Analysis on the Influence of Digital Inclusive Finance on Regional Economic Benefits and Its Factors

Yue Gui[1,a], Tingying Liu[2,b], Xiaowu Wang[3,*]

[1]Chengdu University of Information Technology, Chengdu, China

[2]Communication University of China, Beijing, China

[3]NanChang Hangkong University, Nan Chang, China

[a]1270743451@qq.com; [b]ella445566@sina.com; *promethusa@163.com

Abstract. Based on the digital inclusive financial index from 2011 to 2020, this paper studies the impact of various digital inclusive financial indicators on the provincial economy. With the help of Bayesian cumulative tree model, namely BART model, the importance of indicators is ranked according to the proportion of variables included in the tree and the importance of variables. According to the empirical results, (1) Among the indicators affecting the provincial economy, the use of depth index contributes the most, followed by digitization, payment, coverage, credit and insurance. (2) the introduction of high precision and explanatory and not easy to over-fitting model to verify the impact of digital inclusive finance in the provincial economy and its factor analysis, and achieved good results.

Keywords: Digital Inclusive Finance, Influencing Factors, BART Model.

1. Related theories and concepts

Under the background of digital technology, the development of digital inclusive finance in our country is at the height of national strategy and enters the stage of stable development. In recent years, China 's inclusive finance has made great progress, the first inclusive financial development plan has been successfully concluded. The decline in financing costs has effectively promoted the recovery and development of the theme of small and micro markets in employment stability. It has played an important role in promoting inclusive finance. Digital inclusive finance improves the level of risk management for financial institutions, solves the problem of information asymmetry, provides the public with the convenience of financial services, and overcomes the key problems of financial development, thus promoting economic and financial development. With the advancement of digital inclusive finance, the western region can also enjoy digital financial services. The unbalanced economic and social development between the eastern and western regions is an existing problem. The birth of digital finance has greatly promoted innovation and entrepreneurship in the western region and alleviated the imbalance and insufficiency of development between domestic regions. At the same time, digital finance combined with the platform economy, promote the growth of consumer demand, led to economic development. Wu (2021) found that digital inclusive finance promotes the provincial economy and has a single threshold feature. He and Zhang (2022) used the system GMM method to find that digital inclusive finance has a positive effect on high-quality economic development at both the national and regional levels, and has a stronger role in promoting high-quality economic development in the central and western regions than in the eastern region. Liu Chunlan (2021) used data from 31 provinces across the country as a full sample, selected by the F test and the Hausman test.

2. variable description

2.1 Variable description and processing

In order to compare and calculate the coefficients of the subsequent regression model more easily, we reduce the total index, namely the digital inclusive financial index, by 10,000 times. The index is divided into three aspects : coverage breadth, use depth and digitization degree. In addition, the value of each dimension is equal to the arithmetic mean of the corresponding metric. The table shows the descriptive statistics for the 20 provinces and municipalities in China from 2011 to 2020.

Table 1 Variable statistics table

variable	Mean value	standard deviation	maximum value	minimum value
digital inclusive financial index	0.189	0.001	0.378	0.018
Coverage breadth	0.203	0.083	0.204	0.202
used depth	0.144	0.075	0.411	0.055
digitized degree	0.230	0.071	0.514	0.135

3. Model introduction and technical route

3.1 BART regression model

Bayesian cumulative regression tree model, namely BART model, is a Bayesian nonparametric model. It uses the combination of simple functions to approximate complex functions and has sufficient regularization, which ensures that the model has good performance in accuracy and generalization. The core of the BART model is that the model is a plurality of general binary regression trees and the parameter prior is regularized. The idea is that the sample set generates the cumulative tree model through unsupervised learning and prior parameters. With the help of the Backfitting MCMC sampling algorithm, the convergence of the algorithm is evaluated by the posterior variance estimation drawn by the Gibbs sample number, and the parameters are estimated to generate model fitting. The specific form of BART model is as follows :

$$\begin{aligned} &\boldsymbol{Y} = f(\boldsymbol{X}) + \mathcal{E} \approx \mathcal{T}_1^{\mathcal{M}_1}(\boldsymbol{X}) + \mathcal{T}_2^{\mathcal{M}_2}(\boldsymbol{X}) + \ldots + \mathcal{T}_m^{\mathcal{M}_m}(\boldsymbol{X}) + \mathcal{E}, \\ &\mathcal{E} \sim \mathcal{N}_n\left(0, \sigma^2 \boldsymbol{I}_n\right) \end{aligned} \tag{1}$$

4. Analysis of the influencing factors of digital inclusive financial development on economic benefits

Combined with professional knowledge and literature in related fields, it can be seen that simple linear models cannot accurately interpret the more complex relationships in the economic field. Therefore, this study selects the Bayesian cumulative regression tree model (BART) with both accuracy and interpretability to study the problems in this paper. The BART model uses the posterior distribution P to sample as the model parameter estimation, uses the Gibbs sampling in MCMC, and introduces the Bayesian backfitting technique. At present, existing studies have compared BART regression with random forest, lasso regression, stepwise regression and other methods based on simulated numbers and actual data, and also verified the advantages of BART regression in variable importance identification of feature importance assessment methods. This article takes the above variables as independent variables. Taking the normalized value of provincial current price GDP in 2020 as the dependent variable, the BART regression model is constructed and then optimized by ten-fold cross-validation to determine a more accurate model.

4.1 Descriptive statistics

4.1.1 Data distribution

In order to observe the data distribution more directly, we draw the box plot through statistical software. The results are shown in the following figure :

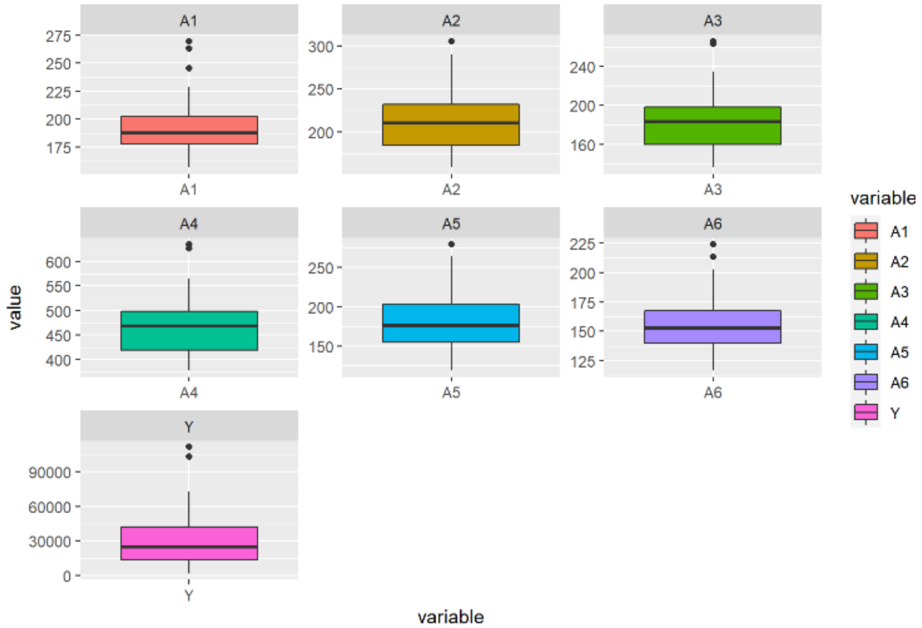

Figure 1 Box diagram of variables

According to the data set data, the variables in this paper are qualitative data. It can be observed from the above figure that all variables have outliers, but the data of variables A2, A5 and A6 are relatively concentrated, and the data of variables A1, A3, A4 and Y are relatively scattered. Variables A1 and A4 show a more obvious skewed distribution.

4.1.2 Variable relevance

Considering that there may be a certain correlation between the respective variables, the correlation heat map is drawn. The results are as follows :

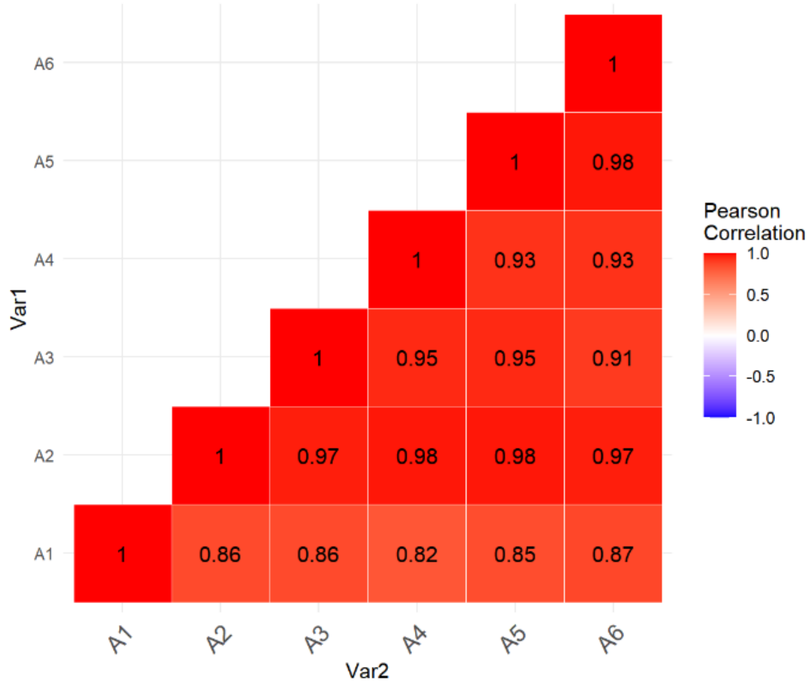

Figure 2 Correlation heat map of variables

The correlation coefficient between a single independent variable and other independent variables is calculated. The output range of Pearson correlation coefficient is -1 to $+1$. It can be seen from the above figure that each correlation coefficient is greater than 0.8. Based on this, it can be seen that there is a very significant positive correlation between each independent variable. Therefore, a simple linear regression model cannot be used to solve the research problems in this paper.

4.2 Construction and analysis of BART model

4.2.1Build BART model

Considering the accuracy of the model, this study divided the data set into 6 : 4 ratios. Using Rstutio to establish the model, the results are as follows :

Table 2 BART model construction results table

Test statistic	numerical value
L2	0.01
RMSE	0.02
Pseudo-Rsq	0.6085
SW test P value	0.72608

It can be observed from the results of the above table that the L2 norm loss function is LSE.Through the boxplot, it can be clearly observed that although there are outliers in the variables of the data set studied in this paper, the index value is 0.01, indicating that the gap between the predicted value and the real value is small. The pseudo-R2 value is 0.6085, the variable interpretation is better, and the model fitting effect is better ; using the test method to test the normality of the model residual, the P value is 0.72608 greater than 0.05, so the null hypothesis is accepted, that is, the residual sequence obeys the null hypothesis, so it conforms to the condition that the residual of the BART model used in this paper obeys the normal distribution.

4.2.2 Cross-validation preferred prior parameters

Table 3 Parameter optimization results

Hyperparameter combination serial number	k	nu	q	num_trees	oos_eror	Diff with lowest
1	2	10	0.75	100	0.024587	0.000
2	2	10	0.75	50	0.024588	0.1404
3	2	10	0.75	150	0.024648	0.2836
4	2	3	0.99	50	0.024677	0.3750
…	…	…	…	…	…	…

Referring to the parameter optimization results in the above table, the hyper-parameter combination of the first row is selected for modeling. The results are as follows :

Table 4 Optimal hyperparameter modeling verification results

Test statistic	numerical value
L2	0.01
RMSE	0.02
Pseudo-Rsq	0.6127
SW test P value	0.74149

With the help of cross-validation parameters and the selection of optimal parameters, the BART model is reconstructed. Each index value is relatively stable and statistically significant, indicating that the model has a good fitting effect.

4.2.3 Model effect test

4.2.3.1 Compare the effect of training set before and after 10-fold cross validation

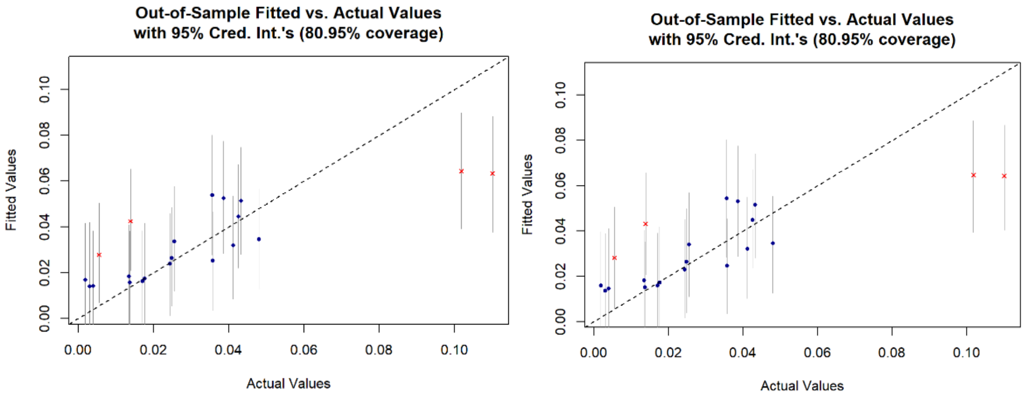

Fig.3 Effect comparison of training set before and after ten - fold cross validation

The color-labeled points and red forks in the above diagram are point estimates, and the line segments are interval estimates. If the interval estimation does not exceed the true value of the independent variable corresponding to the point, the prediction is not accurate, so the red cross is presented in the graph ; if the interval estimates the true value, it indicates that the prediction is accurate, so it appears as a blue point in the figure. It is not difficult to see from the figure that the accuracy of the BART model constructed by using the optimal hyperparameter combination is 80.95 % in the 95 % confidence interval. The model training fitting effect is good, but there is the possibility of over-fitting.

4.2.3.2 Comparison of Test Set Effects Before and After Ten - fold Cross Validation

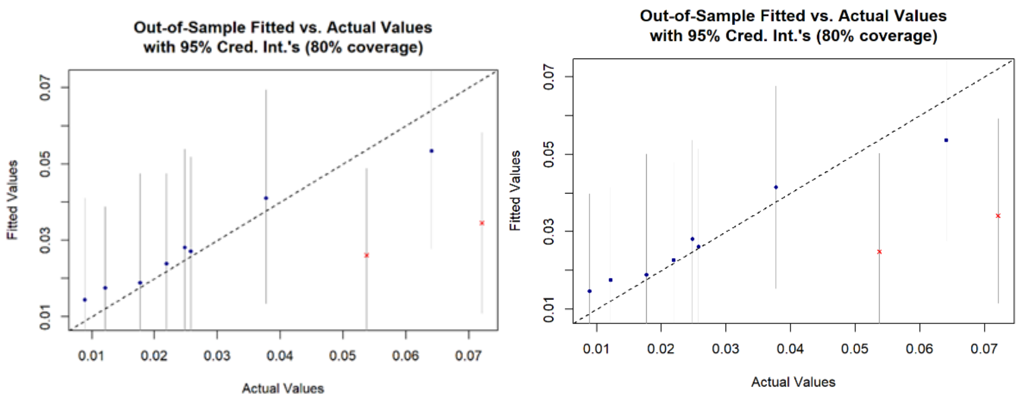

Fig. 4 Comparison of validation set effects before and after cross validation

For the fitting of the test set, it can be seen from the graph that the difference between the test set before and after the ten-fold cross-validation is small, the difference between the internal data of the training set is small, the model is relatively stable, and the accuracy rate is 80 % under the confidence level of 0.05. The model training fitting effect is good, and the comprehensive analysis of the effect before and after the ten-fold cross-validation of the test set shows that the model is not over-fitted.

4.2.4 BART model effects

The model iteration is shown as a graph of the convergence and characteristics of the model used in this study. In the figure, the samples are discarded before the vertical line. By using the tool to evaluate the convergence of MCMC algorithm, it can be seen that the posterior variance estimation is stable with the increase of the number of sampling samples. Through the number of blue lines presented and the average number of nodes per tree in the model and the average tree depth, the blue line is relatively stable, so the number of nodes tends to be stable when the sampling sample increases, and the average tree depth tends to be stable.

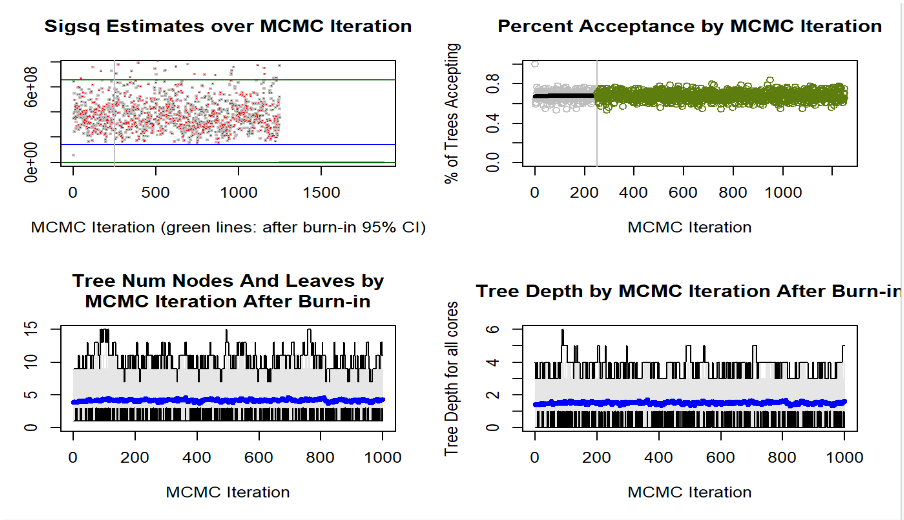

Figure 5 BART model iterative change diagram

4.2.4.1 variable importance

The model is constructed by selecting 50,100 and 150 trees respectively. By calculating the number of occurrences of each variable in the tree, the proportion of BART model variables included in the tree is obtained. The following figure is the importance of each variable, reflecting the degree of influence of each variable on GDP : depth of use > digitization > payment > coverage > credit > insurance.

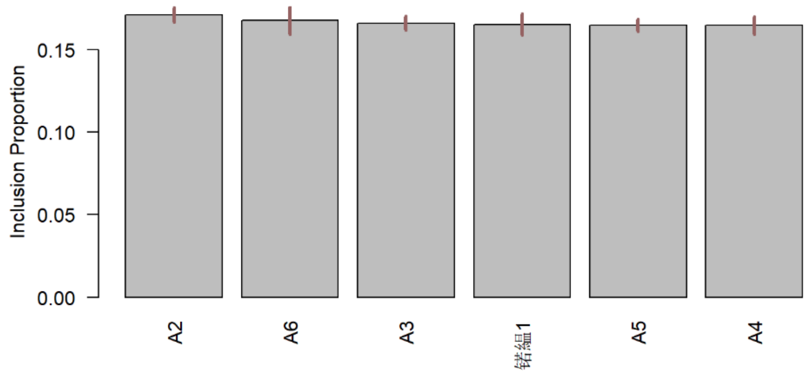

Figure 6 Proportion of variables contained in the tree

4.2.4.2 contribution analysis

The proportion of variables included in the regression tree is the importance of variables, which reflects the contribution of each variable to the BART regression model, so it reflects the degree of influence of each index on GDP. The results are as follows.

Table 5 Variable contribution table

variable	contribution
Using Depth Index	17.12%
Digitization degree index	16.78%
Payment index	16.60%
Coverage breadth index	16.53%
credit index	16.49%
insurance index	16.48%

According to the contribution of each index in the table, the impact of digital inclusive financial indicators in different aspects on the provincial economy is relatively average. Through the Bayesian cumulative tree model, the mechanism of digital inclusive finance affecting regional economic benefits is analyzed. It is found that the mechanism is mainly the depth of use, and the contribution rate is 17.12 %. The degree of digitization has a significant impact on regional economic benefits, with a contribution rate of 16.78 %. The payment index below the depth of use has a significant impact on regional economic benefits.

References

[1] Guan Fei, Gao Xing, Li Maishou. Digital Inclusive Finance, Innovation Driven and Regional Green Economic Growth [J].Credit, 2022,40 (08) : 63-74.

[2] Zhu Hongyu, Meng Defeng. Research on the Impact of Digital Inclusive Finance on Urban Residents ' Consumption [J]. Manager, 2021, (08) : 80-81

[3] Wu Yue. Research on the impact of digital inclusive finance on provincial economic development [D].Instructor : Long Haiming ; chen Bo. Hunan University, 2021.

[4] Tang Yichao, Li Dan, Zhang Dandan, Zhang Li, Pan Dongdong. Analysis of China 's Regional Digitization Level Index and Its Influencing Factors on GDP - Based on BP Neural Network Evaluation Model and BART Regression Model [A]. Chinese Society of Statistical Education, Teaching Steering Committee of Higher Education Institutions of Ministry of Education, National Steering Committee of Postgraduate Education of Professional Degree in Applied Statistics. Collection of Award-winning Papers of National Undergraduate Statistical Modeling Contest in 2021 (7th) (1) [C]. Chinese Society of Statistical Education, Teaching Steering Committee of Higher Education Institutions of Ministry of Education, National Steering Committee of Postgraduate Education of Professional Degree in Applied Statistics : Chinese Society of Statistical Education, 2021 : 675-726

[5] Tao Renjie, Chen Wenxing, Wang Chengwei, Hong Xingjian, Chen Xiongqiang. Spatial-temporal evolution of digital inclusive finance development and its impact on rural revitalization [A]. Chinese Society of Statistical Education, Teaching Steering Committee of Statistics Majors of Ministry of Education, National Steering Committee of Postgraduate Education of Professional Degree in Applied Statistics. 2021