Co-integration Study on the Relationship Between Regional Economic Development and Agglomeration of Rural E-commerce Industry

Yu Sun

Department of Economics and Management, Shandong Vocational College of Science and Technology, Weifang, Shandong, China

Abstract: The correlation between regional economic development and the agglomeration of the rural e-commerce industry is continuously strengthening, yet the economic linkage between the two still requires thorough analysis. This study establishes econometric indicators for the development of rural e-commerce and the agglomeration of the rural e-commerce industry based on spatial econometric methods, and constructs a co-integration relationship model for the years 2009-2021. The research findings are as follows: From the analysis of co-integration rank, there exists a significant co-movement relationship between regional economic development and the agglomeration of the rural e-commerce industry. From the results of the Vector Error Correction Model (VECM), there is a significant positive correction mechanism between regional economic development and the agglomeration of the rural e-commerce industry, indicating a strong positive correlation between the two. According to the Granger causality analysis results, regional economic development is the unilateral Granger cause of the agglomeration of the rural e-commerce industry, while other sequences reject the Granger causal relationship. From the perspective of forecast error variance decomposition, the prediction error of rural e-commerce spatial agglomeration mainly originates from itself and economic growth. Under a 10-period lag situation, the prediction error of rural e-commerce spatial agglomeration attributed to itself remains fixed at 65.19%, economic growth prediction error at 26.26%, and the impact of industrial structure at 8.55%. Regional economic development has become a core force driving the agglomeration of the rural e-commerce industry.

Keywords: Rural E-commerce; Regional Economic Development; Industry Agglomeration; Co-integration Analysis

1. Introduction

The rapid growth of regional economies is strongly supported by a robust industrial base. In the age of the Internet, the e-commerce sector is increasingly becoming a powerful force for transforming industrial structures and accelerating regional economic development. As China's information infrastructure continues to improve, an increasing number of regions are reaping the economic benefits of new e-commerce models, creating a development pattern driven by Internet information. In this context, the rural e-commerce industry has gained prominence.

Firstly, the spread of rural e-commerce helps to reduce information asymmetry, bridge the urban-rural circulation gap, and balance the market between supply and demand. Secondly, rural e-commerce development offers transformation and upgrade opportunities for numerous small and micro enterprises in rural areas. E-commerce platforms enable agricultural and trade enterprises to lower operating costs, reach larger markets, enhance the added value of agricultural products, and positively impact regional economic growth.

Nevertheless, some researchers have observed that the rise of e-commerce has given birth to numerous new formats, leading to the rapid clustering of high-tech industries and creating a significant "Matthew effect," where economically developed regions experience even faster growth. To thoroughly examine the correlation between regional economic development and the clustering of rural e-commerce industries in China, and to explore the causal relationship between the two, this study employs the Vector Error Correction Model (VECM) and forecast error variance decomposition. This approach aims to analyze the cointegration relationship, providing valuable insights for the sustainable development of rural e-commerce in China.

2. Literature Review

In examining the relationship between rural e-commerce and regional economic development, Kadir, N. et al. highlighted that rural e-commerce clusters result from the deep integration of rural informatization and industrialization, with their innovative capacity fostering new rural area construction. They identified weak economic foundations as the main factor hindering rural e-commerce cluster development, suggesting a strong link between regional economic growth and rural e-commerce agglomeration [1]. Liao, Y. et al. used grounded research methods to explore how rural e-commerce industry clusters and regional economic development can progress together at different stages [2]. Various scholars have delved into the reasons behind rural e-commerce agglomeration. For instance, Li, X., Jiang, J., & Cifuentes-Faura, J. emphasized that grassroots initiatives and government support drive rural e-commerce, with information access and online market demand shifts being crucial factors [3]. The development of rural e-commerce in China follows an internal logic of "entry-diffusion-selection-differentiation-entry" combined with external conditions. Su, J. et al. applied the system Generalized Method of Moments (GMM) to assess how e-commerce economic development impacts urban and rural income distribution in China [4]. Their findings indicated that while e-commerce significantly boosted economic participation and income levels among urban and rural residents, it also widened the income gap between these groups. Mei Y and Jiang YQ. examined the causes of rural e-commerce industry agglomeration by studying Taobao villages [5]. They found that government policies, population size, Internet access, and transportation infrastructure are key drivers of Taobao village development in the Bohai Rim region [6]. Government policies, in particular, were identified as the most influential factor, critical for the emergence and sustainable development of these villages, with urbanization processes having varying impacts [7].

In summary, existing research has extensively analyzed the correlation and causation between rural e-commerce development and regional economic growth and identified various factors contributing to rural e-commerce agglomeration. However, there is a lack of studies utilizing spatial econometric methods and co-integration analysis. From a spatial econometrics perspective, as China's rural e-commerce progresses into a high-speed integration phase, industry agglomeration trends become more prominent, yet its economic roles remain underexplored. From a co-integration perspective, there is a significant endogenous effect between regional economic growth and rural e-commerce clustering, necessitating a deeper influence analysis. Thus, this study builds upon existing research by employing co-integration analysis to investigate the econometric relationship between regional economic development and rural e-commerce agglomeration in China, ultimately offering conclusions and policy recommendations.

3. Model Selection

3.1 Model Specification

To validate the econometric relationship between regional economic development and the agglomeration of rural e-commerce industries, this study proposes to use the Vector Autoregressive (VAR) model to examine their long-term dynamic association. This method involves steps such as unit root tests, co-integration tests, impulse response analysis, variance decomposition, and stationarity tests [8]. The most crucial aspect is to introduce the lagged dependent variables of order p as independent variables, conduct a systematic hierarchical analysis of the co-integration relationships between multiple variables, and form a vector autoregressive model. The model equation is as follows:

$$Y_t = \omega_0 + \omega_1 Y_{t-1} + \omega_2 Y_{t-2} + \omega_3 Y_{t-3} + \cdots + \omega_p Y_{t-p} + \varepsilon_t \quad (1)$$

In Equation (1), Y_t represents a column vector composed of four core endogenous variables: regional economic growth (GDP), industrial structure (IS), development of rural e-commerce industry (NCD), and spatial agglomeration of rural e-commerce industry (NSG). φ represents the discriminant matrix corresponding to the coefficients of each lag order, ε_t represents the column vector of random disturbances, where $Cov(\varepsilon_t, \varepsilon) = 0$, t denotes the corresponding period, and p represents the lag order.

3.2 Variable Selection

Due to the lengthy time window required for co-integration analysis, this study selects relevant data from China between 2009 and 2021 as the research sample. To address potential heteroscedasticity issues caused by time series fluctuations in the sample, all endogenous variables are transformed using natural logarithms. The rationale for variable selection is as follows:

3.2.1 Development of Rural E-commerce (NCD)

To accurately reflect the level of rural e-commerce development and ensure comparability across different periods, this study employs the Lilien index constructed using the production factor method to characterize it:

$$EA_{it} = 1 - lilien_{it} = \sqrt{\sum_{i=1}^{n} \frac{LMP_{it}}{TLMP_{it}} \left(\Delta \ln LMP_{it} - \Delta \ln TLMP_t\right)^2} \quad (2)$$

In Equation (2), EA_{it} represents the overall productivity of rural e-commerce in region i at time t, LMP_{it} represents the number of rural e-commerce employees in region i, and $TLMP_t$ represents the total number of rural e-commerce employees nationwide. Generally, a higher value of EA_{it} indicates a higher level of rural e-commerce development in that region. The overall productivity and employment data are sourced from the Ministry of Commerce's “China E-commerce Development Report”.

3.2.2 Spatial Agglomeration of Rural E-commerce (NSG)

Building upon the assessment of rural e-commerce development level, this study employs local spatial autocorrelation variables to analyze the agglomeration of the rural e-commerce industry. Moran’s I index is introduced to characterize this aspect:"

$$Moran'sI = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij(x_i - \bar{x})(x_j - \bar{x})}}{s^2 \left(\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}\right)} \quad (3)$$

In Equation (3), $S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2, \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, y_i$ represents the level of development of rural e-commerce. w_{ij} represents the spatial weight matrix produced by the adjacency matrix. The Moran's I index ranges from -1 to 1. When Moran's I index is greater than 0, it indicates the presence of positive spatial spillover characteristics between regions; when Moran's I index is less than 0, it indicates the presence of negative spatial spillover characteristics between regions.

3.2.3 Regional Economic Growth (GDP)

This study uses the natural logarithm of the annual real per capita GDP for 30 provinces and municipalities in China. Nominal per capita GDP is obtained by dividing the annual GDP of each region by its year-end total population, measured in yuan per person. To account for inflation, the nominal GDP is adjusted using the Consumer Price Index (CPI), with 2009 as the base year, resulting in real regional per capita GDP. The data for GDP and CPI are extracted from the “China Statistical Yearbook”.

Industrial Structure (IS): From the standpoint of regional economic development, optimizing the industrial structure is vital [9]. The advancement of rural e-commerce can harmonize efficiency disparities among various industries, significantly impacting the core variables of this research. Consequently, this study represents industrial structure by the ratio of industrial value-added to GDP, with the CPI from 1980 used as the base period for adjusting the industrial output value [10]. Data on industrial value-added are also obtained from the “China Statistical Yearbook”.

4. Empirical Testing

4.1 Unit Root Test

Before conducting co-integration analysis, it is essential to perform unit root tests on each variable to avoid the phenomenon of "spurious regression" caused by multicollinearity and to ensure the overall stationarity of the econometric model. Using the ADF unit root test in the Stata 15.0 software environment (see Table 1), the null hypothesis is set as "the sequence has a unit root." The results indicate that the sequences GDP, IS, NCD, and NSG accept the null hypothesis at the 10% significance level, and even after taking the logarithm of the aforementioned sequences, they remain non-stationary sequences. Further differencing the sequences once, the results show rejection of the null hypothesis at the 5% significance level, indicating that the aforementioned sequences are all first-order integrated sequences I(1).

Table 1. ADF Unit Root Test

variable	Inspection type(C,T,P)	ADF	1%	5%	10%	Is it stable?
GDP	(C,T,1)	-1.871	-4.270	-3.492	-3.019	no
IS	(C,T,0)	-2.230	-4.721	-3.552	-3.017	no
NCD	(C,T,1)	-1.562	-4.715	-3.571	-3.131	no
NSG	(C,T,1)	-2.031	-4.715	-3.572	-3.132	no
GDP_1	(C,0,0)	-4.231	-3.561	-2.564	-2.031	yes
IS_1	(C,0,0)	-6.032	-3.603	-2.621	-2.029	yes
NCD_1	(C,0,0)	-3.998	-3.594	-2.760	-2.027	yes
NSG_1	(C,0,0)	-4.132	-3.642	-2.813	-2.015	yes

4.2 Co-integration Test

Combining Johansen's Maximum Likelihood Estimation (MLE) method for co-integration testing to determine whether the I(1) vectors of the sequences GDP, IS, NCD, and NSG exhibit a long-term stable relationship. Due to sample size constraints, lag two is used for testing all sequences, based on LR, AIC, BIC, and SC criteria. Additionally, linear trend analysis using the Pantula criterion reveals the presence of only an intercept in the co-integration equation. The results of the co-integration rank test are presented in Table 2.

Table 2. Cointegration Rank Analysis Results

sequence	rank	trace statistics	Eigenvalue statistics	5% critical value	The maximum number of covectors that exist
GDP; IS; NCD; NSG	1	28.1033*		37.26	1
GDP; IS; NCD; NSG	0		38.1301	31.03	1
GDP; IS; NCD; NSG	1		13.0335	24.21	1

The rank test results demonstrate that there is only one linearly independent co-integrating variable among the sequences. The maximum eigenvalue test rejects the null hypothesis of "co-integration rank equals 0" at the 5% significance level. Consequently, with a co-integration rank of 1, the sequences GDP, IS, NCD, and NSG are co-integrated at the 5% significance level, indicating a long-term stable equilibrium relationship among these variables. Using the Maximum Likelihood Estimation (MLE) method, the following co-integration equation is derived:

$$NSG = 1.905 + 0.174GDP + 0.075IS + 0.274NCD \quad (4)$$

The results from Equation (4) show a significant positive relationship between the variables GDP, IS, NCD, and NSG. Specifically, a 1% increase in regional economic growth leads to a 0.174% rise in the spatial agglomeration of rural e-commerce. This indicates that economic development is a key driver of rural e-commerce clustering. Additionally, it is evident that economically advanced regions tend to exhibit higher levels of rural e-commerce development. A 1% improvement in industrial structure results in a 0.075% increase in rural e-commerce spatial agglomeration, highlighting the importance of optimizing industrial structures to promote rural e-commerce, albeit with a smaller impact compared to economic growth. Finally, a 1% growth in rural e-commerce development leads to a 0.274% increase in its spatial agglomeration, underscoring that the primary factors influencing rural e-commerce clustering are its industrial progress and economic status.

4.3 Vector Error Correction Model (VECM)

Since the data series in this paper only consists of 13 statistical periods, this study employs the Vector Error Correction Model (VECM) in the analysis within the vector autoregressive model.

$$\triangle NSG_t = \begin{bmatrix} 1.842 \\ 0.168 \\ 0.068 \\ 0.255 \end{bmatrix} \times VECM_{t-1} + \begin{bmatrix} 1.591 \\ 0.143 \\ 0.103 \\ 0.312 \end{bmatrix} \quad (5)$$

This yields the first equation as follows:

$$\triangle NSG_t = 0.0213 \times VECM_{t-1} + 0.1032 \quad (6)$$

The corresponding coefficients are significantly positive, indicating conformity to a positive error correction mechanism.

4.4 Granger Causality Test

Further analysis is performed to assess whether a Granger causality relationship exists between regional economic development and the spatial agglomeration of rural e-commerce, excluding the influence of rural e-commerce development. To explore potential predictive relationships, lagged values are used to analyze the dependent variables. Based on the AIC and SC criteria, the optimal lag order is identified as 2. The test results are summarized in Table 3.

Table 3. Granger Causality Test Results

Null hypothesis	F	p	in conclusion
GDP is not the Granger factor of NSG	10.2312	0.0216	reject
IS is not the Granger cause of NSG	4.0312	0.1320	accept
GDP is not the Granger cause of IS	0.2394	0.1423	accept
NSG is not the Granger cause of IS	1.1982	0.2130	accept
IS is not the Granger cause of GDP	1.8236	0.2256	accept
NSG is not the Granger cause of GDP	0.3649	0.2597	accept

From the results in Table 3, it is evident that at the 5% significance level, only the rejection of the null hypothesis "GDP is not a Granger cause of NSG" is observed, indicating that GDP is a Granger cause of NSG. Thus, from the perspective of econometric analysis, regional economic development serves as a significant factor driving the spatial agglomeration of rural e-commerce.

4.5 Forecast Error Variance Decomposition

Based on the VAR model, the importance of the interaction between rural e-commerce and economic growth is derived, i.e., the proportion of contributions of each variable to the total contribution.

Table 4. Forecast Error Variance Decomposition

forecast period	Variance decomposition of NSG			Variance decomposition of GDP			Variance decomposition of IS		
	NSG	GDP	IS	NSG	GDP	IS	NSG	GDP	IS
1	0.8035	0.1302	0.0663	0	0.9955	0.0045	0	0	1
2	0.7274	0.2097	0.0629	0.0205	0.9642	0.0153	0.0316	0.0031	0.9653
3	0.7167	0.2124	0.0709	0.0283	0.9554	0.0163	0.0759	0.0232	0.9009
4	0.6776	0.2476	0.0748	0.0215	0.9536	0.0249	0.1227	0.0259	0.8514
5	0.6564	0.2676	0.0760	0.0293	0.9456	0.0251	0.1262	0.0260	0.8478
6	0.6572	0.2627	0.0801	0.0310	0.9424	0.0266	0.1265	0.0259	0.8476
7	0.6664	0.2531	0.0805	0.0337	0.9396	0.0267	0.1273	0.0216	0.8511
8	0.6565	0.2628	0.0807	0.0326	0.9407	0.0267	0.1278	0.0261	0.8461
9	0.6540	0.2626	0.0834	0.0313	0.9418	0.0269	0.1287	0.0261	0.8452
10	0.6519	0.2626	0.0855	0.0373	0.9358	0.0269	0.1290	0.0261	0.8449

From the empirical results in Table 4, it is observed that the forecast error of rural e-commerce spatial agglomeration primarily stems from itself and economic growth. For instance, in the first period, 80.35% of the forecast error in rural e-commerce spatial agglomeration originates from itself, 13.02% from economic growth, and only 6.63% from industrial structure. By the fifth period, the forecast error attributable to rural e-commerce spatial agglomeration decreases to 65.64% from itself, while that from economic growth increases to 26.76%, with industrial structure's influence remaining relatively low at 7.6%. By the tenth period, the portion of the forecast error of rural e-commerce spatial agglomeration originating from itself remains fixed at 65.19%, economic growth prediction error at 26.26%, and industrial structure's influence at 8.55%.

Overall, the self-influencing factors of rural e-commerce spatial agglomeration dominate both in the short and long term. However, as the forecasting period extends, the proportion of regional economic development's impact gradually increases, consistent with the earlier results, demonstrating that regional economic development is a core driving force behind rural e-commerce spatial agglomeration. From the perspective of regional economic development, its own forecast error decreases from 99.55% in the first period to 93.58% in the tenth period, corresponding to an increase in the forecast error of rural e-commerce spatial agglomeration from 0 to 3.73%, and industrial structure's forecast error from 0.45% to 2.69%. This indicates that China's regional economic development exhibits strong autocorrelation, signifying persistent economic momentum. Concurrently, the spatial agglomeration of the e-commerce industry is gradually becoming a significant force driving regional economic development, while the role of optimizing industrial structure in regional economic development is continuously highlighted. From the perspective of industrial structure optimization, its own forecast error proportion is substantial, reaching 84.49% in the tenth period. In contrast, rural e-commerce spatial agglomeration's forecast error increases to 12.90%, and regional economic development's forecast error increases to 2.61%. This underscores that industrial structure optimization possesses long-term inertia, with e-commerce industry agglomeration emerging as a vital driver, and economic development playing a role, albeit less than that of e-commerce industry agglomeration.

5. Conclusion

This study developed a cointegration model for 30 provincial regions spanning from 2009 to 2021, providing a quantitative analysis of the relationships among rural e-commerce development, rural e-commerce industry agglomeration, regional economic development, and industrial structure. The empirical findings indicate a significant cointegration relationship between rural e-commerce industry agglomeration and regional economic growth in China, with the former having a notable impact on promoting the latter. According to the Granger causality test results, regional economic growth is the Granger cause driving the rural e-commerce industry. From the perspective of forecast error variance decomposition, by the tenth period, 65.19% of the forecast error in rural e-commerce spatial agglomeration is attributable to itself, 26.26% to economic growth, and 8.55% to industrial structure. Given that rural e-commerce development in China is still in its early stages with relatively low spatial agglomeration, the current impact of rural e-commerce industry agglomeration on regional economic development is not yet significant. However, from a predictive standpoint, there is substantial room for improvement in rural e-commerce industry agglomeration, which has the potential to significantly boost regional economic growth in China.

References

[1] Kadir, N., & Jusoff, K. (2010). Investigating the Relationship between Tourism and Trade in Malaysia Using Cointegration and Causality Tests. International Journal of Economics and Finance, 2 (1), 138-143.
[2] Liao, Y., Wu, G., & Huang, J. (2023). Examining the Influence of Rural E-Commerce Development on Orchard Expansion through Tele-Coupling: A Study of Pinghe County, Southeast China. Land, 12 (11), 1991.
[3] Li, X., Jiang, J., & Cifuentes-Faura, J. (2023). Logistic Environment and Spatial Spillover Effects on Agricultural Economic Growth: Empirical Evidence from Eastern, Central, and Western China. PLoS One, 18 (7), e0287307.
[4] Su, J., Shen, T., Ma, W. C., & Zhang, J. X. (2023). Assessing the Impact of Rural Low Carbon Logistics on Economic Growth. IEEE Access.
[5] Mei, Y., & Jiang, Y. Q. (2020). Analyzing the Synergistic Development of Rural E-Commerce Industry Agglomeration and Regional Economy under Rural Revitalization—A Case Study Approach Based on Industrial Cluster Life Cycle Theory. China Rural Economy, (6).
[6] Chen, J. (2021). The Effect of E-Commerce Development on Agricultural Industry Agglomeration within the Context of Rural Revitalization: A Guangxi Case Study. Commercial Economy Research, (8).
[7] Gao, F., Fan, L. Y., & Gong, M. G. (2022). Spatial Distribution Characteristics and Driving Factors of Rural E-Commerce Industry: Analyzing Development Cases of Taobao Villages in the Bohai Rim Region. Price Theory and Practice, (11).
[8] Zeraibi, A., Balsalobre-Lorente, D., & Shehzad, K. (2020). Investigating the Asymmetric Relationship between Energy Consumption, Technological Innovation, and Economic Growth: Does Energy

Consumption and Technology Promote Economic Development? Sustainability, 12 (21), 8867.
[9] Long, X. (2024). Addressing Risk Challenges for High-Quality Rural E-Commerce Development. Logistics Technology, (04), 100-102. doi: 10.13714/j. cnki.1002-3100.2024.04.019.
[10] Wang, L., & Teng, H. (2024). Exploring Paths for High-Quality Rural E-Commerce Development in the Context of Digital Business Promotion. Agricultural Economics, (02), 136-137.