

Research on the Contribution of Regional Digital Financial Inclusion Based on PCA-ANN-MLP Algorithm

Pengyu Su^{1, #, *}, Shijia Wan^{2, #}, Dafu Zhu^{3, #}

¹School of Control Science and Engineering, Shandong University, Shandong, China, 250100

²School of Finance, Central University of Finance and Economics, Beijing, China, 102206

³Department of Finance, School of Economics, Xiamen University, Fujian, China, 361005

*Corresponding author: 202000171228@mail.sdu.edu.cn

#These authors contributed equally.

Abstract. Digital Financial Inclusion is a hot topic under the theme of China's economy today. Digital Financial Inclusion is an important supplement and extension of traditional Financial inclusion. Its core idea is to reduce the threshold of financial services and ease the financing difficulties of low-income groups, small and micro enterprises, and other vulnerable groups by providing low-cost and efficient financial services to the public. The empirical results of much literature show that digital Financial inclusion has narrowed the gap between the rich and the poor, which plays a vital role in the high-quality development of China's economy. Therefore, it is important to explore its influencing factors and analyze the influencing mechanism. This study focuses on the three provinces in eastern China and applies the grounded theory approach. Utilizing data from the Peking University database, 17 factors influencing the level of digital inclusive finance are selected. PCA (principal component analysis) is used to reduce dimensions, and ANN-MLP (artificial neural network multi-layer perceptron) algorithm is used to explore the contribution factors of Digital Financial Inclusion, explore important factors, and put forward relevant suggestions. The significance of this study is that taking the three provinces in eastern China as an example, it explores the factors influencing the development of DFI in China, and makes an important contribution to the issue of how to promote the rapid development of digital Financial inclusion in China.

Keywords: ANN-MLP; Financial inclusion; Contribution analysis; PCA.

1. Introduction

As a financial model that can provide better financial services for the poor, digital Financial inclusion has become an important topic of global development, and its vigorous development has made great contributions to poverty eradication and social equity. At the same time, it is also a strong source of driving force for the high-quality development of China's economy. Currently, more and more research has focused on this.

In recent years, much literature on this topic has focused on exploring the policy effects of digital Financial inclusion at the macro level. At the provincial level, NIE&JIANG^[1] used a two-step system GMM and dynamic threshold regression model to prove that digital finance can promote the development of regional innovative technologies and the degree of promotion is related to human resource investment and regional systems, with heterogeneity in time and space. LI&HAN^[2] uses DID double difference method to test the policy effect of Financial inclusion from the county and provincial levels. The empirical research results show that the development of Financial inclusion in the initial stage can reduce the unfair income distribution between urban and rural areas to a certain extent, but this effect is restricted by many factors. More inclusive, diversified, and personalized banking services do not benefit the bottom people. The resource allocation structure of formal financial institutions should be more inclined towards vulnerable groups. The research of WANG&FAN^[3] also found that digital Financial inclusion has a significant carbon emission reduction effect and its intensity changes with the geographical location and socio-economic characteristics of the county.

There are also some studies focusing on micro individuals to explore the improvement of digital Financial inclusion on the level of personal utility. LEI&SHEN^[4] empirically studied the main characteristics of fraud among the elderly in China and the mitigation effect of digital Financial inclusion using the data from the China Health and Retirement Tracking Survey (CHARLS). The research finds that individual cognitive ability and the development of local digital Financial inclusion are important factors that determine whether the elderly are cheated. FU&HUANG^[5] used the Panel data regression method with Instrumental variables estimation to study the heterogeneous impact of digital finance on different types of rural financial demand. The results indicate that the overall improvement of digital finance has reduced the probability of rural productive formal credit demand and also increased the probability of rural consumer formal credit demand, reflecting the multidimensional effects of digital finance in improving efficiency and promoting consumption.

In addition, the latest research has also explored the main factors that promote the development of digital Financial inclusion. WANG&HE^[6] explored the time and space characteristics of digital Financial inclusion development in different regions of China and applied the Weave Thomas model to explore the core driving force of digital Financial inclusion development in 34 provinces of China. Looking at previous studies, it is found that most studies have different development drivers for digital Financial inclusion in different regions. The eastern region is mainly driven by innovative development, the central region is driven by endogenous release, and the western region relies on market applications. However, compared to the first two categories, there are relatively few studies in this area and there is a significant gap. Based on the Panel data of the three provinces in eastern China and the PCA-ANN-MLP algorithm, this paper studies the contribution factors of regional digital Financial inclusion.

This paper makes a new supplement to the research direction of inclusive finance, and clarifies the main factors supporting the development of digital inclusive finance in the three provinces and regions of eastern China, laying the foundation for further in-depth research and promoting the high-quality development of social economy.

2. Construction of Model Indicators.

2.1. Dependent variable (digital inclusive finance)

Digital inclusive finance (DIF) is a financial model that uses digital technology and innovative means to provide financial services to the general public. Its aim is to lower the cost and threshold of financial services through digital channels and technological means, enabling more people to access financial services, especially those who are difficult to reach through traditional financial services, such as low-income groups, rural residents, and small and micro entrepreneurs. DIF provides financing support to small and micro enterprises, promotes entrepreneurship, innovation, and employment growth, and drives sustainable economic development. At the same time, DIF provides convenient, fast, and personalized financial services, improves user experiences, and meets the diverse financial needs of users.^{[7][8]}

In order to more accurately measure the development of digital inclusive finance in China, the Digital Finance Research Center of Peking University and Ant Group jointly compiled the China Digital Inclusive Finance Index. This article selects its released digital inclusive finance total index, coverage breadth, usage depth, digitalization degree as the dependent variable indicators for this study $Y_i (i = 1, 2, 3, 4)$.

In conclusion, Table 1 presents the constructed dependent variable indicator system.

Table.1. Dependent variable table

Primary indicator	Indicator code	Reference
Digital inclusive finance total index	Y_1	[7]
Coverage breadth	Y_2	
Usage depth	Y_3	
Digitalization degree	Y_4	

2.2. Independent variable

2.2.1 Macroeconomic

Macroeconomic research is concerned with the overall operation, growth, and stability of the economic system, instead of the behavior of individual economic entities. Thus, it is concerned with macroeconomic variables such as gross domestic product (GDP), GDP growth rate, and inflation rate. [9]

The development of inclusive digital finance is intimately associated with economic growth. Economic growth can facilitate an improvement in people's income levels, thereby enhancing their capacity and demand for using inclusive digital finance. Furthermore, the economic growth can offer more business opportunities and a wider market to various digital financial technology companies. In addition, the flexibility of macroeconomic monetary policy and interest rate levels affects inclusive digital finance institutions' financing atmosphere and credit circumstances. During times of low inflation, the central bank might implement expansionary monetary policy measures, which can stimulate the inclusive digital finance institutions' development and innovation. Additionally, international trade indicators affect inclusive digital finance. High export totals and trade surpluses signal a country's economic vitality and competitiveness. These attributes entice foreign investment and capital inflows, resulting in a broader funding source and market demand for inclusive digital finance.

In conclusion, this article focuses on the following variables in the macroeconomic dimension: GDP growth rate(X_{10}), economic development level(X_{11}), inflation rate(X_{12}), foreign trade balance(X_{13}), and foreign trade scale(X_{14}).

2.2.2 Mid-level economics

Mid-level economics acts as a bridge between macroeconomics and microeconomics, primarily focusing on relatively small economic units like regions, departments, and industries. Regional economic activities enable mid-level economics to unearth factors that impact the developmental dynamics, market conditions, and product innovation of the digital inclusive finance industry, providing both theoretical and empirical backing to the development of digital inclusivity in finance. [10][11]

Both the developmental level and characteristics of regional economies, in addition to the industry structure, growth rate, and stability of these economies, will significantly impact digital inclusion in finance. Regions with more developed economies have a relatively more stable business and investment environment that encourages the growth and evolution of digital inclusivity in finance. Other factors such as market concentration, entry barriers, and competition levels can significantly affect the market structure and the supply of digital inclusivity in finance. Besides, demographic structure, education level, employment rate, and income distribution are among the social and economic factors that could affect the acceptance, usage behavior, and demand regarding digital inclusivity in finance.

In essence, this paper focuses on nine primary dimensions of mid-level economics, namely regional economic development level (X_{20}), regional inflation rate (X_{21}), urban-rural gap (X_{22}), the degree of government intervention (X_{23}), regional age structure (X_{24}), urbanization rate (X_{25}),

traditional financial development level (X_{26}), industrial structure (X_{27}), and technological innovation (X_{28}).

2.2.3 Microeconomic

Microeconomic research investigates the behavior and decision-making process of individual economic units with a primary focus on topics like supply and demand, and resource allocation. It primarily examines micro-level economic variables such as a household's disposable income, savings, and education level. ^[12]

Households' disposable income serves as the source of funds for consumption and savings. Savings play a crucial role in accumulating wealth and managing risks. Furthermore, higher disposable income and savings promote financial inclusion and economic stability by increasing households' demand for and capacity to use digital financial services and products. A household's education level impacts their financial knowledge, financial management skills, and awareness of risks associated with financial tools. Higher levels of average education in society can enhance people's awareness of and capacity to utilize digital inclusive finance services.

In summary, household income level (X_{30}), wealth (X_{31}) and education level (X_{32}) serve as the key variables for the microeconomic analysis in this study.

Table 2 below provides a summary of the indicator system constructed in this article:

Table.2. Independent variable indicator system.

The secondary-level indicator.	The tertiary-level indicator	Indicator code	The measuring method
Macroeconomic	GDP growth rate	X_{10}	China's GDP growth rate
	Economic development level	X_{11}	The logarithmic value of China's per capita GDP
	Inflation rate	X_{12}	The change rate of CPI for residents with the previous year set as the base period
	Foreign trade balance	X_{13}	The logarithmic value of the trade balance of Chinese goods imports and exports
	Foreign trade scale	X_{14}	The logarithmic value of China's total volume of exported goods
Mid-level economics	Regional economic development level	X_{20}	The logarithmic value of per capita regional gross domestic product
	Regional inflation rate	X_{21}	The rate of change in the consumer price index (with the previous year as the base year) of residents in a specific region
	Urban-rural gap	X_{22}	The Theil index
	Degree of government intervention	X_{23}	The ratio of regional government financial expenditures to regional gross domestic product
	Regional age structure	X_{24}	The dependency ratio of the elderly population to the permanent resident population in a specific region
	Urbanization rate	X_{25}	The urban population divided by the total population
	Traditional financial development level	X_{26}	The ratio of loan balance of financial institutions to regional gross domestic product
	Industrial structure	X_{27}	The ratio of the value of the output of the tertiary industry to the value of the output of the secondary industry
	Technological innovation	X_{28}	The logarithmic value of the number of patent applications in each province
Microeconomic	Household income level	X_{30}	The logarithmic value of disposable income for households
	Household wealth	X_{31}	The logarithmic value of average per capita savings deposit balance
	Household education level	X_{32}	The average number of years of education

3. PCA-ANN-MLP method

3.1. PCA method

PCA is a classic dimensionality reduction method in multivariate statistical analysis. Based on the PCA method, samples can be mapped from the input space to a low-dimensional space through linear or nonlinear mapping to reduce redundant information and convert high-dimensional data into low-dimensional data that is easy to process. This paper uses this method to reduce the dimensions of 17 indicators that may be related to Financial inclusion to avoid dimensional disasters.

3.2. ANN-MLP Method

ANN is a flexible mathematical algorithm that imitates the operation mode of the human brain. It was proposed by psychologist W.S. McCulloch and a mathematical List of logicians W. Pitts in 1943. Neural networks are generally divided into an input layer, hidden layer, and output layer, consisting of a large number of nodes connected. The connection between each two node represents a weighted value passing through the signal, also known as weight. The topology structure of artificial neural networks is shown in Figure 1. The neural network inputs a large amount of input and output into the network, continuously adjusting the weights between each node to achieve the purpose of training and learning. Artificial neural networks have self-learning capabilities and the ability to quickly search for optimized solutions, with strong robustness and fault tolerance. Common artificial neural networks include Multilayer perceptron (MLP) and Radial basis function (RBF).

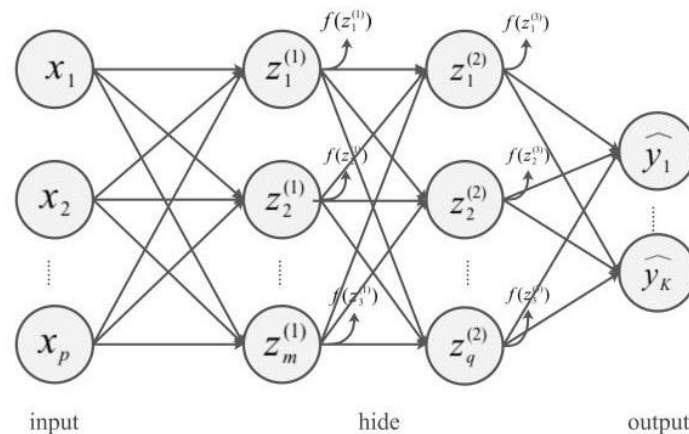


Figure 1. ANN topology diagram

MLP is the most commonly used neural network model, which is a typical Feedforward neural network. It consists of an input layer, a hidden layer, and an output layer, each layer consisting of several nodes directly connected to the previous and next layers. The value obtained from the previous layer is added with some weights, and the sum is converted with the Activation function. The MLP model is a major branch in ANN and holds an important position.

4. Setting of inspection indicators

Algorithm verification is a necessary step to verify the accuracy of a model. The conventional algorithm testing indicators include Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), and R Squared(R2). This article selects MSE as the testing indicator.

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - f(x_i))^2 \quad (1)$$

$$R^2 = \frac{SSE}{SST} \quad (2)$$

MSE refers to the average square of the distance between the predicted value of the model and the true value y of the sample. The smaller the MSE, the smaller the model error; R^2 is used to determine the closeness of the data to you and the regression line, with a value range of $[0,1]$. The closer it is to 1, the better the explanatory power of the variables in the equation for y , that is, the better the algorithm's fitting degree, and the stronger the model's robustness and generalization.

5. Case study

5.1. Research subject selection and data collection

This article investigates the three provinces in Northeast China, namely Liaoning, Jilin, and Heilongjiang, analyzing their relevance to the country's economy, society, population structure, and industrial structure. These regions are noteworthy industrial bases and areas abundant in resources, featuring unique economic development patterns and industrial structures, which are akin to government intervention and promotion strategies. Moreover, based on the Digital Inclusive Finance Index (DIFI) of Peking University, the indices of the three northeastern provinces have substantially dropped in 2020, contrary to the indices of 2015 and 2011, suggesting consistent issues faced in the regions. For this reason, studying the digital inclusive finance of these provinces can build a firm groundwork for future discussions on the digital inclusive finance development for industrial transformation and resources in the regions, demonstrating a certain degree of representativeness.

Panel data from 2011 to 2021 in the three northeastern provinces of China were selected for this study, constructing a panel data model comprising cross-sections and time series. The "Digital Inclusive Finance Index of Peking University" provided the DIFI, and other data were derived from provincial statistical yearbooks, national databases, "China Statistical Yearbook," and China Economic Database, while a few data were missing and filled with the relevant province's statistical bulletin. Thus, the number of observations of this study is 33. Table 3 presents the descriptive statistics of all the variables.

Table.3. Descriptive statistics

Indicator code /unit	Mean value	Standard error	Median	Standard deviation	Minimu m	Maximu m
Y_1 /unit	216.708	16.433	226.40 0	94.403	24.510	357.235
Y_2 /unit	196.186	16.209	194.17 0	93.113	21.120	340.100
Y_3 /unit	209.307	16.182	206.54 2	92.960	24.040	363.898
Y_4 /unit	297.931	20.254	330.21 5	116.348	27.860	420.060
X_{10} /%	6.973	0.313	7.000	1.798	2.200	9.600
X_{11} /unit	4.737	0.019	4.731	0.110	4.560	4.910
X_{12} /%	0.024	0.002	0.021	0.011	0.009	0.054
X_{13} /unit	3.579	0.032	3.623	0.185	3.191	3.830
X_{14} /unit	4.632	0.010	4.619	0.060	4.561	4.782
X_{20} /unit	4.615	0.017	4.614	0.097	4.417	4.813
X_{21} /%	0.023	0.002	0.021	0.012	0.006	0.058
X_{22} /unit	0.063	0.002	0.063	0.011	0.038	0.084
X_{23} /%	0.295	0.009	0.285	0.053	0.213	0.400
X_{24} /%	0.163	0.008	0.157	0.044	0.097	0.267

$X_{25}/\%$	0.595	0.006	0.597	0.034	0.534	0.657
$X_{26}/\%$	1.536	0.058	1.578	0.334	0.860	2.059
$X_{27}/\%$	1.303	0.059	1.359	0.336	0.676	1.953
X_{28}/unit	4.269	0.048	4.291	0.274	3.692	4.904
X_{30}/unit	4.328	0.020	4.330	0.117	4.101	4.545
X_{31}/unit	4.665	0.034	4.667	0.195	4.327	5.050
X_{32}/year	9.724	0.057	9.898	0.330	9.065	10.232

5.2. Dimensionality reduction based on PCA

Based on standardization, this study combines the use of Matlab software to conduct PCA dimension reduction.

To avoid the issues of dimensionality curse and overlapping information in indicators, we performed a dimensionality reduction on the indicators, which ultimately identified five principal components, namely, the Comprehensive Economic Development Index (F_1), Regional Economic Development and Transformation Index (F_2), Regional Inflation Index (F_3), Talent-driven Financial and Trade Development Index (F_4), and Talent-driven Technology Development Index (F_5). These components can explain 94.68% of the variance and satisfy the requirements of dimensionality reduction. Table 4 displays the results of the variances that the calculated principal components explained.

Table.4. Explained variance of principal component

Factor code of the principal component	Eigenvalue	Contribution rate/%	Cumulative contribution rate /%	Including indicators
F_1	10.1934	59.96	59.96	$X_{1j}(j=0,1,2,3); X_{2j}(j=0,1...8); X_{3j}(j=0,1,2)$
F_2	2.6176	15.40	75.36	$X_{20}, X_{22}, X_{23}, X_{27}, X_{32}$
F_3	1.4254	8.38	83.74	X_{12}, X_{21}
F_4	0.9652	5.68	89.42	X_{10}, X_{14}, X_{32}
F_5	0.8939	5.26	94.68	X_{28}, X_{32}

5.3. ANN-MLP Training

The reduced-dimensional results were imported into RStudio and normalized to ensure that all data were positive. The normalized dataset was randomly split into 75% for training and 25% for testing. The ANN-MLP algorithm was run with five principal component variables as input and four numerical inclusive finance evaluation indicators as output. The hidden layer parameters were set to 10-30 with a step size of 1, and iterations were performed based on MSE and R^2 , as shown in Figure 3. The MSE fluctuated around 0.0050, and as the model parameters increased, R^2 gradually approached 0.9950, indicating a good fit.

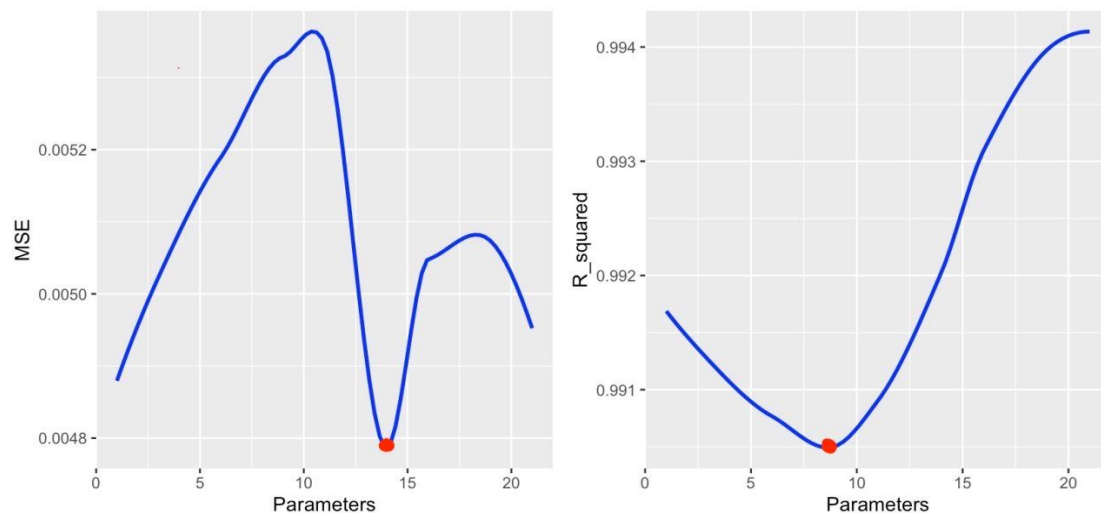


Figure 3. MSE and R^2

The group with the highest accuracy was selected as the final test result, and the importance weight matrix was obtained, which represented the importance of the five principal components for the four dependent variable indicators. The weight matrix was visualized, as shown in the following figure.

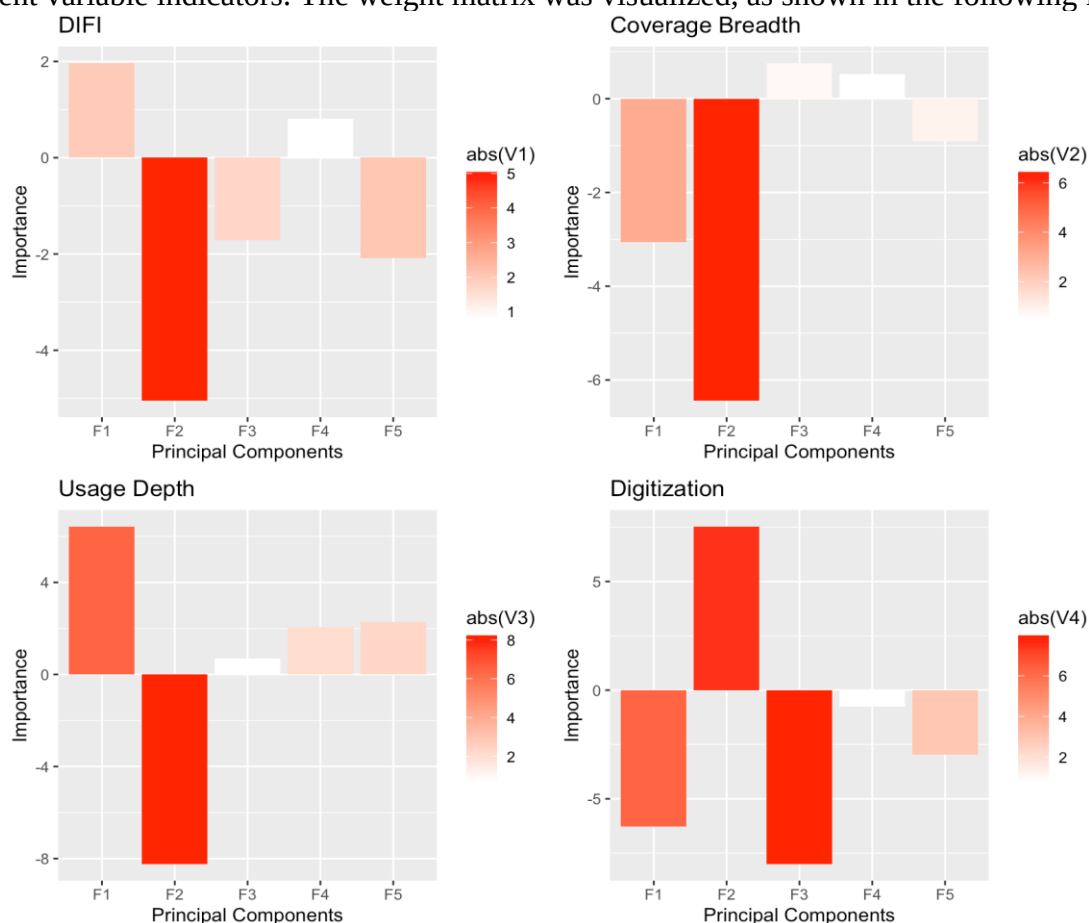


Figure 4. Training Outcome

Analyzing the weight matrix of the dependent variable Y_1 's numerical inclusive finance overall index, factor F_2 had the highest proportion, followed by F_5 . The most important factors affecting the overall index of digital inclusive finance in the region were urban-rural disparities, government intervention, industrial structure, education level, and technological innovation. It can be inferred that the most critical contributing factor was the size of the urban-rural gap. In addition, technological innovation may inhibit the development of inclusive finance to some extent. This may be because technology promotion has a certain degree of lag, and the dividends of short-term technological

innovation cannot benefit residents in poor areas, which deepens the digital divide and widens the gap between urban and rural areas, resulting in a negative impact on the overall index in a certain period.

In terms of coverage breadth, factors F1 and F2 dominated, while the impact of the other three factors was less significant. Therefore, similar to the conclusion drawn from the analysis of Y1, urban-rural disparities, government intervention, industrial structure, and education level were also important factors affecting the coverage breadth of digital inclusive finance in the region. Technological innovation did not have a significant weakening effect in this regard.

In terms of application depth, factor F4 made a significant contribution. Regions with better economic development, higher levels of technological innovation, and larger export trade scales have better application depth of digital inclusive finance. The region's economic conditions and technological level create conditions for better application of digital technology to deepen inclusive finance. The scale of export trade is related to the region's economic development and also reflects the degree of openness and communication in the region. The more frequent business and economic activities, the more conducive it is for new digital technologies to take root and develop in the local area and move towards deeper application areas.

And for digitalization, factor F3 made the highest contribution, and this factor showed a significant negative correlation with the dependent variable. This indicates that the total inflation and regional inflation rate increase will reduce the degree of digitalization of inclusive finance. This conclusion is consistent with the research of Liu Jun et al. in 2023, which stated that "the digital economy suppresses inflation through three channels: low growth in the prices of digital products and services, productivity improvement, and competition among e-commerce platforms."

6. Conclusion

From the above analysis, it can be concluded that factor F2 has a significant impact on the four measurement dimensions of inclusive finance, namely Y₁, Y₂, Y₃ and Y₄. Therefore, if we want to significantly improve the development level of regional digital inclusive finance, we can give priority to the development of influencing factors in F2, that is, narrowing the urban-rural gap, increasing government intervention, accelerating the speed of industrial structure optimization, and improving the level of education. In addition, factor F1 significantly affects the coverage breadth and deepening degree of inclusive finance, and factor F4, that is, the increase of inflation rate, will weaken the digitalization degree of inclusive finance.

In terms of model accuracy, R² of the model gradually converges to 0.995 after 20 iterations, which proves that the model has excellent fitting effect and strong interpretation. In addition, when the number of iterations is greater than 20, the MSE also gradually shrinks and finally changes around 0.005. The calculation results of the above two model test indexes show that the model constructed in this paper has strong energy efficiency, good robustness and generalization.

References

- [1] NIE Xiuhua, JIANG Ping, ZHENG Xiaojia, WU Qing. Research on Digital Finance and Regional Technology Innovation. *Journal of Financial Research*, 2021, 489(3): 132-150.
- [2] LI Jianjun, HAN Xun. The Effect of Financial Inclusion on Income Distribution and Poverty Alleviation: Policy Framework Selection for Efficiency and Equity. *Journal of Financial Research*, 2019, 465(3): 129-148.
- [3] WANG Shoukun, FAN Wencheng (2022). Digital Inclusive Finance and Carbon Reduction: An Empirical Analysis Based on County-level Data in China. *Contemporary Finance and Economics* (11), 53-64
- [4] LEI Xiaoyan, SHEN Yan, YANG Ling. A Study on Fraud Victimization among the Chinese Elderly in the Digital Age: The Role of Internet Usage and Digital Financial Inclusion. *Journal of Financial Research*, 2022, 506(8): 113-131.

- [5] FU Qiuzi, HUANG Yiping. Digital Finance's Heterogeneous Effects on Rural Financial Demand: Evidence from China Household Finance Survey and Inclusive Digital Finance Index. *Journal of Financial Research*, 2018, 461(11): 68-84.
- [6] WANG Xiaohua, HE Wenjin Research on the spatiotemporal characteristics and driving mechanisms of the development of digital inclusive finance. *Agricultural Economic Issues*.
- [7] Guo Feng, Wang Jingyi, Wang Fang, Kong Tao, Zhang Xun, Cheng Zhiyun. Measuring the Development of Digital Inclusive Finance in China: Index Compilation and Spatial Characteristics, 2019, 19(4): 1401-1418.
- [8] Huang Yiping, Huang Zhuo. China's Digital Financial Development: Present and Future[J]. *Economics (Quarterly)*, 2018, 17(04): 1489-1502.
- [9] Zhang Xun, Wan Guanghua, Zhang Jiajia, etc. Digital Economy, Inclusive Finance and Inclusive Growth [J]. *Economic Research*, 2019, 54(08): 71-86.
- [10] Xie Xuanli, Shen Yan, Zhang Haoxing, etc. Can Digital Finance Promote Entrepreneurship?——Evidence from China [J]. *Economics (Quarterly)*, 2018, 17(04): 1557-1580.
- [11] Qian Haizhang, Tao Yunqing, Cao Songwei, etc. Theory and Demonstration of China's Digital Financial Development and Economic Growth [J]. *Quantitative Economics and Technical Economics Research*, 2020, 37(06): 26-46.
- [12] Yi Xingjian, Zhou Li. Does the Development of Digital Inclusive Finance Significantly Affect Residents' Consumption——Microscopic Evidence from Chinese Households [J]. *Financial Research*, 2018, No.461 (11): 47-67.