Correlation Analysis between Regional Online Environmental Information Attention and Local Scale Green Finance

Xinduo Xu *

School of Literature and Science, China University of Petroleum-Beijing at Karamay, Karamay, 834000, China

* Corresponding Author Email: xuxiaoduo@st.cupk.edu.cn

Abstract. In today's world, with the advancement of technology, the living standard of mankind has ushered in a huge leap. However, problems such as environmental pollution are also increasingly apparent, which has attracted great attention from China and other countries. In this context, the research on green finance is particularly important. Combining with the situation that the public expresses their views widely through the Internet in the network era, this paper further explores the correlation between the attention of environmental-related information in the regional network and the local scale of green finance, and establishes the relevant specific model. This paper mainly studies the correlation between the attention of environmental-related information in the regional network and the scale of local scale of green finance through the Spearman correlation coefficient, and explores the trend of the correlation over time. This paper found that there is a significant positive correlation between the two, and this correlation has an increasing trend year by year. Finally, this paper uses a linear regression model and a BP neural network to complete the related modeling. This study shows that the formulation of green finance related policies can reasonably guide people to join in the construction of green finance. Furthermore, the model of this study can also be used to quantitatively predict the impact of residents' participation on the development of green finance to a certain extent, so as to promote the progress of green finance.

Keywords: green finance, public attention, Spearman correlation coefficient, linear regression model, BP neural network.

1. Introduction

With the rapid development of China, people's pursuit of high-quality life is increasing, and high-quality environmental quality is the premise of high-quality life. Therefore, the development of green finance (GF) has become a necessary option for China's future development. Therefore, it is of particular significance to explore the factors affecting the development of GF and establish a specific model between the relevant factors and the development of GF. In addition, with the development of the Internet, the population of Internet users in China has experienced rapid growth. The Internet has become an important way for people to discuss social hot issues. Relevant articles by Shi, Xie, and other scholars have pointed out that public attention has a certain impact on stock returns, green investment, and other aspects [1, 2].

The existing research has a more in-depth discussion on the influence factor of GF, covering a comprehensive range of factors. For example, Yan and Xia explored the impact of economic development potential, economic development level, air quality and other factors on GF based on a new grey relational degree model, and the influence degree of relevant factors is ranked [3]. In addition, the article of Xie and other scholars explored the impact of Regional GDP, regional innovation level and other factors on GF in the Yangtze River Delta region of China through the time fixed effects Durbin model, and concluded that Regional GDP, regional innovation level and air quality are the most important influencing factors [4]. However, most of the above studies focus on economic, environmental, and other factors, while there are few studies on public attention, an important influencing factor. In the relevant research of Liu and Yuan, it is also mentioned that the existing literature rarely involves the influence of public environmental concerns at the macro level [5].

In today's Internet era, with the rapid dissemination of information, the influence of public opinion will become more and more important. In the existing research, Liu and other scholars have studied the impact of environmental concerns on the progress of GF [5]. By establishing a regression model, they concluded that public environmental concern has a significant positive impact on the progress of GF [5]. In addition, they verified the robustness of the conclusion through a series of methods, and explored the relevant influence mechanism [5]. Xie and Ying revealed the significant positive correlation between public attention and green investment by regression analysis and incorporating a number of mechanism variable and control variables [2]. At the same time, they tried to further optimize the results by using two-stage least squares, moderating effect analysis, and other methods [2]. In terms of modeling, Qi and other scholars creatively put forward the application of multi-objective decision-making modeling under the dual carbon goal, and also pointed out the current situation of insufficient research on modeling in the dual carbon field [6].

Based on previous studies, this research collects the search volume of environment related keywords through Baidu index, and uses the collected search volume as an indicator of the attention of environment related information in the regional network. At the same time, this study uses the overall score of China's local green finance in 2021, 2022 and 2023 in the China's local green finance database of the International Institute of Green Finance (IIGF) at the Central University of Finance and Economics (CUFE) as an indicator of the scale of GF [7, 8]. In terms of research methods, this study mainly uses the Spearman correlation coefficient (SCC) to study the correlation between the attention of environmental-related information in the regional network and the local scale of GF, and uses the linear regression model (LRM) and BP neural network (BPNN) to establish the correlation model. At the same time, this study also uses the annual calculation of the SCC to observe the change trend of the correlation between the two over time. So as to achieve the goal of further exploring the correlation between the attention of environmental related information in the regional network and the local scale of GF on the basis of existing research and establishing relevant specific models.

This study can be used as the basis for the formulation of GF related policies. Relevant policies can guide people to join in the construction of GF. Further, the modeling of the attention of environmental related information in the regional network and the local scale of GF can also guide the policy guidance, and then make the GF market flourish.

2. Research Methods

2.1. Data Source and Indicator Description

The theme of this research is the correlation research and modeling analysis between the environmental information attention in the regional network and the local scale of GF.

In the data set of this study, the classification variable is the name of China's provinces (31 provinces, autonomous regions, and municipalities directly under the central government, excluding Hong Kong, Macao, and Taiwan, the same below). The independent variable and dependent variable are the network attention index (NAI) (2021, 2022, 2023) and the green financial scale index (GFSI) (2021, 2022, 2023). The two are used to quantify the attention of environmentally related information in regional networks and the scale of local scale of GF.

For the NAI, this study uses the method of searching keywords through Baidu Index adopted by Xie, Liu and other scholars in relevant research for reference, and uses Baidu Index to count the search volume of relevant keywords to complete data collection [2, 5]. In order to prevent the uncertainty of this study caused by selecting only a single keyword, this study uses the method of selecting multiple keywords adopted by Xie and Ying in relevant studies for reference [2]. This paper selects two environment related keywords, environmental pollution and global warming, in terms of environmental crisis, and two environment related keywords, environmental protection and low carbon, in terms of environmental maintenance. This paper uses Baidu Index to count the search volume of these four keywords in each province from 2021 to 2023, and adds up the search volume of these four keywords as NAI (2021, 2022, 2023). At the same time, taking into account the influence

of individualized attention on the scale of GF will not change with the change of the population base of the region, this study only counts the search volume of relevant keywords without considering the population base of the region.

For the GFSI, to guarantee the precision of the data, this paper selected the China local green finance database of the IIGF, CUFE as the data source [7,8]. The overall evaluation of China's local green finance in 2021, 2022 and 2023 in the China's local green finance database of the IIGF, CUFE is taken as the GFSI of each province in 2021-2023 [7,8].

Table 1 lists the basic data distribution of various indicators.

Table 1. Basic data distribution of indicators

name	sample size	minimum value	maximum value	average value	standard deviation	median
GFSI(2023)	31	9.560	57.050	27.610	12.440	25.130
NAI(2023)	31	44.000	685.000	391.871	154.656	403.000
GFSI(2022)	31	9.010	62.400	29.059	13.517	24.470
NAI(2022)	31	55.000	789.000	446.452	168.381	455.000
GFSI(2021)	31	15.700	71.550	39.613	12.482	36.690
NAI(2021)	31	50.000	833.000	451.968	173.428	463.000

2.2. Method Introduction

SCC is a common method for determining the correlation, especially for the data with non normal distribution (in the subsequent normality test, there are non normal distribution variables in this study). For the sample with sample size n, the two variables are x_i and y_i (i=1,2,3...), and the calculation formula is:

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \quad (1)$$

In the relevant research of Jin and Li, it is pointed out that the SCC can be used for the grade data converted from continuous variable observation data [9]. The specific implementation method is given in SPSSPro [10]. That is, for continuous variables, the original data is assigned a corresponding level according to its average descending position in the overall data [10]. Therefore, the SCC is selected as the index to judge the correlation between the attention of environmental-related information in the regional network and the scale of local scale of GF. And the SCC values of the two were observed annually in this study to obtain the change trend of the correlation between the two from 2021 to 2023. Finally, in terms of modeling, combined with the relevant articles of Shi, Xie, and other scholars, it is pointed out that public attention has a certain impact on green investment, stock returns, and other aspects, this study selects the NAI as the independent variable and scale of GF as the dependent variable [1, 2]. This paper uses the most commonly used LRM to complete the modeling of the relationship between the environmental information attention in the regional network and the local scale of GF. At the same time, this paper considers that the direct influence factor of scale of GF is not only the attention of environmental-related information in the regional network, but also the influence of the attention of environmental-related information in the regional network on the local scale of GF may be more complex. Although the modeling process of the LRM is relatively successful, it does not mean that the two follow a strict linear relationship, and there may be a more complex relationship between them. A BPNN is a form of multi-layer feedforward neural network that employs the error backpropagation algorithm for training. The core principle of this algorithm entails the application of the gradient descent method to minimize the mean square error between the actual output values and the anticipated output values generated by the network. Notably, one of the key strengths of BPNN lies in their robust capability for nonlinear mapping. The most important thing is that the BPNN has a strong nonlinear mapping ability. Consequently, this study also utilizes a BPNN for modeling purposes.

3. Research Results and Discussion

3.1. SCC Correlation Analysis

3.1.1 Data processing and verification

To increase the sample size and ensure the accuracy of the research conclusion, the time index is ignored here, and the relevant data of 2021, 2022 and 2023 are combined (the data of different years in the same province are regarded as different samples) (if there are no special notes, the following analysis in this paper will use this method), and a total of 93 samples are obtained.

The following will explain why the SCC is applied here for analysis from two aspects. First, Fig. 1 shows the relation between the NAI and the GFSI, with the NAI (unit: times) as the abscissa and the GFSI (unit: minutes) as the ordinate.

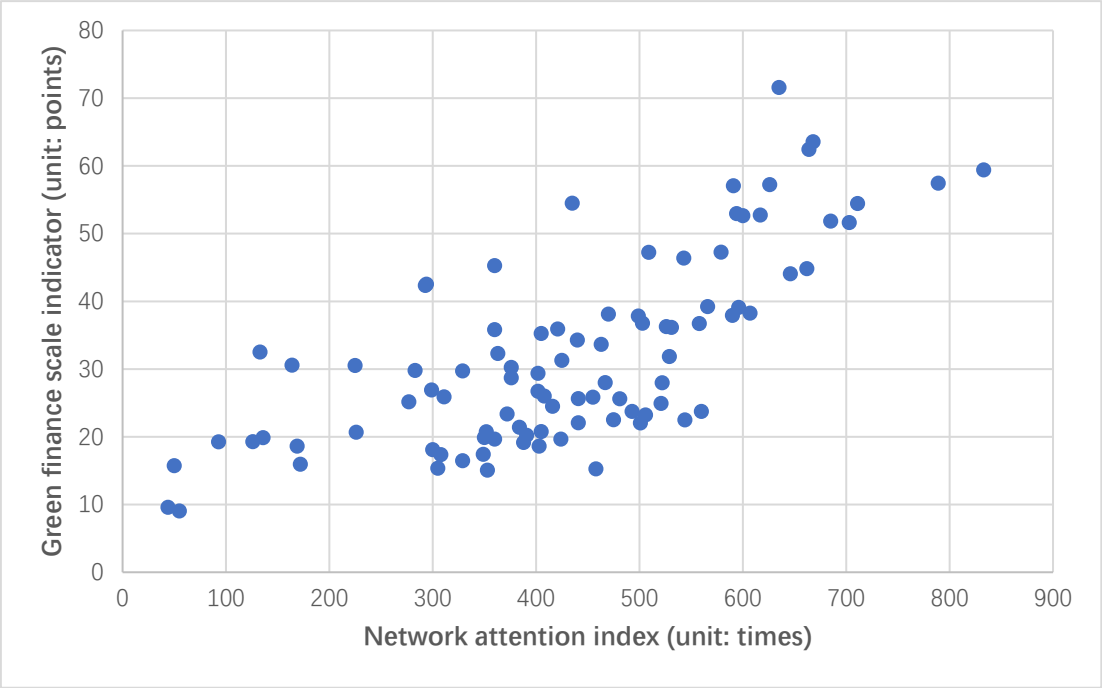

Fig 1. scatter diagram of NAI and GFSI of NAI and GF scale index (Photo/Picture credit: Original).

From Fig. 1, it can be inferred that as the NAI increases, there is a corresponding upward trend in the GFSI. This indicates that there may be some correlation between the two. Therefore, it is imperative to conduct a more in-depth investigation into the relationship between the two utilizing quantitative methods.

Then, a normality test is conducted for the NAI and GFSI, and the analysis results are shown in Table 2.

Table 2. Normality test results

name	GFSI	NAI
sample size	93	93
average value	32.094	430.097
standard deviation	13.776	166.121
skewness	0.751	-0.224
kurtosis	-0.162	-0.019
statistic D value	0.112	0.065
p	0.006**	0.423

* p<0.05 ** p<0.01

From Table. 2, it can be inferred that the sample sizes of the research data all exceed 50, therefore, the K-S test is employed. The GFSI showed a significant ($p<0.05$). This means that the GFSI does not have normal characteristics. NAI did not show a significant ($p>0.05$). This means that the NAI is normal. Therefore, the Pearson correlation coefficient cannot be used here.

To sum up, this research uses the SCC to conduct an analysis of the correlation between the NAI and the GFSI.

3.1.2 SCC analysis

Using the SCC, a correlation analysis was conducted between the NAI and the scale of GF from 2021 to 2023 (the data used here is the merged data ignoring time indicators). To further explore the relationship between the NAI and the GFSI of each province, this study conducted a SCC analysis on the NAI and the GFSI of each province on an annual basis. The analysis results are presented in Table 3.

Table 3. SCC operation results

Project \ year	2021-2023	2021	2022	2023
correlation coefficient	0.690**	0.666**	0.737**	0.771**
p value	0.000	0.000	0.000	0.000
sample size	93	31	31	31

* $p<0.05$ ** $p<0.01$

The data presented in Table 3 clearly illustrates that using the SCC for correlation analysis, after merging the relevant data from 2021 to 2023 without considering time indicators, the correlation coefficient between the scale of GF index and the NAI is 0.690, and it exhibits significance at the 0.01 level. This suggests that there exists a notable positive correlation between the scale of GF index and the NAI.

Meanwhile, the SCCs between the GFSI and the NAI for each year from 2021 to 2023 were 0.666, 0.737 and 0.771, respectively, and all exhibited significance at the 0.01 level.

SCC between the NAI and the GFSI from 2021 to 2023 is shown in Fig. 2.

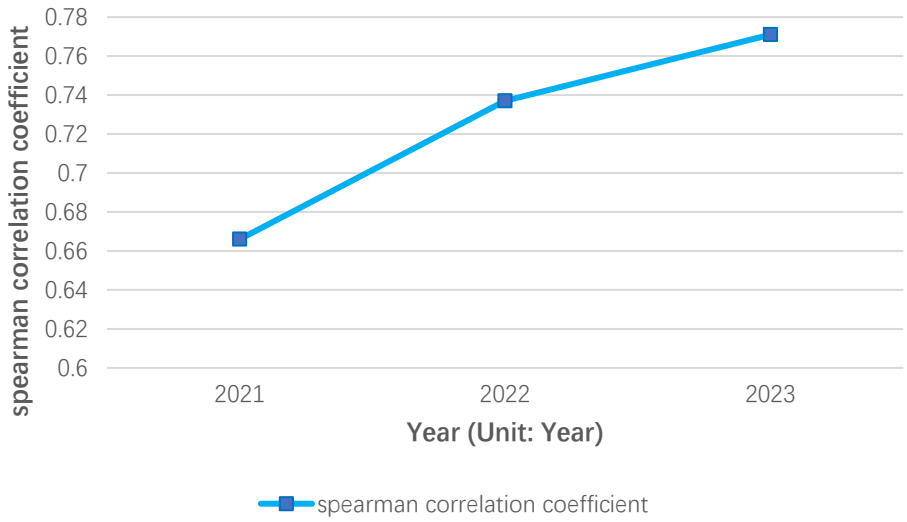

Fig 2. Line chart of SCC between NAI and GF scale index from 2021 to 2023 (Photo/Picture credit: Original).

From Fig. 2 and Table 3, it can be concluded that from 2021 to 2023, the correlation between the attention of environmental-related information in regional networks and the scale of local scale of GF has increased year by year.

3.2. LRM

3.2.1 Model establishment

From Fig. 1, it can be approximately observed that there is a certain linear correlation between the NAI and the GFSI. Therefore, the LRM is used for modeling in this study.

The NAI is used as the independent variable, and the GFSI is used as the dependent variable to model with a LRM. The results are presented in Table 4.

Table 4. LRM results (n=93)

project	*B*	standard error	*Beta*	*t*	*p*	VIF	tolerance
constant	6.593	2.800	-	2.354	0.021*	-	-
NAI	0.059	0.006	0.715	9.756	0.000**	1.000	1.000
R^2				0.511			
Adjusted R^2				0.506			
F				*F* (1, 91)=95.173, *p*=0.000			
D-W value				1.028			

Note: dependent variable=GFSI
* $p<0.05$ ** $p<0.01$

From Table 4, it can be concluded that the linear regression analysis is carried out with the NAI as the independent variable and the GFSI as the dependent variable, the model formula is:

$$\text{GFSI} = 6.593 + 0.059 \times \text{NAI} \tag{2}$$

At the same time, the R^2 value of the model is 0.511, indicating that the NAI can explain the 51.1% change of the GFSI to a certain extent; The model passed the F-test (f=95.173, p=0.000<0.05) when the F-test was conducted on the model, indicating that the NAI must have an impact on the GFSI; The regression coefficient for the NAI is 0.059 (t = 9.756, p < 0.01), indicating a significant positive impact of the NAI on the GFSI. These findings further reinforce the conclusion that there exists a significant positive correlation between the GFSI and the NAI.

3.2.2 Model evaluation

Table 5. ANOVA table (intermediate process)

project	sum of squares	*df*	mean square	*F*	*p value*
regression	8925.551	1	8925.551	95.173	0.000
residual	8534.209	91	93.783		
total	17459.761	92			

From Table 5, it can be concluded that the model passed the F test (f=95.173, p=0.000<0.05) during the F test on the model, indicating that the model construction is meaningful.

A normality test is conducted for the residuals of the above LRM. The findings are presented in Table 6.

Table 6. Analysis results of the normality test

name	sample size	average value	standard deviation	skewness	kurtosis	statistic D value	*p*
Residual of the LRM	93	-0.000	9.631	0.468	-0.167	0.071	0.295

* $p<0.05$ ** $p<0.01$

Considering that the sample size of the research data exceeds 50, the K-S test is utilized. Linear regression residuals are not significant (p>0.05), which means that the original hypothesis (original

hypothesis: normal distribution of data) has been validated, that is, all linear regression residuals are normal.

This shows that the LRM is reasonable. However, as pointed out in the model-building part, the NAI can explain the 51.1% change of the GFSI to a certain extent. This shows that the NAI is not the only factor affecting the GFSI of each province. In fact, many factors such as policy, economy, and environment will have an impact on the GFSI of each province. This is also reflected in the research of Yan, Xie, and other scholars [3,4]. However, the focus of this study is to reveal the correlation between the attention of environmental-related information in regional networks and the scale of local scale of GF. Therefore, this paper does not conduct in-depth research on other influencing factors. Therefore, the LRM established in this paper may not be completely accurate, but this does not mean that the LRM established in this paper is completely meaningless. The LRM established in this paper is established without considering other influencing factors. This model has certain significance in the rough study of the impact of environmental-related information attention on the local scale of GF in the regional network.

3.3. BPNN

3.3.1 model selection and parameter display

Considering the potential limitations of the LRM established previously, which may overlook numerous complex factors, and recognizing the strong mapping capability of the BPNN, this study employs a BPNN for modeling purposes. In this approach, the NAI serves as the independent variable while the GFSI is designated as the dependent variable.

To accommodate the characteristics of the small sample size of the samples, the parameters of the BPNN are set as shown in Table 7.

Table 7. Parameter name

parameter name	parameter value
data segmentation	0.8
data shuffle	yes
cross validation	no
activation function	identity
solver	lbfgs
learning rate	0.1
L2 regular term	1
lterations	1000
number of hidden layer 1 neurons	10
number of hidden layer 2 neurons	10
number of hidden layer 3 neurons	10
number of hidden layer 4 neurons	10

3.3.2 Model evaluation and prediction results display (test set)

The model evaluation of the BPNN is shown in Table 8.

Table 8. Model evaluation

project	MSE	RMSE	MAE	MAPE	R^2
training set	92.248	9.605	7.779	26.828	0.497
test set	91.939	9.588	7.116	22.042	0.522

3.3.3 Model evaluation and prediction results display (test set)

From the model's running results, it can be observed that the performance of the BPNN regression model on both the training set and test set is relatively comparable. This observation suggests that the model possesses a certain degree of generalization ability and does not exhibit any significant signs of overfitting. The difference between the mean square error (MSE=92.248) of the training set and

the test set (MSE=91.939) was only 0.3%, and the root mean square error (RMSE) was 9.605 and 9.588, respectively, which further verified the stability of the model. The average absolute error (MAE) in the test set (7.116) is better than that in the training set (7.779), and the average absolute percentage error (MAPE=22.042%) in the test set is significantly lower than that in the training set (26.828%), indicating that the prediction bias of the model for new data is relatively smaller and the relative error control is better. The decisive coefficient R^2 in the training set (0.497) and the test set (0.522) is at the level of medium explanatory power, indicating that the model can explain about 50% of the variation of the target variable.

In general, the model has basic prediction efficiency.

4. Research Limitations and Prospects

As mentioned above, this study does not consider the possible impact of other factors on the scale of GF. In the future, based on the articles of Yan, Xie, and other scholars, we can further combine the attention of environmental-related information in regional networks with other influencing factors, analyze their impact on the local scale of GF, and analyze the contribution of the attention of environmental-related information in regional networks to this impact [3, 4].

Moreover, the BPNN model used in this study only has basic prediction ability, and the prediction effect is not very good. Combined with related articles of Rahul Saini, the BPNN needs a large amount of data [11]. Although this paper integrates the data from three years, the sample size is still small. In the future, it can be refined on the basis of provinces, and the data at the levels of cities and towns can be further analyzed. It can also be refined or broadened in the dimensions of time, and more accurate results can be obtained through the analysis of a BPNN.

At the same time, this paper notes that the positive correlation between the GFSI and the NAI has an upward trend from 2021 to 2023. In the follow-up, we can further explore the reasons for the enhancement of the positive correlation between the two, and further clarify the impact mechanism of the NAI on the GFSI.

5. Conclusion

Firstly, through the SCC, this paper concludes that there is a significant positive correlation between the GFSI and the NAI. At the same time, combined with the R^2 value of linear regression analysis, it can be seen that the NAI can explain the 51.1% change of the GFSI. Secondly, the change trend of SCC between the GFSI and NAI from 2021 to 2023. The positive correlation between the two is increasing year by year from 2021 to 2023. Finally, a LRM can be established between the GFSI and the NAI. The established model is: $\text{GFSI} = 6.593 + 0.059 \times \text{NAI}$. The model can pass the F test, and the residual distribution also conforms to the normal distribution, indicating that the model has certain reference value. In addition, this paper also establishes the BPNN model for both, and the BPNN model has certain modeling significance. On the basis of existing research, this study further clarified that there is a significant positive correlation between the GFSI and the NAI. Accordingly, when formulating policies related to GF, it can reasonably guide people to participate in the construction of GF, so as to promote the development of GF. At the same time, the model established in this paper can be used to estimate the specific role of guiding measures in promoting the development of GF. This will help policymakers balance the cost and benefit of guidance, to maximize benefits. In addition, this paper also has the limitations of incomplete consideration and a small amount of data. In the future, further research can be carried out by considering various factors and expanding the data set. At the same time, this paper notes that the positive correlation between the GFSI and the NAI has an upward trend from 2021 to 2023. In the future, relevant research can also be carried out to explore the reasons for the enhanced correlation between the two and the specific impact mechanism between them.

References

[1] Shi R., Wang Q., Wang J. How public attention affects stock returns: Based on the responses of Chinese listed tourism companies during COVID-19. Current Issues in Tourism, 2022, 26(21): 3411–3416.

[2] Xie Y., Ying Q. Unpacking the impact of public attention on green investment: Insights from climate sentiment on social media. International Review of Financial Analysis, 2025, 104: 104325.

[3] Yan S., Xia L. A multidimensional grey relational model considering reverse indicators and its application to the influencing factors of green finance in China. Grey Systems Theory and Application, 2023, 13(3): 558–575.

[4] Xie H., Ouyang Z., Choi Y. Characteristics and influencing factors of green finance development in the Yangtze River Delta of China: Analysis based on the spatial Durbin model. Sustainability, 2020, 12(22): 9753.

[5] Liu X., Yuan Q. Does public environmental concern promote the development of green finance? Empirical evidence from 284 cities in China. Journal of Audit & Economics, 2024, 39(3): 107–116.

[6] Qi Y., Huang J., Qi Z. Literature review on the modeling of dual carbon strategy and prospect of modeling based on multi-objective decision making. China Soft Science, 2022, (S1): 204–216.

[7] Wang Y., Ren Y., et al. Annual report of the development of local green finance in China (2023). Beijing: Social Sciences Academic Press (China), 2023: 2–3.

[8] Wang Y., Ren Y., Wu Q., et al. Annual report of the development of local green finance in China (2024). Beijing: Social Sciences Academic Press (China), 2025: 2–3.

[9] Jin L., Li Y. Discrimination of several correlation coefficient and their implementation in R language. Journal of Statistics and Information, 2019, 34(4): 3–11.

[10] SPSSPRO Help Center. Correlation analysis [EB/OL]. 2021. https://www.spsspro.com/help/correlation/#_8%E3%80%81%E6%A8%A1%E5%9E%8B%E7%90%86%E8%AE%BA

[11] Saini R. A review on artificial neural networks for structural analysis. Journal of Vibration Engineering & Technologies, 2025, 13(2).