in operating efficiency and consumer choice.

6.3 Limitations and Directions for Future Research

First, HHI collects the operating income of listed companies by registered provinces, does not cover non-listed companies, and the product market and geographical market that the company actually sells may also differ from the grouping scope. Listing, delisting and changes in operating income records may affect the number of enterprises in the sample. In the future, the external validity of the results can be tested by combining more complete enterprise income, entry and exit and sales data.

Second, the provincial digital economy index is difficult to describe the digital differences between industries, and the index weighting, principal component load and some basic indicators still need to be further verified. In the future, the digital facilities and application data at the industry level can be combined with the technology use indicators at the enterprise level to distinguish the regional digital environment from the enterprise 's own technology adoption.

Third, the fixed effect and placebo tests provide relevant evidence, and there is still the possibility of time-varying confounding factors and two-way effects. In the future, we can use phased digital infrastructure construction or policy pilots to strengthen causal identification, and directly test the path of search cost, financing constraints and platform lock-in in combination with enterprise digital technology use data. If we can observe the order of enterprise entry and exit and income share adjustment, we can further distinguish the scope of enterprise participation and the role of stock enterprise share reallocation.

Acknowledgments

This paper is supported by the 2025 Ministry of Education Humanities and Social Sciences Research Project (Planning Fund) “Research on the Effect of Digital Inclusive Finance on New Quality Productive Forces in Agricultural Counties: Theoretical Mechanism, Empirical Test and Policy Optimization Design” (No. 25YJA790071); the 2024 Guangdong Provincial Education Science Planning Project (Higher Education Special) “Research on the Development Path of Higher Education Clusters in the Guangdong–Hong Kong–Macao Greater Bay Area from the Perspective of Digital Economy” (No. 2024GXJK277); and the 2026 Wuyi University Regional and Country Studies Project “Efficiency Measurement and Mechanism Research on Digital Economy Empowering the Building of a China–Pacific Island Countries Maritime Community with a Shared Future” (No. QYGB2604).

References

[1] Curry B, George K D. Industrial Concentration: A Survey. Journal of Industrial Economics, 1983, 31(3):203-255.

[2] Sutton J. Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration. Cambridge, MA: MIT Press, 1991.

[3] De Loecker J, Eeckhout J, Unger G. The Rise of Market Power and the Macroeconomic Implications. Quarterly Journal of Economics, 2020, 135(2):561-644.

[4] Autor D, Dorn D, Katz L F, Patterson C, Van Reenen J. The Fall of the Labor Share and the Rise of Superstar Firms. Quarterly Journal of Economics, 2020, 135(2):645-709.

[5] Kwon S Y, Ma Y, Zimmermann K. 100 Years of Rising Corporate Concentration. American Economic Review, 2024, 114(7):2111-2140.

[6] Hsieh C T, Rossi-Hansberg E. The

Industrial Revolution in Services. Journal of Political Economy Macroeconomics, 2023, 1(1):3-42.
[7] Goldfarb A, Tucker C. Digital Economics. Journal of Economic Literature, 2019, 57(1):3-43.
[8] Rochet J C, Tirole J. Platform Competition in Two-Sided Markets. Journal of the European Economic Association, 2003, 1(4):990-1029.
[9] Armstrong M. Competition in Two-Sided Markets. RAND Journal of Economics, 2006, 37(3):668-691.
[10] Hagiu A, Wright J. Multi-Sided Platforms. International Journal of Industrial Organization, 2015, 43:162-174.
[11] Bessen J. Industry Concentration and Information Technology. Journal of Law and Economics, 2020, 63(3):531-555.
[12] Lashkari D, Bauer A, Boussard J. Information Technology and Returns to Scale. American Economic Review, 2024, 114(6):1769-1815.
[13] Jones C I, Tonetti C. Nonrivalry and the Economics of Data. American Economic Review, 2020, 110(9):2819-2858.
[14] Bergemann D, Bonatti A. Data, Competition, and Digital Platforms. American Economic Review, 2024, 114(8):2553-2595.
[15] Emery L P. Market Dominance in the Digital Age. Review of Finance, 2025, 29(4):1219-1258.
[16] ZHAO Tao, ZHANG Zhi, LIANG Shangkun. Digital economy, entrepreneurial activity, and high-quality development: Empirical evidence from Chinese cities. Management World, 2020, 36(10): 65–76. (in Chinese)
[17] Cameron A C, Miller D L. A Practitioner's Guide to Cluster-Robust Inference. Journal of Human Resources, 2015, 50(2):317-372.
[18] Cameron A C, Gelbach J B, Miller D L. Bootstrap-Based Improvements for Inference with Clustered Errors. Review of Economics and Statistics, 2008, 90(3):414-427.
[19] MacKinnon J G, Webb M D. Wild Bootstrap Inference for Wildly Different Cluster Sizes. Journal of Applied Econometrics, 2017, 32(2):233-254.
[20] Roodman D, Nielsen M Ø, MacKinnon J G, Webb M D. Fast and Wild: Bootstrap Inference in Stata Using boottest. Stata Journal, 2019, 19(1):4-60.

 Copyright @ STEMM Institute Press