Identification of economic risks in international trade based on RF-Gini coefficients Research

Ruoxin Lu [1, #] , Chang Dong[2,#] , Zhaozhen Lu[3, #,*]

[1]School of Economics and Management, Beijing Jiaotong University, Beijing, 100091, China

[2]School of Economics, Capital University of Economics and Business, Beijing, 100070, China

[3]College of Mathematics, Sichuan University, Chengdu, 610065, China

[*]Corresponding author: zhaozhenlu1@outlook.com

[#]These authors contributed equally.

Abstract. The rapid development of economic globalization objectively promotes the prosperity of international trade, which brings challenges and opportunities for China's economic construction. Based on China's macroeconomic data and foreign trade data, this paper uses the random forest (RF) regression algorithm for training and optimizes iterations for different step sizes by adjusting the number of decision trees to obtain the model with the best fit, and later investigates the strength of each factor's influence on international trade through the Gini coefficient. The results show that the exchange rate (0.128), the Engel coefficient of residential households (0.128), and the consumer price index (0.128) have a greater impact on the international trade economy. Thus, to maximize the stability of international trade risks, the state needs to control the inflation rate, focus on the efficient and coordinated development of the country's internal economy, and work to ensure that the nation has the basic consumption capacity, consumption needs, and create a favorable consumption environment.

Keywords: random forest; Gini; international trade economic risk; China.

1. Introduction

In recent years, with the gradual formation of the global economic integration pattern, international trade has become a powerful means for countries to enhance their economic strength and obtain benefits from the international market. Based on this, enterprises from various countries have gradually joined the ranks of international trade and encountered risks and challenges. Especially in the case of an inevitable global economic recession after the epidemic, how to identify economic risks in international trade has become an urgent proposition to be solved. There are few studies in this area.

Yang and Wang [1] studied the risks of international trade in agricultural products in the post-epidemic era by analyzing the economic data related to agricultural products and mentioned that China should promote the sustainable and healthy development of international trade in agricultural products in the post-epidemic period by improving the level of domestic supply security of agricultural products. Xu et al [2] constructed a digital evaluation index system based on the development of digital industrialization and industry's digitalization, and discussed the impact of digitalization on international trade by combining theory and random forest to determine the influencing factors of international trade. Zhou [3] studied the legal risks in international trade, including the potential risks due to unclear contract terms, the uncertainty of legal procedures and modification of laws and regulations, and proposed measures to prevent legal risks. Wu and Wang [4] explored the method, basis, and process of risk identification in the international trade process and gave quantified magnitude of risk in utility. Peng [5] introduced several common transaction risks in international trade from the perspective of practical operation, including contract approval problems, disputes over the quantity and quality of goods, and the corresponding preventive measures. Xu et al [6] studied how the improvement in the level of economic development and the adjustment of the economic development model based on China's strong national governance capacity under the global value chain trading system have led to a significant increase in China's status, and at the same time,

how China should respond to the risks brought about by the transformation of the international trading system caused by China's elevated status. Liu et al [7] used the new measure of national trade dependence to construct an international trade dependence network, analyzed and visualized its structural features, then inferred the endogenous mechanisms affecting its network evolution based on network theory, and empirically tested the evolution and endogenous mechanisms of the international trade dependence network using a time-exponential random graph model.

Throughout the above studies, international trade economic risks have been analyzed in multiple types and perspectives. However, previous studies tend to focus on the impact of unilateral factors on international trade, lacking research on the combined effects of different factors on international trade. In addition, the analysis of data only stays in the conventional statistical analysis and preliminary regression analysis, and the analysis of data mining is still relatively small. Therefore, this paper first integrated previous research to determine the 12 indicators of the impact of international trade economic risk index system, and then by constructing RF-Gini model, quantitatively researched the strength of each factor for the impact of international trade, so that it can be more comprehensive to know the international trade risk size in the international trade risk identification for enterprises to determine whether to carry out international trade has a certain degree of guidance.

2. International trade economic risk impact indicator system construction

2.1 Measurement of international trade economy

International trade, which refers to the transaction of services across national borders, generally consists of import trade and export trade. International trade is beneficial to the development of a country, regulating the utilization of domestic production factors, improving international supply and demand, restructuring the economy, and increasing fiscal revenues. There are risks in international trade and there are many types of risks that are difficult to classify, however, the ultimate impact of any risk on international trade will be reflected in an increase or decrease in the total amount of international trade imports and exports. Different factors have different degrees of impact on international trade, which will be reflected in the amount of change in the total amount of increase. Thus, this paper uses the inverse of the increase in the total amount of imports and exports of goods ($\mathcal{Y}$), as the risk coefficient, i.e. the dependent variable.

2.2 Construction of impact indicator system

China's international trade is mainly influenced by two aspects: China's national conditions and the world macroeconomic situation. In the case of China's national conditions, China's gross national income ($\mathcal{X}_1$) reflects the country's productivity. Reserves ($\mathcal{X}_2$) reflect the degree of sophistication of China's financial system. Tariffs ($\mathcal{X}_4$) reflect the setting of trade barriers in international trade. The consumer price index ($\mathcal{X}_5$) reflects the high or low prices of goods. The Engel coefficient ($\mathcal{X}_6$) reflects the level of affluence in China. The number of people employed in the transportation sector ($\mathcal{X}_7$) reflects the extent of China's logistics system. The number of Internet users ($\mathcal{X}_8$) reflects how easy it is for China to communicate with the international community. The Shanghai Stock Exchange Index ($\mathcal{X}_9$) reflects the general trend of the Shanghai stock market, which further reflects the economic situation in China. As far as the world macroeconomic situation is concerned, the exchange rate ($\mathcal{X}_3$) system of purchasing power of the RMB to some extent affects the willingness to buy. The number of foreign economic contracts ($\mathcal{X}_{10}$) reflects the number of international trade and the amount of foreign economic contracts ($\mathcal{X}_{11}$) reflects the size of international trade.

In summary, the index system established in this paper is shown in Table 1.

Table 1 Table of index system

Indicator Name	Indicator Number	Indicator Unit	Reference
The inverse of the growth of total exports and imports of goods	$\mathcal{Y}$	No unit	[8]
Gross National Income	$\mathcal{X}1$	Billion	[9]
Reserves	$\mathcal{X}2$	Billion	[9]
100 USD	$\mathcal{X}3$	Renminbi Yuan	[9]
Tariffs	$\mathcal{X}4$	Billion	[10]
Consumer Price Index	$\mathcal{X}5$	1978 = 100	[11]
Household Engel Coefficient	$\mathcal{X}6$	No unit	[11]
Number of employed persons in the transportation industry	$\mathcal{X}7$	People	[12]
Number of people with Internet access	$\mathcal{X}8$	10,000 people	[13]
SSE Composite Index	$\mathcal{X}9$	No unit	[9]
Number of foreign economic contracts	$\mathcal{X}10$	Portions	[14]
Amount of foreign economic contracts	$\mathcal{X}11$	billion	[14]

3. Random Forest-Gini algorithm construction

3.1 Random Forest method

Random Forest (RF) is essentially a classifier composed of multiple decision trees and was first refined and proven as an efficient algorithm by Breiman in 2001. Essentially, it is a combinatorial classifier $\{(\mathcal{X}, \beta_{\hbar}), \hbar=1,\cdots, n\}$ where $\mathcal{X}$ is the individual feature variable, i.e., the input variable. $\beta_{\hbar}$ is a random vector with independent identical distribution, and n represents the number of decision trees contained in the random forest. The random forest is based on the majority voting method of each decision tree classifier to determine the classification results. Using the random forest model, the importance of individual features can be calculated, the importance of individual features can be evaluated, and these features can be ranked according to their importance. Figure 1 shows a schematic diagram of the random forest construction.

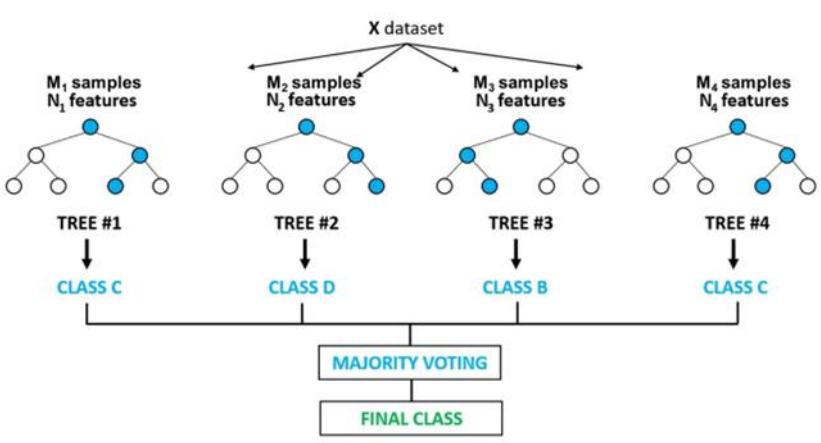


Figure 1 Random forest construction diagram

3.2 Gini Index

The Gini index represents the probability of misclassifying a randomly selected sample in the sample set. The smaller the Gini index is, the lower the probability of misclassifying a randomly

selected sample in the sample set is, which means that the set is purer, and vice versa, the set is less pure. When all samples in the set are of one class, the Gini index is 0. This paper denotes the variable importance scores by *IM* and the Gini index by P_{mk} . The data set in this paper has n features $X_1, X_2, \ldots\ldots, X_n$. Then calculate the Gini index score for each feature X_n which is also known as the nth feature's average change in split node impurity across all decision trees in the random forest. The Gini index is calculated as

$$GI_m = \sum_{K=1}^{K} \sum_{K \neq K} P_{mk} P_{mk'} = 1 - \sum_{K=1}^{K} P^2{}_{mk} \tag{1}$$

where, K indicates that there are K categories, and indicates the proportion of categories K in the node. The importance of feature X_n in node m, i.e., the amount of change in Gini index before and after the branching of node m is

$$IM_n^{GI} = GI_m - GI_x - GI_y \tag{2}$$

where, GI_X and GI_Y denote the Gini indices of the two new nodes after branching, respectively.

If the feature Xn appears in the decision tree i at the node of the set M , then Xn is important in the ith tree as

$$IM_{in}^{GI} = \sum_{M}^{n} IM_{im}^{GI} \tag{3}$$

Assume that there are n trees on RF, then.

$$IM_{in}^{GI} = \sum_{i=1} IM_{in}^{GI} \tag{4}$$

Finally, normalize all the derived importance scores yields.

$$IM_n = \frac{IM_n}{\sum_{i=1} IM_{in}^{GI}} \tag{5}$$

4. The setting of test indicators

As the stochastic model has some randomness, this model may assign different importance weights to each feature each time. As a result, the stochastic model should be trained and tested several times to find the number of models with the best fit and then retain the importance features under that number of models as the final result. In this paper, we use R^2 and MSE test the GNI, reserve, and exchange rate which are indicators to find the number of models under the optimal fit. The following equations are used to calculate R^2 and MSE .

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{6}$$

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 \tag{7}$$

R^2 is the most widely-used coefficient to evaluate the model. R^2 reflects the goodness of fit of the model and is the ratio of the regression sum of squares to the total sum of squares. R^2 values between 0 and 1 and is unitless. The magnitude of its value reflects the relative degree of regression contribution. According to Equation (6), with the sample data already obtained, $\sum_{i=1}^{n} (y_i - \bar{y})^2$ in the expression R^2 is a determined number. Therefore, a larger R^2 (close to 1) means a smaller $\sum_{i=1}^{n} (y_i - \bar{y})^2$,

i.e., a better fit of the model; *MSE* is a measure reflecting the degree of difference between the estimated and the estimated quantity. In general, the metric used to evaluate the point estimate is the function of the distance between the point estimate $\hat{\theta}$ and the true value of the parameter θ. The most commonly used function is the square of the distance, and since the estimate $\hat{\theta}$ is random, the expectation can be derived from this function. Smaller values of the indicator indicate that the difference between the predicted and true values of the model is smaller and that the random forest has a better predictive power for international trade risk.

4.1 Data Collection

Taking the economic situation of China's international trade as an example, based on China's macroeconomic database and China's financial database in the EPSDATA data platform, this paper collects data from multiple perspectives, such as national economic accounting, employment, and population, foreign economic and trade, finance, price index, people's life, transportation and traffic, Internet indicators, finance, etc. Based on the selected data for 10 years from 2011 to 2020, this paper interpolates certain missing values of data. Considering that the independent variables and dependent variables are not uniform, the data will be standardized to Min-Max during the training of the model, and the specific transformation function is

$$X^* = \frac{X - \min}{\max - \min} \tag{8}$$

After completing data preprocessing, retaining two decimal places, and using scientific notation, descriptive statistics for the indicators are shown in Table 2.

Table 2 Descriptive statistics of indicators

Indicator Number	Min	Max	Mean	Mid
y	-4.72E-04	1.65E-04	1.80E-05	5.43E-05
$x1$	4.83E+05	1.01E+06	7.42E+05	7.14E+05
$x2$	-1.75E+04	3.62E+04	1.07E+04	1.28E+04
$x3$	6.14E+02	6.90E+02	6.51E+02	6.54E+02
$x4$	2.56E+03	3.00E+03	2.73E+03	2.71E+03
$x5$	5.65E+02	6.87E+02	6.23E+02	6.21E+02
$x6$	2.82E+01	3.63E+01	3.09E+01	3.01E+01
$x7$	4.59E+06	7.26E+06	6.58E+06	6.80E+06
$x8$	5.13E+04	9.31E+04	7.19E+04	7.10E+04
$x9$	2.12E+03	3.54E+03	2.88E+03	3.08E+03
$x10$	6.38E+03	2.28E+04	1.16E+04	1.05E+04
$x11$	1.42E+03	2.65E+03	2.14E+03	2.26E+03

4.2 Algorithm Training

Based on the SPSS software platform, this paper substitutes the data of each indicator under the years 2010~2020 and uses the random forest regression algorithm for training analysis. In addition, this paper optimizes the model results under different parameters by adjusting the number of decision trees from 100~1000, using MSE and R^2 as judging criteria and optimizing iterations with step size 100. The results are shown in Figure 3.

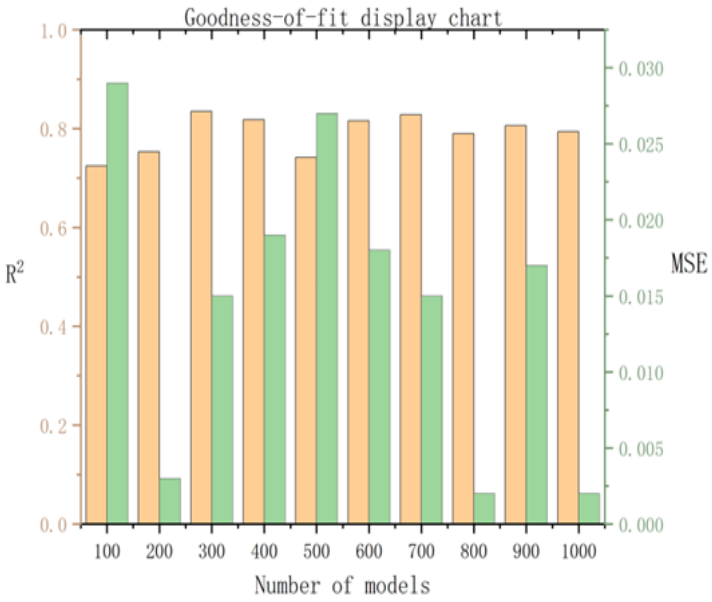


Figure 3 Double $\mathcal{Y}$-axis histogram

To ensure randomness of the decision tree generation, this paper chooses to disrupt and repartition the data for data shuffling. During the training, 70% of the total samples are involved in training, the maximum depth of the tree is set to 10, and the maximum number of leaf nodes is 50. The sampling taken with return, and the minimum number of samples of leaf nodes is 1. Based on the above training results (training set in the table and training set in learning) and the above test, the settings of the indicators are compared for the goodness of fit, and it is found that the best fit is achieved when the number of decision trees is 400, where.

$$R^2 = 0.818 \; MSE = 0.019$$

4.3 Gini Importance Analysis

Based on the training of 5.2, this paper selects the group with the highest goodness of fit from which the Gini index analysis is performed, i.e., the number of decision trees is 400. And when the random forest model is used for classification, this paper selects the node split evaluation quasi as Gini. The more important the feature is, the more obvious the classification result is. The amount of change in the Gini index, and the histogram of feature importance for the fitted optimal group is shown as follows.

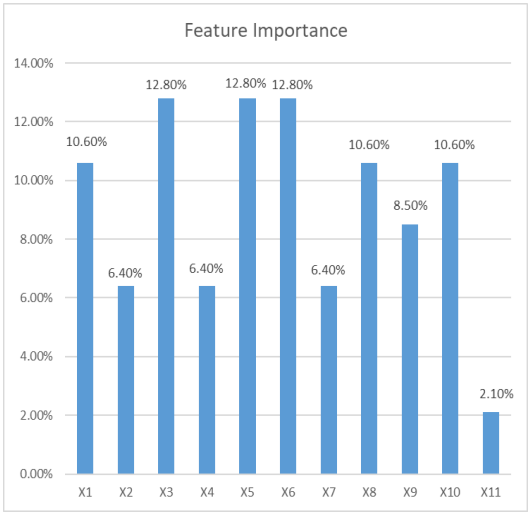


Figure 3 Feature importance histogram

As visualized by the bar chart, the exchange rate (100 USD/RMB), Engel's coefficient of residential households, and consumer price index on the increase in the amount of international trade of the dependent variable are the most important features which greatly affect the inverse of the growth of total exports and imports of goods.

a) Normally, the import of foreign products needs to use foreign currency in a foreign trade, and the way to obtain foreign currency is to use the national currency in exchange. Commodity trade follows the principle of equivalence, so is the currency. And the measure of equivalence is the exchange rate between countries. The exchange rate changes represent the appreciation and depreciation of the currency between countries. When a country's currency appreciates, The amount of import will increase and the amount of export will reduce, and vice versa, which has a direct impact on trade volume.

b) Engel's coefficient is the proportion of total food expenditure to total personal consumption expenditure. Engel's coefficient can reflect the level of affluence, which laterally reflects the ability to invest in consumption expenditure and will directly affect the amount of international trade.

c) The consumer price index, also known as the consumer price index, is an indicator of the change in the price of goods over a certain period and can reflect the degree of inflation over a certain period. When the index grows too fast, it means that prices are changing a lot, which will affect people's decisions about the daily consumption and thus the volume of international trade.

Therefore, it is reasonable that the three factors mentioned above have the smallest Gini coefficient and a large impact on the degree of international trade risk from an economic point of view.

When considering features that have less impact, this paper takes The SSE Composite Index as an example. The SSE Composite Index reflects the changes in the prices of stocks listed on the Shanghai Stock Exchange, and the changes in its value reflect the ups and downs of the stock market, which is characterized by instability, volatility, and speculation. For example, some bonds may have a high value in a very short term but can change dramatically in a very small period, so it is difficult to directly measure how the purchasing power of stock owners affects the amount of international trade.

Suggestions: To minimize the possible risks of international trade, the country needs to formulate scientific fiscal and monetary policies to make the inflation rate as stable as possible in a reasonable range, and be able to use relevant policies to deal with the risks caused by inflation. Also, the government is expected to pay attention to the money supply, interest rates, reserves and other means of influencing the currency, to ensure the maximum degree of currency value stability. Finally, the government should be committed to ensuring that the nation has the basic consumption capacity and demand, and to creating a good employment environment and a safe consumption environment for them.

5. Conclusion

The objective of this paper is to study the impact of international trade risk indicators. Firstly, this paper constructs a system of indicators influencing the economic risk of international trade, establishing the dependent and independent variables, i.e., twelve indicators such as the inverse of the increase in total trade and GNI. Secondly, for the optimal selection of feature indicators, this paper introduces the random forest model as a method of feature selection. The RF-Gini algorithm is constructed for feature selection of the evaluation indicators in the data, and the indicators are tested using R^2 and *MSE* to determine the number of models with the most accurate results. Finally, the data related to eleven indicators such as GNI, reserves, and exchange rate in recent years are collected and research on finding out the indicators with the greatest influence on the inverse of the increase in total trade and qualitative analysis is conducted. The conclusions are as follows.

(1) Exchange rate (100 USD/RMB), household Engel coefficient, and consumer price index have a high impact on international trade risk, and these three influencing factors need to be focused on when conducting international trade.

(2) For future international trade, the state should control the number of imports and exports to ensure the stability of the exchange rate. At the same time, a scientific and reasonable fiscal and monetary policy should be formulated to make the inflation rate as stable as possible in a reasonable range to ensure that the nation has the basic consumption capacity and the maximum possibility of promoting international trade with the risk of international trade reduced.

References

[1] Yang Shengli, Wang Weirong. The risk of international trade of agricultural products in the post-epidemic period and countermeasures to avoid it [J]. Price Monthly,2021(08):65-69.DOI:10.14076/j.issn.1006-2025.2021.08.09.

[2] Xu Huichao, Zhang Manrui, Lu Xiangyi, Zhao Yangyun. The impact of digitalization on international trade development and competitive landscape [J/OL]. Survey World:1-12[2022-07-26].DOI:10.13778/j.cnki.11-3705/c.2022.06.002.

[3] Zhou Wei. Reflections on the prevention of legal risks in international trade [J]. Business Times,2014(22):117-118.

[4] Wu JG, Wang TOS. On the risk evaluation of international trade[J]. Seeking, 2008(02):40-42. DOI:10.16059/j.cnki.cn43-1008/c.2008.02.074.

[5] Peng Ram. Common risks in international trade and their prevention [J]. International Trade Issues,2001(06):54-55+15.

[6] Xu Chen, Wang Jinzhao, Jin Zhida. China's strategic policy exploration in the process of transformation of the international trading system - based on the critical perspective of global value chains[J]. The Economist,2022(04):79-88.DOI:10.16158/j.cnki.51-1312/f.2022.04.011.

[7] Liu, Linqing, Yan, Xiaofei, Yang, Lis, Song, Min. Study on the evolution and endogenous mechanism of international trade dependence network[J]. China Industrial Economics,2021(02):98-116.DOI:10.19581/j.cnki.ciejournal.2021.02.015.

[8] Tian Yilin, Wu Fang. Reflections on promoting high-quality development of China-Africa trade - Analysis of China-Africa trade situation in 2018 and policy recommendations [J]. International Trade,2019(06):12-17.DOI:10.14114/j.cnki.itrade.20190730.004.

[9] Liu S. Y., Li J., Qin A. J.. The game of government discourse in the construction of international trade rules[J]. Journalism Spring and Autumn,2022(01):32-43.

[10] Zhao Jia. The "new" order of international trade? [J]. International Political Science,2018,3(02):64-96.DOI:10.16513/j.cnki.qjip.2018.0011.

[11] Chen Jingquan. A theoretical and empirical study on the relationship between international trade and inflation in China after the exchange rate reform[J]. Macroeconomic Research,2013(03):58-63+80.DOI:10.16304/j.cnki.11-3952/f.2013.03.011.

[12] Li, Nan. International trade transport:A comparative static analysis of basic framework and liberalization reform policies[J]. Business Research,2013(02):183-188.DOI:10.13902/j.cnki.syyj.2013.02.033.

[13] Lu Wei. Leveraging the Internet to innovate the international trade model [J]. People's Forum,2017(31):112-113.

[14] Ai Shangle. Analysis of the new pattern of China's foreign trade development in the context of "One Belt, One Road" [J]. Price Monthly,2021(11):58-63.DOI:10.14076/j.issn.1006-2025.2021.11.09.