J. COMBIN. MATH. COMBIN. COMPUT. 127a (2025) 271–290

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

A Data Mining Approach to Green Financial Product Innovation and Assessment of Environmental and Economic Benefits

Yangyulong Wu[1, ✉]

[1] *School of Business and Management, The Hong Kong University of Science and Technology, 999077, Hong Kong*

ABSTRACT

In the current period, green finance has become an inevitable trend in the development of the financial industry. The study collects the audience demand of green financial products through Octopus collector, uses micro-word cloud analysis system for data de-weighting and Chinese word segmentation, and calculates the keyword weights in the words using TF-IDF algorithm, and realizes the identification of green financial product innovation and economic benefits by combining with multi-dimensional innovation map. Subsequently, the indicators on green financial product innovation and environmental economy from 2007 to 2022 are combed, and a VAR-based econometric model is established to analyze the impact relationship between green financial product innovation and environmental economic benefits. The results show that when the lag period is 10 periods, the contribution of environmental economic benefit itself to the change of environmental economic benefit tends to be 23.79%, while the contribution of green investment products, green bond products and green insurance products to environmental economic benefit tends to be 9.72%, 20.23% and 21.83%, respectively. Green product innovation has a certain influence on the fluctuation of environmental economic benefits, and green bond products and green insurance products have a greater impact on environmental economic benefits.

Keywords: TF-IDF algorithm, multidimensional innovation map, VAR; econometric model, green financial products

1. Introduction

Green finance is an economic activity to support environmental improvement, climate change response and resource saving and efficient utilization [1-3]. Under the tight constraints of resources and environment, the development of green finance has become an important force to support the development of economic green transformation. The scale of green finance in China is growing rapidly, but the products are still

✉ Corresponding author.
E-mail address: 18006479651@163.com (Y. Wu).
Received 12 February 2024; Revised 09 June 2024; Accepted 10 August 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-017

relatively single and not innovative enough, and the degree of fit between finance and green is relatively low [4-5]. Factors such as imperfect laws and regulations and environmental economic benefit system, inadequate green financial support policies, weak innovation consciousness and lack of high-end talents are the main factors restricting the innovation of green financial products [6-8].

"Green water and green mountains are golden silver mountains", protecting the ecological environment and promoting the green development of the economy have become the main tone of China's economic development at present [9]. Green finance guides the allocation of funds and other resources to the environmental protection industry, innovates green financial products, and promotes the development of the environmental protection industry, which not only requires technical professionals, but also requires strong government support for the environmental protection industry [10-12]. It is necessary to improve the green financial laws and regulations, introduce policies to promote the innovation of green financial products, create the market atmosphere of green financial product innovation, and vigorously cultivate green financial innovation talents [13-14]. However, green finance is a quasi-public product, and the environmental protection industry it supports bears a great social responsibility, which not only requires high investment, but also faces a lot of risks such as scientific and technological risks, operational risks and so on [15-16]. Through data mining innovation and design of appropriate green financial products, to meet the financing needs of various types of market players, dispersing financial risks, and improving the environmental economic benefits is an important task facing the current green finance [17-19].

After obtaining the general process of data mining (three stages of data collection, data processing, and data analysis) and the application technology routes for realizing each stage of data mining, the study takes green financial products as the research object to carry out the innovation of data mining audience demand. The evaluation data of the audience on green financial products are collected by Octopus collector, the collected data are de-emphasized and Chinese lexicalized by using micro-word cloud analysis system, and the TF-IDF algorithm is utilized to calculate the weight share of the keywords in the words, and then the multi-dimensional innovation map based on dimensional method coupling is constructed to identify the environmental and economic benefits of the innovations of the green financial products. On this basis, by combing the indicators on green financial product innovation and environmental economic benefits from 2007 to 2022, after passing the smoothness test, Johansen cointegration test and Glenn causality test of the variables, a VAR model is established to analyze the impact relationship between the two with impulse response and ANOVA.

2. Green financial product innovation data mining

2.1. Green financial product innovation data mining process

The process of data mining mainly includes the steps of data collection, data cleaning, data classification, data clustering and data visualization, but in this paper, in order to combine and analyze the data mining with the economic efficiency research, it will introduce the data mining process from the three steps of data collection, processing and analysis.

2.1.1. Data acquisition. Data collection is mainly realized through the technology of web crawler, which is also known as web spider, is a program that continuously obtains various web contents from Internet pages through hyperlinks and document retrieval methods, and it is divided into general-purpose, theme-type and in-depth web crawler according to the different implementation technologies and system structures of crawling, and the three kinds of core technologies of web crawler are similar, and the architectures include crawler scheduling end, web crawler main program and three parts of target data. The three types of web crawlers have similar core technologies, and their architectures include crawler scheduler, web crawler main program and target data, among which the web crawler main program is completed by the collaboration of three modules: URL manager, web page downloader, and web page parser. The web crawler type and structure are shown in Figure 1. Taking the theme-based web crawler technology as an example, the basic steps of crawling data are: firstly, determine the keywords or themes of the crawling program. Secondly, the crawler program continuously downloads web pages on the network and obtains the content, then on the one hand, according to the analysis of the content to find links with relevance, in the sequence to be crawled according to the priority of the unvisited links, and the formation of a cycle of repetition, on the other hand, to calculate the relevance of the content of the page and the theme of crawling, and will meet the requirements of the high relevance of the page is deposited in the database. Finally, the database of accessed pages is judged to determine whether the termination conditions are met, and if the conditions are met, the crawling project is ended.

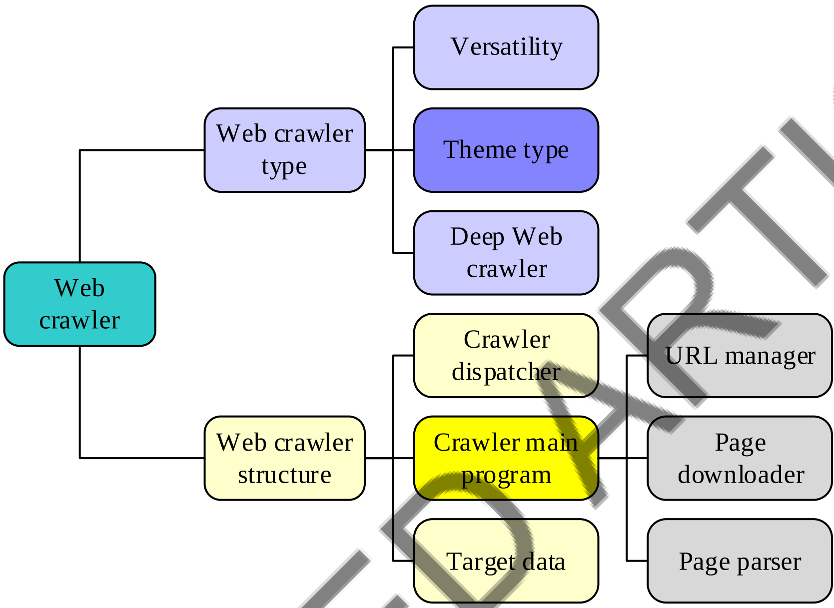

Fig. 1. Web-crawler type and structure

2.1.2. Data processing. In this step of data processing, the main data de-emphasis, de-deactivation of words to get an effective database, and on this basis, for the cleaned data for text segmentation and lexical labeling work.

1) Chinese word separation technology: Chinese word division is the next text classification and analysis of the foundation, there are many fully developed Chinese word division software, such as stuttering word division, Sogou word division, JAVA word division toolkit, Chinese Academy of Sciences Computer Research Institute of Chinese lexical analysis system, micro-word cloud Chinese word division and so on.

The Micro Word Cloud analysis system is powerful, and the flow of data processing using Micro Word Cloud analysis system is shown in Figure 2. The collected comments are imported into the Micro Word Cloud Analysis System in txt format, in which a number of tasks such as Chinese word segmentation, de-duplication, lexical annotation and word frequency statistics can be completed, and the processing time is rapid and the results are accurate.

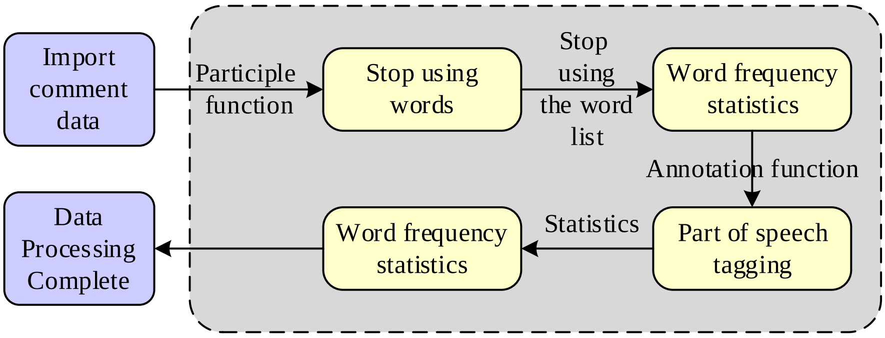

Fig. 2. Micro-word cloud processing data process

2) Stop word technology: Stop words refer to some words that appear frequently but have no practical meaning in the results of the comment sentence segmentation step after the Chinese word segmentation step, and common stop words generally include transition words without any actual meaning, mood particles, adverbs, prepositions, conjunctions and useless punctuation marks, such as "la" "and" "one" "but""! ""@", "%", etc. In the process of removing deactivation words, it is often necessary to complete the steps of deactivating words with the help of a deactivation glossary, which can help filter online review data to improve the effect of deactivating deactivation words.

2.1.3. Data analysis. TF-IDF as a statistical method can be applied in extracting keywords, that is to say, it can be ranked in order according to the magnitude of the TF-IDF value, which is utilized to assess the importance of a certain word to the documents in a document set or corpus, and to distinguish the words with a high weight and frequency in the online comments in order to determine the keywords in the text of the comments [20].

In the TF-IDF method, the phrase TF represents word frequency, which expresses the specific meaning of the frequency of a word in a document, in order to prevent the negative effects brought about by too many words, it is usually necessary to first normalize and standardize the words to obtain the word frequency is calculated as follows: the number of times that a word occurs in a document is divided by the number of total words used in the document, the formula is shown in (1). (1) shown in the formula, the numerator refers to the number of occurrences of a keyword in the document under study, and the denominator refers to the total number of words in the document under study.

Inverse Document Frequency (IDF) is studied in the context of a corpus that represents the number of times a complete word occurs in all the documents, which is used to correct the importance of the word summarized in TF. The formula for calculating the inverse document frequency is shown in (2), where IDF_i represents the inverse document frequency, $|D|$ represents the total number of documents in the corpus, and $\left|\{j : t_1 \in d_j\}\right| + 1$ represents the number of documents containing the word, and 1 is added to the denominator to prevent the case where the number of documents containing the word in the denominator is 0, i.e., the word is not present in all the documents. From the formula can be seen when a word in the document library more common, the more often used, then the denominator is larger, IDF_i value is getting smaller and smaller closer and closer to 0, that is, when a word at the same time exists in a number of research documents, the value of the frequency of the reverse document will be relatively small. Then:

$$TF_{i,j} = \frac{n_{i,j}}{\sum_{k}^{1} n_{k,j}} \tag{1}$$

$$IDF_i = \log \frac{|D|}{\left|\{j : t_i \in d_j\}\right| + 1} \tag{2}$$

Finally, the TF-IDF value can be obtained by multiplying TF and IDF with the following mathematical formula:

$$TFIDF_{i,j} = TF_{I,J} \times IDF_i \tag{3}$$

According to the formula, it can be seen that the value of TF-IDF is directly proportional to the frequency of use of a word in a document, and inversely proportional to the number of times the word appears in the entire document library, using the TF-IDF value can be simple and fast keyword extraction of high-frequency words in the database of users' online comments, and it can efficiently cluster high-frequency words with a higher weight in the online comments of the users, so that the database of useful of user needs become clearer, and the mapping between user needs and product features can be better accomplished.

2.2. *Identification of green financial product innovation opportunities based on DM*

2.2.1. Multidimensional innovation map construction based on dimensional method coupling. The multidimensional innovation map based on data mining is shown in Table 1. In this methodological model, nine innovation dimensions are defined as W={W1, W2, W3, W4, W5, W6, W7, W8, W9}, which correspond to the structural dimension, spatial dimension, environmental dimension, functional dimension, mechanism dimension, material dimension, temporal dimension, dynamical system dimension, and human-computer relationship dimension, respectively [21]. The nine innovation laws are defined as F={F1, F2, F3, F4, F5, F6, F7, F8, F9}, which correspond to the laws of combination and integration, disassembly and removal, local optimization, substitution, dynamization, amicability, self-service, flexibilization, and intelligence, respectively. Coupling a single or multiple innovation dimensions with a single or multiple innovation laws, such a combination can be formed in such a way that an innovation scheme Q_i (i=1,2,3 ......,n) can be formed and thus represented, and the innovation opportunities can be identified from the innovation scheme.

Table 1. Multi-dimensional innovation map based on data mining

	W1	W2	W3	W4	W5	W6	W7	W8	W9
F1									
F2		Q1							
F3							Q2		
F4				Q3					
F5									
F6									Q5
F7			Q4						
F8							Q6		
F9									

2.2.2. Description of Green Finance Product Innovation Opportunities. Product innovation opportunity identification mainly takes online reviews as the research object and utilizes the innovation elements extracted from product-based online reviews to construct a multidimensional innovation map. Product innovation opportunities can be recognized from the multidimensional innovation map, reflecting the usefulness of user needs for product innovation. According to the constructed multidimensional technology innovation map, different product innovation programs can be seen in the figure.

Assuming that the product innovation opportunity is N, the process of coupling innovation dimensions and innovation laws can be visualized from this map, expressed as:

$$N = W_i ※ F_j \tag{4}$$

In Eq. (4), W_i represents the innovation dimension of Green financial products, i is a set, $i \subseteq \{1,2,3,4,5\}$. F_j represents the innovation law of cell phone products, j is also a set, and $j \subseteq \{1,2,3,4,5,6\}$. ※ represents the coupling. The elements in i can take the value of at most 5, and the elements in j can take the value of at most 6. Assuming that a elements are taken from i, the number of subsets after removing the null set in set i is $2^a - 1$, each of which represents one or more fixed combinations of Green financial products innovation dimensions. Taking b elements from j, the number of subsets in set j after removing the null set is $2^b - 1$, and each subset represents one or more fixed combinations of Green financial products innovation laws, so the number of product innovation paths that can be obtained through dimensional law coupling is:

$$C_N = C_5^a\left(2^a - 1\right) \times C_6^b\left(2^b - 1\right) \tag{5}$$

For example, when $a = 2$, $b = 2$ and determine the innovation dimensions as W_1 and W_2 and the innovation laws as F_1 and F_3, i.e., $a = \{\{1\},\{2\},\{1,2\}\}$, $b = \{\{1\},\{3\},\{1,3\}\}$, the number of innovation paths can be obtained as $C_N = \left(2^2 - 1\right) \times \left(2^2 - 1\right) = 9$, and the expression for access to innovation opportunities can be written as:

$$\begin{aligned} W_i ※ F_j &= W_{\{\{1\},\{2\},\{1,2\}\}} ※ F_{\{\{1\},\{3\},\{1,3\}\}} = \begin{bmatrix} W_1 \\ W_2 \\ W_1W_2 \end{bmatrix} ※ \begin{bmatrix} F_1 \\ F_3 \\ F_1F_3 \end{bmatrix} \\ &= \begin{bmatrix} W_1F_1 & W_1F_3 & W_1F_1F_3 \\ W_2F_1 & W_2F_3 & W_2F_1F_3 \\ W_1W_2F_1 & W_1W_2F_3 & W_1W_2F_1F_3 \end{bmatrix} \end{aligned} \tag{6}$$

Online Review Product innovation opportunity identification mainly includes product innovation opportunity description, product innovation opportunity identification, and product innovation opportunity evaluation. It can be further understood that product innovation opportunity description is a necessary step to carry out innovation opportunity identification, and product innovation opportunity description refers to the fact that according to the constructed multidimensional technology innovation map, multiple coupling schemes can be obtained from the map, and since the innovation scheme is a combination of innovation dimensions and innovation laws, it needs to be interpreted in order to describe the innovation scheme in the multidimensional innovation map as a product innovation opportunity.

2.2.3. Green Finance Product Innovation Opportunity Identification. In order to make the identified product innovation opportunities more authentic and convincing, the steps of opportunity evaluation can be carried out selectively according to the actual situation. Since not every product innovation opportunity identified from the multidimensional innovation map is of high value, it needs to be evaluated by using a formula to assess the potential value of each product innovation opportunity. References in this study, to determine the weight of each indicator is determined using the expert scoring method, and the value of each indicator is calculated separately, which in turn helps to obtain the potential value of each product innovation opportunity, and finally to determine the most feasible and thus commercially valuable product innovation opportunities, and to maximize the benefits for the enterprise. The specific process of determining product innovation opportunities can be calculated using the following formulas:

$$Z_i = \alpha_1 * y_{(x_{1i})} + \alpha_2 * y_{(x_{2i})} \tag{7}$$

$$y_{(x_{1i})} = \left(\sum_{m=1}^{n} a_{mi}\right) \Big/ an\left(0 < n \le 9,\ And\ N\ is\ an\ \mathrm{int}eger\right) \tag{8}$$

$$x_{2i} = R_i / C_i \tag{9}$$

$$y_{(x_{gi})} = \frac{x_{gi} - \min(g)}{\max(g) - \min(g)} \tag{10}$$

In Equation (7), Z_i represents the potential value of the i nd innovation opportunity. x_{1i} and x_{2i} represent the potential of the i th product innovation opportunity as judged from technology level indicators and market value level indicators respectively. One is the likelihood of technological realization and the other is the economic benefit. α_1 and α_2 are the weighting coefficients of the two indicators.

Equation (8) represents the calculation of the technical realizability x_{1i} of the ith innovation opportunity, which is calculated by averaging the sum of the ratios of the number of online comments included in the different dimensions included in the i nd innovation opportunity and the total number of online comments. Where a_{mi} represents the number of online reviews of a dimension included in the i th innovation opportunity. n represents the total number of dimensions included in the i th innovation opportunity. According to the formula, it can be seen that the more online comments an innovation opportunity represents, the higher the feasibility of realizing that opportunity.

Equation (9) shows the calculation of the economic efficiency indicator x_{2i} for the i st innovation opportunity. Where R_i represents the expected benefits of the i th innovation opportunity and C_i represents the expected costs of the i th innovation opportunity. The expected benefits and expected costs are determined by the expert scoring method. The economic efficiency indicator is derived by comparing the expected benefits of the positive indicator with the expected costs of the negative indicator.

Equation (10) represents the normalization of the two indicators of innovation opportunities. Due to the different units of measurement of technical realization possibilities and economic benefits, it is necessary to normalize them in order to facilitate the calculation of the potential value of each innovation opportunity.

Through the above calculation process, the score of each product innovation opportunity can be obtained, and based on the score, it is filtered and ranked to determine the best product innovation opportunities, which are also the most practically significant product innovation opportunities.

2.3. *Green financial product innovation data mining results*

2.3.1. Green Financial Product Innovation Keywords Extraction. The initial weights of the green financial product feature words and the similarity and co-occurrence relationships between the words are obtained by using the TF-IDF calculation, and the initial weights of the point features and the edge feature matrices are respectively brought into the formula to obtain the final weights of the candidate green financial product feature words, and the comprehensive weights of the green financial product feature words are shown in Table 2, and the values shown in the results are obtained in the case where the sliding window is equal to 3, and only the top 20 ranked words are shown. The values shown in the results are obtained when the sliding window is equal to 3, and only the top 20 green financial product feature words are shown.

From the analysis of single feature weights, it can be seen that for green bond products, the more important product features such as green, loan, and credit policy have feature weights greater than 1, while the weight coefficients of diversification, trust, coverage, scale, and transaction, which are concerned by consumers, are lower (all less than 0.35). For green insurance products, green, insurance, statistically available, agriculture, insurance scale, payout rate and impact on the environment are the product features that consumers pay more attention to and mention more (all greater than 0.40), while the product's policy (0.1941) and financial support (0.1871) are less concerned by consumers. For green investment innovation products, consumers are mainly concerned about their nature characteristics, such as carbon finance, green, environmental protection, investment, global climate and greenhouse gases, etc., while the methods and trends of investment (tools 0.3929, globalization 0.3909), etc. are less concerned by consumers. Green factors in these three financial innovation products almost always attract consumers' attention, while product features with lower weights do not mean that they do not affect consumers' willingness to buy, but only indicate that they are less popular than product features with higher weights. Financial firms can innovate and improve the above financial product features that are of greater interest to consumers, in order to gain a competitive advantage.

Table 2. The comprehensive weight of the characteristic words of green financial products

Green bond products		Green insurance product		Green investment products	
Feature word	Weight	Feature word	Weight	Feature word	Weight
Green	1.7802	Green	1.3682	Carbon finance	2.862
Loan	1.4383	Insurance	0.9192	Green	2.0257
Credit	1.2852	Statistics	0.5411	Environmental protection	1.4807
Policy	1.0195	Agriculture	0.5263	Invest	1.4682
Welcome	0.7386	Scale	0.4889	Global climate	1.3809
Breed	0.5298	Loss rate	0.4636	Greenhouse gas	1.1424
Market	0.4866	Environment	0.4442	Carbon emission	0.8163
Issue	0.4454	Liability insurance	0.3719	Low-carbon economy	0.712
Product	0.4249	Natural environment	0.3682	Carbon trading	0.565
New type	0.411	Systematism	0.3183	Market	0.5574
Total amount	0.4094	Disaster	0.3061	Green financial system	0.532
Finance	0.4046	Develop	0.286	Gas discharge	0.5184
Safeguard	0.3758	Take seriously	0.2681	Financing	0.5165
Stabilization	0.3549	Environmental protection	0.257	Invest	0.5124
Develop	0.3498	Agricultural insurance	0.2235	Governance	0.4574
Diversification	0.343	Time	0.2217	Develop	0.4431
Trust	0.33	Implement	0.2196	Policy	0.4387
Covering	0.3287	Stabilization	0.2121	Turnover amount	0.4177
Scale	0.3116	Policy	0.1941	Tools	0.3929
Deal	0.3087	Fund	0.1871	Globalization	0.3909

2.3.2. Economic benefits of green financial products. According to the analysis of the value of green financial products, the distribution graph of technical novelty and technology following degree of all green financial products was obtained by using python, and the value assessment of green financial products is shown in Fig. 3, where the x-axis id denotes the product number of green financial innovation products after arranging them in time, the y-axis denotes the technological novelty of green financial products, and the z-axis denotes the technology following degree of green financial products. According to the relationship between the demonstrated economic benefits of green financial products and technical novelty and technical following degree, green financial products with higher technical novelty and technical following degree (both greater than 0.8), i.e., breakthrough green financial products, are clustered in the purple solid circle area, and the green financial products have the highest technical value. Green financial products with higher technology novelty, i.e., new technology green financial products, are clustered within the green solid circle area and have higher green financial product values. Green financial products with low technology following degree and technology novelty degree (both less than 0.2), i.e., low-value green financial products, are concentrated in the blue rectangular area, and the largest number of remaining improved green financial products, i.e., green financial products with high technology following degree and low technology novelty degree.

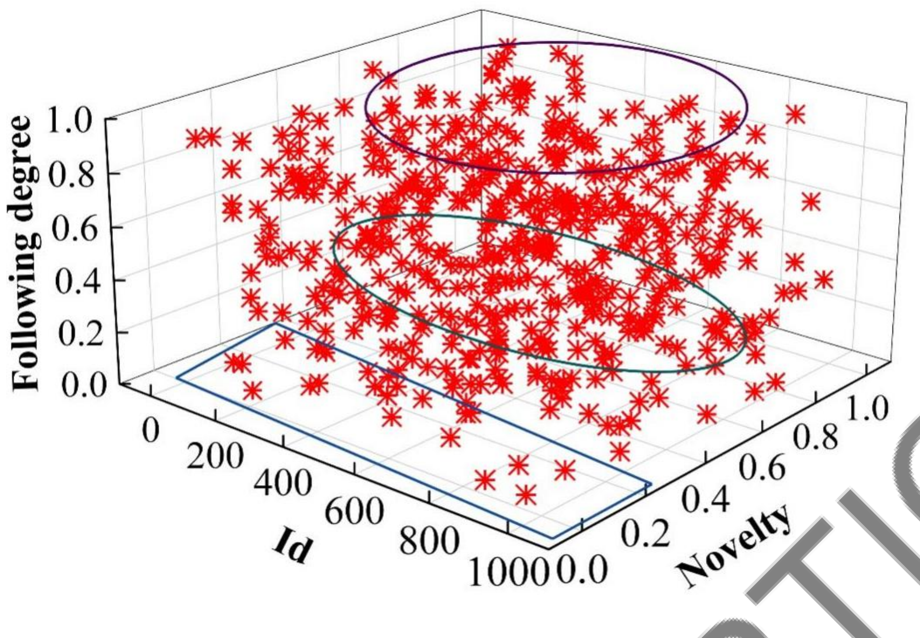

Fig. 3. Value evaluation of green financial products

3. Assessment of environmental benefits taking into account econometric modeling

3.1. VAR modeling for econometric purposes

It is often used to forecast time series as well as to analyze the dynamic interactions of stochastic disturbances on a system of variables [22].

$VAR(p)$ The mathematical expression of the model is:

$$y_t = A_1 y_{t-1} + \ldots + A_p y_{t-p} + BX_t + \mu_t \tag{11}$$

3.1.1. Granger causality test methodology. The Granger causality test is a statistical sense of causality, and the fact that causality in economics must be tested by actual data means that it indicates a relationship of influence between two sequences of variables reflected by mathematical statistics, which means that this test of causality may not be able to determine that a sequence of variables logically determines another sequence of variables [23].

Implementation of Granger causality test: Based on the above definition, we now assume that the realistic information set I_t consists of two smooth time series variables X_t and Y_t, and in order to test whether there exists Granger causality between them, we can construct two models as follows:

Unconstrained regression equation:

$$X_t = p_0 + \sum_{i=1}^{m} p_i X_{t-i} + \sum_{j=1}^{k} q_j Y_{t-j} + \eta_t \tag{12}$$

$$Y_t = \alpha_0 + \sum_{i=1}^{m} \alpha_i Y_{t-i} + \sum_{j=1}^{k} \beta_j X_{t-j} + \mu_t \tag{13}$$

where p_0 and α_0 represent constant terms, m represents the lag order, and η_t and μ_t are white noise sequences. (The corresponding regression equation with constraints in Eq. (13) is:

$$X_t = \alpha_0 + \sum_{i=1}^{m} \alpha_i X_{t-i} + \eta_t \tag{14}$$

$$Y_t = \alpha_0 + \sum_{i=1}^{m} \alpha_i Y_{t-i} + \mu_t \tag{15}$$

If $\beta_j = 0$ holds for all $j = 1, 2, \cdots, k$, then unidirectional Granger causality from X to Y does not occur, and the choice of lags can be arbitrary. This allows us to set up the null hypothesis $H_0 : \beta_j = 0, j = 1, 2, \cdots, k$.

Then regress equations (13) and (15) respectively, and obtain the residual sums of squares EES_1 and EES_2 of equations (13) and (15) by least squares regression, i.e., EES_1 represents the residual sum of squares of the unconstrained regression equation that includes the lagged variable X, and EES_2 represents the residual sum of squares of the constrained regression equation that does not include the lagged variable X, and if T is the total number of observations in the time series Y, construct the following F statistic as follows:

$$F = \frac{(EES_2 - EES_1)/m}{ESS_1/[T - (k + m + 1)]} \sim F_{[m, T-(m+k+1)]} \tag{16}$$

If the unconstrained regression equation (12) and the constrained regression equation (14) are regressed by least squares, and the F statistic of Eq. (16) is constructed by obtaining the sum of squares of the residuals, then, in the same way, if a significance level of α is given, there is a corresponding critical value of F_α, and if $F > F_\alpha$, a confidence level of $1-\alpha$ is rejected H_0. The hypothesis that Y is the cause of X in the Granger sense, i.e., the one-way Granger causality, $Y_t \Rightarrow X_t$. Otherwise, accept the H_0 hypothesis, a change in X cannot be attributed to a change in Y. If the result rejects both null hypotheses H_0, X and Y, there is a Granger causal relationship of two-way feedback, which can be expressed as $Y_t \Leftrightarrow X_t$.

3.1.2. Impulse Response Function. The basic idea of the impulse response function is illustrated according to the $VAR(2)$ model. Then:

$$\begin{cases} x_t = a_1 x_{t-1} + a_2 x_{t-2} + b_1 z_{t-1} + b_2 z_{t-2} + \mu_t \\ z_t = c_1 x_{t-1} + c_2 x_{t-2} + d_1 z_{t-1} + d_2 z_{t-2} + \mu_{2t} \end{cases} \tag{17}$$

$$\begin{gathered} E(\mu_t) = 0, t = 0, \pm 1, \ldots \\ Var(\mu_t) = E(\mu_t \mu_t') = \sum = \{\sigma_{ij}\}, t = 0, \pm 1, \ldots \\ E(\mu_t \mu_s') = 0, t \neq s \end{gathered} \tag{18}$$

Assuming that the above system is active from period 0 and let $x_{-1} = x_{-2} = z_{-1} = z_{-2} = 0$, and let the perturbation terms $\mu_{10} = 0$, $\mu_{20} = 0$ be given in period 0 and 0 thereafter, i.e., $\mu_{1t} = \mu_{2t} = 0$, $(t = 1, 2, \ldots)$ call this period 0 given x to the impulse, and discuss the responses of x_t and z_t below, at $t = 0$:

$$x_0 = 1, z_0 = 0 \tag{19}$$

Substituting the result into equation (17), at $t = 1$:

$$x_1 = a_1, z_1 = c_1 \tag{20}$$

This result is then substituted into equation (17) at $t = 2$:

$$x_2 = a_1^2 + a_2 + b_1 c_1 \tag{21}$$

$$z_2 = c_1 a_1 + c_2 + d_1 c_1 \tag{22}$$

Continuing this calculation, let the result of the solution be:

$$x_0, x_1, x_2, x_3, x_4, \ldots\ldots \tag{23}$$

Is called the response function of x induced by the pulse of x. The same is derived:

$$z_0, z_1, z_2, z_3, z_4, \ldots\ldots \tag{24}$$

is called the response function of z induced by the pulse of x.

Extending the above discussion to multivariate $VAR(p)$ models, in general, the response function induced y_i by an impulse from y_i can be derived as follows:

$$C_{0,ij}, C_{1,i,j}, C_{2,ij}, C_{3,ij}, C_{4,ij}, \ldots\ldots \tag{25}$$

And the i nd row and j rd column elements of C_s can be represented as $(s = 0,1,\ldots\ldots)$:

$$C_{s,ij} = \frac{\partial y_{i,t+s}}{\partial \mu_{jt}} \tag{26}$$

As a function of s, it describes the response of $y_{i,t+s}$ to an instantaneous change in period t holding other and earlier variables constant, which we call the impulse-response function.

3.1.3. Variance decomposition methods. The variance decomposition is to analyze the contribution of each structural shock to the change in the endogenous variable (usually measured by the variance) in order to evaluate the significance of the different structural shocks, along the following lines:

According to equation (26) as well:

$$\begin{aligned} &\left(I - A_1 L - \ldots - A_p L_P\right)\left(I + C_1 L + C_2 L^2 + \ldots = I\right) \\ &I + \psi_1 L + \psi_2 L + \ldots = I \end{aligned} \tag{27}$$

There is:

$$\tilde{y}_{it} = \sum_{j=1}^{n}\left(\psi_{0,ij}\mu_{jt} + \psi_{1,ij}\mu_{jt-1} + \psi_{2,ij}\mu_{jt-2} + \psi_{3,ij}\mu_{jt-3} + \ldots\right) \tag{28}$$

It can be shown that the contents of the individual brackets are the sum of the effects of the j st perturbation term μ_j on y_i from the infinite past to the present time point. Find its variance, since $\{\mu_{jt}\}$ is not serially correlated:

$$\begin{aligned} &E\left\{\left(\psi_{0,ij}\mu_{jt} + \psi_{1,ij}\mu_{jt-1} + \psi_{2,ij}\mu_{jt-2} + \psi_{3,ij}\mu_{jt-3} + \ldots\right)^2\right\} = \sum_{q=0}^{\infty}\left(\psi_{q,ij}\right)^2 \delta_{ij} \\ &j = 0,1,2,\ldots n \end{aligned} \tag{29}$$

$$\operatorname{var}(\tilde{y}_{it}) = r_{ii}(0) = \sum_{j=1}^{n}\left\{\sum_{q=0}^{\infty}\left(\psi_{q,ij}\right)^2 \sigma_{ij}\right\} \tag{30}$$

The variance of $\tilde{y}_{it}$ can be decomposed into n uncorrelated effects, so in order to determine how much the individual perturbations contribute relative to the variance of $\tilde{y}_{it}$ the following scale is defined:

$$RVC_{j\to i}(\infty) = \frac{\sum_{q=0}^{\infty}\left(\psi_{q,ij}\right)^2 \sigma_{ij}}{\sum_{j=1}^{n}\left\{\sum_{q=0}^{\infty}\left(\psi_{q,ij}\right)^2 \sigma_{ij}\right\}}, i, j = 1,2,\ldots,n \tag{31}$$

In practice, it is not possible to use $s = \infty$ the $\psi_{k,i}$ to evaluate, only a limited number of s terms are required. The prediction error for the first s period of the $\operatorname{var}(p)$ model is:

$$\mu_{t+s} + \psi_1 \mu_{t+s-1} + \psi_2 \mu_{t+s-2} + \ldots + \psi_{s-1}\mu_{t+1} \tag{32}$$

Therefore:

$$RVC_{j\to i}(s)=\frac{\sum_{q=0}^{s-1}\left(\psi_{q,ij}\right)^2\sigma_{ij}}{\sum_{j=1}^{n}\left\{\sum_{q=0}^{s-1}\left(\psi_{q,ij}\right)^2\sigma_{ij}\right\}},i,j=1,2,\ldots,n \tag{33}$$

3.2. Selection of indicator variables

The sample data time of this paper is 2007-2022. Through the data mining of green financial product innovation, the indicators needed in this paper are categorized into two aspects, which are green financial product development index indicators and environmental economic benefit index indicators. According to the availability and representativeness of the indicators, combined with the actual situation of the development of green financial products in China, the green financial product development indicators are categorized into green bond products (GBP) = total output value of environmental protection enterprises/total market value of A-shares, green insurance products (GAP) = expenditure on agricultural insurance/total insurance expenditures, and green investment products (GIP) = investment in pollution control/GDP.The environmental economic benefit index (EEB) = GDP per capita + greening coverage of built-up areas.

4. Empirical analysis

4.1. Empirical tests

4.1.1. Variable Smoothness Tests. First of all, it is necessary to verify whether each data under study is smooth, for smooth variables can be directly established regression model, if the variables are non-smooth variables first of all, we must carry out the cointegration test in order to further do the regression analysis, the variables are not smooth will increase the probability of pseudo-regression of the regression equation. In this paper, the four variables of environmental economic benefits (EEB), green bond products (GBP), green insurance products (GAP), green investment products (GIP) are sequentially tested by ADF unit root test, △EEB, △GBP, △GAP, △GIP represent the first-order difference form of the corresponding variables, and the results of unit root smoothness test are shown in Table 3. There are three types of unit root tests: those containing constant term (C), those containing time trend (T) and those not containing the above two are indicated by zero. Based on the results of the unit root test, it can be concluded that the original hypothesis is accepted for the variables EEB, GBP, GAP, and GIP at the 10% significance level, respectively, which indicates that the above vectors have a unit root, and all of them are unsteady vectors. The variables EEB, GBP, GAP, GIP are tested for unit root after first-order differencing, and from the test results, △EEB, △GBP reject the original hypothesis at 10% significance level, △GAP, △GIP reject the original hypothesis at 1% and 5% significance level, respectively, and the variables EEB, GBP, GAP, GIP are all smooth after first-order differencing, so the four variables are first-order single-integrated vectors, and further correlation tests need to be performed on the above variables.

Table 3. Unit root stability test results

Variable	Type	Adf	Significance level (threshold)			P	Conclusion
			1%	5%	10%		
Eeb	(c,0,2)	-2.22578	-4.00436	-3.09891	-2.69046	0.22379	Non-stationary
Gbp	(c,t,0)	-1.68992	-3.92031	-3.06566	-2.67353	0.42493	Non-stationary
Gap	(c,t,1)	-0.15769	-4.8001	-3.79124	-3.34224	0.98681	Non-stationary
Gip	(c,0,0)	-1.55434	4.66786	-3.73323	-3.3103	0.77833	Non-stationary
△eeb	(c,t,1)	-3.49487	4.80006	-3.79116	-3.34228	0.08204	Steady
△gbp	(c,0,0)	-2.66678	-3.95913	-3.08103	-2.68127	0.09856	Steady

△gap	(c,t,1)	-6.89872	-4.80002	-3.79119	-3.34221	0.00004	Steady
△gip	(c,0,0)	-3.88125	-3.95907	-3.08103	-2.68132	0.01622	Steady

4.1.2. Johansen's cointegration test. After the unit root test of the variables, it can be seen from the test results that the variables EEB, GBP, GAP and GIP are all first-order single-integrated vectors, and the Johansen cointegration test can be further done to determine whether there is a long-term equilibrium relationship between the variables EEB, GBP, GAP and GIP according to the test results, and Table 4 shows the results of Johansen cointegration test. In this paper, the cointegration relationship between green bond products (GBP), green insurance products (GAP), green investment products (GIP) and environmental economic benefits (EEB) is judged by the traces test and the maximum eigenvalue test results, under the assumption condition that the original hypothesis is that there are at most two cointegration relationships, from the traces test results, we can see that the critical value of 25.87171 is smaller than the statistical value of 34.7465, at this time The p-value is 0.0083, so the original hypothesis is rejected according to the results of the trace test. From the test of the largest eigenvalue, the critical value of 19.38754 is smaller than the statistical value of 23.5219, at this time p = 0.0082 also rejected the original hypothesis. Under the assumption condition that the original hypothesis has at most three characteristic roots, the original hypothesis is accepted according to the Johansen cointegration traces test and the maximum eigenvalue test. Therefore, at 5% significance level, there are three cointegration relationships between the four variables of EEB, GBP, GAP and GIP, and there is a long-term dynamic equilibrium relationship between the four variables of green bond products, green insurance products, green investment products and environmental economic benefits.

Table 4. The Johansen Cointegration relationship test results

Original hypothesis	Characteristic root	The trace test			Maximum eigenvalue test		
		Statistic	Critical	P	Statistic	Critical	P
None*	0.97544	118.1797	63.8762	0.0000	55.53929	32.11842	0.0000
At most1*	0.91964	62.6406	42.91515	0.0001	37.8932	25.82281	0.0004
At most2*	0.72806	34.7465	25.87171	0.0083	23.5219	19.38754	0.0082
At most3*	0.29444	5.2256	12.51818	0.5652	5.2251	12.51768	0.5642

4.1.3. Glenn causality test. According to the test results of the unit root test method, it can be seen that the variables EEB, GBP, GAP and GIP are first-order single-integrated variables, respectively, and it is also necessary to do Granger causality test before constructing the VAR model, to analyze whether there is Granger causality between the variables GBP, GAP, and GIP and EEB, i.e., analyze whether there is any influence relationship between the innovation of green financial products and the environmental economic benefits. Granger causality test is performed on variables EEB, GBP, GAP and GIP when the lag order is 1, 2 and 3 respectively. If the test result appears $p<0.1$, it means that the original hypothesis is rejected and there is Granger causality between the variables, and if $p>0.1$, it means that the original hypothesis is accepted and it means that there is no Granger causality between the variables. The results of Granger causality test are shown in Table 5. When the order of the model lag is 2, the Granger causality of GIP and EEB is bidirectional, that is, in the long run green investment products have an impact on the environmental economic efficiency and the effect between the two has a lag. The GAP and EEB Granger causality is unidirectional, and when the order of the model lag is 2, the GAP is the Granger cause of the EEB, which indicates that in the long run green insurance products have an impact on the EEB has an impact. When the lagged order of the model is 2, the Granger causality of GBP and EEB is bi-directional, i.e., green bond products are conducive to promoting environmental economic benefits, and the enhancement of environmental economic benefits contributes to the investment of green bond products. The Granger causality of GAP and GIP is bi-directional, and when the lagged order of the model is 3, GAP is the Granger cause of GIP, indicating that in the long run green insurance products have an effect on green investment products.

Table 5. Granger causality test results

Original hypothesis	Lag order	F statistic	P	Conclusion
GIP does not Granger Cause EEB	2	5.00553	0.0736	Refuse
EEB does not Granger Cause GIP	2	4.70015	0.0812	Refuse
GAP does not Granger Cause EEB	2	4.76785	0.035	Refuse
EEB does not Granger Cause GAP	2	0.06126	0.9409	Accept
GBP does not Granger Cause EEB	2	6.98406	0.0121	Refuse
EEB does not Granger Cause GBP	2	4.18672	0.0445	Refuse
GAP does not Granger Cause GIP	3	6.18762	0.0451	Refuse
GIP does not Granger Cause GAP	2	2.18601	0.1627	Accept
GBP does not Granger Cause GIP	1	6.2743	0.0337	Refuse
GIP does not Granger Cause GBP	2	4.22735	0.0716	Refuse
GBP does not Granger Cause GAP	1	3.4378	0.0865	Refuse
GAP does not Granger Cause GBP	1	5.07161	0.0034	Refuse

4.2. VAR model

4.2.1. VAR model construction. Based on the empirical findings of the VAR model the following equation is derived:

$$\begin{aligned} EEB &= 0.27152 * EEB(-1) + 0.00521 * EEB(-2) + 1.22842 * GIP(-1) \\ &+0.27883 * GIP(-2) + 0.33578 * GBP(-1) + 0.06889 * GBP(-2) \\ &+0.37397 * GAP(-1) + 0.04955 * GAP(-2) + 2.42513 \end{aligned} \tag{34}$$

According to the regression results of the VAR model, it can be seen that the regression coefficients of the variables GIP, GAP, and GBP are all positive at the 2nd order of the model lag, which indicates that the effects of the variables GIP, GAP, and GBP on the EEB are positive, while the main purpose of establishing the vector autoregressive model is not to study the relationship between vectors through the coefficients of the regression equations, and the ultimate purpose of modeling is to carry out impulse response and variance decomposition.

4.2.2. VAR model stability test. The unit circle test is to test whether the VAR model is stable or not, if all the eigenvalues are within the unit circle, it can be concluded that this VAR model is stable, on the contrary, this VAR model is not stable. The results of the model stability test are shown in Figure 4. The absolute value of the inverse of each root is not greater than 1, and all the eigenvalues are within the unit circle, so the VAR model established in this paper is stable, and it can be continued to the next impulse response and ANOVA.

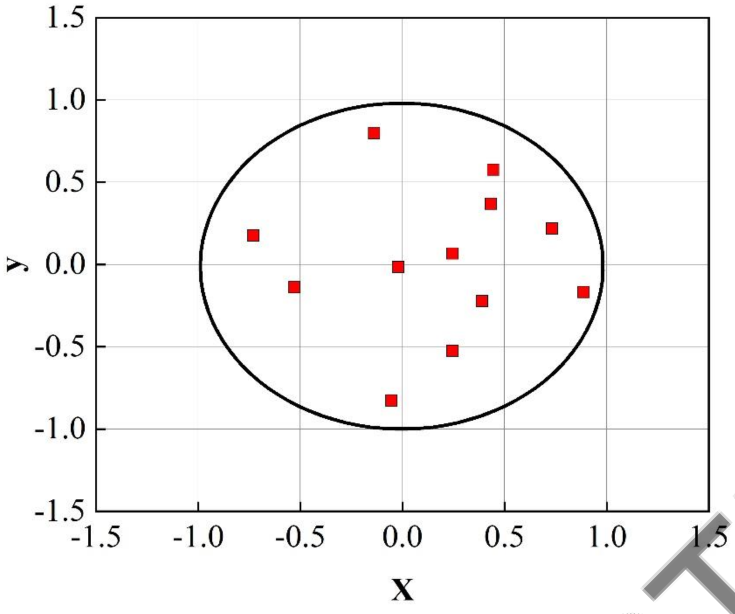

Fig. 4. Model stability test results

4.3. Impulse Response and Analysis of Variance (ANOVA)

4.3.1. Impulse Response. In order to further visualize the degree of mutual influence between the two sequences, we will use the impulse response function to study the response generated by the impact of the environmental economic benefits by itself, the color bond product, the green insurance product, the green investment product, Figure 5 shows the impulse response results, and (a) to (d) are the impulse response of itself, the green bond product, the green insurance product, and the green investment product, respectively.

Figure (a) can be seen, when giving the environmental economic benefits of itself a shock to the impact of economic growth, the fluctuations of the environmental economic benefits are always positive and gradually increasing, in the second period to reach the maximum value and then decreasing, and 3 to 10 period is a steady decline in the state, it can be shown that in the long run, the environmental economic benefits of its own impact on the environmental economic benefits of the impact is positive. Figure (b) to (d) can be seen to green bond products, green insurance products, green investment products for the environmental economic benefits of the impact of fluctuations is more obvious, green bond products on the environmental economic benefits of the impact of positive fluctuations in the 3rd period to reach the maximum value, followed by fluctuations gradually smooth. The impact of green insurance products on the size of the economy is more rapid and the magnitude is also larger, at first it is down, then it rises, some times it is negative fluctuation, overall still shows positive fluctuation. The impact of green investment products on the structure of the economy first rises and then falls, and then continues, showing a negative direction in the short term, but positive fluctuations in the long term.

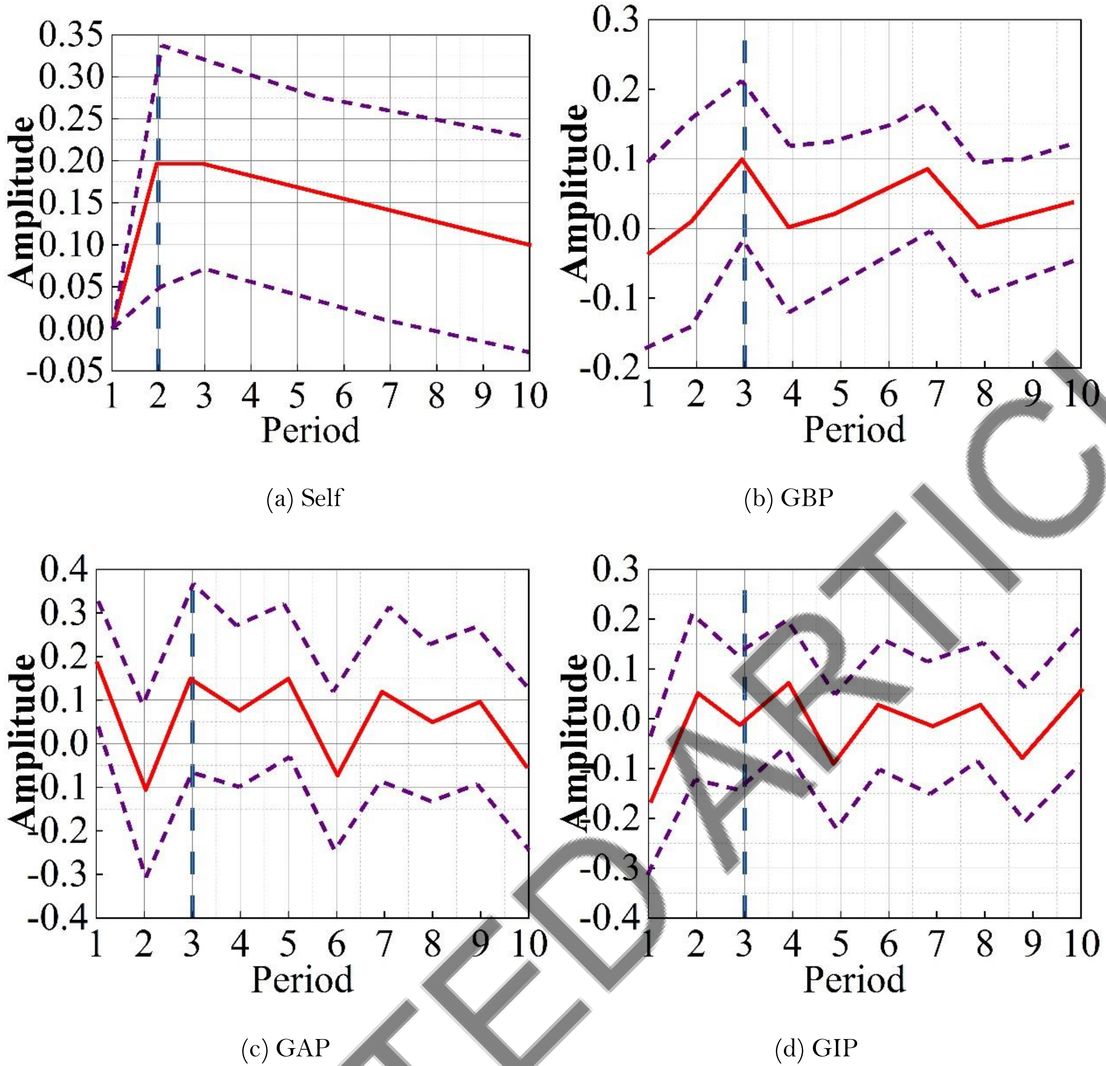

Fig. 5. Pulse response results

4.3.2. Analysis of variance. Impulse response and variance decomposition are two sides of the same coin. Impulse response is a measure of how the endogenous variables in the model respond to the impulse of a variable, while variance decomposition is how to decompose the response of a variable into the endogenous variables in the model, which can specifically examine the size of the contribution of the individual disturbance terms of a variable to its movement. So we use the method of variance decomposition to study how much green bond products, green insurance products and green investment products affect the environmental benefits when green financial products are innovated, and the results of variance decomposition are shown in Table 6. When the lag period is larger, the environmental economic benefit itself, the impact of green investment products on the environmental economic benefit is getting smaller and smaller, and the impact of green bond products and green insurance products on the environmental benefit is getting larger and larger. When the lag period is 10 periods, the contribution of environmental economic benefit itself to the change of environmental economic benefit tends to 23.79%, while the contribution of green investment products, green bond products and green insurance products to environmental economic benefit tends to 9.72%, 20.23% and 21.83% respectively. In summary, it can be concluded that green product innovation has a certain influence on the fluctuation of environmental economic benefits, and green bond products and green insurance products have a greater impact on environmental economic benefits.

Table 6. Variance decomposition results

Period	EEB	GIP	GBP	GAP

1	0.00	1.56	26.33	26.93
2	5.49	1.39	20.25	22.29
3	10.60	7.38	20.73	20.12
4	14.28	6.71	21.7	19.02
5	17.15	6.34	22.59	20.86
6	19.50	7.58	21.73	19.92
7	21.27	10.32	21.73	19.68
8	22.68	10.05	21.88	19.12
9	23.79	9.72	21.83	20.23

5. Conclusion

This paper evaluates the innovation characteristics and value benefits of green financial products based on the general process of data mining, and introduces an econometric model to explore the impact of green financial product innovation on environmental economic benefits. Conclusions are drawn through empirical analysis as follows:

1) In the long run, the impact of the environmental economic benefit itself is positive on the environmental economic benefit. The impact of green bond products on environmental economic benefits shows positive fluctuations and reaches the maximum value in the 3rd period, and then the fluctuations are gradually smooth. The impact of green insurance products on the size of the economy is more rapid and the magnitude is also larger, at first it is down, then it rises, some times it is negative fluctuation, overall still shows positive fluctuation. The impact of green investment products on the structure of the economy first rises and then falls, and then continues, the short term presents a negative direction, but in the long term is a positive fluctuation.

2) When the lag period is 10 periods, the contribution of environmental economic benefit itself to the change of environmental economic benefit tends to 23.79%, while the contribution of green investment products, green bond products and green insurance products to environmental economic benefit tends to 9.72%, 20.23% and 21.83% respectively. Green product innovation has a certain influence on the fluctuation of environmental economic benefits, and green bond products and green insurance products have a greater impact on environmental economic benefits.

3) Green financial product innovation and development can promote the development of environmental economic benefits. Under the influence of the current international situation, it is necessary to pay great attention to and develop efficient, energy-saving and environmentally friendly green financial products to promote the common development of environmental economic benefits, social and resource ecological benefits. In addition, it should continue to increase the cultivation of talents, take the government as the leading role, strengthen cooperation with various universities and explore various modes of cultivating relevant talents.

References

[1] Muganyi, T., Yan, L., & Sun, H. P. (2021). Green finance, fintech and environmental protection: Evidence from China. Environmental Science and Ecotechnology, 7, 100107.

[2] Akomea-Frimpong, I., Adeabah, D., Ofosu, D., & Tenakwah, E. J. (2022). A review of studies on green finance of banks, research gaps and future directions. Journal of Sustainable Finance & Investment, 12(4), 1241-1264.

[3] Sachs, J. D., Woo, W. T., Yoshino, N., & Taghizadeh-Hesary, F. (2019). Importance of green finance for achieving sustainable development goals and energy security. Handbook of green finance, 3, 1-10.

[4] Zhou, X., Tang, X., & Zhang, R. (2020). Impact of green finance on economic development and environmental quality: a study based on provincial panel data from China. Environmental Science and Pollution Research,

27, 19915-19932.

[5] Yu, C. H., Wu, X., Zhang, D., Chen, S., & Zhao, J. (2021). Demand for green finance: Resolving financing constraints on green innovation in China. Energy policy, 153, 112255.

[6] Rahman, M. (2023). The virtuous circle between green product innovation and performance: The role of financial constraint and corporate brand. Journal of Business Research, 154, 113296.

[7] Zhang, D., & Vigne, S. A. (2021). How does innovation efficiency contribute to green productivity? A financial constraint perspective. Journal of Cleaner Production, 280, 124000.

[8] Stucki, T. (2019). What hampers green product innovation: the effect of experience. Industry and Innovation, 26(10), 1242-1270.

[9] Söderholm, P. (2020). The green economy transition: the challenges of technological change for sustainability. Sustainable Earth, 3(1), 6.

[10] Agrawal, R., Agrawal, S., Samadhiya, A., Kumar, A., Luthra, S., & Jain, V. (2024). Adoption of green finance and green innovation for achieving circularity: An exploratory review and future directions. Geoscience Frontiers, 15(4), 101669.

[11] Tolliver, C., Fujii, H., Keeley, A. R., & Managi, S. (2021). Green innovation and finance in Asia. Asian Economic Policy Review, 16(1), 67-87.

[12] Jiakui, C., Abbas, J., Najam, H., Liu, J., & Abbas, J. (2023). Green technological innovation, green finance, and financial development and their role in green total factor productivity: Empirical insights from China. Journal of Cleaner Production, 382, 135131.

[13] Irfan, M., Razzaq, A., Sharif, A., & Yang, X. (2022). Influence mechanism between green finance and green innovation: exploring regional policy intervention effects in China. Technological Forecasting and Social Change, 182, 121882.

[14] Zhou, G., Zhu, J., & Luo, S. (2022). The impact of fintech innovation on green growth in China: Mediating effect of green finance. Ecological Economics, 193, 107308.

[15] Lamperti, F., Bosetti, V., Roventini, A., Tavoni, M., & Treibich, T. (2021). Three green financial policies to address climate risks. Journal of Financial Stability, 54, 100875.

[16] D'Orazio, P., & Popoyan, L. (2019). Fostering green investments and tackling climate-related financial risks: Which role for macroprudential policies?. Ecological Economics, 160, 25-37.

[17] He, B. (2022). Application of data mining technology in enterprise green innovation model construction and path analysis. Computational Intelligence and Neuroscience, 2022(1), 7194171.

[18] Xu, R., Chen, X., & Zhang, F. (2019). Green Technology Innovation and Sustainable Development Based on Data Fusion Mining. Ekoloji Dergisi, (107).

[19] Hsu, C. C., Quang-Thanh, N., Chien, F., Li, L., & Mohsin, M. (2021). Evaluating green innovation and performance of financial development: mediating concerns of environmental regulation. Environmental Science and Pollution Research, 28(40), 57386-57397.

[20] Mihui Kim & Haesoo Kim. (2024). A Dynamic Analysis Data Preprocessing Technique for Malicious Code Detection with TF-IDF and Sliding Windows. Electronics(5).

[21] Jinyuan Xu. (2024). Application Design and Analysis of Network Data Mining in Microsystemic Software. International Journal of High Speed Electronics and Systems(prepublish).

[22] Cao Huimin, Wang Ping, Zhang Surong, Xu Dongpo & Tian Weijun. (2024). Research on the Relationship Between Environmental and Economic Coupling Systems in Bohai Bay Area Based on a Vector Autoregression (VAR) Model. Journal of Ocean University of China(2),557-566.

[23] Eirini Kostaridou, Nikolaos Siatis & Eleni Zafeiriou. (2024). Resource Price Interconnections and the Impact of Geopolitical Shocks Using Granger Causality: A Case Study of Ukraine–Russia Unrest. Journal of Risk and Financial Management(6),240-.